To Maja and Rasmus
As our dependence on IT systems increases, evaluating the dependability of critical IT systems becomes more important. This is especially true for systems that are critical in a crisis situation and whose dependability is an important factor in an organisation’s capability to resolve a crisis. Because crisis situations are typically rare events, both the dependability and the criticality of IT systems during and after a crisis situation are hard to predict.

Part I of this thesis focuses on the analysis of the sensitivity of the reliability of IT systems to changes in their usage. A crisis typically changes the usage of an IT system, which in turn can influence the system’s reliability. With the help of statistical methods the effects of changing usage profiles, modelled through the use of Markov models, can be examined. After a theoretical derivation of the properties of different models for the usage of software systems, the results are validated by applying the models to the data collected from the logfiles of a webserver.

Swedish municipalities also depend more and more on IT systems for their daily work. Because of their important role in the relief coordination during and after a crisis, the dependability of their IT systems during these emergency situations is especially critical for Swedish society. The evaluation of this dependability requires the combination of two kinds of information: how critically needed the IT systems are in crisis situations and how trustworthy the critical systems are.

Part II of this thesis focuses on how two Swedish municipalities deal with these issues in practice. Exploratory case studies involving two municipalities show that an important part of the problem lies in the cooperation and communication between the different actors involved. The study shows a need for concrete methods that can assist governmental actors in assessing and improving the dependability of their IT systems and in integrating this information in their emergency planning.

For this purpose a maturity model for process improvement in this area was developed and evaluated. In the process of evaluating this maturity model a systematic mapping study on the evaluation of maturity models was also conducted.

Finally, a study on how Swedish municipalities use Service Level Agreements (SLA) to communicate about important aspects of information safety resulted in a practical template for internal SLA’s specifically tailored for the needs of Swedish municipalities.
This thesis provides governmental organisations with tools to improve their IT dependability management. These tools are based on solid empirical studies, have been evaluated with the help of practitioners and are now ready for use and extended evaluation by governmental organisations.
ACKNOWLEDGEMENTS

The work presented in this thesis was funded by the Swedish Civil Contingencies Agency under grant for FRIVA, Framework Programme for Risk and Vulnerability Analysis of Technological and Social Systems.

This thesis is the result of a long journey, and there are a lot of people without whom I would never have reached this far. First of all, I wish to thank my supervisors Martin Höst and Per Runeson. Their guidance and encouragement are what enabled me to write this thesis.

Secondly, I wish to thank all my colleagues at the department of Computer Science, for giving me an interesting and welcoming environment to come to work to every day.

Of course, I also want to thank the participants in the interviews and case studies and the many researchers I have met throughout these last years with whom I could discuss my ideas, not in the least my colleagues in the FRIVA project. Their ideas, feedback and the inspiration from all these discussions are an essential part of this thesis.

I also want to specially thank MSB, the Swedish Civil Contingencies Agency for giving me the chance to work with these challenges while at the same time investing in a more dependable Swedish society.

And last but not least, I would not be where I am today without the continuous support and motivation from Kerstin, Maja, Rasmus and my family and friends, both in Sweden and in Belgium. I cannot thank you enough!
This thesis consists of an introductory chapter followed by seven research papers:

- **PAPER I:**
  Sensitivity of Software System Reliability to Usage Profile Changes
  Kim Weyns and Per Runeson
  *Proceedings of the 22nd Annual ACM Symposium on Applied Computing, Seoul, Korea, March 2007*

- **PAPER II:**
  Sensitivity of Website Reliability to Usage Profile Changes
  Kim Weyns and Martin Höst
  *Proceedings of the 18th IEEE International Symposium on Software Reliability Engineering (ISSRE 2007), Trollhättan, Sweden, November 2007*

- **PAPER III:**
  Dependability of IT systems in municipal emergency management
  Kim Weyns and Martin Höst
  *6:th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2009), Göteborg, May 2009*

- **PAPER IV:**
  A Maturity Model for IT Dependability in Emergency Management
  Kim Weyns, Martin Höst and Yeni Li Helgesson
  *Product-Focused Software Process Improvement, PROFES 2010, Limerick, Ireland, June 2010*
• **PAPER V:**
  
  A Review of Methods for Evaluation of Maturity Models for Process Improvement
  
  Yeni Li Helgesson, Martin Höst and Kim Weyns
  
  accepted to the Journal of Software Maintenance and Evolution: Research and Practice, incorporating Software Process Improvement and Practice, John Wiley & Sons, Inc. New York, NY, USA
  
  (published online, awaiting publication in the journal)

• **PAPER VI:**
  
  Evaluation of a Maturity Model for IT Dependability in Emergency Management
  
  Kim Weyns and Martin Höst
  
  Accepted for publication in the International Journal of Information Systems for Crisis Response and Management, Vol 4, Issue 1, 2012

• **PAPER VII:**
  
  Service Level Agreements in Municipal IT Management
  
  Kim Weyns and Martin Höst
  
  submitted to a journal
INTRODUCTION

1 Research Goals ........................................ 2
2 Research Area ........................................ 4
3 Research Contributions ............................... 10
4 Research Methodology ............................... 21
5 Conclusions ........................................... 27
6 Further Research ..................................... 28
Bibliography ............................................. 31

PART I: SENSITIVITY OF SOFTWARE RELIABILITY TO CHANGES IN THE USAGE PROFILE

Paper I: Sensitivity of Software System Reliability to Usage Profile Changes 39
1 Introduction ........................................ 40
2 Related Work ........................................ 40
3 Sensitivity analysis ................................. 41
4 Example System ..................................... 46
5 Alternative System Model ......................... 49
6 Summary and Future Work ....................... 50
Bibliography ............................................. 50

Paper II: Sensitivity of Website Reliability to Usage Profile Changes 53
1 Introduction ........................................ 54
2 Related Work ........................................ 54
3 Data Processing ..................................... 55
4 Markov Model ....................................... 58
5 Sensitivity Analysis ............................... 59
6 Application to the Research Group Webpages ....... 60
7 Summary and Future Work ....................... 63
Bibliography ............................................. 64
CONTENTS

Part II: Dependability of IT systems in Emergency Management 67

Paper III: Dependability of IT systems in municipal emergency management 69
1 Introduction .................................................. 70
2 Background .................................................. 70
3 Related Work ................................................ 71
4 Research Methodology ....................................... 72
5 Findings .................................................... 73
6 Validity Discussion ......................................... 78
7 Proposed Process improvement model .................... 79
8 Conclusions and Future Work ............................. 80
Bibliography ................................................. 80

Paper IV: A Maturity Model for IT Dependability in Emergency Management 83
1 Introduction .................................................. 84
2 Background .................................................. 84
3 Related Work ................................................ 85
4 Methodology ................................................ 86
5 Process Improvement with IDEM3 ......................... 87
6 Maturity Levels ............................................. 88
7 Transition from One Level to the Next ................... 93
8 Relation to Other Maturity Models ....................... 95
9 Evaluation of the Maturity Model ......................... 97
10 Conclusions ............................................... 97
Bibliography ................................................. 98

1 Introduction .................................................. 102
2 Related work ............................................... 103
3 Evaluation framework ..................................... 105
4 Research methodology .................................... 106
5 Results ..................................................... 114
6 Analysis of research on CMM* ............................. 119
7 Discussion .................................................. 120
8 Conclusions ............................................... 122
Bibliography ................................................. 123
Appendix A: Relevant papers from the mapping study .... 125
CONTENTS

Paper VI: Evaluation of a Maturity Model for IT Dependability in Emergency Management 131
1 Introduction ........................................ 132
2 Related work ..................................... 133
3 Background ....................................... 134
4 Evaluation Plan .................................... 138
5 Case Studies ...................................... 141
6 Conclusion and future work ..................... 151
Bibliography ....................................... 151

Paper VII: Service Level Agreements in Municipal IT Management 155
1 Introduction .................................... 156
2 Related Work .................................. 157
3 Research process .................................. 158
4 Data Collection .................................. 158
5 Service Level Agreement Template ........ 162
6 Evaluation ....................................... 168
7 Discussion ........................................ 169
8 Conclusions and Future Work .............. 172
Bibliography ....................................... 173
INTRODUCTION

IT systems have become an essential part of our society. This evolution has not only created new opportunities, but also new threats to our modern society. The presence of IT systems everywhere has made us dependent on IT systems for our daily life. At the same time as the usage of IT systems increases, the number of reports of problems caused by failures of critical IT systems has also increased (Neumann 2006).

One of the common aspects of these failures is the faith in systems that are not sufficiently dependable. The core of the problem is not that these systems suddenly become unreliable, but that we have become critically dependent on a wide variety of systems without analysing whether they are dependable enough and what the consequences could be of a possible failure (Neumann 2006).

The analysis of the dependence on a certain IT system or a set of IT systems requires the combination of two different activities. First, there is a need for an analysis of how critical the IT systems are, or in other words what the consequences would be of the systems failing or being unavailable. Secondly, this requires an analysis of how trustworthy the systems are, or more precisely what the probability is that different kinds of failures will occur. The two analyses require different techniques and expertise. The former requires a deep understanding of the processes in which the IT systems are being used; the latter requires a thorough technical understanding of the workings of the systems and the environment in which the systems are operating. Therefore these two parts of the analysis are often done by different people. A thorough analysis of the dependability of an organisation’s IT systems therefore requires a lot of cooperation. Each of these parts of the analysis is very complex, and the situation is further complicated by the fact that both the criticality and the reliability of a software system are dependent on how the system is used. This usage typically changes over time and can be hard to predict (Musa 1993).

A special case of critical dependence on IT systems is when IT systems are used in emergency management and the recovery efforts after a crisis. Just as in other sectors in our society, IT systems have also become a critical resource in emergency management. In emergency situations, a number of IT systems that are not critical in everyday situations can suddenly become highly critical. This need for a high availability during extreme conditions poses special requirements
on the dependability of these IT systems (Friberg et al. 2011). Typical systems that are not highly critical under normal conditions but that can become critical under crisis conditions are for example telecommunication systems, geographical information systems and demographic information systems such as patient data systems.

Furthermore, the usage of these systems can change drastically and the emergency situations can directly and indirectly influence the reliability of these IT systems. For example, after a major storm the telecommunication infrastructure might be badly damaged and might be further stressed by an increased number of emergency calls. Mobile phone base stations are a typical example of communication equipment that is highly critical in an emergency situation and that is often affected in a crisis situation both by problems with its power supply and by the high number of simultaneous users. Also websites, both from official governmental organisations and from the media, tend to suffer from an increased load in the aftermath of a crisis (Andersson et al. 2005).

This thesis approaches the problem of the assessment of the dependability of IT systems from two different angles. Part I of this thesis focuses on usage models for software systems that can be used to model and predict the sensitivity of the reliability to changes in their usage. Part II of this thesis is closer to current state of practice in IT dependability management. This part focuses on how Swedish governmental organisations deal on a daily basis with incorporating their IT systems in emergency management, with special focus on the cooperation between the different actors involved in the assessment of an organisation’s IT dependability.

1 Research Goals

The research presented in this thesis was conducted as part of the FRIVA, Framework Programme for Risk and Vulnerability Analysis, research project funded by the Swedish Civil Contingencies Agency. All of the research within FRIVA concerns society’s resilience against events that can affect critical societal functions. The research in this thesis is focused on the role of IT systems in emergency management, and more specifically on methods for assessing the dependability of IT systems in such crisis situations.

The goal of the research presented in this thesis was originally formulated as:

**Goal.** To develop and evaluate methods for the analysis of the dependability of IT and communication systems that also take into account the risk of catastrophic events

This goal implicitly contained the following important aspects:

1. To study how current methods used in software reliability engineering could be adapted to take into account the risk for catastrophic events
1 Research Goals

2. To contribute to the research front by developing, improving and evaluating methods for the assessment of software dependability with special focus on rare, but critical events.

3. To study how these methods could be practically used today or in the near future by the main actors in the Swedish emergency management system.

The first two of these aspects are of a more theoretical nature, as they connect closely to the current state of research in the field and are the focus of Part I of this thesis. The third aspect is more closely related to the current state of practice in IT dependability management. Therefore Part II of this thesis is directly based on a close cooperation with actors in the Swedish emergency management system. This part evolved as the main part of the research in this project and of the work presented in this thesis. The first two aspects typically belong in the field of software engineering, whereas the third aspect is more related to the field of information systems research.

The last of the three goals is especially important because typical Swedish governmental actors are different from the organisations that are the traditional target of the latest research in software reliability engineering, such as the telecommunication and avionics sector. Because little was published before about the current methods used at Swedish governmental actors today, the research in Part II of this thesis first starts out with an explorative study to investigate the current status and main problems concerning IT dependability at Swedish governmental actors. Based on these results, this thesis proposes and evaluates concrete methods to help Swedish municipalities and other governmental organisations to assess and improve their IT dependability, based on adapted versions of methods and techniques already being used in the fields of software engineering and information systems research.

These different aspects of the goal of this research can create a conflict between conducting research that can easily be generalized to an international environment and producing results that are easily applicable in a specific Swedish setting. This thesis tries to keep a balance between the two by including some results that are more general but also some results that are more specific for Swedish governmental actors.

For each of the goals, the research presented in this thesis follows systematic research methods. More details on the research methods and the validity of the results are discussed in Section 4.

1.1 Research Questions

For the more theoretical Part I of this thesis, the main research question was defined as:

RQ 1. How can the effect of the changing usage of an IT system in a crisis situation on its reliability be modelled?
For Part II of this thesis, which focuses on methods for IT dependability management, it was immediately clear that very little had previously been published on the current state of practice. Therefore, more exploratory research was necessary before more detailed research questions could be defined. For this exploratory study, discussed in Paper III, the research question was:

**RQ 2.** What are the main problems faced today by practitioners working with IT dependability management in organisations with an import role in the Swedish emergency management system?

The results of this study lead to the following research question for the study presented in Paper IV:

**RQ 3.** How can the cooperation between IT dependability management and emergency management in an organisation be evaluated and improved?

In Paper IV, a maturity model, called IDEM3, is proposed as possible answer to this question. To evaluate this maturity model, Papers V and VI were conducted with the following two research questions:

**RQ 4.** How have previous evaluations of maturity models been conducted (based on the current scientific literature)?

**RQ 5.** How can the IDEM3 maturity model be evaluated?

Finally, Paper VII focuses specifically on Service Level Agreements (SLA), one of the issues identified in Paper III, with the following research question:

**RQ 6.** What elements concerning information safety should be included in a municipal SLA?

## 2 Research Area

### 2.1 Concepts and Definitions

This thesis uses a large number of concepts that are used in different meanings, both in everyday conversation and in scientific literature. To avoid confusion, their exact meaning in this thesis is clarified shortly in this section.

**IT Management Concepts**

For all software engineering concepts concerning dependability we will follow the definitions from Avizienis et al. (2004). This means that dependability is defined as the most general concept, encompassing the more limited concepts of reliability, availability, safety and security. Basically, reliability is mostly concerned with how often failures occur or with the chance of failure-free operation. Availability
measures the percentage of time the system is accessible by taking into account how long the system is not functioning when failures occur. Safety is concerned with the absence of failures causing catastrophic consequences for its users and the environment, while security describes how vulnerable the system is to external threats. The more general concept of dependability takes into account all these aspects and corresponds best to the intuitive notion of how much a system can safely be depended upon by its users. In the work of Neumann (2006), dependability is referred to as trustworthiness, which is then of course the opposite of untrustworthiness.

The results of this thesis are applicable to many different kinds of systems from the field of Information and Communications Technology (ICT) and the terms IT system and information system are used in their broadest sense as defined by Galliers (1992). This means an IT system can include

- physical components such as hard drives, processors and input and output devices,
- software such as the operating system, compilers and applications,
- communication channels such as wireless connections, network cables and telephone lines,
- services used that are offered by other IT systems, such as backup servers or database systems and
- help facilities provided to support the IT system, such as manuals, documentation or trainings.

A complex IT system usually consists of subsystems and the exact scope of the system for which the dependability is being analysed can vary depending on the context in which the dependability is being discussed. In the context of dependability, often the scope of an IT system is considered from a user perspective, this means as all the physical and logical components necessary for the system to fulfil its functionality.

Typical for complex software systems is that the reliability of the systems is highly dependent on how the system is used. Complex software systems typically contain a large amount of code and it is nearly unavoidable that some faults are present in the code (Hatton 1997). The probability that these faults result in failures depend heavily on how and how often the respective part of the code is used by the software’s users. To model this dependency on how the software is used, software reliability engineering uses the concept of usage profiles. Usage profiles, also sometimes called operational profiles, can be used for prioritising testing to the most used components (Musa 1993, Trammell 1995), but also for calculating the reliability of the whole system from the reliability of its components (Siegrist 1988). One way to model these usage profiles in a more formal way is by using
Markov chains (Thomason and Whittaker 1999), a special kind of discrete-time stochastic processes.

Emergency Management Concepts

Also, a number of concepts from the field of emergency management need to be defined. Risk analysis and vulnerability analysis are often mentioned together, or even used to refer to the same process. In this thesis we will nevertheless try to make a distinction between the two.

Risk analysis is concerned with the identification of possible threats, with the probability that these possible incidents occur and with their immediate consequences. Vulnerability analysis on the other hand is concerned with the study of how severe the consequences are of possible incidents on society. For example, risk analysis might study what the possibility is of a large flood caused by a dam failure or by extreme rainfall, while vulnerability analysis would study what the consequences would be of a severe flood and its impact on the population.

From this it should be clear that risk and vulnerability analysis often go hand in hand, since the true value lies in the combination of both. In short, one could say that it is important to reduce the risk for those incidents to which society is the most vulnerable and to prioritise reducing the vulnerability to those events for which the risk is the highest. A more elaborate discussion of the concepts risk and vulnerability can for example be found in Johansson (2007).

The Swedish Civil Contingencies Agency, MSB, defines a crisis as

an event that affects many people and threatens the basic values and functions of society. A crisis is a condition that cannot be handled with normal resources and organization. A crisis is an unexpected, out of the ordinary event. Resolving a crisis requires coordinated action by several actors.

In other words, a crisis is when a combination of events, e.g. natural hazards, accidents or sabotage, result in a situation that negatively affects society in a way that hinders vital society functions. Examples of crises that are included in this definition might be terror attacks, storms, floods, or even a long-term disruption in the national telephone network or the Internet.

Emergency management, also called crisis management, is the process concerned with preparing for and dealing with crisis situations. In Sweden, emergency management is mainly a task of the local and regional governments, a task in which they are supported by MSB. Before the crisis occurs, emergency management includes mitigation and preparation measures. The mitigation process consists of reducing the probability or consequences of possible crises. Preparation is concerned with developing emergency plans. It is clear that risk and vulnerability analysis is an essential part of these two first phases of emergency management.

1translated from http://www.krisinformation.se
Also the emergency response during a crisis is an essential part of emergency management. During this phase the focus is on coordinating all the emergency response resources to minimize the crisis’ effect on society. After a crisis, the part of emergency management concerned with trying to restore the affected society to normal conditions, is called emergency recovery.

Research Concepts

Two other concepts that are central in this thesis are assessment and evaluation. Although these concepts are practically synonyms, they are used differently in this thesis. The term assessment is used in this text to describe the process of measuring the quality of one or more aspects of an organisation’s IT management. The word evaluation, on the other hand, is used to indicate the process of measuring the quality and validity of proposed methods, techniques and frameworks. This distinction is especially important to avoid confusion in Part II of this thesis where methods for the assessment of an organisation’s IT dependability are proposed, and these methods themselves are then evaluated.

A number of different research methods are used in the research presented in this thesis. A case study is a research study that is centred around one case (for example a person, event or organisation), taking into account its context and collecting (mostly qualitative) information from multiple data sources (Robson 2002, Runeson and Höst 2009). This is different from a survey, defined as “a collection of standardized information from a specific population, or some sample from one, usually, but not necessarily by means of a questionnaire or interview” (Robson 2002). In a case study, a survey can possibly be used as one of the methods of data collection. Action research is similar to a case study, but with a more active involvement of the researcher in the organisation. In a case study, the researcher acts only as an observer, in action research, the researcher actively tries to effect a change or improvement in the object of the study (Robson 2002). Finally, a systematic mapping study is a structured literature study in a certain field of interest by mapping relevant research papers to a classification scheme (Petersen et al. 2008). More details on these research methods and their role in this thesis can be found in Section 4.

With these concepts defined, we can now situate the research presented in this thesis more formally within the research fields of software engineering and emergency management. From a software engineering perspective, this thesis deals with the more general focus of dependability, but with a stronger focus on safety than on security. Part I of this thesis is more strictly focused on reliability, and more specifically on the effects of changing usage profiles.

From an emergency management point of view, the work in this thesis concerns the risk and vulnerability analyses of IT systems conducted in the mitigation and preparation phase of emergency management. More exactly, it discusses methods
used for assessing the vulnerability caused by depending on possibly untrustworthy IT systems in the response and recovery phases after a crisis.

2.2 Related work

Each of the papers included in this thesis contains a detailed overview of related work to the research presented. In this section here, the discussion is limited to the related work that is important to large parts of this thesis and not just to individual papers.

The work presented in Part I of this thesis builds directly on the earlier software reliability research on Markov models done by Siegrist (1988) and Goševa-Popstojanova and Kamavaram (2004). The concept of Markov chains has been used extensively in software reliability engineering. For example, Le Guen et al. (2004) use Markov models to generate test cases and to combine the results of the testing to a total reliability estimate. Markov models can also be used to calculate test coverage, as done by Walton and Poore (2000). The sensitivity to usage profile changes in test coverage has been discussed by Wesslén et al. (2000).

Part II of this thesis is closely related to the field of business continuity management, where most of the research is based on large frameworks such as ITIL² (Office of Government Commerce 2007), COBIT³ (ISACA 2000) or CERT-REF⁴ (Caralli 2007). Concerning just the situation for Swedish governmental organisations, MSB has published many reports and recommendations about risk and vulnerability analysis in Swedish emergency management, such as the work by Sundelius et al. (1997) and Hallin et al. (2004), but not so many deal with the role of IT systems. MSB’s publications such as BITS⁵ (2003) and the first results of the survey discussed in the case studies (Kalmelid and Gustavsson 2005) focus often more on security than on reliability of IT systems. An important collection of related work to the results in this part of the thesis consists of reports of famous failures of IT systems in crisis situations, such as those reported by National Research Council (2003), Rahman et al. (2006) and the Swedish National Post and Telecom Agency (Post&Telestyrelsen 2005).

Vogt et al. (2011) have discussed similar problems with IT management as those identified in Paper III of this thesis in Australian and European emergency management organisations. Further, research by Vogt et al. (2011) has shown that the large, international frameworks discussed in the previous paragraph are badly suited for emergency management organisations and simplified, more agile IT governance frameworks are needed to support improvements in their IT management. They propose that these simplified models would be based on larger frameworks like ITIL (Office of Government Commerce 2007) or COBIT(ISACA 2000).

²Information Technology Infrastructure Library
³Control Objectives for Information and Related Technology
⁴CERT Resiliency Engineering Framework
⁵Basic Level for IT Security
Santos et al. (2008) have published a maturity model for the use of information technologies in emergency response organisations. Their model does not cover the dependability of the IT systems in emergency situations, but instead focuses on information management practices. The IDEM3 maturity model described in this paper is most suited for an organisation where the IT services are provided by an IT department that is part of the organisation. For evaluating the resiliency of IT services provided by external suppliers, Bhamidipaty et al. (2007) have developed the Resiliency Maturity Index, a framework for characterizing and evaluating the resiliency of an IT services organization. However this model does not evaluate the relationship between the resiliency of the service supplier and the dependability requirements of the organisation.

The maturity model presented in Part II of this thesis was developed in parallel to a similar maturity model for assessing and improving the emergency management capacity of an organisation (Ek and Borell 2010) which was developed within the same research programme. This maturity model incorporates some of the same concepts as the IDEM3 framework, but has a different focus and does not contain any IT dependability aspects.

One example of work that is directly related to the research in both parts of this thesis concerns the usage of Markov models for doing risk and vulnerability analysis of IT systems at the architectural level (Yacoub and Ammar 2002). They combine Markov models and risk and vulnerability analysis, but the main difference with the work in this thesis is that their analysis method is designed for small, specialised systems, and is therefore less suited for complex systems that are part of a municipality’s emergency management.

Some areas of research that are indirectly related to both parts of this thesis concern risk and vulnerability analysis of critical IT systems, reliability assessment of IT systems with very high reliability and IT systems for emergency management.

In the first of these areas, many different techniques have been developed for risk and vulnerability analysis of critical IT systems. For example, Rae et al. (2005) describe a technique for conducting dependency analysis called critical feature analysis. Also Lawrence and Gallagher (1997) discuss in detail how a software safety hazard analysis can be conducted. Both these papers stress the importance of systematically investigating all possible failure modes. The main difference between this approach and the work presented here is that these methods focus entirely on the system level, while the research in Part II of this thesis focuses on the organisational level. Nevertheless a municipal risk and vulnerability analysis could definitely benefit from such a detailed technical analysis of the vulnerabilities of the most critical IT systems.

For highly critical IT systems a very high reliability is often required. Such high reliability is hard to achieve and equally hard to assess during the testing of these systems. The reliability required is often much higher than what can statistically be shown through traditional testing approaches. A number of different
approaches have been described to deal with this problem. Chan (2004) describes the advantages and disadvantages of accelerated stress testing and how it can be used to discover faults or weaknesses in the systems that normal usage testing would probably not detect. Also Tang and Hecht (1997) and Voas and Miller (1995) discuss the importance of testing outside the normal usage profile to discover rare failure modes of the system. Thomason and Whittaker (1999) discuss the first steps to how even rare failures can be modelled with Markov chains.

The field of IT systems research specifically for emergency management is very broad and has grown substantially in the last decade. Many of the results in this area have been published by the ISCRAM community. The use of IT systems in emergency relief poses special requirements on these systems (Friberg et al. 2011). Many of these systems focus on managing the complexity, or some aspects of the complexity, that is typical for emergency situations (Coskun and Ozceylan 2011).

Finally, a number of authors (Hatton 1997, Neumann 2006) have discussed the general difficulty of constructing and verifying highly reliable systems. They state that we need to learn from past failures and should be more cautious before relying on systems that have not been shown to be reliable, because we currently have no easy way to build highly complex software systems that are guaranteed to be highly dependable.

3 Research Contributions

The work presented in this thesis is divided into two parts. Part I contains theoretical contributions to the field of software reliability engineering and their applications. This first part started from the latest research on reliability models and adapted and extended some of those methods to take into account changing usage profiles.

One of the goals of this research was to contribute with results that Swedish municipalities can directly apply in their IT dependability management. Therefore, there was also a need for research that is much closer to the current state of practice for software dependability management. Since little is actually written about the current state of practice at Swedish governmental organisations, Part II of this thesis focuses on exploring how Swedish governmental organisations today deal with dependability issues of their IT systems in crisis situations, and which problems they are experiencing. The main goal of this part is to provide practical methods and techniques that organisations can use to assess and improve their IT dependability management.

Both parts of this thesis contain complete papers as they were published at or submitted to respective conferences and journals with only minor editorial changes.
Figure 1: Relationship between the papers included in this thesis. An arrow between two papers indicates that the second paper is directly based on (part of) the results of the first paper.

...to create a uniform reading experience throughout this thesis. The papers are presented in the order they were written, as this order also presents the most logical narrative. The relationship between the research presented in the different papers is shown in Figure 1.

3.1 Part I: Sensitivity of Software Reliability

Included Papers

The first, more theoretical, part of this thesis contains two papers presenting the derivation and evaluation of a software reliability model that can be used to assess the effect of changes to the usage profile of a software system on the estimated reliability of the system:

**PAPER I:**

*Sensitivity of Software System Reliability to Usage Profile Changes*

Kim Weyns and Per Runeson

*Proceedings of the 22nd Annual ACM Symposium on Applied Computing, Seoul, Korea, March 2007*

**PAPER II:**

*Sensitivity of Website Reliability to Usage Profile Changes*

Kim Weyns and Martin Höst
**INTRODUCTION**

*Proceedings of the 18th IEEE International Symposium on Software Reliability Engineering (ISSRE 2007), Trollhättan, Sweden, November 2007*

**Other Publications**

In the course of the research presented in Part I, the following paper, not included in this thesis, was also published:

**Sensitivity of Software Reliability to Usage Profile Changes**  
Kim Weyns and Per Runeson  
*Proceedings of the Fifth Conference on Software Engineering Research and Practice in Sweden (SERPS'05), Västerås, October 2005*

This paper is not included here because the main results of this paper are also included in Paper I. The main difference between this paper and Paper I is that this paper only covers the basic software reliability model, and not the more advanced variations discussed in Paper I, leaving more room for a more detailed derivation of the theoretical results and a more extensive treatment of the example system.

**Research Contributions**

Paper I contains the theoretical background, whereas Paper II presents the application of the model to a concrete example. In these papers a software system is modelled through the use of a Markov model (Cheung 1980), this is a state machine with a given number of possible internal states. The usage profile of the system can then be represented as a matrix of transition possibilities between each of these states. A failure of the system is modelled by adding one additional failure state to the model. The reliability of the system can then be defined as the probability for the system to terminate in the goal state and not in the failure state.

This model can be used to calculate the effect of changes to the usage profile of the system on the overall reliability of the system. For small changes to the usage profile, the effect of the changes can be summed to calculate the statistical effect of many small changes with a given statistical distribution, on the total reliability of the system. The goal of Paper I is to present simple formulas that express the relationship between the variation of the usage profile and the variation in the resulting reliability.

Paper I includes such formulas for four different types of changes to the usage profile, but in each case the transition possibilities out of each state must always sum to 1, and can therefore not be changed independently. The simplest case is where every transition possibility between two states is changed by selecting a random other transition possibility to undergo the opposite change. This is referred to as *paired changes*. A second possibility is called *spread changes*, where all transition probabilities are scaled to make the sum equal to 1 again each time a transition possibility is changed. For these two types of changes, all changes are
assumed to be from the same normal distribution with a given standard deviation, which can be set a parameter of the model. For some systems it can be unrealistic to make large changes to very small transition possibilities, therefore a third type of changes, relative changes, was defined. These are similar to spread changes, but with the difference that the size of each change is drawn from a distribution with a standard deviation relative to the size of the changed transition probability. This also means that transitions that were impossible in the original system, remain impossible in the changed system. The most advanced model is where each change between two transition possibilities is drawn from a separate distribution. This is the most flexible model, as it allows for the definition of the most and least likely changes to the usage profile, but this also requires many more parameters than the simpler models above.

Paper I also explains how the formulas can be adapted for systems without a clear goal state, where the dependability is not measured as the reliability (expressed as the possibility of failure), but instead as the availability (expressed as the mean-time-to-failure, MTTF). Also for this type of system, formulas for the sensitivity of the availability to the same 4 types of changes to the usage profile can be derived. This variation of different kinds of system models and different kinds of changes makes the results of Paper I more generally applicable to a wider range of systems and situations.

Paper I presents the reasons for the application of this analysis and derives how the results can be calculated directly without resorting to a lengthy Monte Carlo simulation. Instead, Paper I uses the Monte Carlo simulation on a small example to validate our results and to illustrate the limitations of the developed models. The validity of the different variations of the formulas is shown by again comparing the results to results from simulations on variations of the same example system.

The main conclusions from Paper I is that the presented models can be used to quickly analyse the sensitivity of the reliability of a software system to changes in its usage profile, without using extensive simulations. The disadvantages of this approach are that a detailed usage profile is needed, that the nature of the changes to the usage profile can be hard to predict and that when the changes are too large, the approximate models in these two papers can no longer be applied but more extensive simulations need to be used.

Where Paper I only uses a fictitious example with a usage model taken from literature, Paper II uses data from a real system to show how the models from Paper I can be applied in practice. Paper II shows how server reliability can be measured and analysed with the help of these models. Further, this last paper also contains a thorough discussion of the data processing necessary for the extraction of a usage profile from the logs of a server. This way, Paper II illustrates the possibilities and practical limitations concerning the applicability of the theoretic models discussed in Paper I.

The conclusion of Paper II is that useful results can be obtained by extracting a usage profile from a server log file and applying the models from Paper I.
The disadvantages of this approach are that the usage profile cannot be perfectly reconstructed and that a large amount of pre-processing can be necessary. Both these disadvantages are a direct consequence of the limitations of the data in the available logfiles.

3.2 Part II: Dependability at Swedish Municipalities

Included Papers

Part II of this thesis contains five papers related to the IT dependability practices at Swedish governmental organisations:

**PAPER III:**
Dependability of IT systems in municipal emergency management
Kim Weyns and Martin Höst
6th International Conference on Information Systems for Crisis Response and Management (ICSCRAM 2009), Gothenburg, May 2009

**PAPER IV:**
A Maturity Model for IT Dependability in Emergency Management
Kim Weyns, Martin Höst and Yeni Li Helgesson
Product-Focused Software Process Improvement, PROFES 2010, Limerick, Ireland, June 2010

**PAPER V:**
A Review of Methods for Evaluation of Maturity Models for Process Improvement
Yeni Li Helgesson, Martin Höst and Kim Weyns
Accepted for publication to the Journal of Software Maintenance and Evolution: Research and Practice, incorporating Software Process Improvement and Practice, John Wiley & Sons, Inc. New York, NY, USA

**PAPER VI:**
Evaluation of a Maturity Model for IT Dependability in Emergency Management
Kim Weyns and Martin Höst
Accepted for publication in the International Journal of Information Systems for Crisis Response and Management, Vol 4, Issue 1, 2012

**PAPER VII:**
Service Level Agreements in Municipal IT Management
Kim Weyns and Martin Höst
submitted to a journal
Other Publications

In the course of the research presented in Part II, the following three manuscripts, not included in this thesis were also published:

**AN EXTENDED ABSTRACT DETAILING THE MOTIVATION AND PLANNING OF THE CASE STUDIES PRESENTED IN PAPER III:**

*Software Dependability under Emergency Conditions*
Kim Weyns and Per Runeson

*Extended abstract presented at The 17th IEEE International Symposium on Software Reliability Engineering (ISSRE 2006), Government Track, Raleigh, North Carolina, USA, November 2006*

**A BOOK CHAPTER (IN SWEDISH), PRESENTING A SUMMARY OF THE RESEARCH DISCUSSED IN PAPERS III TO VII**

*FRIVA - risk, sårbarhet och förmåga - samverkan inom krishantering, Chapter 9: Pålitliga IT-system i krishantering*
Kim Weyns, Martin Höst, Yeni Li Helgesson and Per Runeson


**A SERVICE LEVEL AGREEMENT TEMPLATE FOR SWEDISH GOVERNMENTAL ORGANISATIONS (IN SWEDISH), FOCUSED ON INFORMATION SAFETY ASPECTS**

*Service Level Agreement mall för kommunalt IT-stöd*
Kim Weyns and Martin Höst

*published on MSB-owned website [http://informationssäkerhet.se/Stod-i-arbetet/For-kommuner/]*

The goal of these last publications is mostly to disseminate the results to a larger audience outside the research community, which is an important part of the goals of this research project. The first of these publications presented the problems studied in this thesis to an audience with many representatives from governmental organisations, mostly from the USA. The goal of the last two of these publications is to spread the practical results of this thesis directly to Swedish practitioners in the field. Further, the author of this thesis also contributed to an article by Runeson and Falk (2007) in a Swedish, popular science journal concerning the risks of untrustworthy systems to society.

Research Contributions

Paper III presents the results of an explorative study on the current state of practice in IT dependability management at Swedish municipalities. The paper presents the results of two case studies involving IT personnel and safety managers from two municipalities. Before the publication of this paper there was very little empirical data available on the current IT dependability processes being used in practice. The
INTRODUCTION

Case studies focus on understanding the division of responsibilities, the methods currently being used and the problems experienced in IT dependability management today. The problems identified in the course of the research presented in Paper III form the direct basis for all the consecutive papers.

The main conclusions from Paper III are that Swedish municipalities are experiencing a number of important problems when trying to include risk and vulnerability analyses of their IT systems in their emergency management process. The responsibilities for this matter are divided between the emergency managers, the IT personnel and personnel of those departments responsible for the critical IT systems. This division of responsibilities leaves some critical gaps on the borders between the responsibilities of different departments. This causes some important issues to be forgotten because nobody takes responsibility for them. Another problem lies in the communication between the IT department and the rest of the municipality. Because of a lack of a good forum of communication, both parties feel that they cannot get the information they need from the other party. The IT department cannot get the users to understand the threats to the IT systems, while the rest of the municipality feels that the IT department does not understand their needs and that the IT support focuses on the wrong things. Both these problems together make that the IT department is not involved in emergency management and as a consequence the emergency managers sometimes need to have blind faith in their IT systems that might not be as reliable, or to prepare extensive backup measures for IT systems that are sufficiently reliable. With a better communication and cooperation on these issues, IT and emergency management resources could likely be prioritised more efficiently.

Paper IV presents a maturity model that can help Swedish governmental organisations to assess and improve their IT dependability practices. The model is directly based on the conclusions from the empirical studies presented in Paper III and focuses especially on the cooperation between IT personnel and safety managers. The model is presented in Paper IV and evaluated in a practical setting in Paper VI. The evaluation of a maturity model is a complex process and therefore Paper V presents a mapping study on the evaluation of maturity models.

The IT Dependability Management Maturity Model (IDEM3) presented in Paper IV contains 22 attributes and 5 maturity levels. Each of the 22 attributes, shown in Table 1 describe one aspect of IT dependability management, and the attributes are divided over 4 categories, shown in Figure 2. The core of the maturity model consists of a maturity matrix describing each of the attributes at the 5 maturity levels. An organisation that wishes to improve or assess its IT dependability practices should first map their current practices and processes to each of the attributes to assess their current maturity in these aspects. In a next step the organisation can then try to implement higher level practices for those attributes where the organisation was assessed at the lowest level of maturity. By implementing improvements for each of these attributes the organisation can reach a higher level of maturity for their entire IT dependability management.
## Research Contributions

### Table 1: The 22 attributes and the 5 levels of the maturity model

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level 1: Initial</th>
<th>Level 2: Managed</th>
<th>Level 3: Established</th>
<th>Level 4: Quantitatively Managed</th>
<th>Level 5: Continually Improving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>Reactive</td>
<td>Responsive</td>
<td>Preventive</td>
<td>Predictive</td>
<td>Proactive</td>
</tr>
<tr>
<td>Problems that can be identified</td>
<td>Failures in risk analysis procedures</td>
<td>Technical system faults</td>
<td>Insufficient reliability, too high dependence and interdependencies</td>
<td>Failures in risk analysis procedures</td>
<td>Failures in risk analysis procedures</td>
</tr>
<tr>
<td>Basis for improvements</td>
<td>Failing infrastructure</td>
<td>Technical judgement</td>
<td>Standards</td>
<td>Quantitative risk analysis</td>
<td>Safety culture</td>
</tr>
<tr>
<td>Improvements</td>
<td>Few, changes can be both positive and negative</td>
<td>For some systems only, unsustained</td>
<td>For all systems, sustained</td>
<td>Regular, organized</td>
<td>Continuous</td>
</tr>
<tr>
<td>Success</td>
<td>Not repeatable</td>
<td>Repeatable</td>
<td>Sustainable</td>
<td>Measurable</td>
<td>Source for organisational learning</td>
</tr>
<tr>
<td>Success factor</td>
<td>Luck, Competence of key personnel</td>
<td>Personal competence of system managers</td>
<td>Central coordination</td>
<td>Cooperation, Measurements</td>
<td>Continuous improvement offset by everyone</td>
</tr>
<tr>
<td>Risk of IT in emergency situations</td>
<td>Cause of problems identified</td>
<td>Varying between systems</td>
<td>Under control, backup solutions available</td>
<td>Predictable behaviour</td>
<td>Valuable asset</td>
</tr>
<tr>
<td>Dependability problems</td>
<td>Accepted, made clear</td>
<td>Solved on individual basis</td>
<td>Possible problems identified and prevented</td>
<td>Possible problems predicted and prevented in a prioritised way</td>
<td>Continuously prevented</td>
</tr>
<tr>
<td>IT dependability</td>
<td>Low</td>
<td>Mixed</td>
<td>Stable</td>
<td>Controlled</td>
<td>Continually improving</td>
</tr>
<tr>
<td>IT incident management</td>
<td>None</td>
<td>Improvement for one system</td>
<td>Improvement for all systems</td>
<td>Improvements and procedures</td>
<td>Improvement in organisational culture</td>
</tr>
<tr>
<td>IT responsibility management</td>
<td>Ad-hoc</td>
<td>System-based</td>
<td>Formal incident management</td>
<td>Incident Management Procedure</td>
<td>Basis for learning</td>
</tr>
<tr>
<td>Dependability</td>
<td>Known for some systems</td>
<td>Documented</td>
<td>Measurable</td>
<td>Continuously improved</td>
<td></td>
</tr>
<tr>
<td>IT dependability analysis and emergency planning</td>
<td>Separate activities, not combined</td>
<td>Activities with a shared goal, contact for some systems</td>
<td>Input for each other</td>
<td>Measurable values exchanged, direct connection</td>
<td>Naturally combined activities</td>
</tr>
<tr>
<td>Resource of IT personnel</td>
<td>Not included</td>
<td>Include for some systems</td>
<td>Risk and vulnerability analysis for all IT systems</td>
<td>Measurable values included, regularly tested</td>
<td>Naturally included, continuously improved</td>
</tr>
<tr>
<td>Involvement</td>
<td>Nobody</td>
<td>Individual system managers and some stakeholders</td>
<td>All internal stakeholders</td>
<td>All internal stakeholders, some external stakeholders</td>
<td>All stakeholders, internal and external</td>
</tr>
<tr>
<td>Responsibility</td>
<td>'Someone else'</td>
<td>Individual system managers</td>
<td>(delegated by) IT safety manager</td>
<td>Shared by system, process, IT and safety managers</td>
<td>Everyone</td>
</tr>
<tr>
<td>Management and mechanisms</td>
<td>None</td>
<td>Reactive ('Don't do that again!')</td>
<td>Following rules ('Do be like this!')</td>
<td>Double loop learning ('Improve the way you think!')</td>
<td>Everyone</td>
</tr>
<tr>
<td>Organisational learning</td>
<td>None</td>
<td>Reactive ('Don't do that again!')</td>
<td>Following rules ('Do be like this!')</td>
<td>Double loop learning ('Improve the way you think!')</td>
<td>Everyone</td>
</tr>
<tr>
<td>Resource allocation</td>
<td>Reactive, all in</td>
<td>To individual projects</td>
<td>Divided equally over the whole organisation</td>
<td>Optimised for best interests</td>
<td>Optimised for best interests</td>
</tr>
</tbody>
</table>
Figure 2: The 4 categories of attributes and the 5 levels of the maturity model.
Maturity models are a well-established process improvement approach in the field of software engineering, and have lately also been applied in some other fields. The IDEM3 framework is based on a number of similar maturity models already in use in related fields such as CMMI (CMMI Product Team 2006), the CERT Resiliency Engineering Framework (Caralli 2007), the Safety Culture Maturity Model (Fleming 2001) and the IT Business Alignment Maturity Model (Luftman 2003). The main difference between IDEM3 and these models is that IDEM3 was developed specifically to be easy to use with limited resources. The model is sufficiently small that an assessment can be done through a series of individual interviews and one or two focus group meetings. Further IDEM3 is specifically targeted to governmental organisations that have an active role in society’s emergency management system which poses specific requirements on the organisations’ IT dependability management processes. The model was designed to offer a concrete solution to the specific issues identified in Paper III and the attributes of the model are directly based on these earlier empirical results.

The IDEM3 maturity model was partially evaluated in a practical setting in two cases studies at Swedish hospitals in Paper VI. This evaluation involved a partial assessment of the organisation at the first hospital and a full IDEM3 assessment at the second hospital. Each assessment consisted of a series of interviews, followed by a focus group meeting in which the assessment was also evaluated, followed by an evaluation survey among the participants at the focus group meeting. The limited scope of this research does not allow for a full evaluation of the maturity model, which would require many more organisations and a longer timeframe. The evaluation instead focuses only on the assessment part of the maturity model, not on the implementation of the proposed improvements that are the result of the assessment. The evaluation of the maturity model does not mainly focus on the correctness of the assessment, but rather on the value of the assessment for the organisation that is being assessed. The main conclusion of this evaluation was that the IDEM3 model can be used in practice to assess the IT dependability practices of an organisation. The main value of the IDEM3 assessment was shown to lie in the fact that it identifies the strengths and weaknesses of an organisation’s IT dependability management and makes them explicitly visible to the different actors involved while at the same time bringing these different actors together to discuss these issues. Further these strengths and weaknesses are presented in a graphical way that makes it easy for the organisation to prioritise future improvements, for example as in Figure 3.

The systematic mapping study in Paper V gives an overview of the published work on the evaluation of maturity models. The paper proposes a framework to categorize different techniques for the evaluation of maturity models and maps each of the 59 research papers found to this framework. Further the mapping study focuses especially on the different evaluations of the CMM maturity model (Paulk et al. 1993) and its successors. This study serves a logical step between the presentation of the IDEM3 framework in Paper IV and the evaluation of the
maturity model in Paper VI. Paper V makes a distinction between three types of evaluation of a maturity model:

Type 1: a theoretical evaluation of the model by the authors, by analysing properties of the model or by relating the model to previously published literature in the field

Type 2: a static evaluation through feedback from practitioners, but without applying the model in a real-world setting

Type 3: an evaluation through one or more actual assessments based on the maturity model in a real-world setting

Based on this classification framework, the evaluation of the IDEM3 maturity model in Paper VI can be classified as a type 3 evaluation, i.e. an evaluation through an actual assessment in one or more case studies. A type 1 evaluation (a theoretical evaluation of the model by the authors), by mapping the maturity model to different other maturity models from related fields, is included at the end of Paper IV. A type 2 evaluation of the maturity model was performed only informally at different stages during the development of the maturity model.

Finally Paper VII discusses how Swedish municipalities can use Service Level Agreements (SLAs) as part of their IT dependability management. SLAs, one of the aspects of the maturity model presented in Paper V, are often used in IT business alignment, but lately there has also been an increased interest in using

Figure 3: Example of a spider graph representing the maturity of an organisation in each of the model’s attributes.
SLAs in municipal IT management. Therefore Paper VII presents the results of an interview study concerning the use of SLAs at Swedish municipalities. Based on these interviews, an SLA template for Swedish municipalities was developed that focuses specifically on IT dependability management issues. The template was then evaluated through a workshop and a survey.

The background study that forms the basis of Paper VII showed that many municipalities have a strong interest in developing SLAs for their internal IT management, but that it is difficult for them to write effective SLAs. Based on the results of these interviews, together with a number of collected SLAs from Swedish municipalities and SLA templates from other sources, an SLA template was composed that focuses entirely on those aspects that are important for IT dependability management in a Swedish municipality. The resulting 4-page template is in the form of a checklist that organisations can use to select different parts to include when writing their SLAs. The SLA template was positively evaluated by all of the municipalities that participated in the workshop and the template is now available to all municipalities and other interested organisations through the website of the Swedish Civil Contingencies Agency on information security.

3.3 Author Contributions to the Papers

The author of this thesis is the first author of all papers in this thesis except Paper V. For the papers in Part I, the author did all the theoretical derivations and conducted the simulations. The main author was also responsible for the processing of the data in Paper II. Each step was discussed with the second authors of the papers who were the first author’s supervisors. For both papers the main author also contributed most of the text.

For the Papers III, IV, VI and VII, the author planned the case studies with help of the other authors. All interviews were personally conducted and analysed by the author, some of them together with the other authors of the respective papers. For the later papers, the author also became more personally responsible for the idea and the design of the complete studies. The author also wrote the main parts of the text in each of these papers.

For Paper V, the author of this thesis conducted part of the data analysis needed for the mapping study, and wrote considerable parts of the paper.

4 Research Methodology

4.1 Choice of Methodology

This thesis does not only contain the results of the research work but also discusses the research methodology in detail. The research methodology is the most impor-
tant factor in assuring the validity of the results. The research in this thesis uses several different research methods, both more theoretical and empirical. Most of the methodology used in this thesis, such as case studies, surveys or simulations are research approaches that are common both in the field of software engineering (Shull et al. 2007) and the field of information systems research (Galliers 1992). The planning and execution of the research studies is based on the methodology presented by Robson (2002) and for the case studies on the work published by Yin (2003) and Runeson and Höst (2009).

As the research goal for the two parts of this thesis is different, the methodology used is also different. The methodology in Part I is very common in the field of software reliability engineering, while the methodology in Part II is more closely related to the methodology used in information systems research. Because the goal of both parts is to design and evaluate methods for the use in IT dependability management, the research in both parts is also based on design science (Hevner et al. 2004) methodology and guidelines. The conceptual framework for design science applied to information research is illustrated in Figure 4. In design science, foundations and methodologies from the knowledge based are used to develop and evaluate practical solutions to be applied in real organisations. In this way, design science is situated between the domains of social science and fundamental science and combines techniques from both these fields.

![Figure 4: Design science applied to information systems research. Based on a figure by Hevner et al. (2004).](image)

Part I of this thesis is based on the mathematical derivation of statistical properties of Markov models used in software reliability modelling. To assure the validity of the results, the theoretical results are compared to corresponding values obtained from extensive Monte-Carlo simulations on small fictive examples from the literature. To investigate the applicability of the results, the theory is finally also applied to a larger real-life system in Paper II.
This methodology was chosen because it connected well to the work done before on Markov models, and because it allowed for a theoretical study that could be started early in the research process without first setting up contacts for empirical research. Because of the difficulty of obtaining sufficiently detailed usage profiles from real life systems, the first part of the research employed only fictive examples. Only when more research was done on how a detailed usage profile could be derived from the log files of a webserver, the results could be validated on a real-life system.

Part II of this thesis, on the other hand, contains mainly empirical research. Paper III and IV are based on two explorative case studies, complemented with some results from a large survey conducted by the Swedish Civil Contingencies Agency.

Paper V contains a systematic mapping study, a research method that is a variation of the systematic review that has become common in the field of software engineering over the last years (Kitchenham et al. 2009). For this mapping study we followed the methodology recommended by Kitchenham and Charters (2007) and Kitchenham et al. (2011).

Paper VI is based on two new case studies. Unlike in Paper III, most of the data in these studies was collected through telephone interviews complemented with two focus group meetings. The research in Paper VI also included some aspects bordering on action research, as the authors of this paper themselves applied the maturity model to the organisation to conduct an assessment of these two organisations’ IT dependability management practices. Although the involvement of the authors with the assessed organisations was not as deep as in some definitions of action research, they had a more active role than in a traditional case study.

Finally, Paper VII was conducted through a telephone survey among municipalities in Skåne, Sweden’s southernmost province, followed by a focus group meeting. More details on the exact research methodology used can be found in the respective papers.

Throughout the research in this thesis, practical limitations were an important factor in the selection of the research methods and research subjects. The research in Part I of this thesis is mostly theoretical because of the difficulties in collecting data about the usage profiles for real systems. At the start of this project, there were plans to extend this part to include practical studies with data from IT systems that are used by the main actors in the Swedish emergency management system, but it turned out to be more difficult to collect these usage profiles than expected.

For the research in Part II, the case study methodology was a logical choice. A possible alternative would have been to use even more action research, in which the researcher is involved in the studied organisations for a longer time. This would have made it possible to study one organisation in greater detail and to study the effects of changes over time in the organisation’s IT dependability management. Because of the large effort involved in such a study, this would have made the results even more focused on one organisation and would have had a negative
The choice of methodology is closely related to the choice of conceptual framework on which the research is based. Especially in the field of information systems research, many studies build on an explicit system model, such as architectural models, information flow models, life cycle models or organisational models. Part I of this thesis is founded on Markov chain models of the usage profile of a system, which is a detailed form of functional models. Part II of this thesis has a broader focus and is based on organisational models such as maturity models which is a kind of framework used to model an organisation’s capabilities and learning processes, and service level agreements which have their basis in contractual theory.

A typical aspect of critical realism that is strongly present in this thesis is the understanding that the results of design research are strongly context dependent. The methods presented in this thesis only work in certain contexts and it is each time the stakeholders applying these methods that make these initiatives work. Therefore we must always keep in mind who the target is of the designed meth-
4 Research Methodology

The research question in each evaluation of the results is then which methods will work for which stakeholders and why. To study these questions a combination of many different research techniques and methods is necessary, combining information from both objective facts and subjective values.

4.2 Research Validity

In each of the papers included in this thesis, the research validity is one of the main concerns. Details concerning the measures taken to reduce the threats to the validity for each of the studies, are discussed in each of the respective papers. In this section, the discussion is limited to the validity of the overall conclusions of this thesis.

The discussion of the validity of research can be divided in 4 separate aspects (Yin 2003): reliability, internal validity, construct validity and external validity. These aspects of validity are similar to the concepts of grounding of the knowledge acquired through research as discussed by Ågerfalk (2004), who makes a distinction between internal and external grounding of the research concepts and operationalization and the empirical grounding of the application of the research results.

Concerning the reliability of the research (whether the results would be the same if the studies had been conducted by other researchers), it is no longer possible to replicate the exact studies because the case studies capture the processes in an organisation at a certain moment in time. Nevertheless, a number of measures were taken to guarantee the reliability of the studies. When possible, the interviews, whether they were conducted in person, over the telephone or in a focus group, were recorded to make it possible to analyse the exact contents at a later time. Further, to reduce researcher bias, each of the studies involved at least two researchers that analysed the material.

Concerning the internal validity (the validity of the relationship between cause and effect identified by the studies), IT dependability management is a complex field with many factors that influence the final outcome. In case studies, unlike in controlled experiments, the goal is not to prove the direct relations between cause and effect. For example in the IDEM3 maturity model discussed in papers IV and VI, the goal of case studies is not to prove the effect of the attributes in the model on the IT dependability of an organisation. The importance of the attributes is derived from the interviews with the participants and confirmed by the presence of similar attributes in maturity models used in related fields. Ideally, it would be preferred if it was possible to independently measure the effect of each of the attributes in the maturity model on the quality of an organisation’s IT dependability management in a series of experiments that modify one of the attributes and leave all the others unchanged. Of course this is impossible in practice, as there is not even an agreed way to measure an organisation’s IT dependability, except in retrospect over a long period of time.
To address the construct validity of the research in this thesis (whether the studies actually measure what they are claiming to be measuring), multiple sources of evidence were collected in most of the studies. Interviews were conducted with people in different positions in the organisations involved in the studies, and the premise that IT dependability cannot be studied from just the IT department or just from the viewpoint of the safety managers is one of the basic premises of this thesis. Further, in each of the studies, documents relating to the organisation’s IT dependability were collected. The construct validity is further improved through the relating of the approach of this research to similar studies. For example the maturity model is based on other maturity models from a number of related fields, and the proposed SLA template collects elements from a number of other previously published SLA studies.

The most relevant aspect of the validity to discuss in this introduction, is the external validity (the generalizability of the results). Unlike large surveys where the generalizability of the results can be statistically deduced from the sampling process, the generalizability of case study research is not as evident.

First, the question can be asked in how far the results can be generalised to other organisations of the same type as the subjects of the studies, for example to other municipalities in Sweden. There are large differences between individual Swedish municipalities, for example some municipalities have completely outsourced their IT services, which of course has a large influence on their IT management processes. Nevertheless all municipalities have the same responsibilities, and organise similar services for their inhabitants. The contacts with many different municipalities over the course of the different studies, give us some confidence that the results are of importance to many, but probably not all, municipalities. The goal of the maturity model and the SLA template presented in this thesis is then also to be one of the possible tools available to Swedish municipalities.

Secondly, the question can be asked how far the results can be generalised to other Swedish organisations. The results of this thesis are specifically targeted at governmental organisations with an active role in emergency management. This role in emergency management poses special requirements on the organisations’ IT dependability management. Therefore the results presented here are probably also of importance to such organisations such as regional governments or the emergency services. For businesses, where business value is the most important aspects of IT management, there are already many other studies and tools available such as the Business Alignment Maturity Model (Luftman 2003), ITIL (Office of Government Commerce 2007) or COBIT (ISACA 2000).

Thirdly, an important aspect of these results is also whether they can be generalized to an international setting. Both IT management and emergency management are significantly differently organised in other countries. For example, Watson et al. (1997) have shown large differences in the main issues in IT management in different countries and continents. Andersen and Kraemer (1994) have also shown a large difference between the introduction of IT technology in the
public sector between the United States and Scandinavia. They have, for example, shown that the way IT systems are introduced in the public sector in Sweden causes more direct changes in the work processes of the involved organisations. This would imply that they also have a more direct impact on the emergency preparedness of the organisation, making it more important to analyse the organisation’s IT dependability management. Emergency management in each country is coordinated by a national emergency management agency that publishes national guidelines for the different levels of governmental organisations. Sweden places a large responsibility for emergency management on the local level, and less on the national level compared to other countries (Sundelius et al. 1997). This further limits the applicability of the results of this thesis internationally. The generalizability of the results is even more difficult to judge because of the shortage of similar studies on this issue from other countries. But most importantly, one study by Vogt et al. (2011) has shown that similar problems like the ones discovered in Paper III exist in emergency management organisations in Germany and Australia, and also concludes that there is a need for a framework like the maturity model proposed in Papers IV and VI.

5 Conclusions

In Section 1 we stated that the main goal of the research presented in this thesis was to develop and evaluate methods for the analysis of the dependability of IT systems that take into account the risk of catastrophic events. This goal was further refined with three subgoals.

The statistical methods presented in Part I are mostly related to the first two of these three subgoals. In Part I, well-established theoretical methods for the use of usage profiles have been extended to include the sensitivity to different kind of changes in the usage profile. The applicability of these methods has been shown through the application to a real-life example, but this application also shows the limitations of this approach, in particular the difficulty of obtaining a detailed usage profile needed as input for these methods.

The research results presented in Part II of this thesis are directly related to the last two subgoals defined in Section 1. The maturity model and the SLA template are two results that are directly applicable by interested actors in the Swedish emergency management system. The methods have been developed based on empirical studies of the problems that are experienced by these actors today, which were identified in the study described in Paper III. Both of these frameworks are based on common frameworks that are used in related fields and that have been adapted to provide a solution to the most pressing problems that had previously been identified. The IDEM3 maturity model and the SLA template have been developed with input from practitioners, and both frameworks have also been partly empirically evaluated with the help of practitioners from many different organisa-
tions. Both have received very positive feedback from all the respondents. Both frameworks are now ready to be tried by a larger group of organisations and over a longer period of time, to allow the effectiveness of these initiatives to be evaluated more thoroughly.

Together these two parts fulfil each of the goals stated at the start of this research project, although a lot of work remains to be done to help Swedish emergency management actors and organisations to improve their IT dependability management practices.

6 Further Research

Future work on this research can focus on continuing the work in both separate parts of this thesis, but also on combining some of the results of both parts.

6.1 Sensitivity of Software Reliability

A first important continuation of the work in Part I would be to apply the results to a wider range of systems. The main obstacle for this application is the lack of easily available usage profiles. Although usage profiles are a widely used tool in software engineering, few actual detailed usage profiles are collected. The second paper in this thesis shows that even when extensive logfiles are available, the extraction of a usage profile of a system is not an easy task. Many opportunities for further research on the collection of usage profiles are available in this area.

Another logical next step in the research from the first part of this thesis concerns the nature of the changing usage profile over time. More research is necessary on how future changes in the usage profile of a critical system can be predicted. This is especially important when taking into account the occurrence of rare, critical events such as crisis situations typically are. It is known that the usage of an IT system can change drastically during a crisis situation (Andersson et al. 2005), not only concerning the load, but also in the shape of the underlying usage distribution. Nevertheless, little is known on how we can predict the changes to the usage in more detail. This could, for example, be further explored by studying in detail how user behaviour of different systems changes during crisis situations, as was studied for webservers by Andersson et al. (2005).

6.2 Dependability at Swedish Municipalities

There are many opportunities for future work based on the research results presented in Part II of this thesis. The maturity model and even the SLA template have only been evaluated over a short period of time. For a thorough evaluation of these results, more organisations should be involved in trials of these frameworks and the long-term effects of these initiatives should be studied. This way it would be possible to measure whether these methods can lead not only to improved IT
6 Further Research

dependability management practices but also to improved IT dependability outcomes - i.e. an improved capability of the organisation to deal with emergency situations that impact the organisation’s IT systems.

To study the effects of these improvement initiatives on an organisation’s IT dependability management over time, future research studies are currently in the planning phase. For these studies we will shift more from the case study approach to an action research approach.

The maturity model itself has many attributes and a more thorough treatment of each of the attributes and the relations between them could also be a rich source of continued research in this area.

Another possible continuation of this research is to study the generalizability of the results to other governmental actors and even to an international setting. Many of the problems described in the second part of this thesis are likely to be common to many other organisations than just Swedish municipalities and hospitals, even if their responsibilities are different and their internal organisation can be substantially different. This could for example be accomplished through a close cooperation with research projects with similar goals in other countries.

6.3 Combining Parts I and II

Finally, because this thesis consists of two distinct parts, a natural opportunity for continued research would consist in combining results from both parts. The results of the second part show that at this moment most Swedish municipalities do not collect the necessary usage data to use the advanced models developed in Part I.

When, in the next years, Swedish emergency management actors continue to improve their processes for assessing the dependability of IT systems in emergency situations, this might include the measurement and prediction of detailed usage profiles that can be used as input for the models described in the first part of this thesis. At this point the models developed in the first part of this thesis can be a useful tool in the risk and vulnerability analyses conducted at various Swedish emergency management actors. Therefore another possible continuation of the work in this thesis could focus on methods and tools to help municipalities collect the necessary usage profiles from different critical systems.
Bibliography


PART I: SENSITIVITY OF SOFTWARE RELIABILITY TO CHANGES IN THE USAGE PROFILE
Abstract

Usage profiles are an important factor in software system reliability estimation. To assess the sensitivity of a system’s reliability to changes in the usage profile, a Markov based system model is used. With the help of this model, the statistical sensitivity to many independent changes can be estimated. The theory supports both absolute and relative changes and can be used for systems with or without a terminal state. With this approach it is possible to very quickly estimate the uncertainty on the predicted reliability calculated from a Markov model based upon the uncertainty on the usage profile. Finally the theory is applied to an example to illustrate its use and to show its validity.
1 Introduction

Software plays an increasingly important role in today’s society. Reliable software is a prerequisite for most functions in our daily life: power supply, transportation, medical care etc... The software reliability depends not only on the defects residing in the software system, but also on how the software is used, i.e. the usage profile (Musa 1994a). It is often very hard to determine this usage profile exactly and usually it changes over time. However, when the reliability of a software system is estimated, an uncertainty on this usage profile is rarely taken into account.

In this paper we present a quantitative study on the sensitivity of the reliability estimate to changes in the usage profile. After the discussion of some related work in Section 2, Section 3 discusses the theory and in Section 4 this theory is applied to an example from Poore et al. (1993). In Section 5 we adapt our theory for a different class of systems. Finally, in Section 6 the conclusions of this paper are summarized.

2 Related Work

2.1 Usage Profiles and Reliability

The usage profile, or the operational profile, describes the users’ utilization of a software system in terms of the user-initiated events and the probability for these events (Musa 1993). The usage profile is mirrored in the utilization of the software and its components. Thus, a state-based Markov model of the software may be developed, with the probabilities from the usage profile, determining the transition probabilities of the system model. This approach is proposed by Cheung (1980) and later refined by Siegrist (1988a).

In order to use the system model for reliability estimation, an explicit failure state has to be added. The direct unreliability is then expressed as the probability for a failure in each system model state. The reliability of the entire system, for a given usage profile, can then be calculated from the system model (Cheung 1980).

This model can be used to model systems where the reliability can be defined as the probability to terminate in the success state and not the failure state. An example of such a system could be a server where users log in for a session to execute a number of transactions and to finally conclude with a successful termination of the session.

For systems without a terminal success state, the reliability is usually expressed as the mean time to failure. In this paper we first limit the discussion to systems with a terminal success state and then describe how the results can be adapted for systems without such a state in Section 5.
2.2 Sensitivity Analysis

As the software reliability depends on the usage profile, the question arises how sensitive the reliability estimate is to changes or uncertainties in the usage profile. The question has been addressed from two different perspectives: based on software reliability growth models (SRGM) (Musa 1994b, Pasquini et al. 1996, Wesslén et al. 2000), and based on Markov models.

With Markov models Poore et al. (1993) and Yacoub et al. (2004) analyze the sensitivity of a system’s reliability to the reliability of its components. Lo et al. (2005) present an analytic approach which identifies the most sensitive parameter, component reliability and transition flow.

The sensitivity to many changes in the transition probabilities has been discussed by Goševa-Popstojanova and Kamavaram (2004) with two methods based on Markov models: the method of moments and Monte Carlo simulation. Both methods give very good results for small systems, but require a large amount of calculation and hence don’t scale up well for large systems.

Our method is a simplification of the method of moments, and in Subsection 3.6 we discuss shortly how our results can be derived from the more general formulas by Goševa-Popstojanova and Kamavaram (2004). The advantages of our method are that our method is simpler, requires less computation and does not require the explicit estimation of a huge number of covariances between different transition probabilities. This makes our method more practical to use, especially for large systems. The method of moments has the advantage that it is more general and gives more exact results when all the covariances are known. In Section 4 we use the Monte Carlo simulation described by Goševa-Popstojanova and Kamavaram (2004) to check the validity of our results on a small example.

3 Sensitivity analysis

In this section we first shortly repeat some theoretical basics from Siegrist (1988a), then the maximal theoretical sensitivity of the reliability estimate to one change in the operational profile is discussed in Subsection 3.2. Next the statistical sensitivity to many random changes is discussed and some limitations and solutions are further examined in Subsections 3.4 and 3.5. Finally Subsection 3.6 discusses the relationship to the results by Goševa-Popstojanova and Kamavaram (2004).

3.1 Definitions

We use the Markov model as proposed by Siegrist (1988a) to model a system’s usage and reliability. The states of the model can either represent system states or system components between which control is passed.

In the model used here, there are two main assumptions. First, the Markov property means that the future behaviour of the system is determined only by the
current state of the system and not by the history of the system. Secondly, this model assumes that the system contains exactly two terminal states: a success state $t$ and a failure state $f$. This means that a run of the system always terminates in one of these two states. In addition to the two terminal states, the system also contains $n$ transient states $1, 2, \ldots, n$. State 1 represents the initial state.

The dynamics of the faultless system, without the failure state, are described by a Markov chain with state space $1, 2, \ldots, n, t$, and with transition matrix $P$, where $p_{ij}$ is the probability to go from state $i$ directly to state $j$.

In the imperfect system, every state has a designated reliability $r_i$, which means it has a probability $1 - r_i$ of directly entering the failure state $f$. The dynamics of the faulty system are described by a Markov chain with state space $1, 2, \ldots, n, t, f$ and with transition matrix $\hat{P}$, given as follows:

$$\hat{p}_{ij} = p_{ij} \times r_i \quad \hat{p}_{tt} = \hat{p}_{ff} = 1 \quad \hat{p}_{tf} = \hat{p}_{ft} = \hat{p}_{fi} = 0$$

for $i = 1, \ldots, n$ and for $j = 1, \ldots, n, t$.

The method for computing system reliability from these transition probabilities is based on standard Markov chain theory (Kleinrock 1975). Let $\hat{Q}$ denote the restriction of the transition matrix $\hat{P}$ to the transient states $1, 2, \ldots, n$, so the transition matrix $\hat{P}$ without the last two rows and without the last two columns. Then the matrix $V = (I - \hat{Q})^{-1}$ is called the potential matrix of the system. Each value $v_{ij}$ gives the number of expected visits to state $j$ before terminating when the system is currently in state $i$. Since state 1 is the starting state of the system, the system’s expected number of transition periods before terminating is the sum of the elements of the first row of the matrix $V$.

Let $T = (t_i)$ and $F = (f_i)$ be the column vectors containing the probabilities to go directly from a given state to state $t$ or $f$ respectively. Then the probability $s_i$ to finally end up in terminal state $t$ given that the system is currently in state $i$, or in other words the overall chance on success starting from state $i$, can be calculated as follows:

$$S = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = V \times T = (I - \hat{Q})^{-1} \times T$$

Because we assume that state 1 is the initial state of the system, the reliability of the whole system, is simply $s_1$. For each of the states the following equality holds:

$$s_i = \sum_{j=1}^{n} s_j \times \hat{p}_{ij} + t_i$$

Siegrist (1988a) uses this formula is to calculate a maximum to the sensitivity of the system’s reliability to the direct reliability $r_i$ of a state $i$. In the following
sections we build further upon this theory to investigate much further how changes in $\hat{P}$ influence the total reliability of the system.

### 3.2 Sensitivity to One Change

To make a distinction between the old system without the small changes and the new system with the small changes, we use an accent to indicate the variables of the new system. So we investigate the difference between the old reliability $s_1$ of the system with transition matrix $\hat{P}$ and the changed reliability $s'_1$ of the system with changed transition matrix $\hat{P}'$.

Let $\delta_{kl} = \hat{p}_{kl}' - \hat{p}_{kl}$ denote the change to $\hat{p}_{ij}$ and let $\delta_{\text{Rel}} = s'_1 - s_1$ denote the resulting change in the total reliability of the system.

First of all, it’s important to notice that it is impossible to change only one probability $\hat{p}_{kl}$ in the matrix since the sum of the probabilities of one row has to equal 1. One possible solution is to always make two opposite changes within the matrix $\hat{Q}$, to two transition probabilities $\hat{p}_{kl}$ and $\hat{p}_{km}$,

$$\delta_{kl} = \hat{p}_{kl}' - \hat{p}_{kl} = \hat{p}_{km} - \hat{p}_{km}' = -\delta_{km}. \tag{3}$$

For small changes the resulting change in system reliability can then be approximated by

$$\delta_{\text{Rel}} \approx \delta_{kl} \times v_{1k} \times (s_l - s_m) \tag{4}$$

This formula also holds when the states $l$ or $m$ are one of the terminal states, when we consider that the chance of success starting from the states $f$ and $t$ is respectively 0 and 1.

This also means that the reliability is most sensitive to changes in the transition probability $\hat{p}_{kl}$, from the state $k$ that has the higher number of expected visits, to the states with the highest and lowest total chance of leading to success. This corresponds to the findings published by Lo et al. (2005).

Instead of always making two opposite changes, another option is to make a change $\delta_{kl}$ to the transition probability $\hat{p}_{kl}$ and then to rescale the whole row so the sum becomes 1 again. This method we call spread changes as opposed to the paired changes described above. This is equivalent to making $n+1$ paired changes of size $\frac{\delta_{kl}}{1+\delta_{kl}}$ from the probabilities $\hat{p}_{ki}$ ($i \in \{1, \ldots, n, t, f\} \setminus \{l\}$) to $\hat{p}_{kl}$.

When adding together the effect of all these changes and ignoring second order effects, we can approximate the change in the reliability of the whole system caused by such a spread change as

$$\delta_{\text{Rel}} \approx \frac{\delta_{kl}}{1 + \delta_{kl}} \times v_{1k} \times (s_l - s_m) \tag{5}$$

This formula follows directly from equations (4) and (2). For small changes we can ignore the denominator $1 + \delta_{kl}$ and still retain a good approximation. This will simplify the statistical sensitivity analysis in the following sections.
3.3 Statistical Sensitivity to Absolute Changes

In this section we look for the effect of a large number of small changes or uncertainties in the usage profile on the system’s overall reliability. First we look at the most simple case. Here we assume that all the $\delta_{kl}$ have an identical and independent distribution with mean $0$ and variation $\sigma^2_{\delta}$. This is equivalent to assuming an equal absolute uncertainty on all the transition probabilities.

$$E(\delta_{kl}) = 0, \quad \sigma^2(\delta_{kl}) = \sigma^2_{\delta}, \quad k, l = 1, \ldots, n$$ (6)

Because we can not change one transition probability alone, for every change $\delta_{kl}$ we select a random transition probability $\hat{p}_{km}$ on row $k$ that will undergo the opposite change $\delta_{km} = -\delta_{kl}$. This will guarantee that the sum of the transition probabilities from each state is always equal to 1.

From standard statistics we know that:

$$\sigma^2(s_i - s_j) = 2 \times \sigma^2(s) = 2 \times \sigma^2_s, \quad i, j = 1, \ldots, n$$ (7)

Now the total change in system reliability resulting from randomly and independently changing all the elements $\hat{p}_{kl}$ as described above, is the sum of all the changes resulting from $n$ independent changes on each of the $n$ rows. And therefore the total change in reliability will have a distribution with the following variance:

$$\sigma^2(\text{Rel}) \approx 2 \times n \times \sigma^2_{\delta} \times \sigma^2_s \times \|V_1\|^2$$ (8)

Because the total change in reliability is the sum of a large number independent small changes, the distribution is close to a normal distribution with mean 0.

When we also want to account for random changes in the transition probabilities in the vectors $T$ and $F$, equation (7) for $\sigma^2_s$ can simply be extended to include the states $t$ and $f$ with $s_t = 1$ and $s_f = 0$. This would of course seriously increase $\sigma^2_s$ and therefore also the variation on the reliability.

Another option is to apply a random spread change to each of the transition probabilities of the matrix and all the $\delta_{ij}$ have the same distribution with mean 0 and variation $\sigma^2_{\delta}$. Because the different factors in equation (5) are not independent we can not make the same simplification as with paired changes. In this case the standard deviation of the reliability can be approximated by

$$\sigma(\text{Rel}) \approx \sigma_{\delta} \times \sqrt{\sum_{k,l} [v_{1k} \times (s_i - s_k)]^2} \quad k, l = 1, \ldots, n$$ (9)

3.4 Limitations

It is important to understand that the equations (8) and (9) only hold under a number of assumptions. First of all the changes to the system’s transition probabilities
have to be sufficiently small, for equation (4) and (5) to be a good approximation. When the changes become too large, the structural changes in the system model become more important and influence the reliability of the system.

Secondly, this model also assumes that positive and negative changes are equally likely for each of the transition probabilities. In practical examples this is rarely true. For example, most systems contain a large number of transition probabilities equal to zero to indicate impossible transitions. The error on this transition probability can only be negative or zero. When we apply random changes to a system with many zeroes in its transition matrix while only allowing positive changes to the zeros, then the expected change of the reliability will no longer be zero but be biased towards a negative or positive change depending on the structure of the system. But in most cases equation 8 will still be a good indication of the magnitude of the expected reliability change.

3.5 Statistical Sensitivity to Relative Changes

To overcome these problems we can use relative changes, which means we assume a larger absolute uncertainty on larger probabilities. This will of course also make sure that impossible transitions remain impossible. For most systems this is a much better model of the uncertainty on the transition probabilities.

Here we have to use the spread changes discussed before, since it makes no sense to make paired changes relative to the size of one of the changed probabilities without taking the size of the other probability into account.

If we assume an equal relative change or uncertainty \( \sigma_s \) for all of the transition probabilities then the total standard deviation of the overall reliability can be approximated by

\[
\sigma(\text{Rel}) \approx \sigma_s \times \sqrt{\sum_{k,l} [p_{kl} \times v_{1k} \times (s_l - s_k)]^2}
\]

\[k = 1, \ldots, n \quad l = 1, \ldots, n, (t, f).\] (10)

When we want to introduce even more detail into the model and want to model specific uncertainties for different transition probabilities we can adapt the equation above in the following way:

\[
\sigma(\text{Rel}) \approx \sqrt{\sum_{k,l} [\sigma_{kl} \times v_{1k} \times (s_l - s_k)]^2}
\]

\[k = 1, \ldots, n \quad l = 1, \ldots, n, t, f.\] (11)

where \( \sigma_{kl} \) is the standard deviation (or in other words the uncertainty) on the transition probability \( p_{kl} \).
3.6 Relation to the Method of Moments

Our formulas can be seen as a special case of the method of moments presented by Goševa-Popstojanova and Kamavaram (2004). The first order method of moments requires the calculation of all derivatives of the reliability to all the transition probabilities, and further requires the estimation of about $O(n^3/2)$ covariances between the transition probabilities.

This huge number can make the method unpractical to use for systems with a large amount of states. The covariances cannot simply be neglected because the sum of the transition probabilities from each state has to equal 1 which typically creates many negative covariances between these transition probabilities.

To show the relationship between the formulas by Goševa-Popstojanova and Kamavaram (2004) and our formulas, we just have to calculate those first order derivatives of the reliability and calculate the covariances implied by paired or spread changes.

In the derivation of our formulas we assumed that all $E(\delta_{kl})$ are equal to zero. This is a quite logical assumption, since we calculate the reliability with the transition probabilities that we consider most likely. With this assumption the formulas for the first order derivatives of the reliability to the transition probabilities can be reduced to

$$\frac{\partial R}{\partial p_{ij}} = v_{1k} \times s_l.$$  \hspace{1cm} (12)

By using paired or spread changes we are implicitly making assumptions about the covariances between the different transition probabilities. This at the same time makes the formulas less general but also more practical to use when no such detailed information is available. With standard statistical methods the new absolute variances and covariances implied by spread or paired changes can be calculated. For example, spread changes implies

$$\text{Cov}(p_{ki}, p_{kj}) = \sum_{l=1}^{n} \sigma_{kl}^2 \times p_{ki} \times p_{kj} - p_{ki} \times \sigma_{kj}^2 - p_{kj} \times \sigma_{ki}^2.$$  \hspace{1cm} (13)

By incorporating these results into the results previously published by Goševa-Popstojanova and Kamavaram (2004), the formulas can be simplified to the results presented in formulas (9), (10) or (11).

4 Example System

In this section we apply the theory from Section 3 to an example system from Poore et al. (1993). The transition graph of the faultless system can be seen in Figure 1. The reliability is $s_1 = 0.9899$ and the expected number of periods is $\sum v_{1i} = 67.056$. 
4 Example System

Figure 1: Example system without the failure state. Each directed arc, labelled with its transition probability, indicates that control passes from one component to another.

4.1 Statistical Sensitivity to Absolute Changes

With equation (8) we can predict the standard deviation of the system's reliability for 144 random changes with $\sigma_\delta = 0.005$ to be $\sigma(\text{Rel}) = 2.24 \times 10^{-4}$.

To check this result we conducted a Monte Carlo simulation. The resulting reliability change from 500 simulation runs can be seen in Figure 2. The results of the experiments fit very well with the predicted changes indicated by the dashed line. This indicates that the real standard deviation is close to the predicted value.

4.2 Limitations

As we discussed in Section 3.4, the formulas that predict the statistical change in reliability become less precise when the $\sigma_\delta$ increases. For example in Figure 3 we can see what happens when $\sigma_\delta = 0.01$. For 90% of the experiments, our predictions are still quite accurate, but we also notice a group of outliers where the changes have a large influence on the system’s dynamic behaviour with a large change in the reliability as a consequence. This is due to the fact that the matrix $P$ contains some small transition probabilities where a change of a few times $0.01$ can already have a large effect.

4.3 Statistical Sensitivity to Relative Changes

The results of a similar simulation with relative changes of 10% can be seen in figure 4. Here the predicted uncertainty on the reliability is $3.9 \times 10^{-4}$. Working with relative changes allows us to accurately predict the sensitivity to quite large changes. When we would include an equal relative uncertainty of 10% on the
**Figure 2:** Cumulative normal plot of the predicted (dashed line) and experimental (500 dots) sensitivity to 144 random paired changes with $\sigma_\delta = 0.005$.

**Figure 3:** Cumulative normal plot of the predicted (dashed line) and experimental (500 dots) sensitivity to 144 random paired changes with $\sigma_\delta = 0.01$. 
transition probabilities to go to the terminal states, the predicted uncertainty on the reliability would be $1.2 \times 10^{-3}$, which is about 12% of the probability for a run of the system to terminate in the failure state.

5 Alternative System Model

In some kind of software systems there is no clear success state and ideally the system would never terminate, since the only way for the system to terminate is in a failure. For these kinds of systems, reliability is usually expressed as the time between failures of the system (Mean Time To Failure).

These systems can be modelled by a similar Markov chain, but with only one terminal state $f$ (Siegrist 1988b). Here the mean time to failure, expressed as a number of transitions before termination in the terminal state $f$, given that the system is presently in state $i$, can be calculated as

$$MTTF_i = \sum_{j=1}^{n} v_{ij}.$$  \hspace{1cm} (14)

For this model we can make a similar analysis. For example for spread, relative changes we find:

$$\sigma(\text{Rel}) \approx \sigma_\delta \times \left( \sum_{k,l} \left[ p_{kl} \times v_{1k} \times (MTTF_l - MTTF_k + 1) \right]^2 \right)^{1/2} \quad k = 1, \ldots, n \quad l = 1, \ldots, n, (f)$$  \hspace{1cm} (15)
6 Summary and Future Work

In this paper we have made a quantitative study of the sensitivity of the reliability estimate to changes in the usage profile. With this theory it is possible to make a good estimate of the effect of many small changes in the usage profile on the reliability of a system. The main advantage of this methods over the related method of moments from Goševa-Popstojanova and Kamavaram (2004) is that it does not require as much information and is computationally much simpler. The disadvantage is that it is less general and can not be used when very specific information about the covariances between the transition probabilities needs to be taken into account.

When comparing the theory with some experimental results, the results are very good, taking into account a number of limitations described in Subsection 3.4. For absolute changes the formulas only give a good approximation for small changes, but for relative changes the formulas can quickly give a realistic estimate of the uncertainty on a reliability estimate from a Markov usage model.

Further research will be done on the exact nature of the uncertainties in the usage profile of different systems. Another very important goal is also to apply this sensitivity analysis to a real operational system to further investigate the applicability of these results.

Bibliography


Abstract

To measure the reliability of a website from a user’s point of view, the uncertainty on the usage of the website has to be taken into account. In this paper we investigate the influence of this uncertainty on the reliability estimate for a web server. For this purpose a session based Markov model is used to model the usage extracted from the server’s logfiles. From these logfiles a complete user profile can be extracted together with an estimate of the uncertainty on this user profile. This paper investigates the applicability of this kind of Markov model on web server reliability and discusses the difficulties with data extraction from the logfiles. Advantages and disadvantages of this approach are discussed and the approach is applied to data from a university department’s web server to demonstrate its applicability.
1 Introduction

The Internet plays an increasingly important role in today’s society. We have become more and more dependent on the Internet for critical functions in our society. In the event of a crisis in our society the Internet can play an important role in the information spreading (National Research Council 2003). The most important quality attribute when accessing critical information or services on a web server is the reliability. This reliability is very hard to measure because it is a combination of reliability on the client side, on the network and on the server side (Ma and Tian 2007). In this paper we focus on faults caused by broken links between different parts of the website. For the reliability from a user perspective it is not only important how many broken links are present, but also how often they are used.

Just as for most software systems, reliability of a web server should be measured based on the usage profile. It can be hard to determine this usage profile exactly and it often changes over time, especially when the webserver is updated. In this paper we wish to take the uncertainty on this usage profile into account. Therefore we investigate how a Markov model for the usage can be extracted from the logfiles of a web server, and how it can be used to measure and predict the reliability of the web server.

2 Related Work

The reliability of a website could be defined and measured in many ways. Ma and Tian (2007) divide the failures that can occur when accessing information from a website into three categories according to their source: ‘host, network or browser failures’, which are basically failures in one of the technical subsystems from content provider to the user, ‘user failures’, which are faults directly caused by the user and ‘source or content failures’, such as for example faulty or broken links between webpages. This last category, on which we focus in this paper, is specific for the web server domain and in this paper we try to model and measure reliability related to exactly these failures.

Sampath et al. (2004) have successfully used the logfiles of a webserver to generate usage profiles for automated testing, and Goševa-Popstojanova et al. (2006) study the logfiles to measure the reliability from a user perspective. In the log files, as in Figure 1 the HTTP return code of every request indicates whether the server was able to provide the data requested by the user. For example the normal return code is ‘200’ when the content was correctly returned. When the requested document was not found, the return code is ‘404’ and an error page was displayed. The complete list of possible return codes for HTTP version 1.1 is defined in the specification RFC 2616 (Mogul et al. 1999).

With these recorded return codes it can be measured how many of the requests were successfully answered, and Goševa-Popstojanova et al. (2006) called this the
request-based reliability. Because each user session usually contains many requests, Goševa-Popstojanova et al. (2006) also define the session-based reliability, where all sessions that contain at least one failed request are considered as failed.

The main research question in this paper is how sensitive the session-based reliability of a website is to the uncertainty on the usage. Another question we try to answer is how a change in the usage, for example because of the changing structure or content of the website, influences the reliability.

To study this sensitivity we use the techniques described by the authors of this paper in (Weyns and Runeson 2007) for sensitivity analysis of software reliability modelled by Markov chains, which at the same time gives us the opportunity to test the applicability of these methods on a much larger system than before.

3 Data Processing

3.1 Important Issues

Most webservers automatically log all requests to logfiles. Before this logged data can be used to build a usage profile for the web server that can be used in reliability prediction, many issues have to be sorted out. To extract the user profile we take the following steps:

1. collect the logfiles, which are usually very big,
2. filter out web robots and irrelevant document types,
3. generate a list of documents,
4. compile the requests into user sessions,
5. generate a matrix with the amount of transitions,
6. convert the matrix to transition probabilities.

Some of the steps that deserve special attention are discussed in the following paragraphs.
Filtering

When analysing the log data from a web server it can be seen quite fast that a substantial part of the requests are not from actual users, but from all kinds of web robots, commonly known as Internet bots or simply bots, crawling the website. Since we are not interested here in the reliability that the bots experience and because they do not behave as normal users, we need to filter out these requests. This is not a trivial task and can be done in many ways. Tan and Kumar (2002) discuss both simple and more advanced techniques for robot detection. One of the most common ways is by excluding all requests from a specified list of known search bot IP’s. Other types of bots can usually be found by inspecting some statistics from the logfiles to search for irregularities, such as an unusually high number of requests from the same IP address. Some bots might escape detection, but those are then only responsible for a small part of the traffic and can therefore probably be safely ignored.

Other IP addresses we might exclude from the analysis are those addresses that do not represent the user group we are interested in, such as for example the organisation responsible for the updating of the web server. If we are for example only interested in external users it is possible to exclude those IP addresses coming from local computers. This method could also be used to group the users based upon the IP address, for example by country, in case we wish to generate a number of separate user profiles.

Secondly, we need only consider the requests that concern actual pages, and not the images or other elements that are requested while the page is being loaded. From a user perspective, a session consists of visiting a number of pages in a certain order, usually by clicking on hyperlinks to navigate from one page to the next. Therefore we can simply filter the requests by extension. Depending on the structure and technology used by the website, we could for example retain the ‘.html’, ‘.jsp’ or ‘.pdf’ pages but filter out the ‘.jpeg’, ‘.gif’ or ‘.css’ files.

Listing Documents

After the filtering we can compile a list of all documents requested in the logfile. When pages are dynamically generated by the web server, it requires some consideration whether the same page with different parameters is considered as one document or as many different documents. This depends mostly on how strongly the dynamic content influences which pages can be reached from the dynamically generated page. If for example no links to other pages are generated dynamically then the parameters can often be ignored. For certain types of dynamic pages the content generated can depend heavily on session information that is not recorded in the logfiles and then the session of the user can not be reconstructed from the logfiles alone and the techniques used in this paper can not be used when only the logfiles are available.
In a similar way we can opt to group together some similar pages as one state in our model. This could be useful if we do not wish to go into detail about the navigation of users in one or more parts of the website, for example all pages concerning a special project. This is especially useful if this part of the website is only visited a small number of times. Grouping these pages as one node means we can limit ourselves to studying the transitions in and out of this part of the website, and ignore all transitions internally in this group. The same is done by Li and Tian (2003). If part of the website uses frames, it can also be logical to group frames that are always loaded together, or to filter out frames that are often reloaded.

Sessions

From all the remaining requests we can now try to reconstruct the users’ sessions, i.e. all the transitions from document to document by the users when surfing the website. For example the requests in Figure 1 together represent one successful session. Every session begins by a first request, which can be the site’s main page, but can also be another page that was reached directly from an external referrer. Every session ends either by the users leaving the website (which we can only detect by an absence of requests from this user for a given amount of time) or by a request that results in error, after which the session is considered closed.

To group the requests in sessions, we need to group the requests by user and sort them into sessions. There are some problems with this approach. If different users are behind the same firewall or proxy server it might not be possible to discern these users from just the IP addresses. Also, when a user presses the ‘back’-button in his web browser, the last page is not always reloaded and then no request is sent to the web server. The same is true if the user is looking at cached pages being stored either locally on the user’s computer or at another location in the network. These transitions are then of course not logged in the server log. Probably, this is only a small part of the transitions, and can often be safely ignored.

To group requests into sessions we have to first define a session time-out. This means that we consider a request to be the start of a new session if there was no request from the same IP address for a given period of time, for example 30 minutes. A more detailed discussion of the influence of session time-out length can be found in the paper by Goševa-Popstojanova et al. (2006).

To reconstruct the browsing of a user we have two sources of information: the order of incoming requests (Goševa-Popstojanova et al. 2006) and the referrers logged for every request (Li and Tian 2003). For many users this information is consistent and the transitions can easily be deduced as is the case in Figure 1. However, when for example the user’s browser does not indicate the referring URL, or if a user surfs the website in multiple browser windows at the same time, the referrer information can be inconsistent with the order of the requests. When
constructing an algorithm to extract the sessions we have to take this problem into account.

Goševa-Popstojanova et al. (2006) define a failed session as any session containing at least one failed request (with return code 400 or higher). This means that a session is also considered failed if one image or style sheet can not be found. In this paper we employ a slightly different definition. First of all, we focus only on the pages and not their components, so we only consider a session failed if a complete page can not be returned. Secondly, we consider a session concluded after a failure, and therefore no session can contain more than one failure.

It can also be discussed whether a return code of ‘401: Unauthorized’ or ‘403: Forbidden’ should be considered a failure. Rejecting a non-authorised user from seeing a certain webpage should be considered correct behaviour from the web server’s point of view, but the user might experience this as a failure if he was expecting to have access to this webpage. The opposite of a failed session is a successfully concluded session, which is now defined as a series of requests from the same user which does not result in an error code and is concluded by a period of a certain length without requests from that user.

While compiling all the requests into sessions, we can filter out automatic reloads of the same page, since these actions do not reflect real transitions from a user point of view. Further it also makes sense to filter out all failed sessions consisting of only one request. Those failures are caused by a faulty link on external webpages or by mistyped paths. Since these faults are not under the control of the web server, they should not be included in the reliability of the web server. In the same way it makes sense to also exclude normal sessions that consist of only one request, since they do not actually contain any successful transitions through the website.

4 Markov Model

A Markov chain is a discrete-time stochastic process that has the Markov property (Winston 2003). Li and Tian (2003) conclude that Markov chains, extracted from logfiles can also be used as a good model for web usage. Empirical validation of the memoryless property typical for Markov chains shows that Markov chains are a practical and valid simplification.

By going through all the sessions, we can compose a large matrix listing the total number of transitions from every document to every other document. For most websites this matrix contains many zeroes, since most pages are only reachable from a few other pages. The matrix also records how often every document is the start or end of a session, either successful or leading to failure.

By then dividing every row by its sum, this matrix can easily be converted to estimates of the transition probabilities. This matrix of transition probabilities completely defines a Markov Model. With the help of this matrix the reliability
of the website, defined as the possibility that a session terminates successfully (by a session time-out) and not in failure (when an error code is returned), can be calculated with the help of the formulas from Siegrist (1988). If all sessions could be perfectly reconstructed from the logfiles, the total reliability calculated from the Markov model would be equal to the percentage of successfully completed sessions.

5 Sensitivity Analysis

As we saw in the previous section, there is no need for a complicated Markov Model if we just want to calculate the reliability of the web server. However, if we want to analyse the sensitivity of the web server reliability to the changes in the usage, this section details how we can use the theory from (Weyns and Runeson 2007) to give us a good estimate. So the research question we are trying to answer now is how sensitive the reliability estimate is to the uncertainty or to changes in the transition probabilities between the different documents.

When we define an uncertainty on the transition probabilities we have to keep in mind that it is impossible for just one of the transition probabilities to change, since the sum of all transition probabilities out of one state always has to equal one. To solve this problem we use the mechanism of spread changes defined in (Weyns and Runeson 2007). This simply means that we allow the transition probabilities out of one state to change independently and then in the end divide them all by their total sum to make the sum equal to 1 again. This corresponds to modelling changes to the number of transitions instead of the derived transition probabilities.

We also need to define how we estimate the size of the uncertainty on each transition probability. In this article we discuss two options to estimate the uncertainty on the transition probabilities: based on the binomial distribution or based on an identical relative uncertainty on all transition probabilities.

5.1 Binomial Uncertainties

For this option we assume that the usage is the random result of a constant distribution of the transition probabilities. The binomial distribution tells us how we can estimate the transition probabilities from a sample, and how for a large sample the confidence interval can be approximated by a normal distribution.

For those cases where only a small sample is available, the transitions that only occurred a few times or never, the uncertainty on the estimate is much larger, and less symmetric than the simple normal approximation. This is most extreme for those transitions that did not occur during the whole period for which logfiles are available and for which the probability is therefore estimated at 0. Most of these transitions are indeed impossible transitions, but a small part might be from links that were never used during the period of the logfiles but could be used sometime
later. For these cases the uncertainty is much harder to model and since all those transitions only contribute a small part to the overall reliability they are not worth the extra complexity from more advanced formulas such as can be found in (Brown et al. 2001). Nevertheless, we need to keep in mind that the estimate of the total uncertainty might be a slight underestimate if there are many pages for which the amount of requests is very low in the collected logfiles.

The result of applying the formulas of the binomial distribution is that the states that have been visited the least times, have the largest uncertainty. Since the behaviour in these states is also less important for the total reliability, we can expect the total uncertainty on the reliability to be relatively low.

5.2 Relative Uncertainties

The second option for modelling the uncertainty on the transition probabilities takes into account the changing content and structure of the website. After an update of the website, the usage is likely to change abruptly. Then the real variation of the transition probabilities is actually much greater than predicted by the binomial distribution. In this case it is not possible to predict the change in the usage without more detailed knowledge about the planned updates, and even with this information it would still be hard to predict the exact changes in the usage.

If we do not know exactly how the website will be updated we have to resort to a more general model for the uncertainty. It is logical to assume that in general the larger transition probabilities have a larger absolute uncertainty. However, because there are so many small transition probabilities the uncertainty there can not just be ignored. Therefore we propose to use relative uncertainties. This means that we assume an equal relative uncertainty to all transition probabilities measured from the logfiles, for example 10% of their estimated value. Just like in the previous section, this simple assures us that the transition probabilities that were estimated at 0, remain 0 and that we can model the uncertainty as being symmetrical around the estimated value. Both of these conditions need to be fulfilled to allow for a statistical analysis as described in (Weyns and Runeson 2007)

If we would have access to a large quantity of log data and information about past updates, we could statistically analyse the effect of the updates on the usage of different parts of the webservers from the logfiles, but this is out of the scope of this paper.

6 Application to the Research Group Webpages

In this section we describe how we applied the method described above to the website of our research group, of which the latest version can be found on http://serg.telecom.lth.se/. The logfiles of two and a half months were collected and analysed. During those months no important changes were made to the
website and we can expect the usage profile to be quite constant. The collected logfiles contain about two hundred thousand requests, and are together about 38 Mb in size.

6.1 Filtering
Filtering the traffic to remove the traffic caused by bots, reduces the amount of requests by 50%. The website under study contains relatively few images, compared to most commercial websites. Nevertheless, filtering out the images and other documents by file type we can still further reduce the amount of requests we need to analyse by an extra 31%. Most of removed requests consisted of images, CSS-stylesheets and JavaScript source code files.

As discussed in Section 3.1, we can further remove all requests that represent a refresh of the previous request and all failures with an external referrer. This last number is small, but is very significant compared to the total number of failures. Finally we also removed all sessions of length one since they do not represent a real session.

After filtering we are left with about 5% of the requests which represent the real user transitions. When we group these requests into sessions in the next step we add another 2281 implicit transitions that do not show up as requests in the logfile because they represent the leaving of the website at the end of every successful session.

6.2 Sessions
For grouping the remaining request into sessions we chose to base ourselves mostly on the referrer information in every request. When the referrer was unknown or empty, the referrer was replaced with the target of the previous request when this last request occurred in the last half hour. Requests with an external referrer or with an empty referrer after a longer period of inactivity were recorded as the start of a new session. Further transitions to the success state were added for the last document before each period of inactivity.

Just the referrer information does not make a consistent Markov model. To make our model a valid model we have to complete the states for which no outgoing transitions were recorded, with a possible transition to success.

Of the requests that were not filtered out, 99.3 % did not indicate a failure, 66 requests had a return code of ‘403′ indicating a user trying to access a page he is not authorised to see, and only 5 requests were requests with an internal referrer and an return code of ‘404’. No other failure codes occurred. Only those last 5 requests were the result of broken links on the website, caused by only three different broken links. Broken links that lead out of the website under consideration can of course not be detected from the logfiles. Only internal broken links are taken into account here.
When fixing the broken links, we can easily remove them from the model and also use it to estimate the reliability of the website when the broken links have been repaired.

Because the collected logfiles start and end in the middle of the night on local time, there were very few requests from real users close to the start and end time of the logfile. Therefore we detected only one session of which we do not know if it continued beyond the end of the logfile, and there were no sessions starting in the first half hour after the start of the logfile.

We also opted to combine all the administrative pages of the website into one state, as described in Section 3.1. Further, we also combined all pdf-files in each directory into one state since they do not contain links to other pages and are usually quite similar.

### 6.3 Markov Model

Now we obtain a Markov model with 216 states: the 213 documents, a start state, a failure state and a success state. With this model we can very quickly calculate some statistics of the Markov chain we have used to model our system.

The most important statistic is of course the estimate of the reliability of the system. In our system this is calculated to 98.4%, which is of course lower than the request-based reliability. The remaining 1.6% probability of failure can be divided in a probability of 0.1% to end up in an ‘error 404’-page, and 1.5% to end up blocked from an unauthorised page resulting in a ‘403’-error. The average session contains 4.73 transitions, and the session obtained by always following the maximum transition probability is exactly the session shown in Figure 1.

### 6.4 Sensitivity Analysis

With 13,000 transitions, we have 60 recorded transitions for every document, but only an average of 13 transitions for every of the 1042 different transitions that occurred at least once. There are many documents and even more transitions that have appeared only a few times or even only once. For those states and transitions it is impossible to make a good estimate of the transition probabilities. We can still use the binomial distribution to calculate an uncertainty on these transition probabilities, but the results are actually a slight underestimate.

Combining the formulas from (Brown et al. 2001) and (Weyns and Runeson 2007), we find an uncertainty of \( 9.4 \times 10^{-4} \) on the reliability of 0.9837. Mathematically, this uncertainty represents the standard deviation that can be used to compute a confidence interval. This represents about 6% of the probability to end up in failure, and is therefore significant.

A second method to estimate the uncertainty is to consider a relative uncertainty of for example 10% on all the transition probabilities. This roughly represents the intuitive notion that the calculated usage profile is a good indication but
is still a bit uncertain. With the formulas from (Weyns and Runeson 2007) this gives an uncertainty of $8.3 \times 10^{-4}$, or 5% of the probability to end up in failure.

When we would want to calculate the same result by a Monte Carlo simulation, this would require a substantial amount of calculation, and for systems with a few times as many states as the small website under consideration here, the simulation quickly becomes impractical. A 15 minute Matlab simulation on a normal desktop PC produced Figure 2. Each of the dots represents the reliability of a system of which all the transition probabilities have been randomly altered by using a normal distribution with standard deviation of 10% of their original size. The dashed line represents the predicted change in the reliability. The graph shows that the distribution of the simulated reliability conforms well to the predicted normal distribution with the standard deviation calculated above.

7 Summary and Future Work

We have shown how the logfiles of a web server can be used to extract a usage profile from which we can estimate the reliability of the web server from a user’s point of view. This can also be used to calculate the uncertainty on this reliability estimate. By applying the theory to the website of the research group, we have at the same time shown that the formulas from (Weyns and Runeson 2007) can be applied to a larger system in a useful way when usage statistics can be collected.

The disadvantage of this approach is that it requires a lot of pre-processing of the data, before a usage profile can be extracted. At the moment there is little tool
support for this. Further, the approach is limited by the data logged in the logfiles, we can, for example, not detect any broken links leading out of the website.

Further work on this approach could include improving tool support, applying it to more and larger websites and exploring the influence of incorporating more advanced formulas for the confidence intervals from the binomial distribution from (Brown et al. 2001). An important extension could also be to combine this approach with more detailed predictions of how the usage of the website is expected to change in the near future. These predictions could then be used to predict the reliability of the website after changes to the website or when the usage is expected to change substantially.

Bibliography


PART II: DEPENDABILITY OF IT SYSTEMS IN EMERGENCY MANAGEMENT
Abstract

In recent years governmental actors have become more and more dependent on IT systems for their responsibilities in a crisis situation. To avoid unexpected problems with the dependability of IT systems in the aftermath of a crisis it is important that such risks are identified and that measures can be taken to reduce the dependence on systems that could be unreliable. This paper describes two case studies exploring how Swedish municipalities incorporate IT systems in their emergency planning. The study focuses especially on how different actors within a municipality cooperate to analyse the risks of depending on IT systems in critical situations. The study shows that today there is much room for improvement, especially in the communication between IT personnel and emergency managers. Finally, this paper describes the requirements for a process improvement framework that can assist governmental actors in analysing and improving their dependency on IT systems in emergency management.
1 Introduction

In recent years governmental actors have come to depend more on IT systems for all their everyday tasks. For communication, municipalities depend on landline telephone networks, mobile phone networks, web servers, email servers, etc. Other important systems are used for patient administration in health care and social care, school administration or city planning.

Just as for their everyday tasks, governmental actors now depend on all kinds of IT systems for their responsibilities in crisis situations. These systems include not only specially built systems for emergency situations but also the everyday systems described above. The latter category of systems is of special interest, because under normal conditions an occasional unavailability of these IT systems might be acceptable, but during crisis relief, when time is a critical factor, any unexpected unavailability can have disastrous consequences. Therefore, it is important that these IT systems are an integral part of all major risk and vulnerability analyses conducted.

Based on two case studies and a survey, this article presents how Swedish municipalities, with an important active role in crisis relief, include IT systems in their emergency planning and vulnerability analyses. This paper first focuses on the main problems experienced by practitioners today and then shortly describes the requirements for a process improvement framework that could help organisations to improve the way they deal with IT dependability in emergency management.

2 Background

2.1 Dependability

For all software engineering concepts concerning dependability we will follow the definitions from (Avizienis et al. 2004). This means that dependability takes into account all more specific aspects such as reliability, availability, safety and security and corresponds best to the intuitive notion of how much a system can safely be depended upon by its users.

2.2 Emergency Management in Sweden

Swedish emergency management (KBM 2005) is mainly based on the ‘principle of responsibility’, which means that in emergency conditions the responsibilities for everyday matters should still lie with those governmental actors that are also responsible for these matters in normal conditions. Through the principles of proximity and geographic area responsibility, emergency management is in the first place a responsibility of the local governments. Practically, this means that municipalities are the central actor in crisis relief. Only with crises that affect many municipalities the regional governments are directly involved in an operative role.
For their emergency planning, Swedish municipalities receive support from the Swedish Emergency Management Agency (SEMA). SEMA assists the municipalities by educating them about emergency planning and by providing guidelines. Unlike the emergency management agencies in many other countries, SEMA does not have an operative role in crisis relief. An extensive description of Swedish emergency planning at the municipal level can be found in (Hallin et al. 2004).

3 Related Work

Internationally, more and more research is being done on special systems that can be used in crisis relief. The near future will almost certainly see a quick rise in the number of IT systems used in crisis situations. So far most of these systems are only considered as an extra tool in the aftermath of a crisis, but as these tools become more common, emergency responders will also become more critically dependent on them. Therefore it will become even more important to fully integrate these IT systems into emergency management and include them in the vulnerability analyses that are conducted.

In the private sector, a lot of research has been done on improving the cooperation between an organisation’s IT department and the rest of the organisation. For example, Luftman (2003) describes a maturity model for improving this cooperation, which is similar to the process improvement framework we propose in section 7 of this paper. The area of IT management that is most relevant to this research is often called business continuity management, which is concerned with maintaining a reasonable level of service during emergency situations. However, an important difference is that, unlike private actors, governmental actors with emergency relief responsibilities have to attain an even higher than normal level of service during emergency situations, which poses special requirements on their IT management. Conditions during emergency relief operations are often very different from normal conditions. Traditional IT management frameworks often have a strong focus on business aspects and neglect the special needs of organisations with an active role in crisis relief. Many of the problems discussed in this paper can also be found in regular IT management, but because IT dependability is especially critical during emergency situations and because the focus of IT management is mostly limited to normal conditions, the problems with IT dependability management in emergency conditions deserve special attention.

To help Swedish governmental actors with the dependability of their IT systems, SEMA published BITS, the Basic Level for IT Security handbook (BITS). BITS is meant to give Swedish authorities a practical overview of the main administrative measures that can be taken to achieve a minimum level of IT dependability. BITS is based on international standards such as ISO-IEC 17799 (International Organization for Standardization 2005), but BITS is much more suited for small public actors. BITS is also accompanied by BITS Plus, a web based planning tool
that can be used to coordinate the work with the BITS standard. The main disadvantage with using BITS for achieving a higher dependability is that it focuses mainly on security and a lot less on reliability and safety.

4 Research Methodology

The research in this paper combines results from two different sources: data collected from case studies at two Swedish municipalities and the data of a survey conducted by SEMA.

4.1 Case Studies at Two Swedish Municipalities

The main part of this research was conducted in two case studies at two different Swedish municipalities. These municipalities were selected because they had shown an interest in the topic of IT systems in emergency management in previous contacts with SEMA or with other members of our research project.

Municipality A is a large Swedish municipality consisting of a major Swedish city and the surrounding urban areas with close to 125,000 inhabitants and 7500 direct employees. Municipality B on the other hand is a small municipality consisting of two suburbs of a large Swedish city. Municipality A has 6 times more inhabitants, and also from a hazard perspective there are large differences. Municipality A houses a lot of industry and is an important national hub for the transport of dangerous goods. During the last years the municipality has gone through some major emergency situations of different types. Municipality B has a much lower risk profile and has not experienced any major emergency situations in the last 15 years.

To understand how these municipalities assess the dependability of their IT systems in emergency situations, a series of interviews were conducted with emergency managers and IT personnel at both municipalities. Further a number of documents concerning IT strategies, organisational structures and vulnerability analysis were also collected and studied.

For the analysis, all interviews were recorded and transcribed in full. During the transcription they were also translated from Swedish to English. Then, two authors went through all the transcribed text independently and coded (Robson 2002) all excerpts according to the following categories: division of responsibilities, internal communication, service level agreements, risk analysis and practical examples of problems. Afterwards their lists were merged and the excerpts in every category and subcategory were analysed. Since the interviews often returned to the same topic, and because different people in the same organisation were interviewed, triangulation was used to check the consistency of the interviewees’ answers. For the analysis both within and across the two municipalities the technique of explanation building (Yin 2003) was used.
4.2 Survey by SEMA

In May 2005, SEMA conducted a survey among 368 IT security managers at all Swedish municipalities, regional governments and different public authorities. A first analysis of the 230 answers to the survey they received was published shortly afterwards (Kalmelid and Gustavsson 2005).

The goal of the survey was to assess the capabilities of different governmental actors in the field of IT security. Within IT security the survey focused mostly on the methods and standards used and how SEMA’s support towards the governmental actors could be improved. The respondents were also asked to make an assessment of the maturity of their organisation and different members of their organisation in IT security.

For our study, the answers to the survey’s open questions were analysed in a similar way as the interviews in the case studies.

5 Findings

This section contains the main findings from the case studies and the survey. Each of the next sections discusses the conclusions that can be drawn from the excerpts that were coded in to the corresponding categories listed above.

5.1 Division of responsibilities

In both municipalities that participated in the study there is a central IT unit responsible for the maintenance of the IT systems. The final responsibility for most of the systems lies with specific departments that are the main users of the system. This responsibility means they decide about the acquisition, the updates and the evaluation of the systems. This division of responsibilities is logical, but also has a number of problems.

A first problem lies in the evaluation of the dependability of the systems. Since the IT department is responsible for the maintenance they are contacted in case of any problems, but it is not their responsibility to collect failure statistics, as expressed in Quote 1. The system responsible is often not even notified of all the problems, and can not get a full picture of the dependability of the system. In municipality A, the IT department has a help desk that coordinates the maintenance work of the IT department. In municipality B, users contact one of the employees of the IT department directly on their mobile phone, making it even harder to collect failure statistics. Further, concerning the service that is outsourced to external suppliers, some failures are reported directly to the supplier, while others are reported to the supplier through the IT department.
A second problem is that in this organisational structure the separate departments, and in particular the emergency managers, do not have any own technical personnel that can advice them on the technical details that are involved in the administration of the IT systems. This can lead to responsibilities implicitly being shifted to the IT department, just because the different departments do not immediately know how to deal with them. This is for example complained about in the survey as can be seen in Quote 2.

Quote 2
The IT personnel should get better at defining the limits of their area of responsibility to make sure that the responsibility is where it should be. This is necessary to avoid that the focus lies with the technology instead of the processes. We are not good enough at explaining to the different departments that there are some parts for which they must take responsibility. Today the IT department must always take responsibility for IT matters for which no-one else takes responsibility. This is not good. – Survey answer to the question: What do you think the IT personnel could get better at concerning IT dependability?

In municipality B, IT safety is a special responsibility of one of the emergency managers at the municipality. The advantage of this role is that he can lift these safety issues immediately to the highest levels in the municipality. In practice, a problem with this approach is that the IT department feels relieved of all safety responsibilities although their expertise is indispensable for evaluating this safety.

5.2 Internal Communication

An often recurring complaint, in the case studies and the survey, is a lack of real understanding between the IT department and the users, including emergency managers. Users complain that the IT personnel does not understand what they expect of their systems, as for example in Quote 3. The IT personnel on the other hand complains that the users do not understand the risks involved with IT systems, especially concerning security.
Quote 3
We have generators and we can provide backup power to our IT systems for a long time. Assuring the quality of our IT systems is more difficult. We have discussed this a lot, also with our IT technicians, but they often focus on the wrong things. – Emergency Manager, Municipality B

This lack of understanding is a consequence of the communication problems between both parties. Both municipalities under study lacked a forum where the IT department, the users and safety managers could discuss important strategic IT issues together. In the worst case the only communication occurs when a failure of a software system occurs and the IT department has to be notified to fix the problem.

Because of communication problems, many decisions about updates to the systems are made unilaterally and sometimes the other parties are not even notified in advance of the update. Of course, this also means that the risks of these changes cannot be analysed in detail, especially concerning specialised emergency scenarios unknown to the IT department.

Another common complaint about the communication between emergency managers and the IT department is that the communication from the IT department is too technical. Outside the IT department there is not enough technical knowledge to understand the technical details of the system, while the IT department does not manage to communicate their message without resorting to technical details. This adds to the frustration of both parties, and results in the IT department not being consulted as often as necessary for important decisions.

5.3 Service Level Agreements

Both municipalities in the study have some service level agreements, SLAs, with their external suppliers but have no service level agreements at all with their own IT department. Some written communication by which users and the IT department discuss the level of service, would offer clear advantages to both parties. For example in municipality B, the IT department tries to always have some IT personnel reachable to provide service, even in weekends and at night in case there is a need for urgent IT support for critical systems. This level of service is available because the IT department considers it reasonable, but is not explicitly specified anywhere as a guaranteed service. This illustrates again that the IT support is planned without consideration for emergency conditions.

The main advantage of SLAs for the users is that they know what to expect, and what not to expect, from their IT systems. This way they can avoid both depending on unreliable systems and investing in unnecessary backup solutions.
for sufficiently reliable systems. This problem is expressed in Quote 4 from a project manager at municipality A.

**Quote 4**

*If the IT department can explicitly state that they cannot give us any guarantees, we can justify investing some extra millions ourselves to secure our systems. But without any defined service levels, we have no arguments to justify this cost here.* — Project Manager, Municipality A

Even the service level agreements with external suppliers are often not well planned and not adapted to the level of quality actually demanded by the users of the systems. For example at municipality B, the maintenance contract with their supplier of routers guaranteed on-site service within 8 hours. This number was agreed upon many years ago, and nobody recalls exactly why it once was set at 8 hours. The importance of the internal network for the daily operations at the municipality has definitely increased drastically since this decision was taken. This example shows there are no routines in place to regularly re-evaluate important service level agreements.

Service level agreements are closely connected to measurements. The writing of service level agreements forces an organisation to think about how the quality of its IT systems can be measured. Just as both municipalities lack service level agreement for most of their systems, they also lack the possibility to measure the quality of their IT systems. Access to such measurements would give them a possibility to concentrate their resources better to improve the weakest links in their critical systems.

### 5.4 Risk Analysis

Although IT systems can play an important role in the aftermath of a crisis, they are seldom included in the emergency plans and risk analyses that are conducted. Emergency managers would like to include these systems, but they do not manage to do so because of problems in cooperating with the IT department. In municipality A, the emergency management of the social care department is planned in such a way that, if necessary, the department can function completely independent of IT systems. This means, for example, that all critical information is printed out on a very regular basis and communication plans are ready that do not rely on modern technology. As the project manager explained, this is a safe solution, since it means they are prepared for a complete failure of all IT systems, but it is also a serious overhead cost that is only necessary because they do not manage to analyse the risks of depending on their IT systems. If they would manage to include the IT systems in their risk analyses, they would be able to evaluate which
systems are reliable enough to depend upon in different emergency situations, and they could prioritise their resources by focussing on the least reliable and most critical systems. Because the IT systems are not part of the emergency plans, they can also not be used as efficiently in a crisis if they turn out to be reliable after all.

In municipality B, a crisis central was installed with the help of SEMA and a number of external consultants. Although IT systems are a critical component of the equipment there, the IT department was not involved in the development of this room. The IT department also maintains the systems in this room, but they are not involved in any strategic planning of how the systems in this room should be updated or replaced.

When the IT department is not involved in emergency planning, as expressed in Quote 5, they are also not aware of which systems are critical during different crisis situations and they can not correctly prioritise their maintenance work without receiving specific instructions during a crisis. IT systems are also seldom involved in emergency exercises. Useful lessons for emergency management could also be learned from exercises such as regularly trying to restore a system from backup, or measuring the behaviour of the network when one or more routers are disabled.

**Quote 5**

_We are not involved in making emergency plans. It’s not something we think about. And I don’t know what the rules are for prioritised service in an emergency. Nobody told me whether one computer is more important than another._ – IT TECHNICIAN, MUNICIPALITY B

### 5.5 Common Problems

A first major problem that was observed at both the municipalities was the problem with defining who is responsible for evaluating the dependability of the IT systems in crisis situations. This task requires the cooperation between the emergency managers, the IT department and the department owning the system. In practice, because of the communication problems discussed before, this can lead to this issue being overlooked. Especially if the IT department is not involved in the strategical discussions about the IT systems, they limit themselves to the daily maintenance of the systems and only perform technical long-term improvements when explicitly asked.

A second recurring problem is the lack of good supporting tools or standards. BITS (BITS), used by 75% of the municipalities that answered the survey, is more focused on security than reliability, and the focus is therefore more on the systems as separate units, and not on how the systems fit in to the overall activities of the municipality, as illustrated by Quote 6. For this reason, BITS is not ideal.
for a complete dependability analysis, and might even lead to some aspects being forgotten when it is not complemented with other risk analysis methods.

**Quote 6**
*The main disadvantage of BITS is that it uses an object-oriented model for IT dependability, instead of a process-oriented model. This means it considers IT systems as isolated objects, instead of starting from the information processes that are provided or supported by the system.* – SURVEY ANSWER TO THE QUESTION: **WHAT DO YOU THINK COULD BE IMPROVED ABOUT BITS?**

A third problem we observed lies in the emergency managers’ limited understanding of the dependability issues of IT systems. When they want to conduct a risk analysis of the IT systems they need this technical knowledge to be able to understand all the threats to the reliability of the system, their probability and possible consequences. Often it is assumed that the IT systems can be depended upon in a crisis, even if there is no evidence of their reliability.

Finally, a typical problem with IT systems is their fast evolution. New IT systems are installed every year and updates are done even more regularly. Adding new functionality to old systems changes both the reliability of the system and the dependence on the system. Often municipalities do not manage to keep their risk analyses updated to reflect the latest functionality of the IT systems. This is especially important since the dependence on the IT systems is increasing continuously. At first, after a new system has been installed, the system is usually only considered an extra asset that could be useful in a crisis situation, even if it not critically necessary because the old alternatives are still available. At this time the dependability of the system is not critical, but when the users get more used to having the new system around, the alternative systems are neglected and the new systems can become more and more critical. When these changes occur gradually, they are sometimes only noticed too late and systems can become critical without their dependability ever having been seriously evaluated.

### 6 Validity Discussion

A first threat to the validity in this study is the possibility of researcher bias. All the interviews were conducted by the same researchers and the conclusions from the first interviews were used to steer the later ones. To minimise the effect of researcher bias, the interviews were conducted with two researchers present and extra care was given to let the interviewees tell their own story, without guiding their answers. In the analysis of the interviews the possibility of researcher bias was constantly taken into account when building explanations.
A threat to the construct validity that is often present when data is collected through interviews is the possibility that the participants are focusing too much on their own side of the story and give a distorted view of reality. Through the use of triangulation, by interviewing different people at the same municipality and by asking different questions concerning the same topic, the effect of this can be reduced. Overall, our impression was that the interviewees were not afraid at all to talk about problems they were experiencing or had experienced in the past.

Further, when considering the external validity of this study, it is important to reflect on how far the results can be generalized. The case studies studied only two municipalities, but there is good reason to assume that the problems identified in this study are not unique to just these two Swedish municipalities. Because many of the conclusions were very similar for both municipalities, and because they are also supported by the survey which was answered by a majority of the Swedish governmental actors, we believe that it is likely that similar problems can be found in many Swedish governmental organisations with an active role in emergency management. More research is necessary to determine how factors like the size of the governmental actor influence the conclusions of this study.

Although emergency management in other countries is not always organised in the same way as in Sweden, many of the conclusions are general enough to be of importance in an international setting. The increased dependence on IT systems, together with a trend to centralise IT services are commonly found in many countries. The conclusion from this study are therefore likely to be of interest to similar organisations elsewhere that next to their everyday responsibilities also have an active role in crisis relief.

7 Proposed Process improvement model

This study shows the need for a method that can help governmental actors improve in how they manage the dependability of their IT systems. More precisely there is need for a process improvement model that is simple enough to be applied by small municipalities, and that is focussed on stimulating the cooperation between the IT personnel and the emergency managers.

One of the possible solutions is the development of a maturity model. A maturity model is a framework for process improvement that includes a number of maturity levels that can be used to evaluate and improve the capabilities of an organisation. Typically, an organisation first evaluates itself on a number of key process areas, and is then assigned a maturity level based upon this evaluation. In the next step, a number of practices necessary for reaching a higher level of maturity can be selected and goals for the next step in process improvement can be set up. This process can be repeated until a desired level of maturity is reached. The top level of maturity is usually a level where mechanisms for continuous improvement are in place.
Based on the experiences from this paper, it can be noted that a maturity model to help organisations measure their current maturity in dealing with these dependability challenges, needs to fulfil a number of special requirements for it to be useful to governmental actors with limited experience in this field. First, it should be possible to use with limited resources, also for small organisations. Secondly, it needs to contain a well-defined self-assessment tool to help organisations to identify and evaluate the processes they use to monitor the dependability of their IT systems. Further, the framework should offer specific, practical examples of good practices that can be used for improvement. Preferably it should be supported by tools for data collection and analysis. Finally, it is also important that the maturity model includes ideas from maturity models already currently in use in both IT management (Luftman 2003) and in safety culture (Fleming 2001).

8 Conclusions and Future Work

Through case studies and a survey this paper explores the main challenges involved in how governmental actors, and Swedish municipalities in particular, evaluate the dependability of their IT systems in possible crisis situations. The main contribution of this paper is, first of all, that it identifies the main problems areas in the current state of practice and, secondly, that it sketches the requirements for a process improvement framework that can help organisations improve in these areas.

This study shows that the core of the problem does not lie with either the IT systems themselves or the emergency management procedures, but the real problem is the lack of good cooperation to discuss these matters on a strategic level with all involved parties. Therefore, those responsibilities that lie on the border between different people’s areas of responsibility are often given too little attention. This leads to emergency planning that does not incorporate possible dependability problems with IT systems and IT management that does not take into account the special conditions that can occur during emergency situations. These kind of problems cannot be solved by simple measures and require a process improvement effort that involves a large part of the organisation. Because no current framework addresses these issues specifically, we propose the development of a maturity model similar to the process improvement frameworks already in use in some related fields.

Bibliography


Abstract

In many organisations a gap exists between IT management and emergency management. This paper illustrates how process improvement based on a maturity model can be used to help organisations to evaluate and improve the way they include IT dependability information in their emergency management. This paper presents the IDEM3 (IT Dependability in Emergency Management Maturity Model) process improvement framework which focuses on the cooperation between IT personnel, emergency managers and users, to proactively prevent IT dependability problems when the IT systems are most critical in emergency situations. This paper describes the details of the framework, how the framework was developed and its relation to other maturity models in related fields.
1 Introduction

In recent years governmental actors have come to depend more on IT systems for all their everyday tasks. For communication, they depend on landline telephone networks, mobile phone networks, web servers, email servers, etc. Other important systems are used for patient administration in health care and social care, school administration or city planning.

Just as for their everyday tasks, governmental actors now depend on all kinds of IT systems for their responsibilities in crisis situations (Santos et al. 2008). These systems include not only specially built systems for emergency situations but also the everyday systems described above. The latter category of systems is of special interest, because under normal conditions an occasional unavailability of these IT systems is fully acceptable, but in emergency situations, when time is a critical factor, any unexpected unavailability can have disastrous consequences (Fleming 2001, Zimmerman and Restrepo 2006).

Therefore it is important that these IT systems are an integral part of all major risk and vulnerability analyses conducted. This way information about the dependability of the different IT systems can be combined with information about how critical the systems are in different situations (BITS). IT dependability management for organisations with a critical role in emergency situations is a complex process of managing software in terms of IT systems. The occurrence of a number of critical IT incidents in the recent past shows that there is room for improvement. Earlier research (Weyns and Höst 2009) has shown that there is a particular need for improvements with respect to the communication between emergency managers and IT-management. This is a complex problem for which no quick solutions exist that fit all organisations. Instead, organisational improvements in this area must be based on the organisation’s current situation and its goals for the future, that is through a process improvement approach.

This paper presents a maturity model for the coordination of emergency management and IT dependability management. The main focus of the framework is on the cooperation between emergency managers and IT personnel. The purpose of this maturity model is to help organisations to identify, evaluate and improve their IT dependability processes.

2 Background

The maturity model presented in this paper is based on the result of a series of case studies on how governmental organisations deal with IT dependability issues in emergency management (Zimmerman and Restrepo 2006, Weyns and Höst 2009). The main conclusion from these studies was that many organisations today experience problems and frustrations concerning IT dependability in emergency management. The main cause of many of these problems could be traced back to
communication and cooperation problems between the personnel in different roles involved. Further these studies also pointed out a lack of useful tools that support IT dependability improvements across a whole organisation. This maturity model is meant to offer a process improvement model that is simple and general enough to be applicable to many organisations and at the same time effective enough to make a substantial difference in an organisation’s IT dependability practices.

3 Related Work

In the field of IT management a number of international standards and best practice frameworks have been published, among those ITIL (Office of Government Commerce 2007), COBIT (ISACA 2000) and ISO/IEC 27002 (International Organization for Standardization 2005). These frameworks are more suited to be used by large corporations with very large IT resources and are less suited for smaller organisations and often do not take into account the special requirements for organisations with an operative role in crisis relief. Of these frameworks, COBIT is structured as a maturity model. Frühwirth (2009) has discussed the mismatch of software dependability management and industry standards today.

The maturity model presented in this paper is based on a number of maturity models from related fields. The first successful maturity models were developed by the Carnegie Mellon Software Engineering Institute (Konrad et al. 1996). Since the development of the Capability Maturity Model, maturity models have been applied in many other fields. The problems between emergency management and IT management are related to some of the problems in software requirements management and therefore the process improvement methods that have been successfully applied in software engineering can also benefit IT management.

In 2008, SEI published a preliminary version of the CERT Resiliency Engineering Framework (Caralli 2007) for the use in the field of business continuity management with a special focus on IT systems. In the field of IT management, Luftman (2003) presents a simplified maturity model with a strong focus on the business value of IT systems. In the field of safety management, maturity models have also been proposed as a way of assessing an organisation’s safety culture (Fleming 2001), or product design safety (Strutt et al. 2006). Section 8 focuses especially on how each of these maturity models relate to the maturity model presented in this paper. The maturity model presented in this paper does not try to replace any of these maturity models or to cover any of these related fields completely. From each related field, this maturity model contains only those attributes that are specifically important for the dependability of IT systems in emergency management.

Recently, Santos et al. (2008) have published a maturity model for the use of information technologies in emergency response organisations. Their model does not cover the dependability of the IT systems in emergency situations, but instead
focuses on information management practices. The IDEM3 maturity model described in this paper is most suited for an organisation where the IT services are provided by an IT department that is part of the organisation. For evaluating the resiliency of IT services provided by external suppliers, Bhamidipaty et al. (2007) have developed the Resiliency Maturity Index, a framework for characterizing and evaluating the resiliency of an IT services organization. However this model does not evaluate the relationship between the resiliency of the service supplier and the dependability requirements of the organisation.

4 Methodology

To support organisations that want to evaluate and improve their IT dependability practice, this paper presents the IDEM3 (IT Dependability in Emergency Management Maturity Model) process improvement framework.

The research that resulted in the IDEM3 maturity model was conducted in a number of steps: the identification of the attributes, followed by mapping the different levels of each of the attributes to the five levels of the maturity model, then an off-line validation and currently the maturity model is being evaluated in a practical setting. This process is presented in Figure 1.

First, the case studies (Weyns and Höst 2009) that describe the need for this kind of maturity model also resulted in a list of factors that are important for the coordination of IT dependability management and emergency management. These key factors formed the first basis for the attributes of the maturity model.

Secondly, the factors were mapped to the general architecture of a maturity model with five levels as found in other maturity models such as CMMI (CMMI Product Team 2006). For the model to be applicable by small organisations, it was necessary to simplify the structure by replacing the concept of 'key process areas' by the more modest 'attributes' found in the model.
In this mapping the attributes were also compared and complemented with similar attributes found in maturity models from related fields, as described in more detail in Section 8. Before the model was applied in a practical setting, the model was validated with the help of experts and practitioners in the field. Finally, the model is currently being evaluated through the application of the model in a series of large case studies. The validation and the first results from the evaluation are further discussed in Section 9.

5 Process Improvement with IDEM3

This section shortly explains how IDEM3 can be used as part of an organised process improvement effort. First, the model can be used to assess the current maturity of the organisation in dealing with IT dependability in emergency management. For this assessment the current practices in the organisation should be matched with the attributes described in Section 6. The recommended way to do this is to select some of the most critical systems, preferably systems that are quite different in nature and together are representative for the critical IT systems in the organisation. For each of these systems, personnel with different roles should be interviewed individually based on a detailed questionnaire, where they are each asked to describe how they are currently experiencing each of the attributes of the maturity model. The involvement of personnel from different parts of the organisation is an essential part of this maturity model to make sure that IT dependability is not only evaluated from a technical point of view. The interviews should at least include users of each of the systems, system managers, safety managers in the domain where the systems are used and of course IT personnel.

The responses from all these interviews should be analysed in detail by the process manager overseeing the assessment with special attention for differences between the answers of different respondents and between the different systems. The analysis of the interviews should then be the basis for a focus group meeting where the organisation can be assessed on the maturity scale for each of the different attributes presented in the maturity grid shown in Table 1. The 22 attributes are ordered in such a way that attributes are most strongly correlated to those attributes just above and below. Therefore the maturity of an organisation in these 22 attributes can be presented in a spider web diagram, offering a clear representation of the organisation’s strengths and weaknesses.

Finally, after this assessment, the organisation can decide whether the measured level of maturity is sufficient for the organisation. Not all organisations need to aim for the highest maturity level, mostly depending on how critical the role of the organisation is. The process improvement needed to reach a higher level of maturity is a long term project and should be organised as such. This means that a realistic time plan with explicit long and short term goals should be agreed upon. For the long term planning, it is important to realise that after each step
from one maturity level to the next, some time is needed to make sure all proce-
dures are well incorporated in the organisation and to make sure improvements
are not too easily lost again. An improvement of more than one maturity level per
year is probably unrealistic. Organisations should not try to skip certain levels or
to implement a new level too quickly after the previous one since each level builds
on the achievements of the previous level being well understood. For the short
term planning, the organisation can focus most of all on those attribute for which
they received the lowest maturity score. The organisation as a whole should fo-
cus on achieving a stable IT dependability management at this new maturity level.
The actual improvements can be implemented with the help of those project man-
agement mechanisms that are most suited for the organisation in question. While
implementing these planned improvements, it is important that regular critical self-
assessments are held to evaluate the organisation’s progress and to make sure the
selected improvements are correctly implemented and are not easily lost again.
A single assessment based on the IDEM3 maturity model can also be done sepa-
rately from any planned process improvement based on the 5 maturity levels. An
organisation can conduct a one-time assessment of its IT dependability based on
this maturity model to identify its current strengths and weaknesses in this field.
The results of this assessment will then form an excellent basis for a discussion
on how to involve all stakeholders to improve the organisation’s IT dependabil-
ity in emergency management. Unlike with some other maturity models, IDEM3 as-
essment is not meant to be used as a basis for certification or for direct, objective
comparison between different organisations.

6 Maturity Levels

Just like most other maturity models discussed in Section 3, the IDEM3 model has
five maturity levels. The levels have similar names as in these other maturity mod-
els, and the basic idea behind each of the levels are also comparable. The Initial
level is the most basic level, representing an organisation where some critical IT
systems have not been analysed from a dependability point of view and nobody
takes responsibility for initiating a more strategic discussion about IT dependabil-
ity. The second, Managed, level is characterised by an organisation where the
dependability of all critical IT systems is managed on a system-by-system basis
leaving the organisation very dependent on the competence of the system man-
grers for every system.

The third level is referred to as the Established level. This means that the or-
organisation has established a centrally coordinated approach for dealing with IT
dependability. This will usually be established by appointing one central IT de-
pendability manager who distributes standard procedures for dependability anal-
ysis to all system managers. A standardised approach is a prerequisite for being
able to implement future improvements across the whole organisation. A level 3
organisation also has clearly defined roles and responsibilities concerning IT dependability. The fourth level, called Quantitatively Managed, is similar to level 3, but also requires that the centrally coordinated approach is supported by extensive quantitative data collection. Regular measurements and testing with special usage scenarios in mind can make IT dependability statistically predictable and allow for strategic improvements in IT dependability. The Continually Improving level, which is the fifth and final level of the maturity model, is reached by an organisation that can use the feedback obtained from the practices from level 4 to continually improve not only their IT systems, but also their own IT management procedures. IT systems will then be naturally included in risk and vulnerability analyses and their dependability will be regularly re-evaluated.

To define the levels in more detail the levels can be compared on 22 attributes. Of course all these attributes are in some way related and none of them can be changed completely independent of the others. Nevertheless they each add their own focus to the maturity model and stress a special aspect of an organisation’s maturity.

The 22 attributes can be divided in 4 categories: Outcomes, IT management, Cooperation and Organisational Issues. A detailed summary of these attributes can be found in Table 1 and the attributes in each category are also described in the following subsections. The attributes are ordered in such a way that those attributes that are most strongly related are placed next to each other.

### 6.1 Outcomes

The first category of attributes is different from the three other categories in that it contains those attributes that can not directly be influenced by an organisation, but only indirectly. These attributes should mainly be considered as the consequence of an organisation’s maturity, while the other categories are the causes of the maturity level. At the same time the outcome attributes are also the most important because the main goal of this maturity model is improving the outcomes of the IT dependability. This category also contains those attributes that are the most visible to stakeholders outside of the organisation.

The outcomes category contains 9 attributes: Actions taken, Problems that can be identified, Basis for improvements, Nature of improvements, Successes, Success factor, Role of IT in emergency situations, Attitude towards dependability problems and IT dependability.

These 9 attributes together describe the dependability experienced by an organisation at each maturity level, and how the organisation deals with these results.

A level 1 organisation will typically experience many problems with IT dependability and will focus most of its effort on trying to fix the problems as they appear. Because of the lack of an organized approach, some problems will not get solved and implemented changes can cause problems for other parts of the organisation. This will lead to a lot of frustrations, and although many of the minor
# Table 1: Overview of the 22 attributes of the maturity model across the 5 maturity levels

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level 1: Initial</th>
<th>Level 2: Managed</th>
<th>Level 3: Established</th>
<th>Level 4: Quantitatively Managed</th>
<th>Level 5: Continually improving</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. General maturity model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. IT dependability model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Dependability requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Service levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. IT dependability analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. IT incident management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Relationship IT personnel - emergency managers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Presence of IT dependability in emergency plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Involvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Responsibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Management mechanisms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Organisational learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Level 1: Initial
- Level 2: Managed
- Level 3: Established
- Level 4: Quantitatively Managed
- Level 5: Continually improving

**Attributes:**
- General maturity model
- IT dependability model
- Dependability requirements
- Service levels
- IT dependability analysis
- IT incident management
- Relationship IT personnel - emergency managers
- Presence of IT dependability in emergency plans
- Involvement
- Responsibility
- Management mechanisms
- Organisational learning
- Outcomes
problems will not come as a surprise, a serious failure in a critical system during an emergency situation can still do a lot of damage.

An organisation at level 2 will employ a system-by-system approach towards IT dependability allowing it to respond effectively to most of the problems and even to prevent some problems that only affect one system. Lessons learned from problems they experience will often lead to improvements in the affected system only as there is no centralised approach to IT dependability. This method of working leads to a higher dependability than in level 1, but places a large amount of responsibility on each system manager and much will depend on his skill and experience in dealing with the risks of IT dependability problems.

An organisation at level 3, on the other hand, uses a basic centralised approach towards IT dependability. The same basic techniques for risk and vulnerability analysis are applied to all systems and many dependability problems can systematically be prevented. Because of the coordination between different systems, also problems with interdependencies can be detected and dealt with. The main success factor from level 3 on will be the quality of the centrally coordinated dependability measures being used across the whole organisation. This will also make it easier for the organisation to efficiently share important resources such as backup facilities and emergency power supplies between all critical systems.

A level 4 organisation will supplement the basic centralised approach from level 3 with a large-scale systematic data collection and analysis concerning IT dependability. This will make IT dependability more predictable. The data collection will make it possible to measure improvements and their effects and to prioritise the usage of IT dependability resources. A level 4 organisation will also have an improved cooperation between all involved stakeholders which is an important factor for the IT dependability.

Finally a level 5 organisation will continuously work on evaluating and improving its IT dependability. The safety culture in an organisation at level 5 will even make it possible to regularly identify possibilities for improvement in their risk analysis procedures. At level 5, IT dependability is generally working very well and the level of success that can be achieved depends mostly on whether a continuous improvement effort can be sustained throughout the organisation. This makes that IT systems will not only be a source of risks or problems in emergency situations but also a valuable asset that can be depended upon.

6.2 IT Management

The second category of attributes collects those attributes that are directly related to IT management. Unlike some other maturity models, this maturity model does not seek to cover the complete field of IT management, but focuses exclusively on those aspects that are most important for IT dependability in emergency management. This category contains the following 4 attributes: Results of IT incident management, IT incident management, IT dependability management and De-
A level 1 organisation lacks organised IT incident management, and the dependability requirements of most systems will typically never have been analysed. At level 2 incident management is handled for each system separately and for many systems there will be no explicit link to risk analysis or emergency management. A level 3 organisation is expected to have a centralised IT incident management system allowing information sharing between different parts of the organisation. Further centralised guidelines for IT dependability management will require the main dependability requirements for each system to be explicitly documented and available to all stakeholders. From level 4, IT incidents can be analysed in detail and can lead not only to direct improvement in all systems but also to improvements of the procedures used for IT dependability and even lead to improvement in the safety culture of the organisation at level 5. At the two highest levels of maturity dependability requirements for all systems should contain detailed measurable values and these requirements should be updated in the case of changes in the systems’ functionality or usage.

6.3 Cooperation

A third set of attributes concerns the cooperation between the different parties involved in IT dependability. This is in the first place IT personnel, system managers, the system’s users and also the personnel responsible for conducting risk and vulnerability analyses, for example emergency managers. The 4 attributes in the category are: Service level agreements, IT dependability analysis and emergency planning, Presence of IT dependability in emergency plans and Relationship IT personnel - emergency managers. A level 1 organisation will typically lack service level agreements or any other documents clearly linking IT dependability and emergency management. The frustrations and conflicts between different parts of the organisation will hinder a necessary cooperation on these important issues. In a level 2 organisation some of these issues will be taken care of for some systems, while there will be many problems with other systems, mostly depending on whether there are good contacts between the system manager of each system and the IT department. A level 3 organisation is expected to have basic, standardised service level agreements in place for all systems. Further, dependability estimates for all systems will be used as input for emergency management and the requirements discovered while making emergency plans will be used as input in the prioritising the IT dependability activities. From level 4 an organisation’s SLA’s should contain clear, quantitative dependability goals and measurements. The link between dependability requirements and risk and vulnerability analyses for all systems should be explicitly documented. By clearly defining the responsibilities of all parties in detail, all successes will be shared success and when problems should arise the blame cannot just be shifted around as is often the case on the lower levels of maturity. Finally, in a level 5 organisations there is a real
partnership between the different departments cooperating on IT dependability and continuously striving to improve their cooperation.

6.4 Organisational Issues

A last category of attributes collects those issues that concern the whole organisation and how it is managed. There are 5 attributes in this category: Involvement, Responsibility, Management Mechanisms, Organisational learning and Resource allocation.

In a level 1 organisation, in the worst case, nobody is actively involved with IT dependabilities and most stakeholders will feel the responsibility lies with someone else. After an incident, often the blame is shifted around and no learning takes place. In a level 2 organisation, the responsibility for IT dependability lies explicitly with the individual system managers who deal with the issue as they see fit. Therefore learning about IT dependability will mostly happen on an individual basis and improvements will depend on whether the system manager can find the resources to invest in IT dependability for each system. In a level 3 organisation, all the responsibility lies in the first place with central IT safety manager who is responsible for the coordination of IT dependability procedures. The IT safety manager distributes detailed dependability instructions and directions that are meant to be followed strictly by all stakeholders. This coordination allows the organisation to learn as a whole from past failures and successes. In a level 4 organisation, the detailed service level agreements for each system will make it possible for the responsibility to be shared by all actors in the IT dependability process. Through the detailed feedback from the collected data in a organisation at level 4, the organisation can achieve organisational learning by adapting its centralised procedures and guidelines based on measured outcomes. System managers are expected to be experienced enough to be able to apply the centralised guidelines and tools to manage IT dependability without detailed instructions. In level 5 organisations, not only the dependability guidelines are regularly updated, but also the way the organisation learns is continuously re-evaluated. This is called double-loop learning. In a well functioning level 5 organisation everyone will be aware of their own part of the responsibility for IT dependability and resources for improvements in IT dependability can be distributed in a prioritised way.

7 Transition from One Level to the Next

To further clarify the different levels of the maturity model, this section explains the main elements of the transition process from one level to the next. Although not every organisation will be at level 1 initially, and not every organisation will aim for level 5, the levels are meant to be taken successively without skipping over any level. A transition from level 2 to level 4 can only be achieved by first implementing level 3.
7.1 From Level 1 to Level 2

There are no requirements for the first level of maturity, and at this level it is common that there are some critical IT systems for which there is no control over the dependability. For an organisation to rise to level 2 the responsibilities for each system need to be well defined. Usually this will mean that the coordination for all dependability issues is done by the system manager for each system who organises the work with dependability in the way that suits each particular situation best. The main advantage with this approach is the clear definition of responsibilities which makes that the main problems can be discovered and solved. The main disadvantage is that it is nearly impossible for the organisation to evaluate the quality of the dependability analyses done by the system managers since they each use their own methods.

7.2 From Level 2 to Level 3

To go from level 2 to 3, an organisation needs to standardize the way all system managers deal with IT dependability. First an organisational standard needs to be defined and then all system managers need to be instructed in this standard. The standard can be compiled based on national or international standards or on some of the procedures that were already previously used for some IT systems with good results.

7.3 From Level 3 to Level 4

While level 3 is mostly concerned with qualitative data about the dependability of IT systems, level 4 also requires the use of substantial amounts of quantitative dependability goals and measurements. A level 3 organisation might for example classify the availability requirements of a system according to a simple scale, Low-Medium-High, but a level 4 organisation is expected to use more detailed, numeric values. Setting up a central system to collect all service level agreements and to facilitate the analysis of all this data is a requirement for the transition from level 3 to 4.

7.4 From Level 4 to Level 5

Level 5 is characterised by a continuous effort to improve the processes in the organisation. This is only possible if the processes are well understood throughout the whole organisation and even across the borders of the organisations to include suppliers and network operators. To go from level 4 to 5 all procedures from level 4 need to become completely institutionalised throughout the organisation and all stakeholders need to be working together in a natural way. This way the data collected can form the basis for deeper, double-loop learning for the organisation.
This means the lessons learned are not only used to improve the organisation’s dependability practices but also to optimise the improvement process itself.

### 7.5 Commitment Required

It should be clear that there is a large difference between the commitment and resources required of an organisation to reach each level of dependability. Level 1 represents the lowest commitment to IT dependability. Becoming a level 2 organisation only requires a serious commitment from the individual system managers who need to drive IT dependability forward and need to collect input from all other personnel involved. Reaching level 3 maturity requires a regular commitment from all personnel involved with IT dependability to maintain a basic level of IT dependability across the whole organisation. Level 4 is very similar, but requires a larger effort for data collection and analysis. Reaching and sustaining level 5 maturity definitely requires the largest overall commitment to IT dependability, although in practice all efforts for IT dependability should feel more as a natural part of the daily workings of the organisation than as a special effort for IT dependability.

### 8 Relation to Other Maturity Models

As mentioned before, the maturity model presented in this paper is based on a number of maturity models from related fields. Table 2 illustrates how the attributes in the IDEM3 model correspond to similar concepts in these maturity models. For most attributes, similar maturity levels as in IDEM3 can also be found in one or more of these maturity models. The compatibility of the IDEM3 model with each of these models not only makes it easier to combine the usage of this model with the other models, it also increases the validity of each of the attributes and therefore of the whole model. IDEM3 does not in any way try to be an alternative for any of the models presented below, but has a different, very specific focus that is not explicitly present in any of the other models.

Of course, not all attributes can be matched with corresponding areas in all other maturity models. This can be for a number of different reasons. First of all, each of the maturity models referred to here has its own scope, which only partly overlaps with the scope of this maturity model. Therefore there are, for example, no attributes concerning IT management in maturity models from the area of design safety. When an attribute is clearly outside the scope of a certain maturity model, this is marked in Table 2 as n.a., not applicable. Secondly, there are some attributes that are not explicitly mentioned in certain maturity models, for example, organisational learning in all but one of the models. Such attributes are nevertheless generally compatible with these models, they were just not selected as process areas or explicitly used in the description of the different maturity levels. This is marked in Table 2 with a minus sign (-).
Table 2: The primary influence of the maturity model on other maturity models (Lutman 2003, ISACA 2000, Crammel 2007)

<table>
<thead>
<tr>
<th>Organizational Issues</th>
<th>Cooperation</th>
<th>IT management</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Strategic alignment</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. Mature governance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3. Business continuous improvement</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4. Maturity in the IT function</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5. Integration of IT into the business</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6. IT infrastructure that supports business processes</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7. IT services that support business processes</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: The table above shows the primary influence of the maturity model on other maturity models. The influences are represented by the absence or presence of a checkmark (✓) or dash (-) in the table cells.
9 Evaluation of the Maturity Model

IDEM3 is the result of a long development process during which many of the details of the model have regularly been updated. The model has been evaluated and validated in a number of ways. First of all, the case studies (Weyns and Höst 2009) provide an empirical grounding (Ågerfalk 2004) and the relationships with well-established maturity models are a strong external theoretical grounding (Ågerfalk 2004) of the maturity model.

For a further external validation, IDEM3 has, at a number of different occasions, been presented in detail to researchers and practitioners with long experience in the field, such as representatives of the Swedish Civil Contingencies Agency. At each of these presentations the model has received a positive reception, and many practitioners, both from the field of IT dependability and emergency management have expressed an interest in putting the ideas of this model into practice. Their comments and recommendations, both on the form and the details of this model, have all been taken into account in the version presented in this paper.

Further, the model is currently being used to assess certain aspects of IT dependability at two Swedish hospitals and to formulate improvement suggestions. First results of this assessment and the improvements suggested by the model were very positively evaluated by the participating organisations. These four rounds of evaluations give us confidence that the model in its current form can be an effective tool in improving an organisation’s IT dependability in emergency management. The final validation of this model, in the form of a large-scale implementation of this model at a number of organisations, is currently taking place. The practical evaluation of a complete maturity model is in no way an easy task, and proving that the model leads to an efficient improvement in an organisation’s IT dependability requires a huge research effort.

10 Conclusions

This paper has shown that process improvement based on a maturity model can help organisations close the critical gap between IT dependability management and emergency management. The IDEM3 maturity model contains 22 attributes in four categories: Outcomes, IT Management, Cooperation and Organisational Issues. The model is based upon a number of established maturity models from related fields and upon a number of problems identified in an earlier case study.

The maturity model is not a quick fix that will solve all of an organisation’s IT dependability problems. The main value of the maturity model is that it offers a way for an organisation to quickly capture its strengths and weaknesses in how it combines IT management and emergency management. IDEM3 can help an organisation to involve all stakeholders in this process improvement effort and to
visualise its progress. The model has been evaluated and improved based on feedback from experts and professionals in the field, and is currently being evaluated by case studies in the field of application.

**Bibliography**


Abstract

Maturity models are widely used in process improvement. The users of a maturity model should be confident that the weak points of the assessed processes can be found, and that the most valuable changes are introduced. Therefore, the evaluation of maturity models is an important activity. In this paper, a mapping study of the literature on the evaluation of maturity models is presented. Two databases are searched resulting in a set of relevant papers. The identified papers can be classified according to six categories, namely the maturity model under evaluation, type of evaluation, relation of the evaluators/authors to the maturity model, level of objectivity, main purpose of the paper, and size of study. Further, a framework of different evaluations of maturity models is developed, and the relevant papers are mapped to the framework. Finally, the relevant research on the evaluation of the maturity models in the CMM-family is discussed in more detail. The result of this mapping study is a clear overview of how the evaluation of maturity models has been done, and some discussions are provided for further research on the evaluation of commonly used or newly developed maturity models.
1 Introduction

In order to obtain improvements in the software development process and in the resulting product, the software process can be changed with the help of a software process improvement program. When deciding which improvements to implement, there are a number of different changes that are possible and it is in many cases not clear exactly how to change the process in the best possible way. This is one reason why maturity models are used during process improvement. This kind of model is often used as a guide in the improvement work in order to identify which improvements to introduce in the process and at which time. In practice, this is carried out in a number of phases. First the process that should be improved is assessed based on a maturity model. Then the result of the assessment is used to identify which improvements that should be introduced to the process to increase the maturity level of the process.

Some maturity models have been used to a large extent in process improvement in many software development organizations. Examples of this type of maturity models are CMM, CMMI, and the ISO standard based on the SPICE project (e.g. (Jung and Hunter 2001)). These models each have their own advantages and disadvantages, for example as described by Wang et al. (1997). To some extent, these models have been accepted as an international software process assessment state of practice.

In addition to the commonly used models, the basic methodology of using maturity models in process improvement is also useful in other domains, such as in human resources where People CMM has been used. Maturity models are also used in the area of safety culture assessments (Fleming 2001) and IT management (Luftman 2003). Further on, maturity models are also defined for special purposes in cases where new models or adaptations are seen as more suitable than the traditional models. For example, in (Hosny 2004), a set of key practices from CMM is adapted to small organizations. Formulating maturity models is often seen as a useful way of communicating best practices, since it packages the knowledge in a way that makes it useful in improvement initiatives.

Since maturity models are commonly used in process improvement, it is important that they are effective in the sense that they identify the right improvement proposals. The users of the maturity models should be confident that the changes that are introduced really result in improvement and that there are no other changes that would result in significantly more value. This is the basic idea of maturity models, i.e. to outline a path to improvement based on well-proven and accepted improvement steps. If it is not clear that the models are effective and that they point at the right improvements, potential users will hesitate to use them.

An empirical approach to showing that a model or method is effective is to make an evaluation of the model and, based on that, draw conclusions on its effectiveness, and maybe also compare it to other models. Especially in cases when a new maturity model is developed, or a standard model is adapted, it is important
to be able to show that it guides the user to the right improvements. However, it requires significant effort to evaluate this kind of models. If it is required that they are used in a large enough set of improvement initiatives in order to be able to draw significant results about effectiveness, this probably requires much time and effort. Instead, evaluations have also been conducted by other means. There are a number of ways that have been used in the literature, e.g. by investigating the differences between different assessors (Fusaro et al. 1997), or by analyzing the effect on one organization in a case study (Diaz and Sligo 1997).

It is by no means easy for a researcher to decide how to evaluate a newly developed or existing model, or even to understand which evaluation approaches are available. This means that there is a need for a framework, describing the different kinds of evaluations that can be carried out for maturity models. In this paper such a framework is presented together with a mapping study (Kitchenham et al. 2011) where the available literature on maturity model evaluation is mapped to the framework. That is, this paper summarizes what methods have been applied in the literature in order to evaluate maturity models. The result of this review is of course relevant to researchers who are planning to conduct this kind of evaluations. The result is also relevant to practitioners in the area of SPI since it provides an overview of existing evaluation methods, which is useful e.g. when interpreting the results of presented evaluations of existing models. For example, if a model is presented, and it is claimed that it is evaluated based on a set of studies, this study will make it easier to understand the completeness of the evaluations and to identify possible evaluation studies that also could have been conducted.

This means that the main contribution of this paper is not to say which model is the most effective but to summarize the evaluation methods that have been used. It would also be of interest to investigate the effectiveness of the different maturity models, but for most maturity models the data that is available in the selected papers is insufficient for this purpose. In terms of meta-analysis it was only possible to more qualitatively summarize the collection of evaluations of the most widely applied maturity model (CMM) as part of this research. This is discussed in Section 6.

2 Related work

Even if there are no standard evaluation frameworks for evaluating tools, models or frameworks in the field of software engineering, some researchers have done valuable work in this area already.

One such evaluation framework was developed by Kitchenham et al. (2005b) and Kitchenham et al. (2005a). They constructed a risk-based software bidding model and proposed an evaluation framework to evaluate their bidding model. The framework worked well as a validation framework, and first trials confirmed the benefit of the framework, and allowed for some adjustment to the framework
PAPER V: A REVIEW OF METHODS FOR EVALUATION OF MATURITY MODELS FOR PROCESS IMPROVEMENT

according to the evaluation results. The evaluation process was confirmed to be important and the evaluation framework was suggested to be applicable to a broad range of models. This is related to the evaluation of maturity models, for example in that the evaluation task is complicated by the fact that it is very hard to judge whether decisions taken were right or other decisions should have been taken.

Another framework for evaluating tools and models in information system research has been proposed by Ågerfalk (2004). He makes a distinction between internal grounding, external theoretical grounding and empirical grounding. These three categories are closely related to the three types of evaluation proposed in this paper. The types of grounding described by Ågerfalk (2004) are not limited to maturity models or even to process improvement frameworks in general, but can be applied to a wide range of methods, while the types of evaluation presented in this paper are more specific for maturity models.

Pfleeger et al. (1994) reported on the results of the Smartie project (Standards and Methods Assessment Using Rigorous Techniques in Industrial Environments), where they present a framework for evaluating software engineering standards. They discuss the difficulties in evaluating large process improvement frameworks and focus especially on the questions that should be asked during such an evaluation. However, no framework for maturity model evaluation was presented.

Finally, Vaishnavi and Fraser (1998) have published a validation framework for maturity models, which they have used to validate their Formal Specification Strategies Maturity Model. Their framework focuses especially on the requirements for empirical validation of maturity models through application in case studies.

Systematic reviews and mapping studies have been conducted in different studies (Kitchenham et al. 2009) in widely different areas such as cost estimation (e.g. Jørgensen 2004), open source software (e.g. Stol and Babar 2009), and testing (e.g. Engström and Runeson 2011). However, to the best of our knowledge, none of them has looked specifically at the evaluation of maturity models in general. Nevertheless, for the most well-known maturity models, there have been meta-evaluations that summarize selections of previously published results about the evaluation of these specific maturity models. Often the scope of each of the individual evaluations is quite limited because of the large effort involved in evaluating a maturity model, therefore these meta-evaluations actually provide the most complete evaluation of the maturity models. Of course, this approach is only possible for well-established maturity models.

The most famous of these maturity models is probably the Capability Maturity Model Integration (CMMI) (CMMI Product Team 2006). CMMI is a process improvement approach that provides organizations with the essential elements of effective processes. This framework, together with its predecessors and related frameworks, has to a large extent influenced the software industry and the research in software process improvement. McGarry et al. (1998) have published a large study with the result of CMM in over 90 projects. They also explicitly make a
distinction between two kinds of evaluations: through direct measurements of an organization’s results and a more subjective evaluation through surveys among organizations using the framework. Galin and Avrahami (2006) have summarized the results of 19 previous studies on the effectiveness of CMM programs. Both these papers explicitly state the need for more empirical, measurable evaluations of process improvement frameworks. Because of the importance of CMM and because of the large amount of publications available about CMM, Section 6 of this paper discusses the evaluations of CMM identified in this mapping study in more detail.

The SPICE (Software Process Improvement and Capability dEtermination) framework (Emam et al. 1998) aims to deliver the ISO standard of software process assessment. After the first and second versions of SPICE were published, large international SPICE trials were organized to evaluate different aspects of the framework. Jung et al. (2001) and Emam and Goldenson (1995) have published reports summarizing the goals and the results of these evaluations.

From these studies it is clear that there are many different ways to evaluate a maturity model and that there is little consistency in which evaluation methods are currently being used. This shows that there is a need for a broader summary of approaches for evaluating maturity models and classifying how models have been evaluated in the past.

3 Evaluation framework

It is important to distinguish between ‘assessment’ and ‘evaluation’ in the context of maturity models. Maturity models are used to assess and improve development processes and other processes. Also the maturity models themselves can be evaluated and improved and this is the focus of this mapping study. It can be noticed that both assessments and evaluations involve a set of questions that are asked by an investigator. In the assessment the objective is to understand and improve the process and in the evaluation the objective is to understand and improve the maturity model. An evaluation can be conducted independently of assessments or based on the results of assessments.

There are different possible ways to evaluate a maturity model. To classify the different ways of evaluating a maturity model, this paper proposes the following framework consisting of three types of evaluations, based on the experience of the authors in this field:

- Type 1: A type 1 evaluation is conducted “off-line”, only by the authors of the evaluation without involving any outside experts. A type 1 evaluation can be done based on, for example, their knowledge of the processes it is intended to be used with or by comparing it with other similar frameworks. The evaluators could also be the authors of the maturity model itself.
• Type 2: A type 2 evaluation is conducted by involving practitioners, who are the experts on the type of process that is intended to be improved by the maturity model, but who have not been involved in the actual development of the maturity model. In a type 2 evaluation no real assessment is carried out, instead interviews, surveys or simulated assignments can be carried out.

• Type 3: A type 3 evaluation is conducted through real process improvement activities where the maturity model is used in a practical setting.

Type 1 evaluations requires the least effort of the three since they only require that people who have knowledge about it can spend time elaborating different aspects of the model. Type 2 evaluations are more complicated since they require the cooperation of practitioners. Type 3 evaluations can be regarded as the most costly since they require that an assessment is carried out and that the result is used not only to analyze the investigated process but also to analyze the maturity model. That is, if the assessment is carried out in order to evaluate the maturity model, the cost of this evaluation type can be regarded as very high. However, assessments can be carried out with the purpose of at the same time improving the process, and then the expected benefits can be used to justify this cost.

The three types of evaluations imply an order in which they can most logically be used to evaluate a maturity model under development. First the maturity model is evaluated by experts on the maturity model, then by experts on the process that is improved, and finally it is used in improvement programs. For example, a maturity model may be formulated for improvement of an IT management process. It could first be evaluated in a type 1 evaluation where experts on the model evaluate factors like understandability and internal consistency. After that, independent practitioners from both IT management and users of IT systems could evaluate the contents of the model in a type 2 evaluation to validate how well the model correspond with the current state of practice. When these evaluations have been carried out it is reasonable to use the model in a series of assessments where extra effort is spent on evaluation of the developed maturity model, i.e. type 3.

It is of course also possible to iterate or to carry out the evaluations in any order that is seen as useful. In the example above, it can, for example, be necessary update and reevaluate the model based on the feedback from the earlier evaluations.

4 Research methodology

The research in this paper is conducted in two major steps. First a framework for the evaluation of maturity models is defined based on knowledge of the area and commonly cited literature. The defined framework is then used to classify papers in a mapping study of published literature in the area. The mapping study representing the major part of the presented research has two main objectives, both
to identify literature and to investigate the usefulness of the defined framework for classification of research in the area.

The review focuses on the evaluation of maturity models. Mapping studies are to a large extent carried out in the same way as systematic reviews, as described by Kitchenham and Charters (2007). A difference is that while a systematic review is conducted in order to identify best practice in an area based on presented research the objective of a mapping study is more to classify conducted research (Kitchenham et al. 2011). However the systematic approach to identifying relevant research can be conducted in the same way. This research is carried out based on the guidelines for performing presented by Kitchenham and Charters (2007). The procedure of this mapping study includes the following steps: planning, defining research questions, searching the databases, discussion of validity, data extraction and synthesis of the results. These steps are described in the next subsections.

Three researchers were involved in this research, and they are the authors of this paper. From here on the term ‘authors’ is used to mean these researchers.

4.1 Planning

The goal of the mapping study is to find out how the evaluation of maturity models is conducted and to prove the usefulness of the framework introduced in Section 3. A review protocol was developed in the beginning of the mapping study. The protocol includes the research background, the research questions, search strategy, study selection criteria and procedures, quality assessment, data extraction and data synthesis strategies. The intention is that this review protocol should make sure that the study is undertaken as planned and not driven by researcher expectations. In this review protocol, the whole study timetable was not decided from the beginning but the actual timetable of the study and results produced were recorded as the study progressed. The research questions and article identification strategies are described in the following sub sections.

4.2 Research questions

In this research, the methods used for the evaluation of maturity models are classified and mapped to different types in the defined framework presented in Section 3. The result is an overview of which evaluation methods have been used. This can guide researchers in choosing suitable methods for future evaluations of maturity models.

This can be formulated in the following main research questions:

- RQ1: What research has been conducted and reported in the area of the evaluation of maturity models?
- RQ2: To what extent is the framework presented in Section 3 useful for classification of the approaches to the evaluation of maturity models?
• RQ3: How can the framework be further extended in order to support researchers and practitioners developing evaluation approaches for maturity models?

Basically RQ1 is investigated through this mapping study, RQ2 in an extension of the mapping study, where the identified research papers are mapped to the structure of the framework. RQ3 is answered by reflections on the work with RQ1 and RQ2.

4.3 Search strategy and search process

Search resources

This study was planned to find relevant literature about the evaluation of maturity models. Based on the fact that most relevant papers for this study are in the software engineering area, two electronic databases were searched:

• INSPEC: This database is provided by Elsevier Engineering Information Inc. and the Institute of Electrical Engineers (IEE). It includes papers from 1969 to present.

• COMPENDEX: This database is provided by Elsevier Engineering Information Inc. It includes papers from 1970 to present.

Both databases intend to provide a complete coverage of the area, and include papers from all conferences, journals, and publishers (e.g. IEEE, ACM, Springer, and IEE). These two databases are by many experts seen as the leading databases in e.g. Computer Engineering and Electrical engineering and Electronics. Both databases were accessed through Engineering Village (http://www.engineeringvillage2.org).

Searches in the databases were carried out with a search string that is presented below.

Besides these two databases, the Journal of Software Process Improvement and Practice was searched manually. This journal was chosen because it is known to include published software process improvement approaches. The papers of this journal from 1995 to 2009 were searched, i.e. all available issues at the time the database searches were carried out.

Search process

After some tried searches, the following search string was decided on for this study:

\[
\text{((Evaluating WN KY) OR (evaluation WN KY)) AND} \\
\text{((maturity model*) WN KY) OR (CMM* WN KY) OR (SPICE WN KY))}
\]
The first line is intended to make sure that the papers have to do with the evaluation and the second line is intended to make sure that the paper is about maturity models. The term “WN” means existence, i.e. that the work before should exist in some part of the paper. “KY” denotes that certain words should appear in the title, abstract or key word list of the paper.

The search string was used on the two electronic databases on June 30th, 2009. After duplicate papers were removed, 1722 unique papers remained. After removing the papers that are obviously out of the area, 338 papers remained.

In the next step, the Journal of Software Process Improvement and Practice was searched manually on July 23rd 2009. In this journal, 21 relevant papers were found. All of these papers were already included in the list of identified papers, which shows that the initial search was able to identify these important papers.

Afterwards, the references of a selection of the most relevant papers were manually checked and as a result, another 5 papers were added.

**Inclusion and exclusion criteria**

A paper is kept in this mapping study if it satisfies one of the following two criteria:

- The paper is the report of an evaluation of one or more maturity models to identify merits and/or weakness of the models.
- The paper is the report of experience of using one or more maturity models in organizations for process improvement and some feedback was provided in order to improve the models. This includes papers whose main purpose is not to evaluate the maturity model, but still present an empirical evaluation of the used maturity model.

The papers were first reviewed based on titles, abstracts and key words, and they were classified in three different types:

- Relevant papers: if the paper satisfies one of the two inclusion criteria.
- Process assessment papers: if the paper is related to process assessment with maturity models, but not related to the evaluation of maturity models.
- Excluded papers: other papers, which are not relevant to either evaluation of maturity models or process assessment.

One author reviewed all 343 papers and put them into these different types according to the previous criteria. Then the other two authors checked part of the results, and found that some disagreement existed. The conclusion was that the disagreement was large enough to warrant the extra work of rechecking the classification of the rejected papers. Therefore the other two authors reviewed those papers that were excluded or classified as process assessment related, and re-added some papers into the relevant papers group. When there was doubt about
the classification of a paper, it was included in the relevant group, leaving the possibility to discard the paper during the next phase when the full papers were studied. The result of this stage was that 116 papers were classified as relevant to the evaluation and assessment of maturity models.

Classification

The three authors reviewed one third of the papers each, based on the full texts, and classified the papers into three types according to the framework in Section 3. After the first round of classifications, each author had some papers for which he/she wanted to discuss the classification. Then all authors reviewed those papers, and discussed their results together, and found they agreed on the result of the second round classification, so the classification result was accepted by the research team. During the final step of the selection, 55 papers were excluded based on their full texts, 2 papers turned out to be duplicates that were not detected before because the names were slightly different, and 59 papers were selected as relevant papers for this mapping study.

The search process is summarized in Figure 1.

4.4 Classification validity

The goal of this study is to cover as many as possible of the relevant research papers about the evaluation of maturity models. Nevertheless, it is likely that some relevant papers have been missed. This can be attributed to a number of different reasons. First of all, only papers published in English were considered. Because of practical reasons it was not possible to search through the scientific literature in other languages. Secondly, even in the English language there is some ambiguity and the search string described in Section 4.3.2 was chosen to include as many relevant papers as possible, while still returning a practical number of results. This means that some relevant papers that use a different terminology might not have been found. Thirdly, some lesser known journals and proceedings are not included in the electronic databases that were searched and any possible papers published in these collections were therefore not included in the results. A fourth reason is that the terms used were from a typical engineering perspective. It may be the case that the same type of model is used in other domains with other names. Finally, some papers can also have been rejected incorrectly during the selection process from the search results to the final list of relevant papers.

The scope of the mapping study is not limited to maturity models in the field of software engineering, although this is where maturity models have been the most popular and it is also the main focus of the mapping study. Because of the software engineering background from the authors and the way the search string was constructed, the selection of relevant papers is more likely to have missed
Using search string on databases (1722 articles identified)

Excluded articles based on titles and abstracts (338 articles remained)

Search references (5 new articles were found => 338+5=343 articles in total)

Review articles based on inclusion and exclusion criteria (116 articles remained)

Review articles based on full text (59 relevant articles remained)

Manually search Journal of Software Process Improvement and Practice (21 articles found)

Comparison to the 338 already identified articles revealed that no new articles had been found.

Figure 1: Paper identification process
relevant papers from other fields where also a different terminology is more likely to be used.

The following measures have been taken to improve the validity of the research and to minimize the number of missed papers:

- The inclusion and exclusion criteria at every step were explicitly defined and agreed upon by all authors. This makes the results from different authors more consistent and objective.

- For the selection based on abstract, after one author classified all relevant papers, the other two authors rechecked the rejected papers and further added any papers that they considered as possibly relevant. At any time when one author was not certain about the classification, the other authors looked at that particular paper and the final classification was decided when all authors agreed on it.

- To check whether the search string returned all the most relevant papers, the most important journal in the field of interest, the Journal of Software Process Improvement and Practice, was searched manually based on the content of the article and no new relevant papers were found.

- To further complete the list of relevant papers, the most important papers in the list were selected and their reference lists were systematically searched for relevant papers. This resulted in the identification of an extra 5 relevant papers that had not been included in the list before. Three of these papers were from CrossTalk, a journal not included in the databases, the remaining two had not been returned in the automatic search results because they used a slightly different terminology.

These measures together give us a good degree of confidence that most of the relevant papers, at least in the software engineering field, have been identified, although there is a risk that some less influential papers have been missed. Therefore, this mapping study cannot guarantee completeness, but can still be trusted to give a good overview of the relevant literature on the evaluation of maturity models, especially in the field of software engineering.

### 4.5 Data extraction

The data extracted from each paper were maintained through the whole review process. After identification of the relevant papers, the following data were extracted:

- The source (journal or conference)
- Title
### Evaluation type
- Static evaluation only by the authors of the paper (type 1 according to the framework)
- Evaluation by external experts in the field (type 2 according to the framework)
- Evaluation through process assessment (type 3 according to the framework)

### Size, report the following metrics
- Number of case studies for type 3
- Number of independent people involved in type 2

### Relation of the evaluators to the maturity model
- Evaluation of their own MM
- Independent evaluation by others
- Part of an official evaluation of MM (e.g. SPICE trials)

### Objectivity of evaluation
- Objective
- Subjective
- Both subjective and objective

### Purpose of the papers
- Evaluation of MM is the only purpose of the paper
- Evaluation of MM is a main purpose of the paper
- Evaluation of MM is carried out in addition to another main purpose of the paper

### Maturity model of interest
- CMM*
- SPICE
- Special adaptation of a famous MM (e.g. CMM for “small organizations”)
- New, self-developed MM
- Other maturity models, such as BOOTSTRAP

<table>
<thead>
<tr>
<th>Evaluation type</th>
<th>Static evaluation only by the authors of the paper (type 1 according to the framework)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Evaluation by external experts in the field (type 2 according to the framework)</td>
</tr>
<tr>
<td></td>
<td>Evaluation through process assessment (type 3 according to the framework)</td>
</tr>
<tr>
<td>Size, report the following metrics</td>
<td>Number of case studies for type 3</td>
</tr>
<tr>
<td></td>
<td>Number of independent people involved in type 2</td>
</tr>
<tr>
<td>Relation of the evaluators to the maturity model</td>
<td>Evaluation of their own MM</td>
</tr>
<tr>
<td></td>
<td>Independent evaluation by others</td>
</tr>
<tr>
<td></td>
<td>Part of an official evaluation of MM (e.g. SPICE trials)</td>
</tr>
<tr>
<td>Objectivity of evaluation</td>
<td>Objective</td>
</tr>
<tr>
<td></td>
<td>Subjective</td>
</tr>
<tr>
<td></td>
<td>Both subjective and objective</td>
</tr>
<tr>
<td>Purpose of the papers</td>
<td>Evaluation of MM is the only purpose of the paper</td>
</tr>
<tr>
<td></td>
<td>Evaluation of MM is a main purpose of the paper</td>
</tr>
<tr>
<td></td>
<td>Evaluation of MM is carried out in addition to another main purpose of the paper</td>
</tr>
<tr>
<td>Maturity model of interest</td>
<td>CMM*</td>
</tr>
<tr>
<td></td>
<td>SPICE</td>
</tr>
<tr>
<td></td>
<td>Special adaptation of a famous MM (e.g. CMM for “small organizations”)</td>
</tr>
<tr>
<td></td>
<td>New, self-developed MM</td>
</tr>
<tr>
<td></td>
<td>Other maturity models, such as BOOTSTRAP</td>
</tr>
</tbody>
</table>

| Table 1: Information collected from papers |

- Authors
- Publication year
- The evaluation or assessment type: type 1, 2 and/or 3
- Summary of the research, including which questions were solved

To be able to analyze the 59 papers there was a need to classify them in more ways than just according to the framework defined in Section 3. For this purpose further criteria for classifying the papers were defined and discussed by the research team, based on what information was available in the papers. When needed the categories were updated or clarified during the classification process. The result is presented in Table 1.

Based on the criteria for classifying the papers, all relevant papers were reviewed and the corresponding data were extracted.
It is not easy to identify an evaluation as objective or subjective. In this study, an evaluation is identified as objective if it looks at the correlation between maturity level and objective data like faults/KLOC, productivity, etc. An evaluation is identified as subjective if interviews about the maturity model are used to evaluate the model, or if they look for correlation between maturity level and employee satisfaction and how people evaluate the maturity model. Some papers are identified as containing both a subjective and an objective evaluation.

The Carnegie Mellon Software Engineering Institute has published a series of frameworks since the original publication of CMM, for example CMMI (CMMI Product Team 2006). The term CMM* is used in this study to refer to all related maturity models together.

The last categorisation partly overlaps with the third categorisation, especially concerning the self-developed maturity models. The last categorisation is mostly useful for being able to extract those papers that concern CMM* for further analysis in Section 6.

4.6 Data synthesis

The data synthesis was specified in the review protocol from the beginning of the mapping study. When there was any uncertainty about the classification of the primary studies, the issue was discussed by all authors until agreement was reached.

5 Results

5.1 Introduction

Appendix A lists the relevant articles from the mapping study. 59 articles are identified as relevant for this study. In this section, the classification of the articles according to Section 4.5 is discussed.

In Figure 2 the publication years for the identified articles are displayed. From this data it is not possible to identify any clear increasing or decreasing trend. Instead it can be concluded that there continuously have been published articles in the area.

The evaluation type of each paper is shown in Table 2.

Table 2 shows the classification with respect to evaluation type. It can be seen that 12 out of the 59 articles are classified as type 1, 4 articles are classified as type 2, and 50 articles are classified as type 3. It should be noted that a few articles have been classified as more than one type, e.g. article 12 as both type 1 and type 3.

This result shows that most of the cases used to evaluate maturity models are type 3, in which maturity models are evaluated by using them in software process improvement. Only 4 out of 59 cases used type 2, and half of them were combined with a type 3 evaluation.
Figure 2: Histogram of publication year for the identified articles

<table>
<thead>
<tr>
<th>Evaluation type</th>
<th>Identified articles</th>
<th>Number of articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>6, 11, 28, 30, 35, 37, 41, 46, 50, 51, 52, 56</td>
<td>12</td>
</tr>
<tr>
<td>Type 2</td>
<td>1, 4, 54, 59</td>
<td>4</td>
</tr>
<tr>
<td>Type 3</td>
<td>2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 45, 47, 48, 49, 51, 53, 55, 56, 57, 58, 59</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2: Evaluation type of papers
An example of how a type 3 evaluation can be carried out is presented in article 15, where data from already conducted assessments are used to evaluate CMM. Experience reports were collected from a database (PAIS). 707 of 948 reports were about CMM, which means that CMM could be evaluated with respect to internal consistency and correlation between related factors (dimensionality). This is a rather large evaluation with respect to number of included assessments.

An example of a type 1 evaluation is presented in paper 30 where two maturity models are compared by the researchers with respect to simplicity, validity, robustness, prescriptiveness, and analyzability.

5.2 Maturity models

The maturity models that were evaluated in the identified articles are of 5 types: CMM* (i.e. any type of model related to CM, such as CMM, CMMI), SPICE, adaptation of official Maturity Models (e.g. CMM for small organizations), self-developed maturity models, and other maturity models (common models but not so widely used as CMM or SPICE). The classification with respect to evaluated maturity models is presented in Table 3.

<table>
<thead>
<tr>
<th>Maturity model type</th>
<th>Identified articles</th>
<th>Number of articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMM*</td>
<td>3, 4, 5, 6, 9, 13, 16, 19, 20, 22, 30, 31, 34, 35, 36, 38, 41, 43, 44, 45, 47, 48, 50, 52, 57, 59</td>
<td>26</td>
</tr>
<tr>
<td>SPICE</td>
<td>8, 11, 15, 18, 19, 23, 24, 25, 26, 27, 28, 36, 52</td>
<td>13</td>
</tr>
<tr>
<td>Adaptation of official MM (e.g. CMM for &quot;small organizations&quot;)</td>
<td>7, 10, 14, 21, 29, 46, 53</td>
<td>7</td>
</tr>
<tr>
<td>Self-developed MM</td>
<td>1, 2, 12, 17, 32, 33, 37, 39, 40, 42, 44, 51, 54, 56, 58</td>
<td>15</td>
</tr>
<tr>
<td>Other maturity models</td>
<td>49, 55</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3: Evaluated Maturity Models

Table 5 shows that 26 articles were about evaluation of CMM*, 13 were about SPICE, 3 of which were about the evaluation of both CMM and SPICE. 7 were about adaptations of official maturity models. 15 evaluated their self-developed maturity models. One of them compared with CMM also. Only two papers were about evaluation of other common maturity models.

More than half of the studies were about CMM* or SPICE. About one fourth of the evaluations were about self-developed maturity models. Only a few were about adaptation of official maturity models, and even fewer cases were about evaluation
of other common maturity models. In Section 6 the articles presenting research on CMM* are discussed in more detail.

### 5.3 Relationship of the authors to the maturity model

Table 4 shows that 16 of 59 articles were written by researchers that had developed their own maturity model and evaluated it, of course these are nearly exactly the same articles as in the category of self-developed maturity models in the previous sections. Most of these maturity models are from the software engineering field, though this can partly be explained because we focused our search on this area. The evaluation of the maturity models is an important step in the development of it, and it is normal that the first evaluations are performed by the developers of the model.

<table>
<thead>
<tr>
<th>Evaluation group</th>
<th>Identified papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of their own MM</td>
<td>1, 2, 7, 12, 15, 17, 32, 33, 37, 39, 40, 42, 51, 54, 56, 58</td>
</tr>
<tr>
<td>Independent evaluation by others</td>
<td>3, 4, 5, 6, 8, 9, 10, 13, 14, 16, 20, 21, 22, 23, 25, 28, 29, 30, 31, 34, 35, 36, 38, 41, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 55, 57, 59</td>
</tr>
<tr>
<td>Part of the official evaluation of the MM</td>
<td>11, 18, 19, 24, 26, 27</td>
</tr>
</tbody>
</table>

Table 4: Relationship of the authors to the evaluated maturity model for each of the papers

6 of the 59 articles are part of an official maturity models evaluation, and all of those are from the SPICE trials, since this is the only maturity model that has been systematically evaluated in an official set of organized trials.

Most of the articles, 37 out of 59, discuss the evaluation of internationally accepted models by researchers or professionals that were not directly involved in the development process of the maturity model. These articles are most likely to be unbiased and can evaluate how easy the evaluated maturity model is to use as a finished tool for an organization without the involvement of the authors of the maturity model.

This classification is also important because the context in which the evaluation was performed has a large influence on the way the evaluation can be performed. In the evaluations of self-developed maturity models, most studies are quite small because of the huge effort required in this kind of evaluation. The advantage of these evaluations is that the developers of the maturity model are directly involved
in the application of the model and can check that all participants in the study apply the model consistently. In the evaluation of internationally renowned maturity models such as CMM, there is often a lot of data available because many organizations are using these models. However, the difficulty here lies in that the researchers have very little control over how these organizations apply the maturity model, making it hard to compare data from different organizations.

5.4 Objectivity

The evaluation objectivity of relevant articles is shown in Table 5. Articles in the subjective category are those using questionnaires and interviews to investigate the applicability of the maturity models and whether the effects of the maturity model are experienced as positive. The papers in the objective category investigate the direct effect of the maturity model on metrics such as productivity and fault rate.

McGarry et al. (1998) note that many evaluations are based on mostly subjective judgments of the effectiveness of maturity models and that there is a need for more objective evaluations. Table 5 shows that the mapping study identified more subjective than objective papers, but the difference is relatively small. Because both a subjective and objective evaluation are important in the complete evaluation of a maturity model, it is positive that there are many papers in each of the categories and that there are many articles combining both approaches.

<table>
<thead>
<tr>
<th>Evaluation objectivity</th>
<th>Identified articles</th>
<th>Number of articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>5, 6, 12, 13, 15, 18, 19, 20, 23, 24, 25, 26, 31, 32, 34, 38, 43, 45, 55, 59</td>
<td>20</td>
</tr>
<tr>
<td>Subjective</td>
<td>2, 4, 7, 9, 14, 16, 17, 21, 22, 29, 30, 33, 37, 39, 40, 41, 46, 48, 49, 50, 51, 52, 54, 57</td>
<td>24</td>
</tr>
<tr>
<td>both subjective and objective</td>
<td>1, 3, 8, 10, 11, 27, 28, 35, 36, 42, 44, 47, 53, 56, 58</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 5: Evaluation objectivity

It should be noted that this classification is not always easy to make. For the articles of type 3 of the framework, all maturity model assessments automatically contain a subjective factor. This is not what is meant with the subjective category in this study. The subjective/objective classification only takes into account the other variables studied, which are in each study being compared to the measured maturity levels. For some papers it can be hard to extract how some of the study data was collected making the classification of objective or subjective more uncertain. For the articles from type 1 and 2 of the framework the classification of objective or subjective is based on the used evaluation methods.
5.5 Size

The size of each study represents how many persons were involved in the study, how many companies or how many divisions of a company were included in the study, etc.

The extracted size data about the papers shows that it is generally very hard to generalize the size data or even to compare the size of a study presented in one article with another since different measures are reported in the different articles. Therefore the size data cannot be summarized statistically, nevertheless it can be seen that the list of relevant articles contains everything from very small to very large-scale evaluations.

For example, 545 survey participants were involved in a survey about CMM KPA in article 4, and this covers about half of the companies in CMM-based SPI programs at the time. 6 persons in one organization were involved in the evaluation of a requirements quality model in paper 54. In article 42, 18 companies were involved in a case study. The experience of 30 platforms in around 10 divisions was reported in article 33. However, in article 35, only one organization was involved in the evaluation, and in article 29, one case study was done in a virtual software organization.

6 Analysis of research on CMM*

In this work many articles on software process improvement based on different maturity models have been identified and analyzed. Based on this it is natural to say something about the general effectiveness of the models in process improvement. It would, for example, be interesting and valuable to determine the mean value of the effects of improvements based on different models. However, based on the identified articles it is not possible to do exactly this investigation for a number of reasons.

One reason that it is not possible is that it is very different how many studies there are for the each model. For CMM* and SPICE there are rather many articles identified, while there are only a few studies for each of the other models. Another reason is that the focus of the identified studies differs very much. Concerning SPICE, the focus of the studies is, as described above, not really on effectiveness, but more on e.g. consistency of different reviewers when interpreting factors of the model.

Concerning the articles on CMM* they investigate rather different factors, which makes the analysis of effectiveness from the articles impossible. However, it possible and relevant to further analyze the evaluations of the maturity models in the CMM family by summarizing the research that has been conducted. With CMM* we denote all models in the CMM family, i.e. both CMM and CMMI.

26 articles evaluating CMM* have been identified in this study and together they present many different types of evaluations. The largest group of these eval-
ulations (articles 3, 5, 13, 34, 38, 45 and 47) presents metrics collected from many projects across different maturity levels, at one or more organizations. These articles make it possible to statistically analyze the benefits and/or costs associated with an organization reaching a higher maturity level. This is the strongest proof for measurable improvements in an organization from using software process improvement. Article 20 even presents a meta-analysis of a number of these metrics from different independent usages of CMM*. Another similar overview of accumulated evidence of software process improvement in general, including a number of cases using CMM* can be found in (Krasner 1997).

The collection of this kind of metrics is very valuable but is often quite difficult in practice. Therefore, another common way to evaluate the effects of CMM* is with a large survey among organizations that are using CMM*, often with a specific focus. This approach is used in five of the identified articles (articles 4, 16, 22, 36 and 53), where the first and the last article focus especially on the problems faced by small and medium-sized businesses.

Three other articles (articles 9, 35 and 57) present experience reports of using CMM* for process improvement. These articles do not focus on presenting measured benefits or on collecting large amounts of data, but instead on describing in detail the positive effects and practical problems experienced by one organization starting out with software process improvement.

Three articles (articles 19, 31 and 43) present a statistical analysis of the data from a large number of CMM* assessments to investigate the internal consistency of the assessment methods.

All the articles discussed above are clearly type 3 evaluations. A number of type 1 evaluations of CMM* have also been performed. Three articles focus on one specific element of software engineering such as testing (article 6), security (article 41) or software architecture (article 50) and study how the best practices in this area are supported by CMM*. This allows them to identify some very specific possible shortcomings in CMM* and to propose an own framework that complements CMM* and specifically targets this area.

Another type 1 evaluation is to compare the assessment method of CMM* to some other common process improvement methods such as BOOTSTRAP (article 30), GQM (article 44) or SPICE (article 52).

Finally, there is also one article (article 48) that offers a detailed discussion of the advantages and disadvantages of using CMM* levels as a requirement when selecting contractors from the point of view of the US Department of Defense.

7 Discussion

For a more detailed discussion of the distribution of the relevant articles it is also important to look at the combination of different classifications. More than half of the articles present evaluations of CMM or SPICE, and these evaluations are done
mainly as type 3. About one fourth of the articles are evaluations of self-developed maturity models, and these evaluations are also done mainly as type 3.

All reported type 1 evaluations are done independently by others than the model developers. Possibly this type of evaluation has also been done by the developers themselves but then these evaluations have not been reported separately, and only those done by others have been reported. Only two articles used only type 2 evaluation. Both of them are evaluations of self-developed maturity models. Most of the type 3 evaluations are independent evaluations by others, and only few of them are part of the official evaluation. This is reasonable, because with a type 3 evaluation, the model should be used by others and the model’s effectiveness is assessed according to that.

Most of the self-developed maturity models were evaluated by the developers first, because the related articles are classified as evaluation of their own maturity models. This conforms to what we would expect to find from our own experience. Most of the independent evaluations are done regarding CMM and related models. This shows that CMM has been used for quite some time, and quite a few results have been published about it.

Most of the evaluations of self-developed maturity models are side effects of the articles, which shows that the developers of the maturity models usually publish the result of the model evaluation together with their maturity models, and that the evaluation results are only a small part of their articles. When the official evaluations are published, the evaluations of maturity models are usually the only purpose of the articles, and the models themselves are not explained, because the models were published already.

The evaluations of self-developed maturity models are more often subjective than the independent evaluations by others. As expected, all official evaluations are either classified as objective or as both objective and subjective. None of these evaluations are classified as subjective.

Regarding the relation between the objectivity of evaluations and the purpose of the articles, most of the objective evaluations are published as the only purpose of the articles. These objective evaluations are done by others than the model developers, so they usually only report the results of evaluations, but many articles by model developers explain the model and then some result from evaluation. This result is shown by the fact that most of the subjective evaluations are published as a side effect of the articles. Both subjective and objective evaluations are usually published as main purpose of the articles, and at the same time some other evaluations are published as only purpose or side effect, because the evaluation may be published separately or together with the model.

For the papers whose side product is maturity model evaluation, most of them (two thirds) are evaluation of self-developed maturity models. The rest of them are about evaluations of adaptations of official models. This shows that evaluation results are often published together with a model only when the model is new or has been changed for a special purpose. Otherwise there is no need to explain a
model that is already published, or even has been commonly used.

The presented mapping study also shows that there are large differences in the size of the evaluations in the relevant articles. Some articles combine data from hundreds of assessments at many different companies, while others are limited to a few projects at one organization. Generally, it is hard to compare this aspect of the studied papers because not all studies report all details of the data collected. An evaluation with many companies involved is not necessary a larger study than another evaluation with only one company involved, because of the differences in the size of the companies and the number of people involved in the studies.

8 Conclusions

In this paper, a mapping study of the evaluation of maturity models is presented. 59 relevant articles are selected from two electronic databases. The selected articles are classified according to the evaluation types, the relation of the evaluators to the maturity model, the objectivity, the purpose of the articles and the maturity model of interest.

A framework of different evaluations of maturity models is developed, and the relevant articles are mapped to the framework according to their classified evaluation types. The result of the mapping study shows that the type 3 evaluation, i.e. evaluation is conducted by actively using the maturity model in a process improvement effort, is the most commonly used evaluation method. Type 2, i.e. the evaluation involves some external experts from the field but without conducting an assessment, is the least used method. Only those type 1 evaluations done independently by others than the model developers were found in the review. Even though most of the published type 1 evaluations are not done by the developers, the evaluations are still classified as subjective, because the evaluations are done off-line and are not the result from using the model in process improvement.

A lot of results are reported about the evaluations of CMM (including other models in the same family) and SPICE. Especially CMM has been evaluated in many different ways and the advantages and disadvantages of this framework have been documented in many research papers. The evaluations of some self-developed maturity models were also found. It is easy to see that quite a few maturity models have been developed, but the evaluation results for self-developed maturity model are not many. It shows that it is quite difficult to put a new maturity model into use when not much evaluation result has been reported for it. To have more users is very important for the development of a maturity model because it means more feedback, more evaluation results, and the model itself can be improved faster.

This mapping study has looked into relevant articles and extracted data of published evaluation methods of maturity models. This can give a good overview for the researcher and practitioners in this area of how the evaluations have been done,
and it can give them guidance in what kind of methods can be chosen for future evaluations of maturity models.

Future work building on the research presented in this paper could, for example, focus on how this framework can be used to identify possible shortcomings in how some maturity models have been evaluated. This way, further interesting opportunities for the evaluation of some important maturity models could be identified. Currently there is not enough empirical data available in the literature to determine how effective developed models are in different situations in a meta-analysis. This research points to the need for providing this kind of empirical data, and an interesting area for future research would be to do this kind of meta-analysis when such data is available.

Bibliography


Appendix A: Relevant papers from the mapping study


43. Lester, N.G., Wilkie, F.G., McFall, D. and Ware, M.P. Evaluating the internal consistency of the base questions in the express process appraisal. *33rd Euromicro Conference on Software Engineering and Advanced Applications, 2007.*


Evaluation of a Maturity Model for IT Dependability in Emergency Management

Kim Weyns and Martin Höst

Abstract

The IDEM3 maturity model is a process improvement framework that can be used by an organisation to assess and improve their IT dependability management processes. The framework focuses on the coordination of IT management and safety management within an organisation. In this paper, an evaluation plan for the maturity model is presented to evaluate its applicability, assessment accuracy and practical value. Based on this plan, two evaluations were carried out in two case studies at two Swedish hospitals. The first evaluations indicate that the IDEM3 maturity model can be a valuable tool for an organisation to quickly identify the main strengths and weaknesses of the organisation in the field of IT dependability management. Therefore this study provides a strong argument for a further evaluation based on the developed evaluation plan.
1 Introduction

IT systems are a critical resource in most governmental organisations everyday communication and administration. Just as for their everyday tasks, governmental actors depend on all kinds of IT systems for their responsibilities in crisis situations (Santos et al. 2008). These systems include not only specially built systems for emergency situations but also the everyday communication and administration systems. The latter category of systems is of special interest. Under normal conditions an occasional unavailability of these IT systems is fully acceptable, but in emergency situations, when time is a critical factor, any unexpected unavailability can have disastrous consequences (Zimmerman and Restrepo 2006, Fleming 2001). Therefore it is important that these IT systems are an integral part of all major risk and vulnerability analyses conducted. This way information about the dependability of the different IT systems can be combined with information about how critical the systems are in different situations (BITS).

IT dependability management for organisations with a critical role in society is a complex process and the frequent occurrence of critical IT incidents shows that there is a lot of room for improvement. Earlier research (Weyns and Höst 2009) has shown that there is a need for a simple but effective framework to help organisations to address these issues in a structured way.

IT dependability management combines important elements of IT management and safety management in three different ways. First of all, the failure of critical IT systems can be a safety risk by itself, causing an emergency situation when the failure was unanticipated and no backup systems or manual routines are available. Secondly, in the case of an emergency in an organisation many IT-systems constitute an important resource in resolving the crisis, for example internal communication systems and medical information systems. Finally, IT systems also play a crucial role in the communication with the general public in the event of a crisis. This last category contains IT systems used by the traditional media as well as internet news sources and social media.

IT management and emergency management or safety management are different activities within an organisation often involving different parts of the organisation. Therefore there is a need for explicit coordination between these processes to make sure that IT dependability management issues are given the necessary attention (Weyns and Höst 2009).

This paper presents an evaluation plan and the first results of the practical evaluation of the IT Dependability in Emergency Management Maturity Model (IDEM3) for the coordination of emergency management and IT management (Weyns et al. 2010). The purpose of this maturity model is to help organisations to identify, evaluate and improve their IT dependability processes. The main focus of this framework is on the cooperation between emergency managers and IT personnel.
The IDEM3 maturity model has, in this study, been used for the assessment of the IT dependability management processes in two Swedish hospitals. In this way the applicability of the framework and the value of this assessment for the organisation are evaluated based on two case studies.

In this paper first some related work is presented, both to the IDEM3 maturity model and to the evaluation of maturity models in general. Secondly, the IDEM3 maturity model is summarized and the main elements of the framework are presented. Next, a practical evaluation plan for the maturity model is presented. Finally, two case studies based on this evaluation plan are presented in detail.

2 Related work

A number of process improvement frameworks have previously been published, both in the fields of emergency management and IT management. Emergency management is often coordinated on a national level and most countries have a special governmental agency such as the United States Federal Emergency Management Agency (FEMA), Emergency Management Australia (EMA), Public Safety Canada or the Russian Ministry of Extraordinary Situations (EMERCOM).

In the field of IT management a number of international standards and best practice frameworks have been published internationally, among those ITIL (Office of Government Commerce 2007), COBIT (ISACA 2000) and ISO/IEC 17799 (International Organization for Standardization 2005). These frameworks are more suited to be used by large corporations with very large IT resources and are less suited for smaller organisations and often do not take into account the special needs for organisations with an operative role in crisis relief. Vogt et al. (2011) have specifically investigated the use of ICT alignment frameworks in emergency management organisations and found that these frameworks are not well-suited for these type of organisations, mostly because they are not flexible enough. Of these models, COBIT (ISACA 2000) also makes extensive use of a maturity model.

The maturity model evaluated in this paper is based on a number of maturity models from related fields. The first successful maturity models were developed by the Carnegie Mellon Software Engineering Institute (Konrad et al. 1996). Since the development of the Capability Maturity Model, maturity models have been applied in many other fields. In 2008, SEI published a preliminary version of the CERT Resiliency Engineering Framework (Caralli 2007) for the use in the field of business continuity management with a special focus on IT systems. In the field of IT management, Luftman (2003) presents a simplified maturity model with a strong focus on the business value of IT systems. In the field of safety management, maturity models have also been proposed as a way of assessing an organisation’s safety culture (Fleming 2001), or product design safety (Strutt et al. 2006).
Santos et al. (2008) have also published a maturity model for the use of information technologies in emergency response organisations. Their model does not cover the dependability of the IT systems in emergency situations, but instead focuses on the information management practices of the organisation. The framework described in this paper is most suited for an organisation where the IT services are provided by an IT department that is part of the organisation. For evaluating the resiliency of IT services provided by external suppliers, Bhamidipaty et al. (2007) have developed the Resiliency Maturity Index, a framework for characterizing and evaluating the resiliency of an IT services organisation. However this model does not evaluate the relationship between the resiliency of the service supplier and the dependability needs of the organisation.

3 Background

3.1 Background Study

The IDEM3 maturity model (Weyns et al. 2010) presented and evaluated in this study is the result of two case studies at two Swedish municipalities to study their practices and problem areas concerning IT dependability management (Weyns and Höst 2009). The main conclusion from these studies was that many organisations experienced problems and frustrations concerning IT dependability management. The main cause of many of these problems could be traced back to communication and cooperation problems between the personnel in different roles involved. Further these studies also pointed out a lack of useful tools that support IT dependability improvements across a whole organisation. Vogt et al. (2011) have also shown that the IT process improvement frameworks discussed in the previous section don’t fulfil the needs of emergency management organisations.

To make lasting improvements to an organisation’s IT dependability management practices, a process oriented approach is needed. For the IDEM3 maturity model, it was decided not to develop a variation of one of the models above, but instead to develop a new, specially tailored maturity model to offer a process improvement framework that is simple and general enough to be applicable to many organisations and at the same time effective enough to make a substantial difference in an organisation’s IT dependability practices.

The process followed for the development of the maturity model is shown in Figure 3.1. First, the case studies in (Weyns and Höst 2009) that describe the need for this kind of maturity model resulted in a list of factors that are important for the coordination of IT dependability management. These key factors formed the first basis for the attributes of the maturity model.

In a next step, the factors were mapped to the general architecture of a maturity model with five levels as found in many other maturity models. For the model to be applicable by small organisations, it was necessary to simplify the structure by
replacing the concept of ‘key process areas’ by the more modest ‘attributes’ found in the IDEM3 framework.

In this mapping the attributes were also compared and complemented with similar attributes found in maturity models from related fields. Before the model was applied in a practical setting, the model was validated with the help of experts and practitioners in the field. The feedback from these experts resulted in an updated version of the maturity model, as published by Weyns et al. (2010).

Figure 1: The development process of the IDEM3 maturity model. Shown in red, the evaluation of the maturity model, and more specifically the evaluation plan and its first applications in a practical setting, are the focus of this paper.

3.2 IDEM3 Maturity Model

Just like most other maturity models, the IDEM3 model has 5 maturity levels. The Initial level is the most basic level, representing an organisation where some critical IT systems have not been analyzed from a dependability point of view and nobody takes responsibility for initiating a more strategic discussion about IT dependability.

The second, Managed, level is characterized by an organisation where the dependability of all critical IT systems is managed on a system-by-system basis leaving the organisation very dependent on the competence of the system managers for every system.

The third level is referred to as the Established level. This means that the organisation has established a centrally coordinated approach for dealing with IT dependability. This will usually be established by appointing one central IT dependability manager who distributes standard procedures for dependability analysis to all system managers. A standardized approach is a prerequisite for being
able to implement future improvements across the whole organisation. A level 3
organisation also has clearly defined responsibilities concerning IT dependability.

The fourth level, called Quantitatively Managed, is similar to level 3, but also
requires that the centrally coordinated approach is supported by extensive quanti-
tative data collection. Regular measurements and testing with special usage sce-
narios in mind can make IT dependability statistically predictable and allow for
strategic improvements in IT dependability.

The Continually Improving level, which is the fifth and final level of the matu-
ritry model, is reached by an organisation that can use the feedback obtained from
the practices from level 4 to continually improve not only their IT systems, but
also their own IT management procedures. IT systems will then be naturally in-
cluded in risk and vulnerability analyses and their dependability will be regularly
re-evaluated.

To define the levels in more detail the levels can be compared on 22 attributes. Of
course all these attributes are in some way related and none of them can be
changed completely independent of the others. Nevertheless they each add their
own focus to the maturity model and stress a special aspect of an organisation’s
maturity.

The 22 attributes can be divided in 4 categories: Outcomes, IT management,
Cooperation and Organisational Issues. A detailed summary of these attributes
can be found in Table 1 and the attributes in each category are also described in
the following subsections. The attributes are ordered in such a way that those
attributes that are most strongly related are placed next to each other.

The IDEM3 maturity model is much smaller in size than other maturity models
such as the CERT REF (Caralli 2007). The attributes in IDEM3 fulfil a different
role than for example the 26 capability areas in CERT-REF and are much more
specific and smaller in scope. This makes it much faster to assess the maturity
levels of an organisation’s attributes, which can be done directly based on the
description in the maturity grid presented in Table 1.

The main goal of IDEM3 is to be a process improvement tool for organisations
wanting to assess and improve their IT dependability practices. The framework
should be sufficiently simple to allow a quick assessment with minimal resources.
The result of the assessment should be an overview of the assessed organisations
weaknesses and strengths in this area, and a roadmap for sustainable improvement.
Unlike for some of the other maturity models discussed above, the assessment is
not meant to be used as a benchmarking tool for comparing different organisations.
In that case there would be a higher risk for bias in the results and the assessment
would have very high requirements on its correctness. For the IDEM3 maturity
model, the primary value in the assessment is not in exactly determining the or-
ganisations maturity level, but in bringing together all involved parties and starting
a discussion about all attributes of the model, and identifying those attributes were
the organisation needs to focus their improvement efforts.
<table>
<thead>
<tr>
<th>Level 1: Initial</th>
<th>Level 2: Managed</th>
<th>Level 3: Established</th>
<th>Level 4: Quantitatively Managed</th>
<th>Level 5: Continually Improving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>Reactive</td>
<td>Responsive</td>
<td>Proactive</td>
<td>Predictive</td>
</tr>
<tr>
<td>Problems that can be identified</td>
<td>Failures</td>
<td>Inefficient availability, high dependence and interdependencies</td>
<td>Failures in risk analysis procedures</td>
<td>Safety culture</td>
</tr>
<tr>
<td>Basis for improvements</td>
<td>Error</td>
<td>Personal judgement</td>
<td>Standards</td>
<td>Quantitative risk analysis procedures</td>
</tr>
<tr>
<td>Improvements</td>
<td>Few, changes can be both positive and negative</td>
<td>For some systems, not sustained</td>
<td>For all systems, sustained</td>
<td>Regular, organised</td>
</tr>
<tr>
<td>Success</td>
<td>Not repeatable</td>
<td>Repeatable</td>
<td>Sustainable</td>
<td>Measurable</td>
</tr>
<tr>
<td>Success factor</td>
<td>Luck, Competence of key personnel</td>
<td>Personal competence of systems managers</td>
<td>Central coordination</td>
<td>Cooperation, Measurements</td>
</tr>
<tr>
<td>Risk of IT in emergency situations</td>
<td>Cause of problems</td>
<td>Varying between systems</td>
<td>Under control, backup solutions available</td>
<td>Predictable behaviour</td>
</tr>
<tr>
<td>Dependency problems</td>
<td>Accepted, voice change</td>
<td>Faults on individual basis</td>
<td>Identifiable problems identified and prevented</td>
<td>Possible problems predicted and prevented in prioritised way</td>
</tr>
<tr>
<td>IT dependency</td>
<td>Law</td>
<td>Mixed</td>
<td>Stable</td>
<td>Controllable</td>
</tr>
<tr>
<td>IT incident management</td>
<td>Ad-hoc</td>
<td>System-based</td>
<td>Formal incident management procedure</td>
<td>Basis for learning</td>
</tr>
<tr>
<td>IT responsibility management</td>
<td>Systemic, Technical focus</td>
<td>Single initiatives, Technical focus</td>
<td>Base level, linked to requirements</td>
<td>Detailed level</td>
</tr>
<tr>
<td>Dependability Requirements</td>
<td>Unknown</td>
<td>For some systems</td>
<td>Documented</td>
<td>Measurable</td>
</tr>
<tr>
<td>Service levels</td>
<td>None</td>
<td>Basic services for entire organisation and for all systems</td>
<td>SLA with measurable service levels for all systems</td>
<td>Continuously evaluated SLA</td>
</tr>
<tr>
<td>IT dependability analysis and emergency planning</td>
<td>Separate activities, not combined</td>
<td>Activities with a shared goal, contact for some systems</td>
<td>Input for each other</td>
<td>Measurable values exchanged, direct connection</td>
</tr>
<tr>
<td>IT availability</td>
<td>Not included</td>
<td>Identified for some systems</td>
<td>Risk and vulnerability analysis for all IT systems included in plans</td>
<td>Available values included, regularly tested</td>
</tr>
<tr>
<td>IT availability</td>
<td>Conflict</td>
<td>Personal relations</td>
<td>Cooperation between departments</td>
<td>Cooperating, with shared risk and reward</td>
</tr>
<tr>
<td>IT incident management</td>
<td>Nobody</td>
<td>Individual system managers and some stakeholders</td>
<td>All internal stakeholders</td>
<td>All internal stakeholders, some external stakeholders</td>
</tr>
<tr>
<td>Risk management</td>
<td>‘Someone else’</td>
<td>Individual System managers</td>
<td>(delegated by) IT security manager</td>
<td>Shared between System, Process, IT, and Safety managers</td>
</tr>
<tr>
<td>Management and leadership</td>
<td>None</td>
<td>Receive “Don’t do that again!”</td>
<td>Following rules (“Do like this”), Open loop</td>
<td>Following policy (“Think before you act”), Single loop learning (“Imagine what you do”), Double loop learning (“Improve the way you see things”)</td>
</tr>
<tr>
<td>Resource allocation</td>
<td>Reactive, All has</td>
<td>To individual projects</td>
<td>Divided equally over the whole organisation</td>
<td>Divide in prioritised way</td>
</tr>
</tbody>
</table>

Table 1: The 22 attributes of the IDEM3 Maturity Model across the 5 levels
3.3 Evaluation of IDEM3

The IDEM3 maturity model has previously only been evaluated by relating it to other renowned maturity models and with the help of experts from the field (Weyns et al. 2010). The most valuable evaluation of the framework, through the actual application of the maturity model in a real setting, is the topic of the evaluation plan developed in this paper. This type of evaluation should make it possible to evaluate the applicability, the measuring power and the practical value of the IDEM3 framework. This paper presents an evaluation plan for this type of evaluations of the maturity model. Through two case studies based on this evaluation plan, the results of the first practical evaluations of the IDEM3 maturity model are presented.

4 Evaluation Plan

The planned evaluation consists of a complete IDEM3 assessment, where at the end the participants also evaluate the result of the assessment. The process for one evaluation is shown in Figure 4.

![Figure 2: Evaluation process for an IDEM3 assessment](image)

4.1 Scoping

The first step in each assessment is always an exact scoping of the planned assessment. It should be defined exactly which organisation, or part of the organisation, is being assessed. From the start, it is important to assure the approval and support of the management of the organisation. Also, a number of systems need to be selected that will be the topic of the collected data, as it is usually impossible to cover all IT systems used by the organisation. For the result to be relevant, the selected systems should be critical to the organisation, and representative for how IT dependability is managed within the organisation. In this step, it is also possible to limit the assessment to a subset of the attributes of the maturity model, instead of all attributes.

Based on this scoping, a first list of interviewees should be compiled. This should include at least the system managers for the selected systems, IT personnel
responsible for maintenance of these systems and an IT safety manager, or people in similar positions in the organisation.

4.2 Data Collection

In the next step, the necessary data should be collected to make a correct assessment possible. The main source of data consists of the structured interviews with the selected respondents. This can be complemented with the collection of important documents concerning IT dependability management, such as service level agreements, incident reports or strategic IT plans. The interviews should be based on a predetermined list of questions, consisting of a number of general questions and a set of questions for each selected attribute.

The proposed outline for the interviews is shown below, where the questions need to be reformulated slightly to fit each attribute. Each interview starts with an introduction followed by some general questions asking about the role of the respondent related to the studied systems and their experiences with the dependability management of these systems. Then, for each of the studied attributes, the interview contains a number of open questions, allowing the respondents to give their main impressions of how the organisation deals with the attribute and of the strengths and weaknesses the respondent experienced in this area. For each attribute there is also the more specific questions Q5a to Q5e, which can be skipped in case they have already been answered completely by the responses to questions QA1 to QA4. Finally, each interview is concluded with some closing questions. Each respondent is also asked to name a few other possible respondents that should be included in the interviews to get a complete picture of the organisation’s IT dependability management concerning these attributes. This way it is possible that during these interviews additional respondents need to be added to the list.

4.3 Assessment

In the next step, the collected data should be analysed to map the organisation to the levels of the maturity model for each of the attributes. For this step, the collected data should be analysed by the interviewer. Because of the specific questions in the interviews, the analysis should focus mostly on comparing answers from different respondents. When there are clear differences between the studied systems, this should also be studied in more detail based on the collected data.

Based on this analysis, the organisation can be assigned a maturity level for each assessed attribute and an overall maturity level. When many attributes have been assessed, the results can preferable be summarized in a spider graph.
Outline of the semi-structured interviews

Introduction (goal of the interview, scope, confidentiality ...)

Introductory questions:
Q1: Can you describe your role concerning System X?
Q2: How critical is System X for the organization?
Q3: In your experience, has the system been sufficiently reliable, in relation to how critical it is?
Q4: Do you consider that the reliability of the system is managed in a good way?
   a) Which aspects are functioning well?
   b) What could be improved?

For each selected attribute:
QA1: How is attribute A managed for System X?
QA2: Is attribute A important for the reliability of the system?
QA3: What is functioning well, concerning attribute A?
QA4: What could be improved concerning attribute A?
QA5: Do you experience ... concerning attribute A?
   a) attribute A, level 1
   b) attribute A, level 2
   c) attribute A, level 3
   d) attribute A, level 4
   e) attribute A, level 5

Concluding questions:
Q5: Do you have anything else important to add concerning dependability management for System X?
Q6: Who should I interview more, to get a complete picture of the IT dependability management for System X?
4.4 Focus group meeting

The final step of the assessment consists of a focus group meeting where most of the interviewees were present, together with further staff of the organisation with important roles in IT or safety management. During this focus group meeting, the assessor presents and leads a discussion on the results of the assessment for each of the evaluated attributes. For each individual attribute, and for the organisation as a whole, the improvements required to achieve the next maturity level in the maturity model are also discussed.

Because there are no other objective measurements of these attributes of IT dependability management, the correctness of the assessment cannot be objectively evaluated. What can be measured is the correctness of the assessment as perceived by the members of the organisation, the novelty of the identified weaknesses and the perceived value of the whole assessment for the organisations. To measure this, the focus meeting should be accompanied with a survey among the participants to collect their opinion on the result of the assessment.

The assessment is only the first step in a larger process improvement effort where practical improvements are planned, implemented and evaluated. However, this is outside the scope of this evaluation study, which focusses entirely on the evaluation of the assessment.

5 Case Studies

The evaluation plan above was used in two case studies. The methodology used in both case was based on the evaluation plan, although some improvements to the plan were made prior to the second study based on the lessons from the first study. The formal evaluation through the survey form was added to the evaluation plan and was therefore not part of the first case study. The two organisations in the study were selected because they had expressed their interest in assessing their IT dependability management practices to the authors of this paper.

Because of the similarities of the two studies, this section first presents the details concerning the methodology of each of the two case studies. Next the results of the assessment of both organisations are discussed together, followed by the results of the evaluation of the IDEM3 maturity model. Finally, the validity of the case studies is discussed.

5.1 Case Study 1

The first object of study was a large hospital in western Sweden, in the rest of this paper referred to as Hospital A. As with most Swedish hospitals, the studied hospital is a formal part of the regional government of the region it’s located in. As part of this larger organisation, the hospital depends for all its IT services on the regional IT unit that was formed when local IT personnel were joined in one central
IT unit that serves the complete organisation. To limit the scope of this first study, only the situation at Hospital A was assessed. Because this was the first assessment based on the IDEM3 maturity model, it was further decided, in agreement with the organisation being assessed, to limit the scope of the assessment to just two critical systems and to three attributes of the maturity model: Service Level Agreements, Incident Management and Involvement. They represent 3 of the 4 categories in the maturity model and were selected based on the interest of the organisation.

In a first step, the organisation selected two of its most critical systems, two different patient data systems. For both of these systems a number of interviewees were selected by the IT safety coordinator for the region. Because, for practical reasons, the study was to be conducted in one week time the decision was taken to limit the number of interviews to 5: the hospital IT safety coordinator together with for each system the main IT technicians responsible for servicing the system and the system manager of the system. In this study the system manager is a non-technical coordinator for the system responsible for representing the users of the system at the hospital in the cooperation with the IT department.

During the course of one week, telephone interviews with these 5 participants were conducted by the main author of this paper, based on the questionnaire outline presented above. To complement the interviews, a number of important documents relating to IT dependability were collected.

The interviews lasted about 30 minutes each and the interviewees’ answers were recorded on specially prepared forms by the interviewer. These answers were then later transcribed in electronic form for distribution to the other authors of this paper.

The collected data was then independently mapped to the levels of the maturity model by the first two authors and the results were discussed to agree on a preliminary assessment for each of the three assessed attributes.

Finally, the results of the preliminary assessment were discussed during a focus group meeting where most of the interviewees were present, together with further staff of the organisation with important roles in IT or safety management, both from the local and regional level. The results of the assessment were evaluated in a structured discussion about each of the evaluated attributes. The result of the assessment and the evaluation are described in the next section.

5.2 Case Study 2

The second case study was also conducted at a Swedish hospital, in the rest of this study referred to as Hospital B. Hospital B is a smaller hospital than Hospital A, located in a smaller city in the same region as Hospital A. Both hospitals have been part of the same organisation for about ten years, since the counties they were located in where joined together in one administrative region. About 5 years ago a regional IT department was created that provides IT services to both hospitals.
The assessment performed in these case studies is a small part in the regional coordination of IT and safety procedures across the whole organisation.

The procedure for the second case study was very similar to the first. Together with the organisation under study it was decided to focus on 4 attributes, 2 of which were the same as in the first study, and 2 other ones. This made it possible both to compare some of the results with the previous study while at the same time also evaluating other parts of the maturity model.

For the second study, 9 telephone interviews were conducted with 10 participants. On explicit request of two of the interviewees in similar roles, one interview was conducted with them together instead of separately. 9 of these participants were selected by the IT safety manager, one of the participants was added later based on information gathered during the interviews. The interviews lasted 45 to 60 minutes each, a bit longer than in case study 1 since there was one more attribute to discuss. Again, the answers were recorded on specially prepared forms, but this time the interviews were also recorded electronically for greater accuracy when the answers were later transcribed.

Just as in the first case study, the answers were mapped to a maturity level for each of the 4 discussed attributes. Because of the strong relationship between the different attributes, enough information had been collected in the interviews to allow also a rough estimate of the maturity level of the other attributes of the maturity model, although with a higher uncertainty.

Also in the second case study, the results of the assessment were presented and evaluated on a focus group meeting with representatives of the assessed hospital. For the second case study the assessment was also evaluated through a more formal survey among the participants in the focus group meeting. The results of the assessment and of the consecutive survey are presented below.

5.3 Result of the Assessments

The results of the assessment of hospital A on the 3 studied attributes can be summarized as follows:

- **SLA: level 3 maturity**: SLA’s are well established as a tool, but are not regularly evaluated or updated as is required in level 4.

- **Incident Management: level 3 maturity**: Formal incident management procedures are defined and used and some data is collected, but the data is not being analyzed and not available to those who would be interested in analyzing it.

- **Responsibility: level 4 maturity**: Responsibilities are clearly defined and shared between stakeholders. Many stakeholders are actively involved and the personal relationships between them are mostly good with only some minor frustrations. For level 5 the cooperation should be explicitly evaluated and improved over time.
No important differences between the two studied systems were recorded with the same process being used for both systems. For each of the systems and for each of the attributes the level was at least at the ‘established’ level (level 3) showing that the organisation as a whole has an organized approach towards IT dependability for critical systems.

The results of the assessment of the maturity level of hospital B in each of the 22 attributes can be seen in Figure 5.3. The four attributes that were the main focus of the study assessment have been marked in the graph. Note that some attributes in the graph are marked at level 2.5, meaning that they were between level 2 and 3.

The results show that the organisation’s maturity was at least at level 2, while the majority of the attributes was assessed at level 3 and two organisational attributes were assessed at level 4. The results of the study’s 4 main attributes can be summarized as follows:

- **SLA**: *in transition from level 2 to level 3*: general agreements are in place, sometimes referred to as SLA’s. Standardized SLA’s are being worked on. Further, the reliability and security requirements of all systems have been explicitly classified (Basic, High or Very High) to allow IT resources to be prioritized easier.

- **Incident Management**: *level 3 maturity*: Formal incident management procedures are defined and used and much data is collected, but the data is not being analysed and not available to those who would be interested in analyzing it. There is some confusion between the procedures for reporting IT related incidents and reporting patient safety incidents in those cases where they overlap.

- **IT Dependability Analysis and Emergency Management**: *between level 2 and 3*: there is close cooperation between the people working with these two issues but it is not made explicit in all documentation of the systems, through emergency scenario exercises, or through specific system testing for emergency situations.

- **Organisational Learning**: *level 4 maturity*: A lot of effort is spent on organisational learning concerning IT risk management, and many effective ways of spreading information exist. A lot of progress has been made in this area lately. For level 5, some documentation of routines should be kept up to date and the importance of learning should be made more explicit to all personnel.

The overall conclusions of the assessment of the organisation can be summarized as follows. Many attributes are already at level 3 and the first priority should be to also get the remaining attributes stable at level 3. This will mostly require more standardization in these areas, by generalizing good practices that are used in some parts of the organisation to all organisational units.
5.4 Evaluation of the Maturity Model

The first assessment at hospital A was evaluated informally during a focus group meeting during which the results of the assessment were presented and the comments of the participants in the meeting were written down. 10 employees of the assessed organisation, among them most of the interviewees of the first case study, participated in the focus group meeting, which lasted about two and a half hours. The focus group was led by the first author, while the second author took notes for later analysis. During the meeting, each of the three attributes was discussed based on the levels in the maturity model, and only then the result of the assessment was given to evaluate whether it matched the impression of the participants in the meeting. The focus group generally agreed with nearly all the conclusions from the assessment, although the maturity level of the first and third attribute was considered as possibly a bit too forgiving for the occasional problems the organisation was still experiencing in these areas. There was agreement that this could probably be explained by the fact that the two studied systems were not typical systems for these attributes, and that with a different choice of systems the assessment might have revealed more of the problems in these areas. The main conclusion of the first evaluation was that there was definite value in the assessment for the organisation, not only in measuring the maturity level of the organisation in the assessed attributes but even more in bringing together the different actors and
acting as guidance in a discussion about problems and possibilities within these areas. A result of the positive evaluation of the first assessment was also that there was serious interest from the organisation to do a similar assessment at another part of the organisation. Because of the interest from the safety managers from hospital B present at this meeting, it was decided to conduct a similar, but more elaborate, assessment at hospital B.

At hospital B a similar focus group meeting was held to present the results of the assessment but this time the results were evaluated through a survey that all 13 participants filled in (in Swedish) during the forum group meeting or just afterwards.

For each of the 4 assessed attributes and for the overall assessment we asked participants in the survey to rate the following:

- The assessment error, the deviation of the assessment from what they, based on their experience, considered the correct reasonable maturity level for each attribute, on a 5-point scale from ‘much too high’ over ‘reasonable’ to ‘much too low’.

- Agreement with the proposed improvements from strongly disagreeing to strongly agreeing, also on a 5-point scale.

- How much the results of the assessment provided new information to them from ‘nothing new’ to ‘everything was new’, again on a 5-point scale.

- The value of conducting the assessment from ‘no value’ to ‘very valuable’, on a 7-point scale.

Because the last question was the most important for measuring the value of the maturity model, a 7-point scale was chosen instead of a standard 5-point Likert scale. For each of the questions the respondent could also mark a ‘don’t know’ option. For each of the assessed attributes there was also an open question where participants could write what part was most or least valuable in the assessment.

It was stressed to all the respondents that the opinions expressed in the survey were meant to be based on their personal experience, and not on the consensus from the focus group meeting. We can therefore expect some differences in the answers of different respondents. The following graphs present the respondents answers to the grading questions. The sum of all the bars corresponding to the same question sum up to 13, except for 5 of the questions where one of the respondents did not answer the question.

The top left graph in Figure 5.4 4 shows that the assessment was considered correct by nearly all respondents for 4 out of the 5 assessments. Only for the organisational learning attribute the assessment was considered to be too lenient and most participants did not think the organisation had reached maturity level 4 in this area. Review of the interviews after the meeting, shows that this can partly be explained by the fact that much of the organisational learning that is
measured is implicitly present in the assessed organisation although these practices are not explicitly thought of as organisational learning by the respondents. Only half the respondents in the survey were also part of the assessment, while the others were only involved in the evaluation. Therefore, the fact that the assessment for 3 of the 4 attributes fits really well with the personal opinion of most of the respondents, with even both positive and negative deviations for two attributes where some respondents partly disagree, shows that the assessment managed to assess the maturity of the organisation with sufficient accuracy to be useful in the further discussion of these attributes.

The top right graph in Figure 5.4 shows that, for each of the 5 assessment in the case study, there was strong agreement with the improvements proposed by the maturity model. This shows that the maturity model was useful in identifying meaningful improvement suggestions for each of the assessed attributes.

The bottom left graph in Figure 5.4 shows the novelty of the results of the assessment for the participants in the evaluation of the study. Although there are large differences between different attributes and different participants we can conclude that each of the assessment conclusions only contained few new elements or none at all for most of the participants. Some of the suggested improvements were already things that the organisation was actively working on, or that many people were aware of could be improved. This shows that the strength of the maturity model does not lie in that it identifies new and unknown weaknesses in an organisation's IT dependability management. That so many of the conclusions of the assessment were not new to most of the participants of the study is possibly a consequence of the organisation already actively working with improvement of its IT dependability management, which also shows in the relatively high maturity level of the organisation. For an organisation at level 1 or 2, it might be assumed that more of the conclusions will be new to the organisation as less people will be actively working with IT dependability.

The final, bottom right graph in Figure 5.4 is the most important graph and shows that all participants saw a lot to very much value for the organisation in each of the assessments. Only one participant saw no value at all in the assessment of the first two attributes, but then saw very much value in the other assessments. From this participant’s free-text comments it is clear that this is because there was nothing new about these two first results for this participant and the participant considered the improvement suggestions as things that were already actively being worked on. Across all participants and across all 5 assessments, we can conclude that the assessments were experienced as very valuable for the organisation.

Especially this last result is very significant and shows that the participants agreed that the IDEM3 maturity model can play an important role in the assessment and improvement of an organisation’s IT dependability processes. The answers to the open questions show that the most important role of the assessment for the participants is that it is a very good way to involve all stakeholders in an organized discussion and that the spider graphs offers an easy-to-understand overview
Figure 4: Evaluation of the assessment for each of the attributes by the participants in the focus group at hospital B. The vertical axis shows the number of participants that selected a given answer.
5 Case Studies

of the current maturity level of the organisation and the short term improvement goals. The main shortcoming of the maturity model, mentioned in the free-text answers of multiple participants in the focus group meeting, was that although the IDEM3 framework suggests a list of improvements, it does not explicitly say how they should be implemented and who should be responsible. Unfortunately this is not an element that can easily be added to the maturity model without losing general applicability across many types of organisations. The organisation of responsibilities concerning IT dependability can vary a lot between organisations. Also for the method used to implement the improvements, an organisation should use the methods they are already most familiar with, for example Plan-Do-Check-Act (Pittman and Russell 1998).

5.5 Discussion of the results

From the results of these two assessments and the evaluations of these assessments we can draw a number of conclusions. First of all, the case studies show that the IDEM3 maturity model can be used to conduct an assessment of the IT dependability of an organisation with a relatively small effort. An extensive assessment that includes all the attributes of the model and multiple IT systems should be possible with a total interview time of 30 to 40 hours, and should include between 10 and 20 participants. This means that the application of this framework is also possible for smaller organisations that only have a small IT-unit without taking up too many resources.

Secondly, concerning the correctness of the assessment, the case studies show that the maturity model managed to correctly assess the maturity level of the organisation for most of the attributes and to identify the most important improvement areas for all attributes. The focus group meeting at the end of the assessment can be used to fix the any small errors in the maturity levels, if necessary. The case studies therefore show that the correctness of the assessment is more than sufficient to be useful to form the basis for process improvement.

Thirdly, concerning the value of the IDEM3 maturity model, the case studies are very clear. There is strong evidence that the assessment as a whole was experienced as very valuable by all the participants in the study. The evaluation also shows that the value does not, as might first be suspected, lie in the power of the assessment to identify previously unknown shortcomings in an organisation’s IT dependability management. On the contrary, the assessment identified mostly problems that were already known by many of the participants. The value of the assessment lies in its power to bring together all stakeholders in the IT dependability management of an organisation and to present them with a clear summary of the strengths and weaknesses of the organisation in this area. This then forms the basis for process improvement in this area.

The apparent contradiction between the low novelty and the high value of the results of the assessment shows that the most important part of the maturity model
is not the details of each of the attributes, but to provide an organised approach for implementing improvements in IT dependability management. This is also confirmed by the positive response to the spider graph, as it is seen as a practical way to communicate the organisations progress to all involved parties.

5.6 Validity discussion

During the course of this study a number of threats to the validity of the study were identified. First of all, these first evaluations of the IDEM3 were conducted by the authors of the maturity model and therefore some measures needed to be taken to minimize the bias in the results. The two hospitals that were the subject of this case study were not part of the case studies that supplied the data that lead to the development of the maturity model and had no prior experience with the IDEM3 maturity model. The respondents to the surveys used in the second evaluation were given the possibility of answering the survey anonymously and independently of each other. Although politeness could have enticed the respondents to answer more positively, the positive response to the value of the maturity model is strong enough (with many scores of 7 out of 7 on the value question) to allow us to assume that this is not a major factor in the evaluation.

Secondly, to avoid errors in the assessment part of the evaluation, the telephone interviews were recorded when possible, and these recordings were then used in the analysis of the maturity level of the organisation. All interviews were conducted by the first author but the results of the assessment were discussed with the other authors. For the preparation of the interviews, as well as for the planning of some of the other aspects of these case studies, the guidelines from (Yin 2003) were followed.

Thirdly, and most importantly, this first evaluation of the IDEM3 maturity model has only involved two hospitals served by the same IT organisation. Therefore the results cannot be generalized to other organisations or other settings. Because of the large amount of work involved in the assessments and the evaluation, it is impossible to show generalizable results before the maturity model is being independently used by a larger group of users. However, these two case studies show the applicability of the evaluation plan and the positive results provide a strong argument to continue the evaluation and to promote the use of the IDEM3 maturity model to a larger audience with the help of the Swedish Civil Contingencies Agency.

To further check the validity of the results of this study, the final version of this paper has been reviewed and approved by the two hospitals that participated in the study. Several months after the completion of the studies, the organisations themselves report an increased understanding between IT personnel, users and safety personnel, as a result of the discussions started by the assessment based on the IDEM3 maturity model. To really determine the value of the improvements made in the organisation, a new assessment would need to be done, to investigate
if a higher maturity level can be measured. However, this is outside of the scope of this evaluation.

6 Conclusion and future work

The IDEM3 maturity model is a process improvement framework in the field of IT dependability management. The maturity model is based on 22 attributes divided over 4 categories. An organisation can use the maturity model by first using the maturity model to conduct an assessment for each of the attributes by mapping the organisation’s current practices to one of the 5 levels defined for each attribute.

In this paper, an evaluation plan for IDEM3 assessments is presented and applied in two case studies. The goal of this evaluation plan is to enable the evaluation of the usability, the correctness and the value of the IDEM3 maturity model for the assessment of the IT dependability management in an organisation. Although only two small evaluations were performed, these first studies present practical evidence that this framework can be used to quickly and relatively easily conduct an assessment of an organisation’s IT dependability management. The assessment was seen as very valuable by the assessed organisation, especially because it creates a common understanding of the strengths and weaknesses of the organisation and explicitly brings all stakeholders together to discuss and deal with these issues. This is a strong argument for continuing the evaluation of the maturity model.

After updating the IDEM3 maturity model based on the feedback of this study, version 1.0 of the maturity model is ready to be distributed to a wider set of users through the Swedish Emergency Management Agency who are the agency responsible for coordinating information safety in Sweden. Future work with the IDEM3 maturity model could focus on evaluating the assessment with the help of the IDEM3 maturity model at other types of organisations. Another way to evaluate the reliability of the assessment would, for example, be to have an independent assessment of different groups of systems at the same organisation. Future evaluations should also focus on the process improvement aspect of the maturity model, following an organisation during the implementation of IT dependability improvements based on the IDEM3 maturity model.

Bibliography


Abstract

Service Level Agreements (SLA) are considered a good practice not only for IT outsourcing but also for IT management within an organisation. In this paper we study the usage of SLAs in municipal IT management. Municipal IT management traditionally involves a large organisation, often with a low IT maturity, but with high requirements on information security. This study shows that SLAs can enable an organisation to improve its IT management, by better defining the services and responsibilities related to IT management. Further, this study identified an important need for more support for municipalities trying to write and manage their SLAs.

In this study, a series of interviews was conducted with IT managers from a number of Swedish municipalities and based on the SLAs collected an SLA template was proposed that focuses specially on information security issues. Finally the template was evaluated in a workshop with practitioners.
1 Introduction

In most large organisations, the responsibility for servicing the IT systems is given to someone else than the users of the system. IT management can either be centralized to a special IT department or outsourced to an IT service supplier. This way IT services can be used more effectively because of rationalizations and since competence can be centralized (Office of Government Commerce 2007, Larson 1998). Because the IT services are provided by someone else than the organization using the systems, there is a need to negotiate the terms for these services. Aspects like which IT services should be provided by the IT organization, their cost, availability and so on need to be explicitly agreed upon. This type of agreement is often documented in a Service Level Agreement (SLA), which can be seen as a written contract between the organization using the IT system and the IT management organization (Kajko-Mattsson and Makridis 2008).

SLAs can be written both between different organizations, i.e. when an organization outsources IT management to an external contractor, and within an organization between the organisation’s own IT department and the rest of the organisation. In the former case, the focus of the SLA is on the contractual elements and this use of SLAs has been discussed in detail by Goo (2010). In the latter case, the main goal of the SLA is to function as a communication tool between the different partners, especially concerning the responsibilities and the scope of the IT services offered by the IT department. This second use of SLAs is the focus of this study.

When an SLA is written this is done in a negotiation between the different parties. In this process, support is needed concerning which aspects of the IT services to include in the SLA. There is a need to define templates for SLAs which can be adapted to different situations (Kajko-Mattsson and Makridis 2008), more specific than the general guidelines on SLAs offered by frameworks such as ITIL (Office of Government Commerce 2007) or COBIT (ISACA 2000). Outsourcing requires a different kind of SLA than for the management of in-house IT services (Goo et al. 2008). Also, the content of the SLA depends on the type of the organisation and the role of IT systems within the organisation. In this paper, an SLA template is proposed especially for municipalities where IT services are provided by an internal IT department. Municipalities, especially in Sweden where this research was conducted, are typically very large organisations with many different types of IT systems. Many of these IT systems are not only critical to the organisation itself, but support critical functions in our modern society such as social services and education. Further municipalities also have an active role in the emergency management procedures during and after a crisis. This poses special requirements on the services provided by a municipality and therefore on the IT systems it needs to provide these services. Therefore the template proposed in this paper focuses especially on business continuity and information security issues.

This paper presents the result of a research initiative where the literature on SLAs was surveyed, and a set of already developed SLAs was investigated in order
to define a first version of an SLA template specially directed at municipalities. The template was further based on an interview study concerning the current SLA practices used by the municipalities in southern Sweden. This template was then evaluated at a focus group meeting with representatives from the IT departments in 6 Swedish municipalities.

Section 2 first discusses some related work, both on the topic of SLAs and on IT management challenges for municipalities. Secondly, an overview of the research steps taken during this study is presented in Section 3 and more details of the data collection are discussed in Section 4. Next, the main contribution of this paper, the proposed SLA template, is presented in Section 5. Section 6 focuses on the details of the evaluation of this template. Finally, the main results and the limitations of this study are discussed in Section 7.

2 Related Work

2.1 Service Level Agreement

Although the usage of service level agreements is a well-established practice in IT management, few scientific articles have been published that study the actual contents of SLAs. Goo (2010) has studied the use of SLAs as part of outsourcing contracts. Their study of several SLAs used in practice also resulted in a proposed structure for service level agreements used in outsourcing.

ITIL (Office of Government Commerce 2007) has been an important factor in the spreading of service level agreements. ITIL defines both service level agreements (SLA) and operational level agreements (OLA). The difference is that an SLA is an agreement between an IT supplier and customer and an OLA is an agreement between different IT units within the same organisation. Although service management is central in ITIL, ITIL does not define explicitly which elements should be included in an SLA or OLA.

One way to make Service Level Agreements more concrete is to integrate them with the Enterprise Architecture as has been proposed by Correia and Abreu (2009). This leads to a more formal description of the SLA, more suited for the technical aspects of IT management. However, this does not cover the organisational issues associated with service level agreements that are the focus of this study.

Service level agreements are not only restricted to IT services but can also be used to specify other types of services. For example, the use of SLAs in hospitals has been investigated in a number of studies (Berbée et al. 2009, Dib et al. 1998). They conclude that the usage of SLAs has a positive influence in an organisation both on the services provided and on the administration of those services.

The practical SLA management process in four companies has been studied by Kajko-Mattsson and Makridis (2008). They also propose a simplified management process based on their findings and the processes defined in ITIL (Office of
Government Commerce 2007) and COBIT (ISACA 2000). However, their framework focuses only on the process of writing the SLA, not on the actual contents of the SLA.

One field of research focuses especially on the process of writing SLAs, for example with the help of the SEAM method from the field of Requirements Engineering (Wegmann et al. 2008).

2.2 IT Dependability Management for Governmental Actors

This study of service level agreements is part of a larger project studying the IT dependability management at governmental actors. Earlier studies (Weyns and Höst 2009) have identified a series of common problems in this field, directly related to communication problems between IT personnel and the rest of the organisation. The IDEM3 maturity model has been proposed as a process improvement framework for IT dependability management (Weyns et al. 2010). One of the attributes of the maturity model is the use of service level agreements.

Li Helgesson (2009) also proposed Service Level Agreements as an important tool in risk management at public service organisation. The template developed in this study also focuses especially on information security and risk management activities.

3 Research process

The SLA template discussed in this paper is the result of a research study conducted in four steps as presented in Figure 1. The goal of this study was both to better understand the way service level agreements are used in practice today at Swedish municipalities and to develop a template to help these municipalities to write and improve their SLAs. To achieve this goal, first a lot of data was collected from different sources. Secondly, the data was analyzed and processed resulting in the first version of the SLA template. This version was then evaluated at a workshop resulting in the final version of the template presented in this paper. These steps are further elaborated upon in the next sections.

4 Data Collection

In a first phase of this study, to better understand the need and the possible contents of an SLA template, data was gathered from 4 different sources: from the research literature on SLAs, through interviews with IT managers of Swedish municipalities, from service level agreements collected from Swedish municipalities and from Service level agreements available online.
4.1 Research literature

As described in Section 2.1, few scientific articles are available on the actual contents that can be found in a service level agreement. Concerning SLAs for outsourcing, Goo (2010) have studied several SLAs and defined a general structure of an SLA. Their study identified 11 contractual elements, divided over three categories: Foundation characteristics, Change management characteristics and Governance characteristics. These categories and subcategories were also used in the analysis of the gathered service level agreements.

4.2 Interviews with IT managers of Swedish municipalities

Over a period of a few weeks, telephone interviews were conducted with IT managers or Chief Information Officers (CIO) from 22 southern Swedish municipalities concerning their experiences with service level agreements. Service quality management is often a responsibility of the CIO (Lindström et al. 2006) and in smaller municipalities SLAs are often managed directly by the IT manager while in larger municipalities this task can be delegated to a specific service manager within the IT unit. Some of these CIOs actually represented multiple municipalities that shared their IT services.

The interviews were conducted over the telephone and lasted between 5 and 20 minutes, depending on the municipalities experience with service level agreements. The interviews first used some open questions about the municipalities experience with SLAs and concluded with some more specific questions such as how often the SLAs were updated. For the preparation of the interview questions and for the actual interviews, the methodology as described by Robson (2002) was used.
Of the 22 interviewed municipalities, 10 were currently using some kind of internal SLA. Of these 10, 6 were currently completely reworking their SLAs because of important changes in their IT management processes. Of the 12 municipalities that were currently not using SLAs in their IT management, 4 were currently in the process of writing SLAs and another 3 municipalities had long term plans to start using SLAs, 1 municipality had completely outsourced their IT management, and 4 had no interest in using SLAs in the near future. In total, this means that there were 10 municipalities actively working with writing their SLAs, making it a very hot issue.

In all municipalities with some experience from SLA, the initiative for introducing SLAs came from the IT unit and not from the users. The most important reasons for using SLAs was to standardize the IT services offered by the IT unit and to clearly define the scope and responsibilities for the IT services. There were large differences in how SLAs were being used, but smaller municipalities tended to have one or a few SLAs for the complete organisation, while larger municipalities often had separate SLAs for every system or service.

Many municipalities had experienced difficulties in writing effective SLAs and actively sought practical guidelines or templates for writing SLAs. When introducing SLAs, most municipalities had relied on some experiences from neighboring municipalities, but there was no organized regional cooperation on service level agreements as there was on some other IT management issues.

4.3 Service level agreements from Swedish municipalities

Of the interviewed municipalities, four contributed their internal service level agreements to this study. The elements in these service level agreements were categorized based on the categories proposed by Goo (2010). Further, all specific metrics and service levels were extracted from each service level agreement.

Of these 4 service level agreements, one was little more than an agreed price list and budget for the IT services provided. The three other documents were true service level agreements. The only elements from the categorization from Goo (2010) that were present in all 3 agreements were the defining the objectives of the SLA (from the point of view of the service provider), process ownership and the definition of explicit service levels. A number of other elements, among which all the elements in the top category of change management and all elements in the subcategory of measurement were not present in any of the 3 service level agreements. The former can be explained by the fact that change management does not need to be as formally defined in internal SLAs as in outsourcing SLAs, where the SLA is part of a legally binding contract between two organisations. However, the fact that none of the SLAs define any measurement goals or processes, show that there is much room for improvement in the use of SLAs in these municipalities.
A number of specific service levels were present in each of the service level agreements concerning:

- the helpdesk, with attributes such as opening hours, and targets for solving times and response times,
- the backup procedures, and
- the process for ordering new systems.

4.4 Service level agreements found online

To further complement the list of possible elements to include in an SLA, an internet search was made for service level agreements or SLA templates. From all the templates available online, 8 were included in this study, selected because they were all quite extensive and they were from a research institution, had a focus on IT management for governmental organisations or had a strong link to ITIL. The selected templates were collected from the following 8 sources:

1. University of California, Santa Cruz, Information Technology Services, Service Level Agreement (SLA) Template, retrieved December 20th, 2010 from http://its.ucsc.edu/about_us/smt/06_07/sla_template_final.pdf


6. Purdue University (USA), Information Technology Service Level Agreement Template, retrieved December 20th, 2010 from http://www.purdue.edu/bscompt/Projects/SLAs/SLADraftTemplate_111901.doc

These templates, between 6 and 42 pages long, were all much more formal and elaborate than the SLAs collected from the municipalities. In this study we do not make any statements about the quality of these individual templates but only use them as a source for extending the list of possible elements to include in our SLA template.

5 Service Level Agreement Template

The data gathered in the first part of the study, clearly showed a need for a template that could help Swedish municipalities to write and to evaluate service level agreements. Therefore the next step in this study was to compile such a template from the material gathered in the previous step.

Before a template could be compiled, a number of decisions needed to be made about the scope and the form of the template.

First of all, it was decided to compile the template specifically for IT management inside an organisation and not for outsourcing. This meant we would not focus on the contractual details of the agreement, but more on the specification of the services itself, with special attention given to aspects related to information security which had already previously been identified as a weak factor in municipal IT management.

Secondly, it was decided to specifically target the SLA-template towards municipal IT management, even if this meant that the template would be less general. The template is likely to also be useful to other governmental actors that are critically dependant on their information security, and in a lesser degree even to commercial organisations, but this is outside the scope of this study. It was preferred to compile a well-targeted version of the template that filled a concrete need for a small set of organisations than a watered-down version that was only of little use to a large set of organisations.

Finally, some of the studied service level agreements were specific for one IT system or service, others were specific for one administrative part of an organisation and a third category of SLA included the complete IT services of the organisation in one agreement. These differences in the scope of an SLA of course have implications for the elements that need to be included in the agreement.

Especially important for the specification of the SLA is the difference between IT systems that are common to the whole organisation and systems that are specific to one organisational unit. Email systems are a typical example of the former category of systems, which are typically owned by the IT unit of the organisation and offered as a service to the rest of the organisation. The latter category of
systems are typically owned by a different organisational unit and only serviced by the IT unit.

The template needed to be general enough to be applicable in all these situations, therefore it was decided to include even aspects that are important only in one of the cases mentioned above, but to explicitly mention that not all elements in the template are specifically needed in each situation. Instead the template is meant to serve as a list of suggestion of parts that can be included in a service level agreement not a minimum list of compulsory items that always need to be included.

For this process, all elements in the collected service level agreements, were labeled and a complete list of all components, divided over a number of categories was compiled. First the SLAs collected from the municipalities were analyzed and categorized, and afterwards the list was extended with elements from the templates found online. This labeling was based on the SLA-structure proposed by Goo (2010). Except for these general structural elements, also specific service level descriptions were labeled and extracted.

In a next step, all the categories relevant for municipal IT management were selected from the list and ordered in they would most logically appear in a Service Level Agreement to form the basic structure of the document. All elements found at least one of the municipal SLAs were included, complemented with the most relevant additional elements found in the online templates. This resulted in the following outline for the proposed template:

1. Objectives
2. Scope
3. Service levels
   - IT support processes
   - Backup processes
4. Risk and vulnerability analysis
5. Incident management processes
6. Innovation processes
7. References to related documents
8. Follow-up processes
9. Practical matters
10. Appendices
    - Word list
List of systems

This is substantially different from the structure proposed by Goo (2010) in a number of ways. First of all this template has a more limited area of applicability and can therefore be much more concrete. Secondly, all change management characteristics and governance characteristics that were more important for outsourcing contracts have in our template been compressed into the two short final categories while the foundation characteristics have been made more explicit.

Each of these elements of the template is further explained in the next sections. Together these subsections, in the order they are listed, constitute the proposed SLA template.

5.1 Objectives

The first part of an SLA should state what the organisation wants to achieve with writing the SLA. The objective should clearly represent both the viewpoint of the service provider and the service receiver. The objective should be based on common goals within the organisation. Objectives can include:

- Defining different service levels
- Scoping IT services provided by the supplier to the customer
- Communication about actual service levels expected for different services
- Listing responsibilities of both partners on IT management issues
- Defining system and process ownership
- Defining a common vocabulary
- Achieving uniformity across the organisation concerning IT management
- Providing contact information
- Budgeting arrangements and price agreements
- Conflict resolution and prevention

5.2 Scope

Scoping the IT services provided is one of the main functions of the service level agreement. Therefore the second part of the template consists of a list of what is and is not included in the agreement, with an explicit list of responsibilities for both partners. This serves as a summary of the main elements of the SLA. All the important activities and processes in the following sections of the document should be listed as responsibilities of the service provider and/or customer, with explicit references to relevant sections below.
5.3 Service levels

This is the main section of the SLA where the provided service levels are defined. For an organisation such as a municipality with a critical role in society and an active role in crisis relief, there is an obvious need for some IT support to be offered at all times, even during nighttime or holidays. At the same time there are also many systems where such a high level of support is not necessary. Therefore the template proposes to define at least two levels of service, one basic level for all but the most critical systems and one higher level of service for the most critical systems. An explicit list of systems which fall in this critical category should be attached to the SLA. For some systems that are only critical in some specific situations, exceptions to the basic rule can also be defined (e.g. the wage payment system around each monthly payday). For most systems the service levels can probably be limited to availability requirements, support processes and backup services. These are discussed below. For more critical systems also performance requirements might need to be defined (e.g. for the telephone system, the central network or the public website of the municipality). For systems that contain highly sensitive data it might also be useful to define some extra security requirements.

Availability requirements

An availability goal should be set for all basic systems, for example 99%, which means a total downtime of less than 90 hours on a year. This should take into account regular planned service interruptions for maintenance and updates. An important factor influencing availability can be the access to a backup power supply. It should be specifically stated which systems have access to backup power in case of a blackout, for example all critical systems as defined before.

IT support processes

The technical support provided to the users in case of problems is a subject found in nearly all SLAs. There are a lot of service levels associated with this service that need to be explicitly defined in a good SLA:

- opening hours of support
- average or maximum response time
- average solving time
- 'customer satisfaction index': percentage of support requests successfully solved
- status reporting: how the user is informed of the status and completion of a service request
• regular access to IT technician on site: in some organizations each business unit has regular access to an IT technician at an agreed time each week or month to discuss minor problems

• participation of customer in solving high-priority problems

• prioritization of service requests

• availability of user training

**Backup processes**

Another important service that should be defined in detail concerns the backup routines. Backup solutions can often be quite technical, but it is important for users to be aware of the possibilities and limitations of restoring systems after a crash. The main attribute to be defined for backup is the frequency, as it guarantees the maximum amount of time for which information can be lost. On a more technical level, both a frequency for incremental and complete backup of all systems can be defined. Often a daily backup routine is used. For the most critical systems a transaction-based backup can also be defined. Another important service level for a backup service that is often overlooked is how long it takes to restore a system from the backup. Also this should be explicitly agreed upon in a service level agreement. Backup solutions should also regularly be tested to investigate whether systems can reliably be restored.

5.4 **Risk and vulnerability analysis**

All system owners for critical systems should be responsible for conducting a risk and vulnerability analysis of these systems. IT personnel should actively take part in these analyses. Preferably a standardized risk analysis method should be used.

5.5 **Incident management processes**

Incident management can be considered a part of normal IT support, but the management of the most critical IT incidents is important enough to be discussed in a separate section. For all incidents with critical systems, a formally defined incident management process should be followed. This should not only cover the immediate response to the incident but also procedures for following up that lessons are learned from the incident to prevent similar incidents from occurring in the future, even for other systems.

For large IT-incidents a response group should be assembled and the necessary personnel should be available to resolve the problem as quickly as possible and to provide the users with information about the problem and the progress in resolving the problem. Even for emergency situations that are not directly IT-related, the organisation might need to declare a state of emergency and setup a crisis central
in which communication systems are likely to play an important role. Access to IT support for these systems should be included in the crisis plans.

A more detailed study of the requirements for an incident management process compliant with the ITIL framework, has been conducted by Jantti (2009).

### 5.6 Innovation processes

As innovation is often very fast in the field of information technology, the SLA should also define the responsibilities of the SP and SR concerning the introduction of new systems or the update of older systems. This section should describe the processes for ordering new software and hardware products and also for updating existing systems. This can even include special procedures for large IT innovation projects. For the most critical systems, a testing environment should be set up where new updates can be tested before they are deployed to all users.

### 5.7 References to related documents

The SLA is just one of the documents describing details concerning IT management within the organisation. Many organisations also have an IT safety or security policy, an IT handbook, IT user guidelines and technical system documentation. The SLA can be a good place to define who is responsible for administrating all these other documents and where the latest version of these documents can be found. Other documents related to the SLA for which the responsibilities should be described in the SLA include contracts with external suppliers, national or international standards the organisation must follow and national information management laws concerning secrecy and public information.

### 5.8 Follow-up processes

It is important to make sure that the service level goals described in the SLA are achieved in practice. This can only be known if the goals are written in a measurable way and if the service levels are actually being measured. Defining a measurement plan as part of the SLA is a good way to assure that the service levels are written in a way that they can be measured and not on a too abstract level. Measuring the most important service level outcomes can help the organisation to discover problems as early as possible. The measurement plan should also include a description of how the collected data will be analyzed and reported to all involved parties.

Because of the fast evolution in IT the SLA will probably need to be updated on a regular basis. This section of the SLA should also describe how often the SLA will be reviewed and changes will be discussed. The data collected through the measurement process can also give important information about how the SLA needs to be updated. The SLA should also explicitly define the responsibilities
for spreading information about the SLA and about any later changes in the SLA. The latest version should always be available to the users and to the IT-personnel. System owners should make sure that the users are aware of the existence and the contents of the most important points in the SLA, just as the CIO should be responsible for spreading information about the SLA to all personnel of the IT-unit.

5.9 Financial Aspects

Depending on the internal structure of the organisation, a number of practical, financial matters might need to be defined explicitly in a contract between the service provider and the users. In many SLAs these have become the main focus of the SLA, but in a complete SLA the focus should instead be on the level of service, and these financial aspects should only be a part of the contract. These financial aspects can include pricing agreements, licensing issues, insurance aspects and possible compensation for breaching the service levels promised in the SLA. Internally in an organisation, financial compensation for not fulfilling part of the SLA is often not an effective way to deal with this kind of problems. However, it can be useful to require that, as a form of compensation, an investigation into the problem is conducted and reported upon, in a similar way as after critical incidents.

5.10 Appendices

The following appendices should preferably be added to the SLA:

- Glossary
- List of explicit systems ownership for all IT systems
- Contact list for all issues relating to the SLA, on both the service provider’s and the customer’s side
- Classification of all critical systems for Confidentiality, Integrity and Availability based on MSB’s classification system (Swedish Civil Contingencies Agency 2009).
- List of known critical dependencies between systems

6 Evaluation

To evaluate the service level agreement template presented above, we organized a half-day workshop for CIOs and service managers of regional municipalities where the template would be presented and the municipalities could discuss their experience with SLA together. Nine IT-professionals from 6 different municipalities participated in the workshop, 3 more municipalities had to cancel their
participation in the last minute and another 4 CIOs expressed their explicit interest in the material of the workshop but could not participate in the workshop because it collided with other appointments in their calendar.

On the workshop, the results of the background study about SLA were presented, each of the sections of the template was discussed and the municipalities shared their experiences about SLA. There was strong agreement about the usefulness of the template and the need for such a template as all the participating municipalities were currently working with service level agreements.

At the end of the workshop, each of the participants individually filled in a survey about the workshop and the SLA template. On the questions whether they saw a general need for this type of SLA template and whether they would use this particular template in their work, all 9 participants answered strongly positive.

Participants were further asked as which they considered to be the most important parts or functions of a service level agreement. Two elements were present in nearly all respondents’ answers: defining of responsibilities and scoping of the IT services delivered.

The central part of the survey listed all the elements in the SLA template and asked the participants to mark whether they agreed that they should be part of a municipal SLA, on a five-point Likert scale. The responses are summarized in Figure 2. The main conclusion from this data is that there was positive support for each of the listed elements as part of an SLA, with strongest support for Objectives, List of responsibilities, Availability requirements and Backup routines and only marginal support for Insurances and Information security policies.

Finally the respondents were also asked to comment on the usefulness of the complete workshop. There was strong support for the value of the workshop for each of the participants. Most participants listed the possibility to share experiences between the municipalities as the most important aspect of the workshop, together with the actual template presented at the workshop.

7 Discussion

The results of the evaluation of the SLA template in the previous section clearly show that the template in its current form fills a need for many Swedish municipalities. Based on the evaluation of the different parts of the template, no parts of the template were removed. Some minor changes to the formulation of the descriptions of the different parts were made, based on feedback received on the focus group meeting. Also, two small elements were added to the template based on the discussions during the workshop: the appendix with critical dependencies between systems and the requirement of a report as compensation for failing to deliver the promised service levels. For simplicity, they were immediately included in the template presented in Section 5 in this paper but were not included in the first version of the template that was presented at the workshop.
Figure 2: Responses from the survey on the question whether each element should be part of a municipal SLA, on a 5-point Likert scale.
Concerning the validity of the results of this study, a number of issues need to be discussed. First of all, the template itself is specifically directed at Swedish municipalities that have a centralized IT unit inside the organisation. It is outside the scope of this study to investigate the usefulness of the proposed template for other types of organisations. It can be assumed that the template in its current form is likely to also be directly useful for other Swedish governmental actors that are confronted with similar problems concerning IT management and that have a critical role in society. For the template to be equally useful for commercial enterprises the focus would probably need to be shifted to include more business alignment issues and commercial aspects that are less relevant to governmental actors.

Swedish municipalities have a large set of responsibilities that also includes an explicitly defined active role in emergency management. Traditionally they are large organisations with only a small IT unit and low IT maturity. The relevance of the presented template for municipalities and other governmental actors in other countries depends mostly on their organisational structure and their responsibilities.

The template is specifically meant for organisations with an in-house IT service provider. For municipalities or other organisations that outsource a large part of their IT management, service level agreements need to focus much more on contractual issues such as pricing and compensation which are only a minor part of the template presented in this paper. For a more extensive overview of the use of SLAs in outsourcing we refer to the work by Goo et al. (2008) and Goo (2010).

Although the template is specifically directed at municipalities, there are still some ways in which this study is relevant for other organisations. The methodology of this study and the form of the final template should be applicable to other types of organisations where a similar need for a clear SLA template exists. The scarcity of the available published studies on the use of SLAs in different types of organisations shows a need for more research in this area. Further, the issues listed in this SLA template can definitely form a basis for other specific SLA templates, especially in those areas where information security is critical.

A second issue concerning validity concerns the selection of the municipalities involved in the study. Only municipalities in one Swedish region were involved in the interview study, and only few municipalities participated in the focus group. Because the responsibilities of municipalities are the same across Sweden, the results of the study would likely be very similar in case a different region had been chosen. The selection of the municipalities for the focus group was however far from random. Although all municipalities in the most southern region of Sweden were invited to the workshop, only municipalities with at least a general interest in service level agreements and the template are likely to have participated in the focus group. Therefore the positive evaluation of the template cannot be generalized to all municipalities. As the first set of interviews also showed, there is also a substantial group of municipalities with little interest in service level agree-
ments. For them the template is probably of little value except for that it lists a number of issues that should be dealt with in IT management. Nevertheless, the majority of the municipalities contacted in the study showed strong support for the template.

A threat to the validity that should be considered when using a focus group, is that one or a few participants strongly dominate the discussion. To mitigate this risk, all the participants of the focus group individually filled in the survey described in Section 6 and the results of this survey are used as the main data to evaluate the template. Further, during the focus group, attention was given to make sure that all participants contributed to the discussion and to minimise the effect of the researchers present on the conclusions of the discussion.

A final threat to the validity of the results of this paper concerns the completeness of the proposed template. Although the template has been evaluated in a focus group meeting, it is unlikely that all possible additions and improvement to the template were identified during this evaluation. Therefore the proposed template is not considered a one-time publication of a template, but rather as the first version of a living document that can be appended and updated in future revision. Further, the continuous evolution of the field of IT management and is likely to add new requirements for this template. To facilitate the further updating of the proposed template, the template will be published by the Swedish Civil Contingencies Agency (MSB) in a community cooperating on the standardization of information security practices for Swedish municipalities.

8 Conclusions and Future Work

The main result of this study is a proposed service level agreement template for the use in municipal IT management presented in Section 5 of this paper. The template was based on an interview study with municipalities from the south of Sweden, an overview of the SLA research literature and example SLAs and SLA templates collected both from municipalities and different online sources. The template was then evaluated in a focus group meeting with Swedish municipalities. The template focuses especially on information security issues.

The evaluation of this template by the municipalities involved in the study was strongly positive and the study identified a clear need for such a template, with many of the interviewed municipalities currently in the process of writing or reorganizing their SLAs. The template will now be made available for other Swedish municipalities and governmental organisations with the help of the Swedish Civil Contingencies Agency.

Concerning future work, this study also identified a lack of empirical studies on the use of service level agreements in different kinds of organisations. Many other types of organisations are likely to benefit from a similar type of SLA template specifically focused on their IT management needs. Further future work
could also include case studies following municipalities using this template to introduce SLAs into their IT management to evaluate the practical applicability of the template and to further study the process of introducing SLAs in a practical setting.

Bibliography


