Business Driven Prioritization of Service Incidents

Claudio Bartolini, Mathias Sallé

HP Laboratories
1501 Page Mill Rd
Palo Alto, CA 94304
USA

{claudio.bartolini, mathias.salle}@hp.com

Abstract. As a result of its increasing role in the enterprise, the Information Technology (IT) function is changing, morphing from a technology provider into a strategic partner. Key to this change is its ability to deliver business value by aligning and supporting the business objectives of the enterprise. IT Management frameworks such as ITIL (IT Infrastructure Library, [3]) provide best practices and processes that support the IT function in this transition. In this paper, we focus on one of the various cross-domain processes documented in ITIL involving the *service level, incident, problem and change management* processes and present a theoretical framework for the prioritization of service incidents based on their impact on the ability of IT to align with business objectives. We then describe the design of a prototype system that we have developed based on our theoretical framework and present how that solution for incident prioritization integrates with other IT management software products of the HP Openview™ management suite.

1 Introduction

Nowadays, organizations are continuously refocusing their strategy and operations in order to successfully face the challenges of an increasingly competitive business climate. In this context, Information Technology (IT) has become the backbone of businesses to the point where it would be impossible for many to function (let alone succeed) without it. As a result of its increasing role in the enterprise, the IT function is changing, morphing from a technology provider into a strategic partner.

To support this radical transformation, various IT frameworks have been developed to provide guidelines and best practices to the IT industry [1]. In essence, these frameworks address either the domain of IT Governance (CobiT [2]) or the domain of IT Management (ITIL [3], HP ITSM, Microsoft MOF). Whereas the domain of IT Management focuses on the efficient and effective supply of IT services and products, and the management of IT operations, IT Governance is mostly concerned setting the goals and the objectives for meeting present and future business challenges. Most importantly, the IT function needs to leverage both domains to ensure that IT decisions are made on the basis of value contribution. In other words, it is of fundamental im-

portance that the selection among various alternative IT related management options that are available to a decision maker at any point in time is made in a way that optimizes the alignment with the business objectives of the organization.

By propagating business objectives and their relative importance from the IT Governance to the IT Operations and Management as suggested in [1], it is possible to integrate them into the decision support tools used by the various IT functions involved in the different ITIL domains.

In this paper, we focus our attention on a particular process of the ITIL Service Support domain, namely *Incident Management* and we present a theoretical framework for the prioritization of service incidents based on their impact on the ability of IT to align with business objectives. We then describe the design of a prototype system that we have developed based on our theoretical framework and present how that solution for incident prioritization integrates with other IT management software products of the HP Openview™ management suite.

The structure of the paper is as follows. In section 2 we recall the definition of the ITIL reference model, with particular attention to the sub-domains of service level management and incident management. In section 3 and 4, we give a formal definition of the problem of incident prioritization driven by business objectives. In section 5, we describe the architecture of a solution for incident prioritization that integrates a prototype that we have developed with some software tools of the HP Openview™ management suite. Finally, we discuss related work and move on to the conclusion.

2 The ITIL Service, Incident and Problem Management Sub-domain

The Information Technology Infrastructure Library (ITIL) [3] consists of an interrelated set of best practices and processes for lowering the cost, while improving the quality of IT services delivered to users. It is organized around five key domains: business perspective, application management, service delivery, service support, and infrastructure management.

The work presented in this paper focuses on one of the various cross-domain processes documented in ITIL involving the *service level, incident, problem and change management* processes. In particular, we focus on the early steps of that process linking both service level and incident management.

As defined in ITIL [3], **Service Level Management** ensures continual identification, monitoring and reviewing of the optimally agreed levels of IT services as required by the business. Most targets set in a Service Level Agreement (SLA) are subject to direct financial penalties or indirect financial repercussions if not met. It is therefore critical for this management process to flag when service levels are projected to be violated in order for an IT organization to take proactive actions to ad-

dress the issue. To this extent, ITIL defines an *incident* as a deviation from the (expected) standard operation of a system or a service that causes, or may cause an interruption to, or a reduction in, the quality of the service. The objective of **Incident Management** is to provide continuity by restoring the service in the quickest way possible by whatever means necessary (temporary fixes or workarounds).

Incident priorities and escalation procedures are defined as part of the Service Level Management process and are key to ensure that the most important incident are addressed appropriately.

Example of incidents may be degradation in the quality of the service according to some measure of quality of service; unavailability of a service; a hardware failure; the detection of a virus.

3 An Approach to Incident Prioritization driven by Business Objectives

In the incident management process it is of fundamental importance to *classify*, *prioritize* and *escalate* incidents [3]. Priority of an incident is usually calculated through evaluation of *impact* and *urgency*. However, these measures usually refer to the IT domain. The central claim of our work is that in order to achieve the strategic alignment between business and IT that is the necessary condition for IT to provide value, the enterprise needs to drive incident prioritization from its business objectives. This starts from evaluating the impact that an incident has at the business level, and its urgency in terms of the cost to the business of not dealing with it in a timely fashion.

In this section we describe the underlying method that our system follows to derive prioritization values for various incidents. In the development and the deployment of the system, we follow the principle that the *cost of modeling should be kept low*; so that it is easily offset the benefit obtained from the prioritization of the incidents. In this work we restrict the application domain of our tool, although the general techniques that we present are more widely applicable. We only consider incidents generated on detection of service level degradation or violation.

3.1 Calculating the Business Impact of Incidents

Figure 1 depicts an *impact tree* which shows how an incident can impact multiple services and in turn multiple Service Level Agreements defined over those services, hence multiple businesses, organizations, etc.

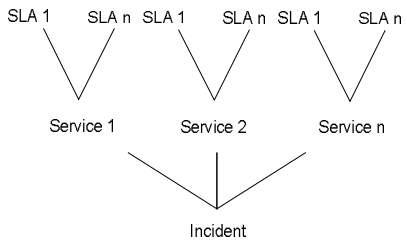

Fig. 1. Impact Tree

In order to assign a priority level to an incident, we start by computing a *business impact value* for it (which we will refer to in the following simply as *impact value*). In general, the impact value of an incident is a function of the time that it takes to get to resolution. We take into account the urgency of dealing with the incident based on how its impact is expected to vary with time. Once the impact values of the various incidents have been computed we prioritize the incidents based on their impact, urgency and on a measure of the expected time of resolution for the incidents.

Among the SLA related business indicators that we take into consideration, there are some quantitative ones such as *Projected cost of violation of the impacted SLAs, Profit Generated by Impacted Customers* and also some quantitative ones such as *Total Customer Experience* defined through the *Number of violations experienced by impacted customers*, etc. Our method requires the definition of *impact contribution maps* over business indicators. Impact contribution maps let us express how much the expected value of each indicator contributes to the total impact of an incident. Because of the assumption that we made above on the normalization of the impact values, all that matters is the shape of the function for any given indicator, regardless of affine transformations. The relative importance among the indicators is going to be adjusted with weights, as it will be clear in the following. As an aside, it should be said here that in order to work with the probabilistic nature of our decision support system, impact contribution maps need to behave like Von Neumann-Morgenstern [4] utility functions, being the calculated impact essentially a measure of the (negative) utility derived from the occurrence of the incident at the business level. Defined this way, impact contribution maps are guaranteed to preserve the preferences of the user among the expected outcomes as a consequence of the incident occurrence. Examples of impact contribution maps are presented in figure 2 and 3.

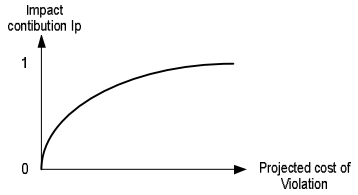

Fig. 2. Impact contribution map for the projected cost of violation of an impacted SLA

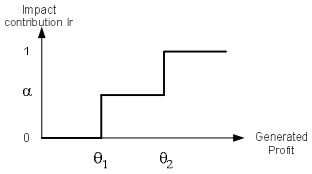

Fig. 3. Impact contribution map for the profit generated by the impacted customer

Figure 2 presents an impact contribution map for the projected cost of violation of an SLA impacted by an incident (measured in dollars, or any other currency). Its meaning is that to a higher projected cost of violation corresponds a higher contribution to the total impact for given indicator. The convexity of the curve symbolizes that the growth rate of the impact slows down as the projected cost of violation grows.

Figure 3 indicates the impact contribution of an incident on the basis of the generated profit by the impacted customers, measured in currency over a given time period (say dollars/year)[1]. It can be noted that three definite regions of profit are defined that correspond to a low, medium and high contribution to the impact. This is equivalent to classifying customers in three categories according to their historical profitability and using that information to prioritize among incidents that impact them so that most profitable customers are ultimately kept happier.

By comparing these two example indicators, we can already see that in the cost of violation example, the value of the impact exhibits a dependency on time. For example, for an SLA guaranteeing a minimum average availability, the longer a system is down, the higher is the likelihood of violating the SLA due to the incident that caused the system downtime. On the other hand, in the customer profitability, there is no such dependency on time, because the values of profitability of the customers are averaged out over a previous history time window and independent of the urgency that is assigned to the incident.

Once all the contributions to the impact are known for a given incident, the information that has been so obtained needs to be integrated over the impact tree, in order to get to an overall impact contribution for each business indicator. For example, in the case of the projected cost of violation of the SLAs, we need to navigate the impact tree and average all the contributions to the impact for all the impacted SLAs. In the next section we are going to walk the reader through an example that will make clearer how this calculation is performed.

The relative contribution of the various business indicators is taken into account by means of a weight that is associated to each business indicator. The formulation of the *incident impact* is as follows. For a set of *n* business indicators, we define $I_j(I, t)$, j =

[1] This measure is supposed to be available through an implemented Customer Relationship Management (CRM) system

1..n as the contribution to the impact of the j^{th} indicator for the incident *i*. w_j is the weight representing the relative contribution of each indicator to the total impact. The total impact *I(i,t)* is given by:

$$I(i,t) = \sum_{j=1}^{n} w_j . I_j(i,t) \quad \text{where} \sum_{j=1}^{n} w_j = 1 \qquad \textbf{(1)}$$

The method described thus far has a very wide applicability. However, at this level of generality, one needs to rely on propagation of information from the operation level to the level of the business indicators, which is a difficult problem to solve in the general case.

In our prototype, the propagation of information from operational metrics to business objectives follows an impact tree similar to the one represented in Fig. 1. We first determine the services impacted by the incident; thence we collate the impacted SLAs.

3.2 Prioritization of Incidents based on Impact and Urgency

Once the business impact of the incidents has been computed, we are faced with the problem of prioritizing them so as to minimize the total impact on the business. Our system requires the use of a *priority scheme*. Together with the definition of a set of priority levels that are used to classify the incidents (defined by the ITIL guidelines for incident management), we require the user to express constraints on what are the acceptable distributions of incidents into priority levels. For any priority level the users can either force the incidents to be classified according to some predefined distribution (e.g. 25%-30% high, 40%-50% medium, 25%-30% low), or define a minimum and maximum number of incidents to be assigned to each priority level. Our method finally requires an expected time of resolution for the incidents that are assigned to a certain priority level, necessary to cope with the business indicators whose contribution to the total impact depends on the time of resolution of the incidents.

The Incident Prioritization Problem

We here present a mathematical formulation of the incident prioritization problem as an instance of the assignment problem. The assignment problem is an integer optimization problem that is well studied in the operation research literature and for which very efficient algorithms have been developed.

Suppose we are required to prioritize between *n* incidents $i_1..i_n$ into *m* priority levels $p_1..p_m$. We introduce a variable x_{jk}, *j=1..m*, *k=1..n* that assumes the value $x_{jk}=1$ if the k^{th} incident is assigned to the j^{th} priority level and $x_{jk}=0$ otherwise.

By observing that the expected impact of each incident can be calculated depending on what priority level it is assigned to, if t_j is the expected time of completion for incidents assigned to priority level *j*, then obviously the impact of assigning the k^{th} incident to the j^{th} priority level is $I(i_k,t_j)$.

The next thing to be noticed is that the constraints that the user imposes on the distribution of the incidents into priority levels can be trivially translated into minimum and maximum capacity constraints for the priority levels. For example, when dealing with n=10 incidents, the requirement that at least 40% of the incidents will be assigned medium priority (assume that is priority level p_2) would read: $\sum_{k=1}^{n} x_{2k} \geq 4$

In general we assign a minimum (c_j) and maximum (C_j) capacity constraint for a priority level j that are symbolized as

$$\sum_{k=1}^{n} x_{jk} \geq c_j \quad and \quad \sum_{k=1}^{n} x_{jk} \leq C_j \quad \forall j = 1..m \tag{2}$$

In order to express the importance of dealing with the most impactful incidents earlier, we introduce a time discount factor λ, $0<\lambda<1$. Introducing time discount gives the desirable property of returning a sensible prioritization of incidents even in cases where the impact of the incidents does not depend on time for any indicator.

The mathematical formulation of the incident prioritization problem (IPP) becomes:

$$(IPP) \quad \min \quad \sum_{j=1}^{m}\sum_{k=1}^{n} e^{-\lambda t_j} I(i_k, t_j) \cdot x_{jk} \tag{3}$$

$$s.t. \quad \sum_{k=1}^{n} x_{jk} \geq c_{jk} \quad and \quad \sum_{k=1}^{n} x_{jk} \leq C_{jk} \quad \forall j = 1..m \tag{4}$$

$$\sum_{j=1}^{m} x_{jk} = 1 \quad \forall k = 1..n \tag{5}$$

$$x_{jk} = 0 \ or \ 1 \quad \forall j = 1..m, k = 1..n \tag{6}$$

The solution of this problem will yield the optimal assignment of priorities to the incidents.

4 A Practical Example of Incident Management Driven by Business Objectives

We now apply the general method to an example that we have modeled in a demonstration of our prototype.

Suppose that our system is used to prioritize incidents based on three business indicators: the *projected cost of violation of the impacted SLAs,* the *profit generated by*

the impacted customers and a measure of the customer experience seen through the *number of service violations experienced by the impacted customers*.

Let's explore more in detail what the definition of each business indicator means.

Projected cost of violation of the impacted SLAs

Our system computes the projected cost of violation through the likelihood of violation that the incident entails for impacted SLAs. For some SLAs there will be certainty of violation, whereas for others (such as service degradation) a value of likelihood depends on the entity of the impact of the incident on the service. In general, as we noted above, the likelihood of violation is also dependent on the time that it will take before the incident is resolved.

In the implementation of our prototype we derive the likelihood of violation from a function that is modeled a priori by looking at the historically significance of a certain value of availability to violating the SLAs in a short successive time frame. More sophisticated methods might be used here; however our system is agnostic with respect to how the likelihood is obtained.

Profit generated by the impacted customer

This is a simpler criterion that would result in prioritizing the incidents according to the relative importance that the customers have on the business, based on the profit that was generated by each customer in a given time period up to the date. If this indicator was used in isolation, it would result in dealing with incidents that impact the most profitable customers first. The value of the profit generated by each customer is supposed to be extracted by an existing CRM system, which Openview OVSD gives an opportunity to integrate with.

Number of violations experienced by the impacted customer

We use this indicator as a measure of the customer experience, which is a kind of more qualitative criterion, although our system must necessarily reduce the qualitative criteria down to measurable quantitative indicators. Therefore in our example, the third business indicator that is used is a sum of the number of violations that have been experienced by the customers with which the SLAs were contracted that are impacted by the incidents. For simplicity of expression, we will consider here all customers being equal, but weights might be added to the computation that would reflect the relative importance of each customer.

Let us now describe the impact contribution functions for an incident *i*

$$i_p(i,s,v(s,i,t)) = 1 - e^{(\frac{-v(s,i,t)}{a})}, \forall s \in SLAs(i) \quad (7)$$

Equation (7) is the impact contribution to the incident i of the projected cost of SLA violation. *v(s,i,t)* is the projected cost of violation for an SLA *s* impacted by the incident *i* when the incident is expected to be resolved within a time interval *t*. The value of the cost of violation is calculated by taking into account the likelihood of violation as described above.

$$i_r(c, p(c)) = \begin{cases} 0 & if \quad 0 \le p(c) < \beta_1 \\ \alpha & if \quad \beta_1 \le p(c) < \beta_2, \forall c \in Customers(i) \\ 1 & else \end{cases} \tag{8}$$

Equation (8) represents the contribution due to the customer generated profit. *p(c)* is the profit that customer *c* yielded in the time period considered.

$$i_k(c, n(c)) = 1 - e^{(\frac{-n(c)}{\gamma})}, \forall c \in Customers(i) \tag{9}$$

Finally, equation (9) is the contribution due to the number of violation for a given customer in a given time period, represented as *n(c)*.

The equations hold for a certain choice of the parameters α, β_i and γ - obviously dimensioned in dollars, dollars and number of violations respectively – We have carried out some experiments to get to a sensible choice of parameters that we will not discuss here as they fall outside the scope of this paper.

The contribution to the total impact of an incident for a given business indicator is computed by averaging all the contributions of each impacted customer and SLA respectively. The averaging weights π express the relative importance of each customer and SLA for computing the total impact contribution of each business indicator. Without loss of generality, in this example, they might be considered uniform.

$$I_p(i,t) = \sum_{s \in SLAs(i)} \pi_{p,s} \cdot i_p(i,s,v(s,i,t)) \quad where \sum_{s \in SLAs(i)} \pi_{p,s} = 1 \tag{10}$$

$$I_r(i,t) = \sum_{c \in Customers(i)} \pi_{r,c} \cdot i_r(c, p(c)) \quad where \sum_{c \in Customers(i)} \pi_{r,c} = 1 \tag{11}$$

$$I_k(i,t) = \sum_{c \in Customers(i)} \pi_{k,c} \cdot i_c(c, n(c)) \quad where \sum_{c \in Customers(i)} \pi_{k,c} = 1 \tag{12}$$

Finally, the calculation of the total impact of an incident i necessary for assigning a priority is carried out through the formula (1), which in this case becomes:

$$I(i,t) = \omega_p I_p(i,t) + \omega_r I_r(i,t) + \omega_k I_k(i,t) \quad \text{where } \omega_p + \omega_r + \omega_k = 1 \tag{13}$$

for a certain choice of the relative importance given to the three business indicators, expressed through the weights ω_p, ω_r and ω_k.

5 An Incident Prioritization Solution

We have built a prototype system that embodies the method described in the previous sections, which we will refer to as the MBO prototype in the following. MBO is

an acronym for Management by Business Objectives, which relates to the more general problem of taking into account business related considerations in the management of IT. In this section, we present a solution for incident prioritization that integrates our prototype with commercially available tools of the HP Openview™ management suite. We begin by briefly describing the features of the Openview components that we used in the integrated solution, and then we present the architecture of the solution, with particular regard to the modifications to the Openview incident handling mechanisms that were necessary for the solution to work.

Overview of the Openview components integrated in the solution

The natural point of integration for our prototype is with the service level management capability of Openview Service Desk (OVSD). OVSD is the tool that falls more squarely in the domains of service level management, incident management and problem management. It allows a user to define a hierarchical service structure with multi-tiered SLA capabilities to describe the relationship between a higher level business service and the supporting operation management service.

OVSD was an excellent starting point for us because it provides most of the links necessary to build the impact tree that we use as the basis of our incident prioritization method. Our MBO prototype complements OVSD by helping the IT personnel faced with the incident prioritization problem with support for their decision based on data and models that are readily available through OVSD.

HP OpenView Internet Services (OVIS) provides monitoring capabilities that are necessary to service level management, as monitoring of availability and response time, along with notifications and resolutions of outages and slowdowns. It builds on a highly scalable and extensible architecture that allows programmers to build probes for a wide variety of data sources.

Architecture of the incident prioritization solution

Figure 4 presents the architecture of the integration of the MBO prototype with Openview Service Desk (OVSD). OVSD receives data feeds from sources as diverse as OpenView Internet Services (OVIS), OpenView Transaction Analyzer (OVTA) and other data feeders. Aside from its reporting activity, the OVSD internal machinery that has to do with service level management -- referred to as OVSD-SLM -- can be summarized in a three step process. The first step is *compliance checking* during which OVSD-SLM seeks to assess whether current measurements comply with existing service level objectives (SLO). This compliance phase uses service level agreements contained in the Configuration Management Database (CMDB) from which are extracted SLOs. Multiple compliance thresholds can be defined for each SLO such as *violation* and *jeopardy* thresholds. This allows for proactive management of degradation of service. The second step is *Degradation and Violation Detection* during which it is detected that a particular metric associated with an SLO has either reported val-

ues that are violating that SLO or meet a jeopardy threshold. In both cases, this leads to the next phase, *Incident Generation*, which reports the violation or degradation as an incident.

At that stage, it is needed to characterize the incident from a business perspective. This is done (step 1) using the MBO prototype prioritization engine. To compute the relative importance of the incident from the business point of view and to prioritize it, the MBO engine fetches (step 2) all the open incidents from the CMDB and extracts the one that have not yet been handled, along with their related SLAs and penalties. Finally, once the priorities are computed (step 3), the MBO engine updates (step 4) all the incidents with their new priorities.

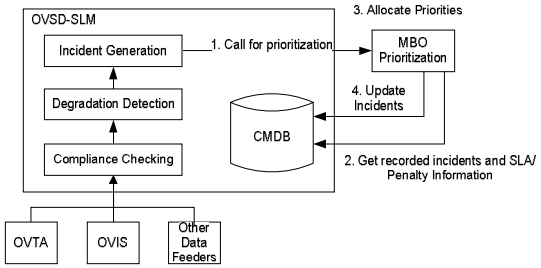

Fig. 4. Integrating SLM with MBO.

For the prioritization solution to work, we had to modify the OVSD-SLM incident handling mechanism so that the MBO prioritization engine is automatically notified on SLA compliance of jeopardy alarms.

6 Related Work

Most of the management software vendors today (such as HP, IBM, Peregrine systems to cite a few) make commercially available tools that are addressed at helping IT managers with incident prioritization. None of them however deals with the problem of driving the prioritization from the business objectives as we do in this work.

One of the few works in the IT management literature that touch on incident management is [5]. However, the aim of this work is quite different from ours, as it concentrates on the development of a specific criteria catalog for evaluating Incident Management for which it provides a methodology.

In any case, we believe that the most innovative aspect of the work here presented is driving incident prioritization from business objectives. From this point of view, among other very valuable works that we cannot review here for space reasons, the most notable in our opinion is [6]. They present a business-objectives-based utility computing SLA management system. The business objective(s) that they consider is the minimization of the exposed business impact of service level violation, for which

we presented a solution in [7]. However, in this work we go far beyond just using impact of service level violations. We provide a comprehensive framework and a method for incident prioritization that takes into account strategic business objectives such as total customer experience thereby going a long way towards the much needed alignment of IT and business objectives.

7 Conclusion

We have shown in this paper that it is possible to integrate the business objectives defined by IT Governance into the decision making process that occurs within the IT Operations and Management functions. We focused our attention on *Incident Management* and we presented a theoretical framework for the prioritization of service incidents based on their business impact and urgency. We also described the design of a prototype system that we have developed based on our theoretical framework and presented how that solution for incident prioritization integrates with other IT management software products of the HP Openview™ management suite. We finally would like to thank Issam Aib for his very valuable comments.

8 References

1. M.Sallé, "*IT Service Management and IT Governance: Review, Comparative Analysis and their Impact on Utility Computing*", HP Labs Technical Report HPL-2004-98, 2004.

2. IT Governance Institute (ITGI), "*Control Objectives for Information and related Technology (CobiT) 3rd Edition*", 2002. Information Systems Audit and Control Association.

3. Office of Government Commerce (OGC), editor. "*The IT Infrastructure Library (ITIL)*" The Stationary Office, Norwich, UK, 2000.

4. J. Von Neumann, O. Morgenstern, "*Theory of Games and Economic Behavior*", Princeton University Press, 1944.

5. M. Brenner , I. Radisic , and M. Schollmeyer, "*A Criteria Catalog Based Methodology for Analyzing Service Management Processes*" In Proc.13th IFIP/IEEE International Workshop on Distributed Systems: Operations & Management (DSOM 2002), 2004.

6. M.J.Buco, R.N.Chang, L.Z.Luan, C.Ward, J.L.Wolf, and P.S.Yu, "Utility computing SLA management based upon business objectives", in IBM Systems Journal, Vol. 43, No. 1, 2004

7. M.Sallé and C.Bartolini , "*Management by Contract*", In Proc. 2004 IEEE/IFIP Network Operations and Management Symposium (NOMS 2004), Seoul, Korea, April 2004