Towards a Framework for IT Service Fault Management

Andreas Hanemann*, Martin Sailer†, and David Schmitz*

*Munich Network Management Team, Leibniz Supercomputing Center, Germany
{hanemann,schmitz}@lrz.de

†Munich Network Management Team, University of Munich (LMU), Germany
sailer@informatik.uni-muenchen.de

Abstract

IT service providers – either research institutions or commercial providers – offer IT services based on the behaviour of resources like the network infrastructure. To satisfy the customer demand for high service quality, which is specified in terms of Quality of Service parameters, the process of diagnosing and resolving resource failures affecting the service quality is a critical issue. Problems with the service quality in this sense differ from problems in the area of network and systems management in that a service may be available, but performing poorly.

In this paper we are addressing a framework to effectively deal with this issue. It consists of components and interactions between them which are required to perform the service fault management. A real-world scenario is used to derive the requirements which have been applied to the component identification. An analysis of the state-of-the-art highlighting the contributions and deficits of existing frameworks and approaches wit respect to their applicability to the framework is also carried out..

Keywords: service management, fault management

1 Introduction

For an IT service provider like a university computing centre it has become a necessity to ensure that its service provisioning is done in accordance to user's expectations. In a commercial IT service provisioning environment these expectations are specified as parameters for indicating the Quality of Service (QoS parameters). Therefore, Service Level Agreements (SLAs) are laid down which contain thresholds for these parameters and penalties for not meeting them. Even though there are in most cases no such SLAs in place in a research environment, the ensuring of the service quality is expected to become more important to justify the IT service funding.

Due to the different nature of problems on the service level in comparison to the resource level we employ the term service quality degradations instead of service faults. Consider an example where the download from a web site is quite slow. As the service is available, this situation cannot be denoted as a fault, but should be specified as quality degradation if a guarantee for the download speed has been provided.

To improve the overall service quality the management of IT service quality degradations is a critical issue. It can be divided into three phases:

Quality degradation detection: In this phase a degradation of the service quality is detected which can either result from a customer report or be detected by the provider's own service monitoring.

Quality degradation diagnosis: Resources which are used for the service provisioning are identified as being responsible for the problems during this phase. Information about the service provisioning, especially different kinds of dependencies are relevant here.

Quality degradation resolution: An appropriate way to deal with the resource problems has to be found. To decide about adequate resolution measures the impact of the resource failures onto customers should be taken into account.

In this paper a *service* is defined as a set of functionalities that are offered by a *provider* to a *customer* with a guaranteed *Quality of Service*. The service can be accessed at a *service access point* and the agreed QoS is specified in a *Service Level Agreement*. The services are provided using other services called *subservices* and *resources* (e.g., network links, network components, end system memory, or end system processes).

The rest of the paper is organised as follows. In Section 2 a large-scale e-mail service provisioning scenario is used to analyse the current situation of IT service fault management and to derive requirements for an appropriate framework. A review of related work is carried out in Section 3 revealing the contributions and deficits of the state-of-the-art in this area. Our framework proposal is presented in Section 4, while conclusion and future work can be found in the last section.

2 E-Mail Service Scenario

The Leibniz Supercomputing Centre (LRZ) is the joint computing centre for the Munich universities and research institutions. It also runs the Munich Scientific Network and offers related services. One of the main services is the E-Mail Service which is available to students and staff of the Munich universities and the LRZ itself.

Even though the service can currently be regarded as best effort service (i.e., without explicit quality guarantees), its proper operation is highly critical due to the amount of users. Therefore, critical situations have to be recognised as quickly as possible to avoid, for example, long waiting queues in the mail processing.

The E-Mail Service is dependent on other services like DNS and the basic connectivity service. Its resources include servers for sending and receiving mail (different identity management solutions at the institutions added some heterogeneity here) and their interconnections.

The fault management for this service is currently performed as follows. A user who experiences a problem with her e-mail account can either contact the LRZ help desk directly or can use the web-based problem preclassification tool *Intelligent Assistant* [DrKa98]. This tool guides the user to traverse a query tree composed of questions (e.g., how the user accesses the service) and tests (e.g., component ping tests) to gain a problem preclassification and in some cases already a solution. The result of this preclassification is forwarded to the LRZ help desk.

Sometimes the problem can already be resolved at the help desk if the customer has made a mistake in the service usage or if the problem is already known and its resolution is under way. Otherwise, a trouble ticket (Remedy ARS Trouble Ticket System) is opened to delegate the problem to other employees responsible for the service. These employees can access management tools like HP OpenView (where an event correlation is performed for using network topology information), IBM Tivoli, and Infovista or examine log files to find the error. The root cause of the problem is reported to the help desk via the trouble ticket system and the user is informed about the service status. The fault recovery decisions are mainly based on the experience of the employees.

In sum, it can be concluded that the service fault management is only partially suitable for an assured timely fault resolution. As a consequence, the steps mandatory for the fault management need to be examined according to the possibility for automation or semi-automation. Therefore, the fault processing steps need to be formalised and a framework to put these steps into effect is required. The following issues need to be addressed by the framework:

1. Which are the components needed for service fault management? How should they interact?
2. How to standardise the fault management information (e.g., service models, resource models, problem reports) in order to be independent from the experience of single employees?
3. Which possibilities exist for optimisation (i.e., automation, semi-automation) in each framework component/fault management step?

3 Related Work

To our knowledge, no standards or approaches can be found in the literature that address a framework completely suitable to the requirements. Nevertheless, approaches and solutions can be identified for parts of the framework which are subject to this section and are arranged in subsections according to the requirements. At first, IT process management frameworks are presented which are related to the general framework design. Related work for the information modelling, i.e. service models and resource models with focus on dependency modelling is referenced afterwards. Finally, partial solutions that could be applied for the framework are reviewed.

3.1 IT Process Management Frameworks

The IT Infrastructure Library (ITIL) [ITIL] is a continuously evolving collection of best practice documents with regard to the service management of an IT service provider. It defines (among others) the process sets of service support and service delivery. Process descriptions are derived from expert knowledge in a particular field and written more or less in prose, so that ITIL can be regarded to be a bottom-up approach. As a consequence, the workflow modelling for our framework cannot be directly derived from these process descriptions.

The enhanced Telecom Operations Map (eTOM) [eTOM] is a business process framework for the telecommunications industry. eTOM is customer-centric and covers a broad range of important processes including processes for strategy, infrastructure, product, and operations. The Service Problem Management (SM&O-A) process deals with the diagnosis and resolution of service problems including the assessment of the impact on customers. However, this process is not described in a formal way, and neither input and output parameters, nor the linking of processes are described explicitly.

3.2 Information Modelling

In order to establish a common understanding of the term *service* a generic service management model (MNM Service Model) has been proposed by our research group [GHHK01]. An important feature of the model is that service management is an integrated part of it. Similar to the service access point a reference point is defined for the exchange of management information between customer and provider (e.g., for ordering of new services, provisioning of service quality reports, information about service failures). It is called Customer Service Management (see section 3.3) access point.

A detailed presentation and classification of QoS approaches can be found in [GaRo04]. The QoS definition and measurement methodology proposed by Garschhammer [Gars04] addresses the issue of implementation independent QoS specification. Therefore, the measurement operations are performed at the service access point independent of the inner structure of the service provisioning.

The Common Information Model (CIM, [CIM]) introduces a management information model that aims at integrating information models of existing management architectures. CIM acts as an umbrella that allows exchanging management information in an unrestricted and loss-free way. This umbrella architecture has been chosen to achieve vendor independence. As a consequence, CIM can be used for the resource modelling.

In the literature the term *dependency* is often used without an explicit definition. In [BKH01] the notion of absent, weak, medium, and strong dependencies is introduced, but no methodology how to assign such values is given.

Caswell and Ramanathan [CaRa99] describe dependencies for services offered by Internet Service Providers. They distinguish between five kinds of dependencies. An execution dependency denotes the performance of an application server process with respect to the status of the host, while a link dependency specifies the service performance with respect to the link status. In case of a web service that is provided on different front-end servers which are selected by a round-robin DNS scheduling the performance depends on the currently selected server (component dependency). An inter-service dependency occurs between services, e.g. an e-mail service depends on an authentication service and on an NFS service. Services and/or server belong to different domains of responsibility which is denoted as an organisational dependency.

While our starting point in this paper is that the dependencies are given, the issue of finding them has also to be addressed. This knowledge is gathered from experts or configuration databases or log files, etc. As changes in the service provisioning are quite frequent, approaches to automatically detect dependencies have been proposed. Examples include the use of neural networks [Ense01] and the analysis of temporal relationships of interactions [GNAK03].

3.3 Related Work for Framework Components

In the area of network and systems management event correlation techniques have proven to be useful for root cause analysis. These techniques are of interest for service quality degradation diagnosis.

Rule-based reasoning: In rule-based reasoning (RBR, [JaWe93,Lewi99] a set of rules is used to perform the event correlation. The rules have to form "*conclusion* if *condition*". The condition contains received events together with information about the state of the system, while the conclusion may consist of actions which lead to changes of the system and can be input to other rules.

The rules in an RBR system are more or less human readable, so that their effect is supposed to be intuitive. Fast algorithms like the RETE algorithm exist to actually perform the correlation. In practice, the rule sets may become quite large which may lead to unintended rule interactions and makes it difficult to maintain the system. In addition, the system is going to fail if an unknown situation occurs which has not been covered by rules yet.

Codebook approach: The codebook approach [YKMY96] uses experience from graphs and coding. The input of this technique is a dependency graph consisting of events and root causes as nodes and directed edges to represent the dependencies. After a graph optimisation has been performed, the graph is transformed into a correlation matrix. The columns in the matrix represent the root causes, while the rows represent the events. In its simplest form the matrix cell entries can either be 1 or 0, denoting the presence or absence of a relationship between event and root cause. Values between 0 and 1 may be used to indicate the strength or likelihood of the dependencies. Techniques from coding theory can be applied for optimisation. For example, some event rows may be deleted if the events do not lead to a root cause discrimination.

This approach has the advantage that it can - in some situations - deal with unknown combinations of events. These can be mapped onto known combinations by using the Hamming distance. Efficient algorithms exist for the codebook-based correlation.

Model-based reasoning: In model-based reasoning (MBR, [JaWe93,Lewi99]) each component of an infrastructure is modelled with respect to its attributes, behaviour, and relation to other models. The behaviour of the whole infrastructure arises from the interaction of the component models. A model can either be a representation of a physical entity or a logical entity. The event correlation itself is a result of the collaboration of models.

This approach does not propose a detailed technique to correlate the events. Therefore, real-world systems often use rules to actually perform the correlation.

Case-based reasoning: In contrast to the techniques presented before the case-based reasoning approach (CBR, [Lewi93,Lewi99]) needs no prior knowledge about the actual configuration. It contains a database of cases which have previously occurred together with the identified root causes. While the first root causes have to be identified by hand, a matching to prior cases is performed at later stages; i.e. the ability to learn is a main feature of this approach.

Service Level Monitoring approaches and tools are used to monitor whether an SLA is met, but they do not deal with the treatment of faults. In the SoLOMon framework [FJP99] a language is defined to specify metrics in an expressive way. The metrics are therefore user-oriented and independent of a specific application. A run-time system was implemented for those metrics which especially aims at achieving scalability. Several commercial tools like Infovista are not only able to monitor the network and systems performance, but can also be used to monitor the service performance.

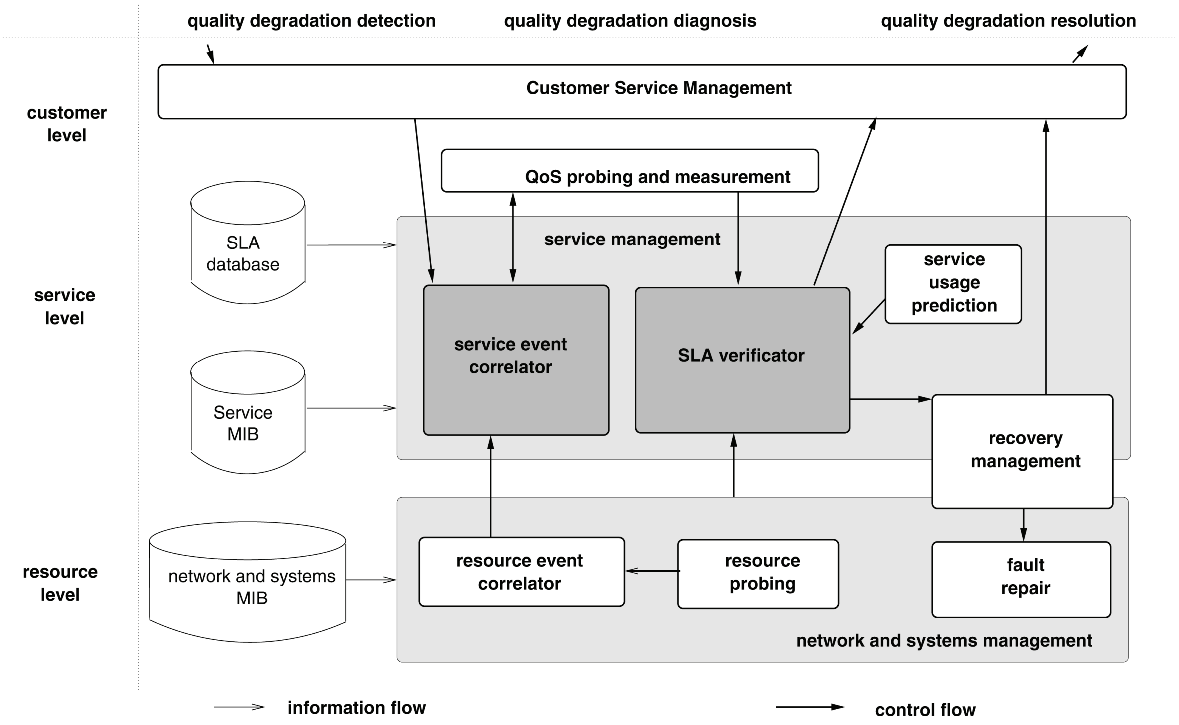

Figure 1: Service Fault Management Framework

An overview of the issues related to Service Level Management can be found in [Lewi99]. Problems arising when dealing with SLAs across domain borders are addressed in [BCS99]. This publication also contains a language to define SLAs.

Salle and Bartolini [SaBa04] approached the management of SLAs from a business perspective (called "Management by Contract"). While other approaches have in focus how to meet service level agreements, the possibility to break an SLA voluntarily is seen as a viable option. A modelling of SLAs and an algorithm to decide which effort should be applied to meet an endangered agreement is presented. A formalisation of the cost of violating the agreement is needed as input. While this approach seems to be appropriate for the modelling of SLAs and to select one of different recovery measures, other issues have not been addressed in a general manner so far. These issues are e.g., the modelling of services and resources, the way the impact analysis of a resource failure should be performed, methods to find possible recovery measures, and how to define costs for not meeting SLAs.

4 Framework Components

In this section an identification of framework components (see Figure 1) is performed which are relevant to the fault management for IT services. They are grouped according to the fault management phases [HAN99].

4.1 Components for Quality Degradation Detection

In the quality degradation detection a distinction is made between the problem detection by customers and by the provider himself.

The *Customer Service Management* is used an interface to the customers. It receives customer reports about service quality degradations and informs the customer about the current service status and recovery measures. An *Intelligent Assistant* is attached to the CSM in order to convert customer reports into formalised events.

A component called *QoS probing and measurement* is needed for the provider to monitor the quality of his own services. This is useful to detect problems prior to the customers and therefore to gain some time to solve the problem. This component treats the service as a black box, i.e. it does not need to have knowledge about the service realisation and performs user interactions for testing. The tests can either be performed on a regular basis or on demand. The component design can be derived from the work in [Gars04] (see related work)

4.2 Components for Quality Degradation Diagnosis

Since association information regarding services and resources is a prerequisite for the fault management process, a component named *ServiceMIB* is required, which acts as a container for service management information. It abstracts a service into an object-oriented entity similar to existing managed objects and provides detailed information about dependencies among services as well as dependencies from services onto resources.

In [HSS04] we proposed the use of event correlation techniques to deal with the customer reports. For doing so, events called *service events* are generated from the customer reports and from the problem detection by the provider's own service monitoring. Rules to correlate the events are derived from the Service MIB. A detailed architecture of this *service event correlator* component (a combination of rule-based and case-based event correlation techniques) is subject to [HaSa05]. The service event correlator is part of the *service management* component. The output of the event correlation is a candidate list of resources which could be the problems root cause.

A *network and systems management* component like HP OpenView or IBM Tivoli has to be in place for the resource management. It contains a *monitoring and probing* component to get information about the infrastructure and uses information about the configuration stored in the *network and systems Management Information Base (MIB)*. This component also has to comprise a *resource event correlator* for dealing with events on the resource level which uses one of the methods presented in the related work. Correlated events are transferred to the event correlator on the service level and are matched to the service events. The candidate list of possible root causes is checked by this component or operation staff.

4.3 Components for Quality Degradation Resolution

To decide about appropriate resolution measures an automated impact analysis [HSS05] is performed by the following components.

The starting point for the impact analysis is a resource failure. It is transferred to the service management where the dependencies contained in the Service MIB are used to identify services which are affected directly (by using the resource) or indirectly (when using another affected service) by the fault. An *SLA verificator component* is used to determine the effect onto agreed service guarantees. As this effect may also depend on the service usage, an additional component - the *usage monitoring* - has to be in place to monitor and forecast the actual service usage.

The impact of the resource failure is forwarded to the recovery management where the consequences of recovery alternatives are determined. The operation staff will then decide by using this indication how to react adequately. For this step the Management by Contract approach can be adapted.

During the fault resolution the Customer Service Management is used to provide the current repair status of the service and to report the problem resolution together with its consequences onto service level agreements to the customers.

5 Conclusion and Future Work

In this paper we have shown the new challenges that arise in the area of IT service fault management which have been motivated by a real-world scenario. Components have been identified which are needed to address these issues and have been put into a service fault management framework. An implementation for large-scale services provided by the Leibniz Supercomputing Centre is currently carried out. In doing so, a quantification of the framework's benefits will be addressed.

Acknowledgements

The authors wish to thank the members of the MNM Team for helpful discussions and valuable comments on previous versions of the paper. The MNM Team, directed by Prof. Dr. Heinz-Gerd Hegering, is a group of researchers of the Munich Universities, the University of Federal Armed Forces in Munich and the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences. Its webserver is located at *http://wwwmnmteam.informatik.uni-muenchen.de*.

References

[BKH01] S. Bagchi, G. Kar, and J. Hellerstein, "Dependency Analysis in Distributed System using Fault Injections: Application to Problem Determination in an E-Commerce Environment", Proceedings of the 12th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2001), Nancy, France, October, 2001.

[BCS99] P. Bhoj, S. Chutani, S. Singhal, "SLA Management in Federated Environments", Proceedings of the 6th IFIP/IEEE International Symposium on Integrated Network Management, Boston, Mass., USA, May, 1999.

[CaRa99] D. Caswell and S. Ramanathan, "Using Service Models for Management of Internet Services", HP Technical report HPL-1999-43, HP laboratories, 1999.

[CIM] Common Information Model, Distributed Management Task Force, www.dmtf.org/standards/cim

[DrKa98] G. Dreo Rodosek and T. Kaiser, "Intelligent Assistant: User Guided Fault Localization", Proceedings of the 9th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 1998), Newark, Delaware, USA, October, 1998.

[Ense01] C. Ensel, "New Approach for Automated Generation of Service Dependency Models", Proceedings of the 2nd Latin American Network Operations and Management Symposium (LANOMS 01), Belo Horizonte, Brazil, August, 2001.

[eTOM] Enhanced Telecom Operations Map. TeleManagement Forum. www.tmforum.org

[FJP99] S. Frolund, M. Jain, and J. Pruyne, "SoLOMon: Monitoring End-User Service Levels", Proceedings of the 6th IFIP/IEEE International Symposium on Integrated Network Management, Boston, Mass., USA, May, 1999.

[GaRo04] M. Garschhammer and H. Rölle, "Requirements on Quality Specification Posed by Service Orientation", Proceedings of the 15th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2004), Davis, California, USA, November 2004.

[Gars04] M. Garschhammer, "Dienstgütebehandlung im Dienstlebenszyklus: von der formalen Spezifikation zur rechnergestützen Umsetzung – in German", PhD thesis, University of Munich (LMU), 2004.

[GHHK01] M. Garschhammer, R. Hauck, H.-G. Hegering, B. Kempter, M. Langer, M. Nerb, I. Radisic, H. Rölle, and H. Schmidt, "Towards generic Service Management Concepts – A Service Model Based Approach", Proceedings of the 7th IFIP/IEEE International Symposium on Integrated Network Management, Seattle, Washington, USA, May, 2001.

[GNAK03] M. Gupta, A. Neogi, M. Agarwal, and G. Kar, "Discovering Dynamic Dependencies in Enterprise Environment for Problem Determination", Proceedings of the 14th IFIP/IEEE Workshop on Distributed Systems: Operations and Management, Heidelberg, Germany, October, 2003.

[HAN99] H.-G. Hegering, S. Abeck, and B. Neumair, "Integrated Management of Networked Systems – Concepts, Architectures and their Operational Application", Morgan Kaufmann Publishers, 1999.

[HaSa05] A. Hanemann and M. Sailer, "Towards a Framework for Service-Oriented Event Correlation", Proceedings of the International Conference on Service Assurance with Partial and Intermittent Resources (SAPIR 2005), Lisbon, Portugal, July, 2005.

[HSS04] A. Hanemann, M. Sailer, and D. Schmitz, "Improved Service Quality by Improved Fault Management – Service-Oriented Event Correlation", Proceedings of the 2nd International Conference on Service-Oriented Computing (ICSOC 2004), ACM, New York City, New York, USA, November, 2004.

[HSS05] A. Hanemann, M. Sailer, and D. Schmitz, "A Framework for Failure Impact Analysis and Recovery with Respect to Service Level Agreements", Proceedings of the IEEE International Conference on Services Computing (SCC05), Orlando, Florida, USA, July, 2005.

[ITIL] IT Infrastructure Library, Office of Government Commerce (UK). www.itil.co.uk / www.itsmf.com

[JaWe93] G. Jakobson and M. Weissman, "Alarm Correlation", IEEE Network, 7(6), November, 1993.

[JaWe95] G. Jakobson and M. Weissman, "Real-time Telecommunication Network Management: Extending Event Correlation with Temporal Constraints", Proceedings of the 4th IFIP/IEEE International Symposium on Integrated Network Management, Santa Barbara, California, USA, May, 1995.

[Lewi93] L. Lewis, "A Case-based Reasoning Approach for the Resolution of Faults in Communication Networks", Proceedings of the 3rd IFIP/IEEE Symposium on Integrated Network Management, San Francisco, California, USA, April, 1993.

[Lewi99] L. Lewis, "Service Level Management for Enterprise Networks", Artech House, 1999.

[LLN98] M. Langer, S. Loidl, and M. Nerb, "Customer Service Management: A More Transparent View to Your Subscribed Services", Proceedings of the 9th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 1998), Newark, Delaware, USA, October, 1998.

[SaBa04] M. Salle and C. Bartolini, "Management by Contract, Proceedings of the 9th IEEE/IFIP International Network Operations and Management Symposium, pp. 787-800, Seoul, Korea, April 2004.

[YKMY96] S. Yemini, S. Kliger, E. Mozes, Y. Yemini, and D. Ohsie, "High Speed and Robust Event Correlation", IEEE Communications Magazine, 34(5), May, 1996.