2017

Framework for Performance Evaluation of Computer Security Incident Response Capabilities

Qutaiba Al Harfi Albluwi
University of Rhode Island, qutaiba.albluwi@gmail.com

FRAMEWORK FOR PERFORMANCE EVALUATION OF COMPUTER SECURITY INCIDENT RESPONSE CAPABILITIES

BY

QUTAIBA ALBLUWI

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMPUTER SCIENCE

UNIVERSITY OF RHODE ISLAND

2017

DOCTOR OF PHILOSOPHY DISSERTATION

OF

QUTAIBA ALBLUWI

APPROVED:

Dissertation Committee:

Major Professor	Victor Fay-Wolfe
	Lisa DiPippo
	Koray Ozpolat
	Nasser H. Zawia DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2017

ABSTRACT

Incident response (IR) is an integral part of today’s computer security infrastructure both at the public and private sectors. The process involves identification of the critical resources, proposing plans for responding to potential breaches and executing effective containment and recovery procedures. The current practices emphasize establishing careful response plans, building technical capabilities and following disciplined procedures for plan execution. This research builds on the above by adding another dimension to the process, namely performance evaluation.

Proposing a framework for the performance analysis of computer security incident response (CSIR) capabilities is the main focus of this research. The various design considerations and challenges to performance analysis of CSIR are investigated. A multidisciplinary survey is conducted to derive lessons learnt and best practices for the design of performance systems. The outcomes of the survey are integrated to the CSIR discipline to produce a development process for constructing performance evaluation models. For each development step, the various design possibilities are investigated to ensure flexibility and applicability to the wide spectrum of CSIR environments.

Expert feedback is used as a method of design validation to ensure conformance to current CSIR best practices. Issues pertaining to how performance evaluation could be incorporated into the current industry practices are also explored. As a notable contribution, the study produces the definition and design considerations for fifty performance indicators that cover the diverse performance aspects of computer security incident response systems.

ACKNOWLEDGMENTS

I am grateful to God, at the beginnings and at the ends, for the boundless Mercy and Care that He bestowed upon me throughout the stages of this project. I submit in love for His limitless Gentleness and His countless bounties.

My sincere thanks to the guidance and support of my supervisor, Dr. Victor Fay-Wolfe. He provided me a space to grow and provisioned me with his genuine trust. His dedication to the field of computer security was inspiring. I also express my sincere thanks to my internal committee member, Dr. Lisa DiPippo, whom I have tremendously learnt from her approach to the study and research of computer science. It was always nurturing to learn and work with her. Special thanks to Dr. Koray Ozpolat, the external committee member, and Dr. Stu Westin, the defense chair. Their contributions in improving this work were valuable.

My special lovely thanks to my mother: Dr. Abla Elhersh and my father: Dr. Salameh Albluwi. They both poured me with their tremendous support and love. I will neither forget my mother's words of encouragement and peace-spirited chats, nor my father's powerful and passionate words of motivation. The splendid support that my wife, Huda Abazeed, had granted me should be remarked. This work would not have been possible without her memorable standing by my side throughout these years. It will also be remembered that my two young sons, Ridwan and Muhamad, grew up accompanying the development of this dissertation. They were the joy that pushed away any stress.

I should also record the numerous fruitful discussions I had with my brother, Dr. Ibrahim Albluwi, which opened doors of questions and brought different perspectives to this work. Extended thanks to my twin sisters: Dr. Ghada and Dr. Najlaa and my brother:

Eng. Tariq for their remote support, prayers, and patience for being away for the past five years.

I would also like to highlight the appreciated support that I received from Dr. Nasser Zawia, the Dean of the Graduate School; Dr. Joan Peckham the chair of the Department of Computer Science and Statistics; and Mr. Melvin Wade, the Director of the Multicultural Center. Their personal care and advice was essential to any success achieved during my studies at the University of Rhode Island.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS v

LIST OF TABLES xi

LIST OF FIGURES xiii

LIST OF ABBREVIATIONS xv

CHAPTER ONE 1

INTRODUCTION 1

1.1 Computer Security Incident Response (CSIR) 1

1.2 The Need for Studying CSIR Performance Evaluation 4

1.2.1 Academic Need 5

1.2.2 Business Need and Impact 7

1.3 Problem Description and Methodology 10

1.3.1 Problem Description and Research Objectives 10

1.3.2 Research Methodology 11

1.3.3 Assessment Methods 14

1.4 Definitions 15

1.4.1 Computer Security Incident Response Terms 16

1.4.2 Performance Evaluation Terms 20

CHAPTER TWO 26

BACKGROUND 26

2.1 Computer Security Incident Response (CSIR) Literature Review 27

2.1.1 History 27

2.1.2 CSIR Models and Processes 35

2.1.3 Computer Security Incident Response Team (CSIRT) 42
2.1.4 Computer Security Incident Response Plan (CSIRP) 45
2.1.5 CSIR Performance Evaluation 49
2.1.6 Performance Evaluation of Computer Security Systems 59
2.1.7 Measuring Security in Computer Systems 67
2.2 Multidisciplinary Literature Review of IR Performance Evaluation 74
2.2.1 Challenges 75
2.2.2 Methods and Practices 83
2.2.3 Performance Indicators and Metrics 89
2.3 Discussion and Lessons Learnt 98
2.3.1 State of the Art of PE in CSIR 98
2.3.2 Lessons Learnt from Other Fields 102
CHAPTER THREE 109
FRAMEWORK DESIGN 109
3.1 Framework Overview 110
3.1.1 Framework Components 110
3.1.2 Framework Development Process 112
3.2 Assumptions 114
A1.1. Civilian Environment 116
A1.2. Team Structure 116
A1.3. Incident Complexity 117
A1.4. Incident Handling Services 117
A1.5. Computer Security Incidents 118
A1.6. CSIR Capability 118
A1.7. Secure Environment 119
3.3 Phase I: Designing a CSIRPE Framework 119
3.3.1 Setting Design Parameters 119
D.1. CSIRT Type 121
D.2. Evaluator Type 122

D.3. Number of Incidents 124

D.4. Incident Concurrency 127

D.5. Analysis Time 128

D.6. Benchmarking 128

D.7. Measurement Type 130

D.8. CSIR Scope 133

3.3.2 Framework Assumptions 134

A2.1. CSIR Model 135

A2.2. Member Dedication 135

A2.3. Central CSIRT 135

A2.4. Internal CSIRT 136

A2.5. Sequential Handling 136

3.3.3 Laying Out Strategies and Policies 136

S.1. CSIR Quality Assurance 138

S.2. CSIRPE Quality Control 139

S.3. CSIRPE Complexity 141

S.4. CSIR Unpredictability 142

S.5. CSIRPE Overhead 143

S.6. Reference Analysis Point 145

3.4 Phase II: Defining a CSIRPE Model 148

3.4.1 Overview 148

3.4.2 Defining Performance Goals 149

3.4.3 Identifying PE Aspects 154

3.4.4 Defining Performance Indicators 159

3.5 Phase III: Measuring CSIR Performance 162

3.5.1 PE Metrics Derivation Process 163

M.1. Metric Type 166

M.2. Performance Metric Scope 166

M.3. Quantifiability 167

M.4. Measurement Tool 168

M.5. Performance Indicators 169

M.6. Dependence 170

M.7. Accuracy 170

M.8. Conditions 171

M.9. Attributes 172

M.10. Interpretation 172

3.5.2 PE Analysis Models 172

N.1. Gap Analysis 175

N.2. Benchmarking 176

N.3. Targeted Analysis 176

N.4. Bottleneck Analysis 177

N.5. Root-Cause Analysis 178

N.6. Goal Achievement 179

N.7. Trend Analysis 181

N.8. Stakeholder Analysis 182

N.9. Predictive Analysis 183

3.5.3 PE Validation Models 184

V.1. Operational Validation (Feedback Systems) 186

V.2. Development Validation (Compliance) 187

V.3. Development Validation (Bootstrapping) 188

V.4. Heuristic Validation (Confidence Rating) 189

3.6 Phase IV: Implementing a CSIRPE Model 190

3.6.1 PE Functional Models 190

3.6.2 Assigning Roles and Responsibilities 194

3.6.3 Integrating CSIRPE into CSIRP 197

CHAPTER FOUR 200

PERFORMANCE INDICATORS 200

4.1 Classifying Performance Indicators 201

4.2 Selecting Performance Indicators 203

4.3 CSIR Performance Indicators 206

CHAPTER FIVE 249

FRAMEWORK SUBCOMPONENTS...249

5.1 The Universal PE Framework (CSIR-UPEF)..249

5.2 The Integrated Analysis & Validation Model (IAV)..254

5.3 The CSIR Balanced Scorecard Model 259

5.4 The SAC Complexity Model 261

5.5 The NFP Unpredictability Platform..266

CHAPTER SIX...271

FRAMEWORK ANALYSIS...271

5.1 Scenario Analysis..271

5.1.1 Scenario I: Confronting Incident Instability 271

5.1.2 Scenario II: Enhancing Incident Declaration Strategy 278

5.1.3 Scenario III: Analyzing Response Time... 285

5.2 Expert Feedback..298

5.2.1 Methodology 298

5.2.2 Expert 1 Feedback 300

5.2.3 Expert 2 Feedback 303

5.2.4 Expert 3 Feedback 307

5.2.5 Expert 4 Feedback 310

5.2.6 Summary of Expert Feedback 314

5.3 Evaluating Framework Features 316

5.3.1 Comprehensiveness.. 316

5.3.2 Flexibility 318

5.3.3 Compatibility.. 319

CHAPTER SEVEN 323

CONCLUSION...323

6.1 Summary of Contributions...323

6.2 Summary of Findings 325

6.3 Future Work 328

APPENDICES 332

Appendix A: Practical Guidelines 332

A.1. Recommended CSIRPE Configuration 332

A.2. Guidelines for Designing Adjustable-Window Framework 333

A.3. Guidelines for Selecting Analysis Reference Point 334

A.4. Guidelines for Identifying CSIR Performance Aspects 335

A.5. Templates for Defining Performance Indicators and Metrics 336

A.6. Basic IAV Model Design 337

A.7. Guidelines for Selection of Performance Analysis Methods 338

A.8. Scenario II Data 338

BIBLIOGRAPHY 342

LIST OF TABLES

Table 1: Summary of Definitions 15

Table 2: Main activities in the IR life cycle of the Hybrid Model.................................. 38

Table 3: Incident Classification Mechanism proposed by the NIST document 42

Table 4: Outline of Action-Based CSIRP Sample.. 47

Table 5: Contents of CSIRP as outlined by [3]... 48

Table 6: VERIS Project Incident Narrative Descriptors.. 53

Table 7: Model for Assessing Threats proposed by [22]... 58

Table 8: Format for describing performance indicators in the field of EMS................... 91

Table 9: Summary of Factors with high impact on CSIR performance evaluation........ 108

Table 10: Framework Components Codes... 114

Table 11: Global CSIR Assumptions... 115

Table 12: Summary of CSIRT services outlined in the CERT document 118

Table 13: Framework Component 1: Design Parameters... 120

Table 14: Comparison between evaluator types design options.................................... 123

Table 15: Comparison between single and multiple-incident PE frameworks.............. 126

Table 16: Comparison between sequential and concurrent PE models 128

Table 17: Comparison between qualitative and quantitative measures 131

Table 18: Global Framework Assumptions 134

Table 19: CSIRPE Issues that require strategic design.. 137

Table 20: Comparison between methods of CSIRPE quality control............................ 141

Table 21: The three principles of the NFP Unpredictability Platform........................... 143

Table 22: CSIRT vs. CSIRP Analysis Reference Points 146

Table 23: SMART Model for Defining CSIRPE Goals 154

Table 24: Methodologies for Identifying PE Aspects.. 155

Table 25: Examples for the Methodologies for Identifying PE Aspects 157

Table 26: Description of the Steps for Defining Performance Indicators 162

Table 27: Design Parameters for Performance Metrics... 164

Table 28: Features of good CSIR performance metrics... 165

Table 29: PE Analysis Methods.. 173

Table 30: Validation Models for CSIRPE Frameworks 186

Table 31: Types of PE Functional Models 190

Table 32: Desired Qualifications of the Performance Officer (CSIRPO) 197
Table 33: Checklist for CSIRPE integration into a CSIRP 199
Table 34: Features of the Universal PE Framework (UPEF) 251
Table 35: Performance Indicators for the CSIR Balanced Scorecard Stakeholders....... 261
Table 36: Flexibility and Preparedness in the NFP Unpredictability Platform 270
Table 37: Scenario I: Definition of the Stability PI 274
Table 38: Scenario I: Definition of Incident Escalation Performance Metric 276
Table 39: Scenario I:Definition of Incident Fluctuation Performance Metric............... 277
Table 40: Scenario II: List of Performance Indicators and Metrics............................... 280
Table 41: Scenario II: Definition of Incident Classification Accuracy 281
Table 42: Scenario II: Definition of the TCIC Performance Metric.............................. 282
Table 43: List of Performance Metrics for Scenario III 285
Table 44: Expert Feedback Survey Themes 300
Table 45: Recommended CSIRPE configuration based on maturity level.................... 333
Table 46: Guidelines for designing an inclusive PE framework 334
Table 47: Assessment questionnaire for selecting analysis reference point strategy 335
Table 48: Recommendations for Identifying PE Aspects.. 335
Table 49: Template for Performance Indicators Definition... 336
Table 50: Template for Formal Definition of Performance Metrics.............................. 336
Table 51: Guidelines for Selecting Performance Analysis Techniques.......................... 338
Table 52: Scenario II: List of Descriptive Metrics 339
Table 53: Scenario II: Data for Annual Number of Incidents (ANI)PM........................ 339
Table 54: Scenario II: Incident Response Fixed Costs(USD) [2012-2016] 339
Table 55: Scenario II: Average Incident Response Operation Costs (USD).................. 339
Table 56: Scenario II: CSIRT Cost Effectiveness Data... 340
Table 57: Scenario II: Average Incident Response Time 340
Table 58: Scenario II: Detection Inaccuracy Cost Performance Metric Data 340
Table 59: Scenario II: Incident Classification Strategy Cost Effectiveness PM Data 341

LIST OF FIGURES

Figure 1: Taxonomy of Performance and Quality Terminologies used in the project 22
Figure 2: Number of CSIRTs worldwide in the period of 1990-2016............................ 31
Figure 3: The three main CSIR publications 32
Figure 4: Timeline of Computer Security Incident Response (CSIRT) development...... 34
Figure 5: Models for Incident Response Life Cycle 37
Figure 6: Main activities in the Identification Phase of the IR cycle............................... 40
Figure 7: Organizational Structure of a CSIRT 44
Figure 8: CSIR Performance evaluation framework Components 111
Figure 9: Performance Evaluation Framework Components Relationships.................. 113
Figure 10: Phase II: Defining Performance Evaluation Model 148
Figure 11: Example of CSIRPE Usefulness Test 151
Figure 12: Example of Combining two methodologies to define PE Aspects 158
Figure 13: Defining Performance Indicators Process 159
Figure 14: Transition from Phase II to Phase III: Measuring CSIR Performance.......... 163
Figure 15: DCA Functional Model 191
Figure 16: PE Monitor Functional Model.. 192
Figure 17: Incremental PE Functional Model with a Single-Mark................................ 194
Figure 18: CSIR Performance Officer Responsibilities... 195
Figure 19: Collaboration Chart for CSIRPO and other officers 196
Figure 20: Higher-level Design of the IVA Model.. 255
Figure 21: Detailed Design of the Integrated Analysis & Validation Model (IVA) 257
Figure 22: The CSIR Balanced ScoreCard Model... 260
Figure 23: Sample of Simplification of the Identification Phase................................... 263
Figure 24: Depiction of the four types of ordered analysis for SCM. Source: [108]. 265
Figure 25: NDDM Component of the NFP Unpredictability Platform 267
Figure 26: Scenario I: Selection of Performance Indicators.. 274
Figure 27: Scenario III: Response Time Components... 288
Figure 28: Scenario III: Detection Time (DT) Components.. 289
Figure 29: Calculation of Incident Detection (DT) Time 290
Figure 30: Scenario III: Incident Handling Time (HT) Components 291

Figure 31: Definition of Incident Containment Factor (ICF) Performance Metric 295

Figure 32: Definition of Weighted Containment Factor Performance Metric................ 296

Figure 33: Basic Design of the IAV model 337

LIST OF ABBREVIATIONS

AHI	Attacker Host Identification
BSD	Berkeley Software Distribution
CEO	Chief Executive Officer
CERT	Computer Emergency Response Team
CFO	Chief Financial Officer
CIA	Confidentiality, Integrity, Availability
CIR	Computer Incident Response
CLR	Competency Lifecycle Roadmap
CSIR	Computer Security Incident Response
CSIRP	Computer Security Incident Response Plan
CSIRPE	Computer Security Incident Response Performance Evaluation
CSIRPO	Computer Security Incident Response Performance Officer
CSIRT	Computer Security Incident Response Team
CVSS	Common Vulnerability Scoring System
DARPA	Defense Advanced Research Project Agency
DCA	Design, Collect, Analyze
DoS	Denial of Service
DDoS	Distributed Denial of Service
EMS	Emergency Medical System
FIRST	Forum for Incident Response and Security Teams
GQIM	Goals, Questions, Indicators, Metrics
IAV	Integrated Analysis and Validation Model
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
IR	Incident Response
IRP	Incident Response Plan
IRPE	Incident Response Performance Evaluation
IRT	Incident Response Team
IT	Information Technology
ITIL	Information Technology Infrastructure Library

KPI	Key Performance Indicator
MANET	Mobile Ad hoc Network
MR	Military Response
NDDM	Non Deterministic Decision Making
NHTSA	National Highway Traffic Safety Administration
NIMS	National Incident Management System
NIST	National Institute of Standards and Technology
NPPO	Nuclear Power Plant Operating
NRF	National Response Framework
PE	Performance Evaluation
PED	Performance Evaluation Database
PI	Performance Indicator
PLO	Programmable Logic Controller
PM	Performance Metric
QoS	Quality of Service
RCS	Relevance, Comparability and Simplicity
SAC	Simplification, Approximation, Cascading
SCM	Supply Chain Management
SET	Stress Exposure Training
SLA	Service Level Agreement
UPEF	Universal Performance Evaluation Framework
VERIS	Vocabulary for Event Recording and Incident Sharing

CHAPTER ONE

INTRODUCTION

1.1 Computer Security Incident Response (CSIR)

Computer Security Incident response (CSIR) is an integral part of today's computer security infrastructure, both at the public and private sectors. The term CSIR refers to the collection of organized activities that address the aftermath of computer security incidents. The CSIR process involves identification of the critical resources that are vulnerable, implementing plans for responding to potential breaches, and executing effective containment and recovery procedures.

The rapid increase of security breaches, the complexity of attacking techniques, the large scope of classified data and critical resources, and the high impact of security incidents on the operational and fiscal resources of business organizations and government units, are examples of factors that contributed to the development of CSIR. These factors triggered the need for systemized and effective methods of handling computer security incident. The computer security community collaboratively developed the incident handling procedures that are known today by CSIR. These procedures were later matured into the form of incident response planning.

There are three main objectives for introducing a CSIR capability to an organization. The first is to quickly detect and properly diagnose a security incident. Second, a CSIR capability helps in minimizing the associated damage of incidents through containment and mitigation procedures. And third, CSIR brings forward effective eradication and recovery procedures from cyber incidents.

Although incident response capabilities vary in their approaches, strategies and policies, five major components could be identified as common between all of them [1] [2] [3]. The five components are: Preparation, Identification, Containment, Eradication and Review.

The Preparation phase defines the goals, identifies what needs to be protected, and sets up policies and procedures of how to handle an incident from the moment it is detected to the moment of full recovery. A notable activity in this step is forming the incident response teams and defining their roles. The objective of this step is to ensure that when an incident erupts, all parties are engaged in handling it, not in figuring out what needs to be done.

The Identification phase refers to the set of mechanisms and procedures to follow for detecting an incident, determining its severity and declaring it.

The Containment phase is concerned with isolating the incident by preventing it from spreading outside its scope, ensuring that the incident doesn't escalate and minimizing the associated damage.

The Eradication phase involves technical procedures to remove the threat from the environment, provide remedies for the caused damage and recover the environment to its normal state.

The final phase: Review identifies potential areas of improvement for future incidents in the forms of best practices and lessons learned.

To demonstrate the disadvantage of not having a CSIR capability, a recent global survey sponsored by IBM [4] found that the longer it took to detect and respond to computer security incidents the higher the loss endured. It was also recorded that having a

CSIR capability decreased the loss of data breach from $158 to $142, i.e. $16 per record. Taking the statistics that the US health sector alone was targeted by cybersecurity incidents against over 112 million record in 2015 [5]; the use of CSIR would have saved 1.792 billion dollars for the health industry alone. Specific incidents can even cost much higher loss, for example the 2013 Target credit card breach is estimated to have cost $200 million [6].

The current CSIR practices emphasize careful design, technical capabilities, and successful execution of incident response plans. This research builds on the work that has been done in the CSIR literature by adding a fifth essential element to the process, namely *performance evaluation*. Performance evaluation refers to the collection of methods and tools used to measure *how good* a specific activity contributed to achieving a goal. Within the context of CSIR, performance evaluation focuses on measuring the effectiveness of designing and implementing a CSIR plan (CSIRP).

This project analyzes the concept of CSIR performance evaluation by inspecting why such evaluation is needed and how it could be achieved. A framework for assessing the effectiveness of a CSIR capability is presented. The framework constitutes of design parameters, development processes, measurement tools, analysis techniques and modeling of various aspects of CSIR performance.

This dissertation is structured in seven chapters. The first chapter provides an introduction to the topic by investigating the need for evaluating CSIR, outlining the research methodology and defining the major terms. The second chapter surveys the relevant publications and is remarked by the multi-disciplinary approach of synthesizing findings of other incident response disciplines into CSIR.

Chapters three to five are centered around the development of the performance evaluation framework. Chapter three provides detailed description of the framework components and parameters. It also highlights the process of building a performance evaluation model for various environments. The main measurement tools used in the evaluation process, called performance indicators, are discussed in chapter four. While chapter five presents five performance evaluation sub-models that can be viewed as extensions to the performance framework presented in chapter three.

The sixth chapter analyzes the proposed PE framework through scenario analysis and expert feedback. Finally, the seventh chapter summarizes the main findings and outlines directions for future research.

1.2 The Need for Studying CSIR Performance Evaluation

In the past fifteen years, the area of computer security incident response (CSIRP) has received substantial attention from the academic and industrial fronts. In academia, researchers in the area of computer security focused on two major areas; the development of incident response models and mechanisms, and proposing technical solutions for the identification and containment of various cyber security attacks. On the industrial front, the attention of security professionals was directed towards three main aspects. The first is drafting and writing computer security incident response plans (CSIRPs) that coordinate incident management between various technical and non-technical teams involved in the process of incident handling. The second is developing industry standards for handling the response to computer security incidents. And finally proposing technical advisories and procedures on handling security incidents within specific environment constraints, or vendor specific configurations. This technical information is normally shared in the form of blog posts and technical reports, normally labeled by the term *white papers*.

The study and implementation of performance evaluation measures within the context of CSIRP is expected to have contributions to both the academic and professional sectors. It mainly builds on the previous contributions and push towards more scientifically sound and cost effective handling of computer security incidents. The following subsections provide insights on how the focus of this dissertation on performance evaluation could enhance the understanding and handling of CSIR.

1.2.1 Academic Need

The vast majority of publications in the area of Computer Security Incident Response (CSIR), have no detailed or technical description of the issue of performance evaluation.. It is only the past decade that the issue of CSIR performance started to get more attention. Nevertheless, there is a distinguishable need for systemized and detailed approaches towards how to conduct such performance evaluation and which measurement methods should be used. In addition, discussing performance is normally presented under the generalities of conducting a post-incident lessons learnt analysis.

The framework proposed in this dissertation aims at providing depth to the aspect of performance evaluation in the context of computer security incident response, that transcends the generalities proposed by previous works. This project will address design considerations, development processes, performance indicators and a variety of implementation considerations pertaining to CSIR.

There are three possible reasons for why the study of performance evaluation in the area of CSIR did not get much academic attention.

First, the study of computer incident response, in its current developed form of being systemized and on the scope of large organizations, is a relatively a new area of study. The previous studies focused on designing models of handling responses, articulating the specifics of each phase, and developing technical solutions to the common cyber-attacks. Since this dedication has produced works that build the foundations within the area of study, it is natural to evolve into studying methods of enhancement. It could be noticed that performance evaluation is perceived as enhancement activity instead of a core activity. Therefore, based on the maturity of the field, discussion of performance evaluation could have been only possible when the field has stabilized in terms of its procedures and processes.

Second, there is an overall wide gap between the contributions of academia and the industry, as the latter is ahead in its complexity and depth of study. This is not surprising as the operational nature of the field dictates that sometimes pseudo solutions or imperfect but working solutions, which might not be accepted by the academic terms, be proposed by the industry for immediate countering to the breaches. However, this has led to an unfortunate outcome of the results being mainly preserved within the classified business documents, which are not normally circulated to academia or shared with the public. There is reasonable reckoning that the industry might have already developed some of its performance evaluation methods, but the locality of these documents remains an obstacle to the growth of the field. However, as stated by a CSIR responder [7], current evaluation methods used in the industry are: “semi-formal, ad hoc and conducted by smart people”.

Third, as will be recorded in the literature review, the study of performance evaluation of incident response is a complex process that depends on a vast number of variables and is usually difficult to normalize. This is a lesson learnt from studying the findings of other disciplines that focus on incident response, like medical emergency. Although, the reasons for reaching such conclusion is different from one field to another, there are several shared challenges across the fields that extend to the area of computer security.

Observing that there is an academic gap between acknowledging the need for conducting performance evaluation and the lack of models analysis was one of the motivations of this dissertation. The study aims at proposing a comprehensive framework that acts as a foundation for the study of performance evaluation the area of computer security incident response. The framework builds on the previous findings in the field, and has the flexibility and extendibility to capture the advancements in the field.

Finally, the multidisciplinary nature of the study is another factor that brings academic significance to the proposed work. The survey of main findings and good practices across fields should be of interest to both security professional who do not need to re-invent the wheel, and to researchers who either study performance evaluation or study incident response across disciplines.

1.2.2 Business Need and Impact

A 2014 survey by SANS found that a quarter of practitioners feel that their organization’s CSIR capability is ineffective and only 9% reported that they felt it was very effective [8]. This perception of ineffectiveness captured by the CSIR technical practitioners is coupled by several concerns highlighted by business owners and administrators [9] that confirm the need for investigating the effectiveness of CSIR.

As shall be demonstrated in Chapter two, the need for measuring performance for incident response is acknowledged across various disciplines. Performance measurement triggers change in performance, or as summarized by [10] [11] it helps to control, budget, motivate, promote, learn and improve the response system

The following points summarize the need for incorporating performance evaluation into incident response handling, from the business standpoint:

[1] *Need for reducing costs:* the cost of handling security incidents is on the rise and the industry is dedicating millions of dollars for handling and containing computer breaches. Despite the observation that the financial implications of not securing the systems of an organization are far more than the cost of imposing security measures [12], it is still crucial for an organization to ensure that the security plans it adopts and implements are cost effective. With the use of performance metrics, an organization can measure the cost effectiveness of its plans and identify areas where cost could be reduced without jeopardizing the security of the system.

[2] *Need for faster responses*: with the high dynamicity of today's IT environments, the time factor plays an important role in responding to security threats. It is common to see threats escalate or even reach an uncontrollable state due to slowness or ineffectiveness of handling the incident. Performance evaluation will enable responders to identify areas where they could improve the effectiveness of the response and propose enhancements based on results of systemized and validated analysis techniques. It is true that performance collection, measurement and analysis occupies time; however, it "ends up saving more time than it consumes" [13].

[3] *Enhancing Decision Making:* findings of performance evaluation provide incident responders and executives with data that empowers them to make effective and accurate decisions [9]. Without using data produced by well-designed performance metrics, most recommendations and reviews are based on speculations and not scientific approaches which might lead to counter effects.

[4] *Need for information sharing:* using performance metrics will enable organizations to build a common language between security professionals and business administrators. At the same time, it helps an organization evaluate its performance compared to other organizations through benchmarking.

[5] *Liability Needs:* Another aspect of industrial needs comes from the trending demand for providing insurance and liability over information security [14] [15]. Insurance companies are demanding objective assessment and quantifiable metrics for enforced security policies and procedures. Having systemized approaches to the evaluation of CSIR capabilities is viewed as an essential component of the liability and insurance processes [7].

[6] *Compliance:* The NIST document [1] considers performance evaluation as one of eight policies that need to be considered when designing and preparing incident response plans. Similarly, the CERT document [2] discusses under the section of quality assurance the need for "checks: measurement of quality parameters". However, both documents do not provide outlines of how to design and implement the measurement process. This calls for a systemized and formalized approach of conducting performance measurement and analysis in order to demonstrate compliance with the above two industry standards.

As this project is finalized, the Forum for Incident Response Security Teams (FIRST) [16] announced its establishment of a special interest group (SIG) for developing metrics for computer security incident response. This demonstrates that not only the industry approves the need for performance evaluation, but also that the industry is taking actual actions towards that direction.

1.3 Problem Description and Methodology

1.3.1 Problem Description and Research Objectives

The research problem of this project is centered around the theme of evaluating the performance of computer security incident response capabilities. More specifically, the project will investigate how to build performance evaluation models for CSIR and how these models could be implemented. This is achieved through presenting a framework for developing CSIR performance evaluation models which is equipped with a variety of performance evaluation analysis and measurement methods adapted from current industry practices and from lessons learnt from other incident response disciplines.

The problem could be more formally described as: *"given a computer security incident response (CSIR) capability, how could an organization construct and apply a mechanism for assessing the performance effectiveness of the capability."*

To address the above research problem, this dissertations attempts to achieve the following objectives:

1. Demonstrate the importance of incorporating performance evaluation in computer security incident response planning and explain how that fits within the CSIR life cycle.

2. Conduct a multi-disciplinary survey of the good practices and major findings in the area of incident response. The results are to be analyzed in reference to computer security incident response for relevance and potential customization.
3. Design a framework for the performance evaluation of incident response capabilities. The framework will provide guidance on identifying what needs to be measured, how to obtain those measurements, and how to analyze the obtained results. The framework will be having the following four features:
    a. *Comprehensive:* The framework captures the performance of the major activities in the CSIR life cycle and provides end-to-end development process.
    b. *Flexible:* The framework is applicable to various environments and can be used under different incident conditions
    c. *Compatible:* The framework is consistent with current industry practices of incident handling
    d. *Multidisciplinary:* the framework incorporates lessons learnt from other IR disciplines.
4. Identify the measurement tools of performance evaluation in the form of performance indicators
5. Provide operational guidelines for how performance evaluation models could be implemented in CSIR environments.

1.3.2 Research Methodology

This project uses four main research methodologies for the design and validation of the proposed performance evaluation framework. The four methods are:

multidisciplinary analysis, system analysis, industry-driven design and expert feedback. These methods were selected to suit the nature of the research problem and to match the approaches used by other researchers who produced publications tackling research problems that resemble the problem statement of this project. A description of the four methodologies is provided below:

Multidisciplinary Analysis: A multidisciplinary survey of incident response in other disciplines, like medical emergency and homeland security, is conducted. The motivation of conducting this survey is to inspect how the other disciplines designed performance platforms for their incident response system. The main findings and best practices suggested by these disciplines are analyzed for their compatibility with the field of CSIR. Next, the relevant findings are integrated as design tools in the construction of a framework for evaluating the performance of CSIR.

System Analysis: The major processes and procedures of computer security incident response system are analyzed. The objective of the analysis is to deconstruct the CSIR system into components and parameters from the performance evaluation perspective. The expected output is identifying which aspects/components of the CSIR system should be subject to performance evaluation and specify which measurement tools should be used for each component.

Industry-Driven Design: The design of the performance framework is to be developed within the context of the best practices of the CSIR processes and procedures. This is achieved by ensuring that all components of the performance framework are in compliance with the two most common and acknowledged documentations of CSIR, namely the NIST publication "Computer Security Incident Handing Guide" (August 2012)

[1] and the CERT publication: “Handbook for Computer Security Incident Response Teams (CSIRTs) (April 2003) [2].

Expert Feedback: A group of computer security incident response and cyber security experts are consulted for their intake into the design considerations and for their feedback on the proposed framework. The feedback, which is obtained through personal interviews, are integrated into the design of the framework or recorded for potential enhancements.

The above research methodology can be summarized as argumentative design supported by expert feedback. Argumentative design refers to proposing various components of the framework associated with supported arguments. In the case of this project, the argument comes from the multidisciplinary survey results, system analysis of the CSIR or expert opinion.

It is noticed that the above research method does not follow the conventional computer science research approach of proposing a model and verifying it through mathematical proofs, algorithms design, software simulation or empirical data. Instead, a reverse approach is used in which system and multidisciplinary analysis is conducted beforehand to derive the design of the framework. This explains the length of the second and third chapter of this dissertation due to the extensive consultation with the literature. However, To ensure the effectiveness of the proposed design, the proposed design is further verified through expert feedback and hypothetical scenario analysis.

Several works in CSIR and related disciplines follow research methodologies similar to the ones adopted in this project. For instance, a collaborative incident response framework is proposed in [17]. The authors designed the framework by analyzing the needs and lessons learnt and is verified through scenario analysis to demonstrate enhancements resulting from using the framework. A similar methodology is also used in [18] which proposed a framework for incident information management. Another publication by CERT [19] proposed a mission-risk diagnostic platform through conducting a multidisciplinary survey to derive best practices that drive the platform design. Other works include [20] that uses interviews and literature survey for development of performance systems, [21] that uses hypothetical scenario analysis for analyzing performance frameworks, and [22] [23] that use expert feedback as an assessment for incident response performance evaluation systems.

1.3.3 Assessment Methods

The proposed performance framework will be evaluated through three main methods:

Scenario Analysis: hypothetical scenarios will be used to demonstrate the functionality and impact of using the proposed performance framework. The scenarios are designed to cover various aspects of the framework applied to a variety of performance issues.

Expert Feedback: Besides expert feedback being integrated into the design and implementation of the proposed framework, expert opinion on various components of the framework will be recorded and discussed.

Framework Feature Evaluation: The four features of the framework, i.e. comprehensiveness, flexibility, compatibility and multidisciplinary feedback, will be evaluated. A demonstration of how the framework satisfies the above features will be presented.

1.4 Definitions

A summary of the definitions of the main terms used in this project is provided in Table 1.

Term	Definition
Incident	“an occurrence of a major event or several events, that demonstrates actual or potential compromise of the system or data, and demonstrates high impact on the organization operations and resources, or the need for an organized response”
Computer Security Incident Response	“the capability to provide end-to-end, cross enterprise management of incidents that affect information and technology assets within an organization”
Computer Security Incident Response Plan	“the documentation of a predetermined set of instructions or procedures to detect, respond to, and limit consequences of a malicious cyber-attacks against an organization’s information system(s).”
Computer Security Incident Response Team	“Group of individuals usually consisting of Security Analysts organized to develop, recommend, and coordinate immediate mitigation actions for containment, eradication, and recovery resulting from computer security incidents.”
CSIR Performance Evaluation	“All activities associated with assessing how well a CSIRT executes CSIR activities as outlined in the CSIRP to achieve the outlined goals”
Performance Framework	“a generic conceptual setting that encompasses several perspectives and methods for the evaluation of computer security incident response capability”
Performance Indicator	“a system quality that reflects some performance behavior”
Performance Metric	‘a measurement tool used for characterizing or quantifying a quality of CSIR performance’. OR “a tool used for measuring a single or multiple performance indicators’
CSIR Effectiveness	“the extent to which CSIR goals are achieved and harm is reduced through efficient use of resources”

Table 1: Summary of Definitions

The basis of the provided definitions comes from the glossary prepared by the Committee on National Security Systems (CNSS) [24]. The subsequent discussion expands on the various available definitions and the rationale for adopting the selected definitions.

1.4.1 Computer Security Incident Response Terms

Events and Incidents

In order to provide organizational support that is both effective and budget-sensitive, the CSIR capability is not expected to respond to all security alerts. For instance, an anti-virus alert or an intrusion detection system (IDS) flag are considered simple activities that do not necessary require the response of a CSIRT. These simple activities are referred to in the literature as *events*. The formal definition is: "any observable occurrence in a network or system" [24]. However, in order to eliminate positive or expected events, like high access to a commercial website during a promotion period, the CSIR field narrows down the term into events with *adverse* nature, i.e. might result in some negative consequences [1].

There seems to be an agreement in the security community that incidents are distinguished from events, where incidents are higher than events, such that an incident constitutes several events or some major event. However, it remains vague about when an incident can be actually identified, leading to various understandings for the term. Several works have provided definitions for the term, like: [24] [1] [25] .Other works like the CERT document [2] has left it to each CSIRT to provide its own definition that fits its operating environment.

The ISO 17799 defines an incident as: "an occurrence that compromises information security" [25]. To provide more precise context for the term "information security”, The NIST [9] standard expands the definition to “is a violation or imminent threat of violation of computer security policies, acceptable use policies, or standard security practices”. The CNSS [24] expands further as “An occurrence that results in actual or potential jeopardy to the confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies. “.

The common theme in the above definitions is the highlighting of the “security” nature of the incident, perhaps to draw distinction from the term: “computer incident” which is more generic [26]. For example, database inconsistency resulting from sever migration is a major computer incident that disrupts the operation of an organization, but it is not of security nature.

I suggest amending the definition of an incident to reflect that complexity nature of security incidents. There are many incidents that satisfy the above definition but are simple to fix and do not require the execution of a CSIRP or the full potential response of a CSIRT. Although this is not explicitly mentioned in the above definitions, it is implied when the above references talk about the organized nature of the CSIRT activities. In that direction, a distinction is to be drawn between simple incidents and major incidents. A major incident is one that either require a coordinated response between various parties, or one that have or potentially have high impact on the organization’s operations or its resources, e.g. budget.

Based on the above discussion, the following definition is proposed for the context of this project: an incident is: “an occurrence of a major event or several events, that demonstrates actual or potential compromise of the system or data, and demonstrates high impact on the organization operations and resources, or the need for an organized response". Thus, just as computer incidents are not considered of interest to CSIR, minor incidents are excluded too.

Computer and Security Incidents

There are four common terms used to describe organized responses to computer security incidents:

- Computer Security Incident Response,
- Computer Incident Response
- Cyber Security Incident Response
- Cyber Incident Response.

The difference is in choosing between: "cyber" or "computer" and whether the term "security" needs to be added.

The above choice of terms is influenced by how various researchers and practitioners understand and use the terms: "computer security", "cybersecurity" and "information security". Some argue for cybersecurity being more generic than computer and information security [27] [28], while other publications use the terms interchangeably [24]. It is also noted in [29] that although the term cybersecurity is used in government agencies to refer to state-sponsored attacks on critical infrastructure or those that reflect serious organized crime, the media and many industry practitioners still use it as a synonym to computer security.

In the major publications in CSIR [2] [1] [30] , the term "computer security" seems to be more preferred over "cybersecurity". For the sake of consistency, this preference will be maintained for the purposes of this project. In addition, the term “computer security incident” will be used over “computer incident” to emphasize the security nature of the handling process.

Based on the above selection, a response plan will be referred to as: “Computer Security Incident Response Plan (CSIRP), compared to “Cyber Security Incident Response Plan” and “Computer Incident Response Plan (CIRP)”. Similarly, the term Computer Security Incident Response Team (CSIRT) will be used to describe the team tasked with CSIR responsibilities.

Computer Security Incident Response

The available definitions for computer security incident response (CSIR) are very similar, and slightly differ by the level of details in the wording of the definition. In this project, a definition adapted from [31] will be used in which CSIR is defined as: “the capability to provide end-to-end, cross enterprise management of incidents that affect information and technology assets within an organization”. Note that, incident management and incident handling are used interchangeably.

The definition of a computer incident response team (CSIRT) introduced by [24] and adopted by NIST [9] is: “Group of individuals usually consisting of Security Analysts organized to develop, recommend, and coordinate immediate mitigation actions for containment, eradication, and recovery resulting from computer security incidents.”. Although, the terms CSIRT and CERT (Computer Emergency Response Team) refers to the same capability, the latter term is currently labeled to entities that act as a single point of contact and focus on coordination responsibilities. Since the focus of this project is incident handling, the term CSIRT is selected.

The definition provided by CNSS [24] for an incident response plan (CSIRP) is adopted in this project, which states: “the documentation of a predetermined set of instructions or procedures to detect, respond to, and limit consequences of a malicious cyber-attacks against an organization’s information system(s).” The focus in this project is on the documentation of how an incident is handled, i.e. actions and procedures, compared to strategic design, business models and security policies. Also, in many organizations a CSIRP is broken into several documents and appendices, For simplification purposes, it is assumed that all incident handling procedures are gathered in a single document.

1.4.2 Performance Evaluation Terms

Performance Framework

The term ‘framework’ is used in this project to refer to the generic conceptual setting that encompasses several perspectives and methods for the investigation of a research problem. This definition is derived from how the term is commonly used in works of computer security that exemplify some resemblance to this project [32] [33] [34] [35]. Based on this, ‘performance framework’ is understood as: ‘a generic conceptual setting that encompasses several perspectives and methods for the evaluation of computer security incident response capability”. Compared to a model which is specific and involves some process, a framework is more abstract and generic. In addition, models are normally developed within a framework.

Performance Evaluation and Quality Management

There is a diverse range of how the following terminologies pertaining to performance are used by various researchers: performance evaluation, performance measurement, performance management and performance appraisal. These terms are also sometimes used interchangeably with quality terms like: quality management, quality control, quality enhancement and quality. The following paragraphs will elaborate on how these terms are defined for the purposes of this project. Further discussions will appear in the survey of articles in this chapter or whenever a term is used for the first time in this document.

The simplest definition of performance is: “how well a person, machine, etc. does a piece of work or an activity” [36]. Based on this definition, CSIR performance can be defined in simple terms as “how well the CSIRT executed the CSIR activities”.

Since measuring “how well” an activity is done can be a subjective matter, different fields view performance in terms of measurements against specific thresholds or standards. For example, the Business Dictionary defines performance as: “The accomplishment of a given task measured against present known standards of accuracy, completeness, cost and speed” [37]. Refining the definition of performance in the context of CSIR, the definition would be “how well the CSIRT executed CSIR activities as outlined in the CSIRP to achieve the outlined goals”.

Although performance measurement is the main performance activity, the term ‘performance evaluation’ is used here as a higher level term that encompasses: performance planning, performance management, performance measurement, performance analysis, performance monitoring and continuous enhancement. Thus, ‘performance evaluation’ and ‘performance evaluation system’ are used as synonyms.

Based on the above, CSIR Performance Evaluation is defined as: “All activities associated with assessing how well a CSIRT executes CSIR activities as outlined in the CSIRP to achieve the outlined goals”. Note that this definition presumes the existence of a

CSIRP and the predefinition of goals. Therefore, it is out of the scope of this project to evaluate CSIRTs acting with no prior planning.

In several contexts, the term ‘performance appraisal’ is used interchangeably with the term ‘performance evaluation’. Performance appraisal is “ a formal management system that provides for the evaluation of the quality of an individual’s performance in an organization” [38]. From the perspective of CSIR, performance evaluation is not a form of performance appraisal, because it is focused on the overall system performance compared to staff performance, and it targets the collective outcomes of the CSIR team not the individual performance of each member.

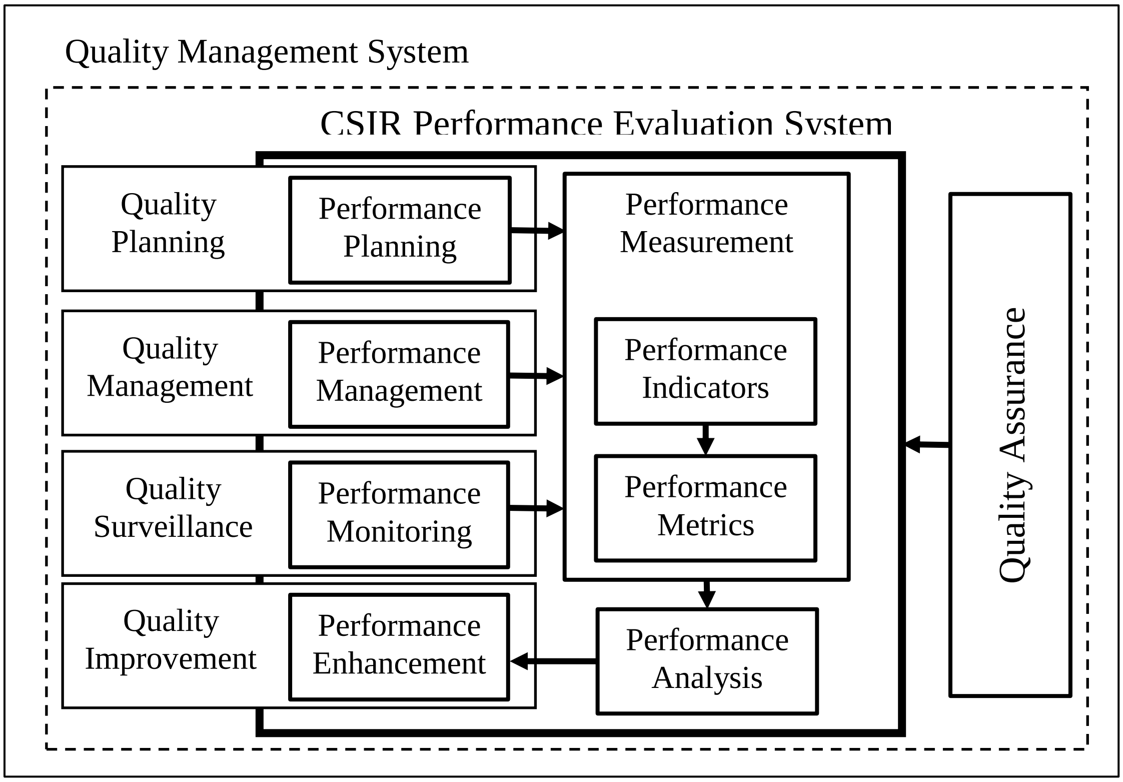


Figure 1: Taxonomy of Performance and Quality Terminologies used in the project

From an organizational perspective, activities relating to ‘quality management’ and ‘performance management’ are very similar in their aims of leading the organization to better productivity and profitability [39]. In this project, quality is viewed as a broader organizational activity that encompasses performance evaluation of the CSIR capability. For instance, CSIR performance planning is one of the activities, but not limited to, that quality planning is concerned with. In Figure 1, a chart that depicts the hierarchy and relationship between the various terms pertaining to quality and performance are presented, as understood for the purposes of this project.

Finally, the terms: ‘performance management’ and ‘performance control’ will used interchangeably.

Performance Indicator

Performance measurement is the process of “quantifying the efficiency and effectiveness of an action” [40]. This measurement is done through what is known as performance indicators (PIs) and performance metrics (PMs). Again, there is a diversity of definitions and usages for the two terms. It is assumed in this project that performance indicators are in higher order than performance metrics, following the GQIM model presented in [41]. Thus, for each performance indicator, there is one or more performance metrics to be used to measure it.

The CSIR system has various qualities and properties that can be subject to measurement. The term ‘performance indicator’ is given to a system quality that reflects some performance behavior, positive or negative. Examples include: response time, response stability, containment effectiveness, recovery cost and detection accuracy. Each of these indicators can reflect a desirable (e.g. short response time, high response stability, low recovery cost …etc) or undesirable (e.g. long response time, low response stability, high recovery cost …etc) system feature that reflect performance. Also, each indicator can be measured in various ways and through various tools. These tools can be quantitative or qualitative, and the number of tools can be singular or multiple.

Performance Metric

The proposed definition of ‘performance metric’ is: ‘a measurement tool used for characterizing or quantifying a quality of CSIR performance’. It could also be defined as: “a tool used for measuring a single or multiple performance indicators’. This definition is aligned with the definition borrowed from measurement theory [42], as both assert the quantitative nature. An example of a performance metric used to measure the detection effectiveness is the difference of the incident severity scale identified during the detection phase and the actual severity scale as approved in the analysis phase. This metric can also be used in the measurement of other performance metrics like incident handling accuracy. Also, this metric is not the only possible tool to measure detection effectiveness. The length of the time period between the first reporting of the incident until the incident declaration can be another metric to measure detection effectiveness.

Effectiveness and Efficiency

Finally, the terms “effectiveness” and “efficiency” continually appear in the definitions of performance activities. Effectiveness is “the extent to which planned activities are realized and planned results are achieved”, and efficiency is: “relationship between the result achieved and the resources used” [43]. Using the aforementioned definition, effectiveness and efficiency can be classified as higher-level performance indicators. It is debatable whether all performance indicators should be mapped to effectiveness and efficiency. Innovation and traceability are examples of system qualities that are not straightforward to be mapped to effectiveness and efficiency.

In some fields like supply chain management (SCM) and engineering systems, it is easy to distinguish between issues related to effectiveness from those relevant to efficiency. However, in the field of incident response, the two aspects are strongly related. Therefore, in this project, a definition of effectiveness that encompasses efficiency will be adopted. For the context of CSIR, performance effectiveness is defined as: "the extent to which CSIR goals are achieved and harm is reduced through efficient use of resources".

CHAPTER TWO

BACKGROUND

The research topic of this dissertation can be linked to two general areas of literature; the first is computer security and the second is performance evaluation. Both of these fields are too vast to be surveyed. Therefore, this literature review will be narrowed down to the following two sub-areas: (1) computer security incident response and (2) performance evaluation of incident response. The common term between these two sub-areas is: ‘incident response’.

This chapter has three sections in which the first two correspond to surveys of major works of the above two sub-areas, and the third provides a summary of main findings along with discussion.

In the first section, an overview of the area of computer security incident response is presented focusing on its development and the current models and standard planning procedures. This is followed by a discussion of performance evaluation within the general context of computer security and within the specific domain of computer security incident response. Finally, an overview of security metrics is presented as a background for the discussion on performance metrics of computer security response.

The second section provides a literature review of how other disciplines, like medical emergency and engineering critical systems, handle the issue of performance evaluation of incident response. As the area of Computer Security Incident Response

(CSIR) is relatively new compared to the other disciplines, this survey aims at borrowing best practices from these fields to the area of CSIR.

The last section provides a summary of the current state of the art in performance evaluation in CSIR, and lessons learnt from other disciplines. The proposed framework, outlined in Chapter 3, will make frequent references to these findings which act as a foundation for the designing and modeling of the framework.

2.1 Computer Security Incident Response (CSIR) Literature Review

2.1.1 History

It is a matter of agreement among computer science historians that issues pertaining to security started to emerge as soon as computers were invented in the 1960s. Some even argue for ancient roots found in cryptology and events associated with phone wiretapping that happened in the late period of 1800s. This rich history was subject to various studies from various perspectives. For example, history of cryptology can be found in [44], history of information security in [45], history of hacking in [46], history of viruses and malware in [47], history of computer crime in [48] [49], and the contemporary rise of Cyber-warfare can be found in [50].

Although some glimpses on the history of computer security incident response (CSIR) can be found in some works like [51] [52], there is no dedicated work that fully trace the history of computer security incident response. However, since issues pertaining to computer security are interconnected, it is difficult to extract a well-identified history for CSIR that is independent of the other sub-fields of computer security. For every computer security incident, dated or contemporary, there are two parties involved: one party that exploit, attack or hack and one party that responds with a counter measure. In such view, it could be argued that CSIR was born on the same day the first incident occurred, which is a matter of debate among historians. Nevertheless, I will attempt to identify some milestones that significantly contributed to the development of CSIR as it is understood today. In that direction, focus will be on activities and incidents that triggered organized or coordinated response in a manner that is similar to how CSIR teams operate today. In my view, the following activities and incidents stand as the major milestones in CSIR development:

The Morris Worm and Creation of CERT (1988): In November 2, 1988 Tappan Morris, a graduate student at Cornell University, released into the internet one of the earliest distributed worms that will have a significant impact on shaping the current understanding of computer security. The worm was designed to be undetectable, and was launched from an MIT computer as a camouflage of the worm origin. The worm exploited some vulnerabilities in the UNIX system, more specifically some versions of the Berkeley Software Distribution (BSD) system, and had the capability to replicate itself to other machines once it has infected a specific machine.

The worm creator claims that the worm was not designed to cause harm, and hence the actual intent remains a speculation. The outcomes of the worm release were unexpected, both to the public and to its own creator, as the scope and magnitude of the worm spread was outstanding. Although the worm did not cause actual loss of data, it caused an economic loss in millions of dollars resulting from thousands of machines being down. It could be considered the earliest distributed denial of service attack, and is sometimes referred to as: the Great Worm. In terms of CSIR, it was the most influential incident that sparked public and official discussion on the need for “collective” and "organized" response to computer security incidents. The incident triggered the establishment of the first Computer Emergency Response Team (CERT) and resulted in the first prosecution under the Computer Fraud and Abuse Act of 1986. For further information on the technical details and the impact of the Morris worm, consult the following two works: [53] and [54].

After two weeks of the Morris Worm release, the Computer Emergency Response Team Coordination Center (CERT/CC) was created. The center is a US federally funded project, mainly funded through the Defense Advanced Research Project Agency (DARPA), and is positioned within the Software Engineering Institute of Carnegie Mellon University. The center was designed to act as a single point of contact for Internet security problems, and as a platform for sharing response information with the government, private sector and public at large. The early documents of CERT draws the analogy to a fire department, as the public can call for assistance to any computer security incident. The concept of Computer Security Incident Response Team (CSIRT) was developed, and CERT can be considered the first CSIRT. In the next decade, hundreds of CSIRTs were formed across the US. Today, most of these CSIRTs are members of one umbrella, the Forum for Incident Response and Security Teams (FIRST). For further information on the history of CERT and its current activities, consult: [55] and [1].

The Wank Worm and Creation of FIRST (1990): The Forum of Incident Response and Security Teams was established in 1989, and started operating in 1990. It was triggered by the WANK worm, which revealed the need for coordination between various CSIR teams across the US. The Wank worm targeted The US Department of Energy and NASA. It is believed that the worm originated from Australia, and it constitutes the first computer breach with a political message, as it displayed an anti-nuclear message. The main objective of FIRST is to coordinate sharing of information and collaboration between various CSIRTs across the world. The mission statement states: “Membership in FIRST enables incident response teams to more effectively respond to security incidents by providing access to best practices, tools, and trusted communication with member teams.” As of the end of 2016, the FIRST website displays membership of 370 teams across 79 countries. For more information on FIRST’s history and activities, consult the following resources: [16].

Using the public data available at the FIRST website, I performed some analysis to see the rate and pattern of membership. Currently, the vast majority of governments and large corporates have membership with FIRST, which could be treated as a reasonable representation of the international appreciation for the need of CSIR. Out of the 370 member CERTs, it was possible to trace the start date of 242 CSIRTs, which is 65.4% of the total number of members; while there were no available data pertaining to the start date for the remaining members. This sample, however, covers the 79 countries which is 100% of the countries that have membership with FIRST. For each member, there is a date of establishment and a date for joining FIRST which is normally later than the former. The earlier date, i.e. the date of establishment, is of more interest because it represents the actual systemized action to implement CSIR.

A chart that represents the number of new members that join FIRST each year, and the number of countries that they originate from is provided in Figure 2.

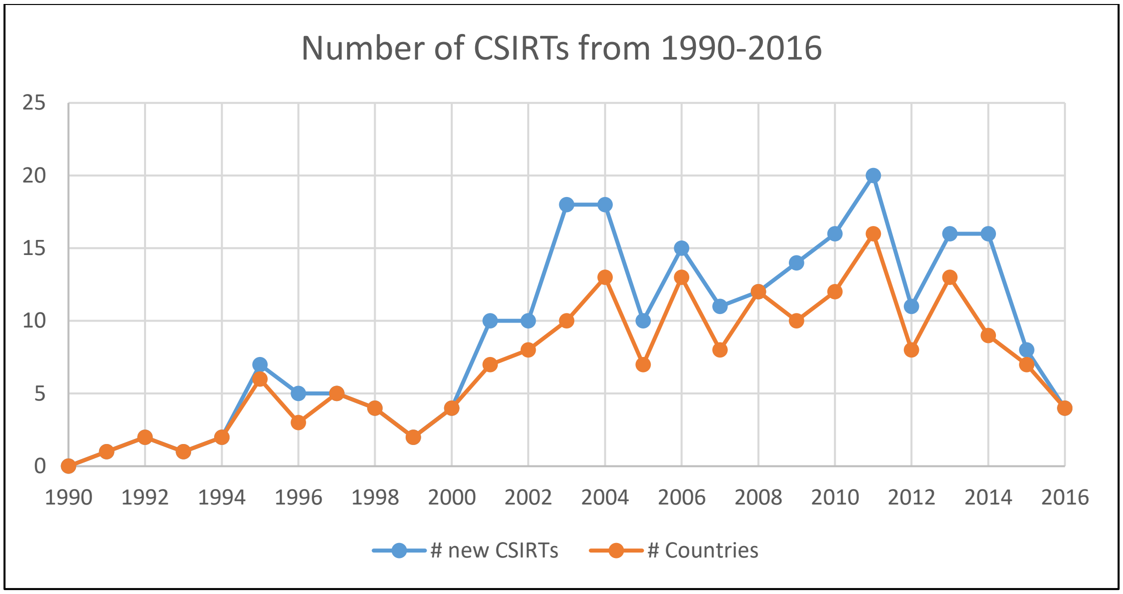
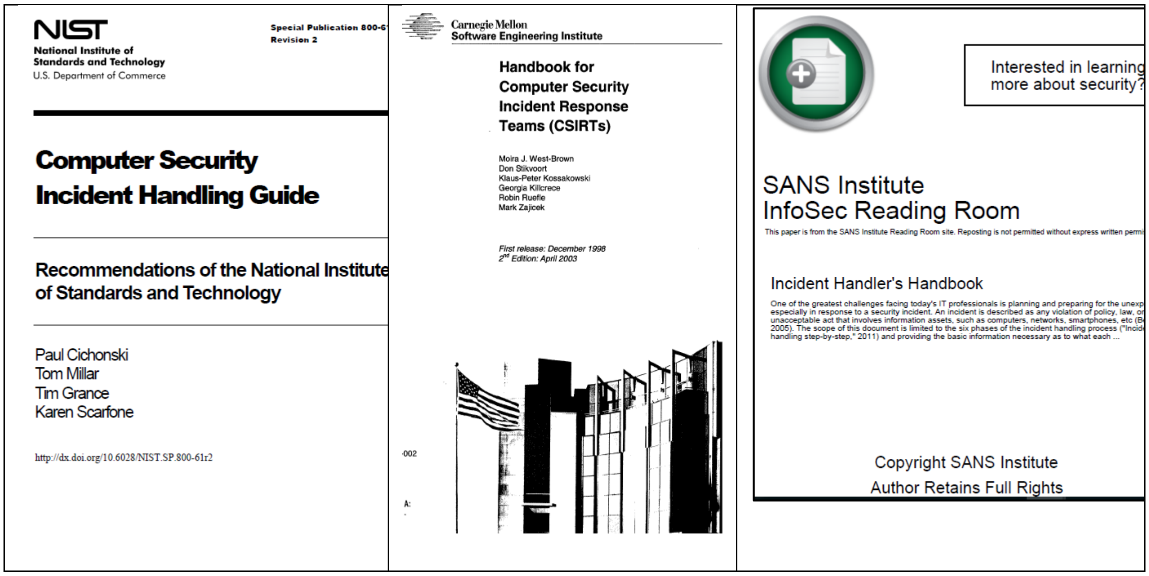


Figure 2: Number of CSIRTs worldwide in the period of 1990-2016

It is noticed that the year 2011 demonstrates the peak of participation, as there were 20 new CSIRTs created in 16 different countries. The second highest year is 2004 with 18 new CSIRTs in 13 countries. I speculate that there was no specific single incident that triggered such high participation in these years. It is mainly due to the nature of security threats on the national level that created undisputable need for creating CERTs to coordinate response at an organized large scale. It also could be credited to the availability of the series of publications from the pioneering CERT at Carnegie Mellon University, which proposed a structure and procedure for creating and executing CSIRTs.

Documentation and Standardization of CSIR: It could be argued that the publication of three main comprehensive documents had a great impact on the development of the CSIR field. The first is the Handbook for Computer Security Incident Response Teams (CSIRTs) [2] by the Carnegie Mellon University Software Engineering Institute. There were two versions released, the first in 1998 and the second in April 2003 which is the one widely used up to this date.



NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Special Publication 800-61
Revision 2

Computer Security Incident Handling Guide

Recommendations of the National Institute of Standards and Technology

Paul Cichonski
Tom Millar
Tim Grance
Karen Scarfone

http://dx.doi.org/10.6028/NIST.SP.800-61r2

Carnegie Mellon
Software Engineering Institute

Handbook for Computer Security Incident Response Teams (CSIRTs)

Moira J. West-Brown
Don Stikvoort
Klaus-Peter Kossakowski
Georgia Killcrece
Robin Ruefle
Mark Zajicek

First release: December 1998
2nd Edition: April 2003

Interested in learning more about security?

SANS Institute
InfoSec Reading Room

This paper is from the SANS Institute Reading Room site. Reposting is not permitted without express written permi

Incident Handler's Handbook

One of the greatest challenges facing today's IT professionals is planning and preparing for the unexp especially in response to a security incident. An incident is described as any violation of policy, law, or unacceptable act that involves information assets, such as computers, networks, smartphones, etc (B 2005). The scope of this document is limited to the six phases of the incident handling process ("Incid handling step-by-step," 2011) and providing the basic information necessary as to what each ...

Copyright SANS Institute
Author Retains Full Rights

Figure 3: The three main CSIR publications

The second documentation is the “Computer Security Incident Handling Guide” published in August 2012 by the US National Institute of Standards and Technology (NIST). Concurrently, the SANS Institute was developing its own documentation, which appeared in December 2011 by the title: “The Incident Handlers Handbook” [30]. These three publications are the three main references used today by the industry for implementing a CSIRT.

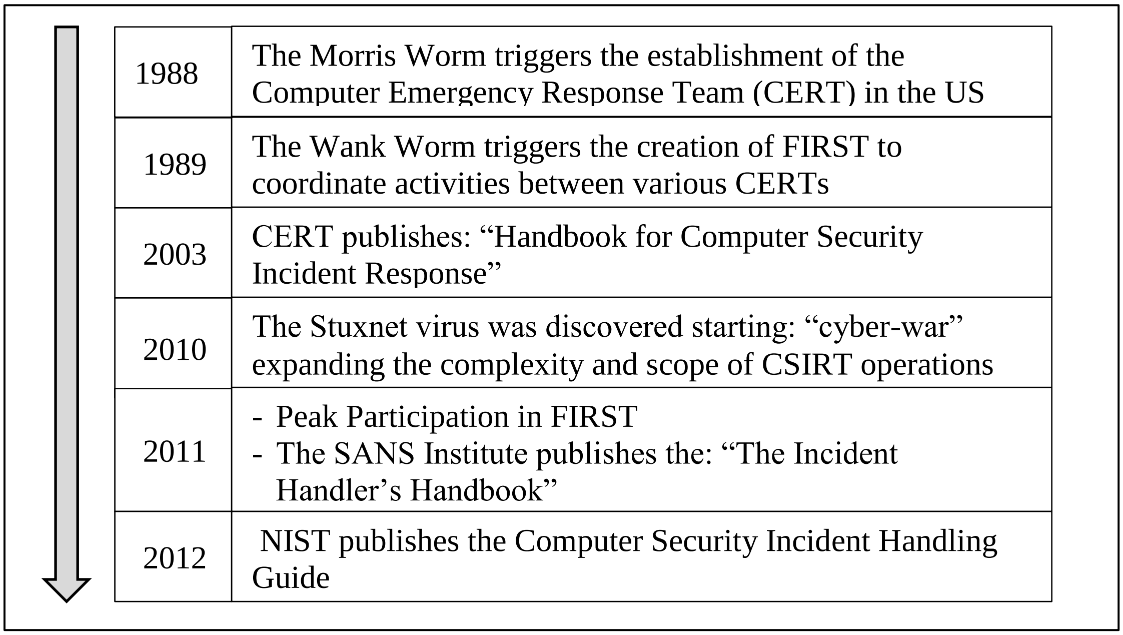
Stuxnet and start of Cyberwar (2010): The Stuxnet malware is considered the first digital weapon with substantial destructive power. The malware hit the Iranian Natanz Uranium enrichment facility in early 2010, causing damage to controllers of centrifuges at the nuclear power plant. Up to this date, no entity had officially called responsibility for the implant of the malware, although several journalists and scientists speculate that it was initiated by a US-Israeli operation to hurdle progress to the Iranian Nuclear Weapon.

From the technical side, the Stuxnet malware targeted specific programmable logic controllers (PLC) such that the malware would only be executed if specific configurations were found on the destination host. The malware is believed to have been transferred through an infected USB exploring its way through worm-like techniques into the intranet of the Nuclear plant. The malware exploited several zero-day flaws, a term given to vulnerabilities exploited by hackers without prior knowledge from vendors about its presence [56]. For further information on the Stuxtnet, the investigative book of the journalist Kim Zitter is an important resource [57].

In my view, Stuxnet and its aftermath are critical to the development of the CSIR field. Although there are several noteworthy state-level cyber incidents prior to Stuxnet [58], the scope and impact of Stuxnet is unprecedented. It takes computer security response into a very sophisticated and advanced stage that is beyond the capacity of CSIRTs as defined by the industry. It is analogous to local police responders tasked to respond to a large scale terrorist attack which require highly trained special forces and the coordination of several federal agencies. It could be argued that two types of CSIR exist today, one on the national level with military like techniques and one on the business level which operated in confined environment and limited budgets. This a divergence from the traditional CERT model in which responding to security threats appear as joint process from the public and private sectors.

Furthermore, with Stuxnet targeting industrial PLCs and being used as an offensive weapon, it adds levels of complexity to the response. The CSIRTs would need to consider lower-level security breaches in the hardware and industrial systems which is not currently of high priority. Also, responding to a backlash of an offensive security breach or dealing with unexpected consequences of an offensive breach would require different techniques than the defensive ones currently used by CSIRTs.

In Figure 4, a summary of the timeline for the above activities that contributed to the development of CSIR is presented.



1988	The Morris Worm triggers the establishment of the Computer Emergency Response Team (CERT) in the US
1989	The Wank Worm triggers the creation of FIRST to coordinate activities between various CERTs
2003	CERT publishes: “Handbook for Computer Security Incident Response”
2010	The Stuxnet virus was discovered starting: “cyber-war” expanding the complexity and scope of CSIRT operations
2011	- Peak Participation in FIRST - The SANS Institute publishes the: “The Incident Handler’s Handbook”
2012	NIST publishes the Computer Security Incident Handling Guide

Figure 4: Timeline of Computer Security Incident Response (CSIRT) development

As this study focuses on performance evaluation in CSIR, it is also interested in identifying the phase at which the field started deliberating the need for integrating performance evaluation in the CSIR process. The conducted survey of the published works reveals that the need for using performance evaluation in computer security incident responses was not acknowledged until a later stage of the field’s development. Most of the early CSIR studies focused on developing models and procedures for handling computer incidents. In these publications, the issue of performance evaluation was not considered a core functionality of the process and hence received minimal or no attention. For example, the remarkable 2003 CERT publication of “Handbook for Computer Security Incident Response Teams (CSIRTs)” lacked any reference to the issue. Another comprehensive IR planning book [3] provides only an indirect reference to performance evaluation in the context of highlighting the importance of performing performance measurements when designing business processes. A similar short indirect reference is found in SANS's Incident Handler's Handbook [30]. The recent publication dates of the latter two works, August 2012 and December 2011 consecutively, suggest that not only the issue of performance evaluation is overlooked in the early works, but it also continue to be overlooked in some major publications in the field.

I found an early reference to "measuring the effectiveness of computer security incident response capability (CSIRC)" in one of the early publications of NIST [59] dated in 1991. However, the term was used to refer to how using incident response would contribute positively to an organization's performance. In other words, the term was used in the context of the effectiveness of introducing a CSIR into an organization, which is now a well-established need, compared to measuring the effectiveness of the response team, which is the interest of this dissertation.

The earliest direct reference to performance evaluation that I was able to trace was in the CERT publication of 2008 [60]. However, the more substantive references [1] [61] [41] had to wait until 2012 to appear. It could be argued that some studies earlier than 2008 made references to performance evaluation in the context of computer security. However, these were made in other areas of computer security like: vulnerability assessment [62] (2007), IT service level agreements [23] [63] (2006) and intrusion detection [64] (2001). A discussion of these references will be presented in Sections 2.1.3 and 2.1.6.

2.1.2 CSIR Models and Processes

Incident Response Life Cycle Models

There are several models for expressing the main activities performed by incident responders. In Figure 5, a depiction of four common models to represent the incident response life cycle is presented. The four models come from NIST [1], US CERT [65] [31], SANS [56] and a hybrid model adopted by some professional experts in the field [66].

The four models express the same methodology and only differ in naming, categorization of activities, and emphasizing some interaction between phases of the life cycle. For example, the term "identification", used by SANS and the hybrid model, and the term "detection" used by NIST and US CERT, are used to describe the same activities performed at the initial stages of the incident. In the US CERT model, "prepare" and "protect" are highlighted as separate activities with strong interaction in order to stress the need to perform both prior to any incident handling. On the other hand, the other models consider "protect" as part of the "prepare" phase. The same could be said about whether "containment" and "eradication" should be considered two separate phases, or one phase. Based on this discussion, choosing which model to adopt is a matter of preference that has little impact on conceptualizing the main IR activities.

In this work, and for the sake of consistency and without loss of generality, the hybrid model will be used. The hybrid model captures the processes of both the NIST and SANS models in a reasonable number of phases, and resembles other industry-based models like: [29] [67]. Compared to the US CERT model that stresses inputs and interaction, the hybrid model stresses the actions, which is a simpler way to describe the IR handling. As there is no statistics on the industry's usage of these models, at least it is confirmed from a CSIR practitioner [66] that this model is currently being used in some large corporates which guarantees that the model is practically used. It also demonstrates high resemblance to other practically used models like the one in [68]. In Table 2, the major activities in the five phases of the IR life cycle as presented in the hybrid model are provided.

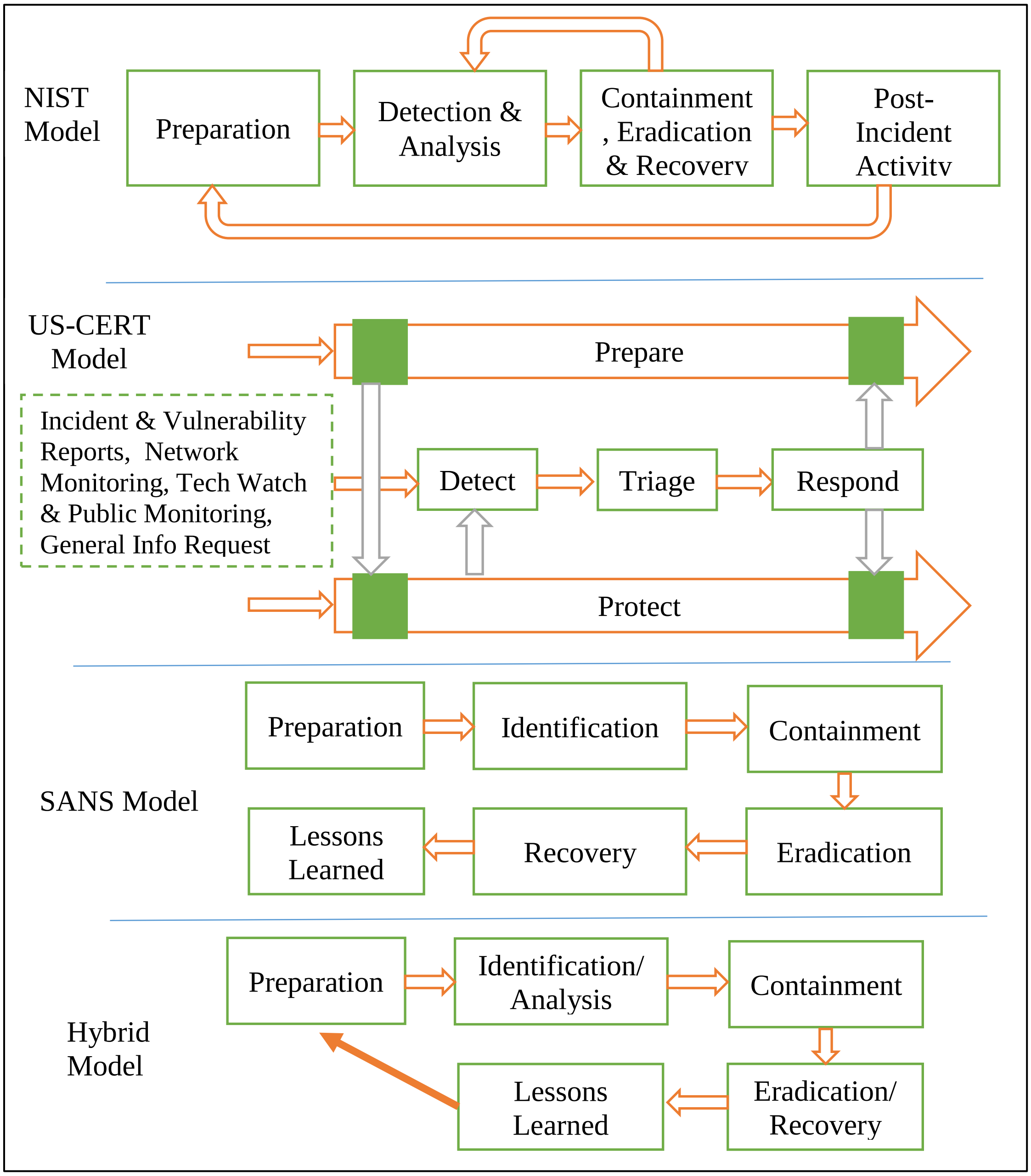


Figure 5: Models for Incident Response Life Cycle

The Computer Security Incident Response (CSIR) can be viewed as a proactive and as a reactive process. It is proactive as the majority of time is spent in planning and preparing for incidents, while it is reactive because of its main objective which is to respond to incidents once they arise.

#	Phase	Activities
1	Preparation	• Review of policies and security processes • Prepare and write an incident response plan (CSIRP) • Training and testing the CSIRP • Prepare hardware and software needed for IR • Approve the CSIRP and raise awareness
2	Identification/ Analysis	• Analyze “alerts” and “indicators” for possible incident • Perform initial assessment and validation • Review available evidence and collect further information • Classify and prioritize an incident • Declare an incident and initiate executing the CSIRP • Document and notify management
3	Containment	• Perform basic triage as outlined in the CSIRP • Isolate “infected” system to ensure minimum loss • Protect critical resources and data • Identify attacking hosts and interrupt any ongoing breach • Collect evidence and document incident timeline
4	Eradication/ Recovery	• Gradually eliminate components of the breach • Restore basic operations of the organization • Identify and mitigate exploited vulnerabilities • Restore to normal operations
5	Lessons Learnt	• Analyze main causes of the incident • Analyze the response compliance with the CSIRP • Analyze response actions for potential improvement • Recommend protection mechanisms and policy updates • Update CSIRP based on lessons learnt • Provide full technical and non-technical documentation

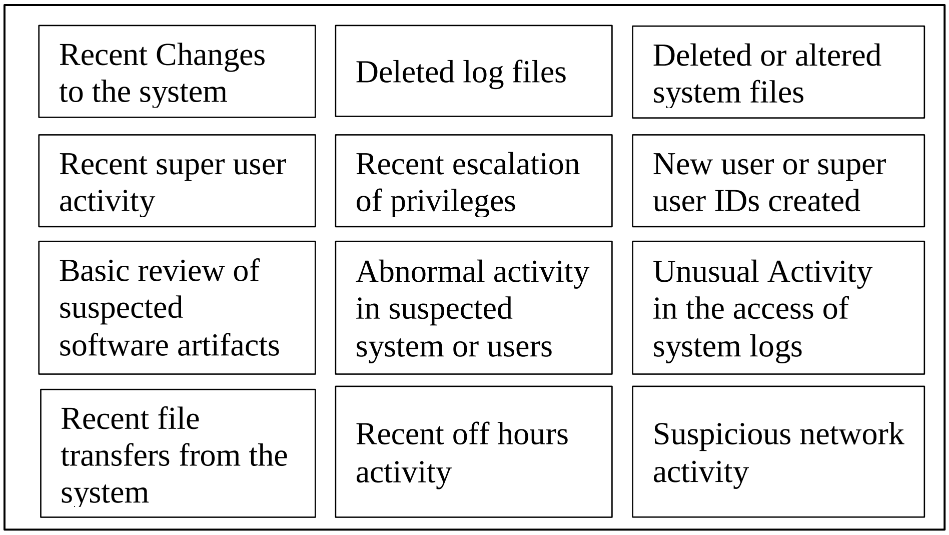
Table 2: Main activities in the IR life cycle of the Hybrid Model

The proactive side can be further divided into two major activities: planning, i.e. planning for incident handling, and security, i.e. ensuring that the environment is secure against breaches. As the field had grown, the second activity pertaining to security assurance and enforcement is delegated to designated security teams, which might or might not be part of the CSIRT. In small size organizations, the whole security team is part of the CSIRT and is tasked with planning, security and incident handling.

Incident Identification Mechanisms

One of the earliest works on CSIRT [69] outlines five general techniques for the identification phase of the IR life cycle. The first is called *differencing*, which is comparing the current state of the system (post-attack) with one that is previously stored (pre-attack) and highlighting the differences to understand the sequence of actions taken by the attacker. The second technique is *finding*, which investigates a specific part of the system during a specific time frame looking for flags and unusual behavior. The third is *snooping*, which is to place some monitors on the system to observe future steps by the attackers. The fourth is *tracking* which relies on the use of logs and audits to backtrack the steps of the attack. And finally, *psychology* which relies on using social engineering techniques and sending specific messages to stimulate the attacker to behave in a specific way.

As the CSIR field had grown and became more sophisticated, more refined technical steps were developed for the identification phase. However, the above five general mechanisms would still apply. In Figure 6, a list of major activities performed in the identification phase is presented. The list is compiled to reflect those outlined in [3]. Note that these activities should be performed before any briefing to the management and after identifying the portion of the system potentially compromised within a time frame.



Recent Changes to the system	Deleted log files	Deleted or altered system files
Recent super user activity	Recent escalation of privileges	New user or super user IDs created
Basic review of suspected software artifacts	Abnormal activity in suspected system or users	Unusual Activity in the access of system logs
Recent file transfers from the system	Recent off hours activity	Suspicious network activity

Figure 6: Main activities in the Identification Phase of the IR cycle

Incident Classification Techniques

There are various classification techniques applied at the early stages of incident handling which work to determine the nature of the incident and guide for a proper response. The classification techniques differ on factors used for categorizing incidents, like: incident source, security target, severity level and expected engagement level.

Classifying incidents based on their source are normally used by CSIRTs operating at the national level or by teams that work closely with prosecutors. Under this classification technique, the technical method for exploitation are irrelevant. For example, state-sponsored or organized hacking will be classified under one category compared to incidents that demonstrate minor crimes conducted by individuals, even if both use the same hacking method. An example of studies that discuss this method can be found in [29]. It could be argued that this classification is not recommended as the main assessment method during the initial stages of the incident handling for CSIRTs not operating at the national level. This is due to the fact that it is practically common that the actual source and motivation are only confirmed at the later stages of the incident handling or deduced through post-incident analysis.

Another classification method is functional [3] and inspects the level of engagement of various parties in executing the plan. Under this method, there are only two levels: Priority 1 and Priority 2 incidents. Incidents classified under Priority 1 are those that require the full support of the CSIRT, full execution of CSIRP and approval of senior management before execution; while Priority 2 incidents require partial execution of the plan and does not need the engagement of all parties. In order to determine which category to classify an incident, the potential harm of the incident would be investigated. For example, incidents that might impact the organization major services or might cause major financial damage would be classified as Priority 1, while incidents that do not impose external interaction as with law enforcement and media can be classified as Priority 2 incidents.

The classification method suggested by NIST [1] is more comprehensive and considers three aspects of the incident: functional impact, information impact and recoverability effort. The functional impacts investigates the incident’s impact on the business operations and services offered by the organization. The information impact inspects actual and potential harm over the organization’s information through the CIA model (confidentiality, integrity and availability). The recoverability estimates the time and resources needed to recover from the incident. A summary of these elements is presented in Table 3. Note that an incident can be classified higher according to one factor but low according to another. It is a decision to be made by the CSIRT to take proper actions based on its investigation of all of the three factors.

	Functional Impact	Information Impact	Recoverability Effort
Level 1	**None** No impact on ability to provide services	**None** no information compromised	**Regular** Predictable time with existing resources
Level 2	**Low** Offer services but lost efficiency	**Privacy Breach** sensitive PII was accessed or exfiltrated	**Supplemented** Predictable time with additional resources
Level 3	**Medium** unable to provide some critical service	**Proprietary Breach** Compromised unclassified or protected data	**Extended** Unpredictable recovery time, external help & additional resources
Level 4	**High** unable to provide some critical services	**Integrity Loss** Lost Sensitive or proprietary data	**Not Recoverable** Recovery not possible, launch investigation

Table 3: Incident Classification Mechanism proposed by the NIST document

2.1.3 Computer Security Incident Response Team (CSIRT)

Types of CSIRTs:

The NIST document outlines three models for structuring a CSIRT [1]: centralized, distributed and coordinating. The centralized CSIRT is the most commonly used model and constitutes of a single team that handle incidents across the organization. The distributed team model is structured in several CSIRTs, each handling a specific geographical segment of the organization or a group of resources. To avoid conflicts, a mechanism for coordination and collaboration is outlined. The Coordinating Team model is structured as a central main CSIRT which provides consultation, with no executive power, to several CSIRTs tasked with incident handling, or part of its operations.

Another method to classify CSIRTs is relative to their association with the constituencies. A CSIRT can be internal or external [70]. An internal CSIRT has all of its members as members of the organization being served, while an external CSIRT is formed with nonmember staff responsible for incident handling through some form of contractual relationship. Note that the meaning of “external members” here is different than “external members” outlined in a CSIRP which refers to external entities, like law enforcement, media and lawyers that collaborate with the core CSIRT team during the response process. The categorization of internal and external is based on the relationship of those tasked with the core CSIR responsibilities within an organization. It is also possible to have a hybrid team of internal and external staff. However, if this is not properly structured with clear outlined responsibilities in the CSIRP, it can leads to conflict of interest and non-cohesion among the team members.

Other methods for classifying CSIRTs can be found in [51].

Organizational Structure

Although CSIRTs can be structured and positioned in an organization in various ways [2] [1] [70], it can be viewed as a body consisting of a core team, several support teams, several extended teams and several external teams [66].

Members of the core team are responsible for carrying the core responsibilities of incident handling, planning and review. The team need to be balanced to reflect technical expertise and knowledge about the business and administration operations of the organization.

The support teams are those that will have substantive involvement in the incident handling process but under the supervision of the core team. These teams normally consist of members whose main responsibilities in the organization are not related to CSIR. The support teams can be further divided into technical and logistical teams. Examples of technical teams include forensics, help desk, networking, data center and system administrators. Logistical teams include: administrative assistants, physical security, transportation, communication …etc. The technical teams can be grouped in one single team or in multiple teams and the same applies for the logistical team.

The extended teams are those members of the same organization that provide indirect support or guidance to the core team. Most important among these teams is the executive and business teams in the organization whom the core team will be regularly reporting and consulting with during the incident response. Other teams include human resources, the legal team and the outreach team. These teams are normally outlined as contacts in the CSIRP to be reached as needed. Finally, the external teams include law enforcement, national and peer CSIRTs, and technical and legal consultants that could be reached if assistance is needed.

A chart that depicts the above relationships between the core CSIRT team and other teams inside and outside the organization is found in Figure 7.

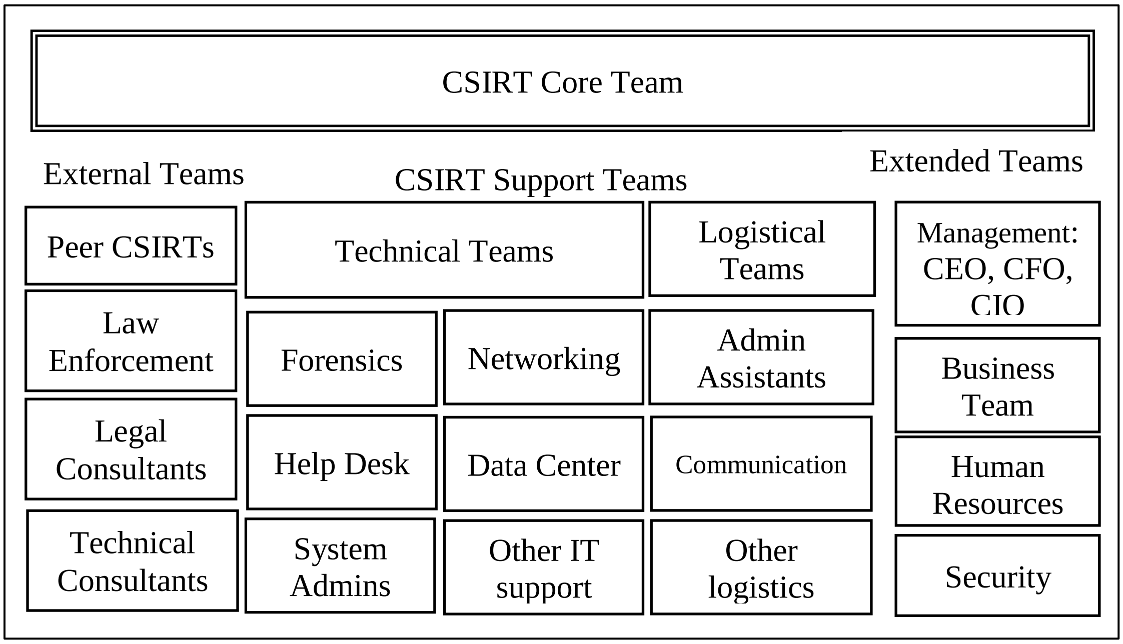


Figure 7: Organizational Structure of a CSIRT

It is noted in [51], that the element of trust is an essential aspect of the culture in which CSIRTs operate. It is a precondition for the numerous collaborations which a CSIRT establishes. Trust is normally built through necessity, opportunity or through introductions by trusted parties. The study also notes that there are four principles that were common about CSIR practitioners interviewed by the study. The four principles are operational independence, reciprocity, confidentiality and transparency. Operational independence mean the CSIRT should operate independently from other policy objectives, focusing only on CSIR objectives. The principle of reciprocity suggest that if a team shares some information, techniques or warning with a party, then it is expected that the other party will do the same as the need arise. Both confidentiality and transparency express one of the two essential professional ethics which CSIR practitioners adhere to.

2.1.4 Computer Security Incident Response Plan (CSIRP)

There are various templates for developing computer security incident response plans (CSIRPs), sometimes also known as CIRP (Computer Incident Response Plan). These templates might vary in the outline, structure and presentation methods, but the minimum requirements of what needs to be included is very similar.

The NIST document [1] lists eight elements that need to be included in any CSIRP. These elements are: (1) Mission (2) Strategies and Goals (3) Senior Management Approval (4) Organizational Approach to CSIR (5) Communication between CSIRT and the rest of organization (6) metrics for measuring CSIR capability and effectiveness (7) Roadmap for maturing the CSIR capability (8) How the CSIR plan fits into the overall organization. It is noticed that the above template emphasizes the organizational perspective and the development roadmap but omits the technical details and the specifics of the IR phases.

This might be explained by the generic nature of the NIST document which emphasizes guidelines and policies over technical procedures.

A sample of a CSIRP outline published at the website of the California Department of Technology is an example of plans that emphasize practicality and documentation method [71]. The sample outlines 17 actions that need to be taken along with the documentation requirements of each action which can be customized based on the nature of the agency preparing the plan. A list of these steps along with a description is provided in Table 4. The advantage of the above CSIRP outline is its practicality and simplicity which makes it accessible by various technical and non-technical personnel. The website lists other forms to be used for performing more technical activities like budget assessment of response. The disadvantage of such approach to CSIRP preparation is the lack of information about the rationale of the outlined steps and that links to the organization's policies and strategies.

#	Title	Description
1	Emergency Contact List	A list of names and phone numbers to be contacted if an incident happens
2	Discovery Reporting	Procedures of reporting the discovery of an incident if discovered by a member of the IT department
3	Public Reporting	Procedures of reporting the discovery of an incident if discovered by a non-member of the IT department
4	Security Office Reporting	Logging information if incident reported to the grounds security office
5	Initial Assessment	Steps taken by the IT department after the reporting in terms of communication within the department and performing initial assessment of the severity of the incident
6	CSIRT assessment	Initial steps by the CSIRT members with regards to incident assessment and setting a plan for response

7	Incident Declaration	Official declaration of an incident under one of the outlined incident categories
8	Response plan execution	Execution of outlined steps suitable for the type of response, e.g. worm response procedure or system abuse procedures
9	Forensics	Application of forensics techniques to extract evidence
10	Immediate Actions	Outline of immediate actions to be taken based on analysis done in Steps 8 and 9
11	Management Approval	Getting management approval to execute the above response recommendations
12	Restoration	Steps taken to restore the system after containing the incident
13	Documentation	Full list of items that need to be documented post-incident
14	Evidence Preservation	Steps to be taken to preserve collected evidence
15	External Reporting	Inform external agencies of the incident like law enforcement, media and lawyers
16	Damage Assessment	Assess damage and response cost
17	Review	Review response and formulate lessons learnt list

Table 4: Outline of Action-Based CSIRP Sample

More detailed CSIRP outlines can be found in [3] and [66]. In [3], a CSIRP is structured into eight major sections: (1) Plan Introduction (2) Incident Preparation (3) Incident Detection, Analysis and Declaration (4) Incident Response Supporting Actions (5) Incident Containment, Eradication and Recovery (6) Post Incident Activity (7) CIRP Roles and Responsibilities (8) Plan Maintenance. A summary of the main topics presented under each of the above sections is presented in Table 5. It is noticed that this outline is comprehensive and combines both organizational and technical details. Some of the public CSIRPs like [72], demonstrates much resemblance to the above outline.

#	Section Title	Contents
1	Plan Introduction	Objective, scope, assumptions, plan execution and command topology, plan ownership, plan structure
2	Incident Preparation	Compliance framework, sensitive data, third party payments, third party services, compromise notification protocols
3	Incident Detection, Analysis and Declaration	Sources of indicators, incident thresholds, incident analysis (business and technical impact), incident categories, incident declaration and declaration
4	Incident Response Supporting Actions	Plan execution, organization and roles, process and rhythm, synchronization and decision making, release of public statements, evidence discovery and retention, liaison with law enforcement
5	Incident Containment, Eradication and Recovery	The Data compromise team, containment procedures (isolation, verification of non-affected systems, third party connections, consequence management), eradication and recovery procedures (remediation, compensations, disaster recovery and business continuity)
6	Post Incident Analysis	Incident Termination (criteria, decision process, evidence retention, response statistics), lessons learnt
7	CIRP Roles and Responsibilities	Accounting, human resources, information security forensics, IT operations center, Investor relations, disaster recovery, internal communications, legal teams, public affairs, retail operations
8	Plan Maintenance	Regular updates, incorporation of lessons learnt, annual testing of the plan

Table 5: Contents of CSIRP as outlined by [3]

Another outline of a CSIRP is provided by [66]. The plan demonstrates similarity to that of [3] in terms of contents and being comprehensive. However, it follows a different structure. The plan consists of four major sections: Charter, Business Documents, Incident Checklist and Appendices. The charter focuses on the authority question and contains information similar to those presented under "Plan Introduction" of Table 5. The Business Document section contains the following information: plan revision history, testing schedule, organization chart, roles and responsibilities, data classification policies, compliance regulations, training requirements and business continuity documents and disaster recovery documents. The third section on “Incident Checklist” contain the specific sets for various types of incidents like: cyber intrusion, denial of service, malicious code outbreak and phishing incidents. The final section on “Appendices” contains documents like: incident classification levels, escalation procedures both internal and external, data preservation procedures, law enforcement support, war room and communications, call tree, incident declaration forms and glossary of terms. It is noticed that this outline is designed such that to facilitate the maintenance and update of the CSIRP.

2.1.5 CSIR Performance Evaluation

This section discusses references made to performance evaluation (PE) in publications dedicated to the study of computer security incident response. The discussion starts with the CSIR comprehensive works that address the issue of PE like [25] [1] [73]. Then the discussion extends to CSIR studies such as [61] [22] [41] [74] [75] [76] that address some aspects relevant to PE, like preparedness and effectiveness of detection mechanisms.

The 2008 CERT report: “Creating and Managing Computer Security Incident Handling Teams (CSIRTs)” [60] is probably the first major work in CSIR to make an explicit reference to performance evaluation. The report calls to: “define methods for evaluating the performance of the CSIRT”. As the report stresses, this is only possible once the response workflow is clearly outlined and incident management processes are well defined. However, the report falls short of providing more information on how to design these “methods” and how it could be integrated into the process.

The above CERT report adopts the US CERT model for the IR cycle, depicted in Figure 5, and hence identifies five main functionalities to be assessed: (1) Prepare (2) Protect (3) Detect (4) Triage (5) Response. The report mentions performance evaluation within the context of the fifth phase: “response”. This conservative usage of the term suggest a limited scope for performance evaluation. According to the report, “response” is limited to actions taken after the detection and triage, compared to the current usage of the term which includes the three activities together, i.e. detection, triage and response. Indeed, some subsequent CERT publications like [41] present a wider scope for performance evaluation to include all of the phases in the IR cycle, a perspective that is followed in this project.

The way that [60] presents performance evaluation raises some fundamental issues relevant to this project. For example, according to the report, the subject of the performance evaluation is the CSIR team. Although this is a reasonable choice, as shall be presented in Chapter 3, this might be confining and it would be better to use the plan (CSIRP) instead of the team (CSIRT) as the main subject of the performance evaluation process. This shall provide several advantages and acts as a more generalized method. In my view, the performance of the team is a subcategory of the overall performance evaluation of the response outlined in the CSIRP. Along the same discussion, the CERT report considers performance evaluation as a method for measuring the performance of the team over a period of time that involves several responses. Again, this can be restrictive and it will be argued that performance evaluation should be generalized such that it could be applied to a single and multiple incidents.

The CERT report provided examples of simple performance metrics to be used like number of reported incidents, response time and amounts of incidents resolved. This acknowledges the importance of performance metrics within the development process of the performance evaluation system. Also, the report suggests that the overall performance of the team should be done in conjunction with the management and the constituency. Both issues, the role of performance metrics and who should conduct performance evaluation will be studied in Chapters 3 and 5.

The NIST “Computer Security Incident Handling Guide” [1] published in August 2012, is an important document guiding the operations of the industries in the area of computer security incident response. The document made two brief but important references to performance evaluation. The first reference is mentioned in the context of policy elements that are needed for response plans. The need for developing metrics to measure the capability and effectiveness of the plan is listed as one of the eight elements that need to be included when creating incident response plans. The guide later mentions that this could be done through data collection. The guide neither provides examples of performance evaluation policies nor describes the type of data that need to be collected, leaving it to the industry to come up with such policies and identify the relevant data.

The second reference, an important one, is found when discussing post-incident activities. The guide suggests that one of the nine elements to discuss in a “lessons Learned meeting” is “How well did staff and management in dealing with incident? Were the document procedures followed? Were they adequate?” Assisting the industry to better answer the above questions is one of the issues of interest to this dissertation.

The guide proposes performing two types of assessments for each incident: objective and subjective. The objective assessment is based on analyzing the collected incident data, while the subjective assessment records the team members and resource owners evaluation of the performance of the response. The guide provides eight examples of what could be done in objective assessment. One of these examples, which is relevant to this project, is: “measuring the difference between the initial impact assessment and the final impact assessment”. One of the goals of the proposed performance framework is to come up with similar assessment methods, with properly derived metrics, that measure the performance of the major activities of the incident response. Using the above example, ‘impact’ needs to be defined and metrics will be proposed on how to measure it in the context of CSIRPs.

The Vocabulary for Event Recording and Incident Sharing (VERIS) community project [73] provides a framework for describing computer security incidents. The project developers describe it as: “a set of metrics designed to provide a common language for describing security incidents in a structured and repeatable manner.” The goal of the project is to help organizations collect and report incidents in a uniform manner that enables sharing analysis within the organization and with external entities.

The project identifies five major areas that are needed to construct an incident narrative: incident tracking, victim demographics, incident description, discovery and response and impact assessment. A summary of the main descriptors/metrics used to describe each of these five areas is presented in Table 6.

The VERIS format is adopted by some major IT service providers, like Verizon, to report incidents. This makes the VERIS platform synchronous with recent threats and current industry practices.

#	Category	Descriptors
1	Incident Tracking	Incident ID, Source ID, Incident confirmation, related incidents, confidence rating, incident notes
2	Victim Demographics	Victim ID, Primary Industry, Country of Operation, State, Number of Employees, Annual revenue, Locations affected, Notes, Additional Guidance
3	Incident Description	Threat Actors (External, Internal, Partner), Threat Actions (Malware, Hacking, Social, Misuse, Physical, Environmental, Error), Compromised Assets (Variety, Ownership, Management, Hosting, Accessibility, Cloud, Notes), Security Attributes (CIA)
4	Discovery & Response	Incident Timeline, Discovery Method, Root Causes, Corrective Actions, Targeted vs. Opportunistic, Additional Guidance
5	Impact Assessment	Loss categorization, Loss estimation, Estimation currency, Impact rating, Notes

Table 6: VERIS Project Incident Narrative Descriptors

Three observations about VERIS could be highlighted which are relevant to this project. First, When an incident is reported, the project suggests giving a 'confidence rating' to the classification of that incident. This provides a practical solution to the observation highlighted by NIST [1] that detecting incidents with accuracy is a very challenging task. The NIST document acknowledges that this high impact on team performance and cost and suggests undergoing thorough initial assessments before declaration. I speculate that if initial assessment procedures are coupled with the confidence rating metric proposed by VERIS, this could be a very valuable information to the CSIRT in terms of allocating resources and responding more efficiently.

Second, the victim demographics descriptors aim to describe the affected organization by giving information like: industry type, country of operation, number of employees, annual revenue and locations affected. To the best of my knowledge, this is the single unique work that attempted to incorporate such information for the purposes of benchmarking. Although the above information is very basic and it is expected that due to its lacking of more information about the organization's CSIR preparedness that it will fall short of providing meaningful comparisons, the VERIS project is commended for taking the first step. If several companies pursue the issue of benchmarking more seriously, this could provide the public with very valuable guidelines of best practices in the area of CSIR.

Third, when performing impact assessment of a security incident the project suggests classifying impact into direct and indirect. Direct impacts are those targeting the assets and can be quantified in currency. On the other hand, indirect impacts are those resulting from a stakeholder's reaction to an incident, e.g. customer, shareholder ..etc. The project admits that the indirect impacts are more difficult to quantify, and leaves it as an open research question. This also imposes challenges to using indirect impacts in performance analysis. As shall be presented in Section 2.2, some works in the performance evaluation of freeway emergencies [77] argued for including only direct measurable impacts in the performance evaluation of IR.

A comprehensive list of "metrics" for measuring the capability of an organization to perform CSIR functions is provided in [74]. The comprehensive document, which appears in 200 pages, aims at providing guidelines to CSIRTs in terms of what needs to be existing in the CSIR capability. The provided metrics are presented in the form of questions to be assessed through scoring. The scoring is mainly in the form of “Yes” or “No”, depending on satisfying a list of “indicators”.

The study provides a remarkable and rich repository for detailed activities associated with the preparation of CSIRP. It could be viewed as a comprehensive guide for measuring the completeness of a CSIRP document, and partially as a tool to measure an organization’s preparedness for handling security incidents. Sample questions include: “is there a financial plan for incident management functions?”, “is forensics analysis performed on constituent systems and networks?”, “do the analytical processes incorporate methods to determine the risk or threat level of a confirmed incident”, and “is there an established business resumption plan to support disaster recovery?”.

Comparing [74] with the performance study provided in this project, the following similarities and differences could be noted:

1- The focus of this project is measuring the effectiveness of a CSIRP and a CSIRT, while [74] focuses on ensuring the preparedness of a CSIR capability. More specifically, [74] focuses on completeness, which is only one aspect of effectiveness as defined in this project.

2- The main evaluation method used in [74] is ensuring the “existence” of an “ability” to perform an activity relevant to IR. However, this project is interested in assessing each of these abilities. For instance, [74] provides a list of actions to ensure that an organization has mechanisms to classify the severity of incidents. On the other hand, this study is interested in how to measure effectiveness of these mechanisms.

3- Major proportion of [74] is devoted to assessing the security measures in an organization, an aspect which is out of the scope of this project.

4- Both works have a similar goal which is to establish a system for assessing CSIR that leads to enhancement. However, [74] is more targeted towards building a CSIR while this project is geared towards maintaining it.

Besides CSIRP completeness, another important activity in the preparation phase of the IR cycle that impacts performance is the preparedness of the CSIRT. It has been noticed [76] that the majority of computer security professionals spend their time in ensuring compliance with various standards and organizational requirements. However, these professionals normally lack the complex technical nature of incident response and the behavioral skills needed to conduct CSIR. The study also notices that CSIRTs who are capable of using several general purpose software in combination to the specialized software performed better in simulated training exercises. This brings forward the complexity of preparing CSIRTs to effectively perform their duties.

In that direction, a proposal to measure performance through measuring the preparedness of a CSIRT is found in [61]. The authors presented “A Competency Lifecycle Roadmap (CLR)” as a model for assessing the preparedness of a CSIRT. The CLR builds on the argument that workforce effectiveness relies on two critical measures: competence and readiness. Competence is measured with respect to an individual, i.e. the level of understanding of a subject matter and the ability to apply a given skill; while readiness is the ability to apply a set of competencies to fulfill the requirements of a real task. The CLR is structured as five activities: assess, plan, acquire, validate and test readiness, and is executed through specific criteria within an environment.

Similar to [60], the performance focus of [61] is the responders. Although the focus of this project is the CSIRP, it is important to highlight the correlation between the two, i.e. the performance of the team and the performance of executing a plan. The readiness and competency of the CSIRT members influence the performance of any IRP execution and vice versa. Distinguishing between the two aspects should be observed when attempting to identify causes of poor performance.

Another distinction between the focus of this dissertation and [61] is that the later views performance, at least in the context of the publication, as a pre-activity process targeting managers and trainers. However, I argue for a broader perspective in which performance is assessed prior to execution, monitored while execution and analyzed post-execution. With that regard, the performance preparedness model presented in [61] could be integrated within the proposed framework as an element of pre-execution assessment.

A subtle but important reference to performance evaluation is found in [75]. The study stands out for highlighting the importance of developing performance metrics for "incident management performance" during the "preparation" phase. The other works limit discussion of the preparation phase to preparedness, and consider performance evaluation as a post-activity in the last phase of the IR cycle. The study calls for "collecting metrics that assess the quality of a process improvement purposes" which is expansion of the requirement to have a policy for performance evaluation outlined in the NIST. Again, there were no details about these metrics as the study focuses on formal representation of the incident response tasks through defining unambiguous relationships, in the context of incident response management.

A simulation study focusing on CSIRTs effectiveness of assessing threats is found in [22]. One of the study's conclusions is that the vast majority of alerts passed to a CSIRT constitute non-threating incidents, mainly due to IDS false alarms, legitimate user activity perceived as threating and other ambiguous network activity. This issue poses as an obstacle for achieving effectiveness. The study suggests careful classification of incidents at the early stages to properly define the threat levels. The proposed method for assessing a threat is to use a simple additive equation of three factors: (1) trajectory of the attack (2) assets targeted and (3) perpetrator. The detailed levels within each of these factors are elaborated in Table 7. Although, the use of an additive metric to measure the attack severity can be misleading, the study is commended for the simplicity of the proposed metric.

Level	[1] Attack Trajectory	[2] Targeted Assets	[3] Perpetrator
1	Targeting no specific entity	No asset	Careless or unknown entity
2	Targeting a specific single entity	A client or set of client assets	Action associated with criminal activity
3	Targeting multiple entities or high-level entities	An infrastructure asset	Action associated with an advanced threat

Table 7: Model for Assessing Threats proposed by [22].

A recent workshop by CERT (January 2015) proposed a model to define and measure goals for an organization [41] based on the GQIM model proposed in [78]. Although the model is generic, it was designed for the purposes of managing computer security activities. The GQIM model defines objectives in which goals (G) are derived. The process then goes into formulating questions (Q), defining indicators (I) and finally metrics (M). The main contribution of the publication is providing detailed information on how to define goals in the context of computer security plans in a manner that permits performance measurement through indicators and metrics. The NIST guide to handling computer security incidents [1] requires the definition of goals and formulation of metrics in CSIRPs, and the GQIM model provides an implementation guide to deriving goals.

The framework proposed in this project borrows from [41] the taxonomical hierarchy, i.e. objectives are higher than goals and indicators are higher than metrics. Since the NIST guide neither distinguishes between goals and objectives nor between indicators and metrics, the above GQIM model does not create incompatibility issues with the guide. It could be viewed as a more refined method for applying the recommendations stated in the NIST guide. Furthermore, as [41] has left out elaborating on how to derive indicators and metrics from goals, the proposed framework aims at filling that gap by providing guidelines for such derivation and providing applied examples of performance indicators and metrics that could be used in CSIRPs.

Finally, a recent report on computer security incident response plans at nuclear facilities was published by the International Atomic Energy Agency [79]. The report considers “measuring the effectiveness of a CSIRP” as an essential element of the post-incident review phase. However, the report does not elaborate on how that could be achieved.

2.1.6 Performance Evaluation of Computer Security Systems

Occasionally, the issue of performance evaluation arises when studying various aspects of computer security and IT management. For example, several publications have studied the performance of intrusion detection systems (IDS) [64] [80] and vulnerability scoring [62] [81]. As this project seeks to design a framework for measuring the performance of CSIRTs, the effectiveness of the detection process will arise and such studies can be used as a resource when deriving performance indicators. However, this project is more generic and is focused on the design of higher level performance measurement techniques which could be applied to a wider range of security incidents compared to designs that effectively handle specific category of security threats.

On the other end, the performance evaluation focus of this project is more focused compared to performance measurement of IT incidents presented in studies like [23] and [82]. For example, an organization's help desk team manages general IT incidents which might or might not involve security incidents of interest to CSIRTs. Compared to incidents handled by CSIRTs, general IT incidents appear in higher frequency, have a relatively small cost and impose less impact on the operation of the overall organization. Therefore, the focus of performance measurement of general IT incidents would be on the average time and cost, while incidents handled by CSIRTs need to focus on both individual and average performance as the cost and impact of each individual incident might be magnificent.

In the following paragraphs, a review of some publications in cyber security that address the issue of performance evaluation will be presented. These papers were selected based on their relevance, from my viewpoint, to the proposed framework in terms of the applicability of the presented ideas to the area of CSIRP performance measurement.

The difficulty of studying performance analysis in computer incident management is highlighted in [82], which states: "The complexity of real-life enterprise-class IT support organizations make it extremely hard to understand the impact of organizational, structural and behavioral components on the performance of the currently adopted incident management strategy, and consequently, which actions could improve it.". The authors presented a decision support model, SYMIAN (SYMulation for Incident Analysis), for improving the performance of incident management which models an organization's IT support as a Markovian open queuing network [63]. The model claims to simulate the effect of corrective measures before their actual implementation.

The scope of the model presented in [82] is similar to those presented in [23] and [83] as the three studies focus on the performance of managing several responses to general IT incidents. These incidents appear in high frequency and have relatively low impact and cost, in contrast to incidents dealt with in CSIRPs. Nevertheless, the proposed model in [82] raises attention to two relevant design considerations.

The first consideration relates to the question of how to layout performance measurement procedures in a CSIRP, especially in the absence of prior measurements. The authors relied on the *what-if-analysis* method to predict the impact of changes before implementation. The technique relies on prediction of system performance based on the theoretical design and suggested handling procedures under various expected incidents. Indeed, most CSIRPs perform a similar analysis when laying out potential attacks and proposing procedures for handling it. Another example of how standard CSIRPs use what-if analysis is compiling a list of secondary contacts, or backup teams, if the primary contacts and team members are not available. However, under the surveyed published CSIRPs, issues related to performance, if ever addressed, would be dealt with in an ad hoc manner by the core team. It is worth considering to use the what-if-analysis for issues related performance and provide basic documentation about it in the CSIRP.

The second consideration relates to performance evaluation modeling. The paper treated performance evaluation in two orthogonal dimensions: effectiveness and efficiency. Effectiveness focuses on the chain process while efficiency focuses on every single support group. This approach could be adopted when a CSIRT is composed of several technical groups like: network, forensics, management …etc. Such approach, which is a common practice in performance evaluation of quality management of chain processes, helps measuring the performance of each team, and the overall system performance. It also helps identifying bottlenecks and aspects of the system that have high impact on performance.

The issue of uncertainty in incident response to intrusions has been addressed in an early study [64]. In automated intrusion detection systems (IDS), alarm messages are generated when some suspicious activity is detected and are brought to the attention of the system administrator. There is a delay between the detection of the activity and the actual response, which henceforth will be referred to by the term “response delay”. This time period is normally used by attackers to further exploit the system. An earlier study [84] has demonstrated through simulation that if attackers get a response delay of ten hours, then they will be 80% successful; and if given twenty hours of response delay then they will be 95% successful. Although these results are outdated, the observation that response delay impact the incident response effectiveness remains valid.

The problem that [64] was interested in is: since IDS systems are automated there is uncertainty whether a flag indicates an actual alarm or a false alarm. The authors identify three sources of uncertainty: uncertainty in detection, uncertainty in classifying the attack and uncertainty in preparing a response. The provided technical details of how to measure the uncertainty is no more applicable, as IDS and Intrusion Prevention Systems (IPS) have undergone several advancements in the past fifteen years. Nevertheless, the question of uncertainty is still valid. For example, when an attack stops, this can be due to action taken by the response team or it could be simply due to factors known to the attackers. It would have major consequences if the response team falsely falls under the impression of proper contamination while in reality the attack was temporarily halted by the attackers. This scenario and others stress the need for careful design of performance measures that are sensitive and conscious of the issue of uncertainty present in the response. In summary, the paper pays attention to the contention between response delay and uncertainty. A team might successfully reduce the response delay but through actions that might be neglectful of some uncertainty factors.

One of the main side effects of improper handling of the uncertainty factor is the stability of the response. A response might escalate or fluctuate throughout the response cycle, depending on the threat assessment of the incident. Therefore, CSIRTs should rely on effective assessment methods that are developed in the various computer security disciplines. One example is assessment of vulnerabilities which is currently standardized in the Common Vulnerability Scoring System (CVSS) [62]. The CVSS model was developed by the Forum of Incident Response and Security Teams (FIRST), and is currently used integrated into the NIST [9] and CERT handbooks. It is the only open system and is characterized by its reliance on only quantitative metrics [81].

The CVSS model categorizes metrics in three categories: base, temporal and environmental. Base metrics focus on the fundamental characterizes of the vulnerability that are constant across environments. Examples include: *Access Vector* metric which measures how a vulnerability is exploited, *Access Complexity* which measures the complexity of the attack required to exploit the vulnerability, *Authentication* metric which measures how many times an attacker need to authenticate before exploiting a vulnerability of a target, *Confidentiality Impact*, *Integrity Impact* and *Availability Impact*. Temporal metrics focus on the variant threat levels of a vulnerability that may change over time. For example, the exploitability metric measures the current state of exploit techniques or code availability. Finally, the environmental metrics measure the impact of a threat on a specific environment. Example metrics include: collateral damage potential, and target distribution.

Models like the CVSS bring its rewards and drawbacks to the study of CSIR performance evaluation. The main reward is that such models produce rigorous metrics which the performance metrics will be derived from. Researchers need not to re-invent the wheel, and can build on the findings of studies in security metrics. For example, assessing potential damage of a vulnerability is a complex task, which impacts any measurement of the effectiveness of the response in reducing harm. However, when a well-research method, like the collateral damage potential metric in CVSS, is provided then a good proportion of the task is fulfilled. Researchers can focus on proper derivation of performance metrics from security metrics compared to full design of metrics. On the other hand, this leaves performance designers dependable on the reliability and accuracy of the security metrics. Inaccuracies resulting from poor design of security metrics are likely to propagate to performance measurements obtained from metrics derived from these security metrics. Striking a balance between building on findings of security metrics which might not have been necessarily designed to serve performance measurements, and proposing metrics with high accuracy would be a challenge.

The previous discussion on the relationship between performance and security metrics incites visiting the broader works on computer security metrics. The field of software engineering had undergone extensive study on the development of software metrics, to assess the quality of each of the development phases of the software cycle . The ISO 9216 standard, with its series of publications [85] [86] [87] [88] , is a major reference that outlines how these metrics are derived and used. The classification method for security metrics is generic and can be applied to the PE metrics.

Metrics can be classified as: quantitative or qualitative, subjective or objective, static or dynamic and direct or indirect. A static metric is obtained during the static state of the software, while a dynamic one is obtained during the execution of the software. A direct metric is independent of other measures, while an indirect metric is dependent on measurements of other metrics. This is the most comprehensive method I found for classifying metrics, which could be borrowed to the field of performance evaluation of CSIR. More discussion on security metrics will be presented in Section 2.1.7.

Another area to consult in the computer security literature is the performance evaluation of security systems. Responses to computer security incident normally involve interaction between human (the team) and non-human (e.g., machines, software) entities, and an effective response requires effectiveness from both fronts. These systems are used throughout the response cycle, including the detection, containment and recovery. The performance of intrusion detection systems, firewalls, forensics tools and backup systems contribute to the overall performance of the system. Again, this is a huge literature and it suffices to show an example, besides the example of intrusion detection provided in this section, of how this literature can be consulted when developing performance systems for CSIR.

In [89], the authors attempt to develop performance metrics to capture the impact of attacks on Mobile Ad hoc Networks (MANETs). Unlike wired networks, MANETS have specific characteristics due to its dynamicity, as nodes and routes continually change. The mobility factor adds more challenges to the reliability of the routes compared to static wireless networks. The performance of MANETs is normally measured through packet delay, packet drop rates and routing effectiveness.

The authors classified attacks on MANETs based on their objective, presenting six types of attacks: denial of service attacks, black hole attacks, flooding attacks, packet dropping attacks, route disruption attacks, wormhole attacks. The technical details of these attack can be found in the following survey of different attacks in MANETs [90]. The authors suggested then to develop metrics such that they measure the impact on two aspects: level of denial of service, and manipulation of network and routing topology.

Ten quantitative metrics were presented. Some metrics can only be used to measure impact on denial of service, like the packet loss ratio metric, and others can be used to measure the topology and routing manipulation, like route length per packet metric. Interestingly, some metrics can be used to measure both like the round trip delay metric. This demonstrates that some performance metrics can be used to measure several aspects of the system performance, an idea which will be used in the proposed framework. The various attacks were executed in a simulated environment with varying parameters. The results showed that some attacks had varying impacts based on number of nodes or number of attackers while the impact remained constant for some types of attacks. This provides MANETS operators with mechanisms to interpret various activities, how to classify attacks and which counter measures to follow in response. This demonstrates that the use of performance metrics can results in enhancements to the response process, as the above results were not clear prior to the use of performance metrics.

2.1.7 Measuring Security in Computer Systems

In this section the notion of ‘how can security be measured?’ will be investigated. As shall be presented, the feasibility and practicality of quantification of security into metrics is debatable among researchers despite its wide spread usage. In the current endeavors about the development of metrics to measure the performance of CSIRTs, the above debate about security metrics arises. Performance metrics evaluating a CSIRT depend on how ‘success’ of achieving a specific task is defined, which is normally expressed through security metrics. Stated in other terms, the calculation of performance metrics will depend on some measurements of security metrics. The fuzzy definition of some security metrics and their questionable practical value can influence how much the outcomes of performance evaluation can be trusted. In the following paragraphs, a survey the findings of some major works in security metrics will be presented followed by my viewpoint on how that impacts the proposed framework.

One of the earliest works addressing the issue of security metrics [91] states that the nature of the field makes security measurement an “operational” process, i.e. it attempts to measure the ability of a system to resist an attack, or the ability of the system to remain free of breaches under specific conditions. Nevertheless, the study notices that most attention is being paid to measuring the safeguards followed during the design and implementation processes, which does not necessarily reflect the security of the system when it actually operates. The authors propose to view “security” similar to “reliability” by contrasting “system failure” with “system breach”. The consequence of adopting such model is producing metrics that are probabilistic in nature, like the expected number of breaches in a period of time, and the probability of a specific mission to be accomplished without security breaches.

Reliability is defined as the probability of failure free operation over a period of time. Such probabilistic definition makes reliability an operational measure compared to other measures that capture the static properties of the system. Comparing reliability and security, reliability measures failures which depend on system faults, which is analogous to security breaches that depend on system vulnerabilities. Another similarity is the high target of making the system “very secure” which is analogous to target of reliability of safety-critical systems. However, failures are mainly accidental while security breaches are normally intentional. Also the notion of ‘time’ is different in both contexts. In reliability, higher measures are proportional to longer times of fault-free, while on security this is not necessarily true as breaches vary in their degree and scope and the role of time in assessment is vague.

The above comparison between reliability and security reveals that uncertainty in both contexts can be expressed through probability. The author warns that in the context of security the interpretation of probabilities is more subjective. For example, an attacker is ‘all-knowing’ of the breach mechanism but does not have certain and complete information about the system, while a system administrator is (ideally) ‘all-knowing’ of the system but uncertain how an attack might happen. This means that a probability can be interpreted differently depending on the viewpoint, in the above example: the attacker or system administrator.

Reflecting on the above discussion, there are three remarks to record. First, during the design of performance metrics a distinguish should be made between operational and static measures. Measuring the preparedness of a system is an example of a static measure, while measuring the execution performance of a specific task outlined in a CSIRP is an example of operational measures. Second, there is a need to deeply understand the notion of time in performance evaluation. Sometimes, time is a crucial measure as when measuring the time from detection to starting the response, while in other contexts time might be misleading as when considering the period of time the responders managed to secure against higher level attacks. This might be due effective security and response measures or due to mere coincidence and absence of targeted sophisticated breaches. Third, as performance evaluation is mainly an operational measure, the use of probabilistic measures is reasonable. However, it is important to define the 'viewpoint' of how these measurements will be interpreted. A measurement read by a technical system administrator might be interpreted differently than if read by the Chief Executive Officer (CEO) or by the Chief Financial Officer (CFO) [92].

An extensive survey, published in 2009, examines the major works in the area of security metrics in the period of 1981 to 2008 [34]. The title of the work: "Quantifiable Security is a Weak Hypothesis" reveals the survey's main conclusion about the hypothesis of "security can correctly be represented in quantifiable fashion".

The study identifies five characteristics of operational security that make it difficult to model and analyze. These characteristics are: (1) the dynamicity of the environment in which threats and security measures adapt to changes from each other (2) Low stationarity as systems and software are continually updating (3) Economics: there are various agents (e.g. attackers and defenders) with conflicting goals and interests (4) Dependence of various components of the system and users on each other (5) Uncertainty. The study then proposes a simple conceptual model to study operational security. Operational security is viewed as a system comprising of three components: (1) systems: structural components, security controls and users (2) Threats: active and passive agents capable of violating the security properties of the system. (3) Vulnerabilities: system properties that allow the exploitation of the system.

Two main outcomes of the study are of relevance to this project. The first remark is counterintuitive and comes from analyzing the ‘perspective’ of security metrics. Perspective here refers to the conceptual viewpoint of what security is which motivates the quantification of metrics. The study classifies perspectives into four major categories: (1) Confidentiality-Integrity-Availability (CIA): which is the classical viewpoint of security as Confidentiality, Integrity and Availability (2) Economics: viewing security as risk and trade-off analysis (3) Reliability: views security as a stochastic process subject to probabilistic analysis of the rate of system failures (4) Other techniques derived from computer science, like viewing security in graph models. The counterintuitive finding is that the CIA was the least commonly used perspective to derive security metrics. The study reasons that this is due to the advancement of the other perspectives in quantification methods as displayed in various disciplines. This made it easier to extend such methods to computer security than to derive metrics through the CIA model which is native to the computer security discipline. In my view, this also demonstrates that deriving security metrics is a challenging task.

The second interesting finding is that the vast majority of works used assumptions that are inconsistent with the nature of operational security. For example, most models assume independence and rationality while practically security systems exert very high dependence and empirical evidence suggests that a relatively large percentage of attacks are conducted through ‘irrational’ entities like automated software. The study acknowledges that assumptions need to be used while modeling, but critiques that these assumptions are diverge from the nature of operational security, causing validation problems to the quantification process. Overall, the survey concludes that proposing security models and metrics was proven to be easy but validating it is proven to be very challenging if not absent. Also, there is a lack of common experimentation methods that enable comparisons between various organizations. Finally, there is a chaos in the taxonomical usage of terms as they are normally borrowed from other fields.

The last point can be re-expressed as the field of security metrics is yet to develop its own identity. This is manifested in the absence of unified terminology and scarcity of validation methods. The five characteristics of operational security described in the above survey stand as obstacles to achieving common taxonomy and sound validation methods. However, looking at the research conducted after the above survey, the field is gradually converging into developing its own common language. The VERIS project is an example of such attempts. In terms of validity, the study is reasonable in its questioning why should descriptive metrics should be trusted without any validation of their practical relevance. Still, the study seems to lean more towards rigors and formal validation methods, which I believe is unnecessary needed when describing security features and threats.

Another extensive survey was published by NIST [93] and appeared at the same year of [34] . The survey affirms the previous observation that much research had been done in the area, but only few metrics have been deemed useful in the practical domain. The authors also note that unlike physics and chemistry which have well formulated metrology systems, the area of IT metrology is still emerging. Therefore, the term ‘metrics’ in the area of computer security has been interpreted in various of ways. The survey notes that security metrics have been used for three main purposes: strategic support (e.g. planning and resource allocation), quality assurance (e.g. during software development) and tactical oversight (e.g. compliance of procedures, policies and regulations). It is noticed that if this classification is to be used, incident response can be classified under the third category.

The study has two discussions which are significant to bring forward while studying performance metrics. First, how is the term ‘metric’ being understood and used in the security metrics literature? The authors found three common usages in the literature. Some researchers use metrics to refer to quantifiable measurements of some aspect of a system. Under this approach, qualitative measures are not considered metrics. The essence of this approach stems from the literal meaning of the term ‘metric’ which is also in alignment of how the term is being used in physics.

The second approach considers metrics as tools designed to facilitate decision making to improve performance and accountability. The focus is on the objective of assessment not the mechanism of assessment, thus quantitative and qualitative would be the same if they used to achieve the same goal. In addition, the terms ‘metrics’ and ‘measurements’ would be used interchangeably.

The third approach differentiates between measurements, which are single-point-in-time views of specific discrete factors, and metrics which are derived by comparing to a predetermined baseline two or more measurements taken over time. In other words, measurements represent actual counting, while metrics represent analysis of measurements. Examples of major works that adopt this last definition are found in [94].

For the purposes of this project, the second approach is more fitting to the research objectives. The objective of developing the PE framework is to enable the CSIRT members and decision makers to advance improvements in the performance. The difference between quantifiable or non-quantifiable metrics and the distinction between measurements and metrics is semantical with no practical implications for the interest of this research.

The second discussion of relevance is the definition of effectiveness. The study define it as: “assurance that the security-enforcing mechanisms of the system meet the stated security objectives.” To refine this board definition to incident response plan, effectiveness can be defined as “assurance that the incident response plan achieves its stated objectives outlined in the CSIRP”. Yet, it is important to note that the above definition only relates to one aspect of effectiveness, which is the IRP design. Other aspects include the IRP execution, the utilization of resources, damage control and cost. In this regard, the notion of “effectiveness” in the context of CSIR needs further investigation.

A notable recent study [42] builds on the survey findings of [34]. The study observes that most of the metrics developed in the field reflect security “management”, compared to security “measurement”. Metrics developed for security management attempt to assess an organization’s adherence to specific standards. On the other hand, the paper investigates the possibility of viewing “security” as a computer system property which then could be subject to measurement in quantifiable fashion similar to physical quantities. A customized model from control engineering is proposed in which computer threats are viewed as “disturbances” to the control. In my view, as the model is dependent on accurate measurement, which is an aspect that is difficult to achieve, the proposed model will suffer from practicality issues.

Another bias in the research of security metrics is highlighted in [95]. The authors noticed that most security metrics focus on preventive and detective measures. However, an argument is made that there is an equivalent need for metrics that measure survivability and restoration capabilities when an incident occur. The argument is based on the acknowledged observation that regardless of how much we attempt to prevent malicious events, incidents will occur. Therefore, a tradeoff between investment on both aspects of security need to be made. Metrics for CSIR performance can be grouped under the above perspective on survivability and restoration metrics.

2.2 Multidisciplinary Literature Review of IR Performance Evaluation

The field of performance evaluation is a well-founded discipline, especially in the fields of engineering and business management, and its findings are implemented across a wide spectrum of disciplines and applications. Various areas of incident response have used performance evaluation to derive its own customized performance frameworks and metrics. Examples of these fields include:

- Medical emergency response, and response to epidemics [96] [97]
- Law enforcement and fire departments incident response [98] [99]
- Natural and environmental disasters responses [100] [101] [102]
- Homeland security incidents [103] [104]
- Engineering systems like critical safety systems and transportation systems [23] [77] [105]
- Business administration and management [83] [106]

- Supply Chain Management (SCM) [107] [108]
- Incident Management Systems [109] [110] [111]

The following literature review scans through some major works in the above disciplines. The aim of survey is to investigate how these disciplines tackled the issue of performance evaluation of incident response within their contexts. Since this work is looking for insights and lessons to learn, focus will be on studies of generic nature compared to those that focus on technical implementation, an aspect which differs from one discipline to another. The proposed survey is neither comprehensive nor extensive, i.e. not all works in each discipline are surveyed and the surveyed works are not reviewed in depth.

The selected works for the survey satisfy one of the following two criteria:

1. The study provides an overview of the major works, summary of main findings or comparison of major tools and methods within that discipline, or is considered one of the major publications in the discipline.
2. The study demonstrates relevance or provides insights to the objectives of this project focusing on developing a framework for CSIR performance evaluation

A survey of about fifty works is presented. For each of these works, a critique or a description of how the work relates to the field of CSIR is presented.

2.2.1 Challenges

The difficulty of measuring the performance of incident response systems is acknowledged across various disciplines. Some of these challenges are unique to each discipline, but as this survey demonstrates, there are common challenges associated to the measurement of performance of the core activities of the IR process. In the following paragraphs, a presentation of the major challenges is provided along with a brief discussion on their applicability to the CSIR discipline.

Challenges of developing performance evaluation systems environmental incident response are highlighted in [100]. Specifically, the study targeted developing performance metrics for responses to oil spills. The study highlights that measuring the effectiveness of responses is “extremely challenging”. The authors discussed several factors behind this difficulty, from which two factors stands out. The first factor is the complexity of establishing a baseline context for measuring effectiveness. This is due to the observation that performance baselines are normally designed on incident-specific strategies, while in reality responses to oil spills are varied and are normally developed in an ad hoc manner in the early stages of a crisis. The second factor is the difficulty of measuring the public perception of successful handling as there are no normative standards of success that are established between decision makers and the public. Indeed, both parties normally provide contradicting interpretations of the response performance.

The field of CSIR seem to suffer from both challenges highlighted in the above study. In terms of the complexity of establishing baseline contexts for measuring effectiveness, this comes from the reality that the types, scales and scopes of cyber-attacks are very diverse; perhaps more than the diversity presented when it comes to oil spills. However, the frequency of oil spills is relatively low compared to the high number of computer breaches, making a distinction between the two disciplines. It is true that the element of surprise is always expected while handling incidents; however, a computer incident is more likely to be similar to a previous incident compared to oil spills incidents. It could be argued that the difficulty of establishing baseline performance contexts are inversely proportional to the frequency of incidents. Based on this argument, it would be relatively easier to create baselines in CSIR compared to oil spills incidents, but more difficult in CSIR compared to medical emergencies and other fields with high volume of incidents. The second factor of contradicting interpretations is also present in CSIR. There is a wide gap between how technical teams, executives and the public perceive the success of responding to computer incidents. This comes back to the fact that there are no well-established tools, i.e. metrics, to measure the performance of CSIR. Such tools need to be precise and disambiguate to avoid the variance of interpretation among various stakeholders.

Differences between measuring performance of manufacturing businesses and service businesses are highlighted in [106]. The authors observe that measuring performance in services is more challenging than in manufacturing. In manufacturing, enhancing performance is achieved through monitoring production processes, monitoring distribution processes, removing waste and limiting variances. In the manufacturing domain, this is possible due to the homogenous nature of production and distribution processes. On the other hand, service businesses operate under high unpredictability conditions in which the needs of customers vary widely. Such environments push executives and senior managers to tolerate relatively higher level of inefficiency and higher costs due to uncontrollable factors. The authors suggest that despite the difficulty of controlling unpredictability, performance measurement should focus on confining variances and deducing similarities between various service requests. The authors also argue that internal benchmarking is more meaningful than external benchmarking between various companies which vary in size, equipment, service level agreements and budget.

From the above description, the field of CSIR is closer to service businesses than to manufacturing businesses. It is very common to face new threats and advanced compromise strategies with high unpredictability. In addition, the current nature of the IT industry in which thousands of equipment, solutions and software that are regularly poured into the market, makes it very difficult to capture everything in the CSIR plan. However, similar to how the authors have argued in the above study, focusing on main threat trends and common attack techniques can help design more efficient CSIRPs. Again, the volume of incidents in service industries are dramatically larger than those handled by CSIRTs. Consequently, the proposed strategy of limiting variance becomes more challenging in CSIR compared to general services.

Another study that draws comparisons between incident response teams in different disciplines [97] highlights that the environment in which responders operate is very challenging. The authors describe the environment as complex, high in information load, wide information diversity, high uncertainty, continuous flow of data and quick change of information. The authors then suggest that understanding the collaborations between team members is crucial for improving the effectiveness of the response team.

The above study, among others, highlights the role of the human factor in the incident response process which could be overshadowed by emphasis on tools and techniques. The response process normally involves a high level of interaction between various technical and non-technical members, both internally and externally. In the contexts of CSIR, the core CSIRT needs to internally interact with IT personnel, engineers, managers/executives, public relations and human resources. At the same time, it interacts with external entities like legal advisors, law enforcement and the media. Therefore, I second the observation made by the above study, and suggest that the field of CSIR needs to conduct thorough studies to understand the collaborations during the response process.

The study of [99] provides a comprehensive survey of the disciplines concerned with emergency response. The study investigates operational research in emergency response over the scope of fifty years in the domains of fire suppression, law enforcement and ambulance response. The study states that "performance metrics in emergency response remain underdeveloped, particularly in the context of large-scale emergencies". The study puts the blame mainly on the complexities arising from multiple stakeholders in the public sectors which have various, if not contradicting, objectives. The study also observes that within emergency response, the planning and implementation of solutions is inseparable, as incidents develop in a unique and semi-structured matter.

There are three points that could be highlighted from my reading of the above survey. First, despite the long legacy and maturity of the field of medical and fire response, performance metrics in these fields remain underdeveloped. This brings a warning to those of us interested in studying the performance of computer incident response. It asserts that the issue of performance evaluation, especially with large-scale incidents, is very challenging and cautions against any claim that a single model or a group of developed metrics would be able to capture the performance of the response system.

Second, the issue of contradicting objectives of stakeholders is no stranger to the field of cybersecurity. The tension between safety and privacy is a classic example. Hence, in order to make sense of any performance evaluation the audience should be clearly identified. This might also require conducting performance evaluation under several contexts.

Third, as the nature of the response bring planning and implementation hand by hand, it is important to carefully reflect on the expectations of the planning process. It would be more realistic to view a response plan as a guiding document more than an instruction manual. At the same time, since responders' time is very valuable during the execution of a response, careful prior planning is crucial. Based on this, a response plan need to be both comprehensive and flexible which is not trivial.

Another unique but relevant challenge brought forward by [98], from the domain of law, is the difficulty of assessing harm. The study aims at developing metrics for cybercrimes, like scale and harm, and comparing them to metrics used in traditional crimes. The author argues that the overall form of crime did not change due to the advent of cybercrimes. The classical classification of crime as (1) crimes against persons (2) crimes against property (3) crimes against state (4) crimes against morality; are the same online and offline. The only difference is the method used; i.e. use of automation in online crimes. However, unlike traditional crimes which are by default one-to-one, online crimes are by default one-to-many which raises interesting challenges. Two challenges were recorded, the ability of law enforcement to apprehend the offenders and the ability to properly measure the scale of harm that may be inflicted. Not only the scale of harm is different but also the degree of harm by a single incident is of great contrast. For example, the average amount lost by a bank robbery in the US is $800, while a single online attack normally results of millions of dollars.

While the legal domain functions quite differently than the technical domain of computer security response, there is much to learn from the legal experience. An analogy can be drawn between a CSIRT working on containing an incident and a law enforcement team engaging in apprehending some offenders. Also, assessment of damage that CSIRTs perform at the early stages of an incident and in post-incident activities is analogous to assessment of harm inflected by crimes which is conducted jointly by law enforcement and the justice system.

I consider the model proposed in the above article for assessing and measuring harm of potential interest to researchers in the area of incident response. The article classifies *harm* under three types: individual harm (like assault), systematic harm (aggregate individual harm or generalized individual harm, e.g. mass shooting) and inchoate harm (i.e. potential harm, e.g. conspiracy). The main task of prosecutors is to properly classify an incident/crime under a specific category, and then work on assessing harm within that category. The article provides long analysis of a hypothetical scenario involving a crime of stealing a password, which is a common computer incident. The article then concludes with the note that coming up with cybercrime metrics is very essential, but it involves a lengthy and complex process.

Unlike the field of law in which harm is normally assessed post-incident, those who work in the field of CSIR normally assess potential harm in the early stages of the incident. The main outcome of such assessment is the categorization of the incident and consequently declaring a response suitable for the severity scale of the incident. Currently, there are two common models for categorizing incidents in CSIRPs. The first model is based on the CIA security model and the second inspects the compromise techniques. The CIA model, which stands for Confidentiality, Integrity and Availability, describes the security aspect of the system that an attack attempts to compromise, regardless of the technique used. The second method, on the other hand, focuses on the attack mechanism regardless of the targeted security aspect. For example, a network worm attack and a compromise resulting from lack of software batch would be classified differently under the second method even if they both result in leak of confidential customer data.

At this current stage of the CSIR's development, it is not clear if using the "harm-model", similar to the one presented in [98], in CSIRPs would outperform the other two methods. However, I foresee that using a hybrid model of the above three models during threat assessment would be helpful for efficient performance of the CSIRT. Also, the harm assessment model presented in the above study can potentially lead to development of more precise harm assessment methods which CSIRTs normally perform at the end of the IR cycle.

Although there are other works that highlight challenges of conducting PE in IR, the above discussion is sufficient for the purposes of this dissertation. The challenges which were not mentioned could be mapped to one of the challenges highlighted above. As shall be highlighted, the challenges of studying the performance of IR can be abstracted in two general challenges: complexity and unpredictability. The high load of information, contradicting objectives, interaction with large groups, difficulty of designing baselines and lack of comprehensive performance metrics are all aspects of complexity. On the other frond, the diversity of incidents, the quick escalation, the instability of the environment, and the evolvement of planning along with implementation are elements that contribute to unpredictability. Hence, any performance evaluation system for the field of CSIR need to address both issues: complexity and unpredictability. Strategies for overcoming these two challenges will be proposed in Chapter Five.

2.2.2 Methods and Practices

It is argued in [13] that the theory of performance improvement originates from three types of theories. The first is economic theories, which are the primary survival and driver for the success of organizations. The second is psychological theories, which view humans as the main source of productivity and trigger enhancements through behavioral and cultural changes. The third is systems theories, which recognizes processes, resources and subsystems that upgrade or degrade performance.

Despite the theoretical nature of the above classification, it reflects practical realities applicable to the field of incident response. For example, CSIR performance analysis heavily involves economic factors that focus on financial gain and loss due to incident handling. However, the performance of the responders themselves is inspected through reviewing relationship, team structure, communication and training. Also, the procedures and policies of the response capabilities are introspected to identify areas impacting performance. Therefore, the above work inspires studying performance through three dimensions: the system, economy and behavior.

A framework for incident response management for the petroleum industry, is presented in [105]. The study was triggered by the status quo of performance evaluation in the industry which is reactive in nature and limits the focus to technical measurements. The authors called for a more proactive socio-technical approach to performance evaluation.

The incident response model which the study analyzes is simple and involves three phases: (1) Prepare (2) Detect and recover (3) Learn. The prepare phase which exhausts most of the team’s time consists of conducting risk assessment, proposing plans and defining roles, raising awareness in the organization, and finally monitoring and adjusting to external dynamics. The second phase involves detecting unusual behavior, analyzing alerts, conducting initial assessment, executing a response, handling escalation and reporting. The final learning phase compromises of identifying sequences of events, stating root causes, proposing recommendations for security improvements and evaluating the overall incident handling process.

It is noticed that the overall structure of the response model is similar to the one used in CSIR. This similarity supports one of the motivations of this project, which is to make inquiry into performance evaluation through the multi-disciplinary approach. It is interesting to find that in the above study the authors called for using performance indicators throughout the response cycle, but did not provide details on how to achieve that. This academic gap seems to be present in several disciplines, and the field of CSIR is among them as highlighted in Chapter One.

In the domain of homeland security, the work of [103] provides in-depth analysis of how to measure the preparedness of an emergency system. The study borrows its analysis methods from the field of reliability engineering and risk analysis, and offers a detailed study case on emergency response to chlorine release. The study was mainly interested in answering the following question: is it possible to predict the performance of an emergency response system for future events? In the context of this research, this could be mapped to the question of how to predict the performance of an incident response plan (CSIRP) before its execution. Proper understanding of the above question can enhance responders' ability to resolve a common dilemma: what caused poor performance: poor planning or poor execution of the plan?

The study is remarked with its genuineness. Previous methods for measuring preparedness can be categorized under two main models. The first model focuses on measuring the capacity of the response system like how many responders are available and how much equipment will be deployed. The study criticizes this model as providing measurement to the inputs of the system not the system itself. The second model focuses on examining the performance of the system through actual utilization of the capabilities to produce response outcomes. Exercises and simulations are used as methods of evaluation. Again, the focus here is on the outcomes of the system, not the system itself.

The study provides another approach to measuring preparedness through what the authors called: “response reliability”. This metric questions how much confidence do responders have in the response system by measuring the likelihood that the response system performs well. Stated in other terms: what is the likelihood that events that prevent the system from performing well will not occur?

The study offers a four step model to measure response reliability (1) define and map the system (2) identify failure modes (3) assess probability of occurrence (4) assess effects and severity. Events are classified based on scale, scope and complexity, then the system is tested to determine the level in which the response plan may perform well in it but may not perform as expected above it.

The details of how chlorine release could be effectively handled and the statistical techniques used to test the equipment are not directly relevant to CSIRT. However, the study does offer an important insight which is to study the preparedness through the system inputs, outputs and system reliability. The current best practices in CSIR advises responders to suggest Plan B for what might not work well during the response. For example, a CSIRP should have list of secondary contacts to be reached if primary contacts were unreachable [12]. If the reliability dimension is added to the analysis of the response system, the system capability would be better captured and performance could be projected accordingly.

Managing critical incident responses on the national level in domains of natural disasters and terrorism were studied in [104]. The authors properly observe the need for distinguishing the management of a single incident from the management of several simultaneous incidents. To the best of my knowledge, this distinction is never made in the published works in the area of computer incident response (CIR), as the focus is on single incidents and the situation of multiple simultaneous incidents is not addressed. As the scale and complexity of breaches is on the rise, this issue deserves more attention, not only in the context of government agencies, but also in the corporate domain. A comprehensive study on protecting national infrastructure from cyber-attacks [112] properly notes that: "it is unlikely that a large organization would not have simultaneous attack scenario to face". The authors also decry the fact that "the notion of managing simultaneous response cases is largely unexplored in conventional computer security".

Another point raised by [104] is the call for the use of performance monitoring such that performance evaluation is viewed as a continuous management task that lasts throughout the life cycle of the incident response. The basic practice in CSIR views performance evaluation as a post-response activity. Although it is not necessary to provide continuous monitoring of CSIRT's performance during a response, it is valid to consider observing performance at different stages of the response. However, this should be applied cautiously as the overhead of continuous performance monitoring could hinder the core duties of the CSIRT.

Since one of the factors that impact the performance of a CSIRT's performance is the ability of the team members to interact with each other, enhancing performance through investigating team dynamics was presented in a recent study (July 2015) [97]. The study observes that there are three fields in which the responders act very similar to CSIRTs. The three fields are: emergency medical systems (EMS), military response (MR) and nuclear power plant operating (NPPO) teams. The authors investigated these fields to develop a model for boosting the performance of CSIRTs through enhancing the team dynamics.

The above study proposed a five-factor model for improving IR teams effectiveness. The five factors are: (1) adaptation (2) collective problem solving (3) trust (4) communication (5) shared knowledge of expertise.

The adaptation refers to the team's ability to efficiently perform under unexpected situations. The study suggests adopting two adaptation modules from the military domain: perturbation training and stress exposure training (SET). In perturbation training, trainees are put in simulated scenarios which are similar to real-life scenarios. The simulation is then repeated several times, in which the trainers disable one of the relied upon resources (e.g. access to phones or internet) to investigate how the team adapt to this unexpected event. Studies have shown teams exposed to this training outperform other teams by 13%. The second training module, the SET protocol in the military, aims at providing individuals with cognitive and behavior skills that help team members maintain effective performance under stressful conditions. The authors also suggested two other training modules from the EMS and MR fields to enhance the coordination of team members to solve problems together. I expect the above training modules to be excellent resources for CSIRTs, especially those operating in the national level or within very large companies. To avoid high costs associated with such training, it is necessary for the organization and the CSIRT team to identify sources of weak of performance and provide backed evidence from previous responses on the need to enhance team effectiveness through such training modules.

The issue of enhancing team communication is nevertheless an important aspect of team dynamics that is less costly but can reflect on tangible enhancements of team performance. The above study highlights that effective communication demands sending messages that are accurate, relevant and timely. To achieve that the study suggests three main methods. The first is to conduct a brief strategy meeting before starting a response. In the military, a 10-minute before start meeting enhanced team communication by 33%. The second method is to use uniform and brief sheets to communicate information when doing hand-offs, i.e. transferring a task from one person to another and from one team to another. In EMS, the use of mnemonics during handoffs decreased errors by 65%. The third method is using regular briefings during the response, similar to the one-to-four minutes briefings conducted in EMS. In these meetings nurses, surgeons, trainees and anesthesiologists communicate updates to an operation using checklists and very brief messages. This has shown to reduce communication failures by 64%. These simple and costless techniques can be adopted by CSIRT for better collaborative execution of CSIRPs.

Another recent doctoral dissertation (August, 2016) [102] investigated the role of volunteers in enhancing the performance of the response of emergency teams to complex incidents. The research focused on responses to earthquakes by analyzing online groups formed during the Haiti (2010) and Japan (2011) earthquakes and then testing the proposed findings during the Nepal earthquake (2015). The study found that communications resulting from volunteer-based online participatory groups were essential in enhancing effectiveness of the responses.

An in-depth theoretical and empirical study of the relationship between the formal organization response and the response conducted by online volunteers was presented. The study also highlighted that validating the credibility of information disseminated through online groups remains a challenge. The study ends with an important remark about the need to view incident response in the networked and digital era in different lens that transcends the current confined method to organizational response. From that perspective, it would be an interesting research work to investigate the role that online groups can play in CSIR. It is a double-edge sword. From one aspect, insights from bloggers and posts from security experts can bring important clues on responding to critical incidents. However, it is also possible that these posts are also read by the breaches themselves which can complicate the on-going breach.

2.2.3 Performance Indicators and Metrics

Watts Humphrey, a pioneer in software engineering said [113]: “quality management is impossible without quality measures and quality data.”. In the performance evaluation literature, the tools developed for measuring the quality of management and data are referred to as performance *indicators* and performance *metrics*. There is a variety of uses for the terms: “indicator” and “metric” across the disciplines and within each discipline. For example, the two terms can be distinguished by the methods used, e.g. qualitative and quantitative, where indicators are qualitative measures and metrics are quantitative, or vice versa. The terms also appear in various taxonomical order, sometimes indicators higher than metrics and sometimes the reverse order is used. Several works also use the terms interchangeably. However, despite the nuances both terms are used to reflect some form of measurement of a quality property. To avoid confusion, and within the context of this section, the two terms are left to their usage within each surveyed article, which might not necessarily conform to the definition outlined in Section 1.4.2.

There are numerous works that explore how to model and develop performance indicators and performance metrics for incident response systems. Focus will be on studies that provide methods for derivation and classification of performance indicators and metrics. In addition, some works that propose performance metrics that could be relevant to the area of CSIR will be surveyed.

The US Department of Transportation’s National Highway Traffic Safety Administration (NHTSA) conducted a five year study to propose a comprehensive list of performance indicators that measure the performance of Emergency Medical Service (EMS) [96]. Performance indicators were classified into seven categories: (1) system design and structure (2) human resources (3) clinical care outcome (4) response (5) finance/funding (6) quality management (7) community demographics. Each performance indicator (PI) is derived from a “Question” that captures some aspect of EMS, and each PI is mapped to one of three measurement types: structure, process and outcome. The type of dispatch system is an example of structure, percentage of patients receiving a specific treatment is an example of process, and the difference of first and last pain scale values is an example of outcome metric.

Question	A question in which an answer provides an aspect of the EMS.
PI Name	Name of performance indicator
Process Path	Which process and sub-process does the PI reflect
Patient Need	What patient need does PI reflect
Type	Measurement Type Structure, process or outcome
Objective	Why is this PI useful
Formula	The equation for the calculation of the PI
Sampling	Is sampling used? If yes, what is the sampling process
# data points	Minimum number of data points needed to calculate the PI
Format	Reporting format: Numerical or graphical
Testing	Formal structured evaluation of the PI (e.g. reliability, validity, difficulty of data collection)
Contributors	List of persons and organizations used in development of the PI
Current Status	Current development status: Amount of work completed to date relative to the final implementation of the PI

Table 8: Format for describing performance indicators in the field of EMS

The study captures EMS performance through 18 questions and 35 indicators. The study also develops a comprehensive method for describing performance indicators in a format that contains 24 “agreed upon” fields. A brief description of the main fields is presented in Table 8.

Another method for classifying performance indicators, which is used in several incident response systems [93] [23] [114], is borrowed from the field of economics [115]. Performance indicators are classified into three categories: leading, coincident and lagging indicators. A leading indicator measures the inputs, coincident indicators measure the system while running, and lagging indicators measure the outputs of the system performance. In [93], the study observed that most metrics currently used in the field of computer security can be classified as lagging indicators.

It is noticed that the difference between the above classification method and the one presented in [96] is only in terminology. Leading indicators are similar to PIs that measure the “structure”, coincident PIs are similar to PIs that measure the “process” and lagging PIs are similar to PIs that measure the output.

The classification of indicators into leading, lagging and coincident is more common in the performance evaluation literature. Nevertheless, classifying PIs based on their scope: structure, process and output is simpler and less technical. This might be appealing to the field of CSIR, as there are several non-technical stakeholders expected to interpret the performance evaluation findings.

Another study classifies performance measurements based on their subjects [116]. Measurements can relate to three different types of subjects: physical, ideal and social. The physical entities are the most common objects subject to measurement and they are confined by time and space. Ideal objects are theoretical objects that transcend time and space. Social objects are found within social constructions and are bound by time and space, like customer satisfaction. Using this classification, CSIR performance metrics would be mainly theoretical (security, impact) or social (team performance).

A survey conducted by [117] studied the performance measurements used in the transportation and emergency services fields across the US. The survey found that “response time” was the main tool used by both fields, although there was wide variance in how various agencies interpreted and defined “response time”. It was also found that performance measurements was used for different purposes. In the transportation field, performance measurements aimed at measuring the effectiveness of the overall incident management system, while in emergency services it was used as a resource management tool for allocating staff and equipment.

There are three relevant findings of the above survey worth highlighting. First, the survey found that all agencies did not create a separate cost budget for planning, monitoring and analyzing performance. All agencies treated performance evaluation as an integrated activity in the incident management system. Second, it was found that only one eighth of the agencies produced periodic performance reports. The rest stored the performance results in a database for access as needed. Finally, the study concluded that attempting to build "one-size-fits-all" approach to performance measurement is not feasible.

This last finding is relevant to this project. The transportation incidents demonstrate high similarity between incidents compared to the wide variance displayed in computer security incidents. If it is infeasible to build a unified PE model for systems like transportation, this would be extended to the discipline of CSIR.

A categorization of metrics for service level agreements (SLA) is provided in [83]. An SLA is a legal contract document that governs the relationship between a service provider with a customer. The document outlines the characteristics of the Quality of service (QoS) that the service provider is committed to deliver [118], along with remedial actions and penalties when the provider fails to keep up to the promised QoS.

In [83], performance metrics are viewed in, what the study calls, a multi-dimensional manner. Three dimensions were defined: service objects, ITIL processes and measurability. With regards to the first dimension of service objects, the study identifies five basic IT object classes: Hardware, Software, Network, Storage and Help Desk. A list of metrics for each object class, with a total of 45 metrics were given. Example metrics include: failure frequency, maximum down-time, availability, failure categorization degree and total service time.

The second dimension uses performance metrics for the eleven ITIL components. The Information Technology Infrastructure Library (ITIL) [119] is an industry standard for IT service management focusing on utilizing resources and enhancing services. The eleven levels, derived from ITIL v2 are: (1) service desk (2) incident management (3) problem management (4) configuration management (5) change management (6) service level management (7) service level management (8) capacity management (9) availability management (10) IT Service Continuity management (11) financial management. The third dimension, measurability, refers to the ability to automate the metric measurements. Three levels were defined: measurable, limited measurability and no measurability.

Studying metrics within in the SLA domain is relevant to CSIR. Indeed, CSIR can be viewed as a special type of service between the CSIRT and the organization, or as a special management process within the ITIL framework. Since the field of IT management is very mature and has gone through several standardizations, the experience of developing metrics offers a good source for CSIR performance metric development. In the above study, the issue of automating metric measurements stands out as relevant. This classification goes beyond the simple categorization of qualitative and quantitative metrics, to questioning the ability to automate the measurement and the analysis of results. Since CSIRTs work under stressful environments, the automation of collecting and measuring performance metrics would allow the team to focus on their main duties with little overhead for performance evaluation. Other than that, the concept of multi-dimension presented above is another fancy method of defining the attributes of performance metrics. The three dimensions can be mapped to the scope, objective and measurement method of the metric.

Two articles in the area of petroleum engineering [23] [105] addressed the issue of incident response in the domain of information security. In [23], the authors provided a list of performance indicators for incident response management. The paper derived the indicators from studies on safety management and from interviews with leaders of the Norwegian Oil and gas industry. The focus of the study was on the performance of management, i.e. the performance of the incident response system over a period of time in which several incidents took place.

The contribution of the above study of relevance to the proposed project is the discussion on how to evaluate performance indicators, i.e. how to know if the developed PIs are "good"? The study suggests five "metrics" to achieve that:

(1) being observable and quantifiable

(2) offers valid measurements

(3) sensitive to change

(4) compatible with other indicators

(5) simplicity

As the CSIR field works for developing its own performance evaluation frameworks, it is essential to ensure that the process of the design and derivation of performance indicators and metrics is sound. Metrics can be viewed as strong weapons that will be used to influence executive and financial decisions. If these metrics are not well designed and correctly used, it could lead to counter impacts on the organization. To avoid ambiguity, "metrics" used to evaluate the soundness of PIs and PE metrics will be called:

“meta-metrics”, a terminology adopted by several studies in software engineering [120] [121] and in computer security [122] [123] .

A study in the area of natural disasters [101] provided a list of metrics for describing incidents and assessing potential responses. Three metrics were proposed for describing a natural disaster and eight metrics to measure the effectiveness of the response. These metrics were designed to be used in computer simulation scenarios, but partially could be applied to real world scenarios.

One metric which stands out as relevant to this project is stability. In the context of natural disasters, stability measures the percentage of causalities per injury type whose condition worsens after being identified by the response team. It also records the percentage of causalities that were not identified by the response team. In the CSIR domain, the above measurement can be viewed through number of “breached” resources instead of causalities. In an ideal scenario, once a CSIRT identifies an incident, the process of containment should start and the incident should not cause further damage. This means that both the identification and containment processes should be perfect, which is practically very difficult to achieve. Since escalations are possible, if not common, the stability metric offers insights on its magnitudes and potential reasons for its occurrence. One simple method to achieve this is to record stability through recording the escalation or de-escalation of attack severity throughout the response process. This could be easily incorporated in CSIRPs as there have clear guidelines on how to define various categories of severity levels.

An article that investigates issues relevant to response to terrorist attacks [124], highlights three aspects that can significantly impact the effectiveness of the response. The first is “just in time logistics”, which refers to the ability of responders to arrive at the scene precisely when needed. The second is “situational awareness” which could be provided to responders through data collected from early warnings. The third is: “enhanced situational awareness” which provides responders with better decision making capabilities.

A study in the field of transportation engineering [77] investigates the performance of response teams to traffic incidents in the Portland (Washington State, USA) metro area. The study aims at studying how much the implementation of the current Portland freeway IR system contribute to the reduction of cost resulting from traffic delays and congestions caused by traffic incidents. Jumping over the technical details which are specific to freeway traffic, highlighting the mapping between cost reduction and the performance evaluation of the response teams is noteworthy. The authors noted that it is very difficult to create a dollar value to benefits of implementing performance evaluation.

It is estimated that 50% of the US highway congestion is caused by accidents. Through “faster” incident response, several benefits could be achieved like: reduced delay, reduced fuel consumption, improved flow of commerce, reduced harm to wildlife and water quality and improved public relations and good will. It is obvious that not all of these could be quantifiable in terms of cost, and hence the focus of the performance of the response team is on reducing the response time. The current cost of the response program is $750,000 and the cost of the performance monitoring system is $75,000. The authors argue that the success of a performance evaluation system is based on cost, i.e. does cost reduction exceeds the implementation cost of the system. For example, it was noted that faster response times resulted in cost reduction of annual vehicle hour delay by 13,000 hours which is mapped to cost savings of nearly $200,000.

In the field of CSIR, determining how performance evaluation could be mapped to cost reduction remains an open question. Although it is intuitive that faster response times to cybersecurity incidents lead to several benefits, there are no current studies that map the response time to cost reduction. It also poses as a more complex question compared to road traffic. For example, it is relatively simple to map the response time of a denial of service attack on an online store through calculating the average cost of purchases per minute. However, this is more challenging when an attack involves having access to customer private data, e.g. social security numbers and credit card numbers. It is very complex to map how a unit time in incident response would map to cost reduction. Therefore, it would be more suitable to use other factors besides cost. For instance, number of affected customers and the importance of the breached data should be used.

It is beyond the scope of this project to provide an empirical correlation between response performance and cost reduction. Nevertheless, the proposed PE framework can be viewed as an important milestone in the path of conducting more rigorous studies on that front.

2.3 Discussion and Lessons Learnt

2.3.1 State of the Art of PE in CSIR

Main Advancements of CSIR

The field of CSIR has undergone several important advancements in the past two decades. Three of these advancements are worth mentioning. The first is overcoming the obstacle of getting the community support to engage in the development and implementation of CSIR. Today, the vast majority of governments and corporates have well-formed CSIRTs and outlined CSIRPs. Although the preparedness of these CSIRTs vary from one place to another, there is a common acknowledgement of the importance of CSIRT and that the cost of functioning without one exceeds the costs allocated to having CSIR capability. It took much efforts from various experts in the field to convince executives and government policy makers to adopt organized CSIR processes, but also the damage caused by various cyber-attacks provided unopposed reality that signaled to managers that they cannot remain unprepared.

The second advancement is the development of several standards and industry documentations for the structuring and operation of CSIRTs. These documents not only facilitate implementation, but also engage the academic and professional communities in important discussions about issues of importance to the development of CSIR.

The third advancement is yet to ripe, but seems to be going in the right direction. It is the collaboration and sharing of information between various CSIRTs. The reality of the interconnected complex internet, in which cyber threats spread and impact all, mandates that all parties interested in security should have strong collaborations and timely robust exchange of information. With the current structure and activities of FIRST, the field is heading towards the right direction. However, investigating the bureaucracy and lack of effective communication between various government bodies is only one example of the amount of work that needs to be done before celebrating this advancement.

Focus on Preparation and Planning

In the past decade, discussion about the quality of CSIR performance started to appear. Prior to that, the field was immersed in laying out the main structures, procedures and processes needed for establishing CSIR capabilities. As of now, the field has passed debating about the benefits of integrating PE in the CSIR process. Many publications now consider PE as an embedded element of the CSIR process [1] [22] [73][25] [61] [41] [74].

Nevertheless, the field seems to focus on the preparation and performance readiness, which is a pre-activity that aims at enhancing performance through better planning. Examples of efforts in that direction include reviewing CSIRPs for completeness and offering training to CSIRT members for better coordination and team work. Indeed, that direction of research is of high value to the development of the field and offers several operational recommendations for CSIRTs.

What seems to be lacking is discussing post-activity performance evaluation. For instance, the need for developing performance metrics is acknowledged, but discussion about what these metrics are, how they could be used and the analysis methods to be used is scarce.

This emphasis on pre-incident compared to post-incident performance evaluation can be due to four factors. First, the CSIR field has developed the culture that successful response is a product of extensive planning. Therefore, it is natural that discussion about performance is done in the planning stage. Indeed, most responders spend most of their time in preparation compared to actual response, so evaluation is directed towards where most resources are devoted [7]. Second, it is possible that many organizations have integrated CSIRPE into their existing quality management system and which has a repository of performance measurement techniques. Third, it is possible that many CSIRTs have developed their PE analysis performs and pools of performance metrics. However, due to the security nature of the field, little has been shared with the public. Finally, it might be due to the natural evolution of the field, in which the need for performance evaluation tools became meaningful only when completing the planning and structuring processes.

The Challenge of Developing Security Metrics

As the development of performance metrics of CSIR highly depend on security metrics, the literature of security metrics needs to be studied thoroughly. The challenges highlighted by [34], were acknowledged along with additional challenges by several later studies like [125] and [126]. It is noticed that some of these challenges are intrinsic to the nature of studying security, like the element of uncertainty. I believe that such challenges will always be present, regardless of the advancements to be made. Such challenges can be termed as *static challenges*. In order to address static challenges, the research community would need to establish balanced protocols for developing, analyzing and later benchmarking data resulting from security metrics. Perhaps, the aspect to be given more attention should be the trade-offs between validity and practicality. By analyzing the CSIR literature, I endorse the current practice that grants precedence to practicality over validation because being operational is a resilient characteristic of computer security. However, there should be some minimum criteria for validation of metrics and data generated from them. This could be achieved by enforcing a meta-metric that specifies the confidence or the trust-level of the measurements obtained through these metrics. The model presented in [42] is an example of such research direction.

The other aspect of these challenges is *dynamic* and could be tackled with further advancements in the field. For example, the complexity of intermixed perspectives to analyze security metrics could be managed with careful design of multi-dimensional analysis platforms that are built on various perspectives. The same could be applied to the challenge of having conflicting objectives based on the requirements of various constituencies. Although the field is slowly consulting other fields for solutions like quality management, game theory and process management in industrial engineering, it is a matter of time until the field develops its own customized platforms of analysis. However, investigating the security metrics literatures makes the author to agree the following remark put forward by [42], which states: “It is anticipated that refining and adjusting the concepts of computer security assessment may take decades and in fact is a challenge for the entire generation”.

2.3.2 Lessons Learnt from Other Fields

Absence of generic methodologies

It is noticed that the various disciplines developed their own solutions to the challenges they faced when developing and conducting IRPE. These solutions are normally customized to the needs of the discipline in the form of models that suit the nature of the problem. This explains why the field of performance of incident response does not have generic models that are applied across the spectrum of various disciplines.

The lack of generic methods and models put the area of Incident Response Performance Evaluation (IRPE) in contrast to other areas of PE which develop and produce generic methods. For example, in computer networking, measuring the performance of a network is measured in tools like throughput, link utilization and packet delay. Scheduling mechanisms and queuing models are developed to enhance performance. Most of these models are generic and are used across various disciplines within and outside of computer science. This creates much interaction between these disciplines in terms of sharing ideas and borrowing solutions. This is different than the current state of art in IRPE in which each discipline seems to be operating in its own world with minimum interaction with the other disciplines.

Two questions arise: why does IRPE lack generic methods despite the extensive work done in various disciplines? And can such generic methods be developed. A possible explanation of the point raised by the first question is that the various IR disciplines have different objectives. For example, the main objective of medical IR is to minimize death and human pain, which is different than environmental IR which is to minimize environmental destruction. Both of medical and environmental IR are also different from help desk incident response which aims at increasing average customer satisfaction. Such unique objectives might have created a perception that the addressed problems are distinct and implicitly pushed against collaboration between these disciplines.

This leads us to the second question, in which I would answer by leaning towards the possibility of developing common models. It is true that the field of IR is very diverse which impacts how PE solutions are developed, but this does not eliminate the fact that there are various similarities that could be deduced from analyzing these disciplines. For example, all incident responders, regardless of their discipline, are interested in achieving faster response time. This does not conflict with the fact that each field has its own definition of response time that fits its context. I believe that some generic models could be developed to enhance team formation, incident declaration and management of major activities in the response cycle. The same could be said about enhancing the communications between responders during an incident and effective assessment of harm in a manner that transcends variance in team dynamics and types of harm.

To the best of my knowledge there is no generic work that discusses the main features and general solutions to performance evaluation in the context of IR. Also, as interest is rising with regards to conducting multidisciplinary research in the area of IR, this should be encouraged and pursued more seriously. It can be expected that findings from such research direction would be rewarding to all of these disciplines.

The Challenge of Complexity

It is observed, as stated in Section 2.2.1, that the two major umbrellas for challenges in the discipline of IRPE are complexity and unpredictability. Both of these challenges also pose as serious challenges to the field of CSIR. Complexity here refers to the presence of numerous contentious factors in the IR process which a CSIRT need to address at the same time, while none of these factors can be eliminated during the IR process.

There are several factors that contribute to complexity. For example, as the storage space of machines is on the increase and as networks are on high speeds, the amount of data present for analysis is gigantic. A responder can neither ignore gigabytes of log files nor neglect the scanning of a significantly large volumes of disk space for remnants of a breach. In that regard, scanning for a software bug in code with tens of thousands of lines is a relatively simpler task than the task of a CSIRT. Another analogy that demonstrates the complexity of a CSIRT operation, is to think of a medical doctor examining a patient suffering from a virus. Doctors normally need to examine the symptoms and recommend medications to counter these symptoms. Imagine if the doctor need to investigate how the patient came in contact with the virus before prescribing medications. There are thousands of possibilities and the prescription of medications become a complex process compared to focusing on confining the virus and treating the symptoms, regardless of its source. Most

CSIRTs need to properly assess the source of the incident and propose solutions at the same time, which is a non-trivial task.

The Challenge of Unpredictability

The element of unpredictability refers to the observation that regardless of how much preparation is being made, the element of surprise is always present. In the field of CSIR, examples of factors that contribute to unpredictability include the fast-paced development of new attack trends and compromise strategies, the difficult of distinguishing between false alarms and actual threat indicators, and the scaling power of cyber-attacks which lead to incident escalation.

The survey found that solutions developed by various disciplines to the challenge of uncertainty can be mapped to the following five main categories of solutions:

1- Designing probabilistic schemes that rely on measuring system reliability [91] and predicting response reliability [103]

2- Using confidence ratings to various decisions taken by the responders [42] [73]

3- Analyzing common threat trends and mechanisms and developing measures to confine variances and deduce similarities [106]

4- Producing response plans that demonstrates completeness [74]

5- Ensuring that responders are trained and demonstrate competency to carry response duties [76] [61]

The above five solutions can be further summarized into two main categories: preparedness and decision making capabilities in nondeterministic environments. These two strategies would be the basis of a framework to address the issue of unpredictability in CSIR which will be presented in Section 5.5.

Note that the term ‘nondeterministic’ is used here differently from how the term is used in computer algorithms [127], in which a nondeterministic algorithm is one that may produces different outputs for the same input, or that produces different outputs based on the enumeration of the set of inputs.

In addition, I argue that CSIR unpredictability is broader than systems that can be modeled through Markovian and stochastic processes. In Markov modeling [128], the system is modeled through states, and transition to the next state is made by probabilistic distributions based on the current available information. On the other hand, IR unpredictability exhibit different inputs that may lead to unknown new states, and consequently the output states may not be outlined through probabilistic distributions. Therefore, nondeterministic models can be as assisting analytical tools not as system models.

It is probably the field of decision making in economics and management that may provide viable recommendations to CSIR. The field discusses nondeterministic scenarios which managers might be faced with and provide analytical tools that can guide future decisions [129]. These tools normally rely on associating risks in the form of probabilities with possible decisions, but may also use other non-probabilistic approaches [130]. Approaches that use such techniques will be referred to as Non-Deterministic Decision Making (NDDM) models.

Is it possible to fully capture the performance of an IR system?

Before conducting the survey, it was anticipated to find a well formulated list of performance metrics that fully capture the incident response performance, especially in the fields like medical emergency which have long practical legacy. However, this proved to be an over simplification of the nature of incident response. It is noticed from the surveyed studies in various disciplines that there was no single study that claimed to provide an exhaustive list of performance metrics that fully assess the effectiveness of the response. Instead, the surveyed studies followed one of two approaches to performance design. In the first approach, an incident response system is analyzed to deduce the *main* duties and then design performance metrics focusing on these major activities. The other approach is to isolate a specific activity or mode of interaction within the incident response system and subject it to extensive review for proper performance measurement. In both approaches, there are aspects of the system which are not subject to performance measurement.

There is something to learn from the above experiences. Instead of attempting to provide precise and complete performance evaluation systems for CSIR, the focus should be shifted towards proper identification of aspects in the response system which need to be subjected to performance analysis for potential improvement. Expecting precise and complete performance evaluation is both expensive and impractical. Thus, the field should ask the following two questions: "what aspects of CSIR should be subjected to performance evaluation?" and "how precise the measurements should be?".

Simple Factors with High Impact

The literature review of both the CSIR and IRPE suggests that there are some factors that need to be thoroughly examined by those interested in developing PE models for CSIR. These factors can be easily confused to be simple and easy to handle, but practically they are complex to measure and analyze. For example, estimating financial loss due to an incident is a very difficult task contrary to how it appears. For example, estimating financial loss due to a vehicle accident is not only limited to damages to vehicles involved in the accident. The financial assessment needs to include estimations of financial loss due to traffic delays which is not a simple task. The same extends to CSIR as estimation financial loss should take account of non-trivial factors like customer trust, privacy violation and public perception.

In Table 9, four issues are identified. These issues are highlighted to be non-trivial tasks of IR as discussed by various disciplines. It is expected that these four challenges will also be non-trivial to the field of CSIR.

Issue	Description
Threat Assessment	How can the actual threat of reported events be precisely assessed? Under which threat level should the incident be classified?
Harm Estimation	At the end of the IR process, how can the actual caused damage be measured?
Communication Effectiveness	As any IR involves high volume of communication between internal and external members, how can the communication effectiveness be measured?
Overall Performance	Since no single performance metric can be used to measure the effectiveness of an IR, how can the results obtained from various metrics be used to provide an overall assessment?

Table 9: Summary of Factors with high impact on CSIR performance evaluation

CHAPTER THREE

FRAMEWORK DESIGN

This chapter presents the theoretical aspects associated with developing frameworks and models for evaluating the performance of a computer security incident response (CSIR) capability. The design considerations and possible development and functional models are discussed. The chapter can be viewed as a design guide that an organization can use to develop its own performance evaluation framework, from the early stages of planning to the actual integration of the framework into the CSIRP. It also can be viewed as a research map of the issues pertaining to CSIRPE that deserve the attention of researchers in the field.

The framework design presented in this chapter is not an extension of a specific work or model. It is a result of the synthesis of the large volume of works surveyed in Chapter 2, both within the field of CSIR and in the other disciplines. It is also guided by the feedback received from professionals in the field about the current practices in the industry. Whenever possible, references will be provided in support of the selected design choices or when further extensions of the discussion are required. However, because the chapter addresses plentiful of issues more emphasis is put on the presentation of the framework components compared to expansion of various possibilities.

The chapter is composed of six sections. The first section provides an overview of the main components used for describing and constructing a CSIR performance evaluation framework. It also highlights the relationships existing between these components providing a holistic perspective to the development process. The second section presents the conceptual assumptions of how CSIR is understood for the purposes of this project. The following four sections provide details for each of the four development phases, where section three addresses the first phase about designing the CSIRPE framework through setting parameters and layout strategies and policies; section four describes how to formally define a CSIRPE model through defining PE goals, PE aspects and performance indicators; section five tackles the issue of measuring performance by discussing performance metrics, analysis methods and validation methods; and the sixth and final section discusses implementation considerations.

3.1 Framework Overview

3.1.1 Framework Components

In order for an organization to develop a computer security incident response performance evaluation (CSIRPE) framework, the organization needs to go through several planning and implementation stages. These stages can be captured in eleven main steps, grouped under four major development phases, as depicted in Figure 8. Each of these steps are referred to as a CSIRPE framework component, or as a PE development step. Below is a brief description for these eleven components:

1. ***Design Parameters:*** Generic parameters which values significantly impact the environment in which PE will be introduced. These parameters can be perceived as factors that can produce various types of PE frameworks.
2. ***Strategies & Policies:*** Approaches to address the issue of PE within the context of CSIR, which are not delimited by simple values. This include paradigms of understanding PE, strategies for addressing CSIRPE and generic policies that are necessary for any successful PE design.

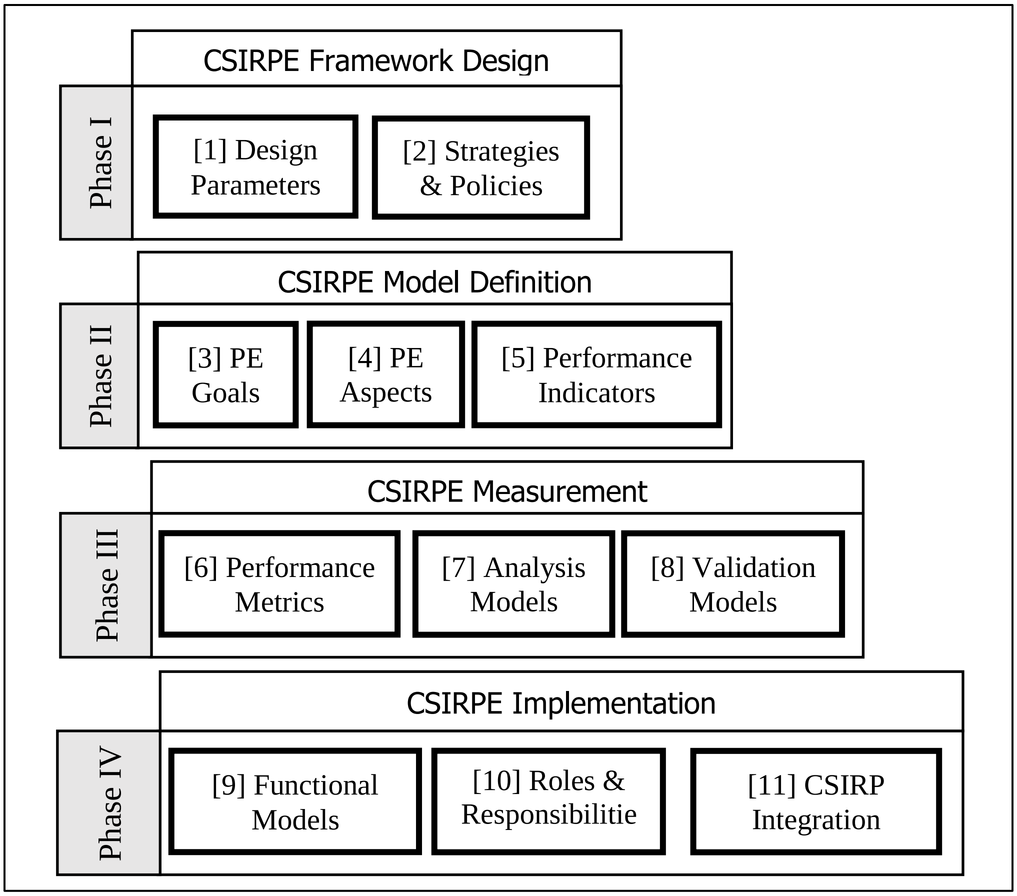


Figure 8: CSIR Performance evaluation framework Components

3. ***Performance Evaluation (PE) Goals:*** The process of defining the performance evaluation goals in terms of what ideally should be achieved and the performance needs of the CSIRT
4. ***Performance Evaluation (PE) Aspects:*** Identifying parts of the CSIR system that need to be subjected to performance measurement
5. ***Performance Indicators (PIs):*** Identifying the desired performance qualities of the system, and factors that determine the quality of the response
6. ***Performance Metrics (PMs):*** The process of deriving and defining measurement tools for assessing various performance indicators

7. ***Performance Evaluation Analysis Models:*** The selection of analysis methodologies for how performance evaluation metrics results need to be interpreted to reflect the performance of the response.
8. ***Performance Validation Models:*** The process of ensuring that the performance measurements are valid and that it will result in correctional activities that enhance performance.
9. ***PE Functional Models:*** Models that determine how performance evaluation activities will be integrated into the incident response life cycle.
10. ***Roles and Responsibilities:*** The process of outlining tasks and assigning performance evaluation responsibilities among the response teams
11. ***CSIRP Integration:*** The process of integrating the designed performance evaluation model into the response plan, i.e. CSIRP.

3.1.2 Framework Development Process

The eleven framework components can be conceptualized as a product life-cycle process starting from a generic performance evaluation framework to a final output of a performance evaluation model implemented in a specific environment. A depiction of this development process is presented in Figure 9. A description of the inputs and outputs of the various development stages is presented below:

Below is a description of the inputs and outputs for the seven components:

- ***CSIR Capability:*** A computer security incident response capability adopted by an organization, that appears in the form of a CSIRT and an approved CSIRP, which lacks mechanisms and tools for measuring the performance effectiveness of incident handling.

- ***CSIRPE Framework:*** A performance evaluation framework customized through specific values of the design parameters and governed by formulated strategies and policies to suit the nature and needs of a specific environment.

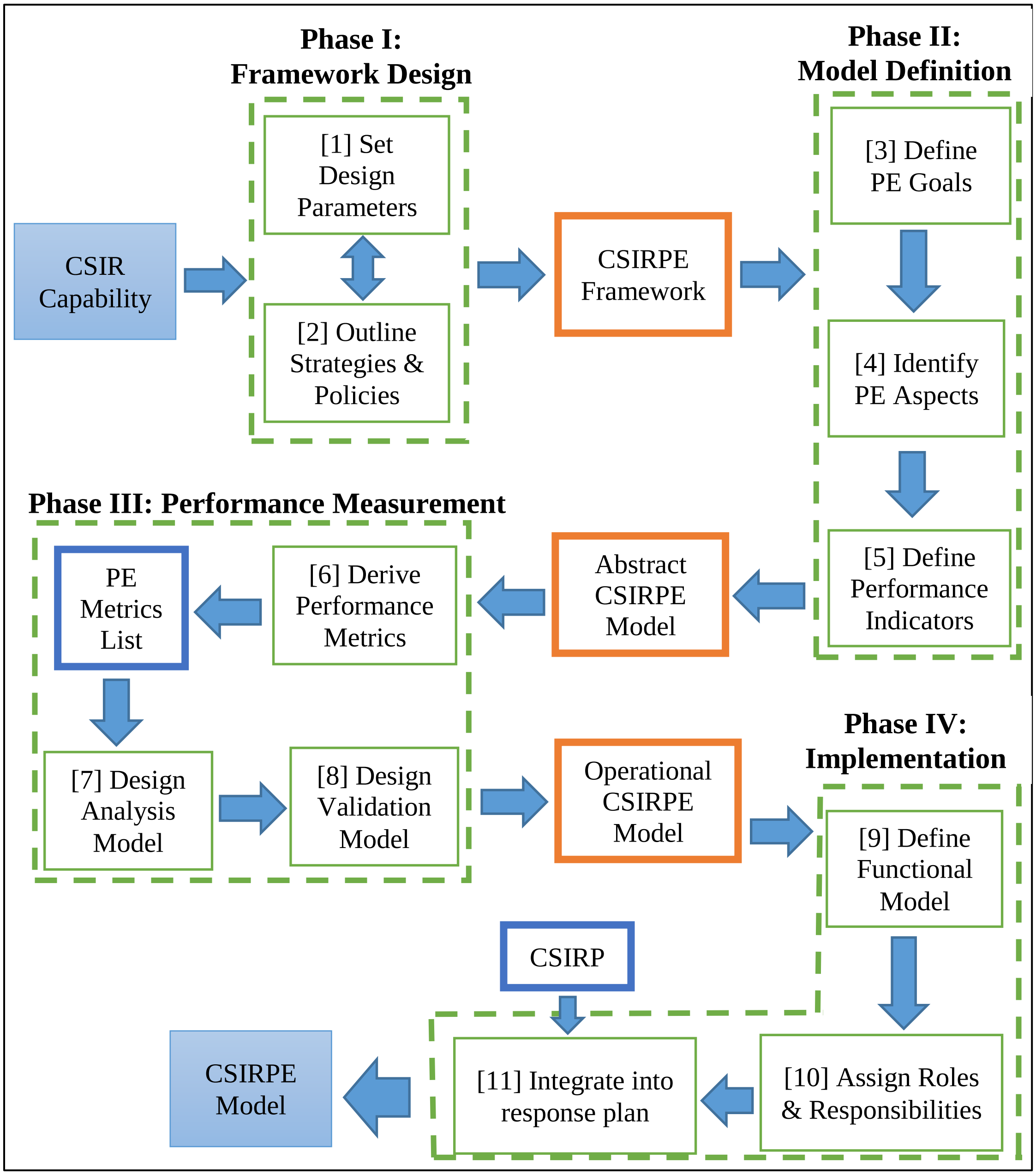

Phase I:
Framework Design
Phase II:
Model Definition
[1] Set Design Parameters
[2] Outline Strategies & Policies
CSIR Capability
CSIRPE Framework
[3] Define PE Goals
[4] Identify PE Aspects
[5] Define Performance Indicators
Phase III: Performance Measurement
PE Metrics List
[6] Derive Performance Metrics
Abstract CSIRPE Model
[7] Design Analysis Model
[8] Design Validation Model
Operational CSIRPE Model
Phase IV:
Implementation
[9] Define Functional Model
CSIRP
[10] Assign Roles & Responsibilities
[11] Integrate into response plan
CSIRPE Model

Figure 9: Performance Evaluation Framework Components Relationships

- ***Abstract CSIRPE Model:*** A theoretical model derived from a predefined framework, that outlines why and what needs to be evaluated within a specific environment. The model is abstract as practical tools that specify how the evaluation will be carried are absent.

- ***PE Metrics List:*** A list of performance metrics, i.e. performance measurement tools, produced to be integrated into a predefined *Abstract PE Model*.
- ***Operational CSIRPE Model:*** a practical model for performance measurement that have defined objectives and operational mechanisms for how these objectives will be achieved in a specific environment
- ***Performance Evaluation Model:*** The final product of the performance evaluation design process which appears in the form of a model equipped with clear objectives, measurement tools and implementation guidelines.

The following four sections will examine each step of the development process. To enable accessible referencing to various aspects of each framework component, the codes introduced in Table 10 will be used for the rest of this dissertation.

Code	Category	Code	Category
A._	Assumption	M._	Metric Design Parameter
D._	Design Parameter	N._	Analysis Model
F._	Functional Model	S._	Strategy
PI._	Performance Indicator	V._	Validation Model

Table 10: Framework Components Codes

3.2 Assumptions

Since there are various understandings about the nature of CSIR and the tasks assigned to CSIRTs, the basic assumptions about how CSIR and the environment in which

CSIRTs operate as understood in this project are outlined in this section. These assumptions are referred to as *CSIR assumptions*, and are summarized in Table 11.

#	Name	Description
A1.1	Civilian Environment	The CSIRT operates in a non-cyberwar environment in which responses are limited to defense compared to a defense-counterattack model
A1.2	Team Structure	The CSIRT is composed of at least two members, compared to an individual tasked with handling IR responsibilities
A1.3	Incident Complexity	Incidents demonstrate some complexity and impact beyond simple security incidents
A1.4	Incident Handling Services	The CSIRT responsibilities are focused on incident handling response and only indirectly tasked with system security
A1.5	Computer Security Incidents	The CSIRT is responsible for handling incidents with some security element compared to regular computer incidents
A1.6	CSIR Capability	The organization either has an established process for handling computer security incidents or acknowledges the importance of CSIR and allocates reasonable financial and human resources.
A1.7	Secure Environment	The CSIRT operates an environment that at least have minimum measures for securing the environment

Table 11: Global CSIR Assumptions

The main objective of presenting these assumptions is to alleviate any ambiguity on how CSIR is understood compared to the several understandings available in the literature. The assumptions can also be viewed analogous to environment assumptions that are normally presented before network models in the computer science literature.

Below is a description of the seven assumptions.

A1.1. Civilian Environment

It was argued in Section 2.1.1 that with *Stuxnet*, the field of CSIR is challenged with a serious level of sophistication. The issue of cyberwar, regardless of how it is defined, is extensively being discussed in the media, among security professionals and government officials. I anticipate that the field of CSIR will be distinctly split into civilian and military like domains, a reality which might have already taken place. The main distinction between the two domains is not limited to the scope level, i.e. major vs. minor incidents, but also extends to the operative method. Traditional CSIR works in defensive mode, while in cyberwar environments the mode of operation is defensive-offensive. For the scope of this project, it is assumed that the CSIR operates in a non-cyberwar environment.

A1.2. Team Structure

Although in some organizations, especially those of small size, CSIR handling can be delegated to a single individual, the reality of security threats mandate the involvement of a team in the incident handling. The field of CSIR acknowledges this fact by giving the term: CSIRT (CSIR *Team*) to responders to emphasize the collective and cooperative nature of the response. The NIST standard [9] highlights this aspect through the definition of a CSIRT which starts with: “A group of individuals”.

Investigating the performance evaluation in a team context is more comprehensive than evaluating the performance of individuals. Perhaps, individual performance evaluation can be viewed as a sub-category of team performance evaluation.

Estimations on the minimum needed number of team members vary. For example, the SANS publication of [131] considers the basic structure of a team to be of two dedicated CSIR members reporting to the CIO, and four support individuals are needed during the incident: network engineer, public relations, legal and human resources. A similar argument is made by [26] for three full time members or 5+ part timers. In order to avoid being restrictive, the most stringent requirement for the number of responders put forward by some works like [67], which is two members, will be used. One of the members should have knowledge of the business and management structure of the organization, and the other of the IT and network infrastructure. Another way to view the two individuals as a team leader and a team member.

Therefore, it is assumed that whenever the term CSIRT is used, the team consists of at least two members.

A1.3. Incident Complexity

This assumption is presented here to emphasize the adopted definition of an incident as outlined in Section 1.4.1. An incident is assumed to demonstrate either a need for sophisticated and coordinated response or significant potential harm to the operations or the data of the organization.

A1.4. Incident Handling Services

There is a wide range of services that a CSIRT can offer to an organization. The CERT document [2] classifies CSIRT services into three main categories: reactive, proactive and quality management; see Table 12. For the purposes of this project, the focus is on the performance quality of the reactive services of a CSIRT. This is consistent with the NIST document [1], page 23, which states: “It is outside the scope of this document to provide specific advice on securing networks, systems, and applications. Although incident response teams are generally not responsible for securing resources they can be advocates of sound security practices.”

It is assumed that whenever performance evaluation is mentioned then it refers to the evaluation of services that are of reactive nature and directly relate to incident handling.

Proactive and quality management services are not considered part of the scope of the study. The evaluation of these services can be done through various existing performance evaluation tools, which are independent of the technical aspects of CSIR.

Reactive Services	Proactive Services	Security Quality Management Services
Incident Handling Incident analysis, IR on site, IR support, IR coordination)	Announcements & Technology Watch Security related information dissemination	Risk Analysis Business Continuity & Disaster Recovery Planning
Vulnerability Handling (analysis, response, IR coordination)	Security Audit & assessment	Awareness building Education & Training
Alerts & Warnings Artifact Handling (analysis, response, coordination)	Development of security tools Configuration & maintenance of security infrastructure & tools	Security consulting Product evaluation or certification

Table 12: Summary of CSIRT services outlined in the CERT document

A1.5. Computer Security Incidents

This assumption is put forward to support the definition of a computer incident provided in Section 1.4.1, which limits the scope to incidents of security nature. It also relieves ambiguity arising from other definitions of computer incidents that include general computer incidents or even non-computing incidents, like the definition used by the SANS publication of [132].

A1.6. CSIR Capability

The development of CSIRPE assumes the pre-existence of a CSIR capability within an organization. If an organization lacks such capability, it is infeasible to discuss quality management and performance evaluation methodologies. However, the presentation of CSIRPE in this work would be of interest to an organization that lacks such capability but is interested in developing one to be equipped with quality measures and PE capabilities. This would need to be reflected in allocation of sufficient funds, gaining management support, and the technical teams acknowledging the importance of incorporating PE as an element to the planning process.

A1.7. Secure Environment

This assumption complements A1.6 by stressing the need for a pre-existing security measures within an organization that is interested in developing a CSIRPE. Having valid and enforceable security policies is a pre-condition for any effective implementation of any CSIRT [26] . The term “secure environment” is used loosely here to refer to the existence of some security measures in the environment that would enable CSIRTs to respond, like logs, IDS, malware detection tools …etc. With the absence of such measures, a CSIRT would be focusing on “securing” the system more than handling an incident, which is contention with A1.4.

3.3 Phase I: Designing a CSIRPE Framework

3.3.1 Setting Design Parameters

At the early phase of constructing a performance evaluation framework for any given environment, there are important decisions to be made which shall shape how the whole framework is structured and developed. These decisions are formulated in this section in the form of design parameters to be presented to the CSIRT members engaging in the planning process. The way how these parameters is set will significantly impact the rest of the design process.

The design parameters were selected such that the values of each parameter could possibly produce a different PE framework. For example, a PE framework that evaluates the performance of a centralized CSIRT demonstrates significant differences than a PE framework for a distributed CSIRT. Similarly, a framework that evaluates incident handling of concurrent incidents has major considerations different than one that evaluates single incidents.

#	Parameter	Description	Values
D.1	CSIRT Type	Will CSIRPE be applied to a centralized, distributed or a customized organizational CSIRT model?	Centralized Distributed Customized
D.2	Evaluator Type	Who is evaluating the CSIRT? Is it the CSIRT performing self-assessment, an internal team within the organization, or an external auditor?	CSIRT Internal External
D.3	Number of Incidents	Does the PE model evaluates individual incidents, several incidents over a period of time, or both?	Single-incident Multiple-incident Adjustable-window
D.4	Incident Concurrency	Is the PE model applicable to single, simultaneous incidents, or both?	Sequential Concurrent Elastic
D.5	Analysis Time	When will the CSIRPE be conducted and translated into corrective actions?	Post-incident Continuous Incremental
D.6	Benchmarking	Will the performance evaluation results be used for benchmarking with results from other departments or institutions?	Standalone Internal-Benchmarking External-Benchmarking
D.7	Measurement Type	Will the performance model use quantitative, qualitative performance metrics or both?	Quantitative Qualitative Both
D.8	CSIRP Scope	Does the PE framework evaluates the effectiveness of the plan design, plan execution or both?	Design Execution Both

Table 13: Framework Component 1: Design Parameters

A list of eight design parameters along with brief description is collected in Table 13. Each design parameters is presented with several values, which can be viewed as performance design decisions to be made by the CSIRT.

D.1. CSIRT Type

The way how a CSIRT is structured has an impact on how it would be evaluated. Therefore, the three structural models: centralized, distributed or coordinated will be mapped to three possibilities for performance frameworks. These frameworks will mainly differ in the selection of performance indicators, functional models and how analysis is conducted.

Practically, the two distinct performance frameworks are the centralized and the distributed. The coordinated model can be viewed as a customized model that derives its features from the two frameworks. To be more generic, the third value has been designated as: “customized” instead of “coordinated” to include other structural models that have features from both the centralized and distributed models.

Compared to centralized framework, the distributed framework is expected to display the following distinctions:

1. *Coordination Effectiveness:* Major part of the performance measurement would focus on measuring the effectiveness of the coordination channels between various parties. For a distributed execution, the coordination is sometimes as important as the actual actions taken by the CSIRT members.
2. *Communications Effectiveness:* Instead of assessing the effectiveness of a centralized command flow, the communication messages need to be examined for potential redundancies and undue overhead. Unlike coordination effectiveness

which inspects the ability of various members to work together, communication effectiveness inspects the contents of the messages and how they are exchanged.

3. *Distributed Analysis:* Each "distributed" entity needs to have its dedicated performance evaluation. Next, the whole response system need to be assessed. The analysis need to identify if performance deficiencies are due to unit or system issues.
4. *Multi-context analysis:* In the distributed model, the evaluation should consider that various parties have different levels of access to information. Therefore, the analysis need to consider multiple contexts. Also, the protocols for establishing trust and confidentiality among various parties would need careful assessment.

Only recently few publications started to examine distributed CSIRTs [133] [17]. This explains the focus of this project on centralized CSIRTs, as specified in assumption A2.3. Awaiting further publications in that area, assessing the PE of the distributed mode of operation can be an extension of this work in a future project.

D.2. Evaluator Type

The entity providing performance evaluation of the CSIR capability can be the CSIRT itself, an internal body in the organization other than the CSIRT, like the quality unit, or an external body. Each choice comes with advantages and disadvantages, and will also impact how the CSIRPE is designed and implemented. A comparison between the three methods is provided in Table 14.

It is noteworthy to mention that this design parameter should be inspected while discussing options for S.1 (CSIR Quality Assurance) and S2 (CSIRPE Quality Control). The three design considerations are interconnected, but each demonstrates slightly different concentration. The CSIR quality assurance focuses on the constituencies' perspective, the CSIRPE quality control on the organization's perspective, while D.2 focuses on the CSIR itself. In addition, S.1 and S.2 are mainly pre-incident activities, while D.2 focuses on the activities during and post to incident handling.

	CSIRT	Internal Unit	External Body
Cost	Low	Low	High
Duration	Relatively short as responders are acquainted with incident details	Depends on the quality management system of the organization	External auditing is a long process
Neutrality	Can be subjective due to "gaming of numbers" [134]	Can be subjective due to emphasizing organizational aspects over technical	Objective
Analysis Focus	Technical and self-learning	Organizational, financial and procedural	Depends on auditor emphasis
Planning Overhead	High, endured by the CSIRT	High but collaborative between various departments	Minimum or none
Ownership	Fully owned by a CSIRT. Ownership of PE enhances positive change [135]	Ownership by quality unit can be mis-interpreted as "finding someone to blame"	No ownership by CSIRT. Can be seen as imposing irrelevant factors

Table 14: Comparison between evaluator types design options

Finally, the "evaluator type" has been presented as a design parameter, compared to a strategic aspect, to highlight that it is an essential decision to be made at the planning stage. It is applicable independent of the absence or presence of quality assurance and quality control capabilities at the organization.

D.3. Number of Incidents

A performance evaluation framework can be designed to evaluate single incidents, i.e. without analyzing the CSIRT's performance over multiple incidents. It could also be designed to evaluates performance over a period of time that constitutes incident handling of several incidents. A more desirable but more complex option is to design a framework that is capable of evaluating single and multiple incidents.

The first option of standalone evaluation platforms is suitable for several contexts. For example, it is appealing to "CSIRT Providers" [136] that need to report their performance to their clients after concerning each incident handling. In such scenarios, the client is interested in seeing metrics that reflect that the delivered CSIRT services met the expected quality of service requirements. The client is not concerned about how the CSIRT performed with regards to other customer's, unless it is presented in a context of comparison confirming that the delivered service is on the same level to that delivered to other clients.

Another potential application for single-incident platforms is special taskforce CSIRTs or ad hoc CSIRTs that are formed to handle a special incident. In such context, there is little interest in how responders perform outside the assigned mission. The same is applied to a CSIRT responding to a unique incident that is rare to happen and require substantive allocation of resources. Here, due to the infrequency of such incidents, analysis over multiple incidents or over specific period of time is infeasible.

There are two main challenges in designing single-incident platforms. The first is the absence of data from other incidents in which the current PE results could be compared to. Therefore, performance metrics need to be designed such that to reflect achievement of pre-defined goals or the fulfilment of specific quality of services requirements outlined in a contract.

The second challenge is the potential of conducting non-objective analysis. The analysis would normally involve analyzing outputs of numerous descriptive metrics. There is nothing to guarantee that how selected or excluded readings from analysis would impact the validity of results. In addition, it is difficult to argue that specific achievements are due to good performance or due to external factors that are distinct to the incident being handled.

The second approach is to design performance platforms that analyze the team's performance over multiple incidents. This seems to be the approach used in the CERT's publication of [60] which implies that PE is measured over multiple incidents over a period of time. This is also the method commonly used for evaluating help-desk teams and IT services that are provided in high frequencies [23] [82] [83].

In multiple-incident platforms, performance metrics would normally exhibit some statistical nature. Performance metrics can be averages, means, medians, variances and frequencies. In such metrics, it is important to define the contexts in which the statistics would provide meaningful results.

For example, does an average response time demonstrates performance if measured over three incidents? Does it provide meaningful reading if not read along with minimum, maximum and the variance values? If the measurement context is not pre-defined, the analysis can also fall short to the non-objectivity drawback.

Examples of contexts that are suitable for multiple-incident platforms include CSIRTs that need to report their performance in the form of annual reports or over a specific number of years. Another example is performance reports that are prepared for strategic planning and budgeting which focus on the overall performance more than results of specific incidents.

A comparison of the advantages and challenges of both approaches is provided in Table 15.

Issue	Single-Incident Framework	Multiple-Incident Framework
Suitable for	Special CSIRT taskforce PE of unique and rare incidents requiring unusual amount of resources CSIRT Providers reporting to each customer	CSIRTs with high number of incident handling CSIRTs that need to report in periodic number of years PE reports needed for strategic and budget planning
Type of PMs	Non-statistical metrics	Statistical metrics like averages and variances
Objectivity	Measures PE through achievement of pre-defined goals and quality of service requirements	Measures collective PE that encompasses all incidents. PE is measured under similar environment conditions
Subjectivity	A response PE can display achievement & quality of service but is poor when compared to other responses	Results of PMs might fail to report poor performance in several areas if average readings are acceptable

Table 15: Comparison between single and multiple-incident PE frameworks

An *adjustable-window* framework is one that is designed to provide performance evaluation for single incidents and also time-analysis for several incidents. In an ideal situation (CSIR-UPEF), the window could be focused into a single incident or expanded to include as many past incidents as needed. This approach combines the advantages of both of the two methods, but would require more resources for the planning and measurement collection.

Some practical guidelines concerning the design of an adjustable-window framework are provided in Table 46 in Appendix A.

D.4. Incident Concurrency

The response techniques of handling incidents that appear simultaneously is different than response techniques to that handle incidents one at a time [104]. This requires that a PE framework that allows for evaluating concurrent handling be supported with different performance measurement mechanisms than a PE framework that is confined to single incidents.

As noted in [112], the current CSIR literature lacks research in the direction of simultaneous handling of incidents. But it is expected that the research community will soon respond to the government and industry needs. For example, the National Response Framework (NRF) requires that planning should assumes: "multiple catastrophic incidents or attacks that will occur with little or no warning" [110].

To contribute towards building the taxonomy, a PE framework that focuses on evaluating responses to incidents that appear exclusively will be called a *sequential PE Framework*. A framework that inspects the PE of handling concurrent models will be called a *concurrent PE Framework*. A model that is flexible to handle both types will be referred to as the *elastic PE Framework.*

Unlike single incidents, concurrent incidents require responses with efficient resource allocation, multi-incident management, priority scheduling, and higher coordination mechanisms. A summary of how this will impact PE design is demonstrated in Table 16.

Issue	Sequential PE Model	Concurrent PE Model
Utilization	The utilization of resources is measured against the maximum potential	The utilization of resources is measured against maximizing overall performance
Scheduling	Not Applicable	The effectiveness of scheduling mechanisms need to be evaluated
Prioritization	Basic evaluation of the CSIRT's ability to manage several tasks sharing same goals	Multi-layer evaluation of the CSIRT's ability to perform tasks pertaining to contending goals, i.e. evaluation of priority classification and scheduling
Contextual Analysis	The context of PE analysis will focus on the available resources	The context of PE analysis will include available resources and presence of the other incidents

Table 16: Comparison between sequential and concurrent PE models

D.5. Analysis Time

This design parameter outlines the time period, in reference to incident handling life cycle, in which the performance evaluation will be conducted. This will directly impact the functional performance model and consequently the analysis techniques.

The most commonly used design approach is to limit performance activities during incident handling to collecting measurements and defer the extensive analysis to the post-incident period. The second approach is to provide continuous analysis during incident handling, which is known as performance monitoring. A third approach, the incremental model, performs partial analysis at specific milestones of incident handling and completes the analysis post-incident. The details and implications of adopting each of these three choices is deferred to Section 3.6.1 when discussing functional models.

D.6. Benchmarking

Benchmarking is one of the most common management tools for measuring performance. It provides performance analysis through comparing the performance of an organization to the best practices of the industry or to is competitors [137]. Such approach is proven to encourage competition which leads to performance enhancement [134]. As there are no public CSIR benchmarks; the term ‘benchmarking’ is used broadly to refer to the collection of activities of comparing a CSIRT’s performance to that of other similar teams, internal or external.

A PE framework that uses benchmarking methodologies is expected to be different than one that does not. When benchmarking is used, an organization will frame its PE methods and metrics around those used in the benchmark. Although this would be restrictive design aspect, the main advantage is the value of the results which are viewed with more validity. On the other hand, a framework that does not use benchmarking would be more flexible and adaptable but would be challenged with a more complex process to provide objective analysis.

A benchmark can be internal or external. It is argued in [106] that internal benchmarking is more beneficial than external benchmarking, a point of view that I support considering the current state of the field. Three main arguments are in support of this view. First, benchmarking analysis is sensitive to numerous parameters of the incident’s environment which are difficult to normalize across diverse environments. Second, several aspects of performance are suppressed when using public benchmarks due to confidentiality and privacy considerations. Third, public benchmarks, especially those designed by commercial entities, tend to give more weight to market demands and investment issues over technical issues. If these three challenges are addressed, then external benchmarking would be a viable option, probably with greater benefits.

Finally, it should be highlighted that this design parameter and *D.3: number of incidents* are mutually exclusive. A single-incident PE framework could be designed with or without benchmarking. For instance, a single incident can be evaluated against a benchmark, and multiple incidents can be analyzed independently without comparison to that of other entities. The same argument is applicable to multiple-incident PE frameworks which can be designed with or without benchmarking.

D.7. Measurement Type

One of the most common classification methods for performance measurements is to group measurements into two categories: quantitative and qualitative [85]. Several disciplines like scientific measurement and financial analysis rely solely on quantitative measures, while other disciplines like sociology and politics extensively use qualitative measures in the form of surveys and polls. The performance measurement disciplines use a mixture of both methods but with slight preference towards quantitative measures.

A summary of the advantages and disadvantages of qualitative and quantitative measures [138] [139] is provided in Table 17.

In the field of CSIR, the main design decision would be whether to include qualitative measurements or not. Meaning, a CSIRPE would either be fully quantitative or be equipped with measurements methods from both types. It is unlikely that an effective PE platform would be designed using only qualitative measures. Therefore, the question is narrowed down to: is it better to design a fully quantitative framework, or should qualitative measures be used?

Answering the above question exposes the polarized viewpoints on a greater debate which the CSIR field inherits, which was highlighted in Chapter 2. Practically, most of the metrics used in the field of computer security are quantitative; however, serious concerns have been raised about forcing quantifiability into the notion of security [34] [42]. Another survey demonstrated that schism exists in the research community concerning the benefits and shortcomings of introducing or excluding qualitative measures [139]. The survey studied the value of IT in organizational performance, a topic of similarity to this project.

	Qualitative	Quantitative
Output form	Text	Numbers
Usage	Understanding, exploring possibilities, analyzing behavior, focused analysis	Fact documentation, measuring outputs, observations, patterns, generalizations
Examples	Satisfaction, trust, transparency, communication, coordination, consistency, traceability (documentation)	Response time, cost, frequencies, counting (e.g. occurrences, infected records), utilization, availability.
Verifiability	Relatively more difficult to verify	Generally easy to verify
Complexity	Normally, simple design but complex analysis	Normally, complex design but simple analysis
Objectivity	Viewed to be more subjective	Viewed to be more objective
Measurement process	flexible and may vary from instance to instance	Well defined and uniform
Automation	More difficult to automate	Relatively easier to automate

Table 17: Comparison between qualitative and quantitative measures

The NIST document [1] sided this debate by dividing assessment methods into subjective and objective measures, without mentioning qualitative and quantitative measures. However, this implicitly acknowledges the need for both, as qualitative measures are normally subjective, and vice versa.

Following the above spirit, I second that both types of metrics should be used as needed. If there is an issue that is naturally easier to understand and analyze through qualitative measures, or if the quantification process results in a complex solution, then use of qualitative measures should not be discouraged. An example of this is metrics that aim at measuring team dynamics like trust, cohesion and collaboration. These aspects are normally easier to measure through qualitative measures.

Nevertheless, whenever quantification is possible, I lean towards the use of quantitative measurements for the following four reasons:

1- Quantitative measures provide more objective and verifiable performance results. The lack of validity for performance results has been highlighted as one of the issues that field of security metrics suffers from, and it is desirable to avoid it.

2- Quantitative measures have the advantage of being easier to automate, an aspect which might be stressing in the field of CSIR that experiences high overhead.

3- It was argued earlier in this project that the field of CSIR is an operational discipline. This sets naturally preference to quantitative measures.

4- Use of quantitative measures facilitates building benchmarks, which is a potential effective method for analyzing the results of CSIRPE.

There are other factors that might influence a CSIRT's decision to use qualitative or quantitative measures. For example, if the organization has its own quality system with clear performance measures, then a CSIRT need to align by the system and use similar type of measures. The same will apply if a CSIRT decides to use benchmarking as the main analysis method, as the choice of the benchmark will strongly influence the type of measurement. Also, many organizations use incident management software systems which might influence the selection of measurements based on their compatibility with the software.

D.8. CSIR Scope

A performance framework can be designed to assess the effectiveness of a CSIRP design or to assess the effectiveness of a CSIRT's incident handling. In an ideal environment (UPEF), a performance model should be capable of conducting both. However, there are several practical and design considerations that can shift emphasis to one aspect over the other. Examples of these considerations include:

1- *CSIR Maturity:* When is performance evaluation introduced to the CSIRT? A newly formed CSIR capability tends to put more focus on planning and design compared to a well-established CSIRT that is interested in identifying areas of improvement

2- *Pro-activeness:* Is there an organizational culture that promotes advanced planning and implementing proactive measures (CSIRP effectiveness) compared to responding to immediate needs and risks. For large organizations, long-term planning is essential; but for starters, small companies focusing on operational needs might be the norm.

3- *Development Model:* Many organizations start with a small CSIRT and slowly grow based on the feedback from deployment. In this approach, operational feedback plays crucial role in the quality enhancement (CSIRT effectiveness).

4- *Analysis Reference Point:* This a higher level question related what is static and dynamic in the underlying environment, the CSIRP or the CSIRT? See *S.6: Reference Analysis Point*.

The above considerations demonstrate that CSIRT and CSIRP effectiveness are not mutually exclusive. They both go hand by hand. However, there are some factors related to the environment and to the management approach that make a performance framework focus on one aspect over the other.

The main implication of this design parameter is how performance indicators and metrics are selected. When focusing on CSIRP design, performance indicators like reliability, preparedness, capacity, completeness, and competency will take precedence in the performance analysis. And when focusing on CSIRT execution indicators like: response time, detection accuracy, containment effectiveness and harm reduction will receive more attention.

3.3.2 Framework Assumptions

Without loss of generality, another set of assumptions are introduced here based on the above design parameters for presentation purposes. These assumptions are referred to as *Framework Assumptions*, and are summarized in Table 18. Without the introduction of this set of assumptions, the framework presentation will contain redundancies and unnecessary lengthy discussions that do not map to actual contributions.

#	Name	Description
A2.1	CSIR Model	The CSIR uses the Hybrid Model
A2.2	Member Dedication	Members of the core CSIRT are tasked with only responsibilities of incident handling without other overlapping responsibilities within the organization
A2.3	Central CSIRT	The CSIRT follows the Central model of organization
A2.4	Internal CSIRT	All of the CSIRT core members are internal and serve as members of the organization being served
A2.5	Sequential Handling	A CSIRT handles one incident at a time, i.e. does not handle incidents concurrently

Table 18: Global Framework Assumptions

A2.1. CSIR Model

As argued in Section 2.1.2, the hybrid model has the features of having reasonable number of phases, being currently used in the industry and focuses on actions compared relationships. Confining discussion to a specific CSIR model is used for presentation consistency and does not impact the major themes discussed in the PE framework design.

A2.2. Member Dedication

In practice, most CSIRT members hold responsibilities within the organization besides CSIR handling. Examples of these roles include: IT managers, Information Security officers, Network Managers and accountants. For simplicity, it is assumed that members of the core CSIRT do not hold other roles within the organization. For instance, when it is referred that a CSIRT communicates with a network manager or a help desk officer, it is assumed that the members are exclusive, while in reality they could be the same individual. This assumption is consistent with CERT [2] which conceptualize a CSIRT in manner that displays very small overlap with the larger security team of the organization.

A2.3. Central CSIRT

Out of the three models for structuring a CSIR outlined in [1], the centralized model is selected in this project for presentation purposes. The central model is currently the most commonly used CSIRT structure model in the industry. Theoretically, in terms of CSIR functions, the other two models, i.e. distributed and coordinated, are similar to the centralized model as the operational difference is in terms of executive power and role assignment, not the CSIR phases or main functions.

A2.4. Internal CSIRT

For simplicity, it is assumed that all members of the CSIRT are also members of the same organization. This assumption ensures that the responders have access to all information relevant to the incident; which is not the case if the CSIRT is external. With external teams, there needs to be disclosure procedures and policies for sharing the information to avoid privacy issues and giving responders access to information which might be sensitive to the organization [140]. Inaccessibility to all relevant information would have an impact on the decision making of the CSIRT and hence its effectiveness. For the scope of this project, this factor is eliminated for presentation purposes but the issue can be of interest to an extension work.

A2.5. Sequential Handling

Although it is reasonable to expect computer security incidents to appear concurrently, there are no published works that discuss how CSIRT should operate under concurrent incidents [112]. Therefore,, it is premature at the current stage to design PE platforms that support concurrent responses. Nevertheless, incident concurrency has been highlighted as a design parameter (D.4) for PE frameworks for completeness purposes.

3.3.3 Laying Out Strategies and Policies

By selecting the proper values for the design parameters, a CSIRT would have taken the first step towards defining its own performance evaluation framework. However, the available framework would have been defined only through parameters which are descriptive of the environment. Another step is needed to refine the framework through how the organization's leadership and the CSIRT envision the role of PE in CSIR.

In this step, which comes after or along with the specification of design parameters, the CSIRT leadership attempts to project its generic plan to address the main issues associated with CSIRPE design.

A strategy is: “a plan to achieve long term or overall objective” and a policy is: “intentions and direction of an organization as formally expressed by top management” (ISO 9000 [43]). Through this understanding, this step aims at outlining the general guidelines of how the PE development should be steered to achieve its forecasted objectives. This is reflected in the form of strategies from which policies are generated or by defining a set of policies that reflect a common strategy.

#	Issue	Descriptions	Strategies
S.1	CSIR Quality Assurance	How can the organization be assured that the designed CSIRP is expected to provide the expected IR performance?	Simulation External Audit Compliance
S.2	CSIRPE Quality Control	Who should oversee the enforcement and implementation of the policies pertaining to CSIRPE within an organization?	CSIRT Quality Unit
S.3	CSIRPE Complexity	What are the strategies to tackle the challenge of CSIRPE being very complex?	Approximation Simplification Aggregation
S.4	CSIR Unpredictability	What are the strategies for CSIRT to perform effectively amid the unpredictability nature of CSIR?	Preparedness NDDM
S.5	CSIRPE Overhead	How can PE be introduced to CSIR without high overhead?	U-Shaped Dev. Automation Comm. Efficiency
S.6	Reference Analysis Point	What is the main reference point of analysis, is it the team or the plan?	CSIRP CSIRT

Table 19: CSIRPE Issues that require strategic design

In other terms, the CSIRT will endeavor here to agree on some approaches, not actual steps, of how it would develop its CSIRPE framework. A list of the issues that require some strategic planning or policy making is presented in Table 19.

After outlining the assumptions, defining the design parameters and sketching the strategies and policies the CSIRT would have finalized the definition of its own performance evaluation framework. The produced framework is suitable for a specific environment and reflects the vision of a specific CSIRT. The following three development stages will evolve in the boundary of this CSIRPE framework. .

S.1. CSIR Quality Assurance

The term quality assurance is used here in reference to methods used to verify that the CSIR capability within a specific environment is what it claims to be. Although several publications use the term loosely or interchangeably with "quality control", the above definition is adapted from the ISO: 8402-1994 standard [141]. In other words, how would it be known that a specific CSIRT is expected to offer a specific level of quality of service during incident handling? Answering this question is of interest both internally, to the higher management, and externally to customers and constituencies.

With absence of direct works that address quality assurance in the context of CSIR, a demonstration that members of the CSIRT had undergone training seem to be common practice, at least from the perspective of the available publications. However, staff training is a method for preparation, perhaps a strong one, but not necessarily a guarantee of the expected delivery of service.

Recently, some works discussed the use of simulation [142] [22] [97]. Simulation can be viewed as a more advanced method of training, and if the outcomes are assessed properly it can provide a satisfactory expectation of quality of service. However, it has been noticed that only very few CSIRTs, mainly military, had undergone simulation training. Also, these simulations only involve the core CSIRT members without simulated interaction with other support teams like the legal, HR and police personnel [14]. Therefore, developing exercises that simulate real-environment remains lagging.

Also, it is noted in [140] that, in the industry, reviewing CSIR processes is becoming a subject of interest for external audit. It is expected that issues like privacy and data security will arise as soon as external auditing gets acceptance in the industry.

Compliance to industry standards is another way to demonstrate quality assurance. Although, the NIST document [1] can be considered a good compliance platform, it is not a standardized system in the formal sense. Therefore, demonstrating compliance with [1] should be treated with caution if it is the only method used for quality assurance.

In summary, it is important for the CSIRT during its early stages of PE development to think about its strategy for quality assurance. Training, simulation, compliance and auditing are the current available options.

S.2. CSIRPE Quality Control

It is a common industry practice to establish a dedicated management unit to oversee the quality management aspects across an organization. This unit is sometimes referred to as “quality office” or “quality unit”. The ISO 9000 standard uses the term: “Quality Management” as the broadest term in which all activities relevant to quality are included, such as: quality planning, quality assurance, quality control and quality improvement [43]. Therefore, some or all activities of performance evaluation of the CSIR capability could be classified under quality management, which is supervised by the quality unit.

It is a strategic decision to be made by the CSIRT about who will oversee the quality control of the CSIRPE. Quality control here refers to activities associated with the enforcement of the quality requirements [43]. The decision will be influenced by the defined PE goals and the existing organizational structure of the CSIRT. This could be fully owned and managed by the CSIRT itself, or by the organization's quality office. Although, managers naturally prefer the centralized management through a single unit in the organization to ensure consistency and facilitates strategic planning, it is noted in [51] that CSIR practitioners prefer to operate independently. They feel that external influences by organizational policy mandates normally contends with the objectives set by CSIRTs which focus on fulfilling CSIR goals.

Practically, the quality management and control tasks will get shared between the CSIRT and the quality unit, but the weight given to each party will vary. In all cases, it is important for the CSIRT to think early about this issue to avoid any undesired consequences. For instance, a CSIRT who have self-administered its CSIRPE might be forced to re-design it in order to meet the requirements set by the management through the quality office.

An important factor that will determine who will evaluate is the intersection level between the various quality components of the organization with CSIR [7]. Several evaluation activities pertaining to CSIR intersect with other IT and organizational activities. Examples include: security resiliency, vulnerability assessment, disaster recovery planning, business continuity planning and risk assessment.

A comparison that demonstrates the advantages and disadvantages of making the quality unit or the CSIRT to administer quality control is presented in Table 20.

Issue	Entity overseeing Quality Control	
	CSIRT	Quality Unit
Flexibility	The CSIRT will have high flexibility in terms of planning, design and quality control	CSIRT will inherit the flexibility or rigidity of the practices adopted by the quality unit
Overhead	The CSIRT will undergo higher overhead to administer all aspects of quality control [15]	The CSIRT overhead will be relatively low and confined to reporting to the quality unit
Ownership	Empowers CSIRT members and advocates viewing PE as an integral activity of CSIR	CSIRT members might view PE as an external or imposed activity
Accountability	The CSIRT will focus on outcomes of interest to the management	The CSIRT will be accountable for its overall PE including inputs, outcomes and processes
Blind-Spots	Might neglect developing PMs that are important to the organization	Might neglect developing PMs that are of technical importance

Table 20: Comparison between methods of CSIRPE quality control

S.3. CSIRPE Complexity

One of the main findings of the survey presented in Chapter Two is that the main two challenges to IR performance evaluation across disciplines is complexity and unpredictability. Since complexity is a sophisticated issue that requires collaborative efforts from researchers and practitioners to produce effective models of analysis and a list of best industry practices, it is beyond the scope of this project to formally model CSIR complexity or provide a rigorous assessment for it.

Nevertheless, a simple model is presented in Section 0 to guide CSIRTs on how to address the issue of complexity in practical terms. The suggested model is referred to as “SAC Complexity Model” referring to the three strategies that are used in deriving the model. The three strategies are :*simplification, approximation* and *cascading*.

The SAC complexity model is proposed as one of the various ways that a CSIRT can overcome the challenge of complexity. Whether a CSIRT adopts this model or uses its own solution to the CSIR complexity challenge, there needs to be enough effort exerted during the planning phase to tackle the issue. Ignoring this step might result in producing a CSIRPE that is also complex to implement and analyze, or producing a CSIRPE that fails to capture the actual performance properties of the CSIR capability due to its impracticality.

S.4. CSIR Unpredictability

As highlighted in the second chapter, unpredictability or uncertainty poses a major challenge to incident response disciplines, including the CSIR. It is beyond the scope of this project to provide modeling for CSIR unpredictability, which requires different analysis techniques than the ones used here. Nevertheless, a platform that focuses on policy and operational guidelines is presented to assist responders in handling this challenge.

A platform for understanding and analyzing CSIR unpredictability is proposed. The framework can be used as a tool to help CSIRTs develop strategies and policies to handle the issue of unpredictability. It also can be used to assist developing performance indicators and metrics that are sensitive to the fact that CSIRTs operate in an environment with high number of uncertain conditions.

The platform is called "*The NFP Unpredictability platform*". The acronym NFP stands for the three principles that are used in the platform, namely: Non-deterministic decision making, flexibility and preparedness, see Table 21.

The three principles are derived from the survey results of chapter two, and are aligned to the requirements set by the NIMS and NRF. The National Incident Management System (NIMS) [109] and its derived National Response Framework (NRF) [110], provide the main principles and strategies that need to exist for management of emergency incidents, including cyber incidents. The two documents are designed to manage incidents of various magnitudes and complexity. The details of the NFP unpredictability platform are provided in Section 5.5.

#	Principle	Description
1	NDDP	The CSIR capability should be equipped with decision making tools suitable for nondeterministic environments
2	Flexibility	The CSIRP should be flexible to allow CSIRT members to adjust to the needs arising from unpredictable factors
3	Preparation	The CSIRT should be prepared in terms of planning and competency to address unpredictability

Table 21: The three principles of the NFP Unpredictability Platform

S.5. CSIRPE Overhead

Adding a performance evaluation component to an existing CSIR system is expected to bring several advantages but will also be accompanied by some undesirable side effects. Overhead is one of these unwelcomed consequences. During incident handling, every minute of the CSIRT's time is valuable. Therefore, introducing the PE module should be sensitive enough not to hinder the core functionalities of the CSIRT.

Since overhead cannot be completely eliminated, focus should be paid to minimizing it as much as possible. It is one of the duties of the CSIRPE planning staff to ensure that the proposed modules do not involve unreasonable overhead, which is a subjective matter depending on the underlying environment.

There are several strategies to minimize the overhead effect, three of them are briefly described hereafter.

1- ***U-Shaped Development Model***: This model suggests that most of the time should be spent at the beginning and at the end, and very little is done at the middle. The term ‘u-shaped’ is used in similar ways in several works [143] [144]. In CSIRPE, this translates to spending most of the time during the preparation and analysis time and aiming for minimizing work-time during the incident handling.

2- ***Automation***: The objective is to minimize performance evaluation tasks through use of automated tools, which are mainly software tools [83]. The following five advantages have been highlighted by [145] with regards to computer security incident response: (1) timely and efficient detection of incidents (2) multi-factor risk assessment (3) efficient notification (4) trend analysis and reporting (5) compliance with privacy laws. In CSIRPE, focus should be on automating the collection of performance metrics readings and performing basic analysis. This has been highlighted as an important aspect to increase effectiveness [146]. However, involvement of automation in planning and during final analysis should be minimized.

3- ***Efficient Communication Models***: The impact of efficient of communication on team performance is highlighted in [147]. The objective here is to minimize communication overhead as much as possible. The existing models suggest, coupling communication with needs, periodical short updates, changing from explicit to implicit communication through the shared mental model, involving the whole team in the derivation of mission and goals, and careful structuring of the team.

S.6. Reference Analysis Point

When analyzing the PE of an organization, there are two methods to frame the analysis context. The first method focuses on the CSIRT and the analysis is framed around the question of: “does the CSIRT’s performance achieve the desired PE goals?”. By making the reference point of analysis to be the CSIRT, the CSIRP is viewed as an outcome of the CSIRT’s performance during the preparation phase.

When using CSIRT as the reference point, the preparedness of the CSIRT through training and planning, the expertise and qualifications of the CSIRT members, the harmony and cooperation between the team members, the decision making capabilities, the tools selected during the incident handling are factors that strongly influence good and poor performance. Issues like effectiveness of security policies, efficiency of handling procedures and robustness of the preventive methods are secondary and can be blind-spots to CSIR performance analysis. This method of analysis is intuitive and is practically used in various disciplines [60] [61].

The second method adopts the CSIRP as the analysis point of reference and the PE question is framed as: “does the implementation of the current CSIRP achieves the desired PE goals?”. Here, the performance of the CSIRT is inspected within the boundaries of the CSIRP. It is expected that if a CSIRT implements the procedures outlined in a CSIRP then that should lead to some degree of good performance, presuming that the CSIRP is well prepared. The main analysis blind-spots here would be the volatility of decision making when needing to go out of script and identifying performance issues arising from poor CSIRP design.

It is clear that there is a high correlation between the CSIRT performance and the CSIRP effectiveness. For instance, poor performance can be due to poor design, poor implementation or both [9]. Therefore, when selecting an analysis reference point, it should be considered as the primary perspective, and the second should be treated secondary; instead of being ignored.

Issue	CSIRT	CSIRP
Scope	CSIRPE can be used in various locations in which the CSIRT operates using various CSIRPs	CSIPRE can be used in all locations in which the CSIRP is enforced through various CSIRTs
Update Frequency	Whenever there is a major update to the CSIRT size or membership, the CSIRPE needs to be updated	Whenever there is a major update to the contents of a CSIRP, the CSIRPE needs to be updated
Analysis Outcomes	Poor performance can be directly mapped to CSIRT preparedness and execution effectiveness compared to policies and CSIRP design effectiveness	Poor performance can be directly mapped to policies and handling procedures compared to CSIRT execution effectiveness
External CSIRTs	If an external CSIRT is invited for assistance, the CSIRPE is used for execution effectiveness not for prevention effectiveness	If an external CSIRT is invited, the CSIRPE is used
System Security	System security and team performance are disjoint entities	The system security and team performance is treated as one entity
Cost	Cost of training all staff, or replacing a team member is relatively high	High costs at the planning stage. Cost of updating CSIRP is relatively low.

Table 22: CSIRT vs. CSIRP Analysis Reference Points

Regardless of the selected reference analysis point, an organization should be able to identify the sources of good or poor performance and map them back to the CSIRP or the CSIRT. However, it is simpler to conceptualize CSIRPE using one of the two reference points which will impact how the CSIRP is documented and how performance metrics are derived. Mixing both methods can result in incoherent planning and analysis and possibly contradicting conclusions. Probably, if an organization seeks intensive PE analysis it is possible to conduct the PE analysis in two phases using a different point of reference for each phase.

Although adopting a reference analysis point can be a matter of design preference, there are several factors that can make one method preferred to the other within a specific environment. Examples of these factors include: company size, CSIR budget, staff retention and CSIRT being internal or external.

A comparison between the two perspectives is presented in Table 22, and a simple guide for which method to select is provided in Section A.3.

A simple questionnaire to guide the CSIRT leadership in selecting the proper reference point is provided in Table 47. If an organization is not certain which reference point to use, I recommend using the CSIRP method [92] [7]. This speculation is based on my observance of the evolution of the field (See for example the note put by the New Zealand CSIRT that consistency, reliability and resilience are established through a CSIRP not the CSIRT [31]). At the early stages of the field development, an organization is likely to have a good IR capability if they hired a well-trained security professional; which is not necessarily true anymore. Much efforts and financial resources are being invested in the preparation of CSIRPs which provide more robust and stable CSIR environment compared to hiring professional experts with low retention probability. Also, the CSIR process is viewed more comprehensively through the lens of the CSIRP which captures the overall IR capability compared to the CSIRT perspective which is sensitive to the qualifications of the CSIRT members.

3.4 Phase II: Defining a CSIRPE Model

3.4.1 Overview

As a CSIRT gets into this second phase of development, it would have achieved two main things from the previous phase. First, setting the design parameters eliminates issues that are not of concern to the team. Second, the team would have inspected issues pertaining to the organization and envisioned a list of strategies and policies which the PE model to be developed need to abide by.

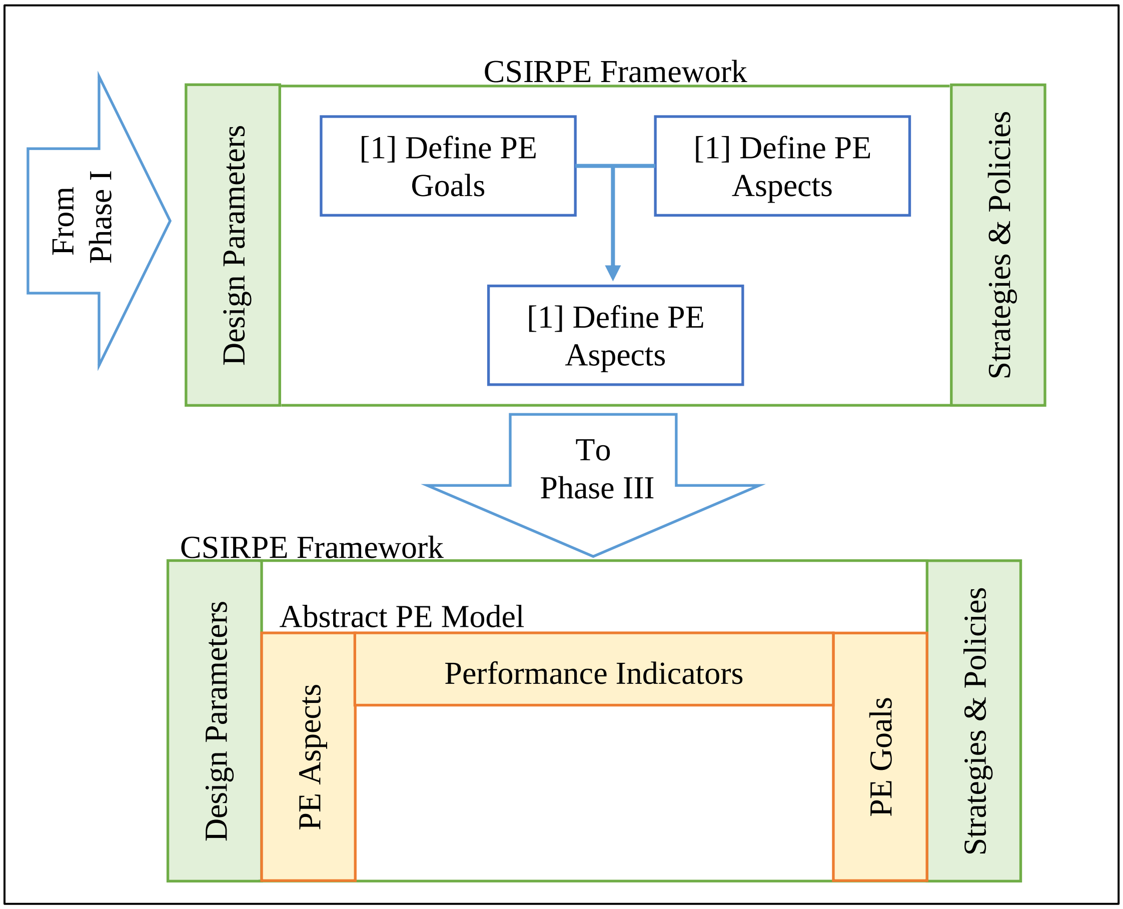


Figure 10: Phase II: Defining Performance Evaluation Model

Through this second phase, a CSIRT would define its own PE model by setting its goals, specifying what needs to be measured and outlining the desired features of the performance. This is depicted in Figure 10, which also shows the inputs and outputs of this development phase.

This phase is composed of three components. The first component, *Defining Performance Goals*, aims at defining performance goals which explain why CSIRPE is introduced to the CSIR capability and how it is expected to enhance it. The second component, *Identifying PE Aspects*, analyzes the CSIR system and identifies areas that would be evaluated. The third component, Defining Performance Indicators, pinpoints factors that determine good and bad performance based on the outlined goals and aspects. The following sections provides the details for the three components.

3.4.2 Defining Performance Goals

There are two perspectives to answering the question of why is PE needed for CSIR. The first perspective is generic and can be applied to any CSIR environment while the second is contextual and can vary depending on the operating environment. The two perspectives can also be treated as a two-stage process of defining goals, starting from general goals to more specific ones attuning to the organizational needs.

Using the model presented in the CERT publication of [41], *objectives* are treated as higher than *goals*. Therefore, the generic goals of CSIR will be called *PE objectives*, and the organizational goals will be called *PE goals*. Using the terminology of the field of management, PE objectives correspond to *official goals* while PE goals correspond to *operative goals* [148].

The main output of this development step is a concise list of goals capable of guiding the derivation of measurement tools and the remaining steps of the development process. An organization may choose to generate two lists, one for objectives and for goals, or simply one maintain one list that is directional and operative.

Defining PE Objectives:

The definition of *PE objectives* is concerned with the general themes and motivations for introducing a performance evaluation module to a CSIR capability. Since these objectives are context free, they could be collaboratively generated by researchers and practitioners. It is recommended that these objectives be publically shared which will contribute towards building best industry practices.

Surveying the current publications and documentations about preparing a CSIRP and comparing how other performance disciplines define their objectives, it can be noticed that CSIR PE objectives revolve around the following three themes (see [19] [7]):

1- *Business Continuity & Growth:* by implementing a PE model an organization will have a system for quality assurance, quality management and continuous enhancement to its CSIR. Deploying such system is expected to contribute to the overall business operations. Without CSIRPE capability, an organization will not be able to maintain its CSIR capability, assure that its delivering a quality service, or be able to recommend effective enhancements.

2- *Effective Resource Allocation:* by implementing a PE model an organization will be able to gain effective allocation of its fiscal and human resources for CSIR. This leads to more cost-effective and better coordinated responses.

3- *Disaster Prevention and Management:* by implementing a PE model an organization will be able to have accurate tools to identify risk factors and operating obstacles during and before a disaster due to cyber incidents. More specifically, the organization is interested in having tools to identify, measure and minimize associated harm due to cyber incidents.

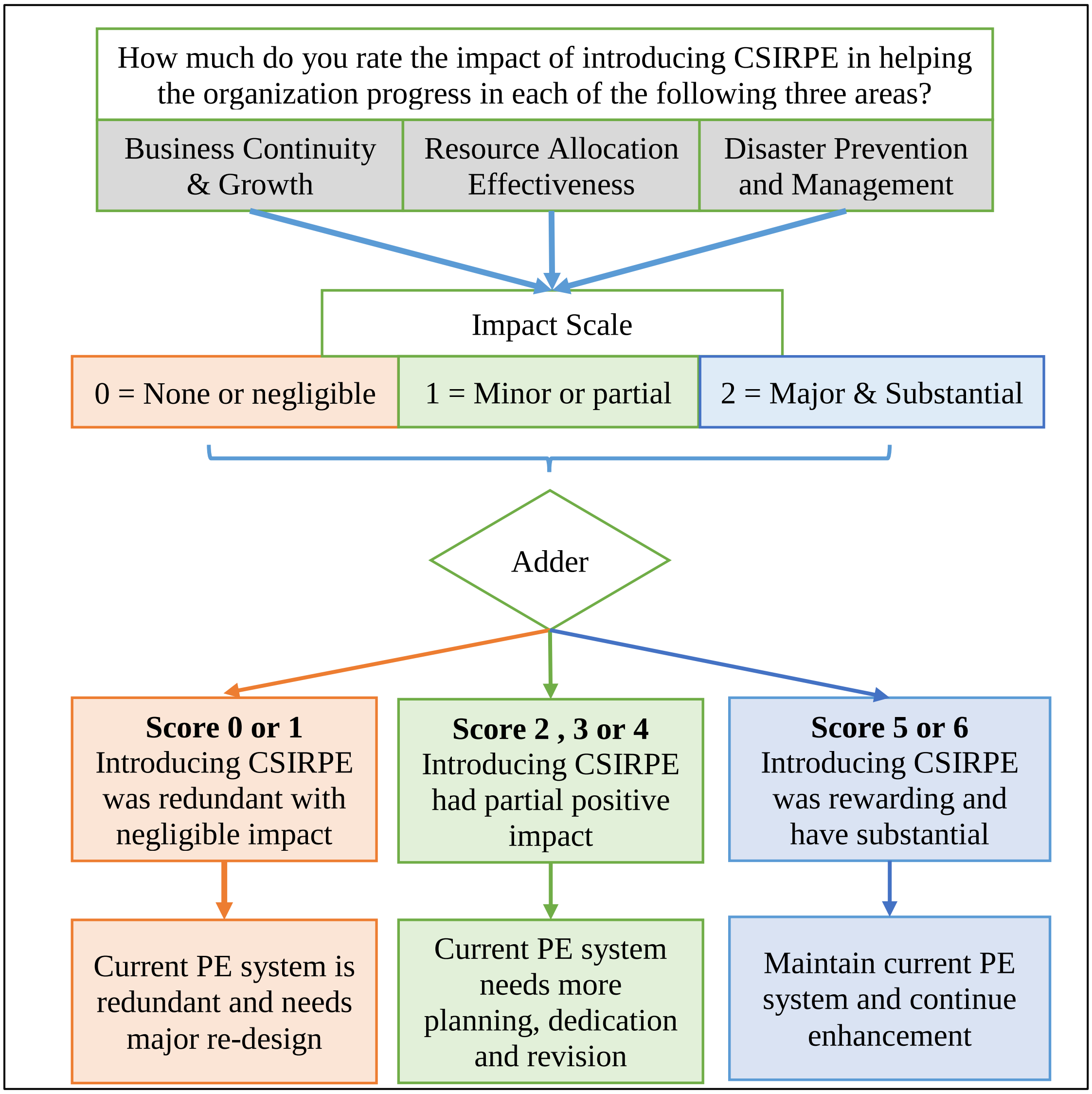

How much do you rate the impact of introducing CSIRPE in helping the organization progress in each of the following three areas?
Business Continuity & Growth
Resource Allocation Effectiveness
Disaster Prevention and Management
Impact Scale
0 = None or negligible
1 = Minor or partial
2 = Major & Substantial
Adder
Score 0 or 1
Introducing CSIRPE was redundant with negligible impact
Score 2 , 3 or 4
Introducing CSIRPE had partial positive impact
Score 5 or 6
Introducing CSIRPE was rewarding and have substantial
Current PE system is redundant and needs major re-design
Current PE system needs more planning, dedication and revision
Maintain current PE system and continue enhancement

Figure 11: Example of CSIRPE Usefulness Test

A CSIRT can treat the above three themes as the starting place from which the PE objectives are derived from. The three themes can also assist in constructing an evaluation test for assessing the benefits of introducing PE into CSIR. The test could be called: *The CSIRPE Usefulness Test*, as it attempts to answer the question: “Did the integration of a CSIRPE module help the organization progress in all or some of the three main goals of

CISRPE: contintuity & growth, resource allocation effectiveness and disaster prevention and management".

The test is generic and qualitaitive and is meant to help redirecting or refining the definition of PE goals, and the higher-level issues of the design. More rigours and quantitative approaches for essessing the effectiveness of introducing CSIRPE can be done in the eighth componenet of this framework on validation models.

Performance Evaluation Goals:

Building on the previous step, a CSIRT needs to contextualize its PE objectives into a list of goals that serves the needs of the organization. These goals need to be relevant and provide operational guidance.

In order for any performance measurement system to be effective, the PE goals need to be derived from or aligned by the organizational goals [149]. In the context of CSIR, there are two main sources in which PE goals need to be derived from and aligned with. The first is the CSIR goals/objectives, which are normally defined at one of the early sections of CSIRP. The second is the mission statement and overall organizational objectives. A CSIRPE that is based on shallow understanding of the CSIR and organizational goals is expected to be ineffective.

Providing an example for deriving from the organizational mission, if a mission statement reflects emphasis on customer satisfaction, then it should be expected that a CSIRPE module will have some contributions towards that. A possible PE goal would be to: "to develop tools to measure customer satisfaction to the organization's response to computer security incidents". Another example for a PE goal derived from CSIR objectives, if CSIRP considers one of its goals "to protect infrastructures X and Y during a compromise"; then a PE goal would be to "develop mechanisms for measuring the effectiveness of a CSIRT in protecting critical infrastructures X and Y during an incident handling".

Refining Goals:

Since *goals* play a crucial role in directing the development phase, distinct attention should be paid to its design and approval process. The stressing question here is "how does a CSIRT know that it has defined the correct goals for its PE module"? I have found that the most commonly used method in performance systems [150] is the SMART model [151]. The name of the model stands for: specific, measurable, actionable (i.e. achievable), relevant (i.e. result-based and realistic) and time bounded. In [41], the model was modified to SMARTER by adding (E) evaluated and (R) reviewed. In my view, this is unnecessary as the amended features reflect the design process more than the goals themselves.

The SMART model is generic and is originally designed for evaluating goals and objectives in the contexts of organizations and projects, not necessarily in the area of computer security or CSIR. Therefore, a customized interpretation of the model is needed for the CSIR context. This is provided in Table 23.

Finally, it should be noted that the definition of goals is a collaborative process between the CSIRT and the management. When a CSIRT defines its own performance goals, it helps integrating the performance activities into the response activities as the developers of the goals are the same individuals that are executing them. It also ensures that the goals are realistic and reflect the operational needs of incident handling. The involvement of the management works as a validation method and assurance that the define goals serve the overall performance of the organization.

	SMART Term	Interpretation in CSIR Context
S	Specific	▪ The PE goal targets a specific *performance aspect* of CSIR (Section 3.4.3) ▪ The PE goal is written in a simple language that could be unambiguously interpreted by the CSIRT and management ▪ The list of PE goals is concise, preferably less than
M	Measurable	▪ The PE goal focuses on outcomes not actual actions ▪ The PE goal can be expressed through CSIR performance indicators ▪ The PE goal can be represented by <u>several</u> PIs. If not, consider expanding or merging with another goal. ▪ The PE goal is expressed in a manner that encompasses several contexts of measurability
A	Achievable	▪ The PE goal is feasible to achieve the PE goal given the CSIRT's competency and available resources ▪ The PE goal is expressed in action-verbs and provides operational guidance
R	Relevant	▪ The PE goal is derived from the CSIR objectives ▪ The PE goal is aligned with the organization's mission and goals ▪ The PE goal does not demonstrate conflict with the policies and procedures outlined in the CSIRP or enforced by the quality unit
T	Timely	▪ The PE goal implies a time-frame, or permits the definition of a time-frame, to achieve the goal. ▪ The PE goal is recognized as short-term or long-term

Table 23: SMART Model for Defining CSIRPE Goals

3.4.3 Identifying PE Aspects

After defining the PE goals, a CSIRT needs to attempt identifying aspects of the IR system that will be subject to performance evaluation. Both goals and aspects will be used to define performance indicators in the consequent step of the CSIRPE development.

The term *aspect* is used here to refer to the collection of actions performed during the CSIR cycle that will be treated as one block in performance evaluation. Examples of aspects include: designing a CSIRP, containing an incident, collaborations with external consultants and CSIRT's reporting to management. Each of the aforementioned examples refer to a group of actions, not a single one, and are normally considered valuable to the CSIRT or consume much of its efforts prior, during or post to an incident response.

There are various methods for identifying CSIR aspects. Five different methodologies are presented in Table 24. The first and third methodologies can be viewed as CSIRT centric, the second and fourth can be viewed as CSIRP centric, while the organizational model is a hybrid one. These methodologies are inferred from the requirements set by the NIST and CERT documents, while the organizational model is borrowed from classical works in the field of performance evaluation [152].

#	Methodology	Description
1	Time Analysis	The CSIR life cycle is viewed as a time line starting from the CSIRP preparation to the end of post-incident analysis. The timeline is divided into time clusters each representing the time period in which a group of activities is performed. Each time cluster is defined as a performance aspect.
2	Phase-Based Analysis	Each phase of the IR life cycle phases defined in a CSIRP is defined as a distinct performance aspect
3	Value Analysis	The main activities of a CSIRT are analyzed and categorized into major and minor based on their importance to the incident handling. Each major activity is defined as a performance aspect, and every group of correlated minor activities are aggregated into a performance aspect.
4	CSIRP Based	The performance aspects are defined in reference to a CSIRP. The two major activities are: CSIRP design and CSIRP Execution. It could be further divided, e.g. escalation protocols.
5	Organizational Model	The whole IR process is conceptualized as a combination of five elements: inputs, processes, products, outputs and outcomes. Each of these can be customized into a single or multiple performance aspects.

Table 24: Methodologies for Identifying PE Aspects

Examples of how each of the above methodologies can be used in identifying aspects is presented in Table 25.

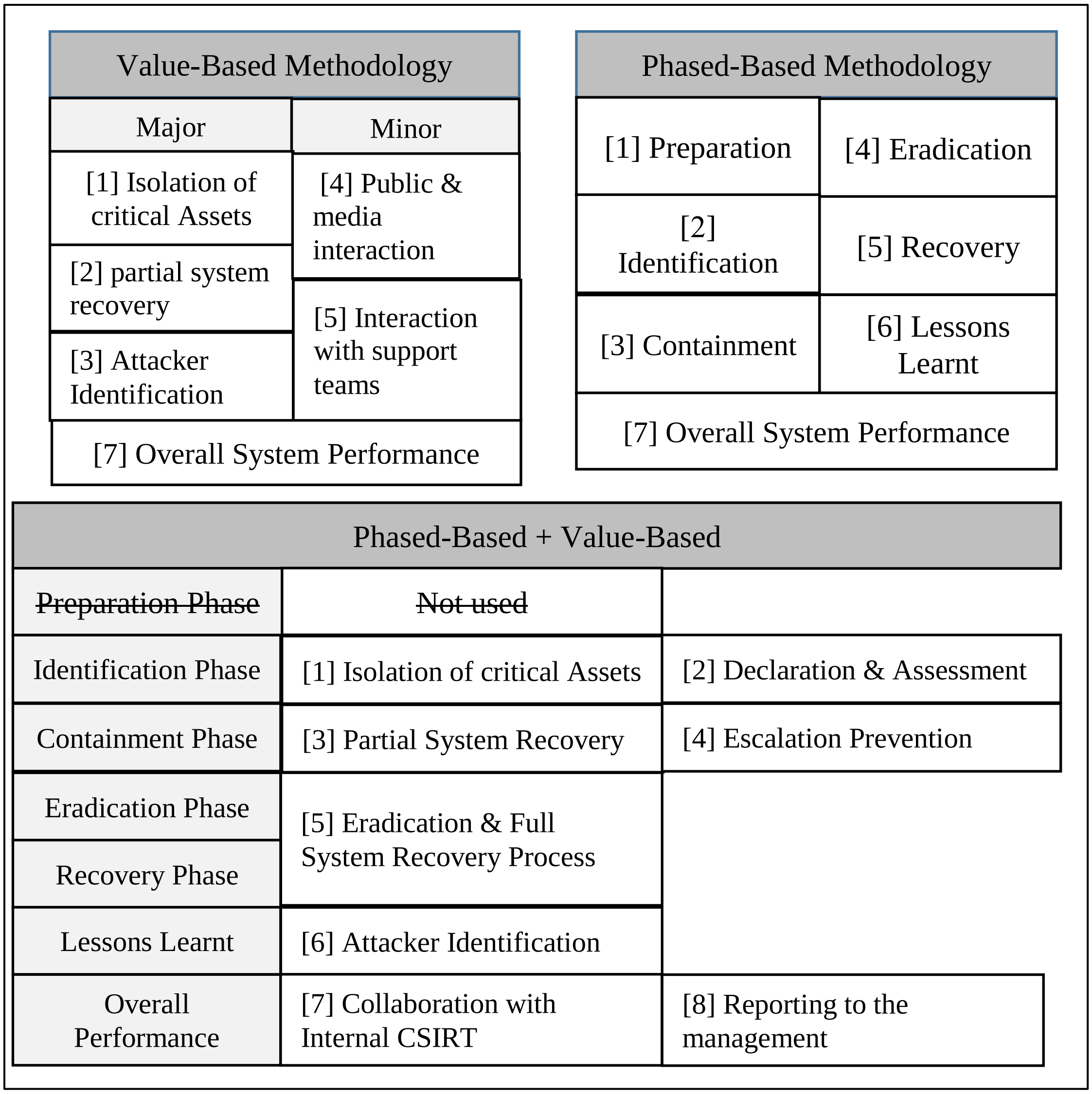
There are advantages and disadvantages for using each of the above methodologies. For example, the time-based analysis is suitable for identifying the major time clusters of incident handling, which can correspond to areas in which the response resources are highly used. It is reasonable to subject these lengthy time clusters into PE evaluation in order to shorten the time period which will consequently lead to better resource allocation. However, the methodology falls short of identifying the moments which are critical to the incident handling but do not necessarily consume long time periods. An example of this is the initial assessment performed during a very short time period at the early stages of the IR life cycle, but have magnificent impact on the effectiveness of the response. It could be noticed that in all of the above methodologies, the overall system is broken down into several aspects which each will be subject to a separate performance evaluation. Such approach would be insufficient unless it is coupled with an evaluation of these aspects when analyzed as a whole. Therefore, regardless of the selected methodology, there should be another aspect defined which is “overall system performance”. This aspect is normally evaluated after the evaluation of the other aspects.

Based on the above note, it would be more effective to use a combination of methodologies to derive the PE aspects. More precisely, it is better to use at least two methods which are of different emphasis, i.e. CSIRT vs. CSIRP. For example, consider an external CSIRT invited to handle an incident which is beyond the capability of an internal team. In such scenario, the CSIRP based model is of less interest because the effectiveness of the CSIRT is focused on the execution not the design of the CSIRP. Therefore, the phased-based methodology could be coupled with the value-based methodology. Using the examples presented in Table 25, a depiction of how the aspects could be merged using the two methodologies is presented in Figure 12.

#	Methodology	Example
1	Time Analysis	Divide the IR timeline into four time clusters: [1] preparation & training [2] incident identification and assessment [3] incident containment [4] incident eradication and system recovery [5] Overall system performance
2	Phase-Based Analysis	Assuming the SANS model is used: [1] preparation [2] identification [3] Containment [4] eradication [5] recovery [6] lessons learnt [7] overall system performance
3	Value Analysis	A CSIRT operating in major municipality might define the following three major aspects: [1] Isolation of critical assets after incident identification [2] Immediate partial system recovery [3] Identification of perpetuators [4] Overall system performance. Other minor activities are classified into two aspects: [5] Public and media interaction [6] CSIRT collaboration with support teams
4	CSIRP Based	[1] CSIRP design [2] Policies defined in a CSIRP [3] CSIRP Execution of containment and recovery procedures [4] Escalation protocols defined in a CSIRP [5] overall system performance
5	Organizational Model	[1] inputs: preparedness [2] processes: identification, containment, recovery [3] products: incident classification (from identification phase) [4] outputs: response Time, response cost [5] outcomes: Damage assessment In addition to: [6] overall system performance

Table 25: Examples for the Methodologies for Identifying PE Aspects

Note how the first phase, the preparation, is ignored because the external CSIRT was not involved in the preparation of the CSIRP. Also, the public and media interaction are not of direct responsibility to the external CSIRT; thus it is changed to reporting to the organization management which shall communicates with its customers and the relevant media. Further, the identification, containment and lessons learnt phases are broken into two aspects, which could not have happened if only the phase-based methodology was used.



Value-Based Methodology	
Major	Minor
[1] Isolation of critical Assets	[4] Public & media interaction
[2] partial system recovery	[5] Interaction with support teams
[3] Attacker Identification	
[7] Overall System Performance	

Phased-Based Methodology	
[1] Preparation	[4] Eradication
[2] Identification	[5] Recovery
[3] Containment	[6] Lessons Learnt
[7] Overall System Performance	

Phased-Based + Value-Based		
~~Preparation Phase~~	~~Not used~~	
Identification Phase	[1] Isolation of critical Assets	[2] Declaration & Assessment
Containment Phase	[3] Partial System Recovery	[4] Escalation Prevention
Eradication Phase	[5] Eradication & Full System Recovery Process	
Recovery Phase		
Lessons Learnt	[6] Attacker Identification	
Overall Performance	[7] Collaboration with Internal CSIRT	[8] Reporting to the management

Figure 12: Example of Combining two methodologies to define PE Aspects

In Table 48, (see Appendix A), a list of practical recommendations for the identification of performance evaluation aspects is presented.

3.4.4 Defining Performance Indicators

The main philosophy behind the use of performance indicators and metrics is the renowned principle: "What's measured can be managed" [106]. The principle is sometimes phrased as: "one cannot improve what one cannot measure" [11] or "if you cannot measure it, you cannot improve it [153]". Therefore, Performance Indicators (PIs) are factors that determine the quality of an incident response (See Section 4.3). In this proposed CSIRPE framework, performance indicators are considered the interface between the performance goals/aspects and the actual measurement tools, i.e. performance metrics.

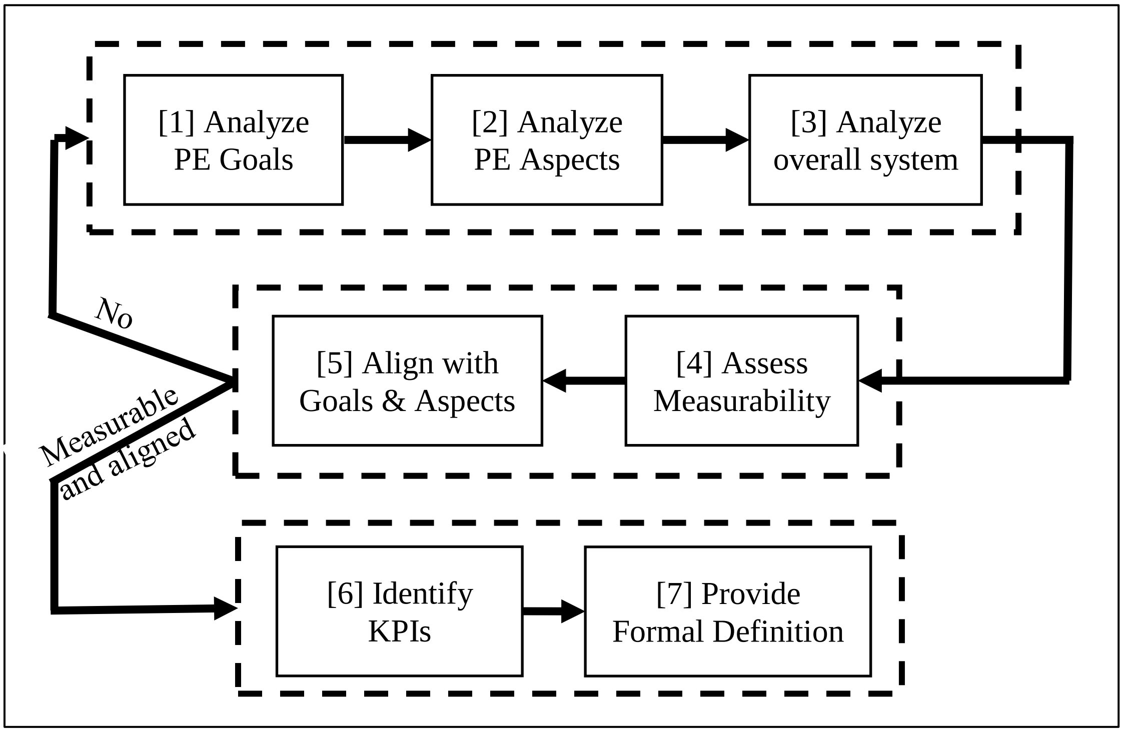


Figure 13: Defining Performance Indicators Process

Using the proposed CSIRPE framework, the two main sources in which PIs will be derived from are the goals and aspects. Performance indicators can be viewed as adverbs for the PE goals, adjectives for the PE aspects or descriptors for the states of response system. Examples include readiness, reliability, stability and length of response time.

It is expected that each goal and aspect will be mapped to at least one PI. It is a common practice to consider PIs with high impact on the performance as key performance indicators (PKIs). Each PI should have at least one performance metric (PM) as a tool to provide quantifiable measurements.

The process of developing performance indicators can be abstracted in seven steps, see Table 26. These seven steps can be grouped under three main stages. In the first stage, the goals and aspects are analyzed to identify the performance factors. The result is a list of desired features about the quality of the response activities and the achievement of goals. Since each goal and performance aspect need to be analyzed separately, the whole response system is then analyzed as one unit for determining the higher level performance factors. The output of the analysis is a list of performance factors, called PIs.

In the second stage, each PI is inspected individually to assess its measurability and its alignment to the goals and aspects. Non-measurable PIs need to be discarded or redefined. The mapping between the PIs and the goals and aspects need to be verified. If a goal or aspect has no mapping to the pool of PIs, then either one PI need to be generated, or the goal need to be re-defined.

The final stage includes the classification and formal definition for each PI. The classification can be based on priority, i.e. KPI and regular PI, or based on factors suitable for the operating environment. A sample template for how formal definition of a PI could be achieved is presented in Table 49 (See Appendix A). Each PI will be properly defined, outlined with the goals/aspects and later with the developed performance metrics. When mapping to several goals/aspects, the formal definition should elaborate the context of interpreting the PI under each goal/aspect, if necessary. More details about the seven steps in the above process are provided in Table 26.

#	Step	Description
1	Goal Analysis	Analyze each PE goal to derive factors that determine the achievement of the goal. This could be done by inspecting the action verbs used in the goal statement and searching for an adverb that describe a good implementation of the action verb, or a noun that describes the state of the system after achieving the goal
2	Aspect Analysis	Analyze each PE aspect to derive factors that determine the quality of the outputs resulting from implementing the aspect's sequence of actions. This could be done by inspecting the nouns used in the aspect statement and searching for an adjective that describes good output, or a noun that describes the state of the system if the actions within a PE aspect were successfully implemented.
3	System Analysis	Examine the overall IR system and think of factors that determine the success of incident handling. This could be done by inspecting the PE objectives, the overall system performance aspect and thinking of the state of the system if multiple PI's (derived from step 1 and 2) were positive.
4	Measurability	Examine the feasibility of using tools to measure the factors obtained from Steps 1 to 3. Focus on tools that are represented in numbers or provide objective assessment. Good PIs need to have two or more measuring tools. Re-examine or discard indicators that are non-measurable
5	Mapping	Examine how each factor relate to the goals and aspects. Each factor should be mapped to at least one goal or aspect. If a factor relates to three or more goals and/or aspects, then it should be mapped to the overall system performance aspect
6	Classification	Distinguish between indicators that are mapped to the overall system performance aspect and those mapped to individual goals and aspects. Also, assign priority levels to these indicators. A simple scheme is to use key performance indicators (KPI) and regular indicators (PI)

7	Formal Definition	For each PI, assign a name, provide description, define scope and priority, outline relationship to goals and aspects and possible performance metrics.

Table 26: Description of the Steps for Defining Performance Indicators

Note that the second stage of the above development process can be considered a validation phase. The two factors that determine that the selection of performance indicators are *valid* is the measurability and alignment to the PE goals. These two factors are assessed during the development process and also later when validating the whole CSIRPE model (Component 8). Furthermore, as PIs are interface between goals and performance metrics, the validation of both indirectly contribute to the validation of the PIs.

Finally, in Chapter 4, a compilation of performance indicators is presented. Each performance indicator is defined, and supported with interpretation guidelines and sample performance metrics.

3.5 Phase III: Measuring CSIR Performance

After completing the second development phase, the main characterization of the CSIRPE framework would be the list of PIs. Each PI represents a desirable performance feature of the response system, and the full list of PIs is supposed to capture the overall performance of the system. The next phase is to design performance analysis and measurement tools. This can be broken into three main steps: deriving performance metrics, defining performance analysis techniques and defining validation techniques, see Figure 14. The details of these three steps are provided in this section.

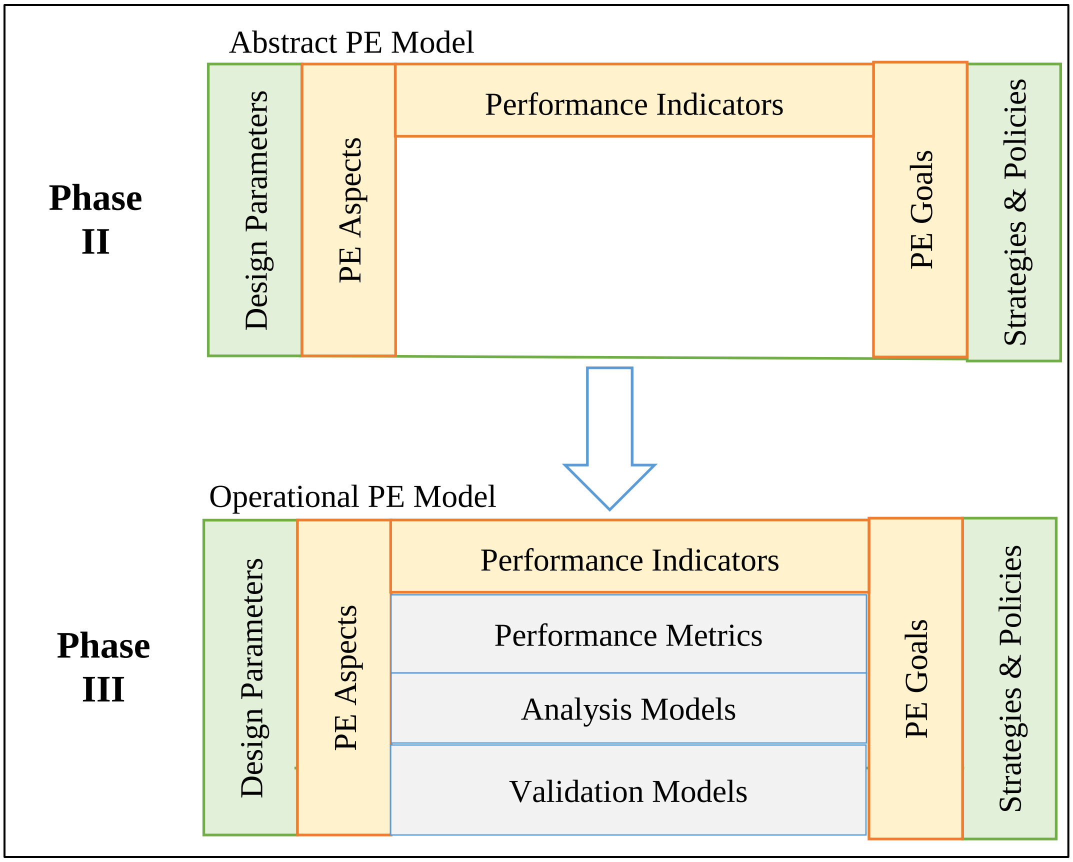


Figure 14: Transition from Phase II to Phase III: Measuring CSIR Performance

3.5.1 PE Metrics Derivation Process

At the end of the response life cycle, the CSIRT will need to prepare a report demonstrating its success based on these outlined PIs. In order to do that, a CSIRT needs to think about the tools that it will use to measure its success in each of the PIs. These tools are called performance metrics.

The IRPE literature, as presented in Section 2.2.3, does not discuss models for deriving performance metrics, as there is no standardized way for achieving this and it is highlight dependent on the system design. Instead, discussion is concentrated on the desirable features and design considerations for performance metrics. A similar discussion in the context of CSIR is provided here.

#	Design Parameter	Values	Description
M.1	Metric Type	Performance Descriptor	Is the metric used for actual performance measurement or as a descriptor, i.e. input, for other performance metrics?
M.2	Scope	Generic Specific	Can this metric be used across all security incidents, or is it applicable to a set of incident categories?
M.3	Quantifiability	Quantitative Qualitative	Does this metric provide qualitative or quantitative measurement
M.4	Measurement Tool	[Type]	What is the type of qualitative or quantitative assessment tool used. If the metric is quantifiable does the numbers represent actual measurement, calculation output, binary values, scale level, …etc.
M.5	Performance Indicators	List of PIs	Is this performance metric used in the analysis of a single or multiple performance indicators?
M.6	Dependence	Direct Indirect	Does this metric depend on measurements obtained by other descriptor or performance metrics?
M.7	Accuracy	Actual Estimation	Does this metric use actual readings or does it involve an estimation or assessment by one of the CSIRT members?
M.8	[Conditions]	[List]	In order to provide meaningful measurement, are there conditions that need to be satisfied before collecting the readings of the PM?
M.9	[Attributes]	[List]	Does the metric have attributes which may have impact on the reading, like instrument or software name, time stamp, estimation method, ..etc
M.10	Interpretation	[Description]	How should the outputs be interpreted? When does it indicate good or poor performance? Are there contextual issues that need to be taken into consideration?

Table 27: Design Parameters for Performance Metrics

Starting with performance metrics design considerations that are relevant to CSIR, a list of ten design considerations is presented in Table 27. Each design issue poses a question to the designers that targets how the metric will be used or analyzed. The list can be treated as a guideline for deriving performance metrics from each PI.

Using a formal definition template that is consistent with the template used for defining PIs, performance metrics can be formally defined through the template in Table 50 (See Appendix A).

Once performance metrics are derived, a CSIRT needs to verify the correctness and usefulness of these metrics. Tools used for this verification are called PE meta-metrics. Several works in other IR disciplines addressed PE meta-metrics [23] [120] [121] [122] [123]. However, only selective list of these meta-metrics are relevant to CSIR. The applicable metrics to CSIR can be classified under three main meta-metrics: relevance, comparability and simplicity, which can be abbreviated as (RCS). A description of these three meta-metrics is provided in Table 28.

Feature	Description
(R) Relevance	• Aligned with a KPI, or several PIs • Compatible with the CSIR capability • Expected to provide results that could enhance performance
(C) Comparable	• Uses objective assessment methods • Provides actual or well-defined estimation range • Demonstrates invariance to some system factors
(S) Simplicity	• Simple to read and analyze • Uses minimum resources and endures little overhead • Unambiguous to interpret by both technical and non-technical members of the organization

Table 28: Features of good CSIR performance metrics

M.1. Metric Type

There are two types of metrics that can be used in PE: descriptive and performance metrics. It is important to distinguish between each type as descriptive metrics can be easily confused as performance metrics which can produce inaccurate PE results.

Descriptive metrics are those used in measuring a specific property of the CSIR system that does not necessarily reflect performance. For instance, the size of the response team and number of infected machines in an incident are examples of descriptive metrics that do not reflect a performance quality in themselves. Descriptive metrics normally provide actual readings and are simple to collect. They are mainly used as inputs for the calculation of performance metrics. On the other hand, performance metrics are directly aligned with the pool of performance indicators through providing measurement mechanisms.

It is desirable for an organization to have a good pool of descriptive metrics which can be recorded by the responders or through automated tools. These descriptive metrics can be used in calculating several performance metrics. It is also important to detect any inaccuracy or inconsistency in descriptive metrics, as that can propagate to the validity and accuracy of performance metrics.

M.2. Performance Metric Scope

Some performance metrics are generic and could be used across all types of incidents. Examples include, incident response time, incident classification accuracy and customer satisfaction. Others are specific to certain category of incidents or under specific conditions. For instance, a server down time or unavailability is a metric that is relevant to incidents involving denial of service attacks. Another example is average response time from external agencies, which can be mapped to the partnership effectiveness PI. This metric is only applicable if the incident handling involved some contact with pre-defined external agencies like law enforcement or peer CSIRTs.

Generic PMs' analysis need always to be conducted in performance reports as they provide some holistic analysis of the response performance. They are also the most commonly used tools in benchmarking. Furthermore, they can be used to identify performance bottlenecks relevant to the management of the incident. On the hand, specific PMs provide more focused performance analysis. They are also used to identify performance issues pertaining to the technical aspects of the system.

M.3. Quantifiability

The advantages and disadvantages of using quantitative and qualitative performance measurement tools were discussed in the design parameter D.7. However, the discussion was focused on the higher level decision of whether to incorporate qualitative measures or not into a CSIRPE framework. In this step, a CSIRT would need to decide for each PI whether a qualitative or quantitative measure needs to be developed, assuming that a framework supports both.

The following considerations need to be taken into account when making the decision:

1- ***Representativeness:*** Which tool provides better capturing (realistic and holistic) of the performance as defined by the PI?

2- ***Simplicity:*** Which tool provides a simpler method for collecting and analyzing the PI?

3- ***Cost:*** Which tool would consumes less of the CSIRT's time and resources?

4- ***Sustainability:*** Which tool is more sustainable in the long run?

5- ***Objectivity:*** Which tool strikes better balance between objectivity and representativeness?

As a general design strategy, it is preferred to support a PI that has several quantitative measures with a qualitative measure, and vice versa. This will ensure that the PI is assessed more comprehensively, and is deemed by some practitioners to provide more effective performance analysis [154].

Finally, several performance evaluation studies suggest that qualitative metrics can acquire the benefits of quantitative metrics if designed through fuzzy logic. Fuzzy models have the feature of being quantified but over a qualitative scale, which provides a reasonable performance reporting. Example fuzzy performance metrics include [155] for measuring flexibility, [156] for measuring individual performance, [157] integration of fuzzy metrics in the balanced score card (BSC) model, and [158] for measuring agility.

M.4. Measurement Tool

When a decision is made that a specific PI is to be measured through a qualitative or quantitative metric, the next decision would be which measurement tool should be used. For example, qualitative metric could be a checklist, survey or a rating on a scale or rubric assessed by specific individuals. Likewise, a quantitative measure can be additive, multiplicative, average, variance or simple counting. Each of these "measurement instruments" have advantages and some shortcomings, and it is a design decision by the CSIRT to define each PM by an appropriate instrument. Although this step might seem simple, it should be noted that subtle changes in the measurement tools can have significant impacts on the performance analysis.

M.5. Performance Indicators

Using the framework proposed in this project, a performance metric needs to originate from some performance indicator. Occasionally, there might be a performance metric discussed in academia or among practitioners that a CSIRT would consider deploying. In that case, there needs to be clear alignments between the PM and the pool of PIs. A "celebrated" PM that cannot be aligned with the outlined PIs is of no practical benefits.

Each performance metric can be mapped to a single or multiple PIs. If a team uses a small number of PMs, then the team should target selecting PMs that either measure a KPI, or multiple PIs.

Whenever a PM maps to multiple PIs there needs to be clear outline of how the PM results should be interpreted in the lens of each PI. For instance, if the network injected with a virus/worm and a large number of machines in various subnets are infected. Let the virus targets transmitting "private" or "secure" data from each machine to an external source. Assume the team uses a metric called: "rate of secured machines" which measures number of machines inspected and successfully cleaned per hour. Also assume that this metric is aligned with two PIs: "eradication effectiveness" and "confidentiality". Under the first PI, a higher rate of secured machines indicate might indicate better eradication effectiveness. The same could be argued about "confidentiality" as the more secure machines the higher confidentiality is maintained. However, if machines have different weights in terms of the "security level" of the stored data, then using the above PM is incorrect. The rate should be calculated using machines with the same level of "security level data".

M.6. Dependence

One of the methods that the ISO 9216 [85] classifies metrics is through the relationship to other metrics. A metric that is independent from other metric measurements is called *direct* metric, otherwise it is called *indirect* metric.

Although it would be desirable for simplification purposes to have all metrics as direct, the majority of practically used metrics are indirect. Additive and multiplicative metrics, and the majority of metrics with statistical properties are examples of indirect metrics.

When deploying indirect metrics, the designers need to insure that all inputs are normalized and variances are minimized. This is a necessary condition for producing correct results, and should be outlined in the definition of the metric. For instance, total response time is an additive metric that uses several time measurements, each pertaining to an elapsed period during the incident handling cycle. If these time periods are collected from several sources, then all of these sources need to share the same definition and understanding of how the incident timeline is divided. This specific simple observation is documented to be one of the issues that caused inconsistencies in performance measurements [117].

M.7. Accuracy

The concepts of accuracy and precision are essential to the theory of measurement. But, as noted by [116] performance measurements operate on the tradeoffs between quality and available resources compared to accuracy and precision. In the real world, it is more meaningful to report to a manager that the unavailability of a resource is between 0.3 and 0.4 compared to an exact value like 0.3435 [159] [160].

To be more specific, it could be argued that most CSIRPE PMs do not require high accuracy to provide meaningful interpretation. However, the CSIRPE system require some metrics to be accurate. This is normally applicable to descriptive metrics like timestamps and counting metrics, e.g. number of compromised records.

When defining PMs for a CSIRPE the following issues need to be considered with regards to accuracy:

1- Does the metric require accurate measurements? Metrics that require exact values should be distinguished from those that do not

2- Whenever accuracy is not mandated, a tolerance gap should be defined. This can appear in the form of a margin of error [78], variance [161] or confidence rating [73].

3- When estimations are used, factors used in the estimation should be recorded

4- Whenever estimation is made, there needs to be a mechanism to identify the practitioner that made the estimation.

5- Does an estimation require a validation? The simplest method is to make multiple estimation by several practitioners and record the variance.

M.8. Conditions

Conditions refer the description of the setting that is needed to collect a meaningful reading for the PM. For example, some performance metrics concerning the team performance would require that the number of team members be above a specific number. Also, volatile data collected during live forensics might require specific processes or network traffic be switched off or temporary disabled while collecting a measurement. The same would be applied to a statistical metrics which might require the domain size be above a specific value in order to make a meaningful reading.

Specifying the conditions is not required for all metrics. However, neglecting specifying the conditions of measurement collection may result in inaccuracies in measurement or interpretation.

M.9. Attributes

The definition of some metrics require the specification of attributes. Unlike the conditions which focus on collection, attributes impact how metric readings are interpreted. For example, the form type and number of surveyed individuals do impact how a survey results are interpreted. Also, in some settings the type of software used in measurement or analysis impact how results are interpreted. Other factors include, who conducted the PM reading analysis, environment context and incident context.

M.10. Interpretation

Each PM needs guidelines on how it should be interpreted in reference of performance. For instance, does a high value indicate good or poor performance? Does the PM reading demonstrates actual measure or a trending measure? Does the variance in the metric reading reflect performance levels or not? Does the metric reading need coupling with other PM readings?

3.5.2 PE Analysis Models

When the structure of the PE framework is designed and it is equipped with measurement tools, it is desirable for the CSIRT to foresee how results produced from deploying the PE framework will be analyzed. Doing this provides more objective assessment that endures less overhead compared to a situation when analysis methods are outlined after the completion of incident handling.

A list of performance analysis methods were extracted from the performance analysis literature, see Table 29. These methods are examined for adaptation to the CSIR field. The list is also used in developing the Integrated Analysis & Validation Model (IAV) proposed in Section 5.2.

#	Analysis Model	Description	Category
N.1	Gap Analysis	Difference between optimal / desired performance and actual performance	(Comparative) Component Analysis
N.2	Benchmarking	Compare incident performance against industry best practices and competitors' performance	
N.3	Targeted Analysis	Perform thorough analysis of a specific aspect of incident handling or confine analysis to a restricted focus to identify potential enhancement venues	
N.4	Bottleneck Analysis	Analyzes the PE results to spot performance bottlenecks, i.e. those that have high impact on poor performance	Deficiency Analysis
N.5	Root-cause Analysis	Asks why poor performance happened.	
N.6	Goal Achievement	Evaluates the success of achieving the PE goals	Holistic (System) Analysis
N.7	Trend Analysis	Analyzes performance over a period of team seeking the identification of recurring performance issues	
N.8	Stakeholder Analysis	Conducts PE analysis from the perspective of various stakeholders	
N.9	Predictive Analysis	Evaluates the system performance under several hypothetical situations	Predictive Analysis

Table 29: PE Analysis Methods

The above nine analysis techniques provide different perspectives to performance analysis. The techniques also vary in the validity degree of the performance results, as some techniques could be applied informally. The main tradeoff of these techniques is objectivity and operationalization, i.e. producing validated results vs. producing results that reflect the operational nature of CSIR and result in actionable plans.

The above techniques can be grouped based on objectivity or based on the analysis subject. The NIST document uses a binary classification for assessment methods in the form of subjective and objective methods. Through this classification, the analysis techniques of N1,N2 and N7 demonstrate a higher level of objectivity due to their reliance on quantified metrics. The remainder of the techniques can be subjective or objective depending on the selected method of analysis.

In terms of analysis subject, the techniques can be categorized into four main categories: component analysis techniques, deficiency analysis techniques, holistic or system analysis techniques and predictive analysis techniques.

Component analysis techniques focus on analyzing separate parts of the system independent from the remaining parts. This allows for more focuses analysis and better identification of technical and/or performance issues. On the other hand, holistic or system analysis techniques view the system as a whole, and analyzes performance through the combined interaction of various system components. The DHS document [162] uses the terms component metrics and enterprise level metrics (ELMs). This is analogous to the usage of component analysis and holistic analysis.

Deficiency analysis targets unfolding performance bottlenecks and factors that impact performance. Deficiency analysis can be applied on the component level or on the holistic level. Finally, predictive techniques are advanced methods that involve comparing actual performance to hypothetical scenarios of what could have been done.

A CSIRT is not required to apply all of the above techniques. Instead, each team should select methods that are suitable for the operating environment in a manner that provides objectivity to the analysis. A list of recommendations of how to select analysis techniques is presented in Table 51 (Appendix A).

N.1. Gap Analysis

Gap Analysis is one of the most common techniques for analyzing performance [163] [164] [165]. The technique measures the difference between the desired performance and actual performance. The results should recognize the aspects of the response that are lacking in terms of performance or the extent of the performance problem [135] .

In the context of CSIR, the first question that arises is what is the *optimal* performance? The theoretical answer can be found in the Universal PE Framework (UPEF), see Section 5.1, in which optimal can be mapped to the *idealistic* values. However, in many situations achieved performance can be too far from this optimal reference, due to practical considerations. Therefore, the optimal performance can be defined in operational terms as *maximum potential* performance. This has been highlighted by [166] as a more realistic method for applying gap analysis. The author referred to maximum potential as: “reasonable” performance.

Yet, the above does not solve the issue, as defining “reasonable or “maximum potential” performance in absolute terms is nontrivial. Potential performance is sensitive to factors related to the operating environment and advancement of technology, which makes the definition relativistic.

The above discussion suggests that the a CSIRT can either:

1- Perform gap analysis over the ideal values and maximum potential values and report both results

2- Perform gap analysis over the maximum potential values, associated with a justification of how these values were calculated and considering some margin of error.

The second issue arises from reading the following example of objective assessment in the NIST document [1]: “measuring the difference between the initial impact assessment and the final impact assessment”. Indeed, this applies gap analysis but from a different angle, which is the difference between perceived values and actual values, or estimated values and actual values. The objective of such analysis is to measure the effectiveness of the estimation techniques of the CSIRT, or the ability of the team to make decisions under unpredictable conditions. Specifically, it could be used in models like the NFP unpredictability platform presented in Section 5.5.

Based on the above, the term ‘gap analysis’ can be refined in the context of CSIR to mean: “difference between optimal and actual performance, or between perceived and actual values for parameters in the incident response”.

N.2. Benchmarking

See Section D.6: Benchmarking.

N.3. Targeted Analysis

Targeted analysis is a generic term used here for any analysis motivated by a specific objective normally targeting a system component or a performance aspect. It is noted in [163] that targeted analysis comes as a third step after gap analysis and root-cause analysis. Thus, targeted analysis is geared towards more precise identification of the causes or towards finding solutions.

N.4. Bottleneck Analysis

Performance bottlenecks are factors that render performance. This can happen through causing system failure or causing the system to operate under poor performance conditions. The removal of these factors normally cause a significant boosting to the system performance. Bottleneck analysis is the method of finding these factors. This process is normally done after gap analysis if it establishes that a specific incident response reflects negative performance characteristics [167]

Bottleneck analysis is normally conducted through analyzing a large pool of post-incident data. However, it could also be performed during the preparation phase when specific tests are performed, like scalability testing [168]. Furthermore, trend analysis data could be essential to the identification of bottlenecks.

Causes of performance bottlenecks could be due to system capacity, response delays or incompetency. The following list of bottleneck causes were reported by several CSIR practitioners to be common [169] [170] [171]:

1- Reliance on tools which might be ineffective or outdated .

2- White noise, i.e. large number of false alarms and redundant alerts

3- Failure to adequately set priorities

4- Lack of competency from responders

There is no standardized method for conducting bottleneck analysis. It has been even argued that it is more of an art than a science [172]. However, the general approach is to conduct analysis over incident data, and especially trend analysis results. The analysis could focus on processes (i.e. series of actions) or on actors (e.g. CSIRT members, software/tools). Identified bottlenecks are analyzed for correlations with various factors which are later validated through a process of elimination.

N.5. Root-Cause Analysis

While gap analysis wants to find if a problem exist, root-cause analysis focuses on what caused it [163]. Although there are attempts to propose formalized models for conducting root-cause analysis that involve digital evidence [173], most organizations use informal or semi-formal root-cause analysis techniques mainly conducted by the technical teams [174]. One of the objectives of the current CSIRT Metrics Special Interest Group is to develop models for root-cause analysis of operational benefit [52]

Three example methods of how root-cause analysis could be applied to the field of CSIR had been suggested by [29]. These methods are: the five-whys approach, why-because analysis (WBA) and cause-and-effect (fishbone) diagrams.

The five-whys approach is borrowed from the Six Sigma DMAIC methodology (Define, Measure, Analyze, Improve, Control) [175]. The approach is simple and can be applied to a wide range of scenarios. It starts by writing down a problem, e.g. performance issue, and then asking why it happened. A tentative answer is provided which is treated as another problem statement and the why question is asked again. In each iteration, the problem is broken down further hopefully up to the root-cause of the issue. In each iteration, the team needs to reach an agreement on the cause or revisit the previous iteration for enhancing the problem statement.

The why-because analysis (WBA) [176] is another iterative approach that seeks causality for accidents. The approach builds a why-because graph (WBG) that analyzes the accident and produces a chart that connects causes to effects. The intermediate connecting items are called factors, i.e. their occurrence contributed to the occurrence of the incident. These factors are categorized and analyzed to construct a list of countermeasures that could prevent the incident from re-occurring.

The fishbone cause-effect diagram is another cause-root analysis tool that is used in Six Sigma, proposed by the Japanese quality scholar, Ishikawa [177]. The analysis method is visual and is used during brainstorming sessions for unfolding the causes of an incident. This method is effective in determining the root-cause if it is a product of several causes. Potential factors are grouped under categories like: rules/policies, equipment/software, environment, people …etc. The output of the analysis is a chain of causes in the form of a process that led to the incident.

N.6. Goal Achievement

Goal Achievement analysis, along with trend analysis and stakeholder analysis are types of holistic system analysis techniques. This category of analysis focuses on investigating the CSIR system as a whole, compared to reporting performance on segregated components of the system.. Using the structure of PE framework presented in this project, holistic analysis techniques should be used whenever a CSIRT selects “overall performance” as one of its performance aspect.

Features of using holistic analysis techniques include:

1- Ensure that performance analysis does not fall into the *fox paradox* [137]. The paradox suggests that good performance in partial aspects of the system might not necessarily reflect good overall performance.

2- Be able to spot performance issues arising from the interaction between various system components [107].

3- Focus on use of sustainable performance measurements compared to component analysis that might use performance metrics volatile to incident circumstances and used technologies [178].

Assessing goal achievement is a common practice in organizations and businesses, both at the scope of long-term strategic goals and at the short-term project-based objectives [151] [19] [78]. The CSIRPE framework presented in this project is goal-oriented as definition of goals is considered the highest system level definition from which performance aspects, indicators and metrics are derived.

The following approaches could be used to evaluate the achievement of goals:

1- A simple Yes/No assessment of whether each goal has been achieved. This is relevant when goals are defined through well-specified targets. Therefore, it is more likely to be used for evaluating strategic sub-goals, or goals set for a specific unit or group of PE activities. An example of CSIRPE goal evaluated through this approach is when a CSIRT sets a threshold time given for the CSIRT from the moment an incident (on a specific severity scale) is detected to the moment of the first CSIRT meeting concerning that incident.

2- A calculation of the percentage of achievement. This could be achieved in two ways: action-oriented or process-oriented analysis. In the action-oriented approach, achieving a goal is mapped to performing a group of activities or sub-targets, regardless of these activities being correlated or exclusive. Level of achievement at each activity is measured, and the total goal achievement is calculated bottom-up. The process-oriented is suitable when achieving a goal is defined through a plan that outlines a process or roadmap towards fully achieving the goal. The percentage of achievement would be a measure of the amount achieved through the process line towards the end-goal.

3- A qualitative achievement analysis. This approach addresses the issue that not all goals can be measured in quantifiable fashion. In such scenarios, qualitative measures could be used to assess how the CSIR achieved the set goals. For instances, goals involving terms like flexibility, adaptability, team cohesion and trust are more likely to be measured through qualitative methods.

N.7. Trend Analysis

Trend analysis conducts performance analysis over a period of time or over a group of incidents. There are two major benefits coming from using trend analysis in performance measurement. First, the analysis transcends good and poor performance due to the unusual circumstances which provides more realistic representation of the system performance. Second, trend analysis captures the effectiveness of the enhancement processes and lessons learnt analysis through demonstrating whether the system performance is improving or deteriorating after the implementation of recommendations.

Although trend analysis is independent of individual incident analysis, it is a good practice to apply performance measurements on both fronts [135]. For instance, after recording response time for a specific incident, the average response time could be inspected for increase or decrease [23].

The benefits of conducting trend analysis in the context of CSIR are summarized by [29] and [74]:

1- Identify patterns of cyber security incidents

2- Detect any increase or change in the number and type of vulnerabilities

3- Identify common factors that influence the occurrence of incidents and proper incident handling

4- Determine the effectiveness of security controls

5- Identify targeted areas of the organization

6- Understand costs and impacts associated with cyber incidents

In addition, note that trend analysis could be performed jointly by multiple organizations that have established information sharing agreements. Such trend analysis can be helpful in determining attack trends and mechanisms, and consequently which countermeasures were effective.

N.8. Stakeholder Analysis

This category of analysis techniques suggest that performance should be evaluated from various perspectives, each corresponding to the perspective of a group of stakeholders. It has been shown that applying financial metrics into non-financial aspects is problematic [134]. A brief discussion is provided below about three common techniques that fall under this category: balanced scorecards, the performance prism and triple bottom line performance system.

The Balanced scorecards performance measurement model [134] is considered one of the early mechanisms triggering multi-perspective performance analysis. The model was originally designed to reform the traditional practice of institutions that rely only on financial performance metrics like return on investment. The model analyzes performance against four perspectives: financial perspective, customer perspective, internal perspective and innovation and learning perspective. The model is widely used by major international corporations. It is also used to derive some information security performance systems [179]. However, the model was criticized for its bias towards the economic aspect of performance [178], along with other practical challenges [180].

The prism performance model [181] attempts to expand the focus on financial stakeholders to include other organizational perspectives. The model analyzes performance through five perspectives: stakeholder satisfaction, strategies, processes, capabilities and stakeholder contribution.

The triple bottom line model [182] expands further to include stakeholders that have no transactional interests with the organization, like local government and human services. The model analyzes performance from three main perspectives: economic, social and environmental. In other words, the model goes beyond business and organizational aspects to include external and indirect stakeholders. It has been argued in [178] that this wide approach creates more sustainable performance systems.

N.9. Predictive Analysis

Predictive analysis techniques are more advanced methods that not only include collected performance data for analysis but also consider hypothetical scenarios of what could have been done. Such methods are used across disciplines for purposes of evaluating performance [82] [155] and for evaluating the reliability [103] and robustness [155] of response systems.

Predictive analysis is bidirectional, i.e. it inspects the past and the future. When analyzing past actions, the focus is on what could have been done better, or what could have been the output if an alternative course of actions were implemented. The future inspection attempts to foresee the impacts of the implementation of some control measures or security policies. The simplest method of inspecting both directions is through the *what-if* analysis approach. Below are some examples of how this approach could be used in the context of CSIR performance evaluation:

1- What-if specific performance metrics were deployed and measured [41].

2- How specific incident activities and their countermeasures impact infrastructure [52]

3- Conducting cost-benefit analysis (CBA) or break-even Analysis for the cost effectiveness of incident handling procedures [183].

4- Expected rewards of implementing preventive measures [31] over the system security and CSIRT performance.

5- Forecasting future performance issues, e.g. failures or bottlenecks, before they arise based on current system states [181]

3.5.3 PE Validation Models

The last step in the framework development is investigating the validity of the performance models. Prior to this discussion, an understanding of what 'validity' means in the context of CSIRPE is presented.

In simple terms, *measurement validity* is ensuring that numbers and scores actually represent what they claim to be [184] [23]. Expressed in other terms, to ensure that users cannot circumvent the results of the performance metrics [135]. Translating this to CSIR terms, validity means ensuring that the performance results correspond to actual incident handling performance.

However, validation has another dimension pertaining to design, which is to ensure that the designed performance system actually captures the expected needs. This is sometimes distinguished by the term 'verification'. This usage is compatible with the wide-spread Six-Sigma model that is used for measuring performance in business and manufacturing environments. The Six-Sigma considers validation of performance models as part of its DFSS (Design for Six Sigma) process which is structured in five phases: define, measure, analyze, design and validate [175].

In this project, the term validation is used generically to include the above two aspects of validation. In addition, the above usage of the term should not be confused with other common usages of validity in the field of computer security which revolve around the concept of data integrity.

As it was documented in the literature survey, several works like [34] have documented that the field of computer security is poor in validation methods. However, this remark was criticized by its narrow definition of validation which correspond to formal methods.

Analyzing works concerning performance validation, the used methodologies could be classified into four types:

1- ***Formal Methods:*** Use rigorous mathematical models for data validation and logical arguments for reached conclusions

2- ***Heuristic Methods:*** Instead of performing complete validation, a probability could be assigned to how much confidence should be put on the results.

3- ***Development methods:*** argue that if the design and process of collecting and analyzing data is valid, then the results are ought to be valid.

4- ***Operational methods:*** disregard theoretical analysis and view the experience of the practitioners as the main validator for what works and what does not work.

It was argued earlier in this project that focusing on formal methods is of little benefits to CSIR practitioners, due to the complexity of the process and the operational nature of the discipline. The later three methodologies will be discussed by giving one or two examples, and highlighting advantages and disadvantages. Four methodologies are presented in Table 30.

#	Validation Model	Description	Category
V.1	Feedback Systems	Validate design through getting operational feedback of what worked and what did not	Operational Validation Method
V.2	Compliance	Review policies, procedures and processes for compliance with industry standards	Development Validation Method
V.3	Bootstrapping	Achieve validation through involvement of management in the development process	Development Validation Method
V.4	Confidence Rating	Assigning a confidence rating to various aspects of the design and analysis	Heuristic Validation

Table 30: Validation Models for CSIRPE Frameworks

V.1. Operational Validation (Feedback Systems)

Validation through operational experience is the main practiced method for enhancing CSIR, which is also extendible to CSIRPE. Team members engage in post-incident analysis to discuss the positive and negative aspects of incident handling. Consequently, positive factors are maintained and negative factors are countered through correcting measures. The process is iterative as it is repeated after each incident handling. For a list of actions performed during post-incident analysis see [26] [59] [1].

The main arguments for using the operational model go back to three factors. First, computer security and so CSIR are operational disciplines. Thus, operational models and techniques resonate more naturally with CSIR practitioners. Second, research in the area of security metrics suggests that validation is difficult and impractical. Since CSIRPE relies on security metrics, this challenge is passed on. Third, since computer technologies and cyber threats are on continuous change, it is easier to develop operational measures instead of long-term planning which might not be applicable when technologies or threat tactics change. In other words, the overhead of keeping robust validation methods is too high taking in consideration the ever-evolving nature of the discipline.

On the other side, relying on operational findings which suffer from subjectivity and are sensitive to short-term factors can be counterproductive. Some classical works in military training asserted that relying on experience for learning does not improve people [185] [13]. Therefore, unless operational feedback processes are evidence-based and are governed by objective procedures, operational feedback outputs cannot be trusted as a validation method [15].

V.2. Development Validation (Compliance)

Validation through compliance is a method to ensure that a developed model is consistent with the industry requirements and best practices. When a CSIRT develops a process or model, e.g. performance model, establishing compliance suggests basic validation that the proposed model is *good*.

Since standards are developed through careful and collaborative efforts of industry and research experts, it provides a sufficient indicator to an organization that its implemented system is *sufficiently good*. This brings the advantage that an organization will grow through the growth of the industry as a whole, and it suggests that the organization need not to devote resources to developing *better* solutions, unless it is necessary.

The disadvantage of relying on compliance as a validation scheme is that standards are normally higher-level and do not necessary address details that might be crucial to the success of the model. This is true about the NIST and CERT documents which only provide very basic information with regards to performance evaluation.

V.3. Development Validation (Bootstrapping)

Unlike compliance which ensures that a developed model meets industry expectations, bootstrapping ensures that a developed model meets the organization expectations. The term bootstrapping refers to a development method in which the developers regularly validate the design with the management and constituents throughout the development stages [129]. The term can also be used to refer to the validation investigation results, through iterative validation with various teams about the findings of each investigation steps [186].

The main advantages of bootstrapping is its simplicity and its assurance that the designed model is aligned with the organizational mission and objectives. However, caution should be observed as relying solely on organizational needs without considering best industry practices may produce ineffective models.

An example of how bootstrapping could be used to enhance CSIR performance system can be found in [161]. The study targeted developing and enhancing system trustworthiness through constructive system design and paying attention to requirements. Trust is defined here as a higher measure composed of security, reliability survivability and other sub-system measurement tools.

Another example applicable to the CSIRPE framework presented in this project, is the validation of performance metrics design. After designing and formulating a list of performance metrics, it is a good practice to review the list and perform a quick review. This review needs to be done by several individuals, other than the developers, to ensure that the list meets the expectations. This review, which is a form of bootstrapping validation, can be done in iterations: first by the technical teams, then by the quality staff and finally by the management and the CSIRPO.

V.4. Heuristic Validation (Confidence Rating)

Heuristic validation methods use techniques that rely on probability instead of clear-cut metrics. It could be argued that since security can never be guaranteed, then so validation of security systems cannot be certain. Using heuristic approaches, systems are validated through expressing how much confidence is being exemplified. Heuristic approaches are used in computer security metrics [162]. For instance, password strength is determined by heuristic approaches to a predefine thermometer.

Using the above understanding, heuristic approaches could be used to validate performance design and also performance results. The design is validated heuristically by inspecting elements like reliability, trust, failures, risks and survivability [103] [183]. Performance results can be validating through assigning a confidence rating to the results, similar to the methods used by [73] [42].

Although heuristic approaches provide practical and simple solutions to a complicated problem, they could suffer from subjective interpretation. For instance, what does it mean that a CSIRT executed a response that is in the range of [90-80%] cost effective. Such statement can be interpreted by different parties, e.g. financial officers vs. technical staff, in different ways. Therefore, heuristic approaches need to follow clear definitions and assessment processes, which could be argued to resemble effective qualitative assessment.

3.6 Phase IV: Implementing a CSIRPE Model

3.6.1 PE Functional Models

While the PE Design Process focuses on *actions* to build a PE model for a specific environment, it only describes actions that are performed at the preparation stage. Another model is needed to address when and how PE actions are executed during the life cycle of a response. These models will be referred to as: “PE functional models”.

The design of PE functional models highly depends on when PE analysis is expected to be performed, which is expressed in the design parameter *D.5: Analysis Time*. Three functional models are presented below corresponding to the three possible values for parameter D.5. The presentation of these models takes into account Assumption A2.1 about the use of the Hybrid Model for the IR life Cycle.

#	Name	D.5 Value	Description
F.1	Design-Collect-Analyze (DCA)	Post-Incident	Measurements collected during incident handling, all analysis done afterwards
F.2	PE Monitor	Continuous	PE is monitored throughout the incident response, frequently assessed and results are fed-back to the CSIRT
F.3	Incremental Model	Incremental	PE is analyzed at pre-defined points during the incident handling process.

Table 31: Types of PE Functional Models

F.1 The Design-Collect-Analyze (DCA) Model:

The Design-Collect-Analyze (DCA) model is the simplest and most commonly functional approach used by CSIRTs. The model concentrates performance evaluation activities to the preparation and lessons-learnt phases, i.e. the first and last phases of the CSIR life cycle. During incident handling, performance activities are confined to collecting measurements for the pre-defined performance metrics, which could be done through automated tools. The main advantage of the DCA model is that it has minimal overhead during incident handling, giving responders more time to focus on the technical aspect of the response. The shortcoming is its lack of mechanisms to make corrective actions during incident handling based on performance data.

The DCA model is to be used when parameter D.5 is set to “Post-Incident”. A graphical depiction of the DCA function model is presented in Figure 15 applies.

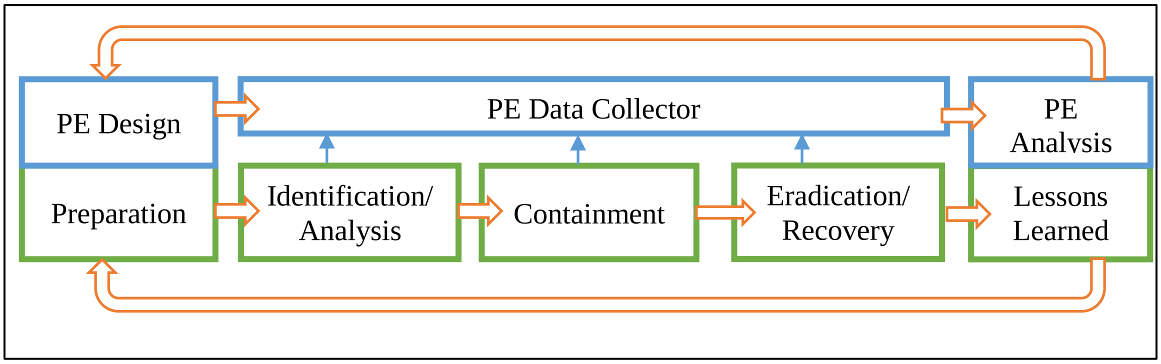


Figure 15: DCA Functional Model

F.2 The PE Monitor Model:

The PE Monitor Model uses an on-going approach to performance evaluation. During incident handling, a “performance monitor” is used to assess and report performance throughout the incident life cycle, see Figure 16. This could be either done through setting up performance triggers that send warnings/alerts about the performance status of the response calling for the attention of the CSIRT leadership to take corrective actions.

The PE Monitor model is suitable for complex incident responses that needs high level of interaction or occupies long response time. In such scenarios, taking corrective actions based on performance results would impact the cost-effectiveness of the response and enhance the response time. For example, it is suggested that PE monitoring be used for the effectiveness of the mitigation of security incidents targeting nuclear facilities [79].

The Monitor model is to be used when parameter D.5 is set to “Continuous”. The type and level of rewards expected from adopting a PE monitor model to the field of CSIR remains shady. At the basic level, there is no immediate industrial need for using such model [7].

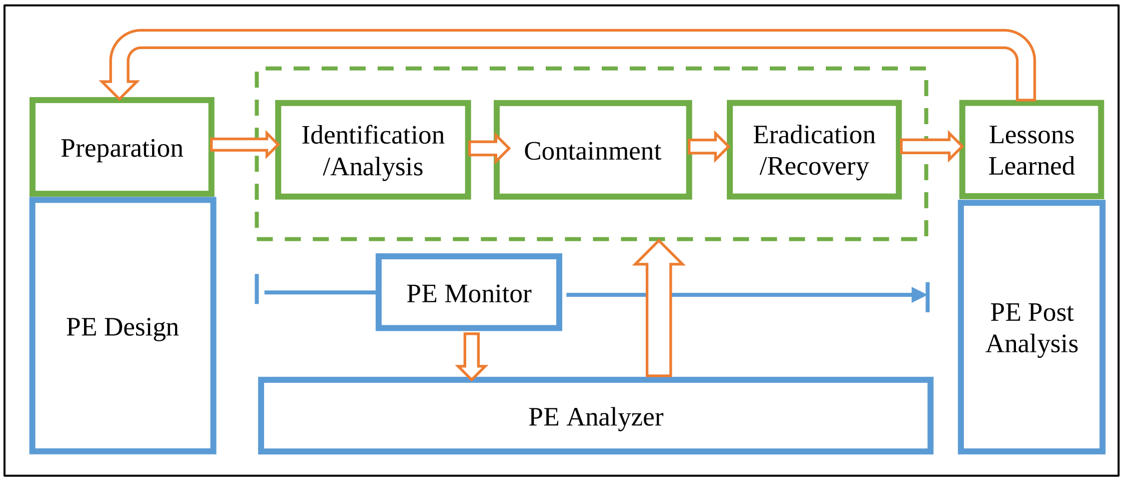


Figure 16: PE Monitor Functional Model

One of the main implications of adopting the PE Monitor Model is the need to develop another type of performance measurements, called interactive PMs. The classical performance metrics are generally classified as diagnostic or mechanistic tools, while interactive PMs are tools for monitoring and control [187]. It has been noted that despite interactive PMs being more “organic” the emphasis of the performance measurement disciplines had been on the diagnostic tools [188] .

F.3 The Incremental Model:

The Incremental Model is a hybrid model that strikes a balance between the benefits of the DCA and PE Monitor models. It follows the basic structure of DCA of condensing performance evaluations at the preparation and lessons learnt phases, with one exception. At specific *marks* of the incident handling, some partial performance analysis is performed for purposes of the on-going incident procedures.

The performance evaluation conducted at these marks are condensed and launched for specific purposes. It can be as simple as a short report on response time and cost, or more informative with a summary of current values to a short list of performance indicators/metrics readings. The main motivation of making such intermediate PE analysis is the early detection of any major deficiencies in performance to enable planning for corrective actions.

The marks at which incremental analysis is performed can be phase-based, i.e. after each phase or a group of phases; at specific time slots during incident handling, e.g. main CSIRT meetings; or at/prior to some major activities, e.g. media release, first executive summary. Since performance evaluation done at these marks should be simple and informative, it can be considered as semi-analysis or partial analysis that will be updated later during the more comprehensive analysis done at the lessons learnt phase.

The incremental model avoids the high overhead associated with continuous monitoring but at the same time keeps the team alert if major performance issues need to be addressed. In its simplest form, the incremental analysis can have a single mark during incident handling. This could be mid-time-point of the incident handling, i.e. after detection and initial containment and before full containment and recovery. A depiction of such scheme is presented in Figure 17.

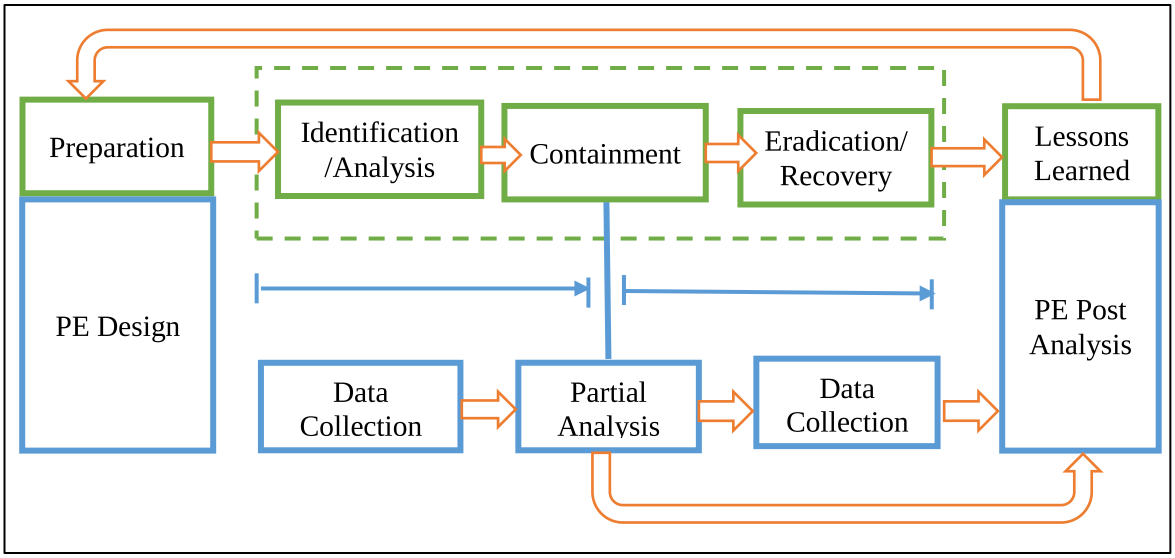
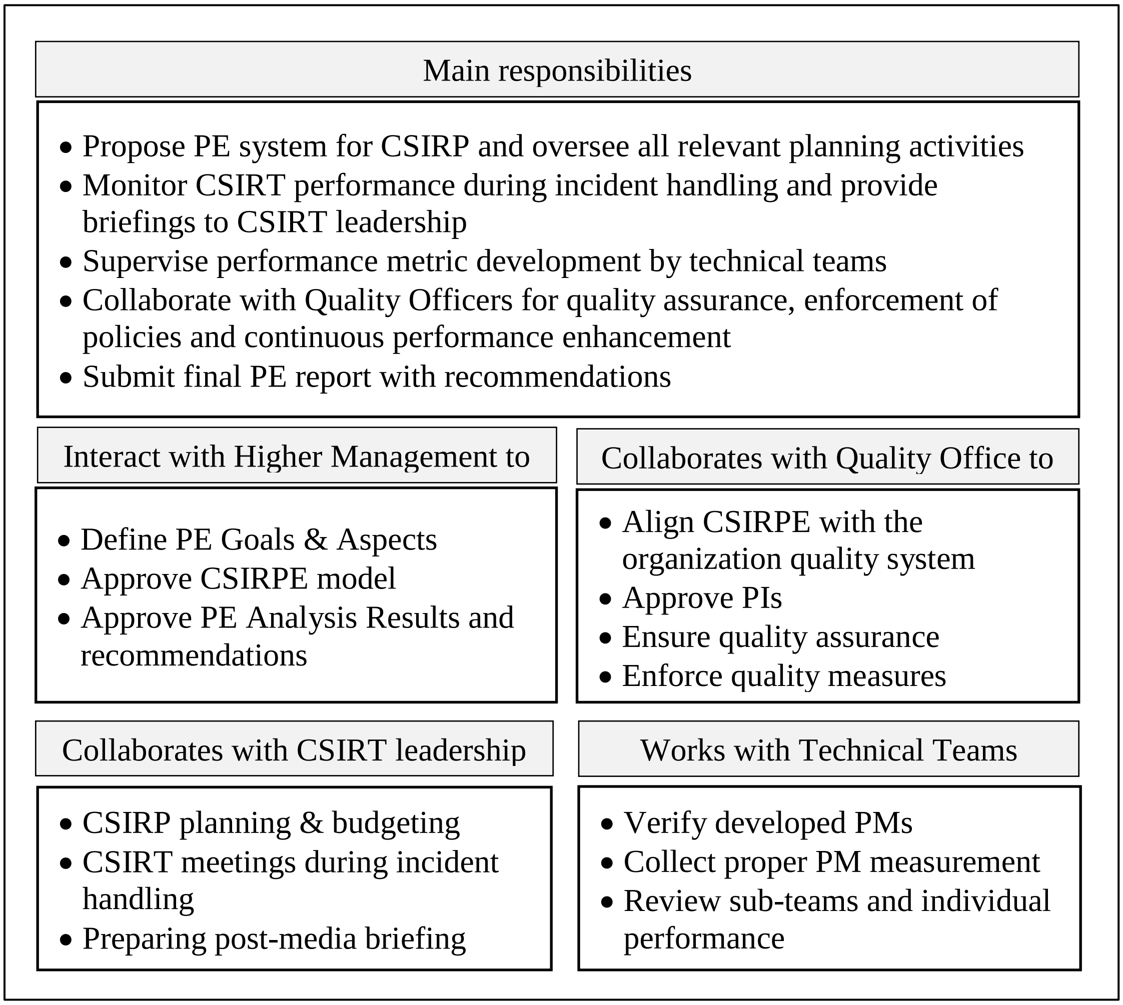


Figure 17: Incremental PE Functional Model with a Single-Mark

3.6.2 Assigning Roles and Responsibilities

Discussion about role assignments could have been presented in earlier sections but was postponed to this part of the dissertation to provide a description of various activities before discussing the expected roles.

There are various activities related to performance evaluation that need to be done prior, during and after incident handling. It is impractical to assume that all of the above duties will be carried by a single individual as this involves high overhead and will result in poor execution. From the management point of view, this is also undesirable because it gives the impression that performance is the responsibility of an individual not the whole team; and it tends to view performance evaluation as a separate activity instead of an integrated activity within the CIR life cycle. However, it is reasonable to assume that the management of the performance evaluation activities will be coordinated by an individual. This individual will be referred to as: “computer security incident response performance officer (CSIRPO)". For simplicity, it is assumed here that the CSIRPO is a CSIRT member other than the CSIRT leader.



Main responsibilities

- Propose PE system for CSIRP and oversee all relevant planning activities
- Monitor CSIRT performance during incident handling and provide briefings to CSIRT leadership
- Supervise performance metric development by technical teams
- Collaborate with Quality Officers for quality assurance, enforcement of policies and continuous performance enhancement
- Submit final PE report with recommendations

Interact with Higher Management to

- Define PE Goals & Aspects
- Approve CSIRPE model
- Approve PE Analysis Results and recommendations

Collaborates with Quality Office to

- Align CSIRPE with the organization quality system
- Approve PIs
- Ensure quality assurance
- Enforce quality measures

Collaborates with CSIRT leadership

- CSIRP planning & budgeting
- CSIRT meetings during incident handling
- Preparing post-media briefing

Works with Technical Teams

- Verify developed PMs
- Collect proper PM measurement
- Review sub-teams and individual performance

Figure 18: CSIR Performance Officer Responsibilities

The CSIRPO collaborates with four main bodies within an organization: the CSIRT leadership, the higher management, the quality office and technical teams. A summary of main responsibilities of CSIRPO is outlined in Figure 18, and a collaboration chart is presented in Figure 19.

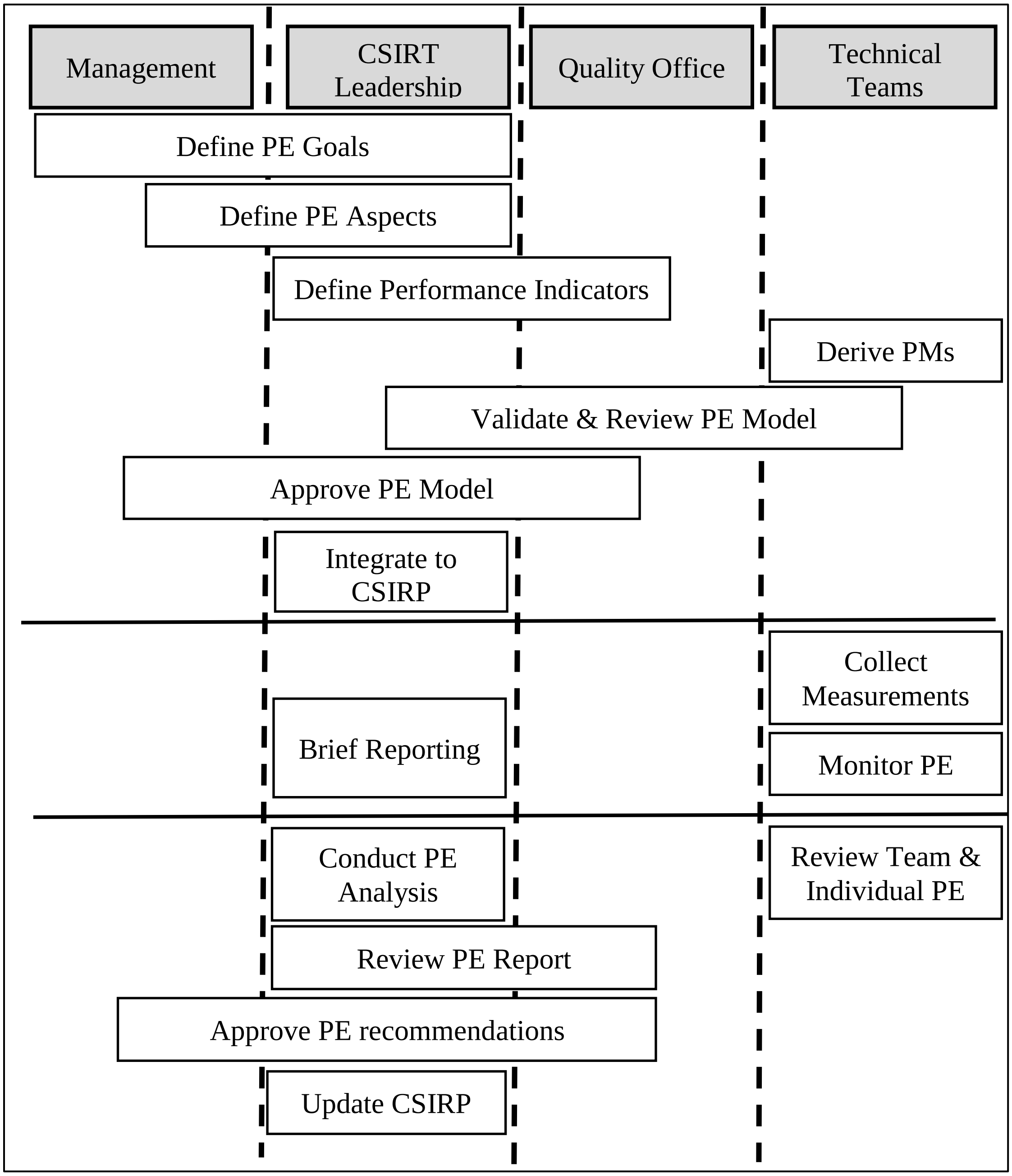


Figure 19: Collaboration Chart for CSIRPO and other officers

There are some basic features that should be satisfied with whoever is to be selected for the "performance officer" role. A list of these features is provided in Table 32. In small CSIRTs, the performance officer can be the same as the leader of the core CSIRT, while in larger CSIRTs, it can be one of the sub-teams leaders. In both scenarios, the performance officer is not hired for the sole purposes of performance evaluation, as it would be more practical and cost-effective to assign that role to an existing CSIRT member who would have other tasks related to incident handling.

	Feature	Rationale
1	Understands the organization’s quality assurance policies and procedures	To design PE framework that is aligned with the organization’s quality assurance system
2	Understands the management structure of the CSIRT	To properly delegate tasks among the CSIRT members, avoid redundancies, and avoid conflict of powers
3	Participates in the preparation or review of the CSIRP	To be able to integrate PE within the CSIR documentation and implementation
4	Values the need for PE within CSIR	Viewing PE as an additional unnecessary burden is detrimental to having an effective design and inspiring team members
5	Has some leadership role within the CSIRT	To be empowered to recommend, and if necessary to enforce, changes
6	Has basic understanding of the technical aspects of CSIR	To help design effective performance metrics that have practical impact

Table 32: Desired Qualifications of the Performance Officer (CSIRPO)

3.6.3 Integrating CSIRPE into CSIRP

The method of integrating a CSIRPIE module into a CSIRP depends on two factors: the type of CSIRP and the type of CSIRPE. As discussed in Section 2.1.4, a CSIRP can be design to reflect the organizational perspective, to be action-oriented, to facilitate maintainability or use a mixture of the three approaches. The selection of the document structure will influence how CSIRPE will be embedded in the document. Also, the complexity of the developed CSIRPE module will influence the level of details that will be presented in the CSIRP.

Transcending these variances, any CSIRP for a CSIR capability that uses a PE module need to at least include the following five elements:

1- A list of PE goals/objectives that outline how a CSIRPE is expected to serve the mission and the organizational goals
2- How does the CSIRPE relate to the overall quality framework of the organization?
3- A higher-level description of the CSIRPE module and its components
4- Which aspects of the CSIR system will be subject to evaluation?
5- What are the key performance indicators?

There are two basic questions on how the above five points combined with the other details of the CSIRPE need to be presented in the CSIRP:

1- Should the CSIRPE be presented in a separate section in the CSIRP, or should its details be integrated across the existing sections?
2- Which information should be presented in the basic CSIR document, and which details should be compiled in a separate appendix?

Answering the above two questions requires drawing a balance between making clear presentation of the objectives and mechanisms of performance evaluation, and making sure that this information does not distract a reader from the understanding the main CSIR components. For instance, it is expected that details about performance metrics and analysis and validation techniques would be put in an appendix.

In Table 33, a checklist for information that needs to be included in a CSIRP, whether in the basic document or in the appendices, is presented. The checklist can be used for updating an existing list or as a guideline when preparing a new one.

#	Checklist Item
1	Does the CSIRP has a listing of the PE goals or a description why CSIRPE is used?
2	Does the CSIRP has a summary of the CSIRPE module?
3	Does the CSIRP describe how the CSIRPE fits into the organization’s quality framework?
4	Does the CSIRT structure include a CSIRPO? If not does the document assign a primary owner for PE?
5	Does the CSIRP describe which aspects of CSIR would be subject to performance evaluation?
6	Does the CSIRP outline the performance indicators and identifies the KPIs?
7	Does the CSIRP has formal definition of all PIs and PMs?
8	Are the PIs aligned to goals, and PMs aligned to PIs?
9	Does the CSIRP outline when PM readings should be collected and analyzed (i.e. functional model)?
10	Does the CSIRP include a section about the analysis techniques that will be used to analyze the PM results?
11	Does the CSIRP outline a mechanism for validating performance analysis results?
12	Does the CSIRP outline when and how PE results will be reported?
13	Does the CSIRP outline mechanisms for documenting and maintaining PE results?
14	Does the section on the review phase of the CSIR life cycle include PE analysis in the lessons learnt and the enhancement process?

Table 33: Checklist for CSIRPE integration into a CSIRP

CHAPTER FOUR

PERFORMANCE INDICATORS

The objective of this chapter is to formulate a list of performance indicators (PIs) that CSIRTs can apply to measure performance and generate performance metrics. Because organizations define their goals in various ways, it is practically difficult to generate a confined list of performance indicators that is applicable to all environments. Therefore, this chapter focuses on developing performance indicators that measure the major CSIR activities and are more likely to be applicable to a "conventional" CSIR environments.

The PIs presented in this chapter were selected through the following criteria:

1- Provide a broad coverage of the CSIR activities, by generating enough PIs that correspond to each phase of the CSIR cycle. Several PIs are also presented to measure the overall performance.

2- Satisfy the major CSIR requirements set by the NIST standard [1] and the CERT document [2]

3- Satisfy the incident response requirements set by the National Response Framework (NRF) [110] and National Incident Management System (NIMS) [109]

4- Capture the major recommendations of works that investigated some aspects of CSIR effectiveness like [61] [74] [9] [97] [52]

For each performance indicator, a definition is proposed that is relevant to how CSIRTs operate. A classification of the PI is provided based on its scope. This is followed by a brief summary of design and interpretation considerations for using the PIs. An example or two is provided concerning possible performance metrics that can be derived from each PI.

The chapter is composed of three sections. The first section discusses methods of classifying CSIR PIs. The second section inspects guidelines of how an organization should select its PIs from the provided list. The last section, which composes most of the chapter content lists provides the definition and detailed information about fifty PIs. The list of PIs is ordered alphabetically.

4.1 Classifying Performance Indicators

Four possible classification schemes could be used to categorize CSIR performance indicators. The first scheme considers the value of the PI to the organizational goals by distinguishing Key Performance Indicators (KPIs) from regular PIs. This scheme was discussed in Section 3.4.4 and is integrated in the framework design. However, this scheme is contextual and PIs will be classified differently depending on where there are deployed.

The second scheme considers the measurement method, i.e. qualitative or quantitative. Despite the wide use of this scheme, it is incompatible with the design of the proposed framework. It is suitable for frameworks that treat PIs and PMs interchangeably; while this project treats PIs as a higher level measurement tool from which PMs will be derived. Performance Indicators are designed to be generic such that various tools could be used to measure it, which can be qualitative or quantitative or contain a collection of both. Therefore, both the first and second approaches are not appropriate for a general classification of PIs that is neutral to the environment.

The third approaches considers the scope of the PI. Performance indicators can be classified as generic or specific. Generic PIs focus on the overall performance of the CSIR system, while specific PIs introspect a specific performance component of the system. A general method to identify these specific components is to map it to the CSIR life-cycle. Using this perspective and using A2.1, specific PIs can be further broken into the following categories: [preparation, identification, containment, eradication & recovery, analysis]. Note that PIs concerning the “analysis” category can be considered as PIs for measuring the effectiveness of CSIR performance frameworks, which is not the focus of this project. Therefore, developing “analysis PIs” is limited for the purposes of this project.

The fourth categorization scheme is derived from the discussion of S.6 about viewing performance in terms of CSIRP or CSIRT. A performance indicator can inspects the CSIR design and preparation, i.e. activities *before* the occurrence of an incident. These PIs are classified as CSIRP PIs, because the CSIRP is the main reference for readiness. On the other hand, PIs that inspect the execution of a CSIRP can be considered CSIRT PIs. These performance indicators are mainly interested in measuring performance *during* the incident handling.

Different methods for categorizing PIs are possible to develop. For example, the CSIR Balanced Scorecard model presented in Section 5.3. can be used to classify PIs based on the stakeholder perspective. However, the classification of PIs in this chapter will be limited to the third and fourth schemes due to their simplicity and potential applicability to a wide range of CSIRTs.

Finally, the above classification schemes should not be treated as *hard* classification techniques. The schemes are considered *soft* as there is a large intersection between various categories. Therefore, categories will be selected based on the relevance strength whenever there is an intersection.

4.2 Selecting Performance Indicators

Out of the large pool of PIs presented in the following section, a CSIRT would eventually settle on a small selection of PIs. This suggests that there needs to be guidelines on how many and which PIs to select.

It is noted in [135] that having a small number of metrics has been proven to provide practical advantages, as it is difficult to focus on more than five to seven indicators at the same time. However, the authors note that this applies to PIs focusing on strategy, while PIs focusing on process can be in hundreds. The NIST document on information security performance measurement suggests using ten to twenty metrics [189]. The threshold of twenty is also endorsed by [190].

Selecting which performance indicators to use depend on multiple factors, like:

1- *CSIRPE Specifications*: The specifications of the developed CSIRPE module impact the selection of PIs, as some PIs might be incompatible or infeasible to measure. For instance, a CSIRPE that relies on quantitative measures (D.7 = "Quantitative) will exclude all PIs that cannot be measured quantitatively. Another example is when a CSIRPE is scoped to measuring the execution effectiveness (D.8 = "CSIRT") is not fit to use PIs that are classified under the "preparation" category.

2- *CSIRT Structure*: For instance, using the sufficiency PI is incompatible with a distributed CSIRT that relies on external entities. Also, a small size CSIRT, e.g. composed of two or three individuals, might not be best to implement PIs concerning communication and team dynamics.

3- *CSIR Goals*: The goals of the CSIR, outlined in a CSIRP, determines areas of CSIR that are of high importance to each specific CSIRT. For example, a

CSIRT order to pay high attention to the protection of classified data would be interested in using PIs like Confidentiality, Containment Effectiveness, and Shielding Effectives; compared to a CSIRT that focus on ensuring the availability of digital resources which may be more interested in PIs like: Availability and Continuity. Some PIs are only applicable to special situations like attacking host identification which only applies to when prosecution is needed.

4- *PI Interdependency*: As a general rule, in order to make reasonable assessment, the analysis need to read results of several PIs at the same time [7]. This is due to the fact that PIs demonstrate high interdependency, in which performing well accordingly to one PI would normally result in performing well according to other

To expand on the above last point about PI interdependency, it is noticed that PIs relate to each other in the following ways:

1- *Inclusive:* a PI can be generic such that several PIs are considered its subcategories. For instance, The Containment Effectiveness PI includes the following PIs: Confinement Effectiveness, Shielding Effectiveness and Mitigation Effectiveness.

2- *Complementary:* Several PIs can measure components of a specific aspect. For instance, the Comprehensiveness and Completeness PIs measure different aspects of CSIRP effectiveness. However, both PIs compositely constitute the overall effectiveness of the plan, and thus are better to be analyzed together.

3- *Contentious:* some indicators attempt to measure aspects of performance that are tension with other PE aspects. For instance, the Conformance and Flexibility PIs demonstrate opposite objectives.

4- *Influence:* Some PIs influence the results of other PIs. For instance, Containment Effectiveness impacts the Stability PI. Another example, detection effectiveness PI impact the result of containment effectiveness.

5- *Exclusive:* The situation in which PIs are independent of each other, and could be analyzed separately. For example, the consistency PI is exclusive to the Availability PI.

From the above discussion, I agree with the recommendation of selecting few, i.e. ten or less, performance indicators. However, these indicators should be considered the key performance indicators (KPIs) that focus on the main PE aspects as defined by the designated CSIRT. At the same time, several other PIs can be used for the purposes of enhancing performance and improving the functionalities of the CSIR as needed. These PIs can be used as needed, and can be consigned to sub-teams in the CSIRT to evaluate their own activities.

With regards to which PIs to select, a CSIRT would start by selecting PIs that are exclusive to each other. Then, based on the defined goals, outlined PE aspects the team expand by adding PIs that are relevant to each other depending on the needs. For instance, a CSIRT that have just prepared a CSIRP might use several PIs that measure the effectiveness of the plan. These PIs can be used for the following years until the plan and its update procedures has reached a maturity state. If the team later finds out, for example, that the detection phase is ineffective, then several PIs under the category of “Detection/Identification” could be introduced.

In summary, a CSIRT needs to have static and dynamic list of PIs. The static list constitutes of few generic KPIs that offer comprehensive assessment of the CSIR capability. On the other hand, the dynamic list constitutes of PIs that are deployed for several used based on the enhancement requirements of the CSIR.

4.3 CSIR Performance Indicators

PI.1.	Accuracy	
Definition	The proximity of predictions and estimations made by the CSIRT during incident handling to actual or reasonable range of values validated post-incident	
Category	Identification	CSIRT
Design & Interpretation Considerations	This PI measures the effectiveness of the CSIRT's decision making capability under uncertain conditions There are two domains for measuring accuracy: 1- Accuracy of estimations made during the incident response to measures validated during post-incident analysis 2- Accuracy of predictions prior-to-incident with actual events during the incident The first domain is the main focus of the Accuracy PI as the second domain is more concerned with preparedness PIs. When deriving PMs from this PI, there are two methods to define accuracy: 1- Compare against actual values (e.g. incident classification) 2- Compare against a reasonable range of values (e.g. estimation of number of compromised records – it is difficult to get exact number at the early stages of the response). Another consideration when defining accuracy PMs, 1- Does the PM measure level of proximity (i.e. how far from actual values), or 2- Was acceptable accuracy achieved (e.g. was the classification accurate? Yes or No)	
PM Examples	Let incident severity be classified in the range: $[S_1, S_2, \dots , S_n]$ Accuracy of incident classification = $\mid S_{declared} - S_{actual} \mid$	
Relevant PIs	Identification Effectiveness	

PI.2.	Adaptability	
Definition	The ability of the CSIR capability to effectively adjust to the change of needs or to changes of incident circumstances	
Category	Generic	CSIRT
Design & Interpretation Considerations	For meaningful application of this PI, the environment should be dynamic (i.e. contains several non-static factors). When designing adaptability PMs, the following two design factors need to be considered: 1- Does the PM considers adaptability to predicted conditions, unpredicted conditions, or both? 2- What is being measured? Is it the adaptability inputs (e.g. member training), the adjustment process (e.g. how did the team adjust), or both When analyzing adaptability PMs: - It is more reasonable to record the ability of the team to adjust to a list of changes more than attempting to measure the "degree of adjustment" - Is failure to adapt a result of poor preparation or inability to adjust to some other factors?	
PM Examples	Prior to defining adaptability PMs several descriptive PMs need to be defined with regards to adaptability dimensions and levels. For example, • adaptability to scalability changes (e.g. [100, 1000, 10,000] infected hosts) • adaptability to incident dimensions (e.g. [incident severity, incident scale, incident type]) Adaptability PMs would measure differences or ratios along multiple adaptability scales [191]. For measuring individual ability to adapt in changing environments see [192]	
Relevant PIs	Flexibility, scalability, survivability	

PI.3.	Attacking Host Identification (AHI)
Definition	The ability of the CSIRT to accurately identify the attacking host(s) in a timely manner

Category	Analysis	CSIRT
Design & Interpretation Considerations	It is a desirable outcome of incident handling to identify the attacking host. However, as indicated by NIST [1] this should not be the focus of responders. Examples of situations that require using this PI: - The incident results in request for legal investigation - The technical requirements for recovery or containment procedures require identifying the attacking host - The incident is targeting government data or assets, especially critical infrastructure. Ability to identify attacking hosts is an indicator of good detection and analysis capabilities of the CSIRT, but the reverse is not necessarily true This PI is applicable when root-cause analysis is used. However, it only measures a specific aspect of root cause analysis, not the effectiveness of the whole analysis process. For example, the root cause of an incident can be determined to be a DDoS and resolving the issue does not require knowing much about the identity of the attacking host. However, legal prosecution requires more specific details To measure effectiveness of the attacking host identification: 1- Level of data collected (e.g. does the data identify the perpetuators? Does the data allow for legal prosecution? Is there information about associated individuals/groups? 2- Is the information accurate? (how much confidence does the analysis provide?) 3- Is it done in a timely fashion? (data is provided when it was needed, e.g. when making media release or court testimony). 4- Was the identification process cost effective?	
PM Examples	*AHI confidence*: a probability is calculated about the confidence of results of AHI process, based on collected facts. For instance, if the attacker is confirmed to be from a specific subnet that contains 10 hosts, then the probability of each is 0.1. (Note: A CSIRT needs to present facts, not make judgements. The first is objective and the second is subjective).	
Relevant PIs	Root-Cause Identification	

PI.4.	Availability	
Definition	The ability of a CSIR to conduct a response that prevents and minimizes the unavailability of service-providing resources	
Category	Eradication & Recovery	CSIRT
Design & Interpretation Considerations	The above definition considers a CSIRT’s ability to protect availability. The above PI can be redefined to mean the availability of the CSIRT itself. Such definition would make the PI a sub-category for the reliability, capacity and preparedness PIs This PI is more generic than the continuity PI. For instance, business continuity can still be achieved with the unavailability of some resources. Unavailability can be either due to the incident itself, or to a decision taken by the CSIR for prevention or remedial purposes. Performance metrics need to distinguish between the two. Availability is normally measured by time percentages. However, it is more reasonable to couple that with business outputs, e.g. number of customers, loss [193] Several levels of availability can be defined based on factors like quality of service or number of users	
PM Examples	availability $= \frac{AST-DT}{AST} \times 100$ *AST*= agreed service time, *DT* = down time (see ITIL v3 [194] [195])	
Relevant PIs	Survivability, Continuity	

PI.5.	Capacity	
Definition	The boundaries of the CSIR system that define the maximum potential outcomes and quality for the offered services	
Category	Preparation	CSIRP
Design & Interpretation Considerations	The above definition need not be confused with how the term is used in some CSIR publications which used “incident response capacity” and “incident response capability” interchangeably [60] [31] [136]. It is desirable to have high system capacity as it reflects good preparation and grants higher confidence in the response system.	

	High capacity systems are also expected to offer faster responses and more adaptability to incident changes. It was practically proven that offering effective and efficient response requires sufficient capacity [111] The Capacity PI is mainly defined in terms of descriptive metrics, which can be mapped to performance metrics. Some benefits for using capacity PMs include 1- Recognizing when the IR system reaches its maximum capacity. In such case, seeking additional or external support is necessary 2- It can be used in developing preventive measures. For example, if networks are working under near maximum capacity, then the likelihood of a successful DoS is high. [1] Some of the factors that could be used in defining capacity PMs: 1- number and availability of responders and support staff 2- Competency and readiness of responders 3- Type and number of equipment and software 4- Type and quality of CSIR offered services. Some limiting factors to capacity include jurisdictions, budget and company size.
PM Examples	Example 1: Defining maximum response capacity (RC_{max}) in terms of reliability probability is explained in [103] Example 2 [111] [109]: 1- Identify CSIR resources 2- Record the capacity of each resource 3- Identify CSIR services 4- Map each service to the resources 5- The capacity of each service is the minimum capacity of the associated resources
Relevant PIs	Preparedness, Utilization, Scalability, flexibility

PI.6.	Communication Effectiveness	
Definition	The ability of the CSIRT to establish communication links that enable efficient transmission of data	
Category	Generic	CSIRT

Design & Interpretation Considerations	Communication effectiveness can be assessed from the following different angles. Note that the above definition emphasizes the first two, but it could be re-defined as appropriate: 1- Are there established communication links and what is their reliability? 2- How much resources and overhead does the communications cost? 3- What is the impact of CSIR communications on incident handling decisions and actions 4- What is the value of the content of communications in terms of correctness and being timely? Some features of effective communication: 1- *Accessible:* to all those involved or affected by the incident [110] 2- *Coordinated:* enables different responders to learn about updates and taken actions throughout the CSIR life cycle 3- *Low Redundancy*: no unnecessary repeated messages are exchanged 4- *Resilient:* withstand and continue to perform after damage absence of resources 5- *Low Overhead:* messages provide concise content that serves the purpose of the communication exchange. 6- *Confidentiality:* observed whenever required Although low redundancy is a desirable feature, this should not be confused with the need to establish alternative or back-up methods of communication as a measure of preparedness [109] Communication effectiveness might be impacted by the level of trust established between various parties [136] it also may be impacted by the level of competency [79] The assessment of communication effectiveness can focus on inter-communication between CSIRT members, intra-communication with constituencies within the organization, and communicating with external agencies and/or the public
PM Examples	The following report discusses some generic metrics for assessing communication effectiveness in incident response [196]. Examples include: financial investment in communications capabilities, capacity of mass notification, frequency of updating contacts and communication links, and compliance with communication requirements (e.g. encryption).

	Several works suggest the need to test and assess the communication between those responsible for patch management and the CSIRT. See [74] [1].
Relevant PIs	Partnership Effectiveness, Team Cohesion

PI.7.	Competency	
Definition	The sufficient mastery of the knowledge, skills and abilities needed to perform CSIR tasks. (adapted from [61]).	
Category	Preparation	CSIRT
Design & Interpretation Considerations	The difference between competency and readiness: - *Competency:* focuses on knowing in terms of technical information and necessary skills - *Readiness:* focuses on applying the acquired knowledge and skills Competency could be defined in terms of competency levels. For instance: 1- Minimum competency [74]: knowledge and skills for performing basic CSIR activities 2- Quality competency: needed to perform diverse CSIR activities with quality outputs 3- Advanced Competency: needed to perform advanced and complex CSIR activities CSIRT members need not only be competent with current practices and tools, but also with new technologies, potential threats and trending solutions. There are several scales for competency. For instance: - *Individual vs. Team:* Is the team as a whole competent as well as each individual? - *Technical vs. non-Technical:* Do all CSIRT technical and non-technical members (e.g. support teams) demonstrate competency	
PM Examples	Continuous Professional Development [7]: What type and how frequent are CSIRT members offered professional development programs? See the CERT publication “Competency Lifecycle Roadmap: Toward Performance Readiness” [61]	
Relevant PIs	Readiness, Preparedness	

PI.8.	Competitiveness	
Definition	The ability of a CSIRT to deliver services in a quality level similar to, or higher than, the quality level offered by its peers	
Category	Generic	CSIRT/CSIRP
Design & Interpretation Considerations	This PI offers a mechanism for comparing the performance of the team to that of its peers. Competitiveness can use: 1- *Absolute measures:* comparing performance against expected quality as defined by standards and best industry practices 2- *Relative measures:* comparing performance against performance of other peer institutions over similar incidents In business settings, this PI is normally measured in terms of financial metrics [197]. However, for CSIRTs this should focus on effectiveness metrics This indicator is best used when benchmarking is adopted as an analysis method Using this PI helps an organization Identify performance areas which are potential for enhancement Using this PI requires minimum normalization at two levels: 1- Institutional Level: a CSIRT is compared to one of similar size and capability, 2- Incident level: comparison is performed over incidents of similar category	
PM Examples	Depends on which areas are selected for peer-comparison	
Relevant PIs	Compliance, Conformance	

PI.9.	Completeness (of CSIRP)	
Definition	Preparing a CSIRP that addresses *all* essential aspects of incident handling in terms of policies, procedures and support.	
Category	Preparation	CSIRP
Design & Interpretation Considerations	Completeness is an indicator for preparedness and readiness For each IR activity, there needs to be some outlined policies or procedures for handling, and how and where to seek support	

	A CSIRP does not necessary need to have pre-prepared solutions. it only needs to have instructions on how to approach all possibilities There are two time frames for measuring completeness: - *Prior-IR Completeness:* can be assessed over the completion of all of the planned activities produced in the design phase. - *Post-IR completeness:* if during incident handling the CSIRT is faced with a procedural issue that has no guidance in the CSIRP then it is a sign of incompleteness. Completeness has another two dimensions: the completeness of the documentation of incident handling and the completeness of performance analysis CSIRP completeness is integrated into the NFP Unpredictability platform presented in Section 5.5
PM Examples	Prior-IR completeness: - Break a CSIRP into components - Translate each component into actions needed for completion - Define completion status levels or percentage scale for completion - Assess the completion of each activity and validate/approve by higher management See the list of metrics outlined in [74]
Relevant PIs	Comprehensiveness, Preparedness, Readiness

PI.10.	Compliance	
Definition	The degree of conformance of CSIR processes and procedures to standards and regulatory specifications	
Category	Generic	CSIRP
Design & Interpretation Considerations	CSIR Compliance has three aspects: 1- Ensuring that the outlined CSIR processes and procedures of the CSIRP is in conformance with CSIR standards 2- Ensuring that an incident handling is in conformance with (non-CSIR) legal and organizational standards and regulatory policies	

	3- Ensuring that CSIR is in compliance with organizational policies There are no formal CSIR standards. However, awaiting formal standardization, the NIST [1], CERT [2] or SANS [30] documentations can be treated as standards. Examples of government and industrial standards that a CSIR might need to comply with include: ISO 27001 on Information technology and security [198] and the Health Privacy Rules (HIPAA) [199]
PM Examples	Compliance can be confirmed through certifications or external auditing
Relevant PIs	Conformance

PI.11.	[CSIRP] Comprehensiveness	
Definition	Issuing a CSIRP that outlines procedures for *all* categories of incidents	
Category	Preparation	CSIRP
Design & Interpretation Considerations	Having a CSIRP that is capable of handling the various types of incidents is a sign of preparedness. The PIs Comprehensiveness and Completeness are complementary to each other. The comprehensiveness PI ensures that the CSIRP is applicable to a wide spectrum of incidents; while completeness ensures that there are detailed procedures for handling the defined spectrum of incidents. The term *Comprehensiveness* alternatively can be used to refer to the performing of comprehensive performance analysis. See the Documentation Effectiveness PI. Due to the variance and the continuous emerging nature of security incidents, it is impossible to outline handling details for every possible incident, Therefore, procedures should be outlined for every group of similar incidents. To achieve comprehensiveness, a “comprehensive” classification scheme is needed, i.e. is capable of mapping an incident to a pre-defined category of incidents, which in turn are associated with handling procedures. The incident classification scheme may differ depending on the business sector (e.g. incidents in health sector can use distinct incident classification from financial institutions). For	

	classification of cyber incidents from the perspective of educational institutions see [200] and for nuclear agencies perspective see [79]. Whenever an incident cannot be classified under the defined categories, a procedure should be outlined in a CSIRP on how the incident should be handled.
PM Examples	Comprehensiveness can be measured through exhaustive testing or simulation for different categories of incidents. For instance, the DHS NCCIC cyber security incident scoring system [201]could be used. A CSIRP need to contain enough information on how to handle each incident level.
Relevant PIs	Completeness, Documentation Effectiveness

PI.12.	Confidentiality	
Definition	The ability of the team to secure and maintain the confidentiality of the data in the system during incident handling	
Category	Generic	CSIRP/CSIRT
Design & Interpretation Considerations	The above definition points to three aspects of confidentiality. When defining this PI in a specific environment, the designers need to specify which points are considered: 1- The CSIRT preventing confidentiality compromise after incident detection 2- The CSIRT mitigating and minimizing confidentiality compromise impacts 3- Conducting a response that conforms with the confidentiality policies The usage of this PI presumes that an information security assessment was conducted in which: 1- Digital data that needs to be retained confidential are identified 2- The confidentiality level of each data is defined 3- Security policies are implemented to maintain the confidentiality of the data. Confidentiality metrics could be defined using: 1- Security metrics used in the CIA model 2- Requirements set by privacy laws 3- Liabilities defined in contracts 4- Information security standard compliance requirements	

PM Examples	Root privilege count metric proposed in [190] can be adapted to measure how many root privileges were compromised, restored or modified during the response. The above metric provides an indication for potential level of compromise to confidentiality.
Relevant PIs	Availability

PI.13.	Confinement Effectiveness	
Definition	The ability of a CSIRT to identify and isolate infected portions of a system and to prevent the negative impacts of the incident of spreading to the healthy portions	
Category	Containment	CSIRT
Design & Interpretation Considerations	This PI is similar to the containment effectiveness PI; however, this PI is focused on preventing the spread, while containment is more generic and involves reducing harm and preventing escalation Ineffective confinement can indicate: 1- Poor identification mechanisms for infected areas 2- Poor isolation mechanisms 3- Flaws in the security architecture, e.g. too many dependencies or inconsistencies. 4- Team incompetency This PI is applicable when the infected parts of the system has been identified. Therefore, inputs from the identification phase would impact measurements of PMs derived from this PI. To measure confinement, it should be clearly defined how infected vs non-infected portions are identified. In that direction, it is possible to define levels of confinement depending the level of infection dripping to the healthy portions.	
PM Examples	Confinement Strength Metric (CSM): 1- Divide the system into “confinement zones”. The smallest zone (Z_1) contain all portions confirmed to be infected. The largest zone (Z_n) contain the entire system. The remaining zones are associated with probabilities of infection. 2- Define infection levels through characteristics of the threat, in which fully infected is denoted by (F_n) and non-infected is denoted by (F_0). 3- A scoring table is to be developed over the values of Z and F. The highest score would indicate the least likely parts of	

	the healthy portion getting fully infected. The lowest score would indicate the highly vulnerable parts getting the lowest level of infection. 4- Assuming that CSM = 1 means full strength, i.e. proper conferment; CSM is defined as 1 minus the score obtained from the scoring table defined in step 3. Higher values of CSM indicate relatively good confinement, while lower CSM values indicate poor confinement.
Relevant PIs	Containment Effectiveness, Shielding Effectiveness

PI.14.	Conformance	
Definition	The ability of a CSIRT to execute a response that abides by the outlined procedures and processes in a CSIRP	
Category	Generic	CSIRT/CSIRP
Design & Interpretation Considerations	In an ideal scenario, a comprehensive and complete CSIRP is sufficient to guide responses to any incident. However, building such CSIRP is practically difficult, expensive and sometimes infeasible. This indicator targets both the effectiveness of execution and effectiveness of the CSIRP design Nonconformance could indicate: 1- Failure or poor performance from the CSIRT 2- Incorrect or inconsistent CSIRP design 3- Incomplete CSIRP design and poor planning 4- None of the above (i.e. it was to not conform, e.g. decision by higher management to overcome a policy or procedure) Conformance can extend to cover other procedures and policies followed in the organization, not only the CSIRP. Measuring conformance is only meaningful when *requirements* are defined. These requirements can be security or quality requirements. An organization can decide for zero tolerance for nonconformity or define a range of tolerance for nonconformity Since conformance is too generic, the definition of this PI can be scaled down to conformance to requirements that impact CSIR performance. It is best to interpret this PI along with the flexibility PI, as both PIs are contentious.	

PM Examples	The degree of conformance can be applied to some aspects like timetables. A PM would provide the difference between actual time and scheduled time A qualitative measure in the form of a checklist that assess the degree of conformance (e.g. minor, major and critical nonconformance). See ISO 9001 [202] and ISO 27001 [203].
Relevant PIs	Compliance

PI.15.	Consistency	
Definition	The ability of a CSIR to follow consistent procedures for detection, analysis and reporting of incidents [109]	
Category	Generic	CSIRP
Design & Interpretation Considerations	Consistency is a desirable feature of CSIR because it reflects good design, enables comparative analysis and facilitates identifying performance obstacles. The CERT document considers consistency essential for CSIR [70] and several works echo the need for consistency in security metrics [123]. It remains challenging of how to ensure consistency in qualitative and subjective measurements [78] Automation is an effective tool for enhancing consistency [145] Consistency domains can be classified as: 1- consistency of procedures (e.g. detection methods) 2- consistency of measurements (e.g. collection methods) 3- consistency of reporting and notification	
PM Examples	Designing quantitative PMs for consistency is non-trivial. Most common methods use a qualitative method of using a checklist that verifies that a specific procedure or data representation is consistent with the response or quality requirements	
Relevant PIs	Conformance, compliance	

PI.16.	Containment Effectiveness
Definition	The ability of a CSIRT to effectively prevent an incident from spreading, escalating, or causing further damage after its detection

Category	Containment	CSIRT
Design & Interpretation Considerations	This PI can be considered the highest-level indicator for the containment phase. The PIs: Mitigation effectiveness, confinement effectiveness and shielding effectiveness are special of this PI. Containment effectiveness strongly impacts incident stability. Therefore, both PIs should be read together. Measuring containment effectiveness is sensitive to accuracy of and effectiveness of the identification / detection processes. Therefore, usage of multiple identification techniques is recommended before starting the containment process [1]. Examples of factors that contribute to containment effectiveness include: 1- Containment time 2- Containment/mitigation rate/level 3- Containment cost 4- Degree of harm reduction (quantification of how much harm is saved through applied containment procedures) 5- Containment robustness (maintain effectiveness despite change of circumstances) The above factors can be defined over a spectrum of containment levels, e.g. partially contained and fully contained.	
PM Examples	See containment factor defined in Section 5.1.3.	
Relevant PIs	Confinement effectiveness, shielding effectiveness, mitigation effectiveness, Stability	

PI.17.	Continuity Maintenance	
Definition	The ability of a CSIRT to execute a response that maintains business continuity	
Category	Generic	CSIRP/CSIRT
Design & Interpretation Considerations	The CERT document considers business continuity and disaster recovery planning as one of the services that a CSIR offers under the category of security quality management [2] Maintaining business continuity is a joint task between the CSIRT and business continuity planning team. [131]. Therefore, the	

	integration and communication between the two teams need to be effective [70] [198]. This PI is especially relevant to incidents that involve denial of service (DoS) attacks [1] One of the common mechanisms used to ensure business continuity during incident handling is called application whitelisting in which safe applications are permitted to run and all other applications are blocked [56]. It is important that maintaining business continuity does not result in loss of digital evidence [204].
PM Examples	Impact of business continuity management on breach cost [4]
Relevant PIs	Availability, Survivability

PI.18.	Coordination Effectiveness	
Definition	The ability of various security and support teams in an organization to collaboratively execute a response that uses effective communication and reporting and efficient allocation of resources	
Category	Generic	CSIRP/CSIRT
Design & Interpretation Considerations	There are various levels of coordination [31]: 1- Coordination among CSIRT sub-teams 2- Coordination between CSIRT and security teams in the organization 3- Coordination between CSIRT and support teams, e.g. logistics and human resources 4- Coordination between CSIRT and management, e.g. executives Coordination must occur throughout the phases of the CSIR life cycle [74] Elements of effective coordination: 1- Leadership, i.e. every task has a leading team to avoid duplication of efforts [2] 2- Centralized point for information dissemination [167] 3- Proper management of task interdependencies [104] 4- Effective communication Preparedness contributes to better coordination among different teams [109]	

PM Examples	A qualitative analysis of the CSIR system that involves inspecting 1- Acknowledgement of shared goals and objectives 2- task ownership (are there tasks with no owners? Are there duplicate owners for the same task) 3- task process associated with sub-outcomes.
Relevant PIs	Partnership Effectiveness, Communication Effectiveness, Team Cohesion

PI.19.	[Response] Cost	
Definition	Amount of financial resources dedicated to responding to an incident	
Category	Generic	CSIRP/CSIRT
Design & Interpretation Considerations	CSIR Costs can be classified into four categories: 1- *CSIR Cost:* Fixed Costs associated with establishing, maintaining and operating a CSIR 2- *Response Cost:* Costs endured due to incident handling including detection, containment and recovery 3- *Incident Cost:* Financial loss due to the incident 4- *Incident Maintenance Cost:* financial resources needed to implement post- incident recommendations Although it is desirable to minimize response cost, this should not be on the expense of effective response. Other indicators like cost effectiveness strike that balance Cost can be actual or estimated (e.g. when averages are used)	
PM Examples	Incident Cost estimation proposed by [205]: *incident_cost = non_productive_cost + response_cost* *non_productive_cost = U*hup*V*DT* *response_cost = A*hap*V*RT* *U* = number of affected users, *A* = number of responders *hup* = average user's page, *hap* average responder pay *DT* = hours of downtime, *RT* = response time *V*= overhead cost (vary from company to another) For another incident cost estimation scheme see: [200]	
Relevant PIs	Cost Effectiveness	

PI.20.	Cost Effectiveness	
Definition	The degree at which the CSIR capability was able to save on costs while achieving the same quality of outcomes [206]	
Category	Generic	CSIRP/CSIRT
Design & Interpretation Considerations	This PI strikes a balance between the organizational need to decrease costs and the need to fulfill CSIRT quality requirements Cost can be compared to latest cost, average cost, or average cost obtained from a benchmark In order for cost calculation to be meaningful, cost needs to be compared to similar incidents in terms of severity level and delivered quality service level	
PM Examples	Cost Saving per data record = V - L/N N = number of compromised records L = total loss due to data breach V = average loss per record due to data breach (benchmark value obtained from [4])	
Relevant PIs	Response Cost	

PI.21.	Detection effectiveness	
Definition	The ability of a CSIR to accurately detect incidents in a timely manner.	
Category	Identification	CSIRP
Design & Interpretation Considerations	The two main elements of detection effectiveness are: 1- Accurate identification of the incident (distinguish between events and incidents) and classify it under the correct severity scale 2- Detecting the incident in a timely manner Other factors that could be used in defining detection effectiveness include: 1- Intelligence capacity 2- Detection cost 3- Incident classification method 4- Mechanism to distinguish between incident precursors (incident might happen in the future) and incident indicators (incident has occurred or is occurring) [1]	

	For the purposes of using this PI, incidents can be classified based on their detection complexity. An example of such classification is: 1- *Detectable attacks*: attacks known to the CSIRT and relevant procedures are defined in the CSIRP 2- *Resolvable attacks*: attacks undefined in the CSIRP, but solutions are available by consulting external entities. 3- *Zero-day attacks* [56]: Attacks are unknown to the CSIRT and no solutions exist yet.
PM Examples	For precise measurement of detection time, the metrics defined by the VERIS project can be used [73]: 1 ***First malicious action time (FMAT):*** Beginning of the threat actor's malicious actions against the victim. ***Initial compromise time*** : First point at which a security attribute (C/P, I/A, A/U) of an information asset was compromised. 2 ***Data exfiltration time(DAT)***: First point at which non-public data was taken from the victim environment. Only applicable to data compromise events. 3 ***Incident discovery time(IDT)***: When the organization first learned the incident had occurred. *Detection time = IDT - FMAT*
Relevant PIs	Containment effectiveness, intelligence capacity.

PI.22.	Documentation Effectiveness	
Definition	The ability of a CSIRT to provide consistent post-incident documentation that is comprehensive and provides effective traceability.	
Category	Analysis	CSIRT
Design & Interpretation Considerations	This is a generic PI that can encompass several PIs like comprehensiveness, consistency and traceability. Documentation comprehensiveness can be defined by: 1- Breadth: There is documentation for each activity taken during incident handling [31]. 2- Depth: There are detailed information recorded for each of the incident handling activities 3- Coverage: documentation is not limited to technical steps. It includes support and secondary activities like legal, financial and media [7].	

	Consistency in documentation contributes to better traceability and allows for comparability and trend analysis Traceability is the ability to quickly access the documented data when needed. Other factors that contribute to documentation effectiveness: 1- Clarity 2- Simplicity 3- Accessibility 4- Conciseness (providing summaries for executives) Note that the above definition does not consider correctness as a factor for effectiveness. It is not required that all documented data be correct because there is a need to document mistakes and wrong decisions and actions for quality purposes.
PM Examples	There are several metrics developed by the Association of Clinical Documentation Improvement Specialists [207] that could be borrowed by CSIR. Examples include: review rate, query rate (e.g. number of queries before getting the right information) and quality/revenue impact.
Relevant PIs	Consistency, Comprehensiveness, Traceability

PI.23.	[Response] Effectiveness	
Definition	A generic indicator about the CSIRT's ability to achieve its desired objectives and to reduce harm in a manner that is timely and uses minimum number of resources	
Category	Generic	CSIRT/CSIRP
Design & Interpretation Considerations	The above definition considers efficiency to be part of effectiveness (See: Section 1.4.2), but a CSIRT may choose to separate the two. In the CSIR context, there are four factors that determine if a specific activity was "effective" 1- Success level of achieving a goal/objective 2- Amount of used resources (financial and nonfinancial) 3- Response Time 4- Level/amount of reduced harm Response effectiveness can also a collective indicator of the effectiveness of the various CSIR phases, i.e.: 1- Preparedness Effectiveness 2- Identification effectiveness	

	3- Containment Effectiveness 4- Eradication/Recovery Effectiveness 5- Analysis Effectiveness
PM Examples	Not Applicable
Relevant PIs	Goal Achievement, Response Time, Response Cost

PI.24.	Eradication Effectiveness	
Definition	The ability of a CSIR to eliminate the key components of the incident with speed and precision. (adapted from [29])	
Category	Recovery & Eradication	CSIRT
Design & Interpretation Considerations	An effective eradication process involves: 1- Fully deactivating the active components of the attack/incident 2- Remedying the system and fixing the vulnerabilities that were exploited 3- Removing the remnants of the incident that may cause the incident to reoccur. 4- Conducting the above with speed and precision [92] Not all incidents would require distinct steps for eradication, as the steps might be included in the recovery [1]. An example is unauthorized access due to password obtained through social engineering. The two most common actions in eradication are patching vulnerabilities and disabling malware code (part of malware analysis processes [208]).	
PM Examples	Since eradication processes are OS and application-specific, designing PMs that are applicable across incidents is not feasible. An example of a PM for eradication that involves vulnerability fixing is the patch dissemination speed discussed in [209]. Scalability and Accuracy are common PMs for measuring the effectiveness of malware analysis techniques [210].	
Relevant PIs	Recovery Effectiveness, Mitigation Effectiveness	

PI.25.	Evidence Retention Effectiveness	
Definition	The ability to extract and retain digital evidence according to the CSIR objectives	
Category	Analysis	CSIRT/CSIRP
Design & Interpretation Considerations	Not only digital evidence is preserved for prosecution purposes, but also for conducting internal intelligence analysis. An “effective” evidence retention is one that is: 1- Cost effective: use least amount of resources 2- Authentic (legal effectiveness): the digital evidence demonstrates authenticity through following proper procedures 3- Timely: availability of digital evidence data when needed 4- Redundant: ensures that redundancy policies are followed to make several backups of evidence [15].	
PM Examples	The NIST document (page 41) suggests two metrics for measuring evidence retention effectiveness: Data Retention Length: For how long is data (e.g. emails) are stored in the storage devices (e.g. 180 days) Storage cost: cost of storage devices and media used for storing evidence.	
Relevant PIs	Forensics Readiness, Intelligence Capacity	

PI.26.	Flexibility	
Definition	The ability of a CSIR system in terms of capacity and readiness to address and adjust to the variations in in incident conditions	
Category	Generic	CSIRP
Design & Interpretation Considerations	The flexibility PI is a higher-level indicator that includes the following three PIs: 1- adaptability 2- scalability 3- capacity (i.e. high capacity) The dynamic nature of CSIR environments makes flexibility a necessity (CERT [2]) Elements of flexibility include: 1- Flexible policies and procedures [2]	

	2- Adaptive performance of CSIRT members [97] 3- The ability to expand and scale [111] 4- Ability to handle any type and scale of incident [109] 5- Ability to carry proactive and reactive measures Flexibilities can be mapped to uncertainties, as each uncertainty requires a specific flexibility [211]. Since flexibility deals with future uncertainty, then it cannot be measured accurately and approximation methods need to be used
PM Examples	A plan need to be tested/simulated for flexibility. The testing need to test flexibility at the micro-level (changing specific conditions of an incident) and at the macro-level (changing major conditions of the CSIR system). Qualitative assessment by experts can be used to report on the CSIRP's flexibility.
Relevant PIs	Adaptability, Scalability, Capacity

PI.27.	Forensics Readiness	
Definition	The ability to conduct forensics investigations and extract digital evidence in a cost effective manner. (Adapted from [204])	
Category	Generic/Identification/Analysis	CSIRT
Design & Interpretation Considerations	This PI is defined in a manner to include all forensics activities. However, it can be broken done to several PIs. The "Evidence Retention Effectiveness" PI is derived from this PI but with more focus on evidence maintenance. This PI also intersects with the "intelligence capacity" PI when applying forensics techniques to extract intelligence information and artifact intelligence sharing [50]. Elements of forensics readiness: 1- The capacity to perform forensics investigations 2- The competency to perform forensics activities 3- The ability to extract and retain digital evidence 4- Determine attacking host and incident root-causes 5- Interaction with law enforcement and legal teams The benefits of having forensics readiness include: minimizing the cost of investigation, blocking attempts to cover internal malicious activities and reducing regulatory and legal disclosure costs [212].	
PM Examples	Metrics for various aspects of forensics readiness can be found in the following works:	

	1- For measuring the relevance of intelligence data to forensics readiness see the framework presented in [186] 2- Cost-benefit analysis techniques for forensics activities are studied in [213] 3- Value of digital evidence extracted from the cloud [214]
Relevant PIs	Intelligence Capacity, Evidence Retention Effectiveness

PI.28.	Goal Achievement	
Definition	A higher level and generic indicator about the extent of a CSIR's achievement of its outlined goals.	
Category	Generic	CSIRP/CSIRT
Design & Interpretation Considerations	Goals here refer to CSIR goals that are derived from the organization's mission Achievement can be measured in several methods like: achievement percentage, threshold surpass, subjective assessment (e.g. surveys), or collective analysis (e.g. cost of achieving multiple goals).	
PM Examples	The CERT publication of [19] provides a platform for evaluating the higher level objectives of CSIR related to the organization's mission. The goals are investigated through assessment surveys focusing on mission *driver*s. Drivers are identified as factors that have strong influence on achieving a mission.	
Relevant PIs	Response effectiveness	

PI.29.	Harm Reduction	
Definition	The ability to reduce or eliminate harm and side-effects of an incident after its discovery	
Category	Generic	CSIRT
Design & Interpretation Considerations	Harm endured due to incidents is normally measured in dollars or lives. In the context of CSIR it is normally the former (i.e. dollars) Estimating harm has similarities to estimating risk. However, risk assessment is more generic. Thus estimating harm can be a viewed as a focused analysis of risk assessment taking the particulars of a specific incident. In addition, risk assessment is performed prior to	

	the occurrence of incidents, while harm reduction measurement is done post incident. To fully quantify the damage of an incident is a challenging task, as highlighted in Chapter 2. However, works like [200] provide practical evidence that producing good approximations is possible.
PM Examples	Calculating how much a response was successful in decreasing harm needs three major estimations: 1- How much actual loss was endured due to the incident (L_i) 2- What is the worst scenario of the damage that would have resulted from the incident (L_{max}) 3- Had the CSIRT not responded, how much additional damage would have resulted? ($L_{avoided} = L_{max} - L_i$) To calculate ($L_{max}$), simple risk assessment techniques that use multiplicative metrics of probability of events by the expected loss [190] can be used. The quantification of harm can be done through loss of revenue, loss of reputation and insurance deductibles [215]
Relevant PIs	Response Cost

PI.30.	Intelligence Capacity	
Definition	The ability of a CSIR capability to collect and analyze intelligence information about current and potential vulnerabilities and breaches in a manner that supports and enhances incident handling	
Category	Detection/Containment	CSIRP
Design & Interpretation Considerations	An intelligence capability of a CSIRT can be divided into two categories: 1- Intelligence gathering and collection capability 2- Intelligence analytics capability. Intelligence collection can be from external sources (e.g. public alerts, information sharing, other CSIRTs) or from internal sources (e.g. information gathered from internal networks). Intelligence analysis provides responders with situational awareness. It aims at knowing motivation, tactics and disruption methods used by attackers [29]. An intelligence capability will also involve how a CSIRT some of its intelligence information with other external agencies,	

	government or business [74]. Therefore, this metric could be linked to partnership effectiveness. Part of intelligence analysis includes artifact intelligence done through forensics analysis [186]. According to a recent SANS survey [8], only one third of business implement intelligence gathering and analysis techniques.
PM Examples	Value of intelligence information can be qualitatively measured through several factors [24]: 1- Robustness level of information assurance (IA) 2- Controls allocated to the protection of collected information like: mission criticality, sensitivity and releasability and perishability. 3- Potential impact of loss of confidentiality, integrity, and/or availability of information A metric for measuring capacity of intelligence gathering is membership with intelligence gathering communities/groups [7]. For example: REN-ISAC for educational and research information [216] and infraGrad for sharing information between FBI and private sector [217].
Relevant PIs	Partnership Effectiveness, Intelligence Capacity, Attacking host Identification

PI.31.	Mitigation Effectiveness	
Definition	The ability to execute ongoing and sustained actions to reduce the probability of, or lessen the impact of, an incident [183]	
Category	Containment	CSIRT
Design & Interpretation Considerations	Mitigation is a higher level indicator that includes confinement/ containment and eradication see NIST [1] and NIMS [109]. Some works, include recovery processes in mitigation recovery [79] Mitigation is mainly concerned with actual threats, but it can also be used for potential risks [19] [24] [25] Mitigation strategies can be developed from analysis of incident data or from advisory/alert data [31] . Actions in the mitigation process include: - Applying threat/risk reducing controls - Applying countermeasures to the security violation - Neutralizing the propagation of the incident	

	- Changing/updating the security infrastructures, e.g. updating firewall filters, IDS signatures, and installing patches [74] - The ability to do the above through assistance of automated software [1]
PM Examples	The DHS publication [162] called researchers to develop baseline measurements for the fraction of infected machines at any moment. Mitigation can be measured through the reduction over time. The absolute/relative risk estimate metric [183] is used to compare the cost of a risk to the cost of mitigating it.
Relevant PIs	Confinement Effectiveness, Containment Effectiveness, Harm Reduction, Eradication Effectiveness

PI.32.	Partnership Effectiveness	
Definition	The extent a CSIRT builds associations with external bodies to exchange information and achieve common goals to serve the CSIR objectives	
Category	Generic	CSIRP
Design & Interpretation Considerations	The CERT document [2] defines the following types of Partnerships: 1- Education or training 2- Out-of-hours coverage 3- Technical expertise 4- Cooperative work 5- Other opinions 6- Point of contact to other teams or experts Partnerships can also be established for project-based objectives, for sharing of information, or for long sustainable relationships. Any assessment should distinguish between these three types. Although assessment of partnership effectiveness can be performed by one of the parties, it is better to be collaboratively performed by all partners involved. Assessments of this PI would either focus on evaluating the level of achievement of the agreed upon goals or measure the gap between the intended and actual level of participation.	

PM Examples	A list of metrics for measuring partnership effectiveness for educational institutes are found in [218]. These metrics are generic and can be applied to CSIR. For assessing partnership that involves information sharing, the following metrics could be used [2]: 1- Confidentiality and secrecy 2- Appropriate use 3- Disclosure 4- Proper acknowledgements
Relevant PIs	Preparedness, coordination effectiveness

PI.33.	Preparedness	
Definition	A generic PI that encompasses CSIR activities performed before the occurrence of incidents which contribute to readiness, competency and efficient allocation of resources.	
Category	Preparation	CSIRP
Design & Interpretation Considerations	Preparedness is the most effective tool to avoiding the risk of cybersecurity incidents and minimizing their negative impact on the business and information security [14]. Elements of preparedness include: 1- Approving a complete and comprehensive CSIRP 2- Deployment of detection and intelligence mechanisms 3- Availability of software and hardware for containment, analysis and recovery 4- Allocation of fiscal resources 5- Establishing partnerships with external agencies 6- Coordination between various internal bodies of the organization 7- Ensuring the competency and readiness of the CSIRT members 8- Adopting procedures for continuous enhancement and professional development 9- Providing communication channels and situational awareness mechanisms Signs of unpreparedness include: 1- Instability of the response 2- Incapacity to respond 3- Facing frequent “surprises” during incident handling	

PM Examples	The DHS publication of [103] is dedicated to measuring the preparedness of CSIR
Relevant PIs	Capacity, Competency, Completeness, Comprehensiveness, Readiness, Flexibility, Reliability, Forensic Readiness, Partnership Effectiveness

PI.34.	Response Public Impact	
Definition	The ability to conduct a response that receives positive response from the public.	
Category	Generic	CSIRT
Design & Interpretation Considerations	Not all incidents have “public” considerations. Some incidents are completely private to the organization with no impact to external entities. The term ‘public’ should be clearly defined. It could be simply defined to include all individuals/entities that are not part of the organization. Alternatively, it could be defined to include all those not participating in the decision making processes related to the incident handling (in this case, employees can be considered part of the public – this intersects with the Constituency Satisfaction PI). Measuring the public impact is challenging, but there several qualitative and quantitative measures that could collectively indicate how the public perceived the response: 1- Media coverage: positive and negative. 2- Post-incident surveys analyzed against pre-incident surveys 3- Letters/emails to the organization 4- Value of the shares in the stock market 5- Post-incident sales compared to sales during similar seasons of previous years 6- Customer retention. A response that is considerate of the public would do some or all of the following: 1- Communicates with the public and responds to concerns 2- Demonstrates transparency in communication 3- Shows that the organization cares about the public concerns (e.g. leak of private data) as much as it cares for the organization’s reputation and financial status. 4- Maintains the established trust with various partners and clients	

	5- Demonstrates compliance to best incident response procedures. Sometimes a team could have executed an effective response, but the public perceive it otherwise. This indicator should help in analyzing factors that contribute to such unfortunate incident side-effect. Since all incidents come with undesirable consequences, it is expected that all incidents will be associated with some negative reception from the public. Therefore, it is better to focus on measuring the extent of this negative consequence compared to “how well” did the public perceive the response.
PM Examples	Customer satisfaction post-incident survey results compared to results obtained from pre-incident surveys.
Relevant PIs	Constituency Satisfaction, Harm Reduction, Communication Effectiveness, Transparency, Response Cost

PI.35.	Readiness	
Definition	The ability to apply a set of competencies to execute CSIR tasks (Definition adapted from [219] and [61])	
Category	Preparation	CSIRP/CSIRT
Design & Interpretation Considerations	Readiness and Preparedness are used interchangeably in most publications. The distinction is made in compliance with DHS terminology [219] [110] and other works like [29] Competency ensures understanding the tasks while readiness ensures the ability to carry out the tasks. Measuring readiness can be broken to measuring two main sub-tasks [61]: 1- Evaluate whether a specific task can be performed as required 2- Evaluate whether competencies can be appropriately applied to tasks	
PM Examples	See readiness assessment methods outlined in [61].	
Relevant PIs	Preparedness	

PI.36.	Recovery Effectiveness
Definition	The ability of a CSIRT to executes a response that fully recovers the affected systems in a quick and cost effective fashion.
Design & Interpretation Considerations	Recovery means restoring the system/environment to its state before the occurrence of the incident but with the addition of preventive measures that disallow the incident from re-occurring. Recovery of affected systems can take long time. Therefore, it is better to divide the recovery process into phases. The phase that restores the basic operations should be "quick" while full recovery should be thorough and intrusive to ensure that all incident impacts are removed. The technical aspect of recovery includes the following: 1- removing the malicious components of the incident and conducting a thorough scan for remnants of the system (this intersects with eradication) 2- Restoring the system from backups and conducting a clean re-building of the environment. 3- Applying measures to prevent the incident from re-occurring (e.g. installing new equipment, vulnerability patching, enforcing new policies ..etc.) 4- Conducting the above steps in quick and cost effective manner. Measuring recovery effectiveness can be divided into several steps each focusing on a separate activity of the above. A recovery process may include monitoring the system for a period of time for re-infection [140] [79]. Measuring the effectiveness of recovery effectiveness can be conducted as a separate activity or as part of evaluating the effectiveness of implementing the business recovery plan [74]. The restoration process can be measured as per the availability of a resource or per the availability of a service.
PM Examples	The metric *recovery_time*. The start of the time period could be: • The time of incident declaration or discovery • The start of the eradication/recovery phase The end of the time period could be: • The time the system is declared fully restored

	• The time the system is fully restored and passed the re-infection monitoring process
Relevant PIs	Eradication effectiveness, Mitigation effectiveness

PI.37.	Reliability	
Definition	The capacity of a CSIRT to operate in failure free mode	
Category	Preparation	CSIRP/CSIRT
Design & Interpretation Considerations	Failures in the domain of CSIR can appear due to unavailability or incompetency The study of reliability includes the CSIRT members, tools, machines, support and other services. Should focus on critical aspects of the system whose absence/failure leads to failure of the whole CSIR system As reliability is normally expressed in probabilities, some aspects of the system can tolerate 97% reliability but others need to be fully reliable	
PM Examples	If there are [x_1 , x_2 , … , x_n] critical tools need for a successful IR, with probability of failure: [p_1 , p_2 , … , p_n] Then the reliability of the tools is: 1- *max* [p_1 , p_2 , … , p_n] For a comprehensive study on how to measure reliability of incident response see [103] and for fault-tree analysis for reliability see [183].	
Relevant PIs	Survivability	

PI.38.	Response Time	
Definition	The ability of a CSIRT to carry out CSIR activities in a quick manner	
Category	Generic	CSIRT
Design & Interpretation Considerations	Almost all incident response time use response time performance metrics. Having a shorter response time is a generic sign of good performance. It may also demonstrate a sign of security maturity [220]	

	Response time can be used as a method for ensuring quality requirements are met. For example, it could be validated post-incident if response time to offering a specific CSIR service exceeded the promised value [2]. Since response time is critical, a CSIR needs to identify activities that could be performed offline, i.e. not during incident handling. Response time can also be geared towards some security services or hardware/software metrics [83] Response time may vary from incident to incident for several factors like [1]: 1- Type of incident 2- Incident Severity 3- Existing service level agreements (SLA) 4- Criticality of resources involved. Because of the diversity of the above conditions/factors, response time needs to be analyzed taking into considerations the totality of the incident. It is better to read it along with other PIs to formulate a correct representation of the response performance.

PM Examples	The VERIS project a list of response time metrics, like the first compromise time, discovery time and containment/restoration time [73].
	The containment/restoration time (P4_RT) which measures the time span of the fourth phase in the CSIR cycle, can be broken into several time slots. The total time is a summative metric.
	$P4\_RT = CT + PRT + FRT$
	CT = containment time
	PRT = partial recovery time
	FRT = full recovery time

Relevant PIs	Response Effectiveness, Goal Achievement

PI.39.	Robustness	
Definition	The ability of the CSIR capability to deliver the same level of QoS throughout various incidents and under various circumstances	
Category	Generic/Preparedness	CSIRT
Design & Interpretation Considerations	There are various definitions of robustness. The above definition considers the aspect not included in the definitions of other PIs. Examples of other usages of the term:	

	1- Robustness is defined in terms of reliability and correctness [24]. 2- Robustness of response is a synonym to response effectiveness [221] 3- Robustness is defined in terms of flexibility, i.e. ability to adapt to unplanned activities [211]. 4- Robustness is defined in terms of stability and ability to react to erroneous inputs to the system [155].
PM Examples	Comparative metrics that measure the delivered incident response quality of service across incidents. This could be used through metrics of other PIs like stability and flexibility.
Relevant PIs	Stability, Flexibility, Adaptability

PI.40.	Root-Cause Identification	
Definition	The ability to successfully identify the factors that cause the incident to occur .	
Category	Identification / Analysis	CSIRT
Design & Interpretation Considerations	Although it is not always necessary for the incident's root-cause to be known to executive effective handling, to be frequently unable to determine the causes is a sign of poor analysis capability [92]. An incident may have a single or multiple root-causes, each can be direct or indirect Knowing the root-cause can help in: 1- Guide the organization where to invest and allocate resources to prevent and mitigate future incidents [4]. 2- Decide if the incident involves criminal activity and hence reporting to law enforcement and legal teams 3- Develop solutions to eliminate the root-cause [19] The root-cause can be: 1- Poor infrastructure or lack of resources 2- Poor defense mechanisms 3- Improper on-going process or policy [79] 4- Inadequate design of service objectives or levels [82] 5- Human error or unaviodable factors Trend analysis is an effective mechanism for identifying root-causes [74]	

PM Examples	The following while paper [222] proposes four metric for measuring the effectiveness of root-cause analysis: 1- Does the event exceed the triggers for failures 2- Completeness of implementation of corrective actions 3- Meeting the success measure for implementing corrective actions 4- Net return on investment for using root-cause analysis
Relevant PIs	Attacking Host Identification

PI.41.	[Constituency] Satisfaction	
Definition	The satisfaction degree of CSIR constituencies on responses conducted by a CSIRT	
Category	Generic	CSIRT
Design & Interpretation Considerations	Constituency satisfaction can be used in one of two ways: 1- Estimating the impact of an incident on its constituencies [200]. 2- Measuring constituency consent on executed response Emphasizing "customer" satisfaction over "conformance" is a characteristic of Total Quality Management (TQM) [39] Although satisfaction is an indicator for overall quality level of performance [223], the results can be misleading. For example, results will be impacted by the level provided to the surveyed individuals, transparency of the CSIR process and the design of the survey form. Therefore, I recommend not relying only on survey tools unless it is supported with validation schemes. There needs to be a careful process for identifying constituencies and grouping them. The use of this indicator should be coupled with an analysis process to identify factors that lead to dissatisfaction of the constituencies. Satisfaction could be measured for individual incidents and over multiple incidents	
PM Examples	The most common method for measuring satisfaction is through surveys and questionnaires	
Relevant PIs	Goal Achievement, Harm Reduction	

PI.42.	Scalability	
Definition	The ability of a CSIRT to deliver service, maintain service quality, or increase capacity in response to incident escalation	
Category	Identification	CSIRP
Design & Interpretation Considerations	Scalability could be measured in terms of system outputs, inputs or both. Scalability is correlated to how incident escalation is defined, e.g. incident size, severity, cost …etc. Scalability can appear as [224]: 1- Scaling up: enhancing current resources 2- Scaling out: adding more resources There are three types of scalability: 1- Ability to respond despite the quality dimensioning proportional to incident escalation. 2- Ability to maintain delivered quality despite incident escalation 3- Ability to effectively increase the capacity of the response system	
PM Examples	Descriptive metrics need to be used to define scalability zones or levels [225]. For type 1: Use two thresholds: maximum capacity and minimum acceptable quality level. Then use a binary metric to evaluate if the CSIRT is prepared to respond within that boundary. This should be validated post-incident. For type 3: Introduce metrics that measure how the change from one capacity level to another is done effectively. Examples: upgrade time and cost effectiveness For using stress testing for measuring the scalability of software and security tools see [226]	
Relevant PIs	Capacity	

PI.43.	Shielding (Protecting Critical Assets) Effectiveness	
Definition	The ability of a CSIRT to protect uninfected critical resources and assets from the impacts of an incident	
Category	Containment	CSIRP

Design & Interpretation Considerations	Shielding effectiveness as defined above is related to containment effectiveness PI, but is different in two aspects: 1- It is focused on critical assets, while containment is generic to all assets 2- It is mainly proactive measures, while containment is mainly reactive A critical resource/asset could be infected or uninfected. If infected: 1- Isolate the resource in order not to infect other resources 2- Mitigate and minimize damage 3- Execute quick and effective eradication and recovery If uninfected: 1- Shield the resource to prevent infection 2- Ensure continuity of functionality with minimum interruption
PM Examples	A matrix with the list of critical resources over the CIA requirements (confidentiality, Integrity and Availability) could be constructed. The grid is filled with timestamps about when each asset is declared “protected”. PM metrics are constructed to provide analysis over the grid. For example, time to declare all assets available, number of records protected. The above metrics could be coupled with confidence ratings, e.g. fully protected, partially protected, unprotected. Also, a scale for the level of “criticality” of the asset could be assigned.
Relevant PIs	Containment Effectiveness, mitigation effectiveness, confinement effectiveness

PI.44.	Stability	
Definition	The ability of a CSIRT to prevent an incident from unexpected escalation and to maintain the response in a controllable state	
Category	Identification	CSIRT
Design & Interpretation Considerations	This PI detects scenarios when the response reaches non-stable states. Analysis of the results can help the team develop solutions that enable future stable response.	

	Instability, as defined above, can, but not limited to, appear in two forms: 1- Unexpected escalation 2- Reaching non-controllable state There needs to be prior definition of what constitute “unexpected escalation”. For instance, sf an incident is forecasted to escalate and the CSIRT develops a response that takes that into consideration, then this does not reflect instability. On the other hand, the unavailability of a staff or a resource (which could have been avoided by better preparedness) can cause a sudden escalation which complicates the response procedures. There needs to be metrics or precursors to identify when a response is considered “uncontrollable” like high variety in the incident severity, contradicting decisions, unreasonable expenditure …etc. The main activities that impact stability are the actions taken right after incident declaration, i.e. the end of the detection phase and during the containment phase. Therefore, the stability PI is strongly related to the Containment Effectiveness PI, but the two are not identical. An incident could be stable but the containment procedures are ineffective, and vice versa.
PM Examples	A metric that measures the variance of the incident classification during the response life cycle. See Section 5.1.1.
Relevant PIs	Containment Effectiveness, Detection Effectiveness, Accuracy

PI.45.	Sufficiency [or Self-Reliance]	
Definition	The ability of a CSIR to execute a response without seeking external assistance.	
Category	Generic / Preparation	CSIRT
Design & Interpretation Considerations	A CSIR that demonstrates sufficiency is an indicator for competency, reliability and good capacity Assistance and support to the CSIR can come from internal support teams or from external entities. The internal teams are considered secondary assets to the CSIR; therefore using them is not a sign of poor performance. However, seeking external assistance “might” indicate poor preparedness. Not all external assistance indicate insufficiency. For example, exchange of information between entities is genetic to the work of CSIR. Assistance from law enforcement might also be necessary.	

	Therefore, insufficiency should be defined in terms of technical and logistical support that could have been prevented through proper prior planning Insufficiency can result from: 1- Insufficient competency and staff training 2- Insufficient tools and instruments 3- Insufficient planning budget 4- Insufficient support services
PM Examples	A simple binary qualitative metric can be used by asking: “did the team seek external assistance”. This can be broken into several sub-questions each corresponding to a component of the response cycle. This helps identify areas in the CSIR that might require enhancement. Insufficiency can be quantized by measuring the impact of external assistance on the success of the incident handling. One possible method to do this is through finding the ratio between the cost of external assistance to the overall incident response cost.
Relevant PIs	Capacity, Reliability, Competency

PI.46.	Survivability	
Definition	The ability of a CSIR capability to preserve essential services in the system despite the presence of compromise (adapted from [227])	
Category	Generic	CSIRT
Design & Interpretation Considerations	The above definition considers survivability as a special case of the availability PI. Availability is more general and focuses on *offering* service, while survivability focuses on completing a mission or ensuring that a specific system cycle is not interrupted. Also, availability can be defined based on quality levels, while survivability focus on availability of a resource while offering the minimum acceptable QoS. Aspects of the system that need to *survive* need to be identified, like network connection, database access, and some critical assets. Another dimension of survivability is the ability of the CSIRT to complete its own mission during incidents [162].	
PM Examples	Tools for measuring network survivability during compromising incidents are outlined in [227]	
Relevant PIs	Reliability	

PI.47.	Team Cohesion	
Definition	Ability of a CSIRT to execute a response in which every team member effectively collaborate with the rest of the team to achieve the outlined goals (adapted from [228])	
Category	Generic	CSIRT
Design & Interpretation Considerations	Depending on the nature of the CSIRT, cohesion can be defined between individuals, i.e. team members [97], or through organizations [109]. There are three PIs that focus on evaluating the effectiveness of how various parties collaborate to achieve the CSIR goals. The *Team Cohesion* PI focuses on members of the CSIRT, the *Coordination Effectiveness* PI focuses on the various groups within the organization and the Partnership Effectiveness PI focuses on collaborations with external entities. Factors in which team cohesion could be defined through [228] [97]: • Members attraction to the team/group (feeling proud to be part of the team) • Trust between team members • Communication between parties • Interaction inside and outside assigned tasks. Team cohesion evolves over time, so measurement should be done over a relatively long period of time.	
PM Examples	A qualitative approach of conducting surveys or interviews with team members to measure cohesion in multidimensional fashion: (see [228])	
Relevant PIs	Communication Effectiveness, Coordination Effectiveness, Partnership Effectiveness	

PI.48.	Traceability (Documentation Effectiveness)	
Definition	The ability of a CSIR to document its actions in a manner that permits accessible reachability to their sources	
Category	Analysis	CSIRP

Design & Interpretation Considerations	This PI measures one aspect of the analysis phase effectiveness, namely the documentation and reporting. Benefits of having good traceability include: 1- Provides means for validation and verification of data and analysis results. 2- Saves the team's time during incident handling by reaching to relevant data in a timely manner 3- Gives opportunity to easily use/apply holistic and trend analysis techniques 4- Facilitates demonstrating compliance Aspects of CSIR traceability include: 1- Traceability from measurements back to the goals [78] 2- Traceability from actions back to policies (ISO 27001 [198]) 3- Traceability from analysis results back to data 4- Data provenance, i.e. source of information [24] 5- Trace performance problems and failures to their sources [135] 6- Traceability of digital evidence through forensics processes [204] 7- Traceability of actions/decisions to individuals
PM Examples	A metric can be developed to count the number of broken links from the source to the source. In situations when data is linked directly to its source, a simple count over number of searches can demonstrate overall traceability. In situations that involve multiple links from source to destination, the distance needs to be measured from the broken link to the source and to the origin.
Relevant PIs	Documentation Effectiveness

PI.49.	Transparency	
Definition	The feature of conducting CSIR activities, reporting results and disseminating incident information in honesty and openings.	
Category	Analysis	CSIRP
Design & Interpretation Considerations	Some benefits of having transparent incident response: 1- Contributes to CSIR sustainability [178] 2- Minimizes the probability of surprises and enhances risk assessment [183]	

	3- Establishes trust between various parties, especially external bodies [51] 4- Provides means for validation and verification 5- Better customer/constituency satisfaction Elements of transparency: 1- Transparency in definitions of authority and autonomy [51] 2- Transparency in sharing of information [81] 3- Transparency in data collection and analysis methodologies 4- Transparency in assessment of risks and caused damage In the above definition, “honesty” means reflecting what actually happened. This leads to transparency being contentious with the need for confidentiality. Balance can be reached by outlining in a CSIRP what information is to be exchanged and how. The transparency and consistency PIs are related to each other, as enhancing one normally leads to enhancing the other. Assessing transparency would require assessing the level of disclosure, accuracy and clarity
PM Examples	The following study [229] suggests using the following four metrics for measuring transparency in the context of organizational management (which can be customized to CSIR): 1- Involvement 2- Feedback 3- Level of details 4- Ease of finding information
Relevant PIs	Consistency, Communication Effectiveness, Accuracy, Traceability, Documentation Effectiveness

PI.50.	Utilization	
Definition	A generic indicator concerning the degree at which a specific CSIR resource or service is used against its capacity.	
Category	Generic	CSIRT
Design & Interpretation Considerations	High utilization of resources can lead to better response time [149], indicate better resource allocation and management, and make more cost-effective responses. Reaching full utilization is an indicator of reaching system capacity which can possibly cause performance bottlenecks.	

	Utilization is an indicator for measuring the effectiveness of the response process not the outcomes [230]. Some resources are “used-as-needed”. Under-utilization of these resources is not an indicator of poor performance, especially if the unavailability of this resource is critical to CSIR incident handling Utilization is also applicable to how responder time and experience is being applied [15]. Good utilization would be reflected in assigning tasks to responders with the proper experience.
PM Examples	Backup_utilization (BU) = ABT / BC - *backup_capacity (BC):* total time (in minutes) needed to backup the full system using full utilization of resources - *actual_backup_time (ABT)*: actual time (in minutes) consumed for backuping the full system during incident handling The BU metric is applicable whenever the recovery process requires backing-up the full system before applying new updates, e.g. vulnerability patches. PM2: Average team member utilization (ATMU): $ATMU = \frac{\sum_{i=1}^{N} h_i}{\sum_{j=1}^{N} H_j}$ *N* = total number of CSIRT members H_x = Available hours of member *x* h_x = actual hours of member *x* used in incident handling
Relevant PIs	Cost Effectiveness

CHAPTER FIVE

FRAMEWORK SUBCOMPONENTS

This chapter builds on the proposed PE framework presented in Chapter 3 by providing a variety of models and methodologies with more emphasis on implementation considerations. Several concepts which were presented in an abstract form in the previous chapter are investigated in this chapter for their potential operational characteristics. The detailed practical aspects presented herein complement the higher level design concepts presented beforehand. However, the themes of generality and flexibility are maintained to allow deriving a diverse range of implementations that suit various environments.

The chapter presents five modules. The first module is a comprehensive theoretical framework for building performance evaluation systems for CSIR. The second module is an integrated model for analyzing and validating performance evaluation results. The third module presents a multi-perspective stakeholder analysis of CSIR. The fourth and fifth models address the challenges of CSIR complexity and unpredictability.

5.1 The Universal PE Framework (CSIR-UPEF)

The *Universal Performance Evaluation Framework* (UPEF) is proposed as a CSIRPE framework that encompasses all of the desirable features of various frameworks. It is an idealistic super-framework that outperforms all other CSIRPE frameworks. It is also the most generic in the sense that all CSIRPE frameworks can be derived from it.

The main feature of UPEF is that it is capable of effectively and accurately assessing the performance of any CSIR. The UPEF is unboundedly scalable, has unrestricted access to resources and can be deployed to any CSIRT, environment or incident. It also conducts comprehensive assessment, produces accurate measurements, and uses a comprehensive list of performance analysis technique, performance indicators and performance metrics. A list of all of the desired features of the UPEF is provided in Table 34.

#	Feature	Description
D.1	CSIRT Type	Can be used to evaluate any type of CSIRT, e.g. centralized, distributed, coordinated or customized
D.2	Evaluator Type	The framework is equally equipped to be used by any type of evaluator (e.g. self-evaluation by CSIRT, internal auditors, or external auditors)
D.3	Number of Incidents	Can be used to analyze performance to a single instance of incident handling or to a collection of incident handling instances
D.4	Incident Concurrency	Can be used to evaluate performance of response to incident occurring simultaneously or in spaced intervals.
D.6	Benchmarking	It can be interfaced with any internal or external benchmark
D.8	CSIRP Scope	Measures both the effectiveness of the CSIRP design and the performance of the CSIRT
A1.1	CSIR Environment	Can be used unrestrictedly in any environment
A1.2	Team Structure	Can be used to evaluate a CSIRT composed of any number of members
A1.3	Incident Complexity	Can evaluate response to incidents of any degree of complexity
F.4	Functional Model	Contains performance monitoring capabilities, is able to provide partial results at any instance of the incident cycle, and produces comprehensive analysis at the end of the incident
S.5	CSIRPE Overhead	It involves no overhead in the collection of PMs and analysis results are produced instantly with no post-incident overhead. Also, there is no overhead

		imposed on the team to design or maintain the CSIRPE.
PI.x	Performance Indicators	Has the capability to use and produce results for all PIs defined in Section 4.3.
PI.1	Accuracy	Provides measurements in optimal accuracy, i.e. full precision

Table 34: Features of the Universal PE Framework (UPEF)

The UPEF is theoretical, i.e. unrealistic to implement. However, it is an effective design and analytical tool for understanding and evaluating the performance of CSIRTs. The author found the UPEF to be useful in developing several components of the CSIRPE framework proposed in this dissertation, due to the complexity nature of the problem. Below is a brief discussion of how UPEF could be used in the design and analysis of CSIRPE frameworks.

The idealistic approach of the UPEF can be mapped to one of the methodologies used for understanding and developing security metrics through idealistic analysis [162]. For instance, in [190] seven idealistic dimensions of security were used to derive security metrics. These dimensions include perfect knowledge of the system, and the attacker knowing nothing about the control system, the system has no vulnerabilities and the security team is capable of instantly detecting an attack and restoring the system. These dimensions are then used to derive operational metrics, or real world measures of these ideals. The UPEF adopts a similar approach.

There are several advantages for the conceptualization of UPEF.

Framework Generator:

Since UPEF provides the most theoretically generic CSIRPE framework, all other frameworks could be generated from it. While developing the CSIRPE framework of this project, I found that there were two theoretical approaches to conceptualizing and defining CSIRPE frameworks.

The first approach is to build a very simplistic model and gradually add parameters to expand the model. Any simplistic model is expected to be unrealistic as it will fail to capture the complex parameters of the problem. The careful expansion of the model through realistic considerations should eventually lead to partial capturing of some practical performance aspects of the system. This approach views performance bottom-up, i.e. starting from measuring system components to the overall performance.

The main advantage of the simplistic approach is its practicality. Most organizations need to start their performance systems small and then grow. It also seems to be the most commonly used approach in the works focusing on measuring security. Indeed, the SAC Complexity model proposed in Section 0 suggests using this approach due to its practicality.

However, the reductionist nature of the approach raises concerns about the validity of measurements and most importantly its practicality [93] [34], which is a counter result to the initial goal. In addition, from the design perspective, in order to build a simplistic model the designers need to generate a long list of assumptions. Since the CSIR environment is very complex, it is very easy to overlook some parameters which might be necessary for producing meaningful performance analysis.

The second approach starts from unrestricted and unbounded framework and gradually add delimiters to bring it from the theoretical to the practical domain. It is a top-down approach that focuses on the overall system performance down to the system components' performance. It is appealing to higher managers who are interested in strategic planning and generic performance indicators. It is also appealing to researchers who endeavor to study the various aspects of the complex CSIR system.

The UPEF adopts the second approach which suits the concept of developing general frameworks compared to specific models. Each of the UPEF features defined in Table 34, when modified, can be used to generate a different type of CSIRPE framework.

CSIRPE Design Tool:

The UPEF could assist a CSIRT interested in building a performance system in three different ways. First, it gives the designers a perspective on the various possibilities to build a CSIRPE. Second, it outlines the desired features of CSIRPE, despite being theoretical. Consequently, it helps in setting the PE goals and setting the design parameters introduced in this project.

Several adopted emergency incident management frameworks set goals that seem unrealistic, demonstrating similarities to the features of a UPEF. For example, the National Response Framework (NRF) [110] aspires that frameworks be designed to handle any emergency situation, regardless of the scale, scope or number of simultaneous incidents; which is realistically very difficult to establish.

Another design benefit that comes with UPEF is its basic structure that can assist software developers interested in developing a CSIRPE generic software solution. These solutions are generic and comes with customization properties to meet specific client needs. It will also provide insights to those developing simulation and automation modules for CSIR.

Analytical Tool:

The UPEF has several benefits to bring forward to CSIR performance analysis. It could be used as a baseline for comparing the benefits and shortcomings of various PE frameworks. It also could be used as a tool to measure the gap between the optimal and practical characteristics of PE frameworks. With the absence of benchmarks, UPEF can also be used as a tool to set optimal values for performance metrics and compared measured results against it. The aforementioned gap analysis would assist in identifying operational factors that deviate performance evaluation from being perfect. These factors could be then subjected to further PE analysis.

5.2 The Integrated Analysis & Validation Model (IAV)

The Integrated Analysis & Validation Model (IAV) is a generic model for using analysis and validation schemes in the process of CSIR performance measurement. The model is generic in two perspectives. The first is its use of a variety of techniques that makes it applicable to various CSIR environments. It is not necessary that a team uses all techniques, as the model can be customized based on the CSIR goals and needs. The second perspective is that the proposed techniques are derived from distinct paradigms that could generate performance results from different perspectives.

The term ‘integrated” suggests that the model is interfaced with the CSIR modules across different phases. It also stresses the need to view performance analysis as an integrated process in the CSIR life cycle instead of treating it as a separate activity applied post-incident.

A higher level design of the IAV model is presented in Figure 20. Starting from the top of the chart, the CSIRPE design and planning activities would produce a CSIR performance evaluation framework. When incidents occur, the collected data is input into the analysis schemes, which produce performance results. These results are used as input to the validation schemes. The right-side of the chart displays a performance evaluation database (PED), which acts the main repository for storing incident and analysis data. The PED is queried by the design and planning activities for enhancement recommendations, and also by the analysis schemes for performing comparative or trend analysis.

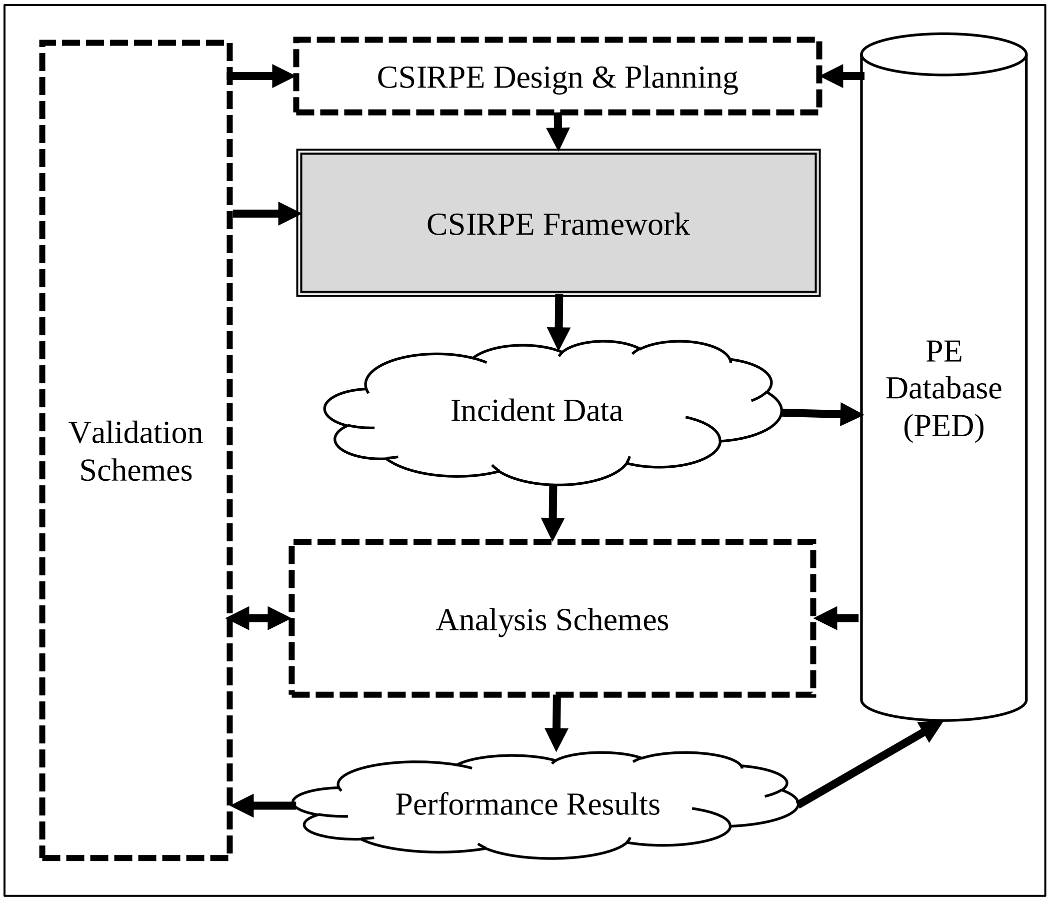


Figure 20: Higher-level Design of the IVA Model

The left-side of the chart displays the validation schemes box, which is fed by the performance results. The validation schemes target, the CSIRPE framework through validating its design and its performance. Note that there is a double edge arrow between the analysis schemes and the validation schemes boxes to highlight the fact that some techniques are used for analysis and validation purposes at the same time.

The detailed component structure of the IAV model is presented in Figure 21. The analysis scheme box is broken into three major interconnected components. The first IAV analysis component is called the *Component Analysis* which targets focused performance analysis into the separate parts of the CSIR system. These techniques normally use comparative methods, like benchmarking or performing gap analysis to compare current performance to expected, ideal, or best industry results. For instance the performance pertaining to the detection/identification phase can be investigated through benchmarking the PE results to that of other peer-institutions (benchmarking) or to an expected/ideal values (gap analysis). The same applies to the other phases of the CSIR life cycle. Overall, this component inspects performance of a specific part of the CSIR system independent from the other parts.

The second analysis scheme component is the *deficiency analysis* component which focuses on identifying sources and causes that negatively impact performance. Example techniques include bottleneck analysis and root-cause analysis. Notice that the *targeted analysis* scheme intersects with the deficiency and component analysis components. Notice also that both components interacts with the PED in bi-directional manner, and feed the feedback system validation scheme.

The third component is holistic analysis which inspects the performance of the CSIR system as a whole. Example of holistic techniques include assessment of goal achievement and performing analysis from different stakeholder perspectives. In addition, the trend analysis technique inspects the overall system in terms of positive or deficient performance over historical windows. The predictive analysis scheme can be used as a method for holistic analysis, validation or both.

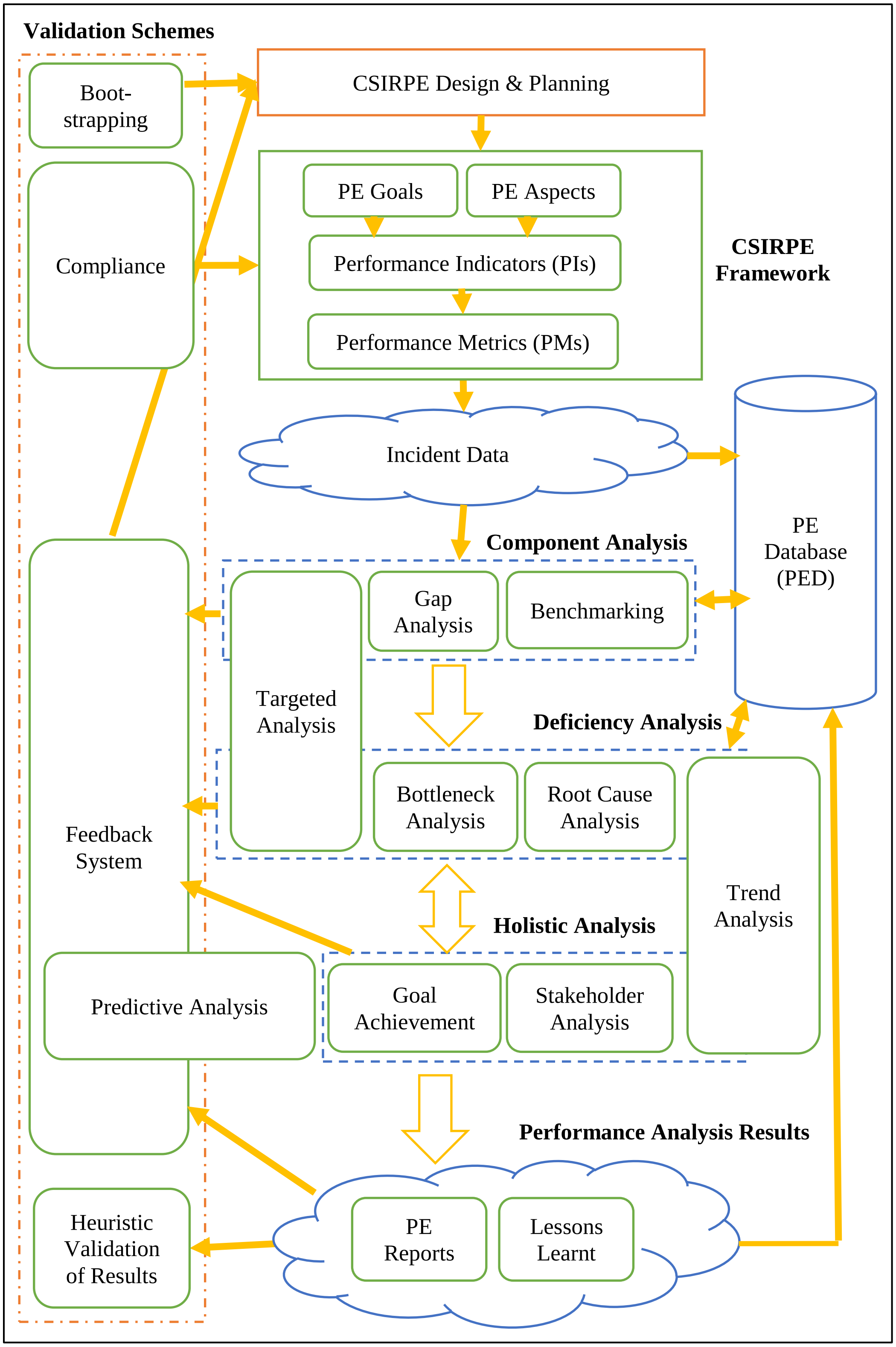


Figure 21: Detailed Design of the Integrated Analysis & Validation Model (IVA)

There are five components to the validation scheme box. Starting from the top, the *boot-strapping* method is used to validate the design and planning activities by continuous interaction with higher management or constituencies. The *compliance* module validates the structure of the CSIRPE in terms of its policies and procedures. The *feedback system* synthesizes incident and performance data to produce practical recommendations for enhancement. Finally, the heuristic method inspects the level of confidence of the accuracy of performance results.

The IAV can be customized in three different ways:

1- *Basic Design:* a CSIRT might choose to have a basic or advance application of analysis and validation schemes. The detailed design presented in Figure 21 presents an advanced design that does not suit most CSIRs. In its most basic form, a CSIRT need only to use a single analysis technique and the feedback system as a validation scheme (see Figure 33 in Appendix A). The basic design can be then gradually enhanced by introducing other

2- *Expansion:* The IAV model is not limited to the modules presented under each component. A team may opt to add its own customized analysis or validation schemes while maintaining the overall structure of the model. For example, a team may choose to use a different comparative technique. In that regard, a new box needs to be added within the 'Component Analysis' box, while maintaining the other parts of the model.

3- *Interaction:* The interaction between various components of the IAV can be modified without affecting the overall functionalities of the model. In its basic design, only the presence of the PED and the feedback system is necessary for conducting analysis and validation schemes. If a new scheme is introduced, it only needs to connect directly or indirectly to the PED or the feedback system.

As demonstrated above, the IAV model can be viewed as an elastic skeleton of how analysis and validation could be used in CSIR performance measurement.

5.3 The CSIR Balanced Scorecard Model

The objective of the CSIR Balanced Scorecard model is to adapt the concept of stakeholder analysis, see Section 3.5.2, to the discipline of CSIR. The name of the model borrows the term *balanced scorecard* from a widely used multi-perspective performance analysis technique used in the industry [134]. The model shifts attention from the uni-perspective of analyzing performance by financial indicators to a wider inclusion of stakeholder perspectives. The aforementioned model uses the perspectives of: financial, customer, internal, and innovation and learning. Following the same spirit, the CSIR Balanced scorecard model presented in this section, attempts to build a method of analyzing CSIR performance from multiple perspectives.

The first question that arises is: who are the stakeholders of CSIR? The answer to that can be found in several works have identified the stakeholders for CSIR [29] [68] [92]. From these works and through analyzing the stakeholders interested in assessing the performance of CSIR, five groups can be identified, each representing a different perspective:

1- *Organizational Stakeholders:* include executives, managers and administrative staff. This group is interested in a response that restores and minimizes interruptions to business operations. The group is also interested in a response that conforms to the organizational policies and procedures.

2- *Financial stakeholders:* include investors, shareholders and financial executives. This group is interested in a response that is cost effective and guards against financial loss.

3- *Technical Stakeholders*: include the CIO, CSIRT, security team, network team, helpdesk and other IT support staff. The focus of this group is conducting a response that utilizes available software and hardware tools and protects the CIA (confidentiality, integrity and availability) of the system.

4- *Client Perspective (Information Stakeholders):* include customers and contractors whose main concern is the protection of their private information and the fulfilment of their contractual terms.

5- *Community Stakeholders:* focus on the public safety at large and the implications of the incident on various dimensions like: legal, law enforcement, social, media …etc. It is also interested in information sharing and raising community awareness.

A graphical representation of the above five perspectives is found in Figure 22.

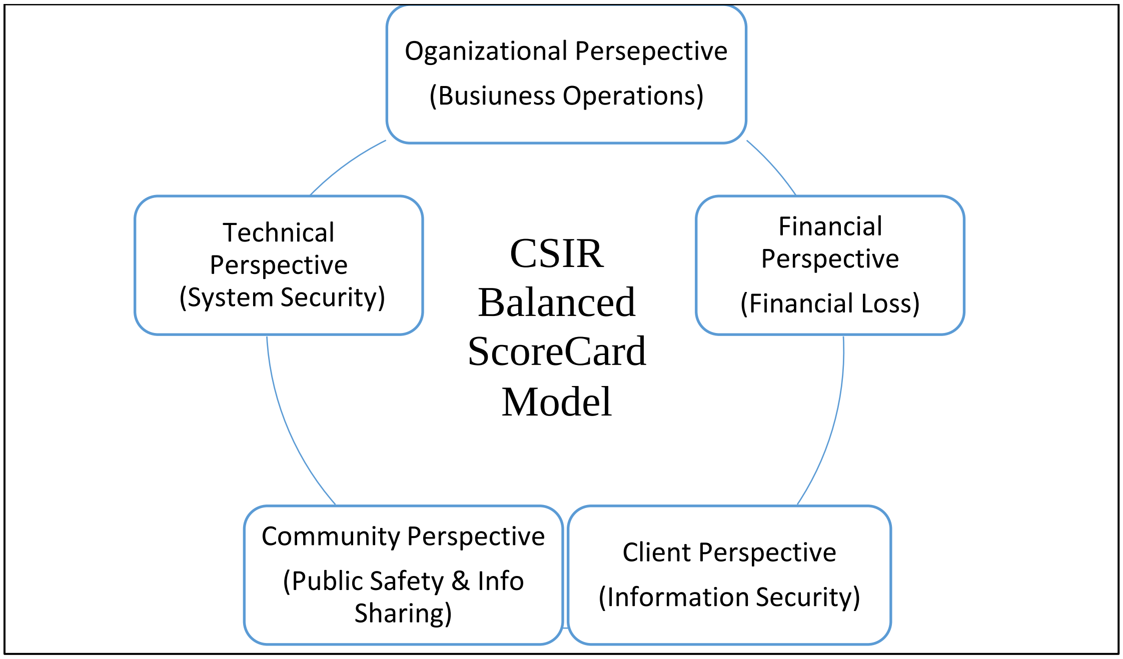


Figure 22: The CSIR Balanced ScoreCard Model

Using the list of performance indicators compiled in Section 4.3, example of PIs that can be used in analyzing each of the above five stakeholder perspectives is presented in Table 35.

Stakeholder	Sample PIs
Organizational Stakeholders	Competitiveness, Adaptability, Flexibility, Goal Achievement, Compliance, Conformance, Consistency
Financial Stakeholders	Response Cost, Capacity, Cost Effectiveness, Business Continuity, Utilization, Sufficiency
Technical Stakeholders	Detection Effectiveness, Containment Effectiveness, Accuracy, Eradication Effectiveness, Root-cause Identification, Stability
Information Stakeholders	Confidentiality, Availability, Constituency Satisfaction, Shielding Effectiveness, Transparency , Documentation Effectiveness
Community Stakeholders	Harm Reduction, Intelligence Capacity, Partnership Effectiveness, Attacking Host Identification, Evidence Retention

Table 35: Performance Indicators for the CSIR Balanced Scorecard Stakeholders

5.4 The SAC Complexity Model

The SAC model is mainly adapted from field of complex systems, but similar approaches are also used in supply chain management. The two main applications of the field of complex systems are the biological systems and engineering systems. A complex system is defined as “system made of a large number of microscopic components interacting with each other in nontrivial ways” [231]. Performance evaluation of CSIR can be viewed through the aforementioned definition of complex system where the microscopic components are the activities performed by the CSIRT or the data portions that need to be analyzed. In that regard, the field of complex systems poses as a potential rich resource for researchers interested in developing PE models for CSIR.

It is noted in [232] that performance evaluation techniques for complex systems can be classified into two categories. The first category uses measurements, benchmarking and prototyping and is suitable for evaluating existing systems; while the second category uses modeling and is suitable for evaluating systems before their actual implementation. Through this classification, CSIRPE would fall under the first category. However, due to the fact that finding and fixing errors in complex systems before they are deployed is much economical than fixing errors afterwards [233], the complex systems field paid more attention to the second category. Therefore, the consultation with the complex systems field needs to be selective to the extent of benefit to CSIRPE.

The suggested model is abbreviated by the “SAC Complexity Model” referring to the three strategies that are used in deriving the model. The three strategies are; Simplification, Approximation and Cascading.

The feature of simplicity is considered an essential element of many incident response systems [111]. The strategy of simplification suggests that in order to measure the PE of an IR system, the system should be analyzed and divided into singular or small activities that can be subject to analysis independent of the overall system. This strategy is intuitive, simple to implement and is backed by research findings. For example, [103] suggests simplification as a method to identify the failure modes of a complex system. The authors found that treating each failure mode independently provides a good approximation of the system performance, despite the simplification.

A sample of how simplification could be applied to the field of CSIR is shown in Figure 23. The diagram uses the hybrid model (See Table 2) and presents a possible simplification of the identification phase. The identification phase is broken into three main activities: incident detection, initial assessment and incident declaration. Incident detection includes the analysis of events and precursors and documenting the initial reporting; while incident declaration involves, assigning a severity scale and following proper procedures of reporting the incident. Focusing on the initial assessment activity, it could be further broken down into three activities (See Table 3), functional impact assessment, information impact assessment and recoverability impact assessment. If a specific CSIR environment uses advanced methods for impact assessment then the three assessments can be further broken into simpler activities.

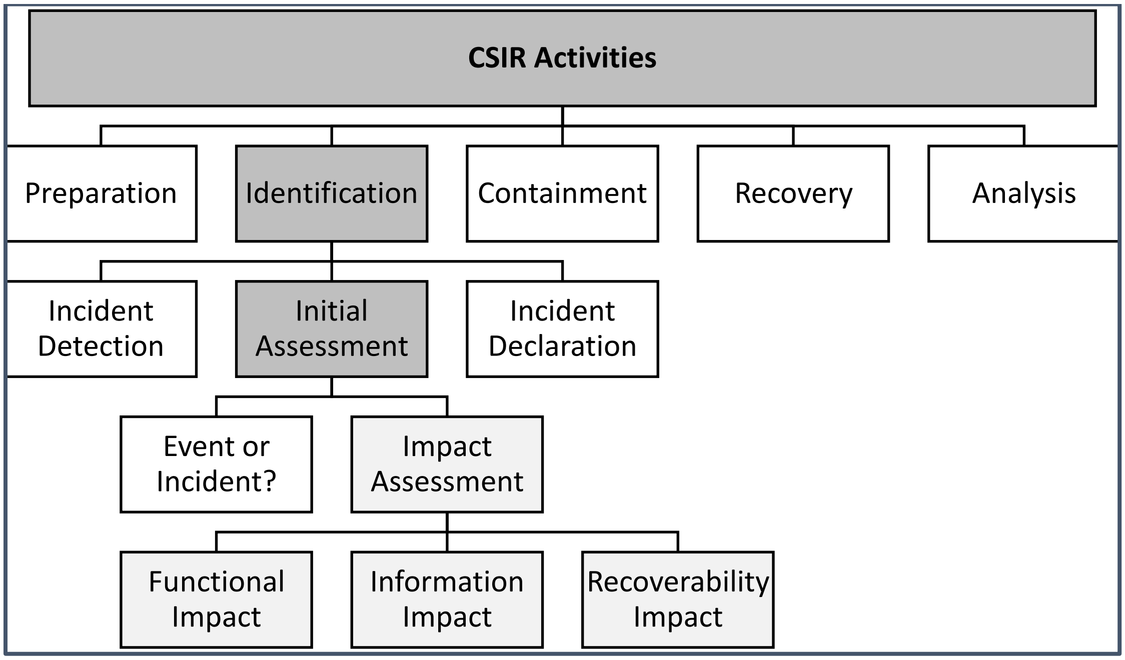


Figure 23: Sample of Simplification of the Identification Phase

Once the CSIR system is simplified into “simple” activities, each could be subject to independent effectiveness assessment. This starts by asking: “Does the CSIRT conduct functional impact assessment effectively?”, as an example. The advantage of this simplification is to enable identify performance issues at the micro-level.

The second principle, i.e. approximation, is stimulated by a lesson learnt from other disciplines that tackled the issue of IRPE that suggests that attempts to precisely measure the system will either fail or produce an analysis platform that is impractical. Therefore, the strategy of approximation here invites CSIRTs to view PE in non-precise terms. It also suggests that CSIRT should seek to analyze the CSIR system in a manner that identifies issues that impact performance more than attempting to develop performance metrics with scales that are sensitive to minor changes. This is backed by the analysis provided in [116] that performance measurements today are about the tradeoffs between quality and available resources in replacement of the original theory of measurements that emphasizes accuracy, precision and objectivity. Another study [137] highlights that approximation is normally needed for feasible use of benchmarking.

The cascading strategy suggests that after performing analysis on singular activities, several activities can be combined for another analysis. The process will continue until reaching the overall performance of the system. This cascading process been recommended as an effective method for the design of performance indicators and metrics [135]. In this referenced work, cascading is performed vertically and horizontally. Vertical cascading is when performance metrics are gradually cascaded up to satisfying a specific strategy. Horizontal cascading applies vertical cascading but across various strategies or domains of analysis.

This cascading strategy is also frequently used in the field of supply chain management (SCM). For example, a model presented in [107] [108], divides the analysis into four types, each corresponding to a phase within the supply chain. The first type is called functional measures and focuses on analyzing specific functions in the process. The second type is called "internal supply chain measures" and analyzes a combination of several functions in the production process. The third type is called "one sided integrated measures” and analyzes the overall system from one perspective like a specific customer or supplier. The fourth perspective is called “total chain measure” and analyzes the system through the inputs of the three previous steps. The four types of this ordered method of analysis are depicted in Figure 24.

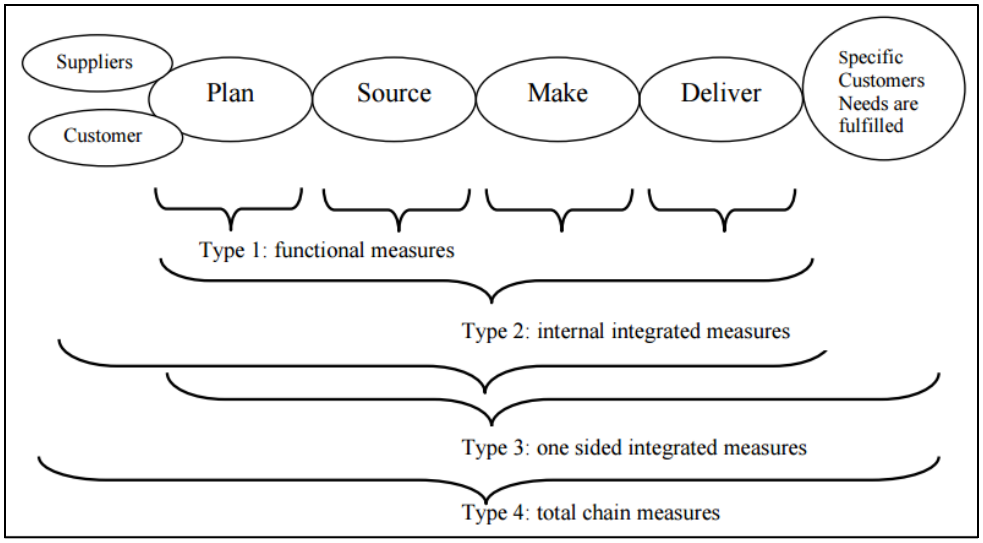

Suppliers
Customer
Plan
Source
Make
Deliver
Specific Customers Needs are fulfilled
Type 1: functional measures
Type 2: internal integrated measures
Type 3: one sided integrated measures
Type 4: total chain measures

Figure 24: Depiction of the four types of ordered analysis for SCM. Source: [107].

There are several challenges to the application of the above SAC model. First, the simplification process which ignores the strong correlation between the various system components might render the performance analysis unrealistic. To overcome this, the approximation strategy might need to be also applied by making estimations of the values of various factors that might impact the performance analysis. For instance, measuring the effectiveness of “impact assessment” is not a simple process of assessing how the outlined procedures are being executed. The analysis is influenced by who does the analysis, the reliability of detection mechanisms, the rate of false alarms, correctness of outlined CSIRP procedures, and documentation effectiveness of previous similar events. Therefore, some estimation of these factors could be recorded to reflect how much confidence is displayed by the PE analysis of that specific component.

A second challenge comes from the fact that compositing several components for analysis is non-trivial [162]. Therefore, although the SAC model suggests the use of approximation, this should not be done without guiding procedures of how approximation is to be applied. These procedures should demonstrate some basic level of objectivity and uniformity. In addition, the cascading should be done carefully such that coherent parts of the system are aggregated together.

5.5 The NFP Unpredictability Platform

The *NFP Unpredictability Platform* is a tool for understanding and analyzing unpredictability for the purposes of conducting performance evaluation in the context of CSIR. The platform is based on three incident response principles: non-deterministic decision making, flexibility and preparedness. In this section, the three principles of the NFP unpredictability platform will be translated into policies and operational guidelines for responders.

The first principle (N) recommends that each CSIR be equipped with a capability that allows for effective decision making in nondeterministic environments. Whether the CSIRT decides to follow a simple-informal process or a validated analytical model borrowed from decision making theories, the capability should have three main features:

1- Decision makers are provided with *timely* and *correct* information. This has been highlighted as one of the critical issues for responders operating in challenging environments [97].

2- Available data is analyzed and summarized to decision makers in the form of probabilities of possible events along with risks associated with potential decisions [129].

3- Available data is analyzed for forecasting purposes and decision makers have outlined procedures of how to enforce proactive measures [110].

The above three steps are graphically modeled in Figure 25.

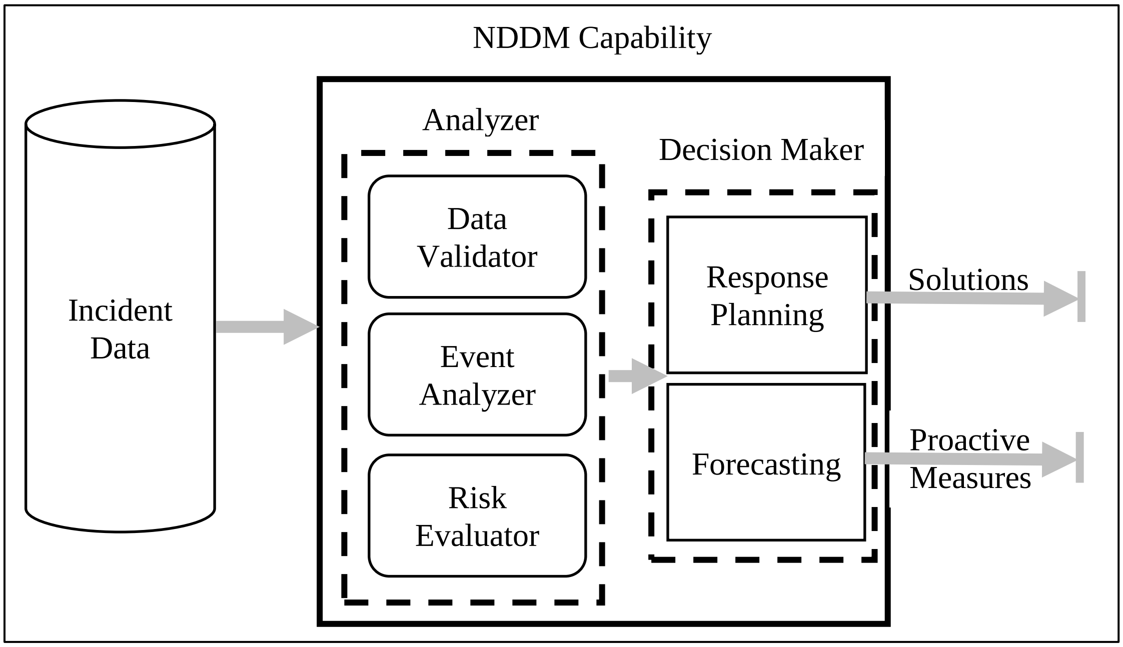


Figure 25: NDDM Component of the NFP Unpredictability Platform

The principle of flexibility (F) has been highlighted in NIMS [109], NFR [110], CERT [61] and in other publications [97]. This principle intercrosses with the principle of preparedness (P), as both aim at ensuring that the CSIR capability is able to respond to unpredictable situations. The term flexibility is used here in contrast to rigidity. A CSIRT can be trained to handle unexpected scenarios, but if the CSIR system, e.g. policies and procedures, is inflexible then there is little that can be achieved. The same is true if an organization has a flexible response system but the team is incompetent to address "out-of-script" situations.

The flexibility principle can be translated into three main policy/strategy elements: empowerment, scalability and adaptability. Empowerment means granting sufficient executive powers to responders to initiate and implement solutions with minimum chains of command. A balance should be made such that executive powers do not exceed pre-defined scopes, do not override specific policies, or contend with higher executive decisions. This concept is expressed in the national response framework (NRF) by entrusting local authorities to respond within their jurisdiction without the need to report to state or federal teams. Therefore, whenever assistance is needed, there needs to be a clear chain of command on how decisions are made among various parties [110].

The concept of scalability focuses on the CSIR capability able to respond to incidents of various magnitudes. Normally non-Scalable response systems are also inflexible. To be flexible, an organization needs first to define its response capacity. Then, there needs to be mechanisms to recognize when the capacity is over-reached [111]. In such scenarios, the CSIRT needs to immediately seek external assistance. It has been highlighted by [66] that failing to do so is considered one of the ‘seven deadly sins’ that a CSIRT can make.

Besides scalability, the system should be adaptable. Adaptability means the capability of responding to the change of needs [109]. Adaptability intersects with scalability when responding to the escalation of incidents. However, adaptability stands out when reliability issues arise or when there is a need for corrective measures. To overcome reliability issues, the CSIR system should avoid dependence on single-point resources or tools by outlining in the CSIRP alternatives and secondary options and contacts. Also, it is common in IR that responders would need to perform ad-hoc planning as needs arise [99]. There needs to be outlined and disciplined procedures for how this could be achieved.

The principle of preparedness (P) is more particular than the generic requirement for CSIRTs to be prepared through having the proper competency and the preparation of a CSIRP. Preparedness here means the ability to respond in stressful and unpredictable situations. This is a joint result of having a complete CSIRP and competent CSIRT.

To reduce unpredictability, the CSIRP should have details about the variance of response steps to different security incidents. The plan also needs to be complete in terms of policy and support. The presence of such guidelines will facilitates the process of responding to the unexpected events. Otherwise, the team's efforts will be diverted from developing solutions to also finding "proper" processes of how to implement them.

The CSIRT's competency require that the team gets proper training in terms of problem solving and collaboration during difficult scenarios. The various modules presented in are a good resource serving this aspect of competency [97].

Another aspect of CSIRT's competency is the for need of effective integration and partnership as highlighted in [109]. Integration here refers to coordination between various internal bodies of the organization, while partnership refers to coordination with external entities. A well-qualified CSIRT can be hindered during incident handling by the ineffectiveness, inefficiency or even unavailability of services and support expected from other departments in or out of the organization. Therefore, the effectiveness of these links should be regularly evaluated to avoid surprises during the response handling.

A summary of the action points pertaining to flexibility and preparedness are available in Table 36.

Strategy/Policy Category	Action Items
Executive Empowerment (F)	Define authority categories and domains with clear delegations of executive powers relevant to CSIR
	Inspect authority channels to omit overlaps and redundancies
	Trust responders with autonomy to implement and execute necessary measures within predefined boundaries
Scalability (F)	Evaluate and record the capacity of the response system
	Outline procedures to recognize when the system capacity is over-reached and how to execute the process of seeking external assistance
Adaptability (F)	Outline procedures for how to respond to incident escalation
	Outline procedures for how to respond to unavailability of staff, tools and support services
	Enhance procedures for how to take corrective measures during the incident handling
CSIRP Completeness (P)	Ensure that the CSIRP has outlined procedures for handling different types of security incidents
	Ensure that the CSIRP has outlined descriptions and contacts of various support services
CSIRT Competency (P)	The CSIRT is trained to operate and coordinate under stressful and non-predictable conditions
	Establish effective integration channels between various bodies within the organization
	Establish effective partnerships with external bodies who might be involved in the CSIR activities

Table 36: Flexibility and Preparedness in the NFP Unpredictability Platform

CHAPTER SIX

FRAMEWORK ANALYSIS

In this chapter, several techniques will be used to analyze and assess the framework presented in the previous chapter. The first section, analyzes various aspects of the framework through hypothetical scenarios. three scenarios are presented in which the first demonstrates how to construct a simple PE system through the development process presented in Chapter three. The second scenario demonstrates how performance analysis could impact technical decisions by CSIRTs through the use of performance metrics. The third scenario involves the analysis of the notion of response time in the context of CSIRT.

The second section provides the expert feedback on the framework. Four experts were interviewed to discuss issues and challenges relevant to CSIRPE and provide a holistic assessment of the framework components presented in this work. The third section demonstrates how the framework displays the three characteristics of being comprehensive, flexible and industry compatible.

5.1 Scenario Analysis

5.1.1 Scenario I: Confronting Incident Instability

Objectives:

To demonstrate how to design a simple PE system for measuring a component of the CSIR system. The scenario focuses on analyzing response stability and provides an alignment of goals/aspects with PIs and PMs. In addition, an example is provided of how several PMs can be derived from a PI. A quantitative approach is used for the definition of PMs.

Background:

HumanDev is a semi-government organization that works with government agencies and the private sector to recruit new college graduates for available work opportunities. The organization maintains two large databases, one for college graduates and one for potential employers. The two databases are connected through a software portal called *Connect*, from which *HumanDev* Staff can create a profile for each client, find suitable work vacancies and later monitor the work progress of each client.

Due to several breaches targeting private personal data stored in the databases, *HumanDev* invested in contracting with a CSIR capability. This CSIR serves several departments in the government ministry that *HumanDev* reports to and maintains a unified CSIRP across the ministry.

Incident Details:

For several incidents, it was noticed that the team would declare an incident on a specific severity level, but then it escalates to higher levels. In some instances, it was recorded that the severity level would be updated several times during the incident handling. During a meeting between senior members of *HumanDev* and the CSIRT, it was acknowledged that responses suffer from instability issues. A committee was tasked with analyzing the previous responses to come up with a list of recommendations to improve the response stability. The committee was also tasked with formulating a mechanism for evaluating the stability of future responses.

Analysis:

The above scenario is confirmed to have issues with incident response stability, which frequently arise in the field of IR [101]. Although it is sometimes expected for an incident to escalate or de-escalate after its declaration, the observation found in the above scenario suggest that incidents would escalate in uncontrollable fashion or would be frequently re-classified during incident handling which indicate signs of ineffective response.

There are several factors that could lead to response instability. Examples include:

1- Inaccurate detection and classification of the incident

2- Incomplete or inadequate containment procedures

3- Ineffective execution of the CSIRP

4- Poor communication or coordination between team members

The analysis of the scenario will focus on the performance evaluation side. A simplified partial PE system is proposed using the process flow of the CSIRPE framework presented in Chapter Three.

The first phase of developing the PE system is to specify the design parameters and general strategies. Assume that the "Basic Environment" settings are used as specified in Table 45.

Using the second phase, since this is a simplified problem space either the definition of a PE goal or a PE aspect would be sufficient. A sample goal statement would be: "to develop mechanisms to measure the stability of the incident response system". Alternatively, the following PE aspect could be used: "Incident Response Stability". In this specific scenario, the derivation of PIs would be similar regardless of which method is used. For simplicity, the aspect "IR stability" will be used.

The next step would be to define performance indicators that map to the IR stability PE aspect. Five indicators are selected, see Figure 26. The PIs are defined as presented in Section 4.3. Note that the first PI focuses on measuring the overall system stability, while the other four PIs map to the four potential causes of instability as mentioned at the beginning of this analysis.

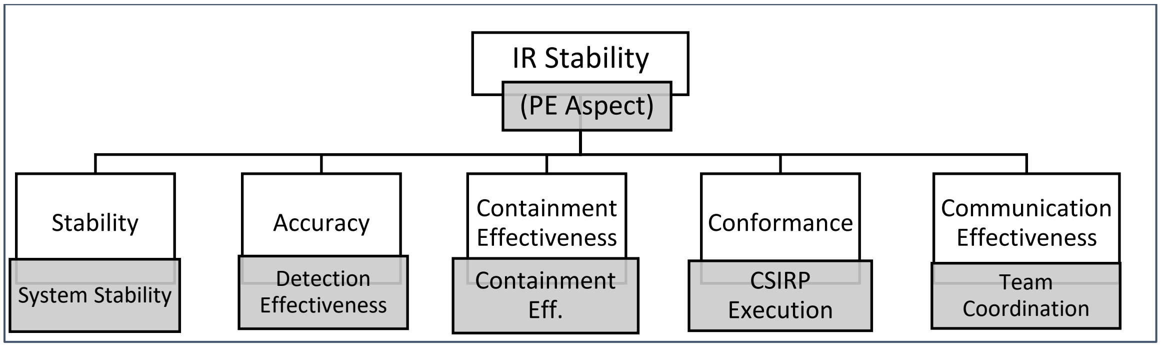


Figure 26: Scenario I: Selection of Performance Indicators

Using the PI Template presented in Table 49, the Stability PI is formally defined as the following:

Name	Stability
Code	ST_{PI}
Description	The ability of a CSIRT to prevent an incident from unexpected escalation and to maintain the response in a controllable state (See PI.44)
Classification	Generic
Priority	KPI
Goals & Aspects	PE Aspect: “IR Stability”
PMs	Escalation: $ST_{PI}.(E)$ Fluctuation: $ST_{PI}.(F)$

Table 37: Scenario I: Definition of the Stability PI

The next step, which is the start of phase III, is to design performance metrics. Two PMs will be defined for the Stability PI. The same procedure could be applied to the other PIs (For an example of an Accuracy PM see Scenario II, and see Scenario IV for an example of containment effectiveness PM).

There are various methods to measure the system stability. Focus here would be to monitor the classification of the incident, across the severity scale, from the time the incident is declared to the time of full recovery. Two metrics will be used, one focusing on measuring the escalation of the incident, and the other on capturing the fluctuations (escalation or de-escalations) of the incident state. Both metrics will use quantitative methods.

Let the severity scale be:

$$\boldsymbol{S = \{S_1, S_2, \dots, S_n\}}$$

Where S_1 represents the least severity level an incident could have and S_n the highest.

Let the severity state of the incident throughout the incident handling life cycle be represented through the following vector:

$$\boldsymbol{SS = [SS_{declared}, SS_1, SS_2, \dots, SS_k]}$$

Where $\boldsymbol{SS_{declared}}$ is the severity of the incident upon declaration, and the rest of the states represent the fluctuations in the severity level during different milestones of executing the incident response plan.

In an ideal and stable response, the value of $\boldsymbol{SS}$ vector would only contain $\boldsymbol{SS_{declared}}$ or in addition to other values such that all the other subsequent values of $\boldsymbol{SS_i}$ are smaller than $\boldsymbol{SS_{declared}}$. In that specific case, it means the incident gradually de-escalated. This could be represented mathematically as:

$$\boldsymbol{SS_j \leq SS_{declared} \quad \forall j : j \in SS}$$

The first performance metric which measures the incident escalation is defined in the following table, which uses the PM Template of Table 50.

Metric Name	Escalation Metric
Code	Escalation: *ST*$ST_{PI}.(E)$
Description	A performance metric to record the occurrence of incident escalation and its extent against a pre-defined severity scale.

	$ST_{PI}.(E) = SS_{max} - SS_{declared}$ Where $SS_{max} = \max\{SS\}$
Classification	Performance metric
Measurement	Quantitative
[Conditions]	There needs to be a pre-defined severity scale *S* of incidents that contain at least three severity levels.
[Attributes]	Severity Scale $S = \{S_1, S_2, \dots, S_n\}$
Interpretation	When no escalation occurs the value of *ST_{PI}.(E)=0,* Because: $SS_{max} = SS_{declared}$ A higher value of $ST_{PI}.(E)$ represent higher magnitude of escalation.
PIs	The stability PI: ST_{PI}

Table 38: Scenario I: Definition of Incident Escalation Performance Metric

The second performance metric, the fluctuation PM, records the changes of the incident state, not necessarily the escalation. For instance, if $\boldsymbol{SS = [SS_3, SS_1, SS_3, SS_4]}$, it represents an escalation of one level, i.e. from SS$_3$ to SS$_4$. However, it is noticed that the incident de-escalated, then went back to the declared level, then escalated before it was finally contained. This fluctuation is not captured through the escalation PM. The following table represents a formal definition of the metric:

Metric Name	Fluctuation Metric
Code	Fluctuation: $ST_{PI}.(F)$
Description	A performance metric to record the changes in the incident severity level through its cycle. $$ST_{PI}.(F) = \sum_{j=0}^{k-1} r_j$$ Where $r_j \begin{cases} 0 & if\ SS_{j+1} \leq SS_j \\ SS_{j+1} - SS_j & if\ SS_{j+1} \leq SS_j \end{cases} \forall\, SS_j \in SS$
Classification	Performance metric
Measurement	Quantitative
[Conditions]	There needs to be a pre-defined severity scale *S* of incidents that contain at least three severity levels.

[Attributes]	Severity Scale $S = \{S_1, S_2, \ldots, S_n\}$
Interpretation	When no fluctuation occurs the value of $ST_{PI}.(F)=0$, Whenever there is an occurrence of escalation, or a de-escalation followed by an escalation, the value of the PM gets incremented by a magnitude equal to the difference between the two states.
PIs	The stability PI: ST_{PI}

Table 39: Scenario I:Definition of Incident Fluctuation Performance Metric

For the analysis of the PM readings, the following remarks could be made:

1- A simple internal benchmark that compares the stability of the current incident to previous incidents within the organization would be sufficient.

2- If the PM results show that instability remains a matter of concern, then a trend analysis of several responses should be conducted to identify patterns and common causes

3- If there is interest to compare the stability of responses with external bodies, e.g. other ministry departments, then it would only be meaningful if the severity scales are very similar in terms of their classification mechanism and the number of defined severity levels. If there is a disparity between the two classification systems, then it would be feasible to compare the presence of instability by recording the occurrence of escalation or fluctuation, i.e. positive values for the escalation and fluctuation metrics.

4- The occurrence of escalation is more likely to happen compared to fluctuation which could occur due to very poor incident handling or within highly dynamic environments. Therefore, the escalation PM should be maintained for longer period, even when stability is maintained; while the fluctuation metric could be applied for a period of time until the fluctuation issue is diminished.

5- It is noticed that when both metrics have the same positive value, then it indicates the occurrence of an escalation. However, if a fluctuation occurs then the value of the fluctuation metric would be higher than the escalation metric. Therefore, both metrics should be analyzed collectively for better understanding of the system stability.

Conclusion:

The above scenario addressed a CSIR system with a specific performance issue, namely stability. Through the use of the CSIRPE framework development process, the issue was addressed through a simple process structure that provided guidance to how the issue should be controlled and measured. The process started from goals to indictors to performance metrics and finally to analysis recommendations. In summary, the scenario demonstrated how CSIR PE issues could be addressed in a systemized manner through the use of the CSIRPE framework.

5.1.2 Scenario II: Enhancing Incident Declaration Strategy

Objectives:

To demonstrate how performance analysis can impact incident response strategies. The scenario involves a situation in which both financial and non-financial performance metrics are to be used. The scenario applies several performance analysis techniques like trend analysis, gap analysis and predictive analysis.

Background:

FastRescue is a computer security incident response team that serves a large business corporation. The core team is composed of five highly talented security professionals, and is partnered with several logistics and support teams inside and outside the corporation to provide as-needed assistance during incident handling. The team operates in a relatively high budget. However, the *FastRescue* services had been viewed positively by the executives who are convinced that despite the high operating costs of *FastResecue*, the financial and management disadvantages of not using such CSIRT would have caused the corporation more damage.

Scenario Details:

In the past year, the corporation was undergoing financial difficulties that require making several cuts to the various departments. A committee consisting of the CIO and a group of risk assessment analysts was formed to review potential cuts to the operation costs of *FastRescue*.

During the assessment, the committee noticed that the team follows a response strategy in which every incident is declared at the highest severity level (Degree Five) and is gradually decreased to the actual severity level. This strategy ensures that no incident is ever declared below its actual severity level. The committee is evaluating whether such strategy is cost-effective.

Analysis:

The above scenario involves a situation in which financial objectives seem to be in tension with CSIRT objectives. In order to conduct performance evaluation of the effectiveness of the top-down approach of incident declaration, several financial and security performance metrics need to be deployed. The committee would need to conduct a trend analysis on data from previous incidents; in addition to predictive analysis on how eliminating the above strategy would impact the cost effectiveness of the incident response.

In order to achieve the above, the trend analysis time domain was defined between 2012 to 2016. The committee defined four PIs from which a total of nine PMs were derived, see Table 40. The definition of these PIs is as defined in PI.21, PI.19, PI.20 and PI.38 respectively.

Performance Indicator (PI)	Performance Metric (PM)		
	Name & Code	Code	Brief Description
Detection Effectiveness (DE)	Incident Classification Accuracy	ICA	Was the incident declared on the correct severity level?
Detection Effectiveness (DE)	Time to Confirming Incident Classification	TCIC	How long did it take to approve or disapprove the initial incident classification?
Response Cost (RC)	Hourly Average Operating Cost	HAOC	What is the average incident operating cost per incident (depending on incident classification)?
Response Cost (RC)	Average Loss	AL	Average financial loss due to incident handling (despite the use of CSIRT)
Cost Effectiveness (CE)	Predicted Loss	PL	Predicted average financial loss per incident due to cyber incidents (without CSIRT intervention)
Cost Effectiveness (CE)	CSIRT Cost Effectiveness	CCE	Average financial savings per incident due to the use of CSIRT
Cost Effectiveness (CE)	Detection Inaccuracy Cost	DIC	Average financial loss due to inaccuracy of initial incident classification
Cost Effectiveness (CE)	Incident Classification Strategy Cost	ICS-CE	Cost effectiveness of incident classification strategy (top-down approach)
Response Time (RT)	Average Response Time	ART	Average incident response time, from incident declaration to incident resolution, depending on incident severity

Table 40: Scenario II: List of Performance Indicators and Metrics

For the purposes of this scenario analysis, only the ICA and TCIC PMs are defined, see Table 41 and Table 42. The remainder of the PMs, which are mainly financial metrics, could be defined in a similar manner.

Metric Name	Incident Classification Accuracy
Code	*DE_{PI}.(ICA)*
Description	A performance metric to compare the actual severity level of an incident to the initially declared level: $$DE_{PI}.(ICA) = \frac{n - \lvert AS - S \rvert - 1}{n - 1}$$

	Where S is the incident severity level upon incident declaration, AS is the actual incident declaration as confirmed later during incident handling, and n is the total number of severity levels (in this scenario 5 levels)
Classification	Performance metric
Measurement	Quantitative
[Conditions]	There needs to be a pre-defined severity scale S of incidents that contain at least two severity levels.
[Attributes]	Severity Scale $S = \{S_1, S_2, \dots, S_n\}$ [In this scenario n = 5]
Interpretation	If the actual and declared severity levels match, then the value of ICA would be 1, meaning full accuracy. If the declared value is on the edge side of the scale, and the actual is on the other edge, then the value of ICA would be 0, i.e. no accuracy.
PIs	The detection effectiveness PI: DE_{PI}

Table 41: Scenario II: Definition of Incident Classification Accuracy

To be more thorough in defining the above performance indicators, several descriptive metrics (See: M.1: Metric Type) are needed. These descriptor PMs could be defined using the same above PM template, with the exception that under the "Classification" field, the value would be "Descriptor Metric" instead of "Performance Metric". A list of these descriptive metrics along with a brief definition is presented in Appendix A: Table 52.

Metric Name	Time to Confirming Incident Classification
Code	$DE_{PI}.(TCIC)$
Description	A performance metric to measure the time in hours from the moment an incident is declared on a specific severity level to the time that level is confirmed or approved to be otherwise. $DE_{PI}.(TCIC) = CT - DT$ Where DT is the timestamp of declaring the incident, CT is the timestamp of approving the severity level of the incident and the above subtraction is time subtraction
Classification	Performance metric
Measurement	Quantitative

[Conditions]	There needs to be a pre-defined severity scale S of incidents that contain at least two severity levels.
[Attributes]	Severity Scale $S = \{S_1, S_2, \ldots, S_n\}$ Escalation: Yes/No (did the incident escalate)
Interpretation	The above metric is suitable when gradual incident classification strategies are used, like the top-down approach. If an incident classification is validated only during post-incident analysis, then this PM should not be used. When using this PM, a CSIRT should be aware to the fact that incidents may escalate. In such scenario, the above time measurement might not be representative of actual events.
PIs	The detection effectiveness PI: DE_{PI}

Table 42: Scenario II: Definition of the TCIC Performance Metric

The committee started by collecting data about the number of incidents per year, categorized by the severity level, see Appendix A: Table 53. Then a break-down of the fixed (Appendix A: Table 54) and operation costs (Appendix A: Table 55) associated with incident handling was recorded. Incident response fixed costs would normally include CSIRT members' salaries, software licenses, equipment, continuous professional development budget and contractual obligations. For simplicity, the fixed costs are minimized to the core team salaries, broken over the period of the five years. Similarly, a summative figure related to the operation costs of the core, logistics and support teams is used for the representation of operational costs. The calculated costs were averaged based on the severity level of incidents, per each hour of incident handling.

Next, the analysts conducted a financial and risk assessment to investigate the overall cost effectiveness of the organization's use of the CSIR capability, see Appendix A: Table 56. For each incident, two financial estimates were made. The first is the actual financial loss caused by the incident despite the use of CSIR, while the second is a predicted figure of how much loss would have been endured if there was no CSIR. The difference between the two numbers, averaged over each category of incidents, represents the cost effectiveness of using the CSIR capability.

The third step was to analyze the technical incident data for time stamps. Two main time stamps were analyzed: the total response time and the time it took to approve or disapprove the initial severity level of the incident, see Appendix A: Table 57.

The final step is to analyze the cost effectiveness of using the top-down approach to incident declaration. To achieve this, two main compound performance metrics were used. The first metric, Detection Inaccuracy Cost (DIC), measures how much unnecessary cost was endured due to classifying an incident on a higher severity level compared to its actual level. This could be calculated through analyzing the data (over the severity scale) of the ICA and TCIC PMs, along with the hourly operating cost and the total number of incidents under each incident category. Note that when an incident is actually on the highest scale (degree five), then there would be no inaccuracy cost, because the CSIRT is operating in its full capacity. On the other hand, when an incident is of the lowest severity level, the inaccuracy cost would be high due to the fact that the team has used too many unnecessary resources at the early stages of the response. The obtained data for the DIC PM is presented in Appendix A: Table 58.

The second performance metric targets assessing the current incident declaration strategy against other possible strategies. A proposal was made to limit the incident classification period to a maximum of three hours, in which the capacity of the CSIR is boosted through providing additional financial, technical and logistical support. An estimation was made about the expected increase in the financial resources based on the statistical occurrences of incidents over the severity scale. This estimation is coupled with a risk assessment of the expected financial loss endured if the CSIRT fails to adequately classify the incident during the three hour period. The addition of both figures summarizes the financial costs of the new proposal. This is compared with the detection inaccuracy cost currently endured due to the top-down incident declaration approach. The difference between the two numbers represent the cost effectiveness of using the top-down incident declaration approach.

The results of the above two PMs are presented in Appendix A:Table 59. As the data displays, the financial benefits of using the top-down approach by *FastRescue* are far more than the condensed identification proposal. Indeed, in this specific scenario, abandoning the above strategy would have costed the company an amount similar to the cost of doubling the size of the CSIRT.

Conclusion:

The above scenario analysis demonstrated how a variety of performance metrics and indicators could be grouped to solve an issue pertaining to performance. The scenario involved a tension between financial and technical objectives. The reckoning that the adopted top-down approach of incident declaration might be costly is reasonable. However, an objective analysis through the use of performance metrics proved otherwise. This asserts one of the foundations in which the framework of this project is built upon, namely the use of performance analysis would lead to objective decision making. The quantitative analysis using the performance metrics provided means for making decisions that are based on operational data compared to relying on speculations and subjective analysis. In this specific scenario, the performance analysis had an impact on the technical functionalities and response strategies of the CSIRT. Had the PE analysis recommended the disapproval of this strategy, significant changes would have been applied to the CSIRP and the overall execution of the incident response activities.

5.1.3 Scenario III: Analyzing Response Time

Code	Performance Metric
AT	Analysis Time
CT	Compromise Time
CST	Classification Time
CTT	Containment Time
DAT	Detection Analysis Time
DCT	Declaration Time
DET	Data Exfiltration Time
DT	Detection Time
DVT	Discovery Time
EHT	Enhancement Time
ERT	Eradication Time
FMAT	First Malicious Action Time
FRT	Full Recovery Time
HT	Incident Handling Time
ICF	Incident Containment Factor
ICT	Initial Compromise Time
MT	Monitoring Time
PRT	Partial Recovery Time
RT	Response time

Table 43: List of Performance Metrics for Scenario III

Objectives:

To analyze the notion of *response time* in the context of CSIR. The scenario will demonstrate how the principles outlined in the SAC complexity model can be used in the development of performance metrics.

Background:

SecureLab is a small but dynamic CSIRT that operates in higher education institution.. The *SecureLab* team is part of a larger CSIRT, *SecureEd* that covers all types of security incidents in the institution. The *SecureLab* team is focused on responding to security incidents targeting the institution's 36 educational computer laboratories. Most of the incidents handled by *SecureLab* are worms and viruses that spread across the network subnets, mainly due to students' attaching external devices like flash drives.

Scenario Details:

In response to claims of slow responses, *SecureEd* is interested in developing a mechanism to measure the response time for incidents defined within the scope of *SecureLab*. The current mode of operation consists of three main steps:

1- *SecureEd* detects, classifies and then pass incident handling to *SecureLab*

2- SecureLab works on the containment, eradication and partial recovery

3- *SecureEd* works on full system recovery, incident analysis and closing the incident.

The objective of measuring the response time is to identify elements of the incident response that might contribute to the ineffectiveness of the response

Analysis:

Measuring response time is a common performance measure across various disciplines. The definition of response time vary depending on the general specifications of the environment in each discipline. Even within each discipline, there are various methods to define response time depending on the specific characteristics of the underlying environment. Failure to recognize this variance of methods may lead to incorrect analysis emerging from inconsistent comparisons.

The principles of the SAC complexity model will be applied to the above scenario to develop a mechanism for measuring response time that suits the environment of

SecureLab. Measuring response time will be divided into measuring the time of completing simple activities (the principle of *Simplicity*), which will be aggregated together to measure more complex activities (the principle of *Cascading*). In addition, the principle of *Approximation* will be used in defining some time-slots within the overall response-timeline.

Using a top-down approach, the response-time can be divided into three main time-slots, see Figure 27:

1- *Detection Time (DT)*: The time it took to detect and classify the incident by *SecureEd* before its passed to *SecureLab*.

2- *Incident Handling Time (HT)*: the time it takes *SecureLab* to contain, eradicate and partially restore the system until it is passed back to SecureEd.

3- *Incident Analysis Time (AT)*: the time it takes *SecureEd* to completely restore the system, formulate recommendations and close the incident.

From the above description, the incident response-time (RT) can be defined as the following:

$$\boldsymbol{RT = DT + HT + AT}$$

Note that the above break-down of incident response time is analogous to the general definition found in the transportation field, which defines the response time into the following three main time-slots **[117]**:

1- Elapsed time from incident occurrence to detection

2- Elapsed time from the point at which the IR team is called out until its arrival on-scene and completes the response activities.

3- Elapsed time to normal traffic flow restoration.

Each of the three timeslots (i.e. DT, HT and AT) need to be subject to further simplification. The following paragraphs investigates each time-slot independently.

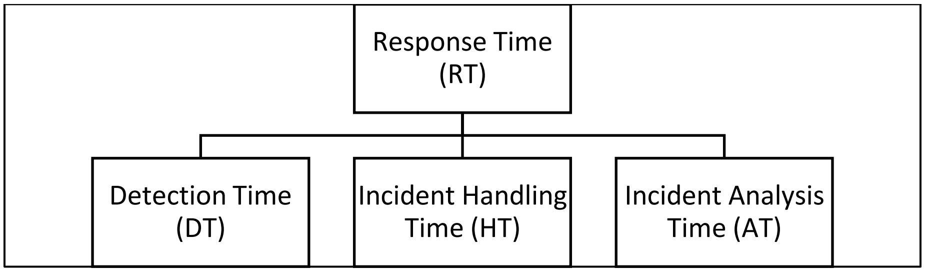


Figure 27: Scenario III: Response Time Components

Detection Time (DT):

The detection time can be divided into two main time-slots. The first is the time an incident is active in the system before it was detected by *SecureEd*; and the second is the time period starting from the time an incident is reported/discovered to the time it is classified, declared and passed to *SecureLab*. The first time-slot will be referred to as: *Compromise Time* (CT) and the second as *Detection Analysis Time* (DAT).

To further break-down the Compromise Time (CT), the time-slots defined by the VERIS project [73] will be used, which are:

1 *First malicious action time* (*FMAT*): Beginning of the threat actor's malicious actions against the victim. Port scans, initiating a brute-force attack, and even physical recon, are a few examples. This is only relevant to intentional and malicious actions.

2 *Initial compromise time* (*ICT*): First point at which a security attribute (C/P, I/A, A/U) of an information asset was compromised.

3 *Data exfiltration time* (*DET*): First point at which non-public data was taken from the victim environment. Only applicable to data compromise events.

Note that the FMAT, ICT and DET could be interpreted differently depending on the nature of the incident. Such discussion is beyond the scope of this scenario. Overall, the FMAT could be viewed as the preparation time of the malicious event, i.e. breach activities before the actual security exploitation; the initial compromise time (ICT) is the timeframe when security has been compromised, despite the fact that no harm has occurred; and the data exfiltration time (DET) is the time when harm is endured before the detection/reporting of the incident. Using the above, the three components are treated as time periods instead of timestamps.

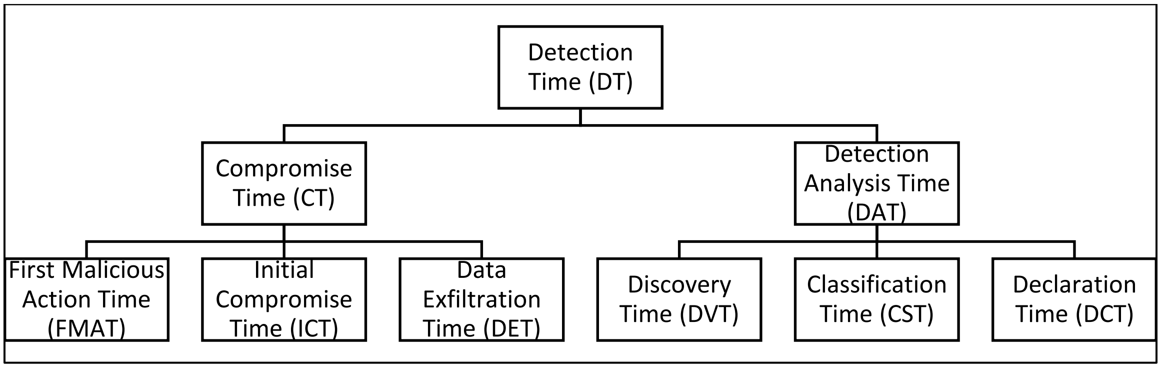

Detection Time (DT)
Compromise Time (CT)
Detection Analysis Time (DAT)
First Malicious Action Time (FMAT)
Initial Compromise Time (ICT)
Data Exfiltration Time (DET)
Discovery Time (DVT)
Classification Time (CST)
Declaration Time (DCT)

Figure 28: Scenario III: Detection Time (DT) Components

The Detection Analysis Time (DAT) can be broken into three time events:

1- *Discovery Time* (DVT): The timestamp of the first reporting of the incident, i.e. when *SecureEd* learnt about the incident.

2- *Classification Time* (CST): The timestamp the incident type and severity is confirmed.

3- *Declaration Time* (DCT): the timestamp when the incident is declared and passed to the *SecureLab* team.

Based on the above definitions the following equations are to be used for calculating the Detection Time (DT):

$$DT = CT + DAT$$

$$CT = FMAT + ICT + DET$$

$$DAT = DCT - DVT$$

Figure 29: Calculation of Incident Detection (DT) Time

When analyzing the above DT metric, having a long compromise time (CT) would suggest that the environment has poor detection mechanisms or weak security measures that would permit a threat to remain undetected for a long time. Ideally, an incident should be detected as soon as it is present in the system, i.e. the value of CT is 0. However, this is practically infeasible. Some studies [234][67] suggest that responders should be available 24x7; however, this imposes several practical challenges and might only be applicable to governments or very large organizations. Therefore, there would be always a time period between the detection of the incident and the actual execution of the response. It is obvious that an effective response would require minimizing this time period.

On the other hand, the detection analysis time (DAT) depends mainly on the readiness of the CSIRT and its effectiveness in executing the CSIRP. Ideally, a team should spend very minimum time in classifying the incident and declaring it. But rushing this could result in inaccurate understanding of the incident and consequently inadequate response measures. Therefore, it is reasonable to assume that there will be a time-period from when an incident is discovered until it is classified and declared. In an effective environment, an incident should be declared as soon as it is classified, i.e. (*DCT-CST = 0*). Yet, the logistics of classifying the incident and handing it to the appropriate team (in this scenario, the *SecureLab* team) would require some time. One of the performance aims of the CSIRT should be to minimize this time period.

Incident Handling Time (HT):

The incident handling time (HT) represent the overall time spent by the team *SecureLab* in responding to the incident. Applying the simplification-cascading process to this time-period can yield the following three main time-periods (Figure 30):

1- *Containment Time (CTT)*: The time-period taken to prevent the incident from spreading beyond its initial infected area.

2- *Eradication time (ERT)*: The time-period to fully remove the malicious components from the system.

3- *Partial Recovery Time (PRT)*: the time-period taken to restore the basic components of the system to their initial state.

The incident handling time (HT) would be an additive metric of the above three metrics, i.e. $HT = CTT + ERT + PRT$

The containment, eradication and recovery steps have many interconnected activities. In addition, each step would involve a series of actions that would vary depending on the type of incident. Therefore, defining time-periods that would separate each step is not intuitive.

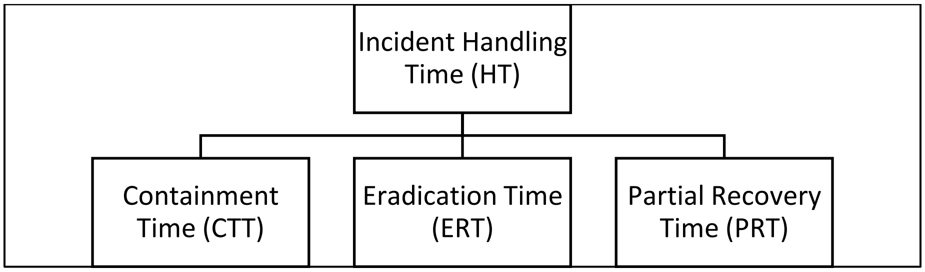


Figure 30: Scenario III: Incident Handling Time (HT) Components

To overcome the above challenge, instead of measuring the time it takes a CSIRT to execute specific activities, timestamps could be recorded when specific *response* *indicators* have taken place. For instance, the *partial recovery time (PRT)* could be recorded based on the fulfillment of one or more of the following response indicators:

- A computer lab is partially functioning such that all software required for conducting a course session are running, despite some auxiliary software (e.g. browsing or media players, printing software) being unavailable
- Instead of focusing on software and applications, recoverability could be measured based on availability of services. For instance, backup or alternative applications could be launched until the original applications are restored.
- A specific number of computers in a lab are fully restored. For instance, at least half of the computers in a classroom are fully restored enabling classes to take place (each two students could share a computer).

The above response indicators are quantitative metrics that rely on simple counting methods. This does not necessarily apply to all response indicators. For instance, the containment and eradication phase do not represent exclusive activities. It is a common practice to get eradication activities initiated while the final steps of the containment phase are still on-going. For such situations, it would be better to use heuristic response indicators that rely on estimations instead of actual measurements. This is where the principle of *Approximation* from the SAC model is applied.

A detailed example is provided below of how the containment and eradication phases could be distinguished. Instead of waiting for full containment before declaring the start of the eradication process, the CSIRT can make an estimation of when the containment process is considered satisfactory to apply eradication measures. This is done through observation some threshold metrics that estimate the containment achievement, e.g. 80%

of containment is completed. In the example below, the metric *containment factor* is defined. When the CSIRT has exceeded a pre-defined threshold, i.e. specific containment factor value, the line between eradication and containment could be drawn.

Containment Factor Performance Metric

Whenever an incident happens, one of the objectives of the CSIRT is to isolate this incident and prevent it from spreading to other points in the network. This is especially important when incidents involve self-replicating or network spreading worms, or distributed denial of service attacks DDoS attacks. The *incident containment factor (ICF)* performance metric aims for measuring the success of the response team in blocking the affected parts of the organization from infecting the healthy parts.

Within a specific environment, a security incident might be global or local. A global incident happens when everyone in that environment gets infected by the incident. Most organizations have security infrastructure empowered with some security policies to decrease the likelihood of such incidents. Yet, it is not infrequent for an organization to be subjected to a global incident. The occurrence of a global incident is a serious indicator that the environment lacks basic defensive mechanisms and the whole security infrastructure should be subject to scrutiny. In local incidents, part of the environment is infected while other parts of the same environment remain healthy.

Note that the above understanding of global and local incidents as described above is similar to the understanding provided in [235] which is different than how the terms are used in the CERT document [2] which defines a global incident as one that involves interaction with external entities like law enforcement, while local incidents remain internal within an organization.

Let the infected areas of the environment be identified through a descriptive metric called incident *locale*. There are various methods to define the *locale*, depending on the type of incident and nature of the environment. In most scenarios, the *locale* could be defined in terms of either *region* or *population*. The *region locale* is the portion of the network, physical or virtual, that is infected by the incident. These portions can be identified as number of hosts connected to a local switch/router, number of hosts within an IP range or within a specific subdomain, or number of departments within an organization. The *population locale* represents the number of users or clients affected by the incident. In some scenarios both locales should be defined, while in others only one is needed for a meaningful description of the incident. This definition of *locale* can be easily incorporated into VERIS [73] under: Incident description → Assets → Variety.

Let *N* denote total number of hosts in an environment. And let $\boldsymbol{N_D}$ represent number of hosts in a region or population infected by the incident as recognized by the detection/identification phase. Note that if $\boldsymbol{N_D}$ is very high then that may suggest weak detection mechanisms and could be used along with the compromise time (CT) metric to trigger a review of the adopted detection tools and mechanisms. The value of $\boldsymbol{N_D}$ can also help in determining the severity of an incident during the initial assessment.

Isolating and preventing an incident from spreading can be mapped to keeping the number of infected hosts after declaration close to $\boldsymbol{N_D}$ as much as possible. The incident containment factor (ICF) performance metric measures the team's effectiveness in achieving this. Note that in order for the ICF to produce correct conclusions, the value of *N* should be large enough. The ICF metric is defined as below:

$$\boldsymbol{ICF} = \begin{cases} \mathbf{0} & \boldsymbol{N_D = N} \\ \mathbf{1} - \dfrac{(\boldsymbol{N_f} - \boldsymbol{N_D})}{\boldsymbol{N}} & \boldsymbol{otherwise} \end{cases}$$

Figure 31: Definition of Incident Containment Factor (ICF) Performance Metric

where $\boldsymbol{N_f}$ is the total number of hosts infected prior to containment. The term $(\boldsymbol{N_f} - \boldsymbol{N_D})$ represents number of hosts infected after declaration and before full containment. In the best scenario, when no hosts were infected in addition to $\boldsymbol{N_D}$, the value of $\boldsymbol{N_f} = \boldsymbol{N_D}$ and the *ICF* will be 1 which means the incident is not spreading which is an indicator of containment. In the worst scenario, if only one host was initially infected and the whole network gets infected, the value of *ICF* will be approximately 0, because the value of *N* is large. For the special case where the whole network was initially infected, the containment factor equals to 0.

Since threats have various compromise techniques, the team should be careful not to assume that if an incident stops spreading then that it means it is contained. Instead the team should rely on the application of containment measures like blocking traffic from specific ports as indicators of containment. Again, the application of these measures do not necessarily guarantee full and effective containment, but they do represent a reasonable approximation of that.

The above definition assumes that all hosts/users in the environment are of equal importance. In practice, this is not the case as hosts/users vary in their importance and impact on business operations. One way to reflect this in the containment factor is to use weighted values according to the following definition:

Let the environment holds the set of hosts:

$$\boldsymbol{H} = \{\boldsymbol{h_1}, \boldsymbol{h_2}, \dots, \boldsymbol{h_N}\}$$

Let the set of infected hosts before declaration be:

$$\boldsymbol{H_D = \{h_1, h_2, \dots, h_D\}}$$

Let the set of infected hosts after declaration and before containment be:

$$\boldsymbol{H_f = \{h_1, h_2, \dots, h_F\}}$$

Let the importance (weights) of every hosts be represented by:

$$\boldsymbol{W = \{w_1, w, \dots, w_N\}}$$

Where $\boldsymbol{w_i}$ represent the weight of the host $\boldsymbol{h_i} \in \boldsymbol{H}$

The containment factor is defined as:

$$\boldsymbol{ICF} = \begin{cases} \boldsymbol{0} & \boldsymbol{H_0 = H} \\ \boldsymbol{1} - \dfrac{[\sum_{j=1}^{F}(\boldsymbol{h_j * w_j})] - [\sum_{k=1}^{D}(\boldsymbol{h_k * w_k})]}{\sum_{i=1}^{N}(\boldsymbol{h_i * w_i})} & \boldsymbol{otherwise} \end{cases}$$

Figure 32: Definition of Weighted Containment Factor Performance Metric

In either case, whether the CSIRT decides to use the standard or the weighted definition, the end of the containment period could be decided by reaching a specific value on the containment factor.

Incident Analysis Time (AT):

When the *SecureLab* completes its task, the incident is passed back to *SecureEd* to finalize the incident, formulate recommendations and declare the end of incident response cycle. This can be divided into three main time-slots:

1- *Full Recovery Time (FRT)*: the time period it takes the CSIRT to fully restore the system to its initial state; probably with enhanced security measures.

2- *Monitoring Time (MT)*: A specific grace period which a CSIRT allocates to ensure that the incident is not reoccurring. The same period could also be used to monitor the impact of introducing new enhancements to the environment as per the recommendation of the post-incident analysis.

3- *Enhancement Time (EHT)*: The time-period the CSIRT spends in formulating a list of lessons learnt and recommendations. Some of these recommendations relate to enhancing incident response and others for improving the environment security. This time period can be further divided into two sub-timeslots. The first is the time the team spends in conducting post-incident analysis until a list of recommendations are produced. The second timeslot is spent in implementing these recommendations, e.g. updating the CSIRP and enforcing new security policies.

It is noticed that the MT and EHT are not disjoint as it is expected that both times will have an overlap.

The Big Picture:

Now that the notion of response time has been deconstructed into several components, these components need to be reconstructed again into a summative figure which is commonly referred to as the incident response time (RT). A CSIRT would need to consider the following issues:

1- When does the response time start? Does it start with the discovery time (DVT) or the declaration time (DCT)? Should the compromise time (CT) be considered?

2- When does the response time end? Does it end after full recovery time (FRT), or should some monitoring time (MT) and enhancement time (EHT) be considered?

3- What does long or short time-periods for the above metrics indicate about the response, the CSIRP and the system security?

4- Does the *SecureEd* team need to use all of these time-periods, or should some of them be summed together?

5- Which metrics can tolerate high values and which ones should the team attempt to minimize?

6- From the implementation perspective, who is responsible for recording each time-stamp? Is this done manually or through automation? .

Conclusion:

This scenario analyzed one of the main concepts of performance evaluation, which is incident response time. The scenario demonstrated how the construction of performance metrics could be achieved through the SAC complexity model, which is built on the principles of simplicity, approximation and cascading. The response time indicator was deconstructed into the measurements of simple activities. Despite the fact that these activities were simple, their measurement was not always straightforward. Although the scenario analysis is generic can be applied to different CSIRT settings, as the major components of the response time (RT) is the same across environments, it remains a matter of design how to define the boundaries of the time and which time-periods should be highlighted.

5.2 Expert Feedback

5.2.1 Methodology

The objective of consulting CSIR experts is not to provide survey results on where practitioners stand with regards to various issues raised in this dissertation. Instead, experts are consulted for their perspective on these issues based on their professional experience . Therefore, each expert's view point might not necessary be the mainstream industrial view.

Four experts were selected in a manner that reflects experience and diversity. All four experts have more than ten years of experience and have substantially engaged with incident response, though in various contexts. The first expert comes from a law enforcement background, the second from corporate and business background, the third from legal investigations and digital forensics background, and the fourth from system administration and computer security management background. All four experts have technical experience in CSIR. In addition, the second and third experts have CSIR leadership and management experience.

Each expert was interviewed separately. The shortest interview took two hours and a half and the longest took four hours and a half. Upon the request of the interviewees, their feedback was recorded anonymously.

Each interview covered the ten themes listed in Table 44.. However, the length and depth of discussion varied depending on the interviewee experience and interest in the topic.

The following sections provide an overview of the main opinions expressed by each expert, followed by a holistic reflection on the collective feedback. Several components of the framework were revisited or modified as per the feedback of the experts. Whenever such modification was made, a citation to the interview was made.

#	Theme	Sample topics
1	Need for CSIRPE	Is there a need for CSIRPE? Does the rewards exceed the overhead? How is it expected to enhance CSIR?
2	Current Practices	How are CSIRT evaluated? Is it done in a systematic way? Does the evaluation focus on pre or post-incident evaluation?
3	Multidisciplinary Survey	How is CSIR different than other disciplines? Should other disciplines be visited for guidance? Why?
4	Challenges	What are the obstacles to developing a CSIRPE? Is it feasible from the first place? Comments on the complexity and unpredictability modeling
5	CSIRPE framework features	Defining: comprehensiveness, flexibility and industry-compliance. Overall comments on the framework four-phases and ten components
6	Parameters & Strategies	Comments on the list of parameters and strategies
7	Performance Indicators & Metrics	What are the main indicators for effective incident response? Qualitative vs quantitative methods.
8	Performance Analysis	Comments on the analysis methods. Who are the stakeholders for conducting the PE analysis?
9	Performance Validation	Comments on the validation methods. Why do security metrics suffer from practicality and validity issues
10	Overall Comments	Who would be interested in this research? How do you view the value and need of this work? Issues of concern and recommendations for enhancement.

Table 44: Expert Feedback Survey Themes

5.2.2 Expert 1 Feedback

Expert 1 [7] is a senior digital forensics and incident response investigator offering consultancy services to state police. (S)He has ten years of experience in that role. The expert received training from several leading industries in the field of digital forensics and from the government bodies. The expert is notable for his/her contributions in the area of forensics and incident response to incidents involving mobile phones and mobile computing devices.

The expert expressed high interest in the research topic acknowledging the need for evaluating CSIR capabilities. Two comments were made with regards to that. First, the need for evaluating CSIR is well-established industrial reality in which the main driving force being liability and insurance. Second, such research direction is mainly beneficial to organizations (i.e. the team as a whole) not individuals (i.e. team members). Therefore, this type of evaluation would face reluctance or resistance, but the benefits would speak for themselves after several years of implementation.

Commenting on the current industrial practices with regards to CSIR evaluation, the expert noted that such evaluations exist but in scattered forms across different evaluation schemes. For instance, different components of CSIR are evaluated under security resiliency, vulnerability assessment, disaster recovery plan, business continuity plan and risk assessment. The more technical evaluations are done in ad hoc fashion by smart people.

It was affirmed by the expert that current evaluation practices focus on the preparation more than the post-incident execution. Using the expert's words: "makes sense, but not good". It makes sense because companies engage in continuous preparation, while post-incident analysis is less frequent. However, it is "not good" for two reasons. First, most preparation evaluation schemes are not comprehensive. The expert noted that only in six instances at the national level (mostly military) there was a full team simulation, i.e. engaging CSIRT with all support teams, e.g. legal and logistics. This makes preparation-phase evaluation incomplete or ineffective. Second, these evaluations tend focus on preventive measures more than actual evaluation of the team performance. Thus, enhancement recommendations are biased towards the technical preventive side.

Several components of the framework were considered by the expert as “too advanced” for the current industrial needs. Examples include: the PE monitoring functional model, the predictive analysis method, validation methods and concurrent incident handling. When asked that some of these components are derived from the NIMS and NFR, the answer was that cybersecurity response is not well-integrated into these national frameworks posing several inconsistencies. This goes back to the fact that these frameworks were not initially designed for cybersecurity response. When consulted if these components should be omitted from the framework, the expert suggested keeping them for research purposes.

The multidisciplinary survey was well-received by the expert. There is much to learn from the other disciplines, speaking from the experience of working with police departments. The expert noted that the only exception is that cybersecurity threats do not pose immediate threat to human life. It was added: “but eventually it will”. The expert saw “no industrial cry” for such multidisciplinary study. However, those working at the state and national levels, like NIMS and NFR, would be directly interested in this type of research.

When asked to comment on how a framework would be considered comprehensive, flexible and industry-compatible, the expert provided the following answers:

1- Comprehensive: a framework that leads to an end point that is useful, i.e. there is a process that produces an output that could be used in the industry.

2- Flexible: a framework that permits responders to develop their own customized solutions (derived from the higher principles/policies). It is also one that does not seek high granularity in measurements

3- Industry-Compatible: a framework that is derived from the NIST and CERT documents. The SANS document was considered insignificant.

When requested to provide a list of the most important performance indicators to evaluate CSIR, the following six indicators were suggested: detection effectiveness, response time, intelligence capacity, professional development, partnership effectiveness and documentation effectiveness. The expert emphasized that indicators should be analyzed in a compound fashion in order to produce reasonable results. It was also noted that qualitative methods are suitable for enhancing the team performance, while quantitative measures are useful for improving the CSIR capability at the enterprise level.

Finally, the expert noted that large companies and research agencies like RAND would be highly interested in the research findings of this project. The expert expressed eagerness to actual implementation of the proposed framework.

5.2.3 Expert 2 Feedback

Expert 2 [66] [92] is a senior manager information security leader who leads a 24x7 global team of security professionals and senior management in breach preparedness assessment and incident response to high stakes breaches. Extensive experience in managed security services (MSS) and security & risk consulting (SRC) across twelve client industries. Areas of expertise include incident response investigation, vulnerability assessment, penetration testing, ISO 27000 series and digital forensics investigations.

As the interview started, the expert was quick in asserting two observations about the topic. First, conducting this research is necessary for businesses, noting that current performance evaluation practices are “sparse and varying”. Second, conducting this type of research is difficult because executives do not like this type of research. It provides evidence for lack of security and preparedness. In other words, despite this topic being necessary there is clear reluctance in the industry about pursuing it.

To support the need for evaluating CSIRTs, the expert gave three examples. The first is the catastrophe of the Catrina Hurricane. If no one is interested in the performance of responders, why was there a public cry!, the expert noted. The second example was cited from the findings of the Ponemon survey on incident cost [4]. Part of the survey conclusions is that decreasing the *time to detect* would lead to a reduction in incident cost. The third example, when an organization needs to higher an external CSIR capability. In such scenario, defining SLAs is not possible without clear performance metrics. These examples provide practical and empirical evidence for the need of evaluating the performance of CSIRTs.

With regards to the multidisciplinary survey, the expert remarked that: “we like to think we are special and different; but we are not”. This is because, at the end of the day, most performance metrics go back to one main factor: *return on investment*. This explains why most organizations quantify performance in financial metrics, which might not necessarily useful for team enhancement.

According to the expert, the main challenge to evaluating CSIRT is that: “we do not know exactly what are the objectives of CSIR”. In other response systems, the main objective is to protect lives which is not applicable to CSIR. When asked to define these objectives, the expert replied with reducing costs, ensuring business continuity and maintaining the CIA (confidentiality, integrity and availability) of the data. Interestingly, these objectives are similar to the three CSIR objectives defined in Section 3.4.2.

With regards to the complexity and unpredictability of CSIR performance evaluation, the expert acknowledged both factors to be a challenge, but they should not stop organizations from building systems. To minimize the impact of these factors, the expert suggested focusing on defining goals and outcomes of the evaluation instead of focusing on the processes and activities which are complex and vary from one incident to another. The expert liked the fact that the framework presented in this project considers goals to be the highest design stage from which the other framework components are derived from. The expert noted though this might seem intuitive, it is not how many in the industry operate.

After being presented with the performance evaluation framework of this project, the expert made the following comments:

1- It is a practical and desirable that the framework is "selection-based", e.g. there are list of parameters, strategic issues and analysis methods to select and customize. This reflects the flexibility of the framework

2- The expert was reserved in describing the framework as "comprehensive", preferring to call it "holistic". The term 'comprehensive' seem to have a negative association in the industry to describe plans that are not-well defined that attempt to achieve too many things but have little benefit.

3- The structure and presentation of the framework is geared towards responders and technical teams, and need further ‘translation’ is needed if it is to be presented to executives and managers.

4- Noting that the industry is full of false and outdated performance metrics, the expert was pleased to see a systematized approach to the definition of PMs.

5- Out of the framework components, the list of performance indicators along with their definitions and interpretation guidelines, would be of most interest and immediate need to CSIRT practitioners.

6- CSIR performance systems should center around the plans (CSIRPs) more than the (CSIRTs), due to the dynamicity nature of teams (e.g. the team now is different than the team after few hours when the shift changes).

7- Although the NIST and CERT documents are a good place to start, most organizations end up having capabilities with features that drift from those outlined in the two documents. Therefore, the expert argued that there needs not be much emphasis put on compliance with these two documents.

8- The framework seems to address too many issues for a responder to read. It is suggested that the framework be broken into several concise documents each to be handed to professional organizations to review (e.g. auditing or cybersecurity companies) and publish in their online platforms.

The above feedback provides excellent points about how the industry would receive and interpret the proposed framework. Nevertheless, some of these points (e.g. Points 3, 7 and 8) are not feasible to apply considering that this project has some academic elements to the research topic. Finally, the expert hoped that such research direction would close the gap between business executives and technical responders through giving both parties clear outline of success measures.

5.2.4 Expert 3 Feedback

Expert 3 [15] has 20+ years of experience as a computer forensic examiner, instructor and manager for a large US federal law enforcement agency, with a focus on financial crimes and seizing/recovering digital evidence from a wide variety of data systems and media. The expert also has 13+ years in academia at the University level teaching computer forensics to undergraduate and graduate students. In addition, the expert has four years in the US Army in a military law enforcement and has published numerous whitepapers and been a frequent presenter in the computer forensic industry in the past 20+ years.

It was highlighted at the start of the interview that the expertise comes from incident response that involves criminal investigations. Incident responders in similar positions operate in environments that have standardized procedures to be followed, cost-free, and focus on evidence extraction, analysis and presentation for prosecution purposes. Overall, the expert described him/herself as a forensics responder noting that the field of digital forensics intersects with CSIR but has the advantage of being a decade older than CSIR.

The current performance evaluation practices were described by the expert as: “depends on the perspective of the team leader”, suggesting its subjective nature. The expert adds that with specific topic (i.e. CSIRPE) there is not much guidance in the industry about how to approach the topic. However, due to the publicity and liability issues related to cyber-incidents, the need for developing PE metrics is arising. The expert notes that

CEOs are now personally responsible for impacts of cyber-incidents, creating a higher-management need for conducting and maintaining evaluation systems.

Three challenges were highlighted. First, the field of CSIR, being a branch of computer security, highly depends on policies. Since policies are continually changing, it is expected that so will evaluation systems. Second, there is an administrative burden associated with performance evaluation. This burden is normally thrown on the technical team, who struggle to see the benefits of this “extra administrative task”. This creates resistance to the development and implementation of evaluation systems. Third, requirements and expectations vary depending on the jurisdiction. This makes developing unified performance models a difficult task.

When asked if the above three challenges makes developing CSIRPE models infeasible, the answer negative. The expert replied: “whenever there are standard procedures, measuring performance is possible”. Despite the field of CSIR having too many variations, there are essential procedures that are common across these environments.

An important observation was made by the expert with regards to a common short-coming of performance evaluation systems in the industry. Several companies conduct surveys, and also provide analysis. But, does this analysis gets translated into actions and enhancements? The expert finds this to be commonly lagging and foresees this as a challenge for CSIR performance evaluation. In the military and criminal investigations, this does not seem to be an issue, because the standard model of operation considers “post-action report” as an essential component of the review system. Referring back to the criminal investigation environment, the expert notes that making post-incident review is well-integrated in the process, despite lacking more guidance on the PE aspect.

The expert views the multidisciplinary survey of specific interest to DHS and similar government-like agencies. Incident response under these agencies involve several teams of different backgrounds. When such interaction occurs across teams, which may be collaborating in the same incident, creates a need to study how performance of these teams contributes to the overall performance of the response.

Speaking about indicators for good incident response performance, the expert highlighted the following five indicators:

1- *Effectiveness of Team assignment:* Was the right person, in terms of competency and experience, being assigned the right task during incident handling?

2- *Evidence Retention adequacy*: how is evidence stored and maintained and are there redundancy measures being taking?

3- *Effectiveness of training programs:* there is a need to continually attend new training programs. A cost-effective model would to be to send one member for professional training who in return would train the other members of the team.

4- *On-site analysis capacity:* how much of the response was conducted on-site compared to bringing machines to the office (i.e. off-site) for analysis.

5- *Information Sharing:* Were the lessons learnt from the incident shared with the internal and external communities?

According to the expert, a comprehensive CSIRPE should have two main features. The first is covering all the phases of the CSIR life cycle and second involving all parties, e.g. CSIRT members, managers, technical teams, support teams …etc. With regards to compatibility, the expert was satisfied with the selection of NIST and CERT, giving slight preference towards the NIST document. Nevertheless, the expert downplays strict observance to these two documents, because the industry of has tens of CSIR operating in a NIST-variant environments which are not necessarily what the original NIST document has outlined.

The final comment made by the expert was: “this is an important and growing field. If you have not done this type of work, someone else would have done it. It is going to take time to accept PE, but soon it will be mainstream”.

5.2.5 Expert 4 Feedback

Expert 4 [174] is a cybersecurity expert with extensive experience in the areas of system administration, networking, secure programming. He/she has eleven years of experience of working with digital forensics teams and contributes to research in the areas of information security and incident response. The expert is notable for his/her contributions in vulnerability patching across various operating systems platforms.

Throughout the interview, the expert voiced strong skepticism to the concept of CSIRT performance analysis. The expert believes that the process is infeasible, involves unnecessary overhead and have little practical benefits. Therefore, the interview focused on unwrapping factors which the expert’s skepticism is founded on.

According to the expert, there are three main reasons for objecting to the research theme of performance analysis:

1- Computer security incidents are very diverse and each incident has its own unique characteristics that make it stand in difference to other incidents. In other words, it is very rare to have two incidents which are identical. Even when incidents display similarities, their environments are different. Because of this diversity, incident response is also expected to be different. Consequently, there are two possible approaches to designing performance systems: either design unique measurement method that suits every incident, or design a generic system that is inapplicable to the spectrum of incidents. In either case, it is infeasible to design performance metrics that will allow comparisons or benchmarking. Since performance analysis is mainly based on trend analysis and benchmarking, it is infeasible to apply to CSIR.

2- Just as there is no standard method to find solutions to cyber security threats, there would be no standard method for measuring the performance of those who develop unique solutions to the incidents. The expert notes that the higher-level goals could be similar, but the underlying processes and procedures are different and most of the time non-systematic. Viewing incident response as a spectrum starting from higher-level management and goal-oriented tasks, to the more specific technical procedures, performance analysis seem to be applicable to the management side of the spectrum not the technical one. Since CSIR is mainly technical, then there is little to expect out of evaluating its performance.

3- Even when it is possible to design performance metrics that are applicable to a group of incidents, the expert thinks that it is unreasonable to infer performance aspects from analyzing the PM data. The expert questions: “why do we infer that a short response time reflects good performance?” A slower response time that produces more sustainable solutions could be better to an organization. Based on this simple example, the use of response time metric can be misleading to the organization. The same criticism is applied to performance metrics that are derived from cost.

Although the above expert feedback has no direct contributions to enhancing the proposed framework, I believe it is important to have proper understanding and serious discussion of the presented viewpoint. This viewpoint represent legitimate concerns that are shared with a sector of computer security professionals in the industry. Even among researchers, this is an extension of the on-going debate about the usefulness of introducing metrics to the field of computer security [34] [42].

To answer the first concern on the diverse nature of computer security incidents, it could be acknowledged that security incidents indeed pose some differences. However, the claim that such differences hurdle the development of performance systems could be disputed. The diversity argument seem to focus on the technical aspect of the response, while ignoring that CSIR is an organic body of several response activities both technical and non-technical. The management of the response, coordination between team members and with outside parties, analysis from several technical point of views, and assessing damage, cost and risk are all essential aspects of CSIR. For the sake of argument, let the technical solutions be unique, there still remain strong similarities in the non-technical side of incident handling that could be subjected to uniform methods of performance analysis.

The technical uniqueness of incidents could also be disputed. Even when incidents use innovative techniques to compromise, strong similarities could be found in the objectives, tactics and potential remedies. For instance, looking at incidents from the CIA model, each incident will target one or a combination of the three security properties: confidentiality, integrity and availability. It is difficult to argue against the fact that many incidents could be detected in similar mechanisms, which are technical in nature. For instance, computational hashing algorithms are used for verifying the integrity of data. The application of these algorithms in detecting a system compromise, which is a technical aspect of CSIR, could be subject to evaluation in terms of detection effectiveness. Another technical example is that most if not all incident responses involve some aspect of analyzing system and network logs. The effectiveness of the tools used in this process can also be subject to performance analysis.

A third example, which could also be used to address the second argument about the difficulty of systemizing CSIR technical aspects, comes from the current maturity state of the field of forensics analysis. Forensic professionals use standardizes methods to collect, preserve and analyze digital evidence. Most incident handling involve some involvement of the digital forensics which could be subject to performance analysis at least from the perspective of readiness and conformance.

Overall, it seems that some objections to CSIRPE can be rooted in having different or more narrow understanding of performance analysis, its methods and application domain. It is true that some technical aspects of the response might not be applicable to a large pool of incidents; however, this does not mean that this is true about all technical aspects of CSIR.

With regards to the third argument about the disassociation of PM measurement and performance effectiveness, the argument could be valid if PM readings are analyzed independently. However, the CSIRPE framework presented in this project ensures that performance metrics are only designed through a process that maps PMs to PIs back to goals. In addition, it was argued that PMs should be analyzed collectively when drawing conclusions about the performance. This means that PM readings are analyzed in the context of a higher structure of performance (goals→ indicators → metrics) and are inspected along with other metrics (e.g. the Cascading principle in the SAC model).

In its early stages, the CSIR field suffered, and it continues to be a challenge to some extent, to convince higher management of the need to establish a CSIR capability despite the strong acknowledgement and recommendation of the security professionals to establish and maintain such capability. The case seem to be reversed in the case of performance evaluation. As the interviews reflect, managers and executives seem to be interested enough in evaluating the effectiveness of the CSIR capability, while resistance may come from security professionals dealing with the technical aspects of the response. The fourth expert feedback is a sample of such expected resistance.

5.2.6 Summary of Expert Feedback

When analyzing the four expert interviews in a collective manner, several common themes and recommendations could be highlighted. Despite the interviews being limited in terms of number of interviewees, considering the long experience of the experts it could be argued that these common themes reflect the industry perspective, at least partially. The following points summarize points of agreement among at least three out of the four experts:

1- Interest in the topic of CSIRT performance evaluation is proportional to the rank of the individual associated with CSIR. More interest is expressed by those holding managerial positions compared to those conducting “lower-level” technical or support tasks. The interest is also proportional to the organization size where larger companies are more likely to adopt performance frameworks. Therefore, individuals with the highest interest in CSIRPE are those holding higher executive positions in environments that deploy a large CSIRT. The least interested group of individuals are technical responders operating in a small CSIRT.

2- Unexpectedly, the experts downplayed the importance of designing a CSIR performance framework that is strictly compatible with the CSIR industry standards. The experts acknowledged the importance of building on the NIST and CERT outlines, however, the vast variations that exist in the industry suggest that more generic methods, not necessarily NIST/CERT adherent, are more useful.

3- The plan, not the team, should be the focus of the performance evaluation of the CSIR. As more governments and businesses have working CSIRPs, CSIR teams can be viewed as experienced staff that execute a plan. This is contrary to the past decade experience in which CSIRT experts were viewed as the origin of plans.

4- The current status of PE indicators and metrics for CSIR can be described as scarce and disorganized. The experts concurred that the list of PIs provided in this dissertation is probably the most notable contribution of the project. The list is expected to be well received by the industry, as it is timely and fulfills an existing need.

5- The overall structure and process flow of the framework developed in this project was approved by the first three experts. The level of details provided for each step was satisfying. Comments of enhancement concentrated on simplifying the framework for business settings as some aspects were considered too advanced for the current industrial needs.

6- Building on the above point, there is a clear gap between how the industry and academia treat the topic of CSIR performance evaluation. The industry is demanding *simple* and *operational* models. On the other hand, academia seem to retain substantiate attention to rigorous models, validation schemes, and mathematical models (e.g. quantification); which are of little interest to the industry.

7- Despite the multidisciplinary survey being academic in nature, the experts viewed that direction of research very positively. The experts acknowledged the uniqueness of CSIR but disregarded the claim that it is too unique to borrow solutions from the other disciplines. The NIMS and NRF were specifically highlighted as immediate beneficiaries of the survey results.

8- Despite the challenges that face the evaluation of CSIR, the experts believe that these challenges are not strenuous enough to consider developing CSIRPE systems infeasible. However, the experts called for careful consideration of various factors and expected the process to take relatively long time before having operational models.

5.3 Evaluating Framework Features

5.3.1 Comprehensiveness

To the extent of the large number of works surveyed in this dissertation, there is no clear definition or outline of the characteristics of a “comprehensive” performance system. Therefore, to demonstrate the comprehensives of the proposed CSIRPE, the term “comprehensive” is analyzed over the common usages of the term in several resources like: [96] [7] [92] [79] [74].

It could be argued that the developed CSIRPE framework exhibits comprehensiveness from four angles: development, component, aspects and perspective.

Development:

The CSIRPE framework provides a complete process of development starting from the early stages of defining goals and strategies to the final steps of implementation and integration. This provides a comprehensive development guide to responders compared to a non-comprehensive framework that would introspect a partial system development.

Component:

The CSIPRE framework unrestrictedly targets the various components of the CSIR system. For instance, the framework encompasses the five-phases of the CSIR life cycle. A framework that focuses on the PE analysis of some phases, e.g. detection or containment could be viewed as non-comprehensive from that perspective.

Aspect

The CSIRPE framework does not limit performance to specific aspects, e.g. reliability or readiness. Instead, it views performance comprehensively by allowing for a holistic analysis that encompasses the various performance aspects.

Perspective

The CSIRPE framework views performance from the perspective of the plan (CSIRP) and the team (CSIRT) (see S.6: Reference Analysis Point). It also considers the viewpoints of various actors, like the organization (e.g. 3.4.2 Defining Performance Goals, S.2: Quality Control), the technical team (e.g. D.1. CSIRT Type, V.1. Operational Validation, 3.6.1 Functional Models) and the management (3.6.2 Assigning Roles and Responsibilities, V.3 Bootstrapping).

5.3.2 Flexibility

Arguing for flexibility will based on the understanding of the concept as in NIMS [109] and the discussion provided in PI.26: Flexibility. The CSIRPE arguably demonstrates flexibility from three angles: incident, design and environment.

Incident Flexibility

The CSIRPE framework could be used to evaluate responses to different incidents, indifferent to their type, scope and severity. Unlike a performance system that is focused on a specific type of incidents, e.g. DoS incidents, the framework focuses on analyzing responses regardless of the nature of incident which gives it a flexible domain of applicability. Note that the restrictions made in A1.1 (Civilian Environment), A1.3 (Incident Complexity), and A2.5 (Sequential Handling) have little practical implications as they avoid special situations that are unlikely to occur in conventional CSIR.

Design Flexibility

The framework grant PE designers a large domain of design possibilities and options. For instance, the selection of PIs and analysis techniques is left to the designers based on its defined goals. Another example is the IAV model (Section 5.2) which allows a CSIRT to select performance analysis methods that suit their capacity and environment. The design parameters and strategies also provide a pool of options for each CSIR to customize to its needs.

Environment Flexibility

The framework is compatible for use with various CSIR environments, regardless of the CSIRT type and capacity (as long as there is a team of responders, see A1.2). It also does not presume the presence of specific policies/procedures (other than A1.7) or the need for having a pre-established quality control unit. This means that a CSIRT can use the framework to construct different PE models for different environments.

5.3.3 Compatibility

The CSIRPE framework is constructed in a manner that is in compliance with the NIST [1] and CERT [2] documents, which are the two most commonly used CSIR industry standards. Since both documents do not address how performance evaluation should be administered, compliance here means that the CSIRPE framework is not in conflict with any of the requirements set by them. In numerous occasions, the CSIRPE framework would address direct requirements set by either document.

Below is a set of two examples, each containing eight examples, that demonstrates computability with the NIST and CERT documents.

Compatibility with NIST

1- The definition of events and incidents (Section 1.4.1) is consistent with the definitions provided in NIST (page 6)

2- The NIST document defines eight policy elements of any CSIR. The seventh element is “Performance Measures” (page 8). The document does not outline procedures for how that could be achieved. The CSIRPE framework can be viewed as a guideline of how to design and implement this seventh policy element.

3- The three selections for the design parameter: *D.1 CSIRT Type* is based on three types defined by NIST (page 13): centralized, distributed and coordinated. The term “customized” was used instead of “coordinated”, but was defined to included coordinated in addition to other structures.

4- The NIST recommends the use of both objective and subjective assessment of incidents (page 40-41). This is adopted in the CSIRPE by allowing for both quantitative and qualitative assessments (D.7: Measurement Type, and M.3: Quantifiability)

5- Assumption A1.4 on Incident Handling Services is borrowed from an assumption made by the NIST (page 23) which focuses the functionality of CSIRTs to incident handling compared to ensuring system security.

6- The definitions and considerations set by the NIST were used in the definition of several performance indicators like: PI.3, PI.16 , PI.21, PI.25 PI.31.

7- The use of the performance evaluation database (PED) in the IAV is inspired by the recommendation set by NIST (page 3) on: "maintain and use a knowledge base of information"

8- The strategy of using automation as a mechanism for decreasing overhead (S.5: CSIPRE Overhead) in incident handling is highlighted by NIST (page 51) as a recommendation.

Compatibility with CERT

1- Section 2.6 of the CERT document on Quality Assurance (page 42) is the category at which the CSIRPE framework would fall. In this section, there is a call for setting quality requirements and measurement mechanisms, which is what the CSIRPE attempts to achieve. A similar reference to quality is also made in page 22 when depicting the CSIRT service and quality framework.

2- Assumption A2.2 (Member Dedication) is consistent with CERT (page 18) which assigns incident handling responsibilities to CSIRT members with little overlap with the other non-incident handling security tasks.

3- The definitions and considerations set by the CERT were used in the definition of several performance indicators like: PI.17, PI.18 , PI.26, PI.32 and PI.38.

4- The first theme of the three suggested PE objectives (see Section 3.4.2) on business continuity and growth is derived from the emphasis put in the CERT document for aligning CSIRT goals with the organizations business continuity plan (page 33).

5- The element of simplicity for performance evaluation systems as highlighted in the SAC Complexity Model (Section 5.4) is explicitly highlighted in the CERT document in page 46.

6- The stakeholder analysis as presented in the CSIR Balanced Scorecard Model (Section 5.3) is consistent with the CERT document call for "Constituents' view on Quality" (page 48).

7- The flexibility element of the NFP Unpredictability Platform (Section 5.5) was designed in accordance to the requirements set by section 2.7.1 of the CERT document (page 50) on "The Need for Flexibility".

8- Both the trend analysis and targeted analysis (Section 3.5.2) are referenced in the CERT document in pages 26 and 81 respectively.

CHAPTER SEVEN

CONCLUSION

6.1 Summary of Contributions

Contributions made by this project can be summarized in the following six points:

First, the project produced an end-to-end process for developing performance evaluation frameworks for computer security response systems. Although the process is designed for the field of CSIR, the general structure and basic details of the development process are generic and could be applied to various incident response fields.

Second, the study laid out a map of the various issues associated with the study of performance evaluations of CSIR. These issues are presented in the form of design parameters, strategies and challenges. The study does not claim to provide a solution for these issues. Instead, the study identifies these issues, provides proper description, summarizes work done towards addressing that issue and proposes basic recommendations that are compatible with CSIR environments.

Third, the project presented a multidisciplinary survey of how various incident response disciplines addressed the problem of team and plan performance evaluation. The value of this survey to the field of CSIR is equipping researchers with lessons learnt from these fields to avoid reinventing the wheel when constructing CSIRPE models. The survey provides evidence that incident response disciplines have much in common in terms of challenges, processes and measurement techniques. This calls for further advancement in that direction of research.

Fourth, the project presented a framework for evaluating CSIR capabilities. The framework can be viewed as a fusion pot of tens of models, techniques and findings of performance measurements. The framework took the skeleton of CSIR life cycle and activities and embarked it with tools that could be used for evaluation. These evaluation tools neither violate the basic CSIR processes nor confine the usage of tools. This design flexibility allows for customization which suits the diverse status quo of CSIR implementations.

Fifth, a list of fifty performance indicators for assessing computer security incident response capabilities was formulated. The list covers the assessment of the major parts of the response system and also the overall system performance. Each indicator was defined and associated with interpretation considerations and examples of potential derived performance metrics. The list gives CSIR practitioners a large repository of PIs to choose from, and saves the effort of defining them and laying out the design considerations.

Sixth, the study made the first step in paving the way towards the implementation of performance evaluation models in CSIR environments. Currently, there are no implemented systemized approaches to CSIRPE and it was infeasible to deploy the proposed framework in real environments. However, the study analyzed a wide list of implementation considerations and made close contact with the industry through integrating the expert feedback into the design and analysis of the framework and through proposing several hypothetical scenario analysis that envision how the framework would function when it is deployed.

6.2 Summary of Findings

The major findings of studying the development of frameworks for performance evaluation of computer security incident response in this project are summarized in the following ten points:

First, the discipline of CSIR had undergone several advancements in the past two decades. It could be argued that the field has reached a satisfying maturity in terms of defining processes and procedures. The technical development of breach preventive and countermeasures is a continuous process, but it is happening now within well-established CSIR capabilities and through a wide network of national and international information sharing.

Second, inspecting the evolvement of the CSIR discipline, it is not surprising that attention to systemized approaches of performance evaluations has only started to flourish. This is supported by several external factors that pushes the industry to adopt organized methods of performance analysis, like liability, insurance and SLA contracting. There are enough indicators that the industry acknowledges the need for conducting CSIRPE. Examples include, the explicit call by the NIST document for developing PE measures and the recent abrupt increase in the number of publications, blogs and white-papers that address issues of effectiveness and PE measurement. However, the urgency and depth of this need may vary depending on the maturity levels of various CSIR capabilities.

Third, discussion so far about CSIR effectiveness could be characterized by being too generic, non-systemized and have narrow scope of performance domain. It was also noticed that the discussion was concentrated on the preparation phase especially in evaluating the effectiveness of a CSIRP. This study recognized the need for having systemized approaches to CSIRPE that provide holistic and focused performance analysis of the CSIR system, be equipped with performance measurement tools, and establishes balance between pre-incident and post-incident performance evaluation.

Fourth, the multidisciplinary survey of performance evaluation systems for various incident response systems reveals that there are many commonalities across disciplines. Despite these similarities, there are little efforts to share experiences across these fields and attempts to build common models of analysis remain scarce. The perception that incident response disciplines are very distinct from each other need to be challenged, and more research should be directed towards understanding the similarities and differences between various IR systems.

Fifth, the complexity of incidents, the large variance in threat techniques, and the unpredictability nature of the response are the main challenges that face responders. This suggest that evaluating the performance of the response require PE systems that are sensitive to the uncertainty nature of incident handling, demonstrate reasonable simplification and are flexible enough to allow application across the domain of incident types and severity levels.

Sixth, the field of security metrics suffers from lack of validation, debatable practical benefit, and unreasonable quantification techniques. Since CSIR performance metrics are built on security metrics, these challenges are passed to the development of performance metrics. Also, despite the several attempts across disciplines to quantify performance measurements, the vast majority of metrics remain qualitative, with the exception of financial metrics which are relatively easier to quantify. Based on this, overcoming the subjectivity nature of qualitative measures is another consideration for the design of CSIRPE measures.

Seventh, the development of performance evaluation systems for CSIR can be broken into four higher-level phases. The first phase defines the generic framework of evaluation by setting the design parameters and outlining the strategies of evaluation. The second phase defines the core of the PE model by defining goals performance aspects and derive the approporiate performance indicators. The third phase focuses on specifying the measurement mechanisms by deriving performance meterics and selecting analysis and validation techniques. Finally, the fourth phase inspects the operational characteristics of the CSIR system and integrates the PE model with the implementation requirements of the environment.

Eighth, developing the framework for evaluating CSIR suggests that CSIRPE is both a science and an art. It is scientific from the perspective of following systemazed approaches and applying obective performance analysis techniques. On the other hand, the artistic aspect is manifest in how the PE system is developed and configured and how PIs are selected and used.

Nineth, there is a tension between the organizational and techincal needs that drive performance evaluations. Organizations are interested in more cost-effective responses that ultimately contribute to better return-on-investment. However, techincal teams are interested in enhancing the detection, containment and recovery processes. Any CSIRPE system need to address this tension by striking a balance between these two domains. This could be achieved by insuring that the CSIRPE is integrated into the organizational qualtiy system but at the same time is developed and maintained by the responders and techincal staff.

Tenth, there is valid skepticism on how much technical benefits are expected out of the deployment of PE systems. Although there are sevearl theoritical projections of the expected outcomes, the actual level of impact on system enhancement would only be known after getting industrial feedback. Therefore, there is a need to monitor and benchmark results of CSIRPE systems in the coming few years.

6.3 Future Work

Extensions of this work can be forked into two directions. The first direction emerges from the design and findings presented in this work, while the second pertain to the general development of the topic of performance evaluation of computer security incident response.

Under the first research direction, there are four main direct extensions of this work, namely the implementation considerations, the development of performance metrics, integration with the VERIS platform and investigating the enhancement procedures.

Perhaps, the most stressing work awaiting the study of CSIRPE is implementing CSIRPE on real-environments. Such implementation would indicate how the industry reacts to CSIRPE measurement in an operational manner. Specifically, there is a need to observe which PIs and which performance analysis techniques produce benefits to CSIR capabilities. If the framework presented in this project is deployed, the following industrial feedback is expected:

1- The industry will demonstrate how the design parameters are commonly set. This leads to refinement of the options available to each parameter, possibly eliminating some parameters and adding others.

2- Testing the development process will reveal which development phase is considered most challenging, which in return will stimulate the attention of researchers.

3- The industry will determine which performance indicators are more valuable to the assessment and enhancement of CSIR.

4- Investigating the actual extent of the variance of performance measurements for the purposes of building common benchmarks and formulating best practices.

The second extension relates to the development of performance metrics. As this work is focused on the development of performance indicators, there is a need to investigate each of these indicators in terms of deriving performance metrics. The PM examples presented in Section 4.3 were provided for elaboration purposes that serve the understanding of the PIs. For each PM, there needs to be more specific and formal definitions that would enable responders to use it and customize it. This opens a wide area for research work, as each PI can be subjected to a dedicated research work in terms of deriving correct and useful PMs.

The third extension relates to the VERIS platform. The current structure and flexibility of the VERIS platform opens the door for potential integration of the CSIRPE framework to the incident definition and sharing. There are three possibilities for such integration. The first, under the block of "Response & Discovery" which consists of five sub-blocks: {incident timeline, discovery method, root causes, corrective actions, targeted vs. opportunistic}, a sixth sub-block could be added by the name: "response performance". The second possibility is under the block of "Indicators" which focuses mainly on indicators of compromise, an additional category could be added under the title of “performance indicator”. The third possibility, under the block of “Impact Assessment” which consists of: {loss categorization, loss estimation, estimation currency and impact rating} a category could be added by the title: “response effectiveness”.

The fourth extension targets a common problem in performance systems that is commonly known as “closing the loop”. A good performance model is capable of producing good assessment and recommendations. However, this would be ineffective unless there is a mechanism to ensure that these recommendations are correctly fed-back to the system with enforcement policies and progress monitoring tools. In other words, mechanisms for quality enhancement based on CSIRPE outcomes need to be carefully studied.

From the general perspective of CSIR performance evaluation, there are several research topics that are yet to be explored. First, there is more work needed to integrate the preparation and evaluation mechanisms of CSIR with the NIMS [109] and NRF [110] frameworks. Although both frameworks integrate cybersecurity incidents, this integration seems primitive and sometimes inconsistent with CSIR processes [14].

Second, there is a scant of works that explore the notions of complexity and unpredictability in CSIR. As this project demonstrated, both notions pose major challenges to the discipline of CSIR and IR disciplines in general. The areas of complex systems, non-deterministic decision making and fuzzy performance systems are examples of disciplines that could significantly enrich the understanding of these two notions under CSIR environments.

Third, there is only limited number of works that address the distributed model for operating CSIRTs. Once more refined understanding is available on the how the CSIR distributive processes are designed and executed, only then the development of performance models is feasible.

Fourth, there is little work done on creating simulation exercises for CSIRTs. The current training programs focus on the technical aspect of CSIR, while real environments have high interaction between several technical and non-technical teams. The presence of simulation exercises will work for the advantage of building performance evaluation techniques. Assessment methods and mechanisms could be tested under simulated environments providing useful predictive expectations of how these methods would impact the functionality of CSIRTs.

Finally, there seems to be much variance on how the industry implements and uses CSIR. In addition, the industry is continually updating its CSIR procedures in response to the evolvement of technologies. For instance, response to incidents that involve cloud computing is slowly developing its own distinctive response techniques. Such update to the operations of CSIRTs would consequently require an update to how evaluation techniques are to be used. At some point, there needs to be a study that identify the common performance issues across CSIRTs and the distinct performance aspects displayed by the various types of CSIRTs.

APPENDICES

Appendix A: Practical Guidelines

A.1. Recommended CSIRPE Configuration

The following table provides suggested steps for how to configure a CSIRPE based on the maturity level of the security system [220], especially the CSIR capability. These configurations provide a starting point for how a CSIRT could configure its PE model in terms of expectations and capacity. However, this needs to be further developed to match the objectives of the team and the available capacity.

Configuration		CSIR Environment		
Code	Name	Basic	Medium	Advanced
D.1	CSIRT Type	Centralized	Centralized or Distributed	Customized
D.2	Evaluator Type	CSIRT	Internal	Internal and External
D.3	Number of Incidents	Single-incident	Multiple-incident	Adjustable-window
D.4	Incident Concurrency	Sequential	Sequential	Concurrent or Elastic
D.5	Analysis Time	Post-incident	Post-incident	Continuous or Incremental
D.6	Benchmarking	None or standalone	Internal	External
D.7	Measurement Type	Mainly Qualitative	Mixed	Mainly Quantitative
D.8	CSIR Scope	Execution	Mainly execution + basic design	Both execution and design

S.1	Quality Assurance	None	Compliance	Compliance & Auditing
S.2	Quality Control	CSIRT	Mainly CSIRT + Quality Unit	CSIRT + Quality Unit
S.6	Reference Point	CSIRT or CSIRP	CSIRP	CSIRP
PI.x	Performance Indicators	Few KPIs	KPIs + other PIs	KPIs and several PIs
N.x	Analysis Methods	Comparative	Comparative + basic deficiency & holistic	Comparative + deficiency + holistic + predictive
V.x	Validation Methods	Operational	Operational and Development	Operational + development + heuristic

Table 45: Recommended CSIRPE configuration based on maturity level

A.2. Guidelines for Designing Adjustable-Window Framework

#	Step	Description
1	Assess the need for an inclusive PE framework	Does your organization really need a comprehensive PE framework? If not, resources will be wasted in planning without any valuable return. If only periodic reporting is needed consider multiple-incident frameworks. If special task force or for directed customers consider single-incident frameworks.
2	Re-define your goals	Create two lists of PE objectives: short-term and long-term. The short-term objectives should be used to derive PMs for single incidents, while long-term goals should be used to derive PMs for multiple-incidents.
3	Increase number of descriptive metrics	Since descriptive metrics will be used in analysis of both single and multiple incidents, increase its pool, increase the pool of these metrics to enable designing a variety of PMs. A reading that seems not useful now might be useful later.

4	Create two lists of PMs for single-incident analysis	In order to analyze incidents individually, prepare two lists of PMs. The first list contains PMs that always need to be used, while the second list contains optional or special-case PMs. This strikes a balance between uniformity of analysis across incidents but also allows for flexible analysis
5	Describe PE setting	For each incident describe the environment and setting in which PM readings were collected and analyzed. This is important to make meaningful comparison between incidents and to ensure that statistical measures are consistent.
6	Define context for interpretation of PMs	For multiple-incident PMs specify the proper conditions for reading the results and if there is a need to analyze multiple readings collectively
7	Conduct two-level analysis	Do not mix single-incident and multiple-incident analysis. First, analyze the incident individually then analyze the incident over the larger spectrum. Record results from both steps, and present results as needed

Table 46: Guidelines for designing an inclusive PE framework

A.3. Guidelines for Selecting Analysis Reference Point

The following simple questionnaire is a tool for CSIRT leadership determine which analysis reference point to use. Having more answers of “Yes” than “No” hint that it is probably better to select the CSIRP as the analysis reference point compared to the CSIRT, and the reverse is true.

#	Question	Answer	
1	Is your CSIRP applicable to several locations compared to a single geographical location?	Yes	No
2	Is the size of your CSIRT composed of three or more members?	Yes	No
3	Did your organization invest enough financial resources in the preparation of its CSIRP?	Yes	No

4	Does the CSIRT has good understanding of the implemented security policies and the overall security of the system?	Yes	No
5	Does the CSIRP demonstrates a good level of completeness and comprehensiveness?	Yes	No
6	Is the likelihood of inviting an external CSIRT to handle an incident low?	Yes	No
7	Is it expected that the team performance will rely on the received training compared to their previous expertise?	Yes	No

Table 47: Assessment questionnaire for selecting analysis reference point strategy

A.4. Guidelines for Identifying CSIR Performance Aspects

#	Key Term	Explanation
1	Selection of Methodology	Before identifying the aspects, think which of the methodologies best fit the nature of the CSIRT/CSIRP of the organization
2	Multiple Methodologies	Select at least two methodologies such that the combination is both CSIRP and CSIRT centric.
3	Parallel Analysis	Perform separate analysis using each methodology. Next, find which one should take precedence and combine the aspects
4	Use of Nouns	Use nouns compared to action verbs when phrasing the aspects
5	Confined List	The list of aspects should not be too short or too long. The number of aspects can be between five and ten aspects. Too little or too many aspects is likely to cause difficulties in the following steps for performance indicators and metrics derivations
6	Overall Performance	Remember to define the “overall system performance” as an aspect.

Table 48: Recommendations for Identifying PE Aspects

A.5. Templates for Defining Performance Indicators and Metrics

The following two tables provide templates for formal definition of performance indicators and metrics. The templates cover the basic requirements for defining the PIs and PMs in addition to other supplementary optional fields.

Name	Term or a brief phrase that identify the PI
Code	[Optional] code to be used if automation is used
Description	A definition of the PI and how it reflects good performance
Classification	Generic vs. Specific (or using another categorization scheme)
Priority	KPI or PI (or another priority scale)
Goals & Aspects	Mapping to the goals and aspects in which this PI represent
PMs	Performance metrics that will be used to measure the PI

Table 49: Template for Formal Definition of Performance Indicators

Metric Name	Term or a brief phrase that identify the PM
Code	[Optional] code to be used if automation is used
Description	A definition of the PM (e.g. formula)
Classification	Performance vs. descriptive metric Generic vs. specific
Measurement Method	Quantitative vs. qualitative Measurement Tool
[Conditions]	Conditions for correct collection of measurement
[Attributes]	Factors that impact interpretation Range of allowed values Accuracy/confidence
Interpretation	How do the PM reading reflect performance?
PIs	Mapping to the Performance Indicators that use this PM

Table 50: Template for Formal Definition of Performance Metrics

A.6. Basic IAV Model Design

When a CSIRT is interested in using a basic design for the IAV model, then it could use the basic skeleton presented in Figure 33. The only necessary components are the feedback system, a database for storing incident data and performance results, and a single analysis technique. The IAV system can be developed later by the addition of various modules as the maturity level of the CSIR capability improves or as needs arise.

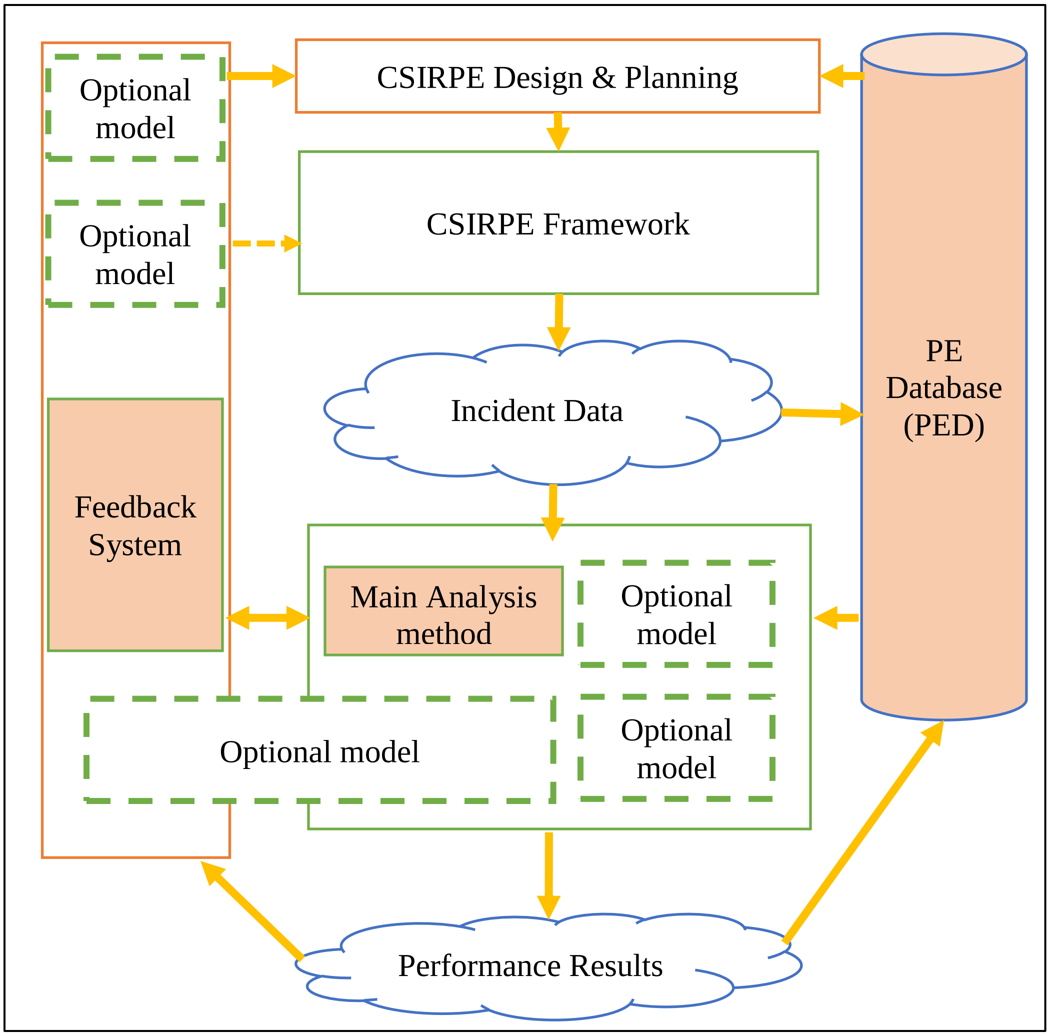


Figure 33: Basic Design of the IAV model

A.7. Guidelines for Selection of Performance Analysis Methods

#	Guideline	Description
1	Diversification	Diversify measurement and analysis techniques to broaden the performance evaluation. As stated by [11]: "there can be no single assessment of accomplishments overall".
2	Avoid Fox Paradox	Avoid the Fox paradox [137] by making sure to analyze system components first and then making an overall PE analysis.
3	Compatibility	Select analysis techniques that are compatible with the selected performance indicators and derived performance metrics.
4	Feasibility	Select analysis techniques that are feasible to apply taking in mind the availability of tools/software, competency of team members and expected cost
5	Future-Centric	Performance analysis should focus on the future, i.e. what to enhance, more than the past, i.e. what went wrong. Solving a past problem that does not have any future enhancements should be ignored. This is especially relevant as technologies advance in quick pace.
6	Target-Based	Predefine targets, thresholds, standards or benchmarks during the preparation phase. Inspect which analysis techniques are suitable to measure performance against these predefined targets
7	Validation	After selecting both analysis and validation techniques, investigate if analysis and validation could be combined together in order to save analysis overhead

Table 51: Guidelines for Selecting Performance Analysis Techniques

A.8. Scenario II Data

PM Name	Code	Brief Description
Declaration Time	DT	The timestamp of the incident declaration
Confirmed Time	CT	The timestamp of when the severity level of an incident is confirmed
Resolution Time	IRT	The timestamp an incident is declared resolved

Incident Severity	S	The initial Incident Severity level upon incident declaration
Actual Incident Severity	AS	Actual incident severity level as confirmed during or post incident handling
CSIRT Salary	CSA	Annual Salary of CSIRT core team
Number of Incidents	ANI	Total annual number of incidents (per severity level)

Table 52: Scenario II: List of Descriptive Metrics

	Degree 5	Degree 4	Degree 3	Degree 2	Degree 1	**Total**
2016	9	18	29	38	58	152
2015	6	14	25	31	51	127
2014	4	9	18	31	39	101
2013	1	7	9	18	42	77
2012	2	5	16	5	25	53
Total	22	53	97	123	215	

Table 53: Scenario II: Data for Annual Number of Incidents (ANI)PM

	Average annual responder salary	Annual CSIRT Total Salary (CSA)
2016	107,000	535,000
2015	103,790	518,950
2014	100,676	503,382
2013	97,656	488,280
2012	94,726	473,632

Table 54: Scenario II: Incident Response Fixed Costs(USD) [2012-2016]

Hourly Average incident Operating Cost in USD (HAOC) [2012-2016]					
	Degree 5	Degree 4	Degree 3	Degree 2	Degree 1
Core CSIRT	440	360	315	270	250
Support Teams	390	340	290	240	180
Logistic Teams	230	180	165	120	120
Total	1,060	880	770	630	550

Table 55: Scenario II: Average Incident Response Operation Costs (USD)

Average Financial loss per incident in USD [2012-2016]					
	Degree 5	Degree 4	Degree 3	Degree 2	Degree 1
Average Incident Loss (AL) (after CSIRT Intervention)	80,000	52,000	30,000	18,000	1,500
Predicted loss (PL) (without CSIRT intervention)	365,000	273,000	191,000	103,000	56,000
Average CSIRT Cost Effectiveness (CCE)	285,000	221,000	161,000	85,000	54,500

Table 56: Scenario II: CSIRT Cost Effectiveness Data

Average Incident Response Activity Time in hours [2012-2016]					
	Degree 5	Degree 4	Degree 3	Degree 2	Degree 1
Average Time to confirm severity level (TCIC)	6.2	5.8	5.5	4.9	4.2
Average Response Time (ART)	32	28	27	19	14

Table 57: Scenario II: Average Incident Response Time

Detection Inaccuracy Cost (Total cost in USD)						
	Degree 5	Degree 4	Degree 3	Degree 2	Degree 1	**Total**
2016	0	18,792	46,255	80,066	124,236	269,349
2015	0	14,616	39,875	65,317	109,242	229,050
2014	0	9,396	28,710	65,317	83,538	186,961
2013	0	7,308	14,355	37,926	89,964	149,553
2012	0	5,220	25,520	10,535	53,550	94,825
Total	0	55,332	154,715	259,161	460,530	

Table 58: Scenario II: Detection Inaccuracy Cost Performance Metric Data

Incident Classification Strategy Cost (ICS-CE)				
	Total detection inaccuracy Cost	Risk cost	Additional detection cost	Strategy Cost Effectiveness
2016	269,349	494,550	241,680	466,881
2015	229,050	388,875	201,930	361,755
2014	186,961	291,675	160,590	265,304
2013	149,553	165,150	122,430	138,027
2012	94,825	154,125	84,270	143,570

Table 59: Scenario II: Incident Classification Strategy Cost Effectiveness PM Data

BIBLIOGRAPHY

[1] P. Cichonski, T. Milar, T. Grance and K. Scarfone, "Computer Security Incident Handling Guide," National Institute of Standards and Technology, U.S. Department of Commerce. , August 2012.

[2] M. J. West-Brown, D. Stikvoort, K.-P. Kossakowski, G. Kilcrece, R. Ruefle and M. Zajicerk, "Handbook for Computer Security Incident Response Teams (CSIRTs)," Software Engineering Institute, Carnegie Mellon University, April 2003.

[3] N. McCarthy, The Computer Incident Response Planning Handbook: Executable Plans for Protecting Information at Risk, New York, US: McGraw-Hill, 2012.

[4] Ponemon Institute, "2016 Cost of Data Breach Study: Global Study," Benchmark research sponsored by IBM, June 2016.

[5] D. Munro, "Data Breaches in Healthcare Totaled Over 112 Million Records in 2015," *Forbes,* 31 December 2015.

[6] S. Chaudhuri, "Cost of Replacing Credit Cards After Target Breach Estimated at $200 Million," *The Wall Street Journal,* 14 February 2014.

[7] Expert_One, Interviewee, *Interview with CSIR police responder.* [Interview]. 29 June 2017.

[8] A. Torres, "Incident Resposne: How to Fight Back," SANS Institute, August 2014.

[9] T. Bailey, J. Brandley and J. Kaplan, "How good is your cyberincident response plan?," *McKinsey on Business Technology,* no. 31, pp. 16-23, Fall 2013.

[10] R. Behn, "Why Measure Performance? Different Purposes Require Different Measures," *Public Adminstration Review,* vol. 63, no. 5, p. 586–606, September 2003.

[11] O. Serrat, "The Perlis of Performance Measurement," Asian Development Bank, Washington, DC., 2010.

[12] J. Luttgens, M. Pepe and K. Mandia, Incident Response & Computer Forensics, Third Edition ed., McGraw-Hill Education, August 2014.

[13] R. Swanson, Analysis for Improving Performance: Tools for Diagnosing Organizations and Documenting Workplace Expertise, 2nd Edition ed., Berrett-Koehler Publishers, February 2007.

[14] Bennett Jones, "Cybersecurity 2017 Report and 2016 Reflections: What Businesses and Boards Need to Know," Bennett Jones, February, 2017.

[15] Expert_Three, Interviewee, *Interview with Senior Forensics Responder.* [Interview]. 20 August 2017.

[16] FIRST, "Forum of Incident Response and Security Teams," [Online]. Available: https://www.first.org/about. [Accessed 11 January 2017].

[17] H. Khurana, J. Basney, M. Bakht, M. Freemon, V. Welch and R. Butler, "Palantir: A Framework for Collaborative Incident Response and Investigation," in *Proceedings of the 8th Symposium on Identity and Trust on the Internet*, Gaithersburg, Maryland, USA, April 2009.

[18] Y. Peng, Y. Zhang, Y. Tang and S. Li, "An incident information management framework based on data integration, data mining, and multi-criteria decision making," *Decision Support Systems,* vol. 51, no. 2, pp. 316-327, May 2011.

[19] C. Alberts, A. Dorofee, R. Ruefle and M. Zajicerk, "An Introduction to the Mission Risk Diagnostic for Incident Management Capabilities (MRD-IMC)," Software Engineering Institute, Carnegie Mellon University, Hanscom, MA, May 2014.

[20] EPA, "Performance Indicators for EPA Emergency Response and Removal Actions," US Environmental Protection Agency (EPA), July 2008.

[21] PMIG, CAM-I, "The Performance Management Maturity Framework (Emerging Issue Paper (EIP))," Performance Management Interest Group (PMIG), The Society of Management Accountants of Canada and Consortium for Advanced Management-International (CAM-I), 2010.

[22] T. Reed, R. G. Abbott, B. Anderson and C. Forsythe, "Simulation of Workflow and Threat Characteristics for Cyber Security Incident Response Teams," in *Human Factors and Ergonomics Society*, October 2014.

[23] M. B. Line, E. Albrechtsen, S. O. Johnsen, O. H. Longva and S. Hillen, "Monitoring of incident response management performance," in *International Conference on IT-Incident Management and IT-Forensics (IMF 2006)*, Stuttgart, 2006.

[24] CNSS, "Committee on National Security Systems (CNSS) Glossary," Committee on National Security Systems (CNSS), National Security Agency, Fort Meade, Maryland, April 6, 2015.

[25] T. Carlson, "Information Security Management: Understanding ISO 17799," International Network Services, 2003.

[26] M. A. Javaid, "Benchmarks for Setting Up CERT," SSRN, September 10, 2013.

[27] R. V. Solms and J. V. Niekerk, "From Information Security to Cyber Security," *Journal of Computers and Security,* vol. 38, pp. 97-102, October 2013.

[28] ISO/IEC, "ISO 2703: Information Technology - security techniques - guidelines for cybersecurity.," ISO/IEC, Geneva, Switzerland, 2012.

[29] J. Creasey and I. Glover, "Cyber Security Incident Response Guide, Version 1," CREST, 2013.

[30] P. Kral, "The Incident Handlers Handbook," SANS Institute, 2011.

[31] R. Ruefle, K. v. Wyk and L. Tosic, "New Zealand Security Incident Management Guide for Computer Security Incident Response TEams," New Zealand National Cyber Security Center, Government Communication Security Bureau, May 2013.

[32] H. J. Schumacher and S. Ghosh, "A fundamental framework for network security," *Journal of Network and Computer Applications,* vol. 20, no. 3, pp. 305-322, 1997.

[33] C. Wang and W. Wulf, "Twoards a framework for security measurment," *Logistics Information Management,* vol. 15, no. 5/6, pp. 414-422, 2002.

[34] V. Verendel, "Quantified Security Is a Weak Hypothesis," in *Proceedings of NSPW'09 New Security Paradigms Workshop*, Oxford, UK, September 2009.

[35] S. Maynard and A. Ruighaver, "What makes a good information security policy: a preliminary framework for evaluating security policy quality," in *Proceedings of the fifth annual security conference*, Las Vegas, Nevada USA, 2006.

[36] Cambridge Dictionary Online, "Performance," Cambridge University Press, [Online]. Available: http://dictionary.cambridge.org/dictionary/english/performance. [Accessed 05 May 2017].

[37] Business Online Dictionary, "Performance," [Online]. Available: http://www.businessdictionary.com/definition/performance.html. [Accessed 5 May 2017].

[38] D. Grote, The Performance Appraisal Question and Answer Book: A survival Guide for Managers, New York, USA: American Management Association (AMACOM), 2002.

[39] D. Gharakhani, H. RAhmati, M. R. Farrokhi and A. Farahamandian, "Total Quality Management and Organizational Performance," *American Journal of Industrial Engineering,* vol. 1, no. 3, pp. 46-50, 2013.

[40] A. Neely, M. Gregory and K. Platts, "Perofrmance Measurement System Design: A literature Review and Research Agenda," *International Journal of Operations and Production Management,* vol. 25, no. 12, pp. 1228-1263, 2005.

[41] K. Stewart, J. Allen, M. Valdez and L. Young, "Measuring What Matters Workshop Report," Software Engineering Institute, CERT., 2015.

[42] J. Zalewski, S. Drager, W. McKeever and A. J. Kornecki, "Measuring Security: A Challenge for the Generation," in *Position papers of the 2014 Federated Conference on Computer Science and Information Systems*, Warsaw, Poland, 2014.

[43] ISO/IEC, "Quality management systems — Fundamentals and vocabulary," International Organization for Standardization, Geneve, 2015.

[44] S. Singh, The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography, Anchor, Repring edition, August 2000.

[45] K. M. M. de Leeuw and J. Bergstra, The History of Information Security: A Comprehensive Handbook, First ed., Elsevier Science, November 2007.

[46] S. Levy, Hackers: Heroes of the Computer Revolution. 25th Anniversary Edition, 1st ed., O'Reilly Media, May 2010.

[47] P. Szor, The Art of Computer Virus Research and Defense, Addison-Wesley Professional, February 2005.

[48] C. Easttom and J. Taylor, Comptuer Crime, Investigation, and the Law, 1st ed., Course Technology PTR, 2010.

[49] M. Kabay, "A brief History of Computer Crime," M.E. Kabay, 2008.

[50] P. Shakarian, J. Shakarian and A. Ruef , Introduction to Cyber-Warfare: A Multidisciplinary Approach, Waltham, MA: Elsevier, Syngress, 2013.

[51] I. Skierka, R. Morgus, M. Hohmann and T. Maurer, *CSIRT Basics for Policy-Makers: The History, Types & Culture of Copmuputer Security Incident Response Teams,* May 2015.

[52] R. Ruefle, A. Dorofee, D. Mundie, A. D. Householder, M. Murray and S. J. Perl, "Computer Security Incident Response Team Development and Evolution," *IEEE Security and Privacy,* vol. 12, no. 5, pp. 16-26, October 2014.

[53] H. Orman, "The Morris worm: a fifteen-year perspective," *IEEE Security & Privacy,* vol. 99, no. 5, pp. 35-43, 14 October 2003.

[54] E. H. Spafford, "Crisis and aftermath," *Communications of the ACM Magazine,* vol. 32, no. 6, pp. 678-687, June 1989.

[55] CERT-CC, "The CERT Coordination Center FAQ," CERT Coordination Center(CERT-CC), Software Engineering Institute, Carnegie Mellon, [Online]. Available: http://www.cert.org/faq/index.cfm. [Accessed 8 January 2017].

[56] A. Kliarsky, "Responding to Zero Day Threat," SANS Institute, InfoSec Reading Room, June 27th, 2011.

[57] K. Zetter, Countdown to zero Day: Stuxnet and the Launch of the World's First Digital Weapon, New YOrk, NY: Crown Publishing Group, 2014.

[58] D. Alperovitch, "Revealed: Operation Shady RAT," McAfee, 2011.

[59] J. P. Wack, "Establishing a Computer Security Incident Response Capability (CSIRC)," National Institute of Standards and Technology, November 1991.

[60] CERT, "Creating and Managing Computer Security Incident Handling Teams (CSIRTs)," CERT Training and Education, Software Engineering Institute, Carnegie Mellon, Pittsburgh, PA, 2008.

[61] S. Behrens, C. J. Alberts and R. Ruefle, "Competency Lifecycle Roadmap: Toward Performance Readiness," Software Engineering Institue, Carnegie Mellon University, September 2012.

[62] P. Mell, K. Scarfone and S. Romanosky , "A Complete Guide to the Common Vulnerability Scoring System Version 2.0," (FIRST), Forum of Incident Response and Security Teams, June 2007.

[63] G. Bolch, S. Greiner, H. de Meer and K. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, 2nd Edition ed., Wiley, 2006.

[64] C. Carver, J. Hill and U. Pooch, "Limiting Uncertainity in Intrusion response," in *Proceedings of the 2001 IEEE workshop on Information Assurance and Security*, West Point, NY, 2001.

[65] G. Killcrece, "Incident Management," US-CERT, Department of Homeland Security, December 19, 2005.

[66] Expert_Two, Interviewee, *Interview with CSIR Senior Consultant (First Interview).* [Interview]. 9 September 2015.

[67] Cyber Security Coalition, "Cyber Security Incident Response Guide," Center for Cyber Security Belguim, November 2015.

[68] IIROC, "Cyber Incident Management Guide for IIROC Dealer Members," Investment Industry Regulatory Organization of Canada, December, 2015.

[69] R. L. Brand, "Coping with the Threat of Computer Security Incidents. A Primer from Prevention through Recovery," NIST, June 1990.

[70] G. Killcrece, K.-P. Kossakowski, R. Ruefle and M. Zajicek, "Organizational Models for Computer Security Incident Response Teams (CSIRTs)," Carnegie Mellon University, Software Engineering Institute, Pittisburgh, PA, USA, December 2003.

[71] California Information Security Office, "Incident Response Plan Sample," [Online]. Available: http://www.cio.ca.gov/OIS/Government/library/samples.asp. [Accessed 26 February 2017].

[72] CMISO, "Computer Security Incident Response Plan," Carnegie Mellon Information Security Office, Carnegie Mellon University, Pittsburgh, PA, February 2014.

[73] VERIS Open Source Project, "Vocabulary for Event Recording and Incident Sharing (VERIS)," [Online]. Available: http://www.veriscommunity.net.

[74] A. Dorofee, G. Killcrece, R. Ruefle and M. Zajicek, "Incident Management Capability Metrics Version 0.1," April 2007.

[75] D. Mundie, R. Ruefle, A. J. Dorofee, J. McCloud, S. J. Perl and M. L. Colins, "An Incident Management Ontology," in *STIDS 20014, CEUR Workshop Proceedings*, Novomber 2014.

[76] J. McClain, A. Silva, G. Emmanuel, B. Anderson, K. Nauer, R. Abbott and C. Forsythe, "Human Performance Factors in Cyber Security Forensic Analysis," in *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015)*, 2015.

[77] R. L. Bertini and M. W. Rose, "Using Archived ATMS Data for the Performance Evaluation of a Freeway Incident Response Program," in *Proceedings of The 7th International IEEE Conference on Intelligent Transportation Systems*, October 2004.

[78] R. E. Park, W. B. Goethert and W. A. Florac, "Goal-Driven Software Measurement - A Guidebook," Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, August 1996.

[79] IAEA, "Computer Security Incident Response Planning at Nuclear Facilities," International Atomic Energy Agency, Vienna, Austria, 2016.

[80] M. Singh, "Performance Evaluation of Intrusion Detection Systems," June 2009.

[81] K. Scarfone and P. Mell, "An Analysis of CVSS Version 2 Vulnerability Scoring," in *Third International Symposium on Empirical Software Engineering and Measurement*, Lake Buena Vista, FL, October 2009.

[82] C. Bartolini, C. Stefanelli and T. Mauro, "SYMIAN: Analysis and Performance Improvement of the IT Incident Management Process," *IEEE Transactions on Network and Service Management,* vol. 7, no. No. 3, pp. 132-144, September 2010.

[83] A. Paschke and E. Schnappinger-Gerull, "A Categorization Scheme for SLA Metrics," in *Multi-Conference Information Systems (MKWI06)*, Passau, Germany, 2006.

[84] F. Cohen, ""Simulating cyber attacks, defenses, and consequences"," *Computers & Security,* vol. 18, no. 6, pp. 479-518, 1999.

[85] ISO/IEC, "ISO 9126-1: Software Engineering - Product Quality - Part 1: Quality Model," International Organization of Standardization, Geneva, Switzerland, 2001.

[86] ISO/IEC, "ISO 9126-2: Software Engineering - Product Quality - Part 2: External Metrics," International Organization of STandardization, Geneva, Switzerland, 2003.

[87] ISO/IEC, "ISO 9126-3: Software Enginnering - Product Quality - Part 3: Internal Metrics.," International Organization of Standardization, Geneva, Switzerland, 2003.

[88] ISO/IEC, "ISO 9126-4: Software Engineering - Product Quality - Part 4: Quality-in-Use Metrics," International Organization of STandardization, Geneva, Switzerland, 2004.

[89] M. Parsons and P. Ebinger, "Performance Evaluation of the Impact of Attacks on Mobile Ad hoc Networks," in *28th International Symposium on Reliable Distributed Systems,* Niagra Falls, USA, September 2009.

[90] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato and A. Jamalipour, "A survey of routing attacks in mobile ad hoc networks," *IEEE Wireless Communications,* vol. 14, no. 5, December 2007.

[91] B. Littlewood, S. Brocklehurst, N. Fenton, P. Mellor, S. Page and D. Wright, "Towards Operational Measures of Computer Security," *Journal of Computer Security,* vol. 2, pp. 211-229, June 1993.

[92] E. Two, Interviewee, *Interview with CSIR Senior Consultant (Second Interview).* [Interview]. 17 August 2017.

[93] W. Jansen, "Directions in Security Metrics Research," National Institute of Standards and Technology (NIST), US Department of Commerce, April 2009.

[94] R. Savola, "A Novel Security Metrics TAxonomy for R&D Organizations," in *Proceedings of the ISSA 2008 Innovative Minds Conference*, Johannesburg, South Africa, January 2008.

[95] H. Rao, M. Gupta and S. J. Upadhyaya, Managing Information Assurance in Financial Services, Hershey, New York: IGI Publishing , 2007.

[96] NHTSA, "Emergency Medical Services Performance Measures: Recommended Attributes and Indicators for System and Service Performance," US Department of Transportation, National Highway Traffick Saftey Administration, December 2009.

[97] J. Steinke, B. Bolunmez, L. Fletcher, V. Wang, A. J. Tomassetti, K. M. Repchick, S. J. Zaccaro, R. S. Dalal and L. Tetrick, "Improving Cybersecurity Incident Response Team Effectiveness Using Team-Based Research," *IEEE Security and Privacy Magazine,* July 2015.

[98] S. W. Brenner, "Cybercrime Metrics: Old Wine, New Bottles?," *Virginia Journal of Law and Technology,* Vols. 9-13, Fall 2004.

[99] N. Simpson and P. Hancock, "Fifty Years of Operational Research and Emergency Response," *Journal of the Operational Research Society,* vol. 60, pp. 126-139, 2009.

[100] T. P. Seager, F. K. Satterstrom, I. Linkov, S. P. Tuler and R. Kay, "Typological Review of Environmental Performance Metrics (with Illustrative Examples for Oil Spill Response)," *Integrated Environmental Assessment and Management,* vol. 3, pp. 310-321, July 2007.

[101] D. E. Brown and D. Robinson, "Development of Metrics to Evaluate Effectiveness of Emergency Response Operations," in *10th International Command and Control Reserach and Technology Symposium*, 2005.

[102] C. H. Park, "Toward a Better Understanding of Complex Emergency Response Systems: An Event-Driven Lens for Integrating Formal and Volunteer-Based Participatory Emergency Responses," August 2016.

[103] B. Jackson, K. Faith and H. Willis, "Evaluating the Reliability of Emergency Response Systems for Large-Scale Incident Operations," RAND, Santa Monica, CA, 2010.

[104] R. Chen, R. Sharman, H. R. Rao and S. Upadhyaya, "Design principles for critical incident response systems," *Information Systems and e-Business Management,* vol. 5, no. 3, pp. 201-227, June 2007.

[105] M. Jaatun, E. Albrechtsen, M. Line, I. Tøndel and O. Longva, "A framework for incident response management in the petroleum industry," *International Journal of Critical Infrastructure Protection,* vol. 2, no. 1, May 2009.

[106] E. P. Harmon, S. C. Hensel and T. Lukes, "Measuring Performance in Services," *The McKinsey Quarterly,* 15 February 2006.

[107] A. Chibba, "Meausring supply chain performance measures - prioritizing performance measures," Department of Business Administration and Social Sciences, Lulea Unviersity of Technology, 2007.

[108] T. Hill, Manufacturing Strategy, New York: Palgrave, 2000.

[109] DHS, "National Incident Management System (NIMS)," US Department of Homeland Security, Washington, DC, December 2008.

[110] DHS, "National Response Framework, Second Edition," US Department of Homeland Security, Washington, DC, May 2013.

[111] Ontario Government, "Incident Management System (IMS) for Ontario," Ministry of Community Safety and Correctional Services, December 2008.

[112] E. Amoroso, Cyber Attacks: Protecting National Infrastructure, First Edition ed., Butterworth-Heinemann, November 2010.

[113] W. Humphrey, "The Watts New? Collection: Columns by the SEI's Watts Humphrey," Software Engineering Institute, Pittsburgh, Penn, November 2009.

[114] J. Hinze, S. Thurman and A. Wehle, "Leading Indicators of Construction Safety Performance," *Saftey Science,* vol. 51, pp. 23-28, 2012.

[115] NBER Macroeconomics Annual 1989: New Indexes of Coincident and Leading Economic Indicators, vol. 4, National Bureau Economic Research, March 1989, pp. 351-409.

[116] P. Micheli and L. Mari, "The theory and practice of performance measurement," *Management Accounting Research,* vol. 25, no. 2, pp. 147-156, June 2014.

[117] K. N. Balke, D. W. Fenno and B. Ullman, "Incident Management Performance Measures," Texas Transportation Institute, The Texas A&M University, College Station, TX, November 2002.

[118] K. T. Kearney and F. Torelli, "The SLA Model," in *Service Level Agreements for Cloud Computing,* Springer New York, September 2011, pp. 43-67.

[119] U. K. O. o. G. Commerce, ""ITIL Service Delivery" and "ITIL Service Support"," IT Infrastructure Library version 3, 2007.

[120] A. Stefani and M. Xenos, "Meta-metric Evaluation of E-Commerce-related metrics," *Electronic Notes in Theoetical Computer Science,* vol. 233, no. Proceedings of the Internnational Workshop on Software Quality and Maintainability (SQM 2008), pp. 59-72, March 2009.

[121] E. Weyuker, "Evaluating software complexity measures," *IEEE Transactions on Software Engineering,* vol. 14, no. 9, pp. 1357-1365, Sep 1998.

[122] W. K. Brotby and G. Hinson, PRAGMATIC Security Metrics: Applying Metametrics to Information Security, Auerbach Publications, 2013.

[123] C. Fruhwirth, S. Biffl, M. Tabatabai Irani and E. Weippl, "Addressing Misalignment between Information Security Metrics and Business-Driven Security Objectives," in *International Workshop on Security Measurement and Metrics (MetriSec),* Bolzano-Bozen, Italy, 2010.

[124] J. J. Carafano, "Preparing Responders to Respond: The Challenges to Emergency Preparedness in the 21st Century," *Heritage Lectures,* vol. No. 812, 20 November 2003.

[125] S. Pfleeger and . R. Cunningham, "Why Measuring Security is Hard?," *IEEE Security and Privacy,* vol. 8, no. 4, pp. 46-54, July 2010.

[126] W. Sanders, "Quantitative Security Metrics: Unattainable Holy Grail or Vital Breaktrhough Within Our Rich?," *IEEE Security and Privacy,* vol. 12, no. 2, pp. 67-69, April 2014.

[127] R. W. Floyd, "Nondeterministic Algorithms," *Journal of the ACM,* vol. 14, no. 4, pp. 636-644, October 1967.

[128] M. Davis, Markov Models and Optimization, Chapman and Hall/ CRC, August 1993.

[129] J. S. Sadaghiyani, "Concepts of Decision Making Under Uncertain, Risky and Deterministic Situations," *The Journal of Mathematics and Computer Science,* vol. 2, no. 3, pp. 529-545, 2011.

[130] Y. Ben-Haim, Info-Gap Decision Theory: Decisions Under Severe Uncertainty, 2nd Edition ed., Academic Press, 2006.

[131] T. R. Osborne, "Building an Incident Response Program to Suit Your Business," SANS Institute, July 3, 2001.

[132] SANS, "Glossary of Security Terms," [Online]. Available: https://www.sans.org/security-resources/glossary-of-terms/. [Accessed 14 March 2017].

[133] M. Cohen, D. Bilby and G. Caronn, "Distributed forensics and incident response in the enterprise," *Digital Investigations,* vol. 8, no. 11th Annual Digital Forensics Research Conference, pp. S101-S110, 2011.

[134] S. A. Likierman, "The Five Traps of Performance Measurement," *Harvard Business Review,* no. October issue, 2009.

[135] W. W. Eckerson, "Performance Management Strategies: How to Create and Deploy Effective Metrics," TDWI Best Practices Report (First Quarter 2009), 2009.

[136] M. Bada, S. Creese, M. Goldsmith, C. Mitchell and E. Phillips, "Computer Security Incident Response Teams (CSIRTs) An Overview," Global Cyber Security Capacity Center, May 2014.

[137] P. Bogetoft, Performance Benchmarking: Measuring and Managing Performance, New York: Springer, 2012.

[138] S. E. Ramona, "Advantages and Disadvantages of Quantitative and Qualitative Information Risk Approaches," *Chinese Business Review,* vol. 10, no. 12, pp. 1106-1110, December 2011.

[139] Y. E. Chan, "IT Value: the great divide between qualitative and quantitative and individual and organizational measures," *Journal of Managmenet Information Systems - Special Issue: Impacts of Information Technology investment on organizational performance,* vol. 16, no. 4, pp. 225-261, March 2000.

[140] T. Proffitt, "Creating and Managing an Incident Response Team for a Large Company," SANS Institute, 2007.

[141] ISO/IEC, "Quality Management and Quality Assurance - a Vocabulary," International Organization for Standardization, Geneve, 1994.

[142] T. F. Brady, "Emergency Management: Capability Analysis of Critical Incident Response," in *Proceedings of the 2003 Winter Simulation Conference*.

[143] L. Carlucci, S. Jain, E. Kinber and F. Stephan, "Variations on U-Shaped Learning," *Information and Computation*, vol. 204, no. 8, pp. 1264-1294, August 2006.

[144] K. Berardo, "Alternatives to the U-Curve Model," Cutlurosity, 2007.

[145] M. Sherjan, "Operationalizing Incident Response with Technology," 24 May 2017. [Online]. Available: https://www.radarfirst.com/blog/operationalizing-incident-response-with-technology. [Accessed 12 June 2017].

[146] A. Torres, "Incident Response: How to Fight Back," SANS Institute, August 2014.

[147] J. Macmillan, E. E. Entin and D. Serfaty, "Communication Overhead: The Hidden Cost of Team Cognition," in *Team cognition : understanding the factors that drive process and performance*, Washington, DC., American Psychological Association, 2004.

[148] C. Perrow, "The Analysis of Goals in Complex Orgnaizations," *American Sociological Review,* vol. 26, pp. 854-866, 1961.

[149] A. Gunasekaran, C. Patel and R. E. McGaughey, "A framework for supply chain performance measurement," *International Journal of Production Economics,* vol. 87, pp. 333-347, 2004.

[150] J. L. Harbour, The Basics of Performance Measurement, 2nd Edition ed., Productivity Press, 2009.

[151] G. Doran, "There is a SMART Way to Write Management's Goals and Objectives," *Management Review,* vol. 70, no. 11, pp. 35-36, 1981.

[152] R. Kaufman, "Preparing Useful Performance Indicators," *Training and Development Journal,* vol. 42, no. 9, p. 80, September 1988.

[153] R. S. Kaplan, "Conceptual Foundations of the Balanced Scorecard," Working Paper 10-074, Harvard Business School, 2010.

[154] J. Becher, "Manage by Walking Around: Qualitative KPIs," 26 August 2006. [Online]. Available: http://jonathanbecher.com/2006/08/26/qualitative-kpis/. [Accessed 11 June 2017].

[155] P. Leitão and F. Restivo, "The use of qualitative indicators for performance measurement in manufacturing control systems," in *Proceedings of the 11th IFAC Symposium on Information Control Problems in Manufacturing (INCOM'04),* Salvador Baia, Brazil, 2004.

[156] A. Shaout and M. Al-Shammari, "Fuzzy Logic modeling for performance appriasal system: A framework for empirical evaluation," *Expert systems with Applications,* vol. 14, no. 3, pp. 323-328, April 1998.

[157] A. H. Lee, W.-C. Chen and C.-J. Chang, "A fuzzy AHP and BSC approach for evaluating performance of IT department in the manufacturing industry in Taiwan," *Expert Systems with Applications,* vol. 34, no. 1, pp. 96-107, 2008.

[158] C.-T. Lin, H. Chiu and Y.-H. Tseng, "Agility Evaluation using fuzzy logic," *International Journal of Production Economics,* vol. 101, no. 2, pp. 353-368, June 2006.

[159] P. Goodwin and G. Wright, Decision Analysis for Management Judgement, 5th Edition ed., Wiley, May 2014.

[160] M. Taghavifard, K. K. Damghani and R. T. Moghaddam, "Decision Making Under Uncertain and Risky Situations," Society of Actuaries, Schaumburg, IL, 2009.

[161] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management,* vol. 45, pp. 427-437, 2009.

[162] DHS, "A roadmap for Cybersecurity Research," Department of Homeland Security, November 2009.

[163] M. Franklin, Performance Gap Analysis, Association for Talent Development, March 2006.

[164] D. V. Tiem, J. L. Moseley and J. C. Dessinger, Fundamentals of Performance Improvment: Optimizing Results through People, Process and Organizations, Pfeiffer, April 2012.

[165] F. S. Wilmoth, C. Prigmore and M. Bray, "HPT Models: An Overview of the Major Models in the Field," International Society for Performance Improvement, September 2002.

[166] R. Chevalier, "Gap Analysis Revisited," *Performance Improvement,* vol. 49, no. 7, pp. 5-7, August 2010.

[167] G. Kilcrece and R. Ruefle, Creating and Managing Computer Security Incident Response Teams (CSIRTs), CERT Training and Education, Software Engineering Institute, Carnegie Mellon University, 2008.

[168] F. Lu, J. Qiu and J. Li, "Performance bottleneck detection in scalability testing". US Patent 20130179144 A1, 11 July 2013.

[169] A. Torres, "Building a World-Class Security Operations Center: A roadmap," SANS Institute, May 2015.

[170] M. Churchman, "Avoiding Incident Response Bottlenecks," 23 March 2017. [Online]. Available: https://www.pagerduty.com/blog/avoiding-incident-response-bottlenecks/. [Accessed 13 August 2017].

[171] M. Hathaway, "Attackers Prey on Incident Response Bottlenecks," 16 March 2016. [Online]. Available: https://community.rapid7.com/community/insightidr/blog /2016/05/13/attackers-prey-on-incident-response-bottlenecks. [Accessed 13 August 2017].

[172] Raj, "Techniques for Performance Bottleneck Analysis," [Online]. Available: http://www.performancetestingfun.com/bottleneckanalysis/. [Accessed 13 August 2017].

[173] P. Stephenson, "Modeling of Post-Incident Root Cause Analysis," *International Journal of Digital Evidence,* vol. 2, no. 2, Fall 2003.

[174] Expert_Four, Interviewee, *Interview with Cyber Security Expert.* [Interview]. 29 June 2017.

[175] T. Kubiak and D. W. Benbow, The Certified Six Sigma Black Belt Handbook, second edition ed., Milwaukee, Wisconsin: ASQ Quality Press, 2009.

[176] J. Sanders, "Introduction to Why-Because Analysis," Dipl.-Inform, February 2012.

[177] K. Ishikawa, Guide to Quality Control (2 Rev Sub Edition), Asian Productivity Organization, February 1986.

[178] G. Hubbard, "Measuring Organizational Performance: beyond the triple bottom line," *Business Strategy and the Environment,* vol. 18, no. 3, pp. 177-191, December 2006.

[179] I. Bernki and K. Prislan, "Measuring Information Security Performance with 10 by 10 Model for Holistic State Evaluation," *PLoS One,* vol. 11, no. 9, 2016.

[180] E. A. Awadallah, "A Critique of the Balanced Scorecard as a Performance Measurement Tool," *International Journal of Business and Social Science,* vol. 6, no. 7, pp. 91-99, July 2015.

[181] A. Neely and C. Adams, "Perspectives on Performance: the Performance Prism," *Journal of Cost Management,* vol. 15, no. 1, pp. 7-15, January 2001.

[182] J. Elkington, Cannibals With Forks: the Triple Bottom Line of the 21st Century Business, Oxford: Capstone, 1997.

[183] DHS Risk Steering Committee, "DHS Risk Lexicon," Department of Homeland Security, September 2010.

[184] R. Adcock and D. Collier, "Measurement Validity: A shared standard for qualitative and quantitative research," *The American Political Science Review,* vol. 95, no. 3, pp. 529-546, September 2001.

[185] R. M. Gagne, "Military Training and Principles of Learning," *American Psychologist,* vol. 17, no. 2, pp. 83-91, February 1962.

[186] G. Ruibin, C. K. Y. Tony and M. Gaertner, "Case-Relevance Information Investigation: Building Computer Intelligence to the Current Computer Forensic Framework," *International Journal of Digital Evidence,* vol. 4, no. 1, Spring 2005.

[187] X. Koufteros, A. (. Verghese and L. Lucianetti, "The effect of performance measurement systems on firm performance: A cross-sectional and longitudinal study," *Journal of Operations Management,* vol. 32, pp. 313-336, 2014.

[188] M. Mura, M. Longo and P. Micheli, "The Effects of Performance Measurement Uses on Organizational Ambidexterity and Company Performance," WBS Accounting Group, Warwick Business School, 2016.

[189] E. Chew, M. Swanson, K. Stine, N. Bartol, A. Brown and W. Robinson, "Performance Measurement Guide for Information Security," Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD, July 2008.

[190] M. McQueen, W. Boyer, S. McBride, M. Farrar and Z. Tudor, "Measurable Control System Security Through Ideal Driven Technical Metrics," in *SCADA Security Scientific Symposium, Idaho National Laboratory*, January 2008.

[191] N. Subramanian and L. Chung, "Metrics for Software Adaptability," in *Proceedings of Software Quality Management (SQM 2001)*, Loughborough, UK, April, 2001.

[192] E. Pulakos, S. Arad, M. Donovan and K. Plamondon, "Adaptability in the workplace: development of a taxonomy of adaptive performance," *Journal of Applied Psychology* , vol. 85, no. 4, pp. 612-624, August 2000.

[193] S. Rance, "Defining Availability in the Real World," itSMF, UK, June 7, 2017.

[194] ITIL, "Information Technology Infrastructure Library (ITIL) Service Design," The Stationary Office, Norwich, UK, 2011.

[195] G. Ritchie, "Introducing ITIL Availability Management Version 2 (White Paper)," Serio Ltd, 2009.

[196] B. Maycock, "Communication Strategies for Incident Response (Mass Notification and Beyond)," Impact Technologies, Inc., 2010.

[197] C. Mann, "Is there a good measure for competitiveness?," in *Is the U.S. Trade Deficit Sustainable?*, Washington, D.C., USA, Peterson Institute for International Economics, September 1999, pp. 95-114.

[198] ISO/IEC, "ISO 27001: Information Technology - Security Techniques - Information Management Systems - Requirements," The International Organization for Standardization, Geneva, Switzerland, October 15, 2005.

[199] HHS, "Standards for Privacy of Individually Identifiable Health Information Regulation Text, as amended," Office for Civil Rights, US Department of Health and Human Services,, October 2002.

[200] Big Ten Academic Alliance, "Incident Cost Analysis and Modeling Project (Final Report)," Report from the CIC Security Working Group to the CIC Cheif Information Officers, Champaign, IL, 2000.

[201] DHS, "NCCIC Cyber Incident Scoring System," Department of Homeland Security: National Cybersecurity and Communications Integration Center (NCCIC).

[202] ISO/IEC, "ISO 9001: Quality Systems - Model for Quality Assurance in Design, Development, Production, Installation and Servicing," International Organization for Standardization, Geneva, Switzerland, 1994.

[203] ISO/IEC, "ISO 27001: Information Technology - Security Techniques - Information Security Management Systems - Requirements," International Organization for Standardization, Geneva, Switzerland, 2013.

[204] R. Rowlingson, "A Ten Step Process for Forensic Readiness," *International Journal of Digital Evidence*, vol. 2, no. 3, Winter 2004.

[205] K. Bechtel, "The Cost of Incident Response," August 2015.

[206] P. Doubilet, M. Weinstein and B. McNeil, "Use and misuse of the term 'cost effective' in medicine," *The New England Journal of Medicine,* vol. 314, pp. 253-256, 1986.

[207] ACDIS, "Association of Clinical Documentation Improvement Specialists," [Online]. Available: www.acdis.org. [Accessed 22 July 2017].

[208] A. Sanabria, "Malware Analysis: Environment Design and Architecture," SANS Institute, January 18th, 2007.

[209] M. Vojnovic and A. Ganesh, "On the effetiveness of automatic patching," in *Proceedings of the 2005 ACM Workshop on Rapid malcode*, Fairfax, VA, USA, November 11, 2005.

[210] M. Grace, Y. Zhou, Q. Zhang, S. Zou and X. Jiang, "RiskRanker: scalable and accurate zero-day android malware detection," in *Proceedings of the the 10th International conference on Mobile Systems, applications and services*, Low Wood Bay, Lake District, UK, June 25-29, 2012.

[211] R. Beach, A. Muhlemann, D. Price, A. Paterson and J. Sharp, "A review of manufacturing flexibility," *European Journal of Operational Research*, vol. 122, pp. 41-57, 2000.

[212] D. Sule, "Importance of Forensics Readiness," *Information Systems Audit and Control Association (ISACA) Journal*, vol. 1, 2014.

[213] Y.-C. Liao, "Process Tracking for Forensics Readiness," *Thesis for Doctor of Philosophy in Information Security*, December 2016.

[214] I. Orton, A. Alva and B. Endicott-Popovsky, "Legal Processes and Requirements for Cloud Forensic Investigations," in *Cybercrime and Cloud Forencis: Applications for Investigation Processes*, Hershey, PA, USA, Information Science Rference, December 2012, pp. 186-234.

[215] D. A. Dittrich, "Developing an Effective Incident Cost Analysis Mechansim," Symantec Connect, June 11, 2002.

[216] REN-ISAC, "Research & Education Networking Information Sharing & Analysis Center," [Online]. Available: https://www.ren-isac.net/. [Accessed 9 July 2017].

[217] InfraGrad, "InfraGrad: Partnership for Protection," [Online]. Available: https://www.infragard.org/. [Accessed 9 July 2017].

[218] C. L. King, "Partnership Effectiveness Continuum: A research-based tool for use in developing, assessing and improving partnerships," Education Development Center, 2014.

[219] DHS, "Government Facilities Sector-Specific Plan: An annex to the NIPP 2013," Department of Homeland Security, 2015.

[220] B. Acohido, "Improving Detection, Prevention and Response with Security Maturity Modeling (SANS Whitepaper)," SANS Institute, May 2015.

[221] M. Bada, S. Creese, M. Goldsmith, C. Mitchell and E. Phillips, "Improving the effectiveness of CSIRTs (Draft Working Paper)," Global Cyber Security Capacity Center, University of Oxford, October 2014.

[222] LCE: Life Cycle Engineering, "Four Ways to Measure the Efectivenss of Your Root Cause Analysis Process," 2013.

[223] R. S. Kaplan and D. Norton, "The Balanced Scorecard: Measures that Drive Performance," *Harvard Business Review,* vol. 70, no. 1, pp. 71-79, January-February 1992.

[224] A. Seovic, M. Falco and P. Peralta, Oracle Coherence 3.5, Birmingham, UK: Packt Publishing, March 2010.

[225] J. Holtman and N. J. Gunther, "Getting in the Zone for Successful Scalability," in *Proceedings of the 34th International Computer Measurement Group Conference*, Las Vegas, Nevada, September 2008.

[226] J. Dykstra, Essential Cybersecurity Science: Build, Test and Evaluate Secure Systems, Sebastopol, CA, USA: O'Reilly Media, 2016.

[227] N. R. Mead, R. J. Ellison, R. C. Linger, T. Longstaff and J. McHugh, "Survivable Network Analysis Method," The Software Engineering Institute, September 2000.

[228] E. Salas, R. Grossman, A. Hughes and C. Coultas, "Measuring Team Cohesion: Observations from the Science," *The Journal of Human Factors and Ergonomics Society,* vol. 57, pp. 365-374, May 2015.

[229] B. Rawlins, "Measuring the Relationship Between Organizational Transparency and Employee Trust," *Public Relations Journal,* vol. 2, no. 2, pp. 1-21, 2008.

[230] D. R. Nerenz and N. Neil, "Performance Measures for Health Care Systems," Commissioned Paper for the Center for Health Management Research, Ann Arbor, MI, May 1, 2001.

[231] H. Sayama, Introduction to the Modeling and Analysis of Complex Systems, Open SUNY Textbooks, August 2015.

[232] C. R. L. Frances, M. J. Santana, R. H. Santana, N. L. Vijaykumar and V. de Carvalho, "Performance Evaluation of Complex Systems: Queuing Statecharts Approach," in *Proceedings of First Workshop on Performance of Computational Systems (WPerformance) / Brazillian Copmuting Conference (SBC),* Florianopolis, SC, Brazil, July 2002.

[233] J. Huang, J. Voeten, P. v. d. PUtten, A. Ventevogel, R. Niesten and W. v. Maaden, "Performance Evaluation of Complex Real-Time Systems: A case Study," in

Proceedings of the third PROGRESS Workshop on Embedded Systems, Jaarbeurs Utrecht, Netherlands, 2002.

[234] CSIRT, "Implementing an Incident Response Team (IRT)," Computer Security Incident Response Team (csirt.org).

[235] S. Kollias, V. Vlachos, A. Papanikolaou, P. Chatzimisios, C. Ilioudis and K. Metaxiotis, "Measuring the Internet's threat level: A global-local approach," in *Proceedings of the International Symposium on Computers and Communications*, June 2014.

[236] K. Kent, S. Chevalier, T. Grance and H. Dang, "Guide to Integrating Forensics Techniques into Incident Resposne," NIST SP800-86 Notes, August 2006.