Benchmarking CSIRT work processes

Ivar Kjærem

Master's Thesis
Master of Science in Information Security
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2005

The MSc programme in Information Security
is run in cooperation with the Royal Institute
of Technology (KTH) in Stockholm.

Institutt for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Abstract

The need for securing the information systems within an organization is well understood today. Organizations implement preventive measures to stop malicious software and attackers at their gates. Larger organizations also establish computer security incident response teams, CSIRTs, this in the recognition that not all attackers or malicious software are stopped. A compromise of the information systems security may cause great damage, and, if not responded to quickly, put the organization out of business.

After establishing a CSIRT, the performance of the team is seldom evaluated, as long as the team resolves the incidents that occur. In this thesis we have developed a set of metrics to measure the performance of the work processes in a CSIRT. The metrics are based on how well different policies and procedures are implemented, time consumption in the incident handling and the results of the investigation conducted.

To show how these metrics can be used to improve the work processes in the CSIRT, a benchmarking experiment was conducted. The benchmarking was accomplished by sending out a questionnaire to several large organizations which we knew had established a team to handle security incidents on their information systems. The answers given in the questionnaire were used as input to the metrics. By using the metrics in a benchmarking between CSIRTs, we have been able to rank the teams, and gained insight in what areas the different teams perform well and where they perform poorly.

Sammendrag (Abstract in Norwegian)

Nødvendigheten av å sikre informasjonssystemene i en organisasjon akseptert i dag. Organisasjoner implementerer preventive tiltak for å hinder angripere og ondsinnet programvare tilgang til informasjonssystemene. Større organisasjoner etablerer også insident respons teams, CSIRTs, fordi man innser at ikke alle angripere blir stoppet før de kommer inn. En kompromittering av sikkerheten til informasjonssystemene kan føre til store ødeleggelser dersom de ikke blir tatt hånd om.

Etter etableringen av et CSIRT blir det sjelden gjort en evaluering av effektiviteten til teamet, så lenge det håndterer de hendelser som oppstår. I denne oppgaven har vi utviklet et sett med metrikker for å måle ytelsen til arbeidsprosessene i et CSIRT. Metrikkene baserer seg på hvor godt ulike policyer og prosedyrer er implementert, tidsforbruket i insident håndteringen, og resultatene av de undersøkelser som blir gjennomført.

For å vise hvordan man kan bruke metrikkene til å forbedre arbeidsprosessene i et CSIRT, ble det gjennomført et benchmarking eksperiment. Benchmarkingen ble gjennomført ved å sende ut et spørreskjema til flere store organisasjoner som man visste hadde etablert et team for å håndtere sikkerhetstruende hendelser på sine informasjonssystemer. Svarene de ulike organisasjonene gav i spørreskjemaet ble brukt som inngangsverdier til metrikkene. Ved å bruke disse metrikkene i en benchmarking mellom flere CSIRTs har vi kunnet rangere teamene innbyrdes, og fått innblikk i hvilke områder de ulike teamene fungerer bra og hvor det er rom for forbedring.

Preface

This thesis is the final work on my Master degree studies at Gjøvik University College (GUC). While studying I have also been employed as a senior instructor at the Norwegian Defense Security Agency (FSA) at Jørstadmoen. The Master thesis has in many ways contributed the work that I do for FSA, and given me opportunity to use the knowledge acquired through the master thesis in practice.

I would like to thank those who have contributed to this project. It would not have been possible to complete without the contribution from the CSIRTs that were kind enough to answer my questionnaire. Thanks to my employer, Norwegian Defense Security Agency, for letting me use work hours to complete my Master degree. And last but not least my supervisor at GUC, Prof. Slobodan Petrović, who has been very helpful, and has been an excellent critic.

Jørstadmoen, 1st June 2005

Ivar Kjærem

Table of Contents

1 Introduction 1
1.1 Topic: Security management 1
1.2 Research problem 1
1.3 Motivation 1
1.4 Research questions 2
1.5 Limitations 2
2 State of the art 3
2.1 Work processes in CSIRT 3
2.2 Measures and metrics 5
2.3 Benchmarking 6
3 Summary of claimed contribution 9
4 Choice of methods 11
4.1 Work processes in CSIRT 11
4.2 Measures and metrics 11
4.3 Benchmarking 11
5 Benchmarking CSIRT work process performance 13
5.1 Suggested CSIRT work processes 13
5.1.1 Preparation 14
5.1.2 Detection 15
5.1.3 Response 16
5.1.4 Recovery 20
5.1.5 Follow up 20
5.2 Suggested measures and metrics 21
5.2.1 Preparation process performance metric 23
5.2.2 Detection process performance metric 26
5.2.3 Initial response process performance metric 28
5.2.4 Containment process performance metric 31
5.2.5 Investigation process performance metric 33
5.2.6 Eradication process performance metric 36
5.3 Benchmarking experiment 37
5.3.1 Preparation 37
5.3.2 Planning 37
5.3.3 Searching 38
5.3.4 Observation 38
5.3.5 Analysis 39
5.3.6 Recommendations 42

6 Discussion 45
6.1 Suggested processes 45
6.2 Suggested metrics 45
6.3 Benchmarking 47
7 Conclusion 49
8 Future work 51
9 Abbreviations 53
10 About the author 55
11 References 57
Appendix A: Questionnaire 61

Figures

Figure 1: Overall incident response process 13

Figure 2: Detection process in detail 16

Figure 3: Response process in detail 18

Tables

Table 1: Metric definition template 23

Table 2: Preparation process performance metric 24

Table 3: Detection process performance metric 26

Table 4: Initial response process performance metric 29

Table 5: Containment process performance metric 31

Table 6: Investigation process performance metric 34

Table 7: Eradication process performance metric 36

Table 8: Measured values for the Preparation process 39

Table 9: Measured values for the Detection process 39

Table 10: Measured values for the Initial response process 40

Table 11: Measured values for the Containment process 40

Table 12: Measured values for the Investigation process 40

Table 13: Measured value for the Eradication process 41

1 Introduction

1.1 Topic: Security management

The management of every company wants security in their information infrastructure. This because information the company possesses represents an enormous value, either because it needs to be available to keep the business going, or it needs to be protected against manipulation or exposure of secure business secrets. It is important to organizations with many and complex information systems, to organize the security process in a way that the security standard is sustained, and not degraded due to fragmented responsibilities. Many organizations choose to establish a centralized resource to support the whole organization, a Computer Security Incident Response Team (CSIRT), and through this resource ensure a competent and consistent handling of security incidents.

1.2 Research problem

Organizations that establish CSIRTs do this to improve security in their information systems. Holm points out in his study [1] that it is necessary to explore how an organization could establish a CSIRT in a way that it improves information security in the organization. After establishing the CSIRT, an organization seldom measures the performance of the CSIRT to evaluate the investment. But it should check whether the resources and efforts in this direction make the quality of the work processes as good as possible. Today, organizations don't have any methods to measure the performance of the CSIRT to establish "Best Practice".

1.3 Motivation

To achieve improved security in their information systems, every organization needs a structured way to handle security incidents in the systems. A security incident should be handled as close to the source as possible, but in large organizations it is not economically viable to have all the competence needed to handle the incident in every part of the organization. It is more efficient and economic to establish a centralized team that can support different parts of the organization in handling a security incident. This CSIRT will be able to handle and follow up security incidents, and support the local security officer in the process of re-establishing the security level in the attacked system.

As a result of a requirement for better economy, reorganization and downsizing, it is important for the organization to make the CSIRT as efficient as possible, and at the same time strengthen the security. In this way the organization will get the best security possible in return for their investment. To utilize the team in the best way, the management needs to know which factors that affect the performance of the team, and it is important to develop methods for measuring performance in CSIRTs. It can be difficult to find a definite measure for information security, but it might be easier to compare teams in order to find the one which is better. To be able to compare different

teams, it is important to find the same factors in the teams that are measurable. Such a comparison is often called a *benchmarking*.

Every organization will benefit from knowing how internal work processes are affected by each other in a positive or negative way. Established CSIRTs could use this knowledge to compare their own processes with the processes in other teams. Through benchmarking it is possible to find out how your own team performs compared to other teams with similar processes. The benchmarking will show which processes are competitive, and which processes are not performing as good as they could. The benchmarking process will also give insight into how processes that perform well are designed, and it can be possible to use this to improve our own processes. The benchmarking process does not only identify processes that don't perform well, but also gives information on how to improve them.

1.4 Research questions

When a CSIRT is established, the management will like to see results from the investment, which means that they need to know how the CSIRT performs. To be able to measure the performance of the team, it is necessary to identify the work processes of the team, and develop metrics for these work processes. The metrics must measure the performance of the organization, and identify what is better or worse. A measure can then be used to compare the performance of work processes in different organizations, and give proof of what team is the best.

The following research questions need to be answered:

1. Which basic work processes exist in a CSIRT?
2. Which metrics could be developed for the performance of the basic work processes?
3. How can benchmarking be used to compare performance in different CSIRTs?

To answer these questions, different approaches are needed. A mixed method approach was used in this thesis. This approach is described by Cresswell in [11]. The first part of the thesis is a literature study to identify the different work processes and activities within a CSIRT. Based on this study metrics for the work processes in a CSIRT are developed. These metrics constitute a quantitative method for measuring the performance of a team, and are used in a case study where the different teams are benchmarked against each other.

1.5 Limitations

An identification of all work processes in a CSIRT could be very time-consuming. This thesis focuses on the processes that are directly related to the incident handling (preparation, detection and response), and metrics for these processes. A particular type of CSIRT, called Internal Centralized CSIRT [43], is given a special attention. This team has responsibility for handling incidents in a well defined part of the organization, and is a part of the organization itself.

2 State of the art

Many sources describe CSIRTs, and the different tasks that CSIRTs can and should perform. The different sources use different terms for the work processes in a CSIRT, and the focus is influenced by the author's interests and bias. Some sources mention measuring CSIRT performance, but none use benchmarking to compare performance between CSIRTs. This Chapter presents a survey of previous work that may be relevant to this thesis.

2.1 Work processes in CSIRT

A report by Killcrece et al. [2], based on a survey among many CSIRTs, shows how CSIRTs can be organized, funded and which processes they include. This is a valuable source in identifying work processes in CSIRTs, and what activities the different processes include.

West-Brown et al. [3] presents an overview of different functions and tasks that a CSIRT could be given. This is meant as a guide for those planning to establish a CSIRT, and the organization establishing a CSIRT needs to adjust the functions and tasks the CSIRT should undertake to adapt to the organization and business culture. This document gives important knowledge about which tasks and functions are present in CSIRTs, but the terms used in this document are not necessarily the terms used by the teams in practice.

Wack [4] presents a list of functions a CSIRT must be able to handle, but does not discuss why the different functions are necessary. The document is mainly focused on the management processes, and those processes needed to establish a CSIRT. The document does not describe in detail the functions directly related to incident handling. This means that the document is less suitable for identifying work processes in incident handling, but is suitable for examining processes in incident management.

In [5], Schultz and Shumway present a guide to how incident handling can be organized for an organization. They use different terms on CSIRT processes and functions than those used in [2, 3]. Their main processes are Preparation, Detection and Reaction. [5] focuses on tasks in both incident handling and incident management. The use of different terms requires each CSIRT to describe what is meant by the term used in the team. If each team has done this it will make it easier to find the processes that can be compared in different teams.

Schultz [8] claims that a CSIRT has an evolution in four phases: Initial, Critical, Established and Post-established. He argues that most CSIRTs end up going in circles in the critical phase, because they are not able to serve the constituency in an efficient manner, they just repeat information that comes from other sources. He proposes changes to get an effective CSIRT that is constantly improving security in the organization. It is important to identify what keeps the CSIRT in this loop, and what

needs to be done to get out of the loop. These factors could be used to measure how far in the evolution described by Schultz, the CSIRT has reached.

In [16], Alberts et al. divides the work processes in a CSIRT into five main processes: Prepare/Sustain/Improve, Protect Infrastructure, Detect, Triage and Response. In addition the report shows which functions are believed to belong to each process. The document [16] is an important contribution to identify work processes in a CSIRT.

In [18], Lucas and Moeller describe an incident response methodology in seven steps. These steps contain mostly the same issues as other methodologies, and the authors recognize that the issues are very much the same even though different methodologies use different terms. The authors of [19] also describe a seven step methodology, but use different terms and organize the activities in a slightly different manner than [18]. Pelkari and Chuvakin describe in [17] a six step methodology. Their methodology also includes most of the activities that are described in other methodologies, but some new activities are included.

Wright [32] points out the importance of the incident response policy, and describes how a useful policy should be designed. In addition, the article describes necessary steps to build an incident response process. This article, however, does not describe the different steps in detail, but it gives an overview of what is needed.

Masurkar has published a series of articles [38, 39, 40, 41] describing the process of establishing and running an incident response group in an enterprise. The first article [38] focuses on establishing an incident response team and developing an incident response policy. Article [39] focuses on the incident response process and the different activities included in incident response. Article [40] gives an overview of the follow-up activities like legal actions, taking inventory and lessons learned. The last article [41] goes more in depth into how to process and analyze incident data and, among other issues, it covers forensics. These articles give a valuable insight into how computer security incident response teams could be established and run. The more technical details, however, are focused on using hardware from Sun Microsystems and Solaris or another Unix-type operating system.

In the framework for incident response from DePaul University [33] the focus is on the need for establishing “security levels”. By security level [33] means a scale to be able to give each incident a classification dependent on the assumed severity of the incident. In addition, the framework describes some roles that need to be defined in the incident response process, and also divides the incident response process into different actions. The actions are only briefly described, but give a useful insight into how incident response can be organized.

Although there are many sources describing different methodologies for incident response, the terms used and the way the activities are organized differ. In this thesis a

response methodology is established that can be recognized in the teams to be benchmarked.

2.2 Measures and metrics

Payne [12] has described how good metrics should be developed. She claims a good metric should be SMART: Specific, Measurable, Attainable, Repeatable and Time dependent. Swanson et al. [13] describe how IT security metrics can be developed and used to facilitate decision making and improving performance and accountability for the organization's security services. The document describes the process in general for the whole organization, but it may serve as a valuable guide in developing metrics for the CSIRT. It is, however, necessary to adapt the techniques in order for them to be used for the CSIRT processes.

Vaughn et al. [14] propose a taxonomy for information security metrics. Alsaker [15] presents a list of indicators for information security published by Centre of Competence for IT in Health services (KITH) in Trondheim. All these documents give important background information that can be used in developing metrics for work processes in CSIRTs.

Wack [4] describes different parameters that can be used to measure the performance in CSIRTs. It is recognized that it can be difficult to find one single measure that gives the value for the CSIRT performance, but statistical analysis of data collected for different incidents is believed to be able to be used for measuring CSIRT performance. In order to be able to improve CSIRT performance it is important to instruct the management which processes function well, and where is a potential to improve. A benchmarking may give the organization information on what to do to improve performance in its own processes. This information could be obtained by studying the processes that perform well in the benchmarking process. [4] identifies few parameters that could be used to measure CSIRT performance, but it is necessary to identify more parameters in order to be able to make a more precise measure and to be able to benchmark different types of CSIRTs.

Grance et al. [6] presents a guide to incident handling that suggests different parameters to measure incident related data, and discusses pros and cons with the different parameters. However, the number of presented parameters is too small, and it is necessary to find more parameters that can be measured.

Brownlee and Guttman describe in [7] which expectations the users within a CSIRT constituency will have to the team. One could say that this is an attempt to describe the user requirements for the CSIRT, by defining which services a user will expect the CSIRT to perform. The quality of the services performed within the constituency will give an indication of the performance of the team. If it is possible to measure the quality of the services the CSIRT performs, that would be a good metric for benchmarking teams against each other, and at the same time it would be of great value for the management.

West-Brown et al. [3] point out that a quality assurance system is necessary for a CSIRT, and describes a framework for such a system. The description also gives examples of different parameters that can be measured to give an indication of the quality of a CSIRT. In most cases a quality indicator will be the same as a performance indicator, and the suggestions listed in this guide are worth considering for use as performance indicators. But this is in no way a finite list of indicators, and it is necessary to find more performance indicators.

2.3 Benchmarking

In the report [2], data from a considerable number of CSIRTs have been collected. This report describes how the different CSIRTs are structured and organized. The report can be used as a basis for a benchmarking experiment between CSIRTs. It gives information on what is the working practice for CSIRTs considering organization, funding, duties and offered services. Comparing a CSIRT to the data given in this report would be a rough analysis of the CSIRT in question. This would, however, only be a rough analysis because, as the report points out, every CSIRT must be adapted to the organization and culture it is supposed to serve.

In [9], Pethia and van Wyk conclude that the CERT system is dependent on the knowledge and experience of the participating teams, i.e. on improving each single team's ability to handle incidents. This sharing of information and knowledge makes the community stronger. If it could be possible to share performance data in the same manner, the community would benefit from this by identifying the "best practice". The CERT system has this potential, but a system to share this kind of benchmarking results for CSIRTs has not been established yet.

Andersen and Pettersen present in [10] a general methodology for the benchmarking process. The book describes both "performance benchmarking" and "process benchmarking". It gives a good introduction to the benchmarking process, and has a practical approach on how to execute the benchmarking. However, it requires that the benchmarking process presented in the book is adapted to the business areas to be benchmarked.

Fogle et al. [20] describe the experience made by a team conducting a benchmarking experiment. The experiment deals with benchmarking software development, and [20] describes the experiences gained by the team in all phases of the experiment. Although it is about benchmarking in quite a different area, much of the experience of what to do and not to do is still useful in benchmarking CSIRTs.

Sole and Bist discuss in [26] the use of benchmarking for the process of writing technical information, and describes a benchmarking process in six steps. The article gives a quick overview of the benchmarking process, and some information on what to do in the different steps. The process described for the benchmarking could be applied to benchmark CSIRTs, but needs to be adapted.

Hagge and Kreutzkamp [27] present a method for benchmarking information systems. The procedure used to derive and describe the scenarios and the exercises is applicable in any benchmarking, and makes it easy to read the different scenarios, and what they include. This approach to describing the exercises was useful for the benchmarking experiment carried out in this thesis.

In [28], Marie and Büyüközkan present a fifteen steps benchmarking process. Their focus is on the first five steps of the process described, and they suggest some methods and tools to use in these initial steps. The explanations give insight into what actions to undertake in the benchmarking experiment carried out in this thesis.

Zairi [35] argues that benchmarking is best suited to measure competitiveness when it is used in a Total Quality Management (TQM) setting, and describes briefly a sixteen step benchmarking method. TQM is a management strategy focusing on enhancing process performance through a deep analysis and understanding of the process internal operation. Benchmarking can be used as a tool in TQM to compare processes between different organizations, to find the best way to perform the different activities. However, benchmarking can also give important information to the management, even if the organization is not committed to TQM. The benchmarking may point out to the management if the organization's processes perform better or worse than the organization it is comparing itself to. This is what Zairi calls a "quick dip approach" [36].

For this thesis, the identified work processes and the corresponding metrics serve as inputs to the benchmarking process. The benchmarking process, as described in [10], consists of five steps: Planning, Search, Observation, Analysis, and Adjustment. All these steps must be executed. The first step serves to identify which process to benchmark, and what metrics to use. Then the metrics are developed for the identified processes. Step two consists of looking for teams or units to benchmark against, and establishing a trusted relationship between the two teams. In this thesis the benchmarking experiment is based on questionnaires presented to several teams, and compares the performance between them.

3 Summary of claimed contribution

This thesis will identify the work processes in a CSIRT that is involved in incident handling, and the basic activities performed in the basic work processes will be described. This gives a common understanding of what activities the different work processes comprise, and clarify what the terms signify.

When the work processes have been identified and described, metrics for the performance of each process will be developed. The metrics will be based on the activities that each process comprises.

To show how the metrics can be used to rank CSIRTs, and to be able to give recommendations to the teams on how to improve the incident response performance, a benchmarking experiment will be conducted. The benchmarking experiment is based on a questionnaire that is sent to several organizations with a CSIRT capability. The questionnaire contains questions which are derived from the metrics. The answers given will be used to calculate the performance for each team.

4 Choice of methods

To be able to answer the research questions, different approaches are needed. Mainly qualitative methods will be used. Qualitative methods are described by Cresswell in [11]. The first part of the thesis is a literature study to identify and describe the different work processes within a CSIRT. The second part of the thesis is developing metrics for the work processes. The metrics will be described according to NIST SP800-55 [13). These metrics will be a quantitative method for measuring the efficiency of the team, and will be used in a case study where two teams will be benchmarked.

4.1 Work processes in CSIRT

To identify the different work processes in a CSIRT, a thorough study of available literature will give the necessary information about the work that has already been done to describe the work processes in a team. There are several sources describing the activities and services a CSIRT should or could perform. However, the different sources often use different terms on the processes and activities. This makes it necessary to describe the different processes in order to clarify what activities and services the processes actually comprise.

4.2 Measures and metrics

After identifying and describing the work processes, it is necessary to derive metrics for each single process. It is desirable to find quantitative measures for all processes, as this will simplify the benchmarking between teams.

Several sources describe how to develop and describe metrics. For this thesis the guide published by NIST will be used. The NIST guide has some limitations, and it will be necessary to add some fields to describe reliability and validity properties of the metric.

4.3 Benchmarking

A general methodology for benchmarking is presented in [10]. The identified work processes and the corresponding metrics will be input to the benchmarking process. The benchmarking process consists of five steps: Planning, Search, Observation, Analysis, and Adjustment. All these steps must be executed. The first step is to identify which process to benchmark, and what metrics to use. This will be identified through the literature study in the first part of the thesis, and the metrics developed for the identified processes.

Step two is searching for a unit or team to benchmark against, and establishing a trusted relationship between the teams. In this thesis the benchmarking experiment will consist of several participants. To find organizations that wanted to participate in the experiment, several large public and private organizations in Norway were contacted. The organizations were asked if they had established a CSIRT capability, and if they were interested in participating in a benchmarking experiment.

The observation phase consists of the measuring of the parameters in the teams when handling an incident. The measurement is based on a questionnaire that is developed from the different metrics. The answers given to the questionnaire is the actual measurement.

In phase four the answers are analyzed and the performance of the teams is calculated. The calculated result is used to rank the teams. The results will also reveal any relevant difference in the performance between the teams. It is necessary to identify what causes this difference, and what effort can be imposed to improve performance. If one chooses to implement changes to improve performance of the team, the adjustment phase starts. After adjusting procedures the benchmarking process can be repeated to see if the adjustment had the wanted effect on performance. In this thesis the benchmarking is not repeated, but recommendations on how to improve process performance is given.

5 Benchmarking CSIRT work process performance

This Chapter is divided into three subsections. Chapter 3.1, identifies and describes the basic CSIRT work processes related to the incident handling. Chapter 3.2, defines metrics for the work processes described in Chapter 3.1. In Chapter 3.3, the benchmarking experiment is described. The benchmarking experiment compares the work process performance of the CSIRTs, using the metrics defined in Chapter 3.2.

5.1 Suggested CSIRT work processes

There are several sources which describe the duties and services a CSIRT should or could perform, and how teams could be organized to solve their tasks in the best possible way. This means there is a lot of information that can be used in CSIRT processes identification, but different documents often use different terms on the processes and services, and do not always agree on what activities belong to which processes. The different sources also focus on different parts of incident handling, which partially depends on the authors' bias and area of interest. This makes it necessary to suggest a set of processes in order to clarify what the process or service actually comprises, and to describe the most common activities in incident handling. We have chosen to divide the overall incident response process into the five sub-processes shown in Figure 1. Different sub-processes will be described in more detail in the following chapters.

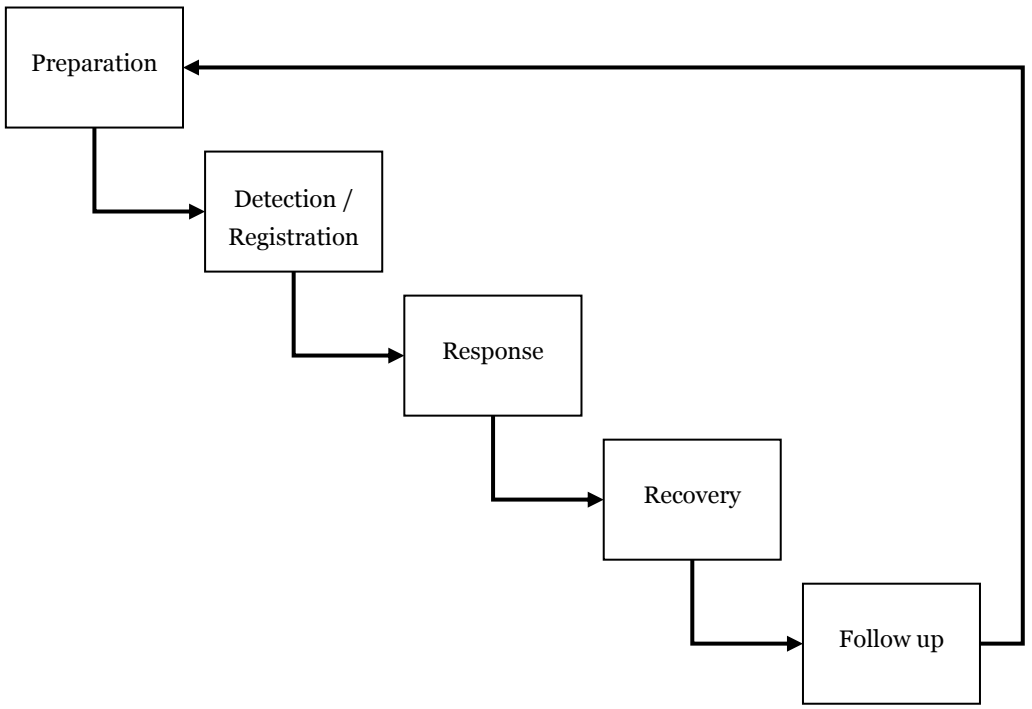

Figure 1: Overall incident response process

This description of work processes in a CSIRT is not intended as a guide to establishing a CSIRT. The planning and implementation of CSIRTs is described in detail in [3-6], [16-19]. The description of work processes in this chapter is intended to describe the most common activities in a CSIRT, and to serve as a basis for comparing different CSIRTs.

5.1.1 Preparation

Preparation can encompass several different activities, ranging from establishing a security policy to awareness training. Establishing policies and procedures are important, as they describe how incidents are supposed to be handled. Policies are an important way to inform employees/users about what acceptable use is and what is not. The response team also needs a policy for acceptable response and management reporting. Such a policy is a corner stone in the effective response team. A policy should also include a section concerning responsibility. Describing who is responsible for what action/process, and what decisions can be made by the different stakeholders, makes the work in the team easier and improves the performance. The incident response team requires a methodology. The methodology describes how the team is supposed to function, and which strategy is to be used for handling incidents.

The incident response team has to know the environment of the organization: what systems are implemented in the constituency, and what architecture is used. Mapping the environment makes the team members more able to identify where an incident could be contained or where a countermeasure could be deployed.

Awareness training is important to educate users in correct use of the systems, and what not to do. Awareness training also educates users in how to detect and respond/report suspicious behavior in the information systems. Establishing a simple way of reporting suspicious behavior is critical for an effective incident handling. Through awareness training and proper information, the organization establishes a security culture. Many organizations let the incident handling team do the security awareness training.

Evaluating the system security and implementing tools for monitoring or testing security are also a part of the preparation process. Having configured the system both for performance and security requires that trade offs are made. Knowing where trade offs are achieved, and what measures have been implemented to mitigate the risks are important information sources for the incident response team. Security evaluation may include vulnerability assessment and security audit. In a broad sense this evaluation might be called a risk analysis. A procedure for managing patches and security patches in particular, improves organization's security.

Establishing and maintaining internal and external contact lists, telling who is to be contacted in different cases, save time when an incident occurs, and thus improve overall security for the organization.

5.1.2 Detection

There are many different ways in which a CSIRT gets to know about an incident. Most incidents are reported, orally or in written form, by end users or system administrators. Some incidents are even reported by external contacts like business partners or cooperating CSIRTs. To get an efficient reception of these reports, it is vital that the CSIRT defines a well designed procedure to do this, and focuses all these reports to one point in the team.

Other ways to detect incidents are through intrusion detection systems, antivirus software or other security or management software. In [30], different approaches to detecting intrusions in computer systems are summarized, and in [31] it is described how honeypots can be deployed to give valuable information about malicious activity against a network, or collection of networks.
Personnel operating or administrating this software can belong to the CSIRT, or they may belong to other parts of the organization. If the personnel do not belong to the CSIRT, the incident is usually reported as described above, but if the operators belong to the CSIRT there is a need for a procedure describing how these incidents should be registered.

Having an application for registering reported and detected incidents is a useful tool. This application should allow the team members to register all necessary data about the incident, link to relevant reports or documents and write continuous comments about how the incident is handled.

All reported incidents should be registered by the team. Like in all processes the CSIRT encompasses, documentation is important through the whole process. Having a structured way to document an incident and the proper tool for saving the documents is crucial. This makes it easier to track what has been done with the incident, and search through previous incidents.

A separate activity will do a quick classification to check if the registered incident is actually a security incident. If it is not, the incident should be closed or reassigned to the appropriate system administrator. If it is a security incident it should be classified according to an established scheme for the organization. This classification will give information about how to handle the incident.

Incidents that are considered to be security incidents should be correlated with other events, to check if there are more incidents / events that are concerning the same system or service. While doing this, it is important to document what is done. If necessary, the chain of custody should be maintained, in case the incident is to be handed over to the police later, so no evidence is destroyed.

Figure 2 shows the details of the detection process. When these activities are completed, the incident is handed over to the response process.

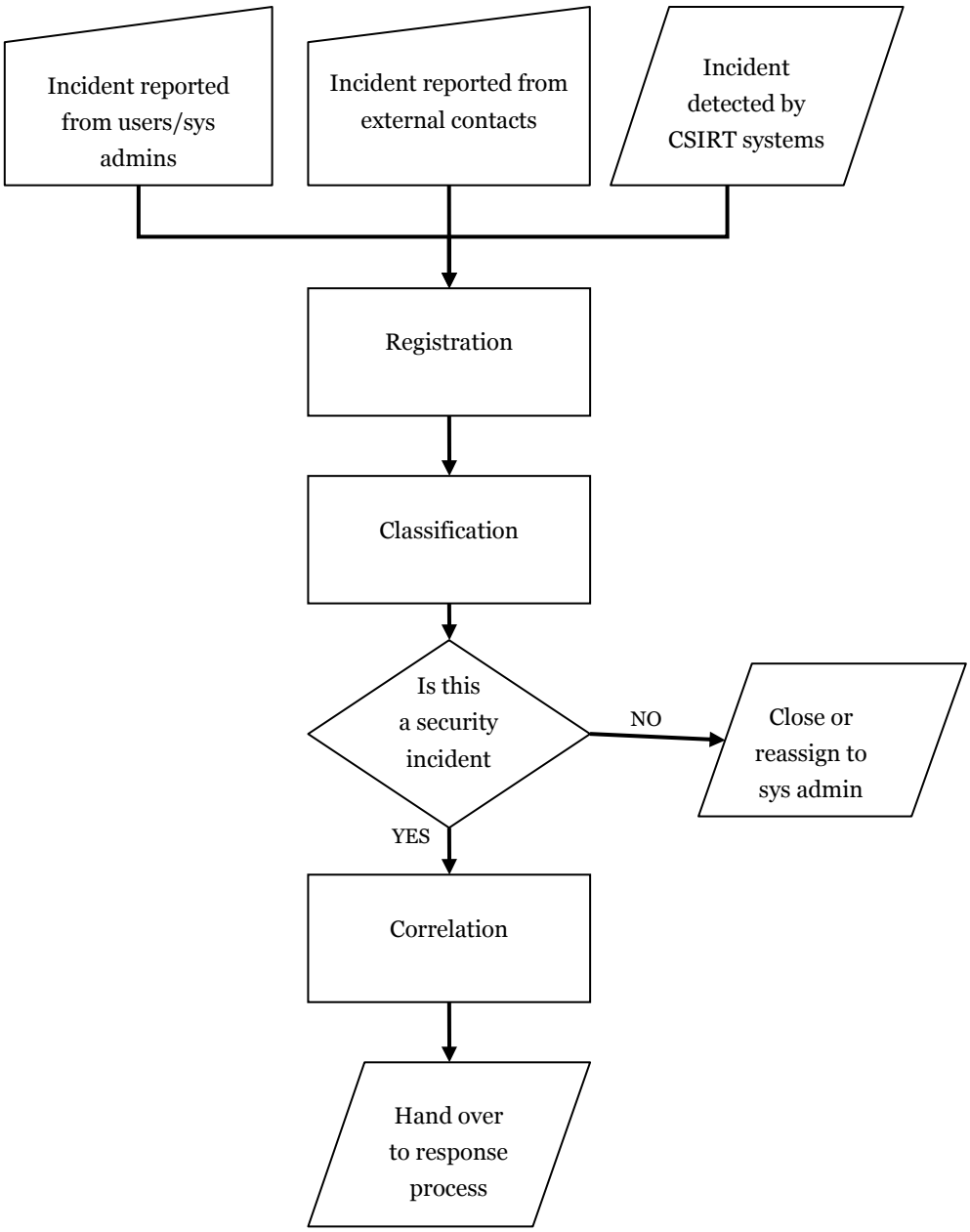

Figure 2: Detection process in detail

5.1.3 Response

The response process is built up of several activities. The detailed process is shown in Figure 3. An important task for the CSIRT is to inform management and other stakeholders. This should be done regularly and in all stages of the incident handling. A procedure should describe how often the management is to receive information, and set milestones where information always should be passed to the management. Other

stakeholders requiring information could be users, authorities or administrators. The procedure describing who is to get information must also describe which information the different stakeholders should receive, and the form the information should be presented in. For effective communication internally in the team and with external contacts it is important to have a common language to describe the incident. Howard and Longstaff together with the CERT/CC [22] have developed a baseline language with terms and taxonomy for computer security incidents that the CSIRT can use when communicating internally or with external contacts.

Initial response / Triage

In initial response the first activity must be to determine the type of incident, and categorize the incident. This will make it easier to assign the event to personnel with the correct expertise. This activity will be a more thorough classification than the one performed in the detection process. [21] presents a method for systematically classifying computer security incidents. It classifies intrusions both according to the technique used and the result of the intrusion. Another, and perhaps more extensive, classification is described in [24]. This taxonomy is more holistic, and consists of four dimensions: attack vector and the main behavior of the attack, attack targets, vulnerabilities and payloads. Finally [29] presents a defense-centric taxonomy based on attack manifestation that the authors claim is more effective predictor of the detector's ability to detect specific attacks. It is important to stress the need for the classification scheme for the CSIRT to use at communicating with both internal and external stakeholders.

Assessing the impact of the incident is crucial in order to be able to prioritize the incidents, by ordering them so that the incident which is supposed to cause the most damage is handled first. In this activity, it is important to include personnel with competence and of course management representatives. To make the assessment as correct as possible, all relevant data should be collected from available sources. What sources are available depends on the system and organization. The information that should be considered includes at least what systems are affected, their criticality, sensitivity of compromised information and what level of access did the attacker attain. The number of networks and hosts that are compromised will give important information, and help to decide where countermeasures could be applied. The information about what vulnerability is exploited, or if there are several vulnerabilities (attack vectors) that are exploited would make an important contribution to the assessment. If knowledge of existence of this vulnerability in other systems or hosts within the organization exists, this should be taken into account. An immediate increase in audit information capture is necessary to gather as much information as possible about the incident. To save data for future analysis a backup of the affected systems, or at least of the identified compromised files, is useful.

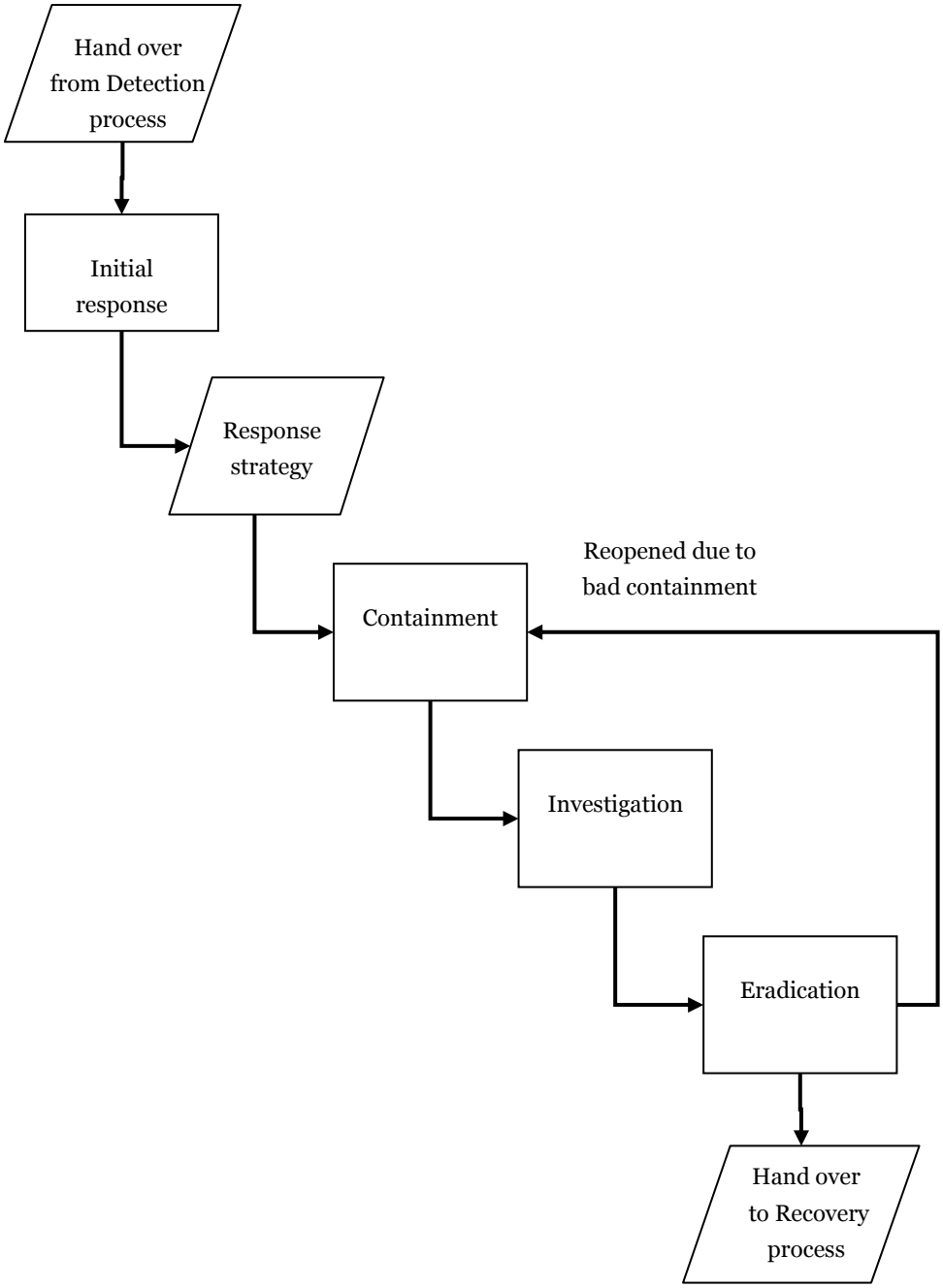

Figure 3: Response process in detail

The initial response activity ends with the development of a response strategy. This strategy must consider the totality. Business factors, legal issues, technical possibilities and public relations must all be considered when the response strategy is formed. If it is possible to get more information about the attack and/or the attacker, this would be valuable in defining the response. This could be information about who the attacker is, whom he/she is working for, what his/her skills are, is the attack pinpointed to our

system or are other organizations affected as well, and so on. A quick analysis of system anomalies and events contributes to this information. The response must be focused on minimizing overall loss, and keeping system and user downtime to a minimum. The strategy presented to the management must also describe the potential drawbacks of the chosen response.

Containment

By containment we understand activities contributing to deny further malicious or unwanted activities. These activities include denying or restricting access (i.e. deleting or disabling accounts, disabling services), blocking traffic (i.e. changing firewall rules), increasing monitoring, deploying decoy servers and in worst case disconnecting or shutting down systems. What activities or countermeasures are chosen depend upon the strategy developed in the previous phase.

When commencing in the containment activity one should always follow the response strategy. This is because the strategy is formed out of a complete picture of the organization. Not only technical factors are considered but also business factors, legal issues and others. Impact of the incident versus criticality of the system affected should be taken into account in the response strategy.

Automated tools, like the distributed system presented in [23] and the tools used in the experiment described in [25], are also popular. These can be tools that automatically reconfigure parts of the system, thereby responding much quicker than a manual reconfiguration. This can, however, be exploited by attackers as an efficient denial of service (DoS) attack. Launching a minor attack that makes this tool reconfigure the firewall to block what should be legitimate traffic is a quick way to make one launch a DoS attack on oneself. Another automated tool is the forced file integrity check. This tool, however, consumes CPU cycles, and may affect system performance.

Investigation

When the incident is contained within a host or a part of the network, it is time to do a more thorough investigation. The events related to the incident should be analyzed and as much information as possible should be gathered from the available sources. What one wants to find out is what happened. Of special importance is the following: what vulnerability was exploited, where and how? When did the different actions take place, and on which hosts? If possible it would also be very interesting to know who exploited the system and why.

The available sources can be of different types. Event logs, security logs etc. on the computers, firewall and IDS logs and, if available and needed, logs for access control systems. To gather this information there is a need for close cooperation with system administrators. It might also be necessary to use computer forensics expertise if available.

Eradication
When all relevant information is gathered for the investigation, cleaning up is necessary. Removing files that have been put on the system by the attacker is critical, as they may contain programs that can damage the systems in the future. The same holds for checking the configuration for changes, and correcting them, if necessary. The goal of these activities is that the attacker cannot benefit from a weakening in the system configuration at a later time. These activities are dependent on cooperation with system administrators.

5.1.4 Recovery

The main issues in the recovery phase are getting business up and going, and removing vulnerabilities exploited in the incident.

The major concern is getting business going, as this is the production of the organization. Businesses can bankrupt if information systems are unavailable for a period of time. Minutes can perhaps be tolerated, hours may cause serious damage and days means end of business. Rebuilding the systems and restoring data is vital for the organization. The systems should be rebuilt in a prioritized order to ensure that the critical systems are handled first.

When systems are up and running a careful review of the running services might identify that some are not necessary. These services should be stopped, as they open a new attack vector for an opponent, and are not used by the organization. Installing patches and corrections according to patch management procedures fix vulnerabilities, making it harder to exploit them. If it's decided to implement countermeasures, like security software or hardware, that should be applied in this phase.

When these steps have been taken it is time to reconnect the system. If it is necessary to reconnect the system before installing all the patches, corrections or security features, it is reasonable to assign responsibility for the process of implementing these corrections to one person, and then track progress in the process through regular reporting.

5.1.5 Follow up

After handling an incident, it is important to document the process, and the response chosen. The documentation makes it possible to trace the incident timeline, and the team response.

Every incident handled by the response team needs to be reported to the management with an assessment of the impact the incident caused in operations. The management would also like to know what has been done to reduce the risk of the incident happening again. It might also be necessary to inform the users about the incident, both by informing about the applied corrective measures, and by informing about the exploited vulnerability. Feedback to the department or user who reported the incident is always appreciated.

An internal evaluation of the incident handling is needed to review the procedures used and learning from the incident. Learning from handled incidents is one of the best sources of knowledge that members of the team can get. A thorough internal review of the incident increases the team's ability to handle incidents in the future, and also improves quality in future responses. If this evaluation finds that documentation, policy or procedures are not mitigating the threat as they should, a revision is needed. Updating documentation and procedures should take place immediately after the incident has been evaluated.

If the incident occurred through exploitation of vulnerability in software or hardware that is used in other systems within the organization, these systems should be audited to see if the same vulnerability has been exploited, or if the vulnerability is possible to exploit. If so, this should have generated a new incident.

5.2 Suggested measures and metrics

After identifying the basic work processes, it is necessary to develop metrics for each single process. It is desirable to find quantitative measures for all processes, as this will simplify the benchmarking between teams. In this thesis a unit called Incident Response Performance (IRP) is proposed and used for the metrics. The metrics calculate the IRP from the answers in the implementation evidence, using the formula given in the metric description. The IRP is expressed as a number in the range [0-1], where 1 is considered to be the best achievable result.

To calculate the IRP, it is necessary to convert the answers to the questions into numerical values. If the answer is a single *yes* or *no*, the value *yes* is assigned the value "1" and *no* is assigned "0". When the given answer has more options, a scale is used. These questions have five options, and the values for the options are 0, 0.25, 0.5, 0.75 and 1. "0" is assigned to the option that is supposed to be the worst performance, and "1" is assigned to the option that is supposed to be the best.

Some questions depend on other questions, i.e. there might be a yes/no question if a procedure exists, and then a follow up question to state how well the procedure is known. For this type of questions the value for the first question is multiplied with the value for the second question in order to obtain the result. Other questions get the answer as a time or a percentage. The percentage is easily converted by dividing by 100, thus getting the value in the right range. When getting time as the answer more calculations are needed. As it is impossible to find the maximum value for the time use for the teams, we use the measure values to create a reference value. First, it is needed to find the biggest and the smallest time value for the given question. We call these values MAX and MIN. To normalize this parameter, MAX and MIN are summed together giving the reference value of the parameter. To calculate the parameters contribution to the IRP, the answer is divided by the reference value, giving a normalized value between 0 and 1. For most questions, the best performance is the one with the lowest time consumption. To achieve the right value for the IRP in these

cases, it is necessary to subtract the normalized value from 1. This produces the best result, i.e. the lowest time consumption, to get the value closest to 1.

When all these values have been calculated, they are summed. The sum is then divided by the number of values to give the IRP.

Yee argues in [34] that finding one single value for security ranking or a single security partial ordering is not useful in practical security work, because making the measurements is very difficult as the target changes continuously. [34] suggests that a multi-faceted or multi-dimensional security measure is more useful. One can agree with [34] that finding one single value for measuring security is perhaps very difficult. However, in this thesis trying to find one single value to describe the performance of the incident handling capability of an organization, does not describe the organization's security state. The incident handling performance value is one of the dimensions in a multi-dimensional security measurement for the organization.

The suggested metrics will describe the performance of each single process. As the processes will differ from organization to organization original processes are suggested that comprise activities that almost every CSIRT perform. [13] suggests a form for documenting and describing metrics. This source is used here as a basis for the description of the metrics, and a sample metric is found in Table 1. In [37], Mathisen adds two fields to the NIST metric description form. These two fields contain comments to validity and reliability. This is adopted here, and the validity field and the reliability field have been added to the metric definitions.

Table 1: Metric definition template

Metric number	Unique identifier for the metric, expressed by a capital letter and a number. Example: A-2, identifying main area A and metric number 2.
Critical element	The critical element states the name of the metric.
Unit/Metric	States the unit the metric will use.
Purpose	Purpose states the purpose of this metric.
Implementation Evidence	Implementation evidence lists what security controls to check and how to check them.
Frequency	Frequency is a suggestion to how often the metric should be calculated in a continuous improvement program.
Formula	Contains the formula needed to calculate the value of the metric.
Data Source	Data source states where data can be found.
Indicators	A description of the metric and how the metric should be analyzed.
Validity	Evaluation of the validity of the metric. Is the metric measuring what we want to know?
Reliability	Evaluation of the reliability of the metric. Will the result show the same value if we measure again, or is it possible to get incidental errors?

5.2.1 Preparation process performance metric

The preparation process performance metric is based on checking if the organization has policies and procedures for testing and improving security. It is not enough to check if these policies and procedures exist, it is also of interest to examine how well these policies are known in the organization, and if the policies and procedures are usually used. The metric is defined in Table 2.

Table 2: Preparation process performance metric

Metric number	A-1
Critical element	Preparation process performance
Unit/Metric	IRP
Purpose	Evaluate the overall performance of the preparation for security incident handling within the organization
Implementation Evidence	**A.** How often is information security tested? According to the policy ☐ ☐ ☐ ☐ ☐ Never **B.** Does your organization have a procedure for testing and implementing patches in information systems (patch management)? Yes ☐ No ☐ **C.** How well is this procedure known in the organization? Well known ☐ ☐ ☐ ☐ ☐ Known only by a few **D.** Does your organization have a policy for handling computer security incidents? Yes ☐ No ☐ **E.** How well is this policy known in the organization? Well known ☐ ☐ ☐ ☐ ☐ Known only by a few **F.** Is there a policy describing who to contact when a security incident is detected? Yes ☐ No ☐ **G.** How well is this policy known in the organization? Well known ☐ ☐ ☐ ☐ ☐ Known only by a few

	H. Does the incident handling team have a pre-produced contact list? Yes ☐ No ☐ **I.** How often is this contact list checked and updated? Weekly ☐ ☐ ☐ ☐ ☐ Never ☐ Don't know
Frequency	Every six months
Formula	(A+B*C+D*E+F*G+H*I)/5
Data Source	CSIRT
Indicators	As the organization develops policies and these become known throughout the organization the score will increase. If the personnel handling the incidents have a contact list, this is good, but the list needs to be updated frequently. If this is done, the score will increase. All results are normalized to meet the IRP unit with a value in the range [0-1]. A natural development for an organization is an increase in the score as information security and incident handling mature, and more and more security controls are implemented. The best performance is the score of 1.
Validity	The parameters checked in this metric contribute to the performance of the team if the controls are in place. It is probably a lot more parameters that can contribute to this, to make the measurement more accurate. However, for practical reasons, only a few parameters are used in this metric. Other possible parameters to use could be security awareness in the organization, skill of the CSIRT personnel and so on. Although, with the parameters used the validity of the metric should be good.
Reliability	The collection of data measurement is obtained through asking questions to personnel. The reliability is dependent on the subject answering the same question each time. This can be achieved through a careful development of the questions. With the questions used in this metric the reliability should be good.

5.2.2 Detection process performance metric

The metric for the detection process performance is based on checking how the incidents are detected and registered by the organization. The existence of procedures for registering and documenting incidents within the organization affect the performance because these procedures state what information is needed in the different steps, and how this information should be formatted. A person taking over the investigation of an incident then has structured documentation from the detection up to the latest events.

A concern is the quality of reports that the CSIRT get. The quality of the reports, depend upon the right events being reported. To ensure the right events is reported, it is necessary to raise the security awareness in the organization, and educate users in what to report. If an IDS is deployed, it is important that the sensors are placed in a way that they will detect security events in the system. To find measure for the quality of the reported events has not been the focus of this thesis.

Classification of an incident makes it quicker to find the right personnel to handle the incident, and assigning the right resources.

Time is also crucial in the detection process. It is important that the registration and classification do not delay the work too much, because the incident may escalate in time and cause more damage. This is the whole idea of responding to incidents - not giving the opponent enough time to spread and destroy information and infrastructure.

Table 3: Detection process performance metric

Metric number	A-2
Critical element	Detection process performance
Unit	IRP
Purpose	Evaluate the overall performance of the detection and registration of security incidents within the organization
Implementation evidence	**A.** How long is the average time between the moment when the incident is discovered (the team gets the call) and the moment when the incident is registered and assigned to a team member? Time:_____ **B.** Does the incident response team have a procedure for registering new incidents? Yes ☐ No ☐

	C. How often is this procedure used when registering incidents? Always ☐ ☐ ☐ ☐ ☐ Never **D.** Does the team have a procedure that defines how to document an incident? Yes ☐ No ☐ **E.** How well do you think this procedure describes the documenting process? Very well ☐ ☐ ☐ ☐ ☐ Poorly **F.** Has the team established a classification scheme for incidents? Yes ☐ No ☐ **G.** How well does the classification scheme cover different types of incidents? Covers all incidents ☐ ☐ ☐ ☐ ☐ Covers only a few incidents **H.** What is the average time from the moment when the incident is assigned to the moment when it is classified? Time:_____
Frequency	Every three months
Formula	$((1-A/(A_{MIN}+A_{MAX}))+B*C+D*E+F*G+(1-H/(H_{MIN}+H_{MAX})))/6$
Data source	CSIRT
Indicators	When incident handling is new to an organization, it is likely that several of the controls checked by this metric are not in place. As the organization handles more incidents, the need for formalized ways to detect/register incidents and document them becomes clearer. As these controls are implemented the score increases. With the controls implemented and more experienced personnel the time used decreases, causing the score to increase. A well designed classification scheme is an important tool when describing incidents. A team that has just

	been established may have a classification scheme, but the scheme does probably not cover all the incidents that may occur in the future. A classification scheme develops over time, thus increasing the score. The best performance is the score of 1.
Validity	The parameters checked in this metric contribute to the performance of the team if the controls are in place. Several sources, like [3], [5] and [17], point out that a structured way of documenting the incident handling is important. Checking if organizations do this, supports the validity of the metric. Time indicates how fast the incident handling is progressing, but it does not state the quality of the work. It is possible to handle incidents quickly, but with poor quality. This weakens the validity, but if the other controls are implemented, it is likely that the incident handling has good quality in this process.
Reliability	The collection of data measurement is obtained through asking questions to personnel. The reliability is dependent on the subject answering the same question each time. This can be achieved through a careful development of the questions. The questions asking for a time estimate may cause the reliability to decrease. Answers to these questions are dependent on accurate registration of time used to complete the different tasks. The different teams may include different activities into the processes, and the time measured in one team may then differ from the time measured in another team because they include different activities in the process.

5.2.3 Initial response process performance metric

The main purpose of the initial response process is to develop a response strategy. This strategy should give the personnel handling the incident enough information to make the correct decisions when investigating the incident. The preformance in this activity is dependent on the team having policies that clearly state what priorities the management have regarding different types of incidents, and the understanding of how the impact of different types of incidents will affect the organization's business operations. If the organization previously has handled similar incidents and documented them, or has prepared procedures for handling this type of incidents, this would significantly improve the organization's performance in handling the specific incident.

Establishing a common language for incident handling reduces misunderstanding among team members, and between the team and the management. This reduces the possibility of wrong actions during the incident handling, and improves performance.

The time spent preparing a response strategy, is reduced if policies and procedures are prepared, and therefore gives an indication on the performance of the team. Unclear policies and diffuse procedures need to be addressed in this stage of the incident handling, and therefore require longer time to develop a response strategy.

Table 4: Initial response process performance metric

Metric number	A-3
Critical element	Initial Response process performance
Unit	IRP
Purpose	Evaluate the overall performance of the initial response to security incidents within the organization
Implementation evidence	**A.** Does the team have written procedures for how to handle different types of incidents? Procedures exist for all types of incidents ☐ ☐ ☐ ☐ ☐ Procedures do not exist **B**. Is there a policy enabling the team to set the severity of an incident? Yes ☐ No ☐ **C.** Does the team have a scheme for prioritizing incidents? Yes ☐ No ☐ **D.** Has the team established a common language for describing incidents? Yes ☐ No ☐ **E.** How useful do you think this common language is for communicating inside the team? Very useful ☐ ☐ ☐ ☐ ☐ Not useful **F.** What is the average time to develop a response strategy from the time of detection? Time:_____
Frequency	Every three months

Formula	$(A+B+C+D*E+(1-F/(F_{MIN}+F_{MAX})))/5$
Data source	CSIRT
Indicators	Having implemented the security controls improves the team's performance, and increases the score. If the team has developed procedures for all kinds of incidents this improves performance significantly. Having developed procedures and implemented security controls decreases the time used to develop a response strategy, thus enabling the team to move to the next step in the incident handling process. The desired goal for the metric is 1.
Validity	Implementation of the controls has a great impact on performance of the initial response activity. The metric only checks if a policy is present, and if there is a scheme for prioritizing incidents. The quality of the policy and the scheme are not measured. This decreases the validity of the metric, as the quality affects the performance of the process. Establishing a common language for incident handling improves internal and external communication when handling an incident. However, this might not be needed if the personnel are experienced, and have worked together for a long time. This weakens the validity. Time indicates how fast the incident handling is progressing, but it does not state the quality of the work. It is possible to develop a response strategy quickly, but with poor quality. This weakens the validity, but together with the other controls, it gives an indication on the performance of the process.
Reliability	The collection of data measurement is obtained through asking questions to personnel. The reliability is dependent on the subject answering the same question each time. This can be achieved through a careful development of the questions. The questions asking for a time estimate may cause the reliability to decrease. Answers to these questions are dependent on accurate registration of time used to complete the different tasks. Different teams may include different activities into the processes, and the time measured in one team may then differ from the time measured in another team because they include different activities in the process, thus decreasing the reliability.

5.2.4 Containment process performance metric

The performance of the containment process is very much dependent on the skill of the personnel carrying out the incident handling. This is, however, very difficult to measure on a general basis. Some teams may be very skilled in handling incidents in a Unix or Linux environment, but be lost handling incidents on systems running Windows. The skill is also very dependent on earlier experience. Measuring this "skill" does not give an indication on how the team performs when handling incidents generally.

There are, however, some measures that can give a general indication on the performance. How detailed and useful the strategy developed in the initial response activity is an indication of containment. The response strategy should be as detailed as possible, but this is, however, not a guarantee that the strategy is effective. If the strategy only addresses the main areas of the investigation, something may be skipped and not taken care of. With a detailed strategy the time to develop the strategy may be longer, but the chance of getting a quick and complete response to the incident is better.

Automated tools for incident response are getting more and more popular. These tools can be very effective and give a quick response. On the other hand, if these tools make wrong decisions, the organization may face losses in the form of money or reputation. The true/false ratio will therefore influence the performance of this response.

The time spent containing an incident is a good indicator on the performance of the incident handling. Skilled personnel, a good response strategy and automated tools make it possible to achieve a quick containment of the incident and a good basis for further investigation. A quick containment is however not a guarantee for good containment. It is therefore important to consider the ratio of reopened incidents due to bad containment. If the containment activity is not performed well, the incident spreads beyond the contained parts, and it is necessary to contain the incident once more.

Table 5: Containment process performance metric

Metric number	A-4
Critical factor	Containment process performance
Unit	IRP
Purpose	Evaluate the overall performance of the containment of security incidents within the organization
Implementation evidence	**A**. How detailed is the response strategy? All details are described ☐ ☐ ☐ ☐ ☐ Only main areas are addressed

	B. How effective do you find the chosen strategy (on average)? Very effective ☐ ☐ ☐ ☐ ☐ Ineffective **C.** Does the organization use automated response tools? Yes ☐ No ☐ **D.** What is the true/false ratio for the automated tools? _____% **E.** What is the average time to contain the incident? Time:_____ **F.** What is the ratio of reopened incidents caused by bad containment? _____%
Frequency	Every three months
Formula	$(A+B+C^{*}D+(1-E/(E_{MIN}+E_{MAX})+(1-F))/5$
Data source	CSIRT
Indicators	A detailed response strategy is likely to increase performance. However if the personnel doing the incident handling find the strategy to be ineffective, the performance decreases. A detailed response strategy that the personnel find effective is desirable, but a mix in detail and effectiveness is acceptable. If automated tools are used, the response is quick, and performance increases. But if the true/false ratio for the automated tools is low, the performance decreases. The time used to contain the incident should be as short as possible and small values for the time estimate increase performance. But if the quality of the containment is bad, it is necessary to reopen the case, and perform the containment process over again. If the ratio of reopened cases increases the performance of the process decreases. The best performance is the score of 1.

Validity	The level of detail in a response strategy, and the effectiveness of this strategy are very subjective. A skilled operator would perhaps like a strategy without much detail, and still find it effective, as a less skilled operator might want a very detailed response strategy, and still find it ineffective. This decreases the validity of the metric. The presence of a strategy and the effectiveness of automated tools will give an indication of the effectiveness in the containment process. In addition, the time used to contain the incident is related to the performance, and the ratio of reopened cases will give an indication of the quality of the work done in the containment process, thus increasing the validity of the metric.
Reliability	The collection of data measurement is obtained through asking questions to personnel. The reliability is dependent on the subject answering the same question each time. This can be achieved through a careful development of the questions. The true/false ratio of an automated tool is easy to obtain, and therefore strengthens the reliability. The questions asking for a time estimate may cause the reliability to decrease. Answers to these questions are dependent on accurate registration of time used to complete the different tasks. The different teams may include different activities into the processes, and the time measured in one team may then differ from the time measured in another team because they include different activities in the process, thus decreasing the reliability. It can be difficult to tell if the reopening a case is due to bad containment or if the incident is in fact another attack. This is decided manually when reopening the case, and depends on the person investigating the case. This decreases the reliability of the metric.

5.2.5 Investigation process performance metric

After the containment is completed there is a need to investigate the incident more thoroughly. This investigation is necessary to identify what vulnerability was exploited, how the vulnerability was exploited, who was the attacker and what was the motive. This is important information to prevent future incidents. The ratio of the incidents where this could be identified is used to measure the performance of this activity.

The time used to investigate also gives a performance indication, as the management usually wants the investigation to be quick since the systems need to be restored and put back into production. These performance measures may work against each other, as it is necessary to use more time to identify the vulnerability and the attacker.

Table 6: Investigation process performance metric

Metric number	A-5
Critical factor	Investigation process performance
Unit	IRP
Purpose	Evaluate the overall performance of the investigation of security incidents within the organization
Implementation evidence	What is the percentage of the incidents in which the following could be identified: **A.** What vulnerability was exploited? _____% **B.** Could the exploit be isolated? _____% **C.** Could attacker be identified? _____% **D.** Could motivation be identified? _____% **E**. What is the average time to investigate an incident? Time:_____
Frequency	Every three months
Formula	$(A+B+C+D+(1-E/(E_{MIN}+E_{MAX})))/5$
Data source	CSIRT
Indicators	The goal of the investigation is to establish what vulnerability was exploited and how it was exploited. This in order to be able to patch the system to avoid similar incidents in the future. As the ratio of incidents where this can be identified increases, the performance improves. It is important to identify the attacker and his/her motivation in order to remove the cause for the attack. An improved ratio

	of identified attackers and motivation increases the performance of the process. The time used to investigate the incident should be as short as possible and small values for the time estimate increase the performance. But if the quality of the investigation is poor, it is not possible to correctly identify the vulnerability, the exploit, the attacker and the motivation. The ideal team uses a short time to investigate and correctly identify the other parameters, but a more inexperienced team uses longer time to identify the parameters. The desired score is 1.
Validity	The goal of the investigation is to establish what was exploited and how, together with who did it and why. Measuring the ratio of these parameters gives a good validity of the performance. To measure the time might weaken the validity, as with some incidents it is vital to find out what was exploited and who did it, thus requiring longer time to investigate. With other incidents, the organization's priority might be restoring the systems to minimize the downtime of business operations.
Reliability	The reliability for this metric depends on the organization's registration of incident data, and the quality of the incident handling. If the organization is able to correctly identify the values for the parameters used in the metric, the reliability is good. However, it is often difficult to correctly identify attacker and motivation, as attacks are often launched over the Internet, resulting in a poor reliability for this metric. The questions asking for a time estimate may cause the reliability to decrease. Answers to these questions are dependent on accurate registration of time used to complete the different tasks. The different teams may include different activities into the processes, and the time measured in one team may then differ from the time measured in another team, because they include different activities in the process, thus decreasing the reliability. It may also be difficult to tell exactly when the investigation is ended. New information may appear, making it possible to continue the investigation, after the incident is handed over to the eradication process.

5.2.6 Eradication process performance metric

It is difficult to find a performance parameter for the eradication activities, as the activities are very system and incident dependent. But time is as always crucial. It is important for the organization that the systems affected by the incident are put back into production as soon as possible. To measure the overall time used from the moment when the incident is detected to the moment when the eradication activity is finished gives an indication on the performance.

Table 7: Eradication process performance metric

Metric number	A-6
Critical factor	Eradication process performance
Unit	IRP
Purpose	Evaluate the overall performance of the removal of security incident cause within the organization
Implementation evidence	**A.** What is the average time from the moment an incident is detected until the system is healthy again? Time:_____
Frequency	Every three months
Formula	$1\text{-}A/(A_{MIN}+A_{MAX})$
Data source	CSIRT
Indicators	The time used to handle the incident from detection to eradication, indicates the performance. An organization wants the time to be as short as possible, as this indicates low downtime for the business operations. A short time gives a higher performance score.
Validity	Measuring time is often giving poor validity, as there are many factors affecting the usage of time through the incident handling process. If the teams using the metric have the same activities defined in their incident handling processes, the validity increases.
Reliability	To get good reliability, it is important that all teams registering this information have the same understanding of when to start the time registration, and when to end the registration. This is difficult to achieve, and the reliability is considered poor.

5.3 Benchmarking experiment

5.3.1 Preparation

Benchmarking is a method for improvement. The aim of benchmarking is to compare the system's own performance or practice with other organizations having the same or similar processes, and then learn from those performing better. In [10], we find an almost philosophical definition of benchmarking:

> "Benchmarking is the art of being humble enough to admit that there is someone who is better than you, and at the same time being wise enough to learn to become equally good or even better."

Benchmarking is a supplement to the continuous improvement process in the organization. The improvement process usually has an internal focus with improvement ideas from employees or quality circles. Benchmarking on the other hand requires the focus to be external, looking for how other organizations perform. This assumes that there are organizations out there performing better.

Several different models exist for conducting a benchmarking experiment, but they all contain the basic activities described at different levels of detail and different number of steps. To conduct the benchmarking experiment we have chosen to use the benchmarking model described by Andersen & Pettersen in [10]. This because we had some earlier experience in using this model. The model contains five steps that need to be completed. The five steps are: Planning, Searching, Observation, Analysis and Recommendations. These are described in detail in the following sub-chapters.

5.3.2 Planning

The planning stage of a benchmarking process is the most crucial part of the benchmarking according to [10]. The planning includes deciding on what process to benchmark, establish a benchmarking team, document the process that is to be benchmarked and develop metrics for the process. This thesis demonstrates how benchmarking can be used to evaluate the performance of incident handling teams, thus the process to benchmark is the incident handling process. We have also chosen to limit the study even more, and concentrated on the preparation, detection and response activities in the incident handling process.

The benchmarking experiment was conducted by the author alone, and was based on responses to a questionnaire.

Documentation of the process to be benchmarked was presented in Chapter 3.1 in this thesis. It is a general documentation of the incident handling process, and describes the general activities that a CSIRT needs to carry out. This description is not intended to describe the complete list of activities that a CSIRT may conduct, but describes the most common activities that one can find in almost every CSIRT.

The final activity in the planning stage is developing metrics for the incident handling processes. The metrics could be based on quality, time, cost or any other quantifiable

parameter, or several parameters. The metrics that will be used in this experiment are described in Chapter 3.2.

5.3.3 Searching

In this stage the aim is to find a suitable partner to benchmark against. First, one must define the requirements that possible benchmarking partners should satisfy, and then search for as many candidates as possible. The search should give a list of possible candidates. From the list, one or more candidates should be communicated. To get a good benchmarking process it is vital that the benchmarking partners share the same goal for the benchmarking, and that there is a high level of confidence between the partners.

For this thesis, several large organizations in Norway have been communicated for a possible contribution in the benchmarking experiment. The choice of communicating large organizations is based on the fact that smaller organizations and companies seldom have resources to establish a CSIRT, and more often depend on their Internet Service Provider (ISP) to handle security incidents.

To avoid problems with information being "company confidential", the benchmarking was based on information the organizations most likely can release. By choosing this approach to the benchmarking process the demand for a high level of trust is not that important, making it easier to get answers from the organizations. In this experiment we have also chosen to keep the different organization anonymous, to avoid unfavorable information being related to them.

The organizations were first contacted by e-mail asking for a point of contact for a benchmarking experiment as part of this research. Those organizations that replied with a point of contact were contacted again by sending a questionnaire to the contact. A total of 16 organizations were contacted initially, of these 7 answered with a point of contact.

5.3.4 Observation

In the observation stage the aim is to observe how the benchmarking partners conduct their work. There are several methods and techniques that can be used. For the thesis we have chosen to use questionnaires, as this would save time and resources in conducting the benchmarking. Other suitable methods are on-site observation of the incident response team or interviews with CSIRT personnel.

The downside of using a questionnaire is that the different persons answering the questionnaire may understand the questions differently. An on-site observation or an interview would give more accurate results. To minimize this problem, the questions asked in the questionnaire need to be carefully selected. The validity and reliability of the questions are crucial to the results of the study. The questions used in the questionnaire are derived from the metrics developed in Chapter 5. The questionnaire is presented in Appendix A.

The questionnaire was sent out to the organizations who reported a point of contact. Of these organizations 6 returned the completed questionnaire.

5.3.5 Analysis

In this stage the collected data are sorted, normalized and analyzed. The main objective in this stage is to identify gaps between the participating organizations, and if possible identify why there is a gap in the performance.

Tables 8-13, present the measured values for the processes in the following order: Preparation process, Detection process, Initial Response process, Containment process, Investigation process and Eradication process. The last column in each table contains the calculated IRP values for the respective process. These values are obtained by using the formulas from the metric descriptions. Some questions have a value outside the range [0-1]. These questions are time estimates, and have the unit "hour". This applies to the questions A and H in Table 9, question F in Table 10, question E in Table 11, question E in Table 12 and question A in Table 13.

Table 8: Measured values for the Preparation process

	Question									
	A	B	C	D	E	F	G	H	I	IRP
Team A	1	1	1	1	1	1	0,75	1	0,75	0,9
Team B	0,5	1	0,25	1	0,25	1	0,5	1	0,25	0,35
Team C	0,25	0	0	1	0,25	1	0,5	0	0	0,2
Team D	0,5	1	0	1	0,25	1	0,25	1	0,25	0,25
Team E	0,75	0	0	1	0,5	1	0,5	0	0	0,35
Team F	0,5	1	0,75	1	0,5	1	0,5	1	0,25	0,5

Table 9: Measured values for the Detection process

	Question								
	A	B	C	D	E	F	G	H	IRP
Team A	0,5	1	1	0	0	0	0	0	0,40
Team B	24	1	0,5	1	0,75	1	0,75	2	0,28
Team C	0,5	1	0,75	1	0,5	0	0	0	0,45
Team D	0,5	0	0,25	1	0,5	0	0	0	0,30
Team E	1	1	0,75	0	0	1	0,5	2	0,36
Team F	0,5	1	1	0	0	1	0,5	1	0,46

Table 10: Measured values for the Initial response process

	Question						
	A	B	C	D	E	F	IRP
Team A	0,75	1	1	0	0,75	1	0,73
Team B	0,75	1	1	0	0,75	12	0,57
Team C	0,25	1	0	0	0,25	3	0,40
Team D	0,5	1	0	0	0,5	2	0,47
Team E	0,5	1	1	0	0,5	1	0,68
Team F	0,75	0	0	1	0,5	1	0,43

Table 11: Measured values for the Containment process

	Question						
	A	B	C	D	E	F	IRP
Team A	0	1	0	0	4	0	0,56
Team B	0,25	0,25	0	0	2	0,05	0,47
Team C	0,25	0,5	0	0	18	0	0,37
Team D	0,5	0,25	1	0	4	0	0,51
Team E	0,75	0,75	0	0	4	0	0,66
Team F	0,25	1	0	0	8	0	0,57

Table 12: Measured values for the Investigation process

	Question					
	A	B	C	D	E	IRP
Team A	0,5	0,5	0,5	0,5	4	0,57
Team B	1	0,5	0,5	0,5	2	0,68
Team C	0,5	0,5	0,5	0	12	0,41
Team D	0,8	0,8	0,7	0,1	24	0,50
Team E	0,75	0,75	0,5	0,25	8	0,59
Team F	0,8	0,8	0,3	0,1	12	0,51

Table 13: Measured value for the Eradication process

	Question	
	A	IRP
Team A	2	0,92
Team B	2	0,92
Team C	24	0,08
Team D	18	0,31
Team E	12	0,54
Team F	12	0,54

To compare the overall performance of the incident handling process, the results for each team are put into a multi-dimensional vector, starting with metric A_1 as the first dimension (Table 8) and continuing up to metric A_6 as the last dimension (Table 13). We have then got a six-dimensional sample-vector:

$V = (A_1, A_2, A_3, A_4, A_5, A_6)$

When putting the calculated IRP values (Table 8-13) for each team into vectors, the sample vectors were as follows for the different teams:

Team A: $V_A = (0.9, 0.4, 0.73, 0.56, 0.57, 0.92)$
Team B: $V_B = (0.35, 0.28, 0.57, 0.47, 0.68, 0.92)$
Team C: $V_C = (0.2, 0.45, 0.4, 0.37, 0.41, 0.08)$
Team D: $V_D = (0.25, 0.30, 0.47, 0.51, 0.5, 0.31)$
Team E: $V_E = (0.35, 0.36, 0.68, 0.66, 0.59, 0.54)$
Team F: $V_F = (0.5, 0.46, 0.43, 0.57, 0.51, 0.54)$

The ideal team would have a sample vector like this:

Ideal team: $V_I = (1, 1, 1, 1, 1, 1)$

We can compare the teams with each other, to find the relative distance between the teams, or it is possible to compare each team to the vector of the ideal team. We have chosen to compare each team to the ideal team, and use the Euclidian distance to compare the sample vectors. The Euclidian distance is given by:

$$E(X, Y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

The two vectors to compare are defined as **X** and **Y**. The dimension of the vectors is denoted by *d*. When we compare the teams to the ideal team the vector for the ideal

team is **X**, and the sample vectors for the teams is **Y** for each calculation. This gives the following results:

$E(V_I, V_A) = 0{,}91$
$E(V_I, V_B) = 1{,}23$
$E(V_I, V_C) = 1{,}70$
$E(V_I, V_D) = 1{,}52$
$E(V_I, V_E) = 1{,}19$
$E(V_I, V_F) = 1{,}23$

We see that the team A is the team with the score closest to the ideal team. Team C is the team with the worst performance of the teams. If we group the results that are close to each other we get four groups. Group I with only team A, group II with teams B, E and F, group III with team D and group IV with team C. This can be used as a classification of the team performance.

Group I: Teams in this group have a good incident response performance
Group II: Teams in this group have an average incident response performance
Group III: Teams in this group have an incident response performance below average
Group IV: Teams in this group have a poor incident response performance

5.3.6 Recommendations

For the preparation process we see that the team A performs significantly better than the other teams, scoring 0.9 and the other teams scoring from 0.2 to 0.5. This indicates that the team A has implemented more security controls in this process than the other teams, and that the policies and procedures implemented are well known to the organization. To improve their own performance, the other teams should try to learn from the team A, and implement more security controls. Generally the team A scores better than the other teams on the ability to make policies and procedures known in the organization.

In the detection process the results do not differ as much as for the preparation process with results spread out evenly from 0.28 for the team B to 0.46 for the team F. The teams C and F score almost equal (0.45 for the team C and 0.46 for the team F). The teams B and D perform slightly worse than the other teams (0.28 for the team B and 0.3 for the team D). All the teams have the potential to improve by implementing procedures and classification schemes. The cause for the poor performance of team B is a significantly longer registration and classification times than the other teams.

In the initial response process, the results differ more. The results vary from 0.40 for the team C, to 0.73 for the team A. It seems obvious that the teams C, D and F have a potential for improvement. Most of the teams have not established a common language for incident handling. Implementing a common language would prevent misunderstanding and improve performance in the incident handling process. The teams C, D and F also lack a scheme for prioritizing incidents. Implementing such a

scheme is likely going to improve performance, at least if the team is handling several concurrent incidents.

For the containment process, the difference between the best team (the team E with 0.66) and the worst team (the team C with 0.37) is 0.29. The teams A, E and F are performing evenly with scores from 0.56 to 0.66. Only one team has implemented an automated response tool. The team do, however, not measure the true/false ratio of the tool, and the score is not affected. Implementing automated tools can speed up the response, but performance of an automated response depends heavily on the true/false ratio of the tool, and knowledge of this ratio is important to be able to evaluate the performance of the tool. To improve performance in this process most teams can improve the level of detail and the usefulness of the response strategy. The team C also suffers from a significantly longer containment time than the other teams.

In the investigation process the difference is between the worst performance, the team C with 0.41, to the best performance, the team B with 0.68. The results of the investigation depend on the experience and education of the personnel investigating the incident. With more experience and education, it is likely that the organization improves the results. The time spent investigating the incident also affects the result, as it is likely to improve the results by using more time to investigate. However, this does not conform to management requirements to get systems back in operation. From the questionnaires we see that the teams that use more time to investigate get low scores for this process. The same teams do not score better than the other teams on the results of the investigation. This might be caused by inexperienced personnel.

The biggest difference in performance is for the eradication process where the team A and the team B are the best with a score of 0.92, and the team C is far behind with the score of 0.08. For this process, only one security control was checked. If the teams have significant difference in performance for this control, this reflects in the performance score for this process. As the control checked is time estimate, a difference in how this time is registered might cause significant difference in the reported result. The time to clean the system is dependent on the experience and competence of the personnel handling the incident. To improve score, the organization can educate and train incident handling personnel.

By studying the sample vectors we can clearly see that the team A performs well for all the six processes, perhaps with a possibility of improving performance in the detection process. The team C is the team with the worst performance. In all processes except the detection the team C has the poorest performance. In the detection process the performance of the team C matches the other teams' performance. The team C has a huge potential for improvement.

6 Discussion

6.1 Suggested processes

Large organizations often find it useful to establish a CSIRT to handle incidents on their information systems. Depending on the organization, the CSIRT might get other responsibilities in addition to the incident response process. This can include security awareness training, system administration, computer forensics and others. To get an absolutely correct measurement of the performance of the CSIRT, metrics should have been developed for all these activities. The benchmarking experiment should have included all activities, giving the full picture. However, not all CSIRT teams include all these activities. Most teams have one or more activities that are not directly involved in the incident response, but the content of the activities vary from team to team.

The Processes that are suggested in this thesis as incident response processes are those activities that have been identified as being a part of the actual incident handling. Through a thorough study of sources on incident handling and response the suggested processes include activities that seem to be a consensus that should be included in the incident response process.

It is always possible to argue that some activities are missing, but the suggested processes should cover most aspects within the pure incident response process. It all depends on one's own points of view. The different sources will most likely include more activities depending on the author's focus. The presented activities represent a least common denominator of the effective incident response team. Some might also argue that some activities are present that should not have been there. It is certainly possible to run a CSIRT without all these processes, but this can affect the performance of the team.

6.2 Suggested metrics

The suggested metrics are mainly based on checking the implementation of policies and procedures. Other parameters that are checked are time consumption to perform certain activities and result ratios for automated tools, reopened incidents and investigation results. The validity and reliability of these metrics depend heavily on the quality of the registration of these values.

Checking if a policy or procedure is developed gives an indication that this has been a focus for the CSIRT, but measuring how well known the policy or procedure is, gives a better measurement of the actual implementation, thus strengthening the validity of the metric. On the other hand, a registration of how well known a policy or procedure is in the whole organization is very dependent on the person answering. A person knowing the policy is more likely to give a high score than the actual state, and a person not knowing the policy will most likely give a lower score. This weakens the reliability of the metric.

Reliability will also depend on the answers being as objective as possible. Often persons answer what they would like the actual state to be, meaning they unconsciously give the answer that they think will look better, which weakens the reliability. If the persons answering the questionnaire holds the same positions in the teams this would strengthen the reliability, compared to the answers being given by persons with different positions, as different positions can result in different views.

The time estimates used in the metrics can strengthen the validity of the metric. If the time estimate is low, this is a preferable result, meaning that the team has completed the activity quickly. However, it does not state the quality of the work done. By checking time estimates together with the implementation of procedures, it also gives an indication of the quality of the work. If procedures are well implemented and used, a low time estimate indicates a quick response with good quality, as the team clearly is well prepared. But if the procedures are missing, a low time estimate indicates a response with poorer quality, as developing a correct and suitable response takes more time. Everything that can be prepared in advance saves time when an incident occurs.

The quality of the time registration affects the reliability of the metric. If time consumption is carefully registered for activity, the reliability is good. On the other hand, if the time registered is only a rough estimate, the reliability is weaker. The time registration may also differ significantly between different teams, depending on which activities the teams include in their processes. Teams that include more activities naturally use more time in the incident handling process.

The time registration is also very dependent on the service-level defined for the CSIRT. Some organizations consider their information systems a critical resource, and have a 24/7 CSIRT capability. This reduces the response time when an incident occurs. Organizations that have a CSIRT operational only during the working hours might have significantly longer response time, as incidents can occur during the evening or night when there is nobody present to handle the incident. This affects the reliability of the metric.

Checking the true/false ratio of automated tools, the ratio of reopened incidents and investigation results strengthen the validity of the metric. These ratios indicate the quality of the response actions. However, the reliability of these parameters is dependent on the experience and knowledge of the personnel doing the incident handling, as it is vital that the registration is correct. Inexperienced personnel have a greater chance of drawing the wrong conclusions on the cause of an incident.

One factor that was not measured in this experiment was the influence of experience and knowledge within the CSIRT. Experience and knowledge that the CSIRT personnel have acquired have a significant influence on the CSIRT performance, and may compensate for the lack of formal policies and procedures. But if the team is dependent on the experience alone, and has no written policies or procedures, and does not document incident handling, the team is vulnerable when personnel quits or

become unavailable for other causes. New members of the team depend on the written documentation to quickly adapt to the team and the methods used, thus responding quickly and correctly to incidents.

6.3 Benchmarking

Benchmarking is one method that can be used to improve process performance in an organization. A lot of other methods exists, and can be used. Most organizations have a continuous improvement process. This process usually has an internal focus, and uses quality circles, process analysis, brainstorming groups or other methods to improve work processes. This is important to continuously improve the work processes. However, the improvement is slow. By using methods with an external focus the possibility of getting improvement increases much quicker. This is because external sources are not familiar with the work processes, and will more easily be able to see new ways of doing things, thus creating improvement. Benchmarking is one of the methods with external focus. Another recognized method is process reengineering. The difference between these two methods is that benchmarking is using the existing process as the basis for the improvement, while process reengineering starts from the beginning, designing a whole new way of conducting business. By choosing benchmarking, it is possible to use the parts that perform well, and change and improve those parts that perform poorly.

Benchmarking can be used as a stand-alone method. It can even be used with an internal focus. With an internal focus the benchmarking is conducted between earlier performance measures for the organization and new measurements. The result will tell the management if process performance has improved or not. It is not possible to see if this is a good or poor performance comparing with others having equal or almost equal processes. In this adaptation the benchmarking method will be very much the same as the self-assessment method. When used to compare performance between two or more organizations, the real benefit of benchmarking becomes clear. It is now possible to see if there are organizations that perform significantly better than the organization in question. It is also possible to learn what the other organizations do to perform better, thus making it possible to improve.

To get the maximum benefit from benchmarking, it is necessary to have a high level of trust between the benchmarking partners. This means that it can be difficult to benchmark two competing organizations. On the other hand, when benchmarking computer security incident response the benchmarking partners are not competing with each other. In incident response all teams have the same goal, and they may benefit from the cooperation. There is a need for protecting information for the organization they serve, and this is perhaps the biggest obstacle in cooperation between CSIRTs. Development of an international infrastructure as proposed in [42] would facilitate a closer cooperation between individual CSIRTs and organizations like CERT/CC and Forum of Incident response and Security Teams (FIRST).

In this thesis, the method of generating sample-vectors and calculating the Euclidian distance was used to compare the overall performance of the teams. This method seems to be easy to adopt and use in this kind of benchmarking experiments where all the results are measured at the same time and collected at one single point.

7 Conclusion

Many large organizations choose to establish a CSIRT to be able to quickly resolve a computer security incident. Many publications describe how to plan and implement a computer security incident capability, but none of them describes a method for measuring the performance of the CSIRT.

A study of publications on incident response shows that a wide range of activities can be assigned to a CSIRT. In this thesis, we have suggested the least common denominator for what work processes a CSIRT should have, based on what seem to be the activities that most publications consider to belong to a CSIRT.

By analyzing different activities, we have developed metrics to measure the performance of each individual work process, and proposed a unit called Incident Response Performance (IRP). These metrics are based on checking how well the organizations have implemented policies and procedures that are accepted to be necessary to have an efficient incident response, how much time is used to resolve an incident and results of the incident response. By checking these parameters, we get a good indication of the team's performances.

In the experiment, we showed how these metrics can be used in a benchmarking between different CSIRTs, and how it is possible to conclude that one team performs better than another. From the measured performances for each process, we constructed a sample-vector for the team, and used the Euclidian distance to calculate the teams' distance from the desired goal. We could then compare their distances from the goal to decide who had the best overall performance.

This thesis demonstrates that using the defined metrics in a benchmarking can be a valuable tool for a CSIRT's improvement program. It can be used as a stand-alone method, but would be the most effective if implemented as a part of a continuous improvement program.

8 Future work

To continue the work performed in this thesis, the development of metrics for other CSIRT work processes will be necessary. When performance metrics for all CSIRT work processes are developed, it is possible to do a complete benchmarking of the overall performance of CSIRTs.

The suggested metrics use parameters as implementation level of policies and procedures, time consumption, and result ratios as a basis. Another factor that will affect the performance of a CSIRT is the level of knowledge and experience of the CSIRT personnel. Developing a metric to measure knowledge and experience of a CSIRT, and how this affects the performance of the team, will be an important supplement to the metrics suggested in this thesis.

For this thesis, only a small number of CSIRTs have contributed. An implementation of these metrics and benchmarking method on a larger scale within a trusted community of partners may confirm the usefulness of benchmarking CSIRT performance.

9 Abbreviations

CERT	Computer Emergency Response Team, synonym with IRT and CSIRT
CERT/CC	Computer Emergency Response Team / Coordination Centre, at Carnegie Mellon University, Pittsburgh, Pennsylvania
CMU	Carnegie Mellon University, Pittsburgh, Pennsylvania
CSIRT	Computer Security Insident Response Team, synonym with IRT and CERT
DoS	Denial of Service
FIRST	Forum for Incident Response and Security Teams
FSA	Forsvarets Sikkerhetsavdeling, Norwegian Defense Security Agency
GUC	Gjøvik University College, Norway
IRP	Incident Response Performance, Unit used to measure incident process performance
IRT	Incident Response Team, synonym with CERT and CSIRT
ISP	Internet Service Provider
IT	Information Technology
KITH	Kompetansesenter for IT i Helsesektoren, Centre of Competence in Health services, Trondheim, Norway
NISLab	Norwegian Information Security Laboratory, Gjøvik University College
NIST	National Institute of Standards and Technology, US Department of Commerce
SEI	Software Engineering Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania
TQM	Total Quality Management, management strategy

10 About the author

Ivar Kjærem is an army officer employed with the Norwegian Defense Security Agency (FSA) at Jørstadmoen. He has a bachelor degree in computer science from Buskerud University College at Kongsberg, and has studied Total Quality Management at Gjøvik University College. Kjærem lectures information security at the Norwegian Army University College of Engineering and is participating in different security projects aimed at securing defense information infrastructure. He is also a member of an inspection team supervising military units using classified information systems, and guides personnel in information security.

11 References

[1] Holm, Ola (2004): Risikohåndtering av informasjonssystemer i dynamiske omgivelser, Masteroppgave, Høgskolen i Gjøvik

[2] Killcrece et al (2003): State of the Practice of Computer Incident Response Teams (CSIRTs), Technical Report CMU/SEI-2003-TR-001, Pittsburgh, Pennsylvania, Networked Systems Survivability Program, Software Engineering Institute, Carnegie Mellon University

[3] West-Brown et al (2003): Handbook for Computer Security Incident Response Teams (CSIRTs) 2nd edition, Handbook CMU/SEI-2003-HB-002, Pittsburgh Pennsylvania, Networked Systems Survivability Program, Software Engineering Institute, Carnegie Mellon University

[4] Wack (1991): Establishing a Computer Security Incident Response Capability (CSIRC), NIST Special Publication 800-3, Computer Systems Laboratory, National Institute of Standards and Technology (NIST)

[5] Schultz & Shumway (2002): Incident Response, A strategic guide to handling system and network security breaches, New Riders, Boston, Indiana, USA, ISBN 1578702569

[6] Grance et al (2004): Computer Security Incident Handling Guide, NIST Special Publication 800-61, Computer Security Division, National Institute of Standards and Technology (NIST)

[7] Brownlee, Guttman (1998): Request For Comments 2350, Expectations for Computer Security Incident Response, http://www.ietf.org/rfc/rfc2350.txt

[8] Schultz (2004): Incident response teams need to change, Computers & Security, 23(2) s87-88

[9] Pethia & van Wyk (1991): Computer Emergency Response – An international problem, Austin, Texas, Proceedings of the 13th international conference on Software engineering.

[10] Andersen & Pettersen (1995): Benchmarking, en praktisk håndbok, TANO AS, ISBN 8251833434

[11] Creswell (2003):Research design, Quantitative, Qualitative and Mixed methods approaches, 2nd edition, Sage publications Inc., ISBN 0-7619-2441-8

[12] Payne (2001): A Guide to Security Metrics, SANS Security Essentials GSEC Practical Assignment

[13] Swanson et al (2003): Security Metrics Guide for Information Technology Systems, NIST Special Publication 800-55, Computer Security Division, National Institute of Standards and Technology (NIST)

[14] Vaughn et al (2001): Information Assurance Measures and Metrics, State of practice and proposed taxonomy, Mississippi State University

[15] Alsaker (2004): Indikatorer for informasjonssikkerhet, Trondheim, Kompetansesenter for IT i helsesektoren (KITH), KITH rapport 08/04, ISBN 82-7846-225-9, http://www.kith.no/vedlegg/19224/R08-04_Indikatorer_ informasjonssikkerhet.pdf

[16] Alberts et al (2004): Defining Incident Management Processes for CSIRTs: A work in progress, Technical Report CMU/SEI-2004-TR-015, Pittsburgh Pennsylvania, Networked Systems Survivability Program, Software Engineering Institute, Carnegie Mellon University

[17] Peikari & Chuvakin (2004): Security Warrior, O'Reilly & Associates, ISBN 0596005458

[18] Lucas & Moeller (2004): The Effective Incident Response Team, Addison-Wesley, ISBN 0201761750

[19] Mandia & Prosise (2003): Incident Response & Computer Forensics 2nd ed., McGraw-Hill/Osborne, ISBN 007222696X

[20] Fogle et al (2001): The Benchmarking Process; One Team's Experience, IEEE Software, September/October, pp. 40-47

[21] Lindqvist & Jonsson (1997): How to Systematically Classify Computer Security Intrusions, Proceedings of the 1997 IEEE symposium on security and privacy, IEEE Computer Society Press, Los Alamitos, California, May 1997, pp. 154-163

[22] Howard & Longstaff (1998): A Common Language for Computer Security Incidents (SAND98-8667), Sandia National Laboratory, USA

[23] Musman & Flesher (2000): System or Security Managers Adaptive Response Tool, DARPA Information Survivability Conference, Jan 2000, Integrated Management Services Incorporated Research Center, Arlington, Virginia, USA

[24] Hansman (2003): A Taxonomy of Network and Computer Attack methodologies, Department of Computer Science and Software Engineering, University of Canterbury, Christchurch, New Zealand

[25] Kewley & Bouchard (2000): DARPA Information Assurance Program Dynamic Defence Experiment Summary, IEEE Transactions on systems, man and cybernetics, Part A: Systems and humans, Vol 31, No 4, July 2001

[26] Sole & Bist (1995): Benchmarking in Technical Information, IEEE Transactions on professional communication, Vol 38, No 2, June 1995

[27] Hagge & Kreutzkamp (2003): A Benchmarking Method for Information Systems, Proceedings of the 11th International Requirements Engineering Conference

[28] Marie & Büyüközkan (1997): Methods and Tools for First Five Steps of Benchmarking Process, Innovation in Technology Management - The Key to Global Leadership, Portland International Conference on Management and Technology (PICMET '97), Portland, Oregon, USA

[29] Killourhy et al (2004): A Defence-Centric Taxonomy Based on Attack Manifestations, Proceedings of the 2004 International Conference on Dependable Systems and Networks (DSN'04)

[30] Gonzalez (2002): Current Approaches to Detecting Intrusions, Unpublished assignment for Doctoral course DCIS1200, www.scis.nova.edu/~luisg/pdf/LG02092003.pdf

[31] Hoepers et al (2003): Honeynets Applied to the CSIRT Scenario, 15th Annual Computer Security Incident Handling Conference, Ottawa, Canada

[32] Wright (2001): How to Design a Useful Incident Response Policy, http://www.securityfocus.com/printable/infocus/1467, last updated Sept 18th 2001, used source Sept 18th 2001

[33] DePaul University (2002): A Framework for Incident Response, Information Security Team, DePaul University, Chicago, Illinois

[34] Yee (2001): Security Metrology and the Monty Hall Problem, Position Paper for 1st Workshop on Information-System-Security Rating and Ranking, Williamsburg, Virgina, USA, May 2001, http://www.bennetyee.org/ucsd-pages/pub/metrology.pdf

[35] Zairi (1994): Benchmarking; the best tool for measuring competitiveness, Benchmarking for Quality Management and Technology, Vol 1, No 1, pp. 11-24, MCB University Press

[36] Zairi (1992): Competitive Benchmarking; An executive guide, Technical Communication Publishing Ltd, Letchworth, ISBN 0946655510

[37] Mathisen (2004): Measuring Information Security Awareness, A survey showing the Norwegian way to do it, Norwegian Information Security Laboratory (NISLab), Gjøvik University College

[38] Marsukar (2003): Responding to a Customer's Security Incidents – Part 1: Establishing Teams and a Policy, Sun Microsystems, Sun BluePrints OnLine, www.sun.com/blueprints/0303/817-1795.pdf, last updated March 2003

[39] Marsukar (2003): Responding to a Customer's Security Incidents – Part 2: Executing a Policy, Sun Microsystems, Sun BluePrints OnLine, http://www.sun.com/blueprints/0403/817-1796.pdf, last updated April 2003

[40] Marsukar (2003): Responding to a Customer's Security Incidents – Part 3: Executing a Policy, Sun Microsystems, Sun BluePrints OnLine, http://www.sun.com/blueprints/0903/817-3733.pdf, last updated Sept 2003

[41] Marsukar (2003): Responding to a Customer's Security Incidents – Part 4: Executing a Policy, Sun Microsystems, Sun BluePrints OnLine, http://www.sun.com/blueprints/1003/817-4002.pdf, last updated Oct 2003

[42] West-Brown & Kossakowski (1999): International Infrastructure for Global Security Incident Response, Pittsburgh Pennsylvania, CERT/CC, Software Engineering Institute, Carnegie Mellon University

[43] Killcrece et al (2003): Organizational Models for Computer Security Incident Response Teams (CSIRTs), Handbook CMU/SEI-2003-HB-001, Pittsburgh Pennsylvania, Networked Systems Survivability Program, Software Engineering Institute, Carnegie Mellon University

Appendix A: Questionnaire

Questionnaire about computer security incident handling

Lillehammer 10th mars 2005

Introduction

Many organizations chose to establish some sort of Computer Security Incident Response Team (CSIRT), to handle security incidents that occur within their information systems. When these teams are established, there are however few methods to measure the performance of the team. A metric for measuring performance would make it easier for the team and management to use the resources in a way that strengthen the information security. My master thesis will address this, and present a method for benchmarking CSIRTs.

My name is Ivar Kjærem and I am writing my MSc Thesis this semester at Gjøvik University College (GUC). I am currently employed at the Norwegian Defense Security Agency as senior instructor in information security. The title of my thesis is:

”BENCHMARKING CSIRT WORK PROCESSES”

The next pages contain a questionnaire. I hope you will spend a few minutes filling out the questionnaire. The answers will be used to support my thesis. All information will be treated as confidential, and your company will not be associated with data in the thesis.

As a reward for supplying background data you can receive a copy of the thesis if you like. Hopefully this can be a useful tool in you organization and the efficient handling of security threats.

If you will not support this work, please let me know. If there are other persons in your organization who is better suited to answer these questions, please forward the questionnaire and let me know so I can update my contact list.

Please return questionnaire by Thursday 7th April 2005 to:

- Email: ivar.kj@online.no
- Mail: Ivar Kjærem, Bakliveien 62, 2625 Fåberg

Best regards,

Ivar Kjærem
Phone: 99 09 48 37

In this questionnaire please mark the box you find to be most suitable. Some questions require you to make an estimate.

Computer Security Incident Response Team

1. Does your organization use a standard for information security?
i.e. BS7799, CC, Forums Standard of good practice

Yes ☐ No ☐

2.a Does your organization have procedures for regularly testing and evaluating information security?

Yes ☐ No ☐

2.b How often is information security tested?

According to the policy ☐ ☐ ☐ ☐ ☐ Never

3.a Does your organization have a procedure for testing and implementing patches in information systems (patch management)?

Yes ☐ No ☐

3.b How well is this procedure known in the organization?

Well known ☐ ☐ ☐ ☐ ☐ Known only by a few

4.a Does your organization have a policy for handling computer security incidents?

Yes ☐ No ☐

4.b How well is this policy known in the organization?

Well known ☐ ☐ ☐ ☐ ☐ Known only by a few

5.a Does your organization have a predefined group for handling computer security incidents?

Yes ☐ No ☐

5.b Is incident handling the only task of this group?

Yes ☐ No ☐

5.c How many persons are in the group?

Number:_____

5.d Is it clearly defined what responsibility the team has?

Clearly defined ☐ ☐ ☐ ☐ ☐ Not defined

5.e What is the average educational background in the group? How many years education after finishing high school?

Years:_____

Incident handling preparation

6.a Is there a policy describing who to contact when a security incident is detected?

Yes ☐ No ☐

6.b How well is this policy known in the organization?

Well known ☐ ☐ ☐ ☐ ☐ Known only by a few

7.a Does the incident handling team have a pre-produced contact list?

Yes ☐ No ☐

7.b How often is this contact list checked and updated?

Weekly ☐ ☐ ☐ ☐ ☐ Never

☐ Don't know

7.c What is the average time to reach people when trying to contact them?

Time:_____

Incident detection and registration

8. How many security incidents does your organization handle per week (avg)?

Number:_____

9. How long is the average time between the moment when the incident is discovered (the team gets the call) and the moment when the incident is registered and assigned to a team member?

Time:_____

10. What is the true/false ratio for reported incidents?

_____%

11.a Does the incident response team have a procedure for registering new incidents?

Yes ☐ No ☐

11.b How often is this procedure used when registering incidents?

Always ☐ ☐ ☐ ☐ ☐ Never

12. Does the team use adapted software to document the incident and actions taken?

Yes ☐ No ☐

13.a Does the team have a procedure that defines how to document an incident?

Yes ☐ No ☐

13.b How well do you think this procedure describe the documenting process?

Very well ☐ ☐ ☐ ☐ ☐ Poorly

14.a Has the team established a classification scheme for incidents?

Yes ☐ No ☐

14.b How well does the classification scheme cover different types of incidents?

Covers all incidents ☐ ☐ ☐ ☐ ☐ Covers only a few incidents

14.c What is the average time from the moment when the incident is assigned to the moment when it is classified?

Time:_____

Initial response to incidents

15. Does the team have written procedures for how to handle different types of incidents?

Procedures exist for all types of incidents ☐ ☐ ☐ ☐ ☐ Procedures do not exist

16. Is there a policy enabling the team to set the severity of an incident?

Yes ☐ No ☐

17. Does the team have a scheme for prioritizing incidents?

Yes ☐ No ☐

18.a Has the team established a common language for describing incidents?

Yes ☐ No ☐

18.b How useful do you think this common language is for communicating inside the team?

Very useful ☐ ☐ ☐ ☐ ☐ Not useful

19. What is the average time to develop a response strategy from the time of detection?

Time:_____

Containment

20.a How detailed is the response strategy?

All details are described ☐ ☐ ☐ ☐ ☐ Only main areas are addressed

20.b How effective do you find the chosen strategy (on average)?

Very effective ☐ ☐ ☐ ☐ ☐ ineffective

21.a Does the organization use automated response tools?

Yes ☐ No ☐

21.b What is the true/false ratio for the automated tools?

_____%

21.c Can you estimate the impact of false actions by the automated tools have on business (in NOK)?

NOK:_____

22. What is the average time to contain the incident?

Time:_____

23. What is the ratio of reopened incidents caused by bad containment?

_____%

Investigation

24. What is the percentage of the incidents in which the following could be identified:

What vulnerability was exploited? _____%

Could the exploit be isolated? _____%

Could attacker be identified? _____%

Could motivation be identified? _____%

25. What is the average time to investigate an incident?

Time:_____

Eradication

26. What is the average time from the moment an incident is detected until the system is healthy again?

Time:_____

If you have any comments / remarks, please use this box.

If you would like to receive a copy of the MSc Thesis please specify contact information in this box:	
Organization:	Name:
Phone:	Email:

Thank you for completing this questionnaire!

Please return the questionnaire by **Thursday 7th April 2005 to:**

ivar.kj@online.no

or with regular mail to:

Ivar Kjærem, Bakliveien 62, 2625 Fåberg