



Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
SCIENZE DELL'INGEGNERIA
INGEGNERIA DELL'INFORMAZIONE

CICLO XXI

COORDINATORE Prof. Stefano Trillo

Business-driven IT Management

Settore Scientifico Disciplinare ING-INF/05

Dottorando
Dott. [Claudio Bartolini](#)

Tutore
Prof. [Cesare Stefanelli](#)

(firma)

(firma)

Anni 2006/2009

Business-driven IT Management

Claudio Bartolini

March 2009

Business-driven IT management (BDIM) aims at ensuring successful alignment of business and IT through thorough understanding of the impact of IT on business results, and vice versa.

In this dissertation, we review the state of the art of BDIM research and we position our intended contribution within the BDIM research space along the dimensions of decision support (as opposed of automation) and its application to IT service management processes.

*Within these research dimensions, we advance the state of the art by 1) contributing a **decision theoretical framework for BDIM** and 2) presenting two novel **BDIM solutions in the IT service management space**. First we present a simpler BDIM solution for prioritizing incidents, which can be used as a template for creating BDIM solutions in other IT service management processes. Then, we present a more comprehensive solution for optimizing the business-related performance of an IT support organization in dealing with incidents.*

Our decision theoretical framework and models for BDIM bring the concepts of business impact and risk to the fore, and are able to cope with both monetizable and intangible aspects of business impact. We start from a constructive and quantitative re-definition of some terms that are widely used in IT service management but for which was never given a rigorous decision: business impact, cost, benefit, risk and urgency.

On top of that, we build a coherent methodology for linking IT-level metrics with business level metrics and make progress toward solving the business-IT alignment problem. Our methodology uses a constructive and quantitative definition of alignment with business objectives, taken as the likelihood – to the best of one's knowledge – that such objectives will be met. That is used as the basis for building an

engine for business impact calculation that is in fact an alignment computation engine. We show a sample BDIM solution for incident prioritization that is built using the decision theoretical framework, the methodology and the tools developed. We show how the sample BDIM solution could be used as a blueprint to build BDIM solutions for decision support in other IT service management processes, such as change management for example.

However, the full power of BDIM can be best understood by studying the second fully fledged BDIM application that we present in this thesis. While incident management is used as a scenario for this second application as well, the main contribution that it brings about is really to provide a solution for business-driven organizational redesign to optimize the performance of an IT support organization. The solution is quite rich, and features components that orchestrate together advanced techniques in visualization, simulation, data mining and operations research. We show that the techniques we use - in particular the simulation of an IT organization enacting the incident management process – bring considerable benefits both when the performance is measured in terms of traditional IT metrics (mean time to resolution of incidents), and even more so when business impact metrics are brought into the picture, thereby providing a justification for investing time and effort in creating BDIM solutions.

In terms of impact, the work presented in this thesis produced about twenty conference and journal publications, and resulted so far in three patent applications. Moreover this work has greatly influenced the design and implementation of the Business Impact Optimization module of HP DecisionCenter™: a leading commercial software product for IT optimization, whose core has been re-designed to work as described here.

This thesis is composed of two parts: in part I we give a brief introduction to the discipline of BDIM (chapter 1); we review the state of the art of research in BDIM and describe some of the open challenges (chapter 2); we present a decision theoretical framework for framing BDIM decision problems that arise in IT service management processes (chapter 3) and we present a methodology and a tool consistent with the framework for approaching the general problem of linking IT metrics with business objectives in a principled way. In part II we present BDIM solutions constructed using the building blocks described in part I: first we give a simple BDIM solution for incident prioritization (chapter 5) that can be used as a template for similar solutions in other IT service management processes; then we present a more comprehensive BDIM solution for organizational re-design of an IT support organization aiming at maximizing its performance with respect to business objectives (chapter 6); finally we discuss the impact of our research and draw conclusions (chapter 7).

Part I: The problem of Business-driven IT management

1 What is business-driven IT management (BDIM)?

Business-driven IT management (BDIM) is the application of a set of models, practices, techniques and tools to map and to quantitatively evaluate interdependencies between business performance and IT solutions and using the quantified evaluation to improve the IT solutions' quality of service and related business results.

*Having a fully virtualized, secured, automated infrastructure, on top of which business critical applications and services run, does not guarantee that IT contributes to the success of the business. Successful alignment of business and IT requires a thorough understanding of the impact of IT on business processes and business results, and vice versa. Besides the conventional IT metrics such as availability and response time, one needs to look at key performance indicators (KPIs), metrics that have significance from the point of view of the business supported by the IT. It is of fundamental importance that the selection among various alternative options available to an IT decision maker is made in a way that optimizes the alignment with the business objectives of the organization. This is obviously true at the IT strategy level for decisions on portfolio management and financial management that pertain to a CIO or equivalent executive level. However, the enterprise stands to benefit if decision made by other stakeholders (IT managers and IT practitioners) are informed as much as possible by the overall objectives of the business. This aspect is termed **business impact analysis** and is one of the foci of the **Business-driven IT management (BDIM)** research agenda.*

*Another aim of BDIM research is **enhancing IT executives and managers' decisions making** through is the ability to predict the consequences of courses of*

actions. An example of such decisions is resource calibration. IT departments need to be looking at grouping people into pools of expertise that can be assigned out to workloads that exist across IT, rather than siloing resources. This is particularly pressing when a CIO or IT leader faces a need to decrease staffing levels. Most workforce reduction work is done manually, according to subjective criteria and prone to huge disruptions in service. When making technicians redundant who may not be top performers, but are the only ones with the skills to run parts of the infrastructure, what is the impact of these decisions on the levels of service provided by IT? When an IT department intends to outsource or out-task certain roles, skills or functions, how can one predict what they can expect as a result in terms of quality of service (or cost, speed, etc.)? What-if scenario and predictive analysis allow IT stakeholders to predict the effect of likely courses of actions before making investments and applying costly measures.

The contribution of Information technology (IT) to business value creation is currently a hotly debated topic. IT is expected to bring value to the business, as is attested to by the introduction of Control Objectives for Information and Related Technologies (COBIT) [1] and Sarbanes-Oxley (SOX) Act [2] compliance requirements. To meet such expectation, IT management methodologies, tools and processes have had to evolve in maturity. Evolution has been made possible with the IT Service Management (ITSM) practices recommended by the process-oriented Information Technology Infrastructure Library – ITIL [3] framework. Other IT management frameworks have been developed on the basis of ITIL by HP, IBM and Microsoft, among others [4]. Automated, ITIL-based IT management procedures are also being proposed for adaptive or autonomic computing platforms [5][6]. In order for IT to help the business achieve its goals, ITSM decisions must be steered by business-oriented measures and objectives. These measures can be inferred from key corporate performance indicators. They can be of a financial nature (revenue, cost ...) or different (personnel utilization ...) or even reflect intangibles such as impact on the company's image. When ITSM uses business measures for decision support instead of (or in addition to) conventional, technical measures (availability, throughput and response time) to signal smoothness of IT operation, we call it Business-driven IT management (BDIM).

BDIM is a new IT management research area [7][8][9][10]. The term BDIM appears to have been first used in [11] in a utility computing context. In general, however, BDIM refers to a new culture and a set of new tools and decision-making processes that explicitly focus on making IT help the business. BDIM attempts to gauge the impact that IT has on the business and aims at rethinking IT management from this perspective, be this in an operational, tactical or strategic context. The

introduction of just a few, functionally limited, BDIM support tools in the market (see [12] for instance) so far, signals that BDIM is still embryonic. On the other hand, BDIM can cover a lot of ground, in autonomic or conventional IT infrastructures as well as in manual IT management procedures. The proceedings of [7][8][9][10] offer an illustrative – albeit non-exhaustive – list of the many areas BDIM can successfully be applied to. There is growing interest in BDIM because of the benefit that it promises to bring to the business.

A significant research effort is required in order for BDIM to mature and be consolidated as the mainstream approach for IT management. We argue that the effort should be mainly directed at investigating the feasibility and options of spreading BDIM applications to cover all ITIL management processes; at fulfilling autonomic self-management needs; and at meeting business requirements and expectations necessary for effective IT governance.

1.1 BDIM applications

Following [13] and [14], we define BDIM as the application of a set of models, practices, techniques and tools to map and to quantitatively evaluate interdependencies between business performance and IT solutions and using the quantified evaluation to improve the IT solutions' quality of service and related business results. Conceptually then, BDIM enacts IT quality improvement and control with business metrics as objective functions – which are related to IT performance metrics through IT–business linkage mapping functions.

Application of BDIM may be carried out in six steps:

1. **Identify business objectives and business-level metrics of interest** – these could be about revenue, cost, inventory turnaround time, etc. Monetized business metrics are of particular interest, since they are

understood in most business settings and allow algebraic operations such as addition.

2. **Select (technical) performance metrics in the context of the IT management scenario of interest.** Scenarios include IT service management processes, autonomic computing platform self-management, software engineering, IT projects, and strategic planning, IT governance, etc.
3. **Model the relevant entities in the scenario of interest**, their attributes and their relationships; and **quantify IT–business linkage**, i.e., estimate the impact that the IT scenario solution has on the business metrics, or, the other way around, how business metrics can lead to IT parameters.
4. **Validate model**, making required enhancements in the model itself and in its associated IT-business linkage quantification.
5. Use the validated model to *support decisions concerning IT solution* in scenario of step 2.
6. For the scenario of interest, **evaluate gains in business results**. Compare gains to business goals. In case discrepancies are unacceptable, make adjustments in the IT solution.

Note that the IT solution could still be in the making, as in a proposed project. In this case, the above steps would evaluate possible business gains (e.g., *return on investment, ROI*) when the project is realized (thus, step 6 may apply to its design). Notice also that by automating the above steps and by looping through them, one does in effect get a BDIM control loop. When all steps are automated, BDIM control may be encapsulated into autonomic computing infrastructures to enact online, on-the-fly self-management. The focus of this dissertation is however on **BDIM solutions**

aiming at providing decision support to human agents rather than on autonomic solutions aiming at taking the human out of the loop. This way of operating is particularly appropriate when tackling decision problems that IT managers and IT staff face in **IT service management processes**, which is our application domain of choice for this thesis.

Step 3 in the methodology described above is where the hardest challenges often reside when building and operating BDIM solutions. To that end, an appropriate BDIM IT-business linkage model – or simply, a BDIM model – must be used. Models for solving a BDIM problem essentially describe relations between business and IT measures. For instance, one such relation could be a function that yields revenue from IT service availability estimates. BDIM model features and resources are to be determined by research efforts which include constructing and validating BDIM models while simultaneously studying IT governance and IT management best practices and methods for eliciting knowledge and policies related to IT and business decisions. The resulting models also depend on the situation and goals of the model's intended usage – such as in the case of autonomic computing or in a decision support tool for IT managers to use.

2 State of the art and research challenges in Business-driven IT management

The purpose-driven definition of BDIM that we have adopted makes it so that many works presented in the areas of distributed systems, network and system management, economics of IT and organizational behavior can be considered as BDIM applications ante litteram.

*By studying those contributions, along with other more recent ones whose authors consciously position in the BDIM research agenda space, we realize that many interesting challenges remain open, in the two orthogonal dimensions of **modeling for BDIM**, and advancing the state of the art at the intersection of BDIM with the disciplines of **autonomic computing, IT service management, and IT governance**. We present an overview of these challenges, and decide to focus our contribution in **BDIM decision support in IT service management**, and contribute decision theoretical framework and models that bring the concepts of **business impact and risk** to the fore, as we will see in later chapters.*

We review the research literature on BDIM and highlight some open research challenges along the two dimensions of *modeling* and *applications*. These two dimensions are not properly orthogonal since the details of a BDIM model depend on the type of IT scenario, business goals and strategies and on the model's intended applications. However for the sake of discussing our research agenda, we dedicate separate sections to addressing construction and validation of BDIM models (section 2.1) and their possible applications domains, broadly characterized as *autonomic computing*, *IT service management (ITSM)* and *IT governance* (sections 2.2 to 2.4). In each of these application domains, examples can be found of BDIM-like solutions aimed at addressing both *automation* and *decision support* problems. Because of the nature of these disciplines, autonomic computing is naturally more concerned with automation, as IT governance is with decision support. In the ITSM space, BDIM solutions address both automation and decision support. Figure 1 conceptually depicts the space of BDIM problems in the areas of consideration.

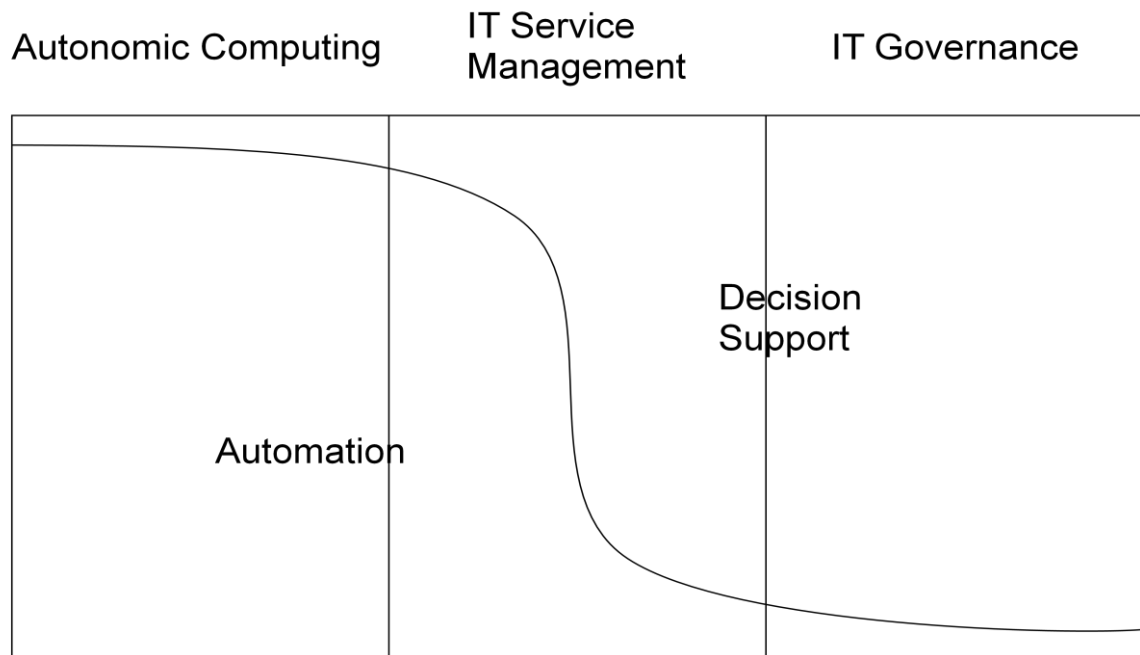


Figure 1: A space of possible BDIM applications

2.1 Research on BDIM models

Technically, the construction of BDIM models is challenging for several reasons:

a) **reusability challenge:** the very different characteristics of the many possible IT scenarios and of the various IT management processes seem to offer no chance for comprehensive and generic BDIM models, leading to very little reusability of models.

b) **model accuracy challenge:** deriving appropriate and accurate values to instantiate the parameters of the model is a very hard task, whether they be estimated by mining historical data or obtained by eliciting knowledge from domain experts and users. As a result, decisions will need to be made based on incomplete or incorrect information on parameter value and even accuracy levels. Research challenges here include using data mining from logs of operations, applications and processes, knowledge acquisition and estimation techniques – such as that discussed in [15] for BDIM applications.

c) **predictive analysis challenge:** the consequences of courses of action on the IT infrastructure and services will need to be predicted and this depends on the infrastructure component or service itself, the workload, the people involved, the schedule for deployment, the business processes affected, etc.

d) **risk analysis challenge:** prediction of consequences will likely need to include risk analysis over a long enough observation period. Strategic impact for instance, may be barely noticeable over months or years. How is one to validate BDIM models in this case?

e) **cost of modeling challenge:** Last but not least, the cost of modeling needs to be kept to a minimum.

The construction of BDIM models of varied complexity, using diverse techniques, has been addressed in [5][6][16][17][18][19][20][21][22] as well as by Web service

QoS modeling efforts, when considering business-IT alignment. The BDIM model used in some of these papers takes into account availability and response time Service Level Objectives (SLOs) at the IT services layer and map SLO violations to lost revenue at the business layer through a decrease in business process throughput. The work in [18][19] adopts utility functions as business measures. The paper in [20] considers IT infrastructure cost. Modeling of finer grained details of IT service is carried out through a Customer Behavior Model Graph (CBMG) in [6][16]. Solutions to the BDIM models have been obtained through queuing theory in [17][20], while simulation is preferred in [6][16]. A closer look at these models reveals several motivating research challenges, listed below and unpacked in the following sub-sections:

- Modeling of all aspects of IT: infrastructure, processes and tools, people
 - Modeling (legacy) IT infrastructure
 - Modeling of IT and business processes
 - Modeling human behavior
- Modeling of financial aspects
 - Modeling of monetizable costs and benefits
 - Modeling of intangible costs and benefits
 - Modeling risk
- Modeling viewpoints of multiple stakeholders
- Extensive validation

2.1.1 Modeling (legacy) IT infrastructure

How does one model legacy infrastructure or at least consider the cost of moving existing services to new platforms? Mappings between IT metrics and business metrics need to be investigated more thoroughly – to include details of the IT

infrastructure such as the network – and widely to include all important IT dependencies affecting business processes or the business in general. For instance, how does one map between CPU and storage performance measures to database transaction throughput to business process throughput to generated revenue? How is cost accounted for? This can also be done in the other direction: how does one relate business requirements down to IT requirements and provisioning decisions? In a similar line, it would be of use to have BDIM models that assist in both a priori and a posteriori decisions such as tuning application parameters to meet business requirements (for example, the number of load-balanced machines running an application server that yields an acceptable tradeoff between business result objectives and IT infrastructure cost). Resulting BDIM models can thus be used in what-if tools to support decisions.

2.1.2 Modeling of IT and business processes

A BDIM modeling effort might be eased by reusing Business Process Management (BPM) models, possibly adjusted to fit BDIM requirements. Business process models may also provide for better drill-down facilities of BDIM tools. While one is navigating between layers, it may actually be useful to top management (or IT management) to see intermediate entities affected by IT. Actually, a good research problem to think about is: what entities should exist because one wishes to see them while drilling down? In other words, there are two reasons for inserting an entity (or whole layer) in the model: i) it helps to link IT to the business; ii) it is useful for the user to see it while drilling down and, thus, make better decisions.

Thomas Schaaf and colleagues [23][24] have undertaken a research effort aimed at exploiting the ITIL definition of IT service management processes, categorizing them

by how much or how little process automation is possible in each of them and where process management tool support can be brought to bear.

2.1.3 Modeling human behavior:

People are one of the principal components of IT. How can BDIM models be enhanced to accommodate labor or productivity issues? As noted in [11], labor costs are usually bigger than hardware costs. Diao et al.'s work in [25][26], moving from the observation that automation does not come for free, proposes a model for predicting labor cost through IT management complexity metrics.

2.1.4 Modeling of financial aspects: monetary costs and benefits

Some BDIM models presented in the literature link a provider's revenue stream to the throughput of the customer's supported business processes (usually in an e-commerce scenario [20][22]). These models consider compensation arrangements whereby the provider collects a fee for each successfully processed e-commerce transaction (revenue is accrued at some rate whenever the IT service provided is up). On the other hand, whenever the associated SLA is violated, the provider may pay the customer a (fixed) penalty. Notice that the IT-business financial link in these referenced models is somewhat naïve: IT service stoppages stop the provider's revenue stream. Closer examination of SLA compensation clauses may also help to improve the financial aspect of BDIM models. Other more realistic IT-business linkage models – which consider pay-per-use, pay-per-volume or subscriptions besides (simultaneous) treatment of both customer and provider sides – have been examined in a Web services context [27] and should be looked into for use in BDIM models.

2.1.5 Modeling of financial aspects: intangible costs and benefits

Intangible aspects such as customer satisfaction, employee morale and others have been tackled by research in management sciences and organization behavior (see for example [28]). A naïve way of taking advantage of that in BDIM models could be by deriving (somehow arbitrarily) corresponding monetary values for metrics representing intangibles. In [5] we used a different approach and made a first step towards treating monetary and intangible costs and benefits within a single coherent framework where the financial dimension is just one of the many dimensions of the business to be modeled, and costs and benefits are measured with respect to a quantitative measure of alignment with business objectives. This thesis improves and completes the framework sketched in [5].

2.1.6 Modeling risk

IT Managers also frequently ponder risk before making decisions. It seems worthwhile to approach any BDIM model from a risk point-of-view. Further, managers frequently have to decide with incomplete knowledge or information or by juggling multiple variables. BDIM models may be enriched if re-examined under the light of disciplines such as decision-making based on incomplete knowledge, decision in the presence of uncertainty, multi-attribute utility theory [29], the Delphi method [30], in addition to interviews and questionnaires (which seem to be the preferred options so far). An interesting application of ethnography to autonomic computing appears in [31].

Sauvé et al. [32][33][34] presented one of the first BDIM models applied to an IT service management process that dealt with risk a first-class concept.

2.1.7 Modeling viewpoints of multiple stakeholders:

Metrics have been calculated from the point-of-view of only a few stakeholders – typically service providers. It is of interest to handle business metrics for several stakeholders at once. Examples of stakeholders include: data center operator, application service provider, business service provider, business process owner, IT service user, customers, IT manager, CIO, CFO, and CEO. Publishing a commented list of business measures of interest to stakeholders in different industry segments, identifying associated business process patterns and supporting IT services would be of much use. Early attempts at this can be found in [35] and [19], which respectively investigate cascading Balanced Scorecard [36] to represent the point of view of multiple stakeholders, and make use of Balanced Scorecards to define business objectives for driving decision support for IT management. Investigating automatic ways of producing metrics from raw data is another interesting topic. Finding whether data mining works in extracting parameter values from logs of operations, applications and processes and the IT services that support them will serve many modeling and validation efforts.

2.1.8 Extensive validation:

BDIM models require case studies, reports on validation efforts and general information on realistic workload characteristics and on business process structure and their dependencies on IT. This need arises for properly building and tuning comprehensive BDIM applications. A BDIM control scheme that can be generalized to tune any application support is presented in [37]. For consolidation, BDIM needs proof-of-concept experiments for services, products and tools.

This section has not exhausted the list of BDIM modeling needs. The next three sections, discussing BDIM application research challenges, will help uncover additional needs.

2.2 Research on BDIM and Autonomic Computing

Autonomic computing [38] is a young research area gaining prominence. The paradigm of autonomic computing could be summarized as modeling artificial systems after self-regulating biological systems. Autonomic artificial systems can be looked at from an intentional standpoint and seen as maximizing their own welfare. A natural application of BDIM to autonomic computing is therefore one where a BDIM approach is used in deriving the measure of welfare that the biologically inspired systems aim at maximizing (usually through the implicit or explicit definition of utility functions).

Autonomous IT management decisions using policies to maximize the business value of IT services are discussed in [6][16][18][22][38][39][40]. The authors in [37] propose an approach for automated enforcement of SLAs by using IT-level feedback loops to maximize profits – revenue as percentage of completed transactions minus rebate to customers who experience bad QoS. In [40], we proposed reviewing SLAs and the enterprise IT balanced scorecard (BSC) in order to keep the cost of eliciting knowledge about the business value of the service low (SLA information is obtained from the Configuration Management Data Base – CMDB [3]). Aiber et al.'s paper [6] also considers the dynamic and autonomous optimization of IT infrastructure parameters through policies that attribute traffic handling priorities according to high-level business objectives such as revenue increase. [22] considers optimization policies for a shared utility computing environment which supports multiple third-party applications subject to SLA performance targets – in terms of maximum

throughput and minimum response times. The resulting model is solved using mathematical and simulation techniques. Menascé et al. [16] describe a policy-based web server resource management system whose goal is to increase revenue by attempting to reduce the number of users that abandon transactions due to poor site performance. Keller et al. [18] propose a scheme for scheduling changes to be executed automatically assuming that impact measures are given as input. [41] looks into financial penalties for service-level violations in an e-Business scenario. The problem studied is that of task/resource scheduling in order to reduce penalties. The authors innovate by tackling the scheduler problem from a business perspective, using monetized business measures, but their approach considers the provider side only. Each of these papers addresses specific autonomic requirements. Comprehensive analyses of autonomic computing research needs appear in [11][38].

In [11], we offered vision for an adaptive IT infrastructure and its requirements, discussing business priority aware technologies for provisioning, monitoring, decision making and control required for autonomic computing and that are under development in academia and industry. To contribute to that development, it is worth re-visiting capacity management problems to consider allocation of a fraction of a physical component. This is required since autonomic computing allows for infrastructure virtualization.

As for any IT technology, drivers for BDIM dynamic provisioning are analytics, data collection, integration, automation and visualization. A research challenge is thus the elicitation of BDIM technology requirements for each of these drivers which could be effected by conducting surveys among CIOs and other IT and business executives of major IT users. Other topics in need of results include: techniques for

business-driven real-time composition and provisioning of Web Services, and methodologies for business-driven capacity planning.

2.3 Research on BDIM over IT service management processes

BDIM applications are possible over many of the IT Service Management processes. In the previous sub-section, we have reviewed some of the BDIM solutions that are more amenable to automation and as such could be seen as application of autonomic computing as well as touching on the relevant IT service management processes. In this section we review applications of BDIM on IT service management processes [3][4], where – as opposed to autonomic computing – people are an indispensable component. Therefore the kind of solution proposed aim at providing **decision support** rather than **automation** (compare Figure 1). A generic research opportunity here is to investigate conditions for running IT service management processes using business measures or utility functions as objective functions – we call this *BDIM over IT service management*.

Optimally designing service level objectives (SLOs) from a business perspective by pondering the cost of the IT infrastructure needed to support the IT services and the losses incurred from service degradations is considered in [20]. The focus is on capacity planning – Service Level Management (SLM) interdependencies. Other investigated IT service management processes include incident management [5][19], service level management [17][20], capacity management [6][16][20], change management [18][42][43][44] and security management [45]. [42] addresses the problem of business-driven planning and scheduling changes. The BDIM model used there links IT availability metrics to monetary loss due to availability service level violations. An example illustrates how the derived metrics may support change management decisions in order to plan and schedule changes to minimize adverse

business impact (which can reach hundreds of thousands of dollars). The authors of [45] address intrusion detection from a BDIM perspective. They propose estimating the business value of a given asset under attack and then deciding what to do next. As security threats are detected, response costs are compared against damage costs. Only when the damage cost is expected to offset the cost of responding should one try to stop an attack. Other studies claim annual gains of millions of dollars due to the introduction of BDIM-like methodologies [6][20].

In [5], we introduced a BDIM method and a corresponding reasoning engine, Aline that can be applied across different IT service management processes, with an example application to incident management. Aline computes the business alignment for each of the possible IT management options. Options are then ranked based on their values of the utility to the business. In this thesis we take that work further by completing the framework and giving a more comprehensive example. In the same spirit, we presented an example for business-driven change scheduling in [44], where we used an alternative model for estimating the cost to the business of the various scheduling options.

Although single-process solutions have been developed, the *BDIM over IT service management* approach faces a much harder challenge when considering multiple interacting IT service management processes simultaneously. The complexity of modeling the interactions of incident, problem, change, security and continuity management has not been tackled yet to the best of our knowledge.

One aspect of *BDIM over IT service management* that makes it particularly challenging is the need of incorporating models of people's behavior into BDIM decision support tools. These challenges might be interesting for HCI (human-

computer interface) researchers thereby bringing about cross-fertilization between the areas.

The research contributions we have reviewed so far in this section are not dependent on the nature of the relationship between the IT function and the business that it supports. However, when the business *outsources* its IT function to a different enterprise [46], the relationship between the two is precisely and thoroughly formalized through sourcing contracts, opening up new and interesting opportunities for BDIM research. A core IT service management instrument for managing IT sourcing contracts – as defined in ITIL [3] - is the *Service Level Agreement (SLA)* by which certain promises are made to clients about the quality or performance objectives of the service provided and how the provider is to be compensated / penalized. Definition of SLA contracts are discussed in [46][47]. (Similar concepts apply when the IT function is provided in house: ITIL defines *Operational Level Agreements (OLA)* that though not as precisely formalized as SLAs, can still be used as instruments to capture the nature of the relationship between business and IT offering similar opportunities for BDIM research.)

Schmidt [47] combines service contracts with workflow concepts. The use of the customer's business processes as a basis for the contract ensures a customer-oriented service view. Workflow concepts allow both specifying non-ambiguous contracts and constructive instructions for usage and management of services by the customer. Using Transaction Cost Theory and Incomplete Contract Theory in six case studies and interviews with contract managers and legal experts, outsourced SLA contracts have been studied in [46] to provide insight into 3 important tasks: identification of possible future scenarios for inclusion in contracts (this is difficult due to the incompleteness of the contract); cost analysis in an “outsourcing versus internal”

decision and costs of transition to another provider. The approach taken in the paper is qualitative rather than quantitative, and its scope does not include the definition of SLA contract clauses (such as SLOs) or the selection of IT services for SLA contracting. The authors in [20] bring in cost considerations in an approach to design and negotiate SLAs for an e-commerce environment. Both the provider's and the client's points of view are taken into consideration. Studies that quantify cost and offer comprehensive, business and technical guidelines for service / SLO definition will be nice complementary contributions to the results in [46][47][20]. Further developing models to account for both outsourcing contract parties (provider and customer) simultaneously is likely to intersect with the research on models for business processes as suggested in [46].

Alves et al. [48] worked at the intersection of risk and IT offshoring, presenting a framework and a tool for assessing and minimizing risk in offshoring of IT projects.

Managing outsourcing relations is a tough problem because of the lack of transparency and difficulty of explicating all the possible ramifications of the contract execution flow. Future attempts at tackling this problem may unveil ways for BDIM to cross-fertilize the research in Business Process Management (BPM).

2.4 Research on BDIM and IT Governance

IT Governance applies to the alignment of IT with a company's mission, strategic goals and expected results. The new ITIL documentation uses the term "stereo vision" to refer to business and IT alignment [3]. IT Governance is concerned about IT delivering value to the business and that IT risks are mitigated. While value delivery is focused on the creation of business value, risk management is focused on the preservation of business value [49]. COBIT is an IT governance framework [1]. COBIT and ITIL are not mutually exclusive and can be combined to offer more

powerful support for IT Governance [50][51]. Sallé [51] remarks that *IT Governance* defines “what” the IT function should pursue or achieve and *IT management* suggests “how” it will achieve it. The fact is that the IT function can have multiple ways of being integrated in the enterprise, and must work on three levels [52]: *strategic* – where decisions are made, policies are set, rules are created, financial goals are defined and plans are prepared; *tactical* – where decisions are implemented; and *operational* – where implemented decisions become active. BDIM can tightly integrate IT management to IT governance.

IT Governance issues concerning project development and (long term) planning of the corporate IT infrastructure and services are of interest here. IT projects have been the object of several studies that attempt to show their contribution to business performance measures such as productivity, consumer value or decision velocity (by flattening corporate hierarchical organizations, for instance) [53][54]. The measures considered however, do not usually encompass all aspects of the business nor is there unquestionable verdict on IT benefits [54]. Clearing up doubts on IT contribution to business and coming up with a single, generic, summary measure remain challenging. It is worth noting that the IEEE International Conference on Exploring Quantifiable Information Technology Yields (IEEE EQUITY [55]) was established recently to quantify the investment yield of existing or proposed IT projects. Research challenges that EQUITY hopes to address include cost estimation, auditing, outsourcing, portfolio management, productivity, project business impact assessment, ROI determination and IT Governance [55]. IEEE Equity can thus be seen as a sister conference to the BDIM Workshop [7][8][9][10]. Research efforts on planning aspects for IT from a business perspective have focused on IT-business alignment and on stand-alone strategic issues. An early and notable example is [56], with a balanced

scorecard-based framework for an integrated business and IT planning and evaluation process. The paper addresses two problems: time lag (IT usually lags business planning) and IT-business communication. For addressing the first, the authors propose common performance measures and measurement procedures; and, for the second, that some measures be monetized. A monetized metric (potential financial loss) and a balanced scorecard-based analysis that relates IT SLA violations and strategic business objective misses were proposed in [21], and by us in [5]. The approach in [21] supports IT investment decisions that improve business results even in complex IT service-business process interdependency scenarios. Non-dimensional – i.e., pure scalar, specially non-monetized – alignment measures tend to lose business semantics: a higher valued alignment option (say 0.97) may lead to lower business results (such as revenue) in absolute terms, if multiple business dimensions are being considered as in the case of sales and production. However, that work does not address the observation made by Kephart in his autonomic computing manifesto [38] “*utility functions are an attractive lingua franca for high-level objectives ... [but] ... humans find them difficult and awkward to specify*”. In [5] we addressed exactly that point and deriving credible utility functions from business objectives is one of the main contributions of this dissertation, building on that early work.

Operations and tactics can affect strategy: if an e-commerce site shows availability problems, the on-line selling strategy is likely to present poor results. A research opportunity in the IT governance domain is thus to build and solve corresponding BDIM strategic models that link operational and strategic aspects of IT, including investments. Support for IT investment decisions (which are obviously linked to IT project selection) has been considered by several authors [57][58][59][60][61][62]. Most of the work concentrates on examining the potential return on investments [58].

Dutta and Roy in [62] propose to link IT projects to business benefits by capturing interactions in the business process flows in which they are embedded. A system dynamics solution technique is used. The literature also documents other techniques from Management Science – Return On Investment (ROI), Total Cost of Ownership (TCO), Economic Value Added (EVA) [63] and Activity Based Costing (ABC) [64], Information Economics [65], Balanced Scorecard [66] and Real-options Theory [67] – when applied to *IT projects*, yet to become products or services. A validated, comprehensive approach that encompasses most of these techniques would be an interesting result of efforts on IT project evaluation. The survey in [68] offers directions for tackling IT ROI research challenges. Yet another research challenge is to produce support models for investments in *operational* IT services to improve business results. To that end, the same management science techniques above, which are easily communicated to IT and business executives alike, can be used as anchors for the business layer of BDIM IT-business linkage models..

Lastly, a whole area also related to IT governance and that offers many intriguing BDIM research challenges is what we call *BDIM Economics* – the study of economic relationships between IT and the business to answer questions like: What are IT limits in contributing to a company’s results? What can be optimized on the business side (not IT) with a BDIM model (by varying product price, say, in order to do the best with whatever IT infrastructure is available)? Where to invest a given amount of dollars on IT to make a business difference?

2.5 Discussion

From our definition of BDIM, our survey of the state of the art of BDIM research and exploration of open research challenges, it is apparent that the whole space that we touch on in this chapter is way too vast for anyone to try and embrace it all. In this

thesis we will focus our contributions in the space of **BDIM decision support over IT Service Management processes** (Figure 2).

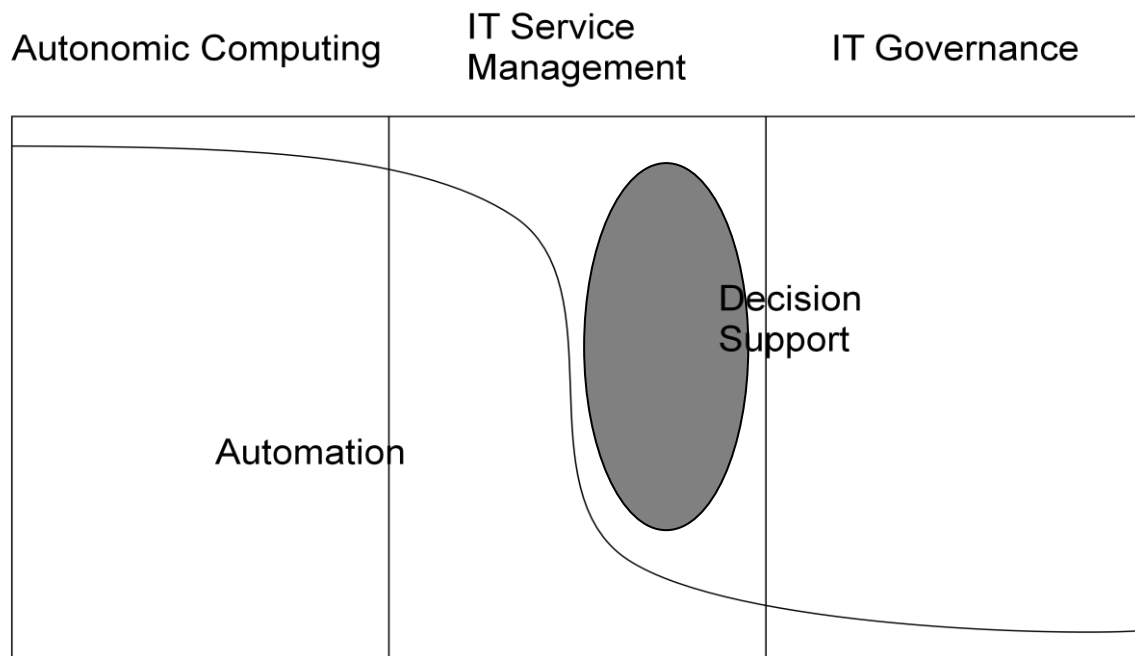


Figure 2: Scope of this dissertation

From the point of view of the modeling challenges, this thesis tackles and advances the state of the art in BDIM modeling by presenting a decision theoretical framework that models **business impact** aspects of BDIM including **monetary** as well as **intangible** costs and benefits. Moreover, our framework deals with **risk** consistently with the economic / decision theoretical definition of risk.

3 Enabling Decision Support in IT Service Management

The scientific literature on IT Service Management and related disciplines abounds with excellent contributions tackling economic and business-related issues. In order to organically address economic and business-related issues in IT service management, a change in perspective is necessary that promotes objectives, preferences, policies of the agent making the choice to the fore, relegating the actual solution techniques to play a – still indispensable, but – secondary role in the background.

*In this chapter we introduce a decision theoretical framework for providing decision support by taking into account economic and business issues in IT service management. We start by providing a constructive definition and characterization of terms such as **benefit, cost, impact, risk, urgency**, so that they can be used in decision support for IT management processes based on quantitative analysis of alternative options. We then indicate ways of mapping these quantitative measures to representations of preference decisions, using business objectives as reference, to put in place the basic abstractions to deal with decision support in IT service management.*

IT managers who are responsible for aspects of IT service management (ITSM) processes are continuously faced by difficult choices. Consider the problem of an IT support manager wanting to reduce by 10% the average duration of service incidents. How can she decompose the problem into more digestible chunks? Even as the original decision is broken down into individual components, a few factors make the choice hard. First there is usually no perfect knowledge about the state of the IT (or the portion of it within the boundaries of relevance for the problem at hand). The complexity of IT systems and networks is such that sometimes one has to do with a very shallow understanding of the various components and their interactions, and therefore of the consequences of the courses of action that one may undertake. What will happen to the response time of our application when we upgrade our database server? Furthermore, even when the relevant portion of the IT systems is fairly well understood, there may be uncertainty on the consequences of the actions, or non-predictability of boundary conditions. What is the likelihood that hackers will exploit a known security threat before a patch becomes available and gets deployed on IT systems enterprise-wide? Or even, what impact will the new policy on password expiration have in terms of how many operators will write their password down on post-its stuck to their monitors?

For a moment, let's assume that we know how and can model all that. The problem we are addressing here is that most of the time IT managers don't know how to compare likely consequences of different courses of action, and this is what makes the choice hard. In other words, they don't know what they want. When needing to make tradeoffs in prioritizing changes which will forcedly result in some breach of service level agreements (SLAs), which SLAs is best to breach? In this chapter, we lay the

foundation for an organic approach aimed at addressing questions such as these, business-related and economic in nature.

The ITSM scientific literature is rich in examples of point solutions to these decision problems, as extensively discussed and cited in section 2.3. Most of the time, the contributors perceive and present – and rightly so – the value of the solutions as residing in the innovative techniques that they used to solve the problem. Typically the storytelling goes something like this: “we have perfected this outstanding technique for {planning / scheduling / prioritizing / ...} applied to {capacity management / load balancing / incident management / ...}. Suppose that the preferences of the user are expressed through the following {utility function / logic predicates representing goals}...”. Because the perceived value of the solution is elsewhere, the contributors do not spend a lot of effort synthesizing the utility functions or goals, let alone running reality check on them. Utility functions and logical predicates representing goals are more often chosen in order to make calculations simpler than for being a high-fidelity representation of the actual user preferences and goals.

In order to organically address economic and business-related issues in ITSM, a change in perspective is therefore necessary that promotes objectives, preferences, policies of the agent making the choice to the fore, relegating the actual solution techniques to play a – still indispensable, but – secondary role in the background.

Decision theory [69] comes in handy when trying to address questions such as the above. It provides us with a set of tools to tackle decision problems such as the ones stated above, provided that one has a fair grasp of the likely outcomes of the alternative choices, and that one knows what he or she wants, or prefers. The preferences of the agent making the choice can be represented as a **utility function**

over the domain of possible outcomes. The problem is now eliciting those preferences from the agent. When the agent is an individual, preference elicitation is a long and expensive process, often carried out through making the user navigate through a huge information space and obtaining information from them in the form of comparisons or ranking (see for example [70]). When the agent represents an organization – such as an organization running IT infrastructure on behalf of a business – the preference elicitation problem is simplified by the fact that organizations need and tend to spell out their business objectives and business policies, and those could be either used (nearly) as they are for logic-based reasoning, or used to derive utility functions to enable quantitative reasoning. This is the basis for the approach we sketch here.

As described in the previous chapter, the IT Information Library (ITIL, [3]) and the Common Objectives for IT and related Technology (COBIT, [1]) are the most comprehensive efforts to date in providing guideline for infrastructure, management and security of IT systems. ITIL defines a number of processes for IT service delivery and IT service support, and gives guidelines for their effective implementation, including basic guidelines for orienting choices and tradeoff in executing the processes, such as suggesting to measure impact, risk, urgency, priority, costs and benefits among other characteristics of various aspects of the processes. COBIT defines a number of processes for IT management and IT governance, and provides indicators for assessing performance (key performance indicators – KPIs) and alignment with business objectives (key goal indicators – KGIs).

Although they provide an excellent starting point for driving the choices that IT organizations face, they do present a number of limitations. On one hand the definitions of the measures of impact, risk, etc. are vague and not necessarily quantitative. (Compare ITIL's approach to the incident prioritization problem, in the

box below. More importantly the measures do not adjust to the extremely dynamic nature of the organization and the systems (for example the priority of a service incident is set at creation and only reviewed in extraordinary cases). On the other hand, the definition of the performance and goal indicators is aimed at providing a high-level, organization-wide metrics which only reflect the performance of the IT organization in executing the process, rather than the alignment with the business. Just because all application servers are happy and running, it doesn't mean that the business is thriving, and vice-versa.

In incident management, ITIL recommends that incident priority be based on the urgency (the acceptable delay to the user or business process) and the impact (the extent of the deviation from normal service level, in terms of the number of users or business processes affected). ITIL also recommends to segment urgency and impact in three or more categories (e.g. high, medium and low) and to derive priority from the priority coding matrix represented in Table 1.

Table 1. Priority Coding Matrix

		Impact		
		high	medium	low
Urgency	high	critical	high	medium
	medium	high	medium	low
	low	medium	low	planning

From ITIL [3]

In the remainder of this chapter, we begin to sketch our conceptual framework for providing decision support by taking into account economic and business issues in IT service management. We start by providing a constructive definition and characterization of terms such as benefit, cost, impact, risk, urgency, priority, so that they can be used in decision support for IT management processes based on

quantitative probabilistic analysis of alternative options. In the second part of this dissertation (starting at chapter 5) we will then indicate ways of mapping these quantitative measures to representations of business objectives and business policies, to put in place the basic abstractions to deal with decision support in business-driven IT management.

3.1 Enabling Decision Support in ITSM: a decision theoretical framework based on a constructive and quantitative re-definition of some terms

The first thing to define to guide our decision theoretical framework for business-driven IT management is to derive suitable **utility functions** (see text box). Utility functions are used in decision theory to express a decision maker's preferences over a set of expected outcomes. A refresher of utility function is provided in the text box below. Utility functions are quantitative instruments, therefore in this and the following subsections of this chapter we will give quantitative definitions of aspects that are related to business performance, such as business impact, cost, urgency and risk and build over them to construct suitable utility functions to drive our decision support methods and tools.

Utility functions (adapted from wikipedia.org)

While [preferences](#) are the conventional foundation of [microeconomics](#), it is often convenient to represent preferences with a utility [function](#) and reason indirectly about preferences with utility functions. Let X be the **consumption set**, the set of all mutually-exclusive packages the consumer could conceivably consume (such as an [indifference curve](#) map without the indifference curves). The consumer's **utility function** $u : X \rightarrow \mathbf{R}$ ranks each package in the consumption set. If $u(x) \geq u(y)$, then the consumer strictly prefers x to y or is indifferent between them.

For example, suppose a consumer's consumption set is $X = \{\text{nothing}, 1 \text{ apple}, 1 \text{ orange}, 1 \text{ apple and } 1 \text{ orange}, 2 \text{ apples}, 2 \text{ oranges}\}$, and its utility function is $u(\text{nothing}) = 0$, $u(1 \text{ apple}) = 1$, $u(1 \text{ orange}) = 2$, $u(1 \text{ apple and } 1 \text{ orange}) = 4$, $u(2 \text{ apples}) = 2$ and $u(2 \text{ oranges}) = 3$. Then this consumer prefers 1 orange to 1 apple, but prefers one of each to 2 oranges.

In microeconomic models, there are usually a finite set of L commodities, and a consumer may consume an arbitrary amount of each commodity. This gives a consumption set of \mathbf{R}_+^L , and each package $x \in \mathbf{R}_+^L$ is a vector containing the amounts of each commodity. In the previous example, we might say there are two commodities: apples and oranges. If we say apples is the first commodity, and oranges the second, then the consumption set $X = \mathbf{R}_+^2$ and $u(0, 0) = 0$, $u(1, 0) = 1$, $u(0, 1) = 2$, $u(1, 1) = 4$, $u(2, 0) = 2$, $u(0, 2) = 3$ as before. Note that for u to be a utility function on X , it must be defined for every package in X .

A utility function $u : X \rightarrow \mathbf{R}$ **rationalizes** a preference relation \preceq on X if for every $x, y \in X$, $u(x) \leq u(y)$ **if and only if** $x \preceq y$. If u rationalizes \preceq , then this implies \preceq is complete and transitive, and hence rational.

3.1.1 Cost, income, net cost

When thinking of business-driven considerations, the first concepts that come to mind are **cost** and **income** to the business, *usually expressed in monetary terms*. Observing that – whatever the quantitative indicators of choice – **net cost** can be defined as cost minus income, without loss of generality we can do away with talking about income and concentrate on (net) cost.

We choose to concentrate on the negative side (cost) of the equation, because most often management options will come at a cost. We could just as well have made the opposite choice.

3.1.2 Business impact

Monetary measures work quite well for measures such as revenue brought in by IT services, or cost to a department either defined as chargeback or through SLA that defines economic terms and conditions. However, a comprehensive framework for business-driven IT management must take into account **intangible costs and benefit** to the business, such as loss of reputation, increase in customer satisfaction, employee morale, etc... All of these can be *monetized*, for example by eliciting from the decision makers the monetary cost due to loss of reputation for providing worse than expected quality of service to some customers. Taking a more general approach, we'll mimic decision theory's definition of utility functions in defining a **business impact function** for expressing the net benefit to the business of the outcomes of making given decisions and/or following given courses of action. Again we will assume the definition to apply to a net business impact. However, given the meaning most often associated to impact of a negative effect of an outcome, we'll take business impact to have a greater value for a less preferred outcome.

One question arising at this point is: "*whose preferences are we taking into account?*" For example, when Microsoft makes a security patch available, it is in the interest of the IT department to make sure that the software is patched in the office productivity machines belonging to lines of business. From the point of view of the line of business, this preventive patch management might be a disruption and carrying negative impact, however, from the point of view of IT, it saves more work later.

Here, we assume that the net benefit that we aim at maximizing is that of the decision maker. In business-driven IT management, that responsibility usually sits with the IT department, or that of the role responsible for its management. The IT department supports numerous customer, both internal (other units of the same organization or enterprise) and external (as it's the case for example in an outsourcing context). While estimating the net benefit to the IT customers of the different outcomes is important and useful, we decide to use the IT customer utility as an input variable to the calculation of the net benefit to the IT department. For instance, tradeoffs are often necessary that needs to consider choices that might disrupt one or more lines of business, but still may benefit the enterprise as a whole by allowing IT to save time and money.

3.1.3 Utility functions harmonizing tangibles and intangibles (cost and business impact)

Since the objective of the task at hand is to derive useful quantitative criteria to use as utility functions to drive decision support, it is important to harmonize the concepts of cost and business impact. In other words, we want our decision theoretical framework to be able to seamlessly cope with both tangible and intangibles benefits and cost to the business.

There are three ways to do so: 1) **monetize intangibles**: associate a monetary cost to figures of business impact; 2) **understand business impact of financial cost**: treat the financial dimension as one of the possible way to optimize the business; 3) **define a suitable utility function** that contains aspects of both, such as a linear or convex combination of suitably derived quantitative measures.

Let's talk a little about all three.

Monetize intangibles

As we said in section 3.1.2 on business impact, intangible cost/benefits to the business can usually be *monetized*. This is could be done for example by eliciting from the decision makers the monetary cost due to loss of reputation for providing worse than expected quality of service to some customers (as advocated for example in [21]).

After monetizing intangibles, cost (including now intangible costs in terms of adverse business impact) can be used to derive a suitable utility function to choose between alternative options. An obvious such utility function is $u = -cost$; more in general u will be a monotonic function defined over cost measures (as cost increases, utility decreases and vice-versa).

Understand business impact of financial cost

Taking the dual approach to monetizing intangibles, one could consider the financial perspective is just one of the many dimensions that the business would want to optimize for business-driven IT management solutions. This is a less obvious choice, and the choice we make in this thesis. Besides our approach that we will describe in detail in the next chapter (first introduced in [5]), to our knowledge there are no previous examples of it in the literature.

Likewise, in a dual fashion to what discussed above, when financial cost is used as one component of business impact, then business impact itself can be used to derive a suitable utility function. E.g. $u = -impact$. Again, more in general u will be a monotonic function defined over business impact measures (as business impact increases, utility decreases and vice-versa).

Define utility functions over both cost and business impact

Finally, one could just decide to harmonize tangibles with intangibles by defining utility functions that contains aspects of both. E.g. $u = -w_{cost} * cost - w_{impact} * impact$ for appropriate choices of the relative importance weights w_{cost} and w_{impact} taken as positive real numbers (hence the minus signs in the formula).

For each of the three previous harmonization alternatives, we gave default examples of utility functions. In the following we will see more interesting examples of utility functions built over cost that can codify *risk attitudes* of decision makers, as well as their preferences. This is the subject of the next sub-section.

3.1.4 Risk

When estimating cost and business impact as defined above, there will necessarily be a degree of **uncertainty**. This uncertainty in business impact estimations is what characterizes **risk**. Jacques Sauvé [33] suggests making a distinction between *epistemic uncertainty* and *stochastic uncertainty*. The former has to do with how much one knows or doesn't know about factors affecting cost, business impact and therefore utility. The latter has to do with the fact that even if we assume perfect knowledge about cost and business impact, there will be variability in the possible outcomes following a given course of action (and their utility thereof).

In our framework we take the economic / decision theoretical definition of risk. Risk is defined as a *second order momentum* over utility measures, assuming uncertainty on them and basic knowledge over utility distributions. In order to get to a quantitative treatment of risk as a second order momentum over utility, a few alternative definitions would all do. For reasons that will be clearer in the following,

we take **risk** to be defined as **variance of the utility function over a given set of outcomes**.

This definition behaves well with respect with all the possible ways to harmonizing tangibles and intangibles that we discussed above. For example, when utility is defined only on financial cost (including monetization of intangibles), the definition of risk is consistent with financial risk. When utility is defined only on business impact (including the financial perspective of the business) then **risk becomes a measure of the uncertainty of business impact estimates for given courses of action**.

We observe these definitions of risks are consistent with intuition in that that the higher the variability of cost or business impact in a given situation, the riskier the situation.

An important side effect of choosing a definition of risk that is a measure of the uncertainty is that we can now define utility functions over either cost or business impact, or a combination of the two that can encode information about the risk appetite of the decision makers. Readers that are familiar with the basics of utility and risk in decision theory, might already appreciate this point when dealing with utility of money outcomes: a linear (more properly: affine) utility function over a domain measured in money represents a **risk neutral** attitude. Convex and a concave utility functions represent **risk averse** and **risk seeking** attitudes respectively. (For readers who are not so familiar with these concepts, see the text box below.) The extension of these concepts from monetary value to measures of business impact as we have defined is obvious.

Risk attitudes (adapted from wikipedia.org)

A person is given the choice between two scenarios, one with a guaranteed payoff and one without. In the guaranteed scenario, the person receives \$50. In the uncertain scenario, a coin is flipped to decide whether the person receives \$100 or nothing. The expected payoff for both scenarios is \$50, meaning that an individual who was insensitive to risk would not care whether they took the guaranteed payment or the gamble. However, individuals may have different **risk attitudes**. A person is:

- **risk-averse** if he or she would accept a payoff of less than \$50 (for example, \$40), with no uncertainty, rather than taking the gamble and possibly receiving nothing.
- **risk neutral** if he or she is indifferent between the bet and a certain \$50 payment.
- **risk-seeking** (or **risk-loving**) if the guaranteed payment must be more than \$50 (for example, \$60) to induce him or her to take the guaranteed option, rather than taking the gamble and possibly winning \$100.

The average payoff of the gamble, known as its [expected value](#), is \$50. The dollar amount that the individual would accept instead of the bet is called the [certainty equivalent](#), and the difference between the certainty equivalent and the expected value is called the [risk premium](#).

In [utility](#) theory, a participant has a utility function $U(x)$ where x represents the value that he might receive in money or goods (in the above example x could be 0 or 100).

Time does not come into this calculation, so inflation does not appear. (The utility function $u(c)$ is defined only [modulo linear transformation](#) - in other words a constant factor to be added to the value of $U(x)$ for all x , and/or $U(x)$ could be multiplied by a constant factor, without affecting the conclusions.)

The graph shows this situation for the risk-averse player: The utility of the bet,

$$E(u) = (U(0) + U(100)) / 2$$

is as big as that of the certainty equivalence, CE , in this case $U(40)$.

For instance $U(0)$ could be 0, $U(100)$ might be 10, $U(40)$ might be 5, and for comparison $U(50)$ might be 6.

The risk premium is $(\$20 - \$10) / \$10$ or 25%.

In the case of a wealthier individual, the risk of losing \$100 would be less significant, and for such small amounts his utility function would be likely to be almost linear, for instance if $U(0) = 0$ and $U(100) = 10$, then $U(40)$ might be 4.0001 and $U(50)$ might be 5.0001. The above is an introduction to the mathematics of risk aversion. However it assumes that the individual concerned will act entirely rationally and will not factor into his decision non-monetary, psychological considerations such as regret at having made the wrong decision.

Often an individual may come to a different decision depending on how the proposition is presented, even though there may be no mathematical difference.

From there it follows that decision makers – depending on their risk attitudes - may not choose all the time the highest expected outcome. They may choose a lower risk

option over a higher expected outcome, depending on their risk attitude. We'll discuss interesting applications of this last point in our discussion on risk in section 3.4.1 at the end of this chapter.

3.1.5 Urgency

Given our quantitative definition treatment of utility in terms of cost and business impact above, we observe that utility usually varies with time, and most of the time this happens in foreseeable ways. In some cases, the cost and business impact of service incidents or delayed changes to the system will be immediate and constant. Such is the case of a black-out, or a sudden interruption of the systems. In other cases, the situation may worsen with time (e.g. a detected trend in some service level so that it can be foreseen that at some point an SLA violation will result. In some cases it may even happen that the problem will go away if one does nothing about it (e.g. problems due to outdated software that is scheduled to be updated). Obviously there is a cost associated with not fixing the problem, but the point is that in every case we have a profile of impact in time, that might let us decide that it's just more convenient to wait until the next scheduled release before doing anything.

The variability of the impact rate in time is perceived through a sense of **urgency**. Urgency means that fixing the problem will take time, and we must estimate how long it will take to fix it, with a certain degree of confidence. Moreover the deep meaning of urgency is more related to how soon one should start responding to a problem than how soon does this problem need to be solved. An estimate of the time necessary for fixing the problem must be taken into account. If one expects to incur in an SLA violation in two days, and the problem is estimated to take a day or so to fix, then there is not much time to wait. This can be further refined taking into account resource calendars. If the deadline is two weeks from now, but next week has been

marked as out-of-office for the resource, then the latest possible time to complete the work is at the end of the current week. And if the estimated workload is one day, things need to get underway at latest by Thursday. Notice that having a workaround available changes things, because the workaround can mitigate the impact of the incident.

In sum, **urgency characterizes the variation of utility in time**. When faced for example with a prioritization problem in the context of business-driven IT management, an agent will use their perception of the entity of the cost and business impact, and a sense of its urgency, and they will use that to prioritize the tasks at hand. When the problem is a simple, triage-like prioritization problem it would be overkill to try and guess the whole utility profile and resource availability profile..

3.2 Mathematics of risk and urgency

In the following subsection we give a mathematical formalization of our framework for decision support, by giving quantitative definition of risk and urgency, based on utility functions.

3.2.1 Risk is the variance of utility

Having observed above (3.1.4) that the concept of risk is associated to the spread of possible values of utility for a given situation (the wider the spread, the riskier the situation), we define **risk as the standard deviation of utility**. Given a probability distribution characterized by $pdf(\omega)$ over a set Ω of possible outcomes, $\omega \in \Omega$, on defining a utility function u such that $u(\omega)$ is the utility of a possible outcome $\omega \in \Omega$, and indicating with $\langle u \rangle$ the expected utility over Ω , risk is defined as:

$$risk = \int_{\omega \in \Omega} \sqrt{u^2(\omega) - \langle u \rangle^2} pdf(\omega) d\omega$$

It is easy to see that this definition of risk behaves well with the characterization of risk attitudes described above, though demonstrating it is beyond the scope of this work and can be taken as a simple exercise in game theory.

3.2.2 Urgency is the first derivative on time of utility

Often, outcomes will have a time component that describes them. Even when that is not the case, the distribution of the possible outcomes as a random variable might have a time component (stochastic, non-ergodic variable) and thence will utility. We define **urgency as the rate of change in time of utility** for a given situation. The sign of urgency is set so that the faster a situation is degenerating (utility decreasing), the higher is the urgency associated to that situation. Urgency is then defined as

$$urgency = -\frac{\partial}{\partial t} \langle u \rangle = -\frac{\partial}{\partial t} \int_{\omega \in \Omega} u(\omega) pdf_t(\omega) d\omega$$

where we assume that preferences stay constant over time and therefore $u(\omega)$ does not depend on time t . We express the dependency on time of the distribution of possible outcome as $pdf_t(\omega)$.

3.3 Worked examples

Before closing this chapter with an in-depth discussion of our decision theoretical framework for BDIM, we present a worked example to exemplify the concepts expressed so far. In this example we follow the convention most often followed in the academic literature of choosing financial cost as the measure of the utility of a situation, along with arbitrary assignment of cost values to intangible aspects impacting the business. On top of that we will choose suitable utility functions. We'll revert to our preferred choice of defining utility over business impact in the next chapter.

Let's suppose that an IT manager working for a financial institution is faced with the following situation: a problem has occurred which disabled the credit checking service used for determining whether customers applying for loans are credit-worthy. We want to define a utility function to model the fact that until that problem is fixed, this is costing the business 40\$ per hour.

Let's assume further than an upgrade of the credit checking service is scheduled for one hundred hours (about 4 days) from now, and that the upgrade will make the problem disappear. Our choice for modeling the space of possible outcomes¹ is to use the time at which the credit checking service will resume as a random variable. The situation we are considering does not introduce any variability in the space of the possible outcomes, and we can model it with "*credit checking services resumes at time 100 hr with probability 1*". Our utility-as-cost function for a risk neutral individual ($u = -cost$) tells us that this situation has a cost of 4,000\$ with probability 1, which obviously gives an expected cost of 4,000\$.

There's one important observation to make about accuracy of these dollars estimates. Obviously, the resulting monetary values have to be taken with a pinch of salt. What saves us however is that the point of the exercise is almost always not to come up with credible figures of what will be the cost (or business impact) of a given option. Such figures will more often be used for comparing two or more alternative options with one another. All this means that, even if the absolute figures are way off, their relative comparison might still make sense.

¹ Note that in this case the set of possible outcomes Ω is the time interval between now – posited time 0 *hr* – and one hundred hours from now – time 100 *hr*.

3.3.1 Risk

Now let's suppose that our IT manager has an option of getting Tom, a skilled technician, to work on the problem. Tom will take a time that is uniformly distributed in $[20 \text{ hr}, 40 \text{ hr}]$ to fix the problem (representing the distribution over the set of possible outcomes for the option "call Tom"). Tom's intervention costs the business 2,800\$. The business impact distribution of the option "call Tom" is therefore given by composing the business impact function for the credit checking problem with the distribution of possible outcomes and adding that component to the 2,800\$ cost of making the call. It is easy to verify that the cost is uniformly distributed in the interval $[3,600\$, 4,400\$]$, yielding an expected cost of 4,000\$². To a risk neutral IT manager, the two options would be equivalent. However, it's intuitive that a risk averse IT manager would go for the "wait it out" option rather than "call Tom", given that there is a 50% likelihood that calling Tom will result in savings that do not offset the cost of the option. (A risk-seeking IT manager would opt for waiting instead.)

Things become more interesting when managers have to make trade-off of expected cost versus risk, as they will respond differently depending on their attitude to risk or risk appetite.

Suppose that on hearing about the situation, Tom is willing to lower his cost to 2,700\$ per call. The distribution of possible outcomes stays the same. However, its cost distribution is different and lower, due to the lower fixed cost (uniformly distributed in $[3,500\$, 4,300\$]$). Let's now compare again the "wait it out" vs. "call

² We want to note once again that in the general case, the expected impact will **not** be equal to the impact of the expected outcome. However, in our example, 4,000\$ is also the impact of the expected outcome of fixing at 30 *hr*, because of linearity of the impact function over the set of outcomes.

Also observe that the nice additive property of cost – or business impact – of options only applies if the options are genuinely independent, as is the case here for the cost of calling the skilled technician and the cost incurred because of the open problem. When modeling alternative options and their cost, attention should be paid to this fact.

Tom” options. If the manager is risk neutral, she will just compare the expected income in the two cases. Because this time “call Tom” has a lower expected impact (3,900\$ vs. 4,000\$), she will go for that option. However, if we now take into account risk, the situation becomes more complicated. Suppose that the IT manager’s risk profile characterizes a risk-averse decision maker represented by the convex

$$\text{utility function } u = \sqrt{\frac{8000\$ - \text{cost}(t)}{4000\$}}.$$

The expected utility for “wait it out” is 1. The expected utility for the “call Tom” option is now given by

$$\int_{t \in T} t * u(\text{cost}(t)) dt = \int_{20 \text{ hr}}^{40 \text{ hr}} \sqrt{\frac{8000\$ - (2500\$ + 40 \frac{\$}{\text{hr}} * t)}{4000\$}} dt$$

that as the reader can verify amounts to slightly less than 1. Which means that an IT manager with a utility function so defined over cost would choose the less risky option of waiting for the credit checking system upgrade rather than spending money for an option that might not result in a gain (it’s easy to verify that the breakeven fix time is 32.5 hr, and there’s a high enough likelihood - 37.5% - that Tom’s job will exceed that duration).

3.3.2 Urgency

Let’s return to the original situation with the problem at the credit checking system, the impact of which we estimated to be 4,000\$ with certainty (that is *stochastic* (un)certainty – we admittedly gloss over *epistemic* (un)certainty here, compare [33]). Suppose now that the same IT manager has one more trade-off to make. A different and unrelated problem is affecting the email infrastructure. The (*intangible, monetized*) cost of that because of loss in personnel productivity is estimated in 80\$ per hour: $\text{cost} = 80\$/\text{hr} * t$. A patch release that will solve the problem is

scheduled for 50 hours for now. It can be seen here that the utility-as-cost profile of this situation is that the problem will also cost 4,000\$ with certainty. However the situation with the email problem is intuitively more urgent than the one with the credit checking system problem (and has therefore a higher value of $urgency = -\frac{\partial}{\partial t} < u >$). This is due to the faster rate of worsening of the situation, represented by its negative utility growth rate.

3.4 Discussion

This chapter presented a decision theoretic framework to enable decision support in business-driven IT management.

The approach of defining utility and utility functions over the space of possible outcomes appears in numerous works in the academic literature, see for example [20][22]. While in some cases, the choice of the particular utility functions reflects a need by the author of having an analytically tractable optimization problem to solve, in other cases such as [20][22], more conscious efforts are made towards choosing utility functions that genuinely reflect net benefit to the decision maker. The problem in this latter case becomes one of structuring the functions so that preference elicitation does not become so hard to be a showstopper. This becomes apparent in the case of **intangible** benefits, as estimates for monetization of intangible benefits are arguably precise. Further along this thesis – starting in chapter 4 – we will build on this preliminary framework to present a principled way of determining business impact functions from business objectives in a balanced scorecard [5] that goes a long way towards greatly simplifying the knowledge elicitation process, as we'll move the discussion from monetizing a (huge, potentially infinite) space of outcomes to eliciting a set of preferences between a finite, small, set of business objectives. In our framework, financial cost will be just another dimension over which to optimize,

similarly to how in a balanced scorecard [36], the financial perspective is just one of the four perspectives of the business.

3.4.1 Risk

The main contribution of this chapter is the treatment of risk as a depending concept from cost and business impact, through the definition of utility. Our definition of risk as the variance of the utility distribution (and therefore indirectly of the cost and impact distributions) is consistent with the intuition that wider spreads of possible outcomes leading to wider spreads of cost/impact values carry a higher risk. The important aspect here is really to have defined risk as a measure of the variability of utility. Alternative definitions of risk that are consistent with ours include the characterization of risk in BDIM as value-at-risk, as suggested by Sauv e [34]. A full comparison of these alternatives is beyond the scope of this dissertation. Here it will suffice to make a couple of observations. First: assuming full knowledge over the distribution of utility (or cost, or business impact), one could derive risk-as-variance from value-at-risk and vice-versa. Second: it could be argued either way on whether capturing risk through variance comes more or less natural to decision makers than the value-at-risk alternative. The same goes for other sound decision theoretical definitions of risk such as risk as standard deviation, or max variability of the utility distribution.

The rationale behind our choice to characterize risk as the variance of utility distribution is that risk behaves best of all other alternatives with respect to combining utility from different options, regardless of whether it's utility as cost, utility as business impact or any combination of the two. Given two or more options (not necessarily independent), their expected utility (either computed as cost or business impact or combination thereof) is the sum of the expected utility values of the options

taken separately. More interestingly, *under some certain basic assumptions*³ on the cost (business impact) distributions, **if the options are independent, their combined risk-as-variance is also the sum of the values of risk for each option taken separately.** This is not necessarily true for any other alternative (decision theoretically sound) characterization of risk. If the options are not independent, then – under the same basic assumptions – then, assuming the **co-variance** of the cost (business impact) of any couple of options if known, one can still easily calculate the risk as variance of the combination of the options.

Note that expected utility and risk for an option are the two first order moments of the utility distribution for that option. In order to solve decision problems in first (and second) approximation where options are assumed to be independent, it will therefore suffice to consider expected utility (again either as cost or business impact or combination thereof) and risk for each option. This is usually a much simpler proposition than having to fully define a utility function over the space of possible outcomes and derive distributions for possible outcomes for each options. Note again than when mutual dependencies between options are assumed, they will be modeled through co-variance values for each pair of options⁴.

We do not take in consideration characterizations of risk that are not consistent with decision theory. One such characterization is the oft-used “probabilistic risk”, calculated by estimating the probability of (usually undesirable) outcomes and

³ This is true for instance when two different options are assumed to entail outcomes drawn from the same distribution, which will most often be the case in practical applications of our framework.

⁴ Readers familiar with the classic Markowitz portfolio theory [71] can appreciate similarity between that theory and this framework. We are currently working an application/extension to this framework to deal with IT portfolio optimization based on classic portfolio theory, but this is beyond the scope of this dissertation.

multiplying that for the loss attributable to the outcome occurring. It's easy to see that "probabilistic risk" is strictly speaking a measure of expected utility and not of risk.

3.4.2 Urgency

In our worked example, we limited our consideration on urgency to the simpler of the two ways in which the business impact function depends on time: that is through the characterization of the outcomes. To make our decision support framework more complete, it would be necessary to study the other case: that is when outcomes are described as stochastic, non-ergodic variables. However we had to leave this outside the scope of this dissertation. Our treatment of the quantitative aspects of urgency as a first order derivative of impact is anyway a first step in the right direction as it enables ways of making tradeoffs between impact and urgency that had not been made explicit before. This could be done for example by defining utility functions that take into account both expected business impact and (expected or instant) urgency. Decision makers could be characterized by their "urgency sensitivity" and their profile could be used for tuning the relative importance of perceived impact vs. urgency in defining utility functions. Another interesting way of going about this would be to define an "efficient frontier" over impact vs. urgency considerations, and present the decision maker with a minimal set of options each having the property that no options are available that are both of higher impact and higher urgency. In other words, for the options in the efficient frontier if an option exist that has higher impact, then that has necessarily higher urgency and likewise the other way around.

4 An approach to business-IT linkage

In this chapter we build on the framework introduced in the previous chapter in order to make progress towards a generic solution for business-IT linkage, or the connection between IT metrics and metrics and objectives of relevance to the business level.

In our discussion of business impact in the previous chapter (section), we underlined the need for eliciting business impact functions that are less than arbitrary, as these will be used in BDIM solutions to solve the following class of decision problems: computing a quantitative figure of business impact for alternative options available to IT managers.

In order to obtain a quantitative figure of business impact, we first present an information model that is articulated around a set of key business concepts: objectives, key performance indicators (KPI), and perspectives. The terminology used in this information model borrows where possible from the lexicon of the COBIT [1] (Common Objectives for Information and related Technology) framework and from balanced scorecard [36].

*On top of that model, we define **alignment** with a given business objective the measure of the **likelihood** – given the best knowledge about the current situation – **that the objective will be met**. From there we can compute alignment with a set of business objectives, assuming knowledge about the relative importance of the objectives. From there, we derive a suitable business impact function in function of alignment with a given set of business objectives and demonstrate that this function has the properties that were laid out in the previous chapter.*

Finally, we introduce Aline, a reasoning engine for the computation of alignment with business objectives. Aline uses the measure of alignment thus derived as a (negated) value of the business impact to the business of carrying which is used to rank the alternative courses of action. On ranking the options, Aline returns a suggestion on what course of action to take, substantiated by the evidence that it has for assessing the alignment with respect to the business objectives.

*In the development and the deployment of the solutions, we follow the principle that **the cost of modeling should be kept low**; so that it is easily offset by the benefit obtained from the decision support.*

In this chapter we build on the results of the previous chapter to define a quantitative utility function to drive choices among alternative options in business-driven IT management solutions. Our choice is to use a constructive and quantitative definition of business impact as our choice for harmonizing tangibles and intangibles benefits and costs to the business. We will define that business impact functions in terms of alignment with business objectives. In order to do so we start creating introducing building blocks in terms of key concepts of importance to a business.

4.1 The IT Management by business objectives (IT-MBO) information model

For each of the various IT management domains the generic decision problem is specialized into a decision problem that pertains to that domain. This requires a mapping of the domain specific concept onto the generic concepts that are defined in the MBO information model.

The IT Management by business objectives (IT-MBO) information model (introduced in [5], depicted in Figure 3) is articulated around a set of key concepts: **objectives, key performance indicators (KPI), and perspectives**. The terminology used in our information model borrows where possible from the lexicon of the COBIT [1] (Common Objectives for Information and related Technology) framework and from balanced scorecard [36].

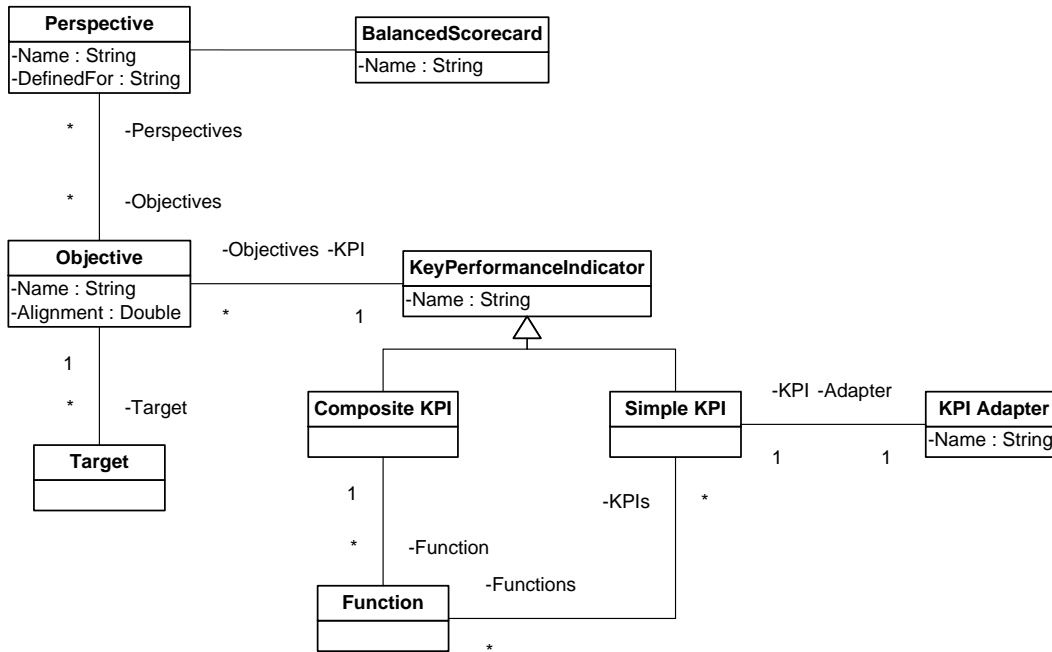


Figure 3: IT Management by Business Objectives (IT-MBO) information model

COBIT is a framework addressing the management's need for control and measurability of IT. It provides a set of tools and guidelines to assess and measure the enterprise's IT capability for the principal IT processes. Balance scorecard is a tool for management that enables organizations to clarify their vision and strategy by capturing them into actionable objectives

In the remainder of this section, we briefly describe the principal concepts defined in the IT-MBO Information Model.

4.1.1 Business Objectives

COBIT introduces key goal indicators (KGI) as measurable indicators of the business objectives. In our model the objectives are the corresponding concept to COBIT's KGIs. They are represented by expressing one or more target values_i over a key performance indicator, or KPI – see below.

4.1.2 Key Performance Indicators

As defined by COBIT, key performance indicators (KPI) are measurable indicators of performance of the enabling factors of IT processes, indicating how well the process enables the goal to be reached.

4.1.3 Perspectives

Perspectives are used to bundle objectives together that concern a certain angle of the business. The concept of perspectives is borrowed from the balanced scorecard [36]. A balanced scorecard defines four perspectives: **financial**, **customer**, **business process** and **learning and growth**. Our model defines a perspective as a first class object, not limiting its usage to the traditional balance scorecard model. Perspectives do not represent a partition over the set of objectives defined. An objective can belong to more than one perspective.

4.1.4 Examples

An example of an objective defined through the model is “*the aggregate service revenue generated over the current three-month period must be above 100,000 \$*”. This is modeled in IT-MBO by defining a KPI ι representing the aggregate cost of SLO penalty paid over the current three month period, represented by a dollar amount. The target⁵ of the objective is the region in the KPI space characterized by the inequality $\iota < 100,000\$$.

⁵ It has to be noted that in this example – as in all the examples given in this thesis - we define only one target region per objective. We do so in order to help the flow of the discussion and without loss of generality. However, the more general case will see more than one target per objective. An example of objective with multiple targets is one defining a first threshold of acceptability and a further threshold that represents a stretch goal. Example: revenues for the quarter must increase 15% year over year, with a stretch goal of 20% increase. When multiple targets are defined for an objective, the measure of alignment with the objective needs the definition of importance weights for all the target regions.

An example of perspective is a financial perspective, containing objectives such as the one listed above on the aggregate cost of SLO violations, or an objective that defines a target over a KPI representing the aggregated revenue generated in a given time period. A customer perspective could contain objectives defining targets over some KPIs representing quantitative measures of the customer satisfaction (measures of TCE: total customer experience), and so on.

Perspective	Financial		Customer
weight	80%		20%
Objective	Aggregated revenue in three month period	Aggregated cost of penalties for SLA violation in three month period	Total customer experience
weight	40%	60%	100%
adjusted weight for perspective	32%	48%	20%

Table 1: Perspectives, Objectives and Importance Weights

IT-MBO assigns importance weights to objectives and perspectives. As we will see later, the weights are used to compare business impact values over different objectives. The weight assigned to one perspective is propagated down to the objectives belonging to that perspective, as exemplified in **Error! Reference source not found.**

4.2 Aline: deriving Business Impact from Alignment with Business Objectives

Our alignment engine, Aline, reasons over the objectives and KPI defined through the IT-MBO information model in order to assign a value of alignment to a given course of action among the ones available to the IT manager. Because in the decision

support process we use this value of alignment to derive the business impact (see chapter 3) of carrying out a given course of actions, we require a formal and quantitative definition of alignment, which we give in the next subsection.

4.2.1 Alignment with business objectives as the likelihood that objectives will be met

We define the alignment with a given business objective as **the measure of the likelihood** – given the best knowledge about the current situation – **that the objective will be met.**

Before applying this definition of alignment, let us discuss why it behaves better for our purposes than other definitions of alignment that are sometimes used.

Let us recall the simple objective given in the previous section: “*the aggregate service revenue generated over the current three-month period (KPI) must be above 100,000\$ (target)*”. Let us suppose that 2 months into the period, the aggregate revenue figure amounts to 60,000\$.

A naïve measure of the alignment is derived by dividing the current amount by the target threshold, obtaining a figure for the ‘alignment’ of $60,000\$ / 100,000\$ = 60\%$. There are a number of problems with this definition. To begin with, the measure so defined does not take into account how far into the time period the aggregate value of penalties is measured. It is obvious to anyone that an organization is much better poised to meet (i.e. aligned to) its objective if the figure reads 60,000\$ one month into the period than two months. But the naïve definition of alignment would miss this.

Having disposed of this, one possible improvement is to take the time dimension into account and compare the revenue/time interval figure generated so far with the one that characterizes the target. In this case, the situation given in the original

example would amount to an alignment measure of $(60,000\$ / 2 \text{ mo}) / (100,000\$ / 3 \text{ mo}) = 90\%$. In the example where the figure of 60,000\$ refers to a one month period, the alignment measure would be 180%. The problems with this definition, in reverse order of gravity are:

1) it is difficult to associate an interpretation of a value for the alignment that exceeds 100% or results in a negative figure (which is possible with this definition);

2) it's really difficult to compare alignment across objectives (how to compare between an alignment of 1000% and 2000%?),

3) (*most importantly*) it does not require one to take into account foreseeable events that might impact the likelihood of the organization to eventually meet their objectives. About this third problem, suppose that in the example above (60,000\$ at 2 months, 90% aligned) it's known that the third and last quarter the revenue slows down because of seasonality of the business. Now the figure for the alignment derived with this method is completely useless, as it is evident that it will be utterly improbable that the organization meets its objective.

It's easy to see that our definition of alignment (Figure 4) behaves well against all the objections made to the alternative definitions. First off, the alignment always results in a figure between 0 and 1, which makes it easier to compare among alignment figures for different objectives. Most importantly, our definition copes well with the "seasonality" problem that was highlighted by the last variant of the given example. By reminding the reader that alignment is defined as the likelihood – to the best of one's knowledge – that the objective will be met, suppose an estimation is made that the revenue for the last month is uniformly⁶ distributed in the interval

⁶ Used here for simplicity of calculation

[0\$, 45,000\$]. The likelihood of meeting the objective is equal to the likelihood of posting a top line figure for the last month of the quarter in excess of 100,000\$ – 60,000\$ = 40,000\$, that amounts to $(45,000\$ - 40,000\$) / 45,000\$ = 11.11\%$.

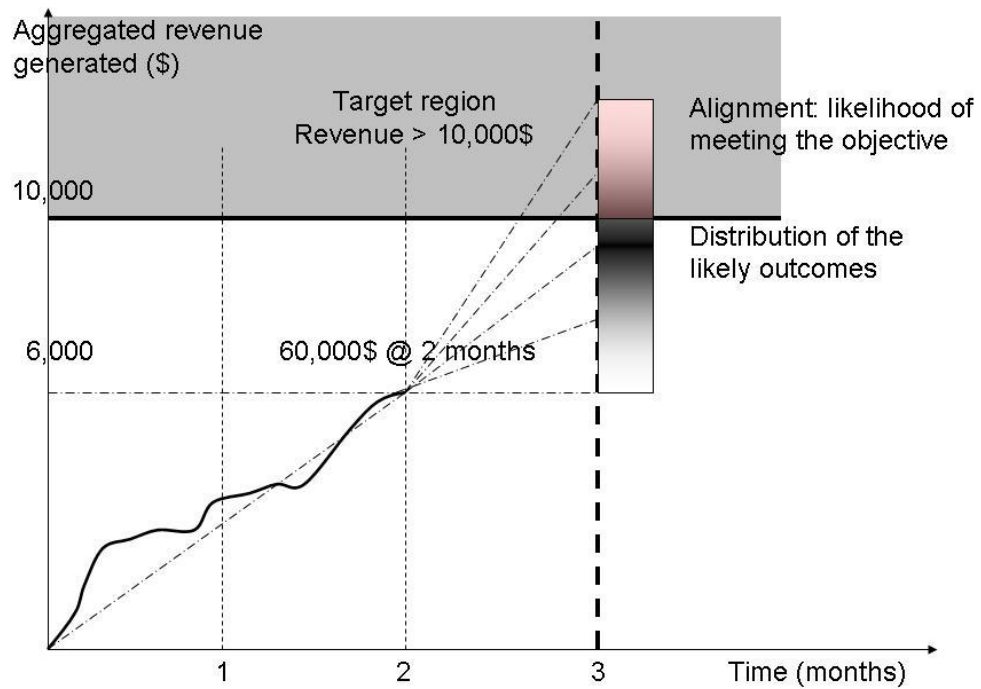


Figure 4: Alignment as likelihood of meeting the objective

From the definition of alignment used here, and the kind of business objectives that we consider, it follows that our method requires some estimate of the future value of the KPIs. The estimate is captured as a distribution of probability over the relevant KPI spaces. In the rest of this chapter, we refer to such an estimate as a likely outcome. An outcome is characterized by the distributions of probability over the KPI spaces that it entails.

Our working hypothesis is that the actual method that is used to estimate the likely outcome does not matter so much as long as there is a simple way to estimate the

likelihood of meeting the objective. Once again, the principle we follow is that we keep the cost of modeling low. Our hypothesis is that very sophisticated models will only add marginally to the accuracy in the computation of utility.

When building viable BDIM solution, it is important to validate this hypothesis by carrying out experiments to determine the sensitivity of any measure of the goodness of the decisions suggested to the complexity of the methods used for determining workable figures for the alignment. In the worked example above, three alternative methods to estimate the alignment of the seasonal-sensitive organization to its revenue objective could be:

- elicit knowledge from a business expert (or a pool thereof) through the question “how likely do you think it is that you’ll post a revenue figure in excess of 40,000\$ dollars this month
- use the something like the uniform distribution model given above
- use a more complex method that keeps into account a great number of variables, such as Box-Jenkins’ ARIMA [72].

Our conviction is that the simplest method will be “just good enough” for the analysis that the decision making engine will have to perform. Accuracy is not the most important quality of the prediction, as it would be for a system that predicts revenue for business managers, where an error greater than 3% would be considered bad. The most important quality required to the model in our framework is that it be **low cost**. In a case where the complex prediction models would give figures of say 11.11% and 13.245%, using a ”guesstimate” interval of say “15% to 20%” may still result in good enough suggestions. As we hint in figure 2, large increments in the complexity of the model (to which the cost of modeling is proportional) result in decreasing marginal quality of prediction.

The simplest model for prediction suggested above opens an interesting avenue for the usage of prediction markets, or information markets in IT management (see text box).

Prediction markets (adapted from wikipedia.org)

Prediction markets (also known as **predictive markets, information markets, decision markets, idea futures, event derivatives, or virtual markets**) are [speculative](#) markets created for the purpose of making predictions. Assets are created whose final cash value is tied to a particular event (e.g., will the next US president be a Republican) or parameter (e.g., total sales next quarter). The current market prices can then be interpreted as predictions of the probability of the event or the expected value of the parameter. Prediction markets are thus structured as [betting exchanges](#), without any risk for the bookmaker.

People who buy low and sell high are rewarded for improving the market prediction, while those who buy high and sell low are punished for degrading the market prediction. Evidence so far suggests that prediction markets are at least as accurate as other institutions predicting the same events with a similar pool of participants.

Many prediction markets are open to the public. [Betfair](#) is the world's biggest prediction exchange, with around \$28 billion traded in 2007. [Intrade](#) is a for-profit company with a large variety of contracts not including sports. The [Iowa Electronic Markets](#) is an academic market examining elections where positions are limited to \$500. [TradeSports](#) are prediction markets for sporting events. [The simExchange](#), [Hollywood Stock Exchange](#), [NewsFutures](#), the [Popular Science Predictions Exchange](#), [Hubdub](#), [The Industry Standard](#)'s technology industry prediction market, and the Foresight Exchange Prediction Market are virtual prediction markets where purchases are made with virtual money. [Bet2Give](#) is a charity prediction market where real money is traded but ultimately all winnings are donated to the charity of the winner's choice.

4.2.2 Using alignment with business objectives to derive a business impact function

The objective used in the example above was useful to compare our definition of alignment with alternative ones, though not so directly relevant to IT service management. Let's now switch to an example that is more relevant to ITSM to see how we use alignment with business objectives to derive a business impact function to rank alternative management options. The objective we consider here is "the

aggregate cost of paid penalties for not meeting SLOs over the current three months period (KPI) must be below 10,000\$ (target)''.

We consider a situation where, two months into the period, the aggregated penalty cost amounts to 6,000\$. For the sake of this example, we'll use a prediction model that assumes that the aggregated penalty cost over the last month of the quarter is uniformly distributed in [1500\$, 4500\$] (but a guesstimating oracle would do just as fine, as discussed above). The current measure of the alignment is therefore computed to 83.33% (the probability that the aggregated cost of penalty in the last month will exceed 4000\$). At this point an incident occurs that is likely to disrupt the service being provided to a customer. The penalty associated to the service disruption amounts to 2000\$. The options available to the IT manager are defined by the priority value to assign to the incident. If a) the incident is dealt with the highest priority, it is expected to result in a 25% probability of incurring in the penalty. If b) the incident is dealt with with lower priority the likelihood of ending up breaching the SLA is 75%.

Assuming risk-neutrality, these two options are characterized by expected new values of aggregate cost of penalty of a) $6,000\$ + 25\% * 2,000\$ = 6,500\$$ and b) $6,000\$ + 75\% * 2,000\$ = 7,500\$$. Assuming independence of incidents, the distribution of the aggregated cost of penalty for the last month is unchanged. The resulting alignment resulting by acting on each option is therefore 66.66% and 33.33% respectively. For option a), the alignment is equal to the likelihood that the aggregate cost for the last month does not exceed $10,000\$ - 6,500\$ = 3,500\$$. For b) the threshold figure is 2,500\$.

Recalling from chapter 3 that we use business impact with a negative connotation (cost minus benefit), we derive business impact from the complement to one of the alignment to one business objective:

$$impact = 1 - alignment$$

It is easy to see that our chosen **business impact** function represents – given the best knowledge about the situation – **the likelihood that the objective will not be met**. The lower the value of alignment, the higher the business impact that is felt and vice-versa.

By applying a linear combination of the business impact of the option over the set of the objectives - each taken with its given weight representing the relative importance of the objective to the business – we obtain a (Von Neumann – Morgenstern, see text box below) utility function that we can use to rank the management options.

To exemplify, assume that three objectives were defined, and given weights representing their relative importance. After calculating the alignment of the two options given above with respect to all the objectives, a total business impact value is calculated for each of the options (Table 2).

Objectives	Aggregate revenue	Aggregate penalty	Customer satisfaction	Utility
Weights	32%	48%	20%	
Alignment of Option a: <i>assign high priority</i>	86%	67%	90%	Score: 0.7768 Alignment: Good
Alignment of Option b: <i>assign low priority</i>	86%	33%	50%	Score: 0.5536 Alignment: Fair

Table 2: Calculation of business impact for alternative options

Von Neumann – Morgenstern utility functions (adapted from wikipedia.org)

In older definitions of utility, it makes sense to rank utilities, but not to add them together. A person can say that a new shirt is preferable to a baloney sandwich, but not that it is twenty times preferable to the sandwich.

The reason is that the utility of twenty sandwiches is not twenty times the utility of one sandwich, by the [law of diminishing returns](#). So it is hard to compare the utility of the shirt with 'twenty times the utility of the sandwich'. But Von Neumann and Morgenstern suggested an unambiguous way of making a comparison like this.

Their method of comparison involves considering probabilities. If a person can choose between various randomized events (lotteries), then it is possible to *additively* compare the shirt and the sandwich. It is possible to compare *a sandwich with probability 1*, to *a shirt with probability p or nothing with probability 1-p*. By adjusting p, the point at which the sandwich becomes preferable defines the ratio of the utilities of the two options.

A notation for a *lottery* is as follows: if options A and B have probability p and 1-p in the lottery, write it as a linear combination:

$$L = pA + (1 - p)B$$

More generally, for a lottery with many possible options:

$$L = \sum p_i A_i.$$

By making some reasonable assumptions about the way choices behave, von Neumann and Morgenstern showed that if an agent can choose between the lotteries, then this agent has a utility function which can be added and multiplied by real numbers, which means the utility of an arbitrary lottery can be calculated as a linear combination of the utility of its parts.

We cannot stress enough the fact that the business impact values are only useful as an indication of which option will be preferable to achieve a better alignment with the business objectives. The values so obtained are not meaningful per se: they only make sense as an arbitrary utility value used to rank the options. This observation reinforces

the hypothesis of the relative unimportance of the particular method used for prediction to the final decision.

An alternative way of perceiving the utility value of the available options is through a monetization process, which states a mapping between the utility values calculated through the alignment definition, and a monetary measure of the perceived goodness of the option. This is useful in that it allows instant comparison with measures of the monetary cost of executing the option. However, it's very important to note that the output of the monetization process is not meant to be an accurate monetary evaluation for the option, but rather just an input to the ranking process of the available options.

Without loss of generality, whatever the method chosen for forecasting the value of the KPIs at the end of the period, we will indicate the default outcome with $pdf_{default}(kpi)$.

4.2.3 Modeling cost as one of the dimensions of alignment

Recall from section 3.1.3 that there are three ways of combining impact and cost: 1) monetize intangibles: associate a monetary cost to figures of business impact; 2) understand business impact of financial cost: treat the financial dimension as one of the possible way to optimize the business; 3) define a suitable utility function that contains aspects of both, such as a linear or convex combination of suitably derived quantitative measures.

With our method, we choose the second option: that is **we treat the financial dimension as one of the possible ways to optimize the business**. In the cost of penalty associated to SLA violation example that we gave above, it is clear how a KPI can be defined over a financial dimension and how a value of alignment can be computed based on the definition of a target value. The usual objection to this method is that for financial measures “more money” is always better than “less money”.

While that is obviously a truism, we counter that when looking at a balanced scorecard, the perceived utility for financial measures is often subjected to a big jump when crossing a target region threshold.

Moreover we observe that our method mitigates the quantization effect over financial KPIs by allowing the definition of multiple targets for one KPI, corresponding to the concept of *stretch goals*. In this way, it is possible for example to define two successive targets over the objective in the example above: cost of violation should be less than 10,000\$ (base goal) and cost of violation should be less than 8,000\$ (stretch goal). These are treated to all effect as separate objectives even though the obvious dependencies between them are exploited in the calculation of alignment.

4.2.4 Modeling the impact of courses of action on the business objectives

In our path towards building viable business-driven IT management decision support solution is a way of modeling the impact of alternative courses of action on the business objectives. In this sub-section we describe the core of the alignment computation of our methodology, implemented by Aline.

To follow our principle of keeping the cost of modeling low, we have to be very selective in choosing the relevant information to model. In general, the complexity required in modeling the effect that actions taken by the IT manager have on the KPI - and consequently on the business objectives – can quickly grow beyond acceptability. The main inventive step is to break down the chain of dependencies between actions and KPI values by inserting an intermediate step in the middle. We identify **episodes** that can have an impact on the KPIs. Unlike the actions that can be taken, the episodes that we model are usually of few different types and can be described quite easily in terms of the metrics underlying the KPIs. For the sake of an example, let us now

introduce a couple of business objectives and their relative KPIs that we will refer to time and again within this chapter. The first objective states that the aggregate cost of penalty for SLO violation in a three month should be less than 10,000\$. The second objective states that the total number of SLO violations for all customers in a three month period should be less than 15. We note here that whatever the actions that an IT manager can take, the only relevant episodes to the KPIs above described are SLO violations. On identifying the relevant episodes, the calculation of the alignment for any given course of action is therefore split in two steps:

- 1) **estimate the likelihood that a given course of action will result in a relevant episode** (e.g. SLO violation) will take place;
- 2) calculate the alignment with the business objectives both when the episode takes place and when it does not; and use these values to **compute a final measure of the alignment given the likelihood of the occurrence of the episode** that was computed in the previous step.

The method here described results in much simplified calculations because the only information that is required for a given course of action is what is needed to estimate the likelihood of the episode. In the second part of this thesis we will fully work out an example of how to effectively utilize Aline in building a solution for incident management. Here we skip the first step and assume that likelihood estimates will be available that link a given course of action with a given episode. We describe in detail the calculations prescribed by the methodology to compute the final measure of alignment.

First, we need to model the impact that our episode, a **likely SLO violation**, has on the KPIs described above. The description of the impact has an uncertainty component, therefore lends itself better to be expressed through probabilistic

measures, and Aline is able to cope with that. However, here and in the following we use deterministic functions to express that dependency, as it's the character of our example and because it gives us simplicity of exposure. For example, when considering the effect of a likely SLO violation on the aggregated cost of penalties KPI, we represent it through the function

$$f_{impact}(kpi_{penalty}) = kpi_{penalty} + penalty(SLO)$$

The formula is intended to mean that when a SLO violation occurs, the value of the KPI measuring the aggregate cost of penalty is increased by the penalty relative to the SLO violation. Similarly, considering a customer related KPI that measures the the number of violations experienced by all customers in a period of time, we write:

$$f_{impact}(kpi_{violation}) = kpi_{violation} + 1$$

The outcome that follows an SLO violation is therefore obtained by composing the effect of the SLO violation with the default outcome (no SLO violation) as indicated by the following formula

$$pdf_{violation}(kpi) = pdf_{default}(f_{impact}^{-1}(kpi))$$

We now determine the likely outcome of closing an incident by a given time. We have already determined the likelihood of SLO violation λ in function of the time taken to close the incident. The likely outcome is given by the combination of the default outcome if the violation doesn't occur with the modified outcome if the violation does occur.

$$pdf_{combined}(kpi) = (1 - \lambda) pdf_{default}(kpi) + \lambda pdf_{violation}(kpi)$$

The method followed is illustrated in Figure 5.

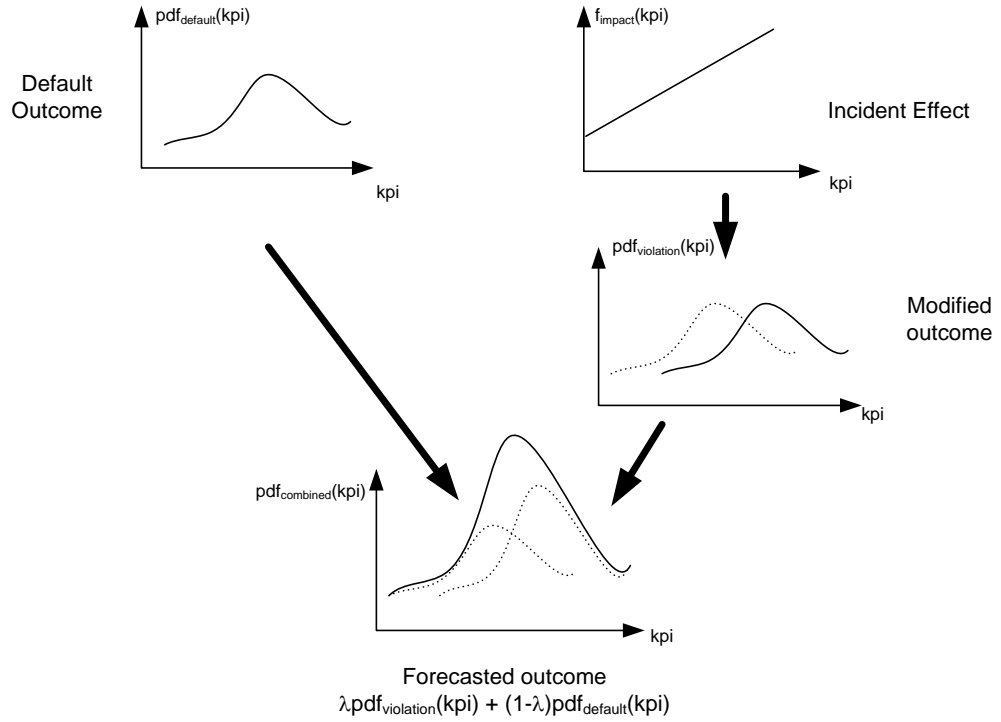


Figure 5: Composition and combination of likely outcomes

Computation of alignment

To compute the alignment of the forecasted outcome with a given business objective, we simply integrate the probability density function for the outcome within the target region defined by the objective in the KPI space (illustrated in Figure 6).

$$\text{alignment}_{\text{objective}} = \int_{kpi \in \text{Target}(\text{objective})} \text{pdf}_{\text{combined}}(kpi) dkpi$$

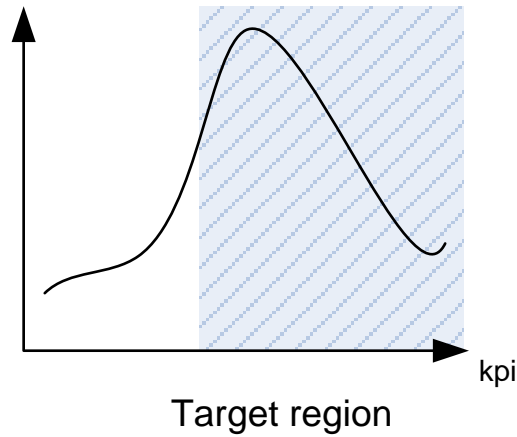


Figure 6: Integrating the probability density function over the target region

The alignment of the outcome with all the business objectives is finally obtained by summing the contribution of the alignment with each of the business objectives, each taken with their own importance weight.

$$alignment = \sum_{\forall objective} weight_{objective} \cdot alignment_{objective}$$

It can be noted that because of the linearity of the combination operated above, we also could first independently compute the default alignment and the alignment in case of violation; and later linearly combine them with the likelihood that the violation will in fact happen. The end result does not change, and in this way the calculations are much simpler and faster to carry out, since it is much simpler to deal with scalars than with probability density functions, and the number of integrations required is dramatically reduced.

4.2.5 Implementation of Aline using the Mathematica™ kernel and J/Link toolkit

Since the Aline core algorithm makes extensive use of numerical integrations, on a first version of the prototype we had considered a limiting the definition of probability

distribution functions to piecewise linear function, as that considerably simplified the calculations involved.

However, we eventually implement a version of Aline that is able to cope with arbitrary probability distribution functions of any type that can be defined in Mathematica™ by Wolfram Research [73]. Aline's java code interfaces and makes call into the Mathematica™ kernel to perform the numerical integrations described above using the Wolfram J/Link toolkit [74].

4.2.6 Building BDIM solutions for IT service management using IT-MBO and Aline

Before we conclude this chapter, we show how the IT-MBO information model and the Aline can be used to build BDIM solutions for decision problems in IT service management. In order to employ the methodology and the tools in a full BDIM solution, IT-MBO and Aline need to be complemented with corollary components that are specific to the domain under study. In particular, we need a **forecasting module** that can estimate the likely outcomes in terms of KPI value (what we called the default outcome in the section on impact modeling). Once Aline is given the default outcome then it can compute the alignment for each alternative course of action, but the default outcome is necessary for bootstrapping the calculation.

The problem of forecasting from time series is well studied in the literature [72]. It has to be noted that, since the output of the forecasting module is only one of a chain of steps that Aline goes through, accuracy in the prediction is not so important here.

For some domains where the space of possible courses of action is limited, the alignment engine could be used in a brute-force mode to calculate the alignment of each of the options and choose the course of action which results in the optimal alignment with the business objectives. However, life is not always so simple. When

the number of alternative courses of action is susceptible of combinatorial explosion (as we will see is the case for incident prioritization), the approach that we take is to model the decision problem according to multi-attribute utility theory ([29]). In cases like these, it will be necessary to develop a **solver** module that tackles the decision problem using the output of Aline as its input to calculate the utility of alternative options as alignment with the business objectives.

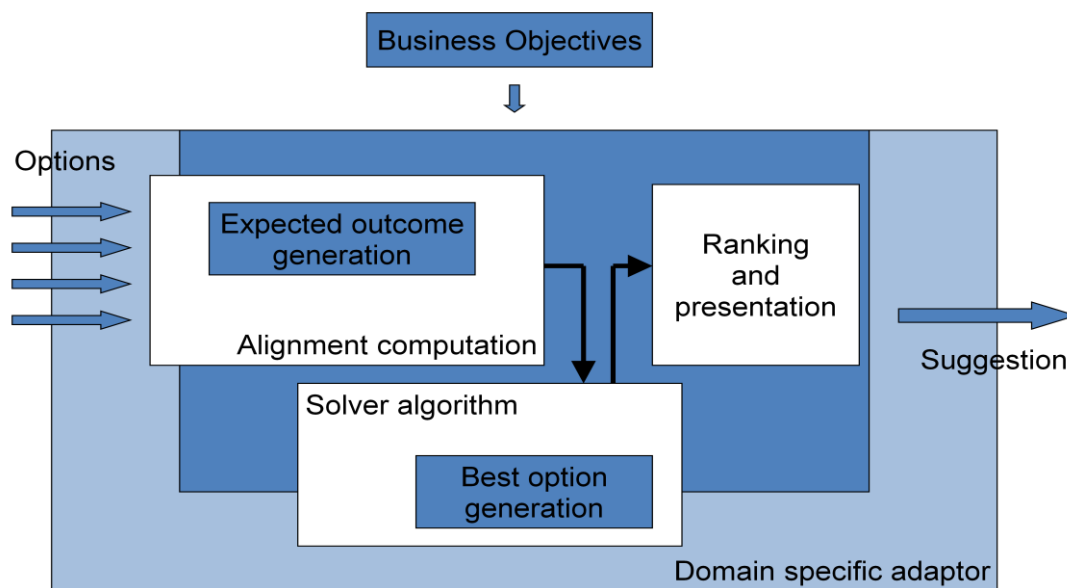


Figure 7: Conceptual architecture of a solution built around Aline

The conceptual architecture of a solution built around Aline is represented in Figure 7. The alternative options that are available to the IT manager are passed on to Aline (the alignment computation module) to obtain a value of utility. Aline in turn uses a forecasting module (expected outcome generation) to calculate the value of alignment according to the method described above. The output of Aline is then used by a solver module that determines the best option among the ones available.

In part II of this thesis (chapter 5) we will describe the specialized solver and forecasting module that are necessary for a fully fledged BDIM solution for incident prioritization.

4.3 Discussion

In this chapter we presented our methodology for dealing with IT-business linkage, and a set of artifacts that embody it, building on the framework presented in the previous chapter. We showed how to create business impact functions that are not arbitrary, but based on a quantitative figure of alignment with a set of business objectives.

The IT-MBO information model and the Aline alignment engine can be used as reusable components for building BDIM solutions. However, in order to employ our methodology in a full solution, these reusable components need to be complemented with corollary components that are specific to the domain. In particular we need a forecasting module that can estimate the likely outcomes in terms of KPI value (what we called the default outcome in the section on impact modeling). In the second part of this thesis, starting from the next chapter, we will show how to build such solutions.

The main way in which our methodology advances the state of the art is that it provides a coherent way of dealing with both tangible and intangible objectives in a way that is natural to people who are used to instruments such as the balanced scorecard. The added advantage of our methodology is that the objectives that are used to compute alignment (and therefore business impact) are the same that IT executives are used to negotiating over. This way of dealing with eliciting preference is therefore far preferable to arbitrary monetization of intangible objectives as in previous works in the BDIM literature. It is more natural to argue over the relative importance of a small number of objectives than it would be to argue over any possible arbitrary monetization of outcomes.

Part II: Solutions for Business-driven IT Management

5 A representative BDIM solution for decision problems in IT service manager: business-driven prioritization of incidents

Having laid out the bases in part I of this thesis for building decision support tools for BDIM decision problems in the IT service management space, in this chapter and in the next we will demonstrate how the decision theoretical BDIM framework (chapter 3), and the Aline alignment computation engine (chapter 4) can be put to work to build such solutions.

In this chapter we present the first instance of such solution: a decision support tool for the prioritization of service incidents. This simple, self-contained BDIM solution can be seen as a template for building BDIM solutions to recurrent decision problems in IT service management. In [43] and [44] we presented another such example, applied to the problem of scheduling IT changes, which we will not present in detail here. In the next chapter we will present a much more comprehensive BDIM solution aimed at optimizing the performance of an IT support organization in its incident management process and help-desk function.

ITIL [3] 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 restore the services in the quickest possible way. 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.

In the incident management process it is of fundamental importance to classify, prioritize and escalate incidents. ITIL suggests that priority of an incident be 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, 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.

The BDIM solution for incident prioritization that we present in this chapter assigns priority levels to a set of service level degradation incidents so as to maximize the alignment with a given set of business objectives. This component is described in more detail in [5], specializing the model for business-IT linkage described in chapter 4.

The prioritization mechanism exploits Service Level Agreements (SLAs) contracted with customers, each containing a set of Service Level Objectives (SLOs), and suggests how to deal with the incidents so as to maximize alignment with a set of business objectives (that in turn are linked to metrics related to SLO violations).

5.1 Proof of concept BDIM solution for IT incident prioritization using IT-MBO and Aline

The problem solved our BDIM solution for the incident prioritization problem is to **assign priority levels to a set of service level degradation incidents so as to maximize the alignment with a given set of business objectives**⁷. Let us begin by recalling some of the lexicon that we use during the description of the incident prioritization solution.

A Service Level Agreement (*SLA*) is contracted with a *customer* and contains a set of *Service Level Objectives (SLO)*. Each SLO specifies an acceptable range of values for a given system metric, through the definition of a *violation threshold* for the SLO. A *penalty cost* is associated to SLO violation, which occurs when the metric value surpasses the violation threshold. Besides the violation threshold, a *jeopardy threshold* is also specified. Metric values are obtained by probes deployed by the management system and monitored by a monitoring component. A *degradation of service level incident* for an SLO occurs when the monitoring component reports on a metric value surpassing the jeopardy threshold for the SLO. An incident management system collects and organizes the information on the degradation of service level incidents by assigning priority values to them together with other information on the lifecycle of the trouble ticket associated to the incident. The problem that we solve is to suggest how to deal with the incidents so as to **maximize the alignment with a set of business objectives**. In this work we only consider incidents generated on

⁷ Expressing this once again in terms of our decision theoretical framework described in chapter 3, we describe a business impact function in terms of the complement to 1 of said alignment, and then we use that as utility function to solve optimally the incident prioritization problem.

detection of service level degradation or violation, although the general techniques that we present are more widely applicable.

To solve the incident prioritization problem, we operate on the following steps:

1. Compute the likelihood of violation of an SLO in function of the time taken to close a jeopardy incident, using an ad-hoc module developed for the incident prioritization problem;
2. Compute the alignment with the business objectives in function of the likelihood of violation, using Aline;
3. Prioritize the incidents in function of the alignment with the business objectives, using a solver module based on Integer Linear Programming (ILP).

5.1.1 Likelihood of SLO violation as a function of the time of closure of the service degradation incident

We make the assumption that the IT manager is required to specify a time value that represents the expected time that it will take for the system to move from the jeopardy state to the violation state if no measures are taken (expected time from jeopardy to violation). We assume an exponential distribution of the time from jeopardy to violation if no corrective actions are taken. The parameter of the exponential distribution, λ , is defined as the inverse of the expected time from jeopardy to violation.

The plot in Figure 8 represents the cumulative distribution function of the distribution of probability associated with the time from jeopardy to violation. The analytic form of the curve is given by the equation $p = 1 - e^{-\lambda t}$, where p represents the probability of violation if the incident is closed after t from its starting time. In the example given, for a value of $\lambda = 1/3$, corresponding to an expected mean time of 3

hours, the probability of a violation occurring within 4 hours from the jeopardy alarm if no corrective measures are taken equals to $1 - e^{-4/3} = 76\%$.

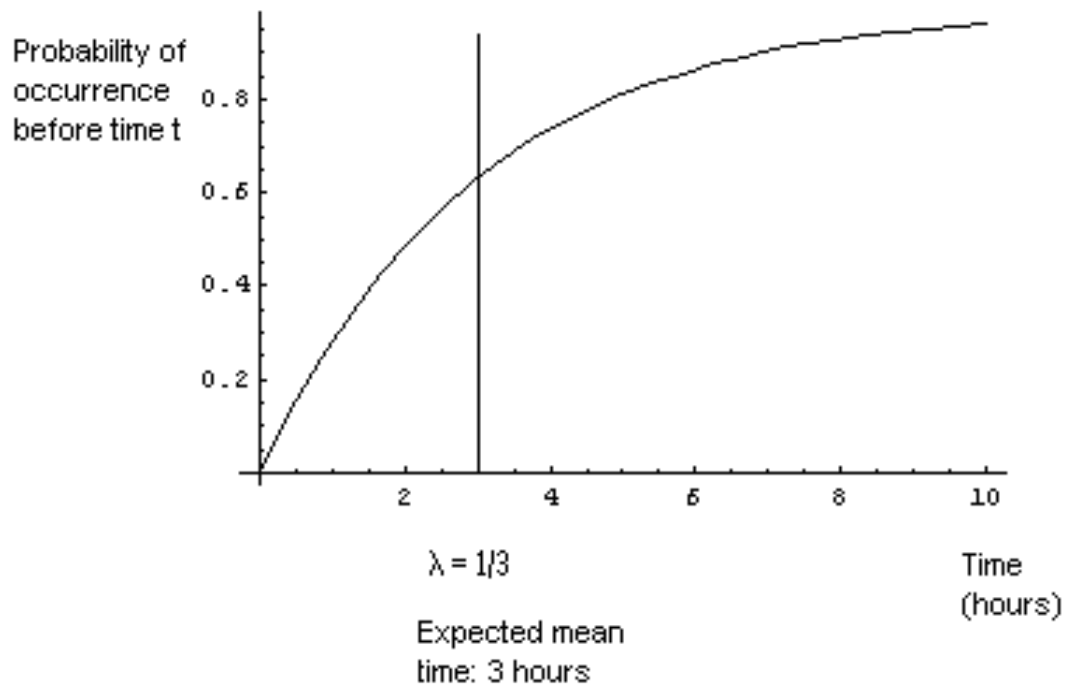


Figure 8: Cumulative probability distribution function of time from jeopardy to violation

5.1.2 Alignment with the business objectives in function of the likelihood of violation

Aline is invoked to compute the alignment with the business objectives. The alignment computation steps are described in detail in the previous section on the alignment engine. Here we exemplify the computation by presenting a fully worked out example.

Worked example of alignment computation

Let's consider a service degradation incident that relates to an SLO for which the expected time from jeopardy to violation is three hours. From the previous subsection, the likelihood of SLO violation expected from closing the incident in four hours is given by $1 - e^{-4/3} = 76\%$. The penalty associated to violating the incident is set at 1,000\$.

Let's assume that two simple business objectives have been defined. The first objective states that *the aggregate cost of penalty for SLO violation in a three month should be less than 10,000\$*. The second objective states that the total number of SLO violations for all customers in a three month period should be less than 15. The objectives are deemed to be equally important, therefore each carrying an importance weight of 1/2.

Two months into the period, the current readings of the KPIs are 6,000\$ in penalties paid for 10 SLO violations. For simplicity's sake and without loss of generality we here assume that the forecasting module predicts a default outcome characterized by uniform distributions with extremes [7,500\$, 10,500\$] for the penalty KPI and {13, 14, 15, 16, 17} for the violations KPI.

The alignment with the business objectives entailed by the default outcome is computed as the probability of meeting the objectives given the expected default outcome.

For the first objective we obtain an alignment measure of 5/6 (such is the probability that the penalty KPI wind up being below 10,000\$ given that it's distributed uniformly in [7,500\$, 10,500\$]). For the second objective it's obvious to observe that the alignment measure is 2/5. The alignment with all business objectives is therefore $1/2 * 5/6 + 1/2 * 2/5 = 31/60 = 0.517$ or 51.7%.

The effect of the likely SLO violation on the KPIs is to increase the penalty figure by 1,000\$ and the number of violations by 1. This would therefore define an outcome characterized by uniform distributions with extremes [8,500\$, 11,500\$] for the penalty KPI and {14, 15, 16, 17, 18} for the violations KPI. If the violation were to occur, that would entail level of alignment of $1/2$ and $1/5$ for each objective respectively, and therefore an alignment with all business objectives of $1/2 * 1/2 + 1/2 * 1/5 = 7/20 = 0.350$, or 35%.

Since the violation is expected to occur with a 76% likelihood, the measure of alignment for the combined case becomes: $24% * 0.517 + 76% * 0.350 = 0.390$, or 39%.

Once again, the reader should not read too much into the actual figure for the alignment, besides considering it a useful way to discern among alternatives.

Forecasting

As far as the forecasting module is concerned, the method is only relatively sensitive to the accuracy of the prediction of the forecaster, as we noted above. This means that that MBO's suggestions are good if the accuracy is just good enough.

The simple method that we have chosen for this system predicts the mean value of the KPI at the end of the period simply by extrapolating its current value. The forecasted value of the KPI is considered to be normally distributed with mean calculated as above and variance set at a sensible customized value (for example the square of 5% of the mean value). For the example given above, when the revenue KPI reads 60,000\$ two months into the period, we will forecast its value at the end of the three month period to be normally distributed with mean 90,000\$ and a variance of $(4,500\$)^2$, and therefore characterized by

$$pdf_{default}(kpi_{revenue}) = NormalDistribution(90,000\$, (4500\$)^2)(kpi_{revenue})$$

A full validation of the solution would require experimenting with alternative forecasting modules and carrying out sensitivity analysis experiments to assess the goodness of the prioritization that we obtain. However, since our main objective in this chapter was simplicity of exposure in explaining a template method that could be applied to similar BDIM solutions in other IT service management domains, we'll leave it at that.

5.1.3 Incident prioritization to maximize alignment with business objectives

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 [3] 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 maximum closing time for the incidents that are assigned to a certain priority level.

The incident prioritization problem

We here present a mathematical formulation of the incident prioritization problem as an instance of the **generalized assignment problem**. The generalized 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 [75].

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 alignment 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 alignment yielded by assigning the k^{th} incident to the j^{th} priority level is given by the alignment of closing the incident by the time t_j , which we know how to compute from the previous sections. We'll call this measure of alignment $a(i_k, t_j)$ for short

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 $n=200$ incidents, the requirement that at least 40% of the incidents will be assigned medium priority (assume that is priority level p_2) would read:

$$\sum_{k=1}^n x_{2k} \geq 80$$

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

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

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

$$(IPP) \quad \min \quad \sum_{j=1}^m \sum_{k=1}^n a(i_k, t_j) \cdot x_{jk} \quad (2)$$

$$\text{s.t.} \quad \sum_{k=1}^n x_{jk} \geq c_{jk} \quad \text{and} \quad \sum_{k=1}^n x_{jk} \leq C_{jk} \quad \forall j = 1..m \quad (3)$$

$$\sum_{j=1}^m x_{jk} = 1 \quad \forall k = 1..n \quad (4)$$

$$x_{jk} = 0 \text{ or } 1 \quad \forall j = 1..m, k = 1..n \quad (5)$$

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

Incident prioritization algorithm

Here is a pseudo-code description of the incident prioritization algorithm

Input:

A number of SLOs, each modeled with the following information:

- Violation threshold for an SLO
- Jeopardy threshold for an SLO
- Penalty cost for SLO violation
- Expected time from jeopardy to violation (if no corrective measures are taken)

A set of priority levels for incidents, modeled with the following information:

- Constraints over the number of incidents to be assigned to each priority level
- Expected maximum closing time of incidents assigned to the priority level

A number of service degradation incident, modeled with the following information:

- Incident start time (the time when a jeopardy alarm was raised)

Output:

A complete prioritization of the incidents that assigns to each of them a priority level

Steps:

1. Compute the default alignment with business objectives (section 4.2.2)
2. For each incident:
 - 2.1 For each priority level:

2.1.1 Compute the likelihood of violation of this incident if assigned to this priority level (section 2, using the expected maximum closing time for the priority level)

2.1.2 Compute the alignment with business objectives if a violation occurs (section 4.2.4)

3. Solve the incident prioritization problem (section 4, using the values of alignment derived from the previous step)

5.2 Demonstrator of the BDIM solution for Incident prioritization

In this section we describe a demonstrator of BDIM solution for incident prioritization. There are two roles in our story: **business manager** and **IT manager**. The responsibility of the business manager is to define business objectives, their importance weights, and the key performance indicators (KPIs) over which the objectives are based. The IT manager is responsible for the configuration and use of the incident management solution. Behind the scenes, and taking a secondary importance in this story, there is a third role, the **solution architect**. The solution architect is responsible for the design and maintenance of the incident management solution by ensuring proper customization of the Aline and development and maintenance of corollary components such as the *KPI forecaster*.

The story is narrated in three parts. The first part is about the **business view** of the solution. This is where the business manager defines business objectives, their importance weights and the KPIs. The second part is the view of the **solution architect**, who helps setting the link between business objectives and relevant IT episodes. Finally, the third part is about the configuration and usage of the incident prioritization solution by the **IT manager**. Notice that the first and second part of the demo could equally apply to a different solution for some other domain of IT service

management, such as change management for example. The third part is specific to the incident prioritization application.

5.2.1 Part I: Business view of the solution

Define business objectives

The business manager defines business objectives according to the IT-MBO information model (section 4.1). Each business objective is defined through the definition of a target region for a key performance indicator. The objectives are given weights that are used by the Aline engine to derive the overall alignment of the various options with all the business objectives.

In our example two business objectives are defined (Figure 9). The first states that *the aggregated number of violations for all customers over a three month period has to be less than 23*. The second says that *the aggregated cost of penalty for all customers over the same period has to be less than 1300k\$*. In the example the first objective is given an importance weight of **0.8**, and the second **0.2**.

Define key performance indicators

The business manager defines key performance indicators (KPIs) over metrics, the values of which will be collected in our demo by accessing a database and showing the operation of aggregation.

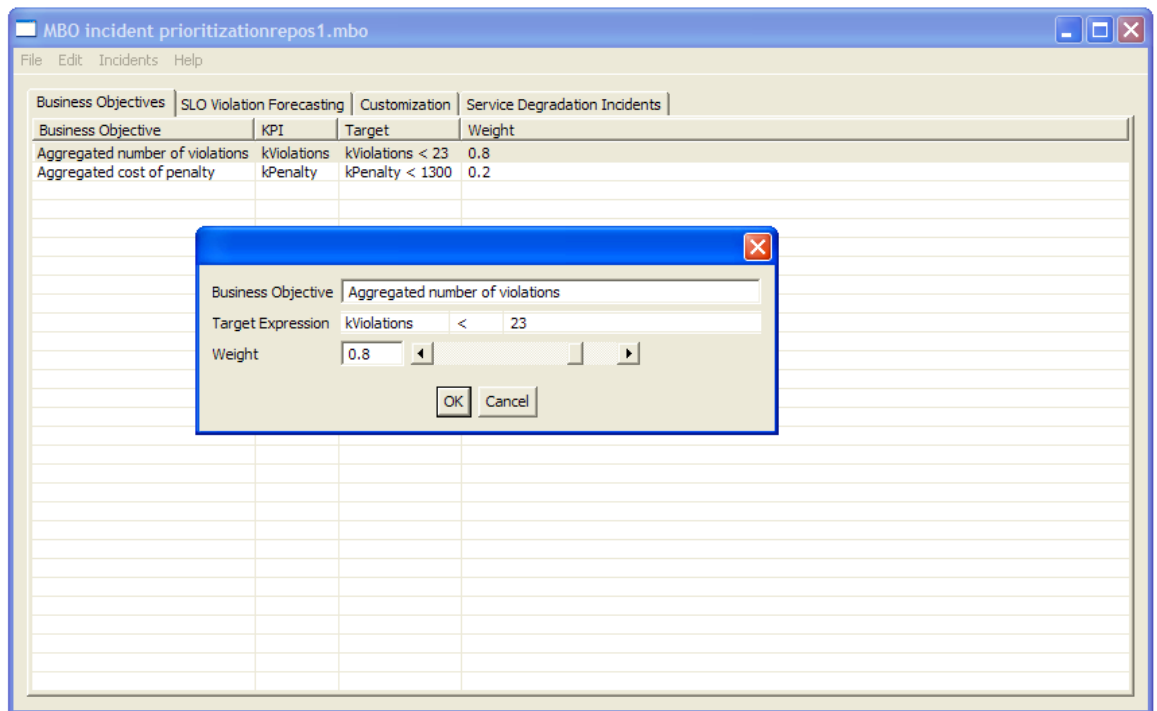


Figure 9: Define business objectives

Forecast SLO violations

The solution architect – outside the scope of this demo – had developed a forecaster module, which is used to forecast the values of the key performance indicators. In this demo we show the forecast KPI values. In the example in , for two months, we have experienced **15** SLO violations for an aggregated penalty figure of **800,000\$**, as can be seen in Figure 10.

The forecaster can then show the expected values of the KPIs at the end of the period, together with the uncertainty around them. In our example, it returns **1,250,000 ± 10%** for kPenalty and **22 ± 2 for kViolations**.

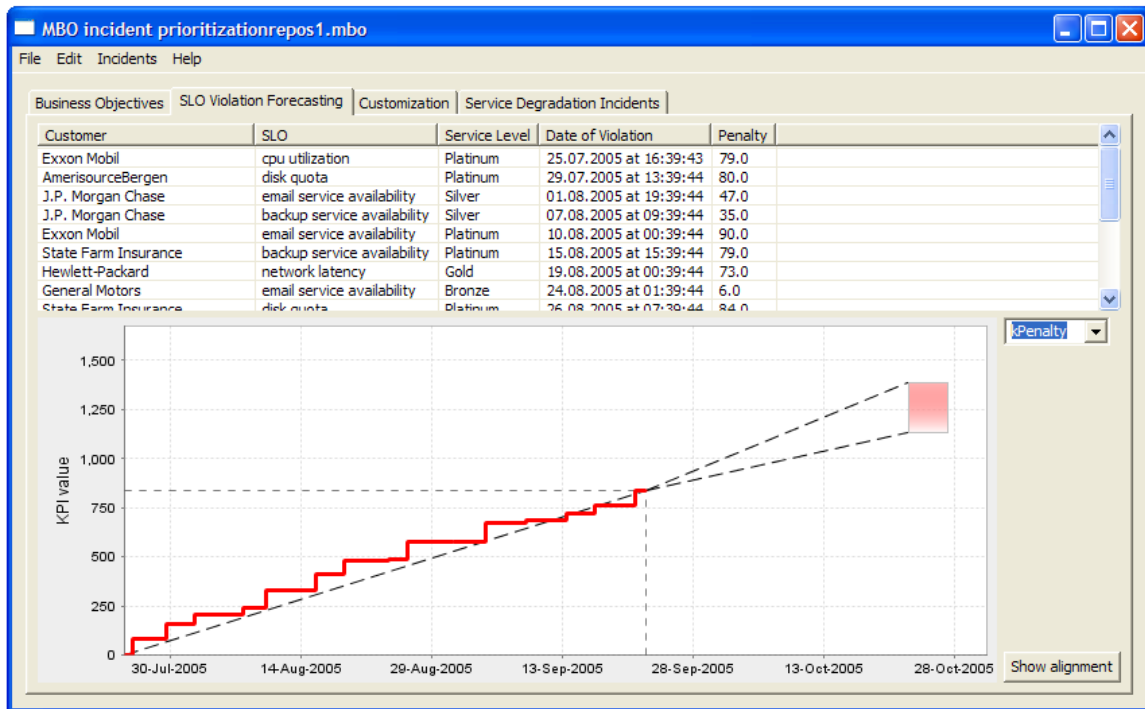


Figure 10: SLO violations forecasting

Current alignment

The output of the forecaster is then used to compute the current alignment with the business objectives. The current alignment is going to be used by the alignment engine as a basis for calculating the loss in alignment brought about by each of the possible alternatives – in this case alternative incident prioritizations – that in turn represents the utility of the alternatives.

The graph in Figure 11 shows a graphical representation of the computation of the current alignment based on the time series of the values of the KPIs, the target region defined by the objectives and the probability distribution that is the output of the forecaster.

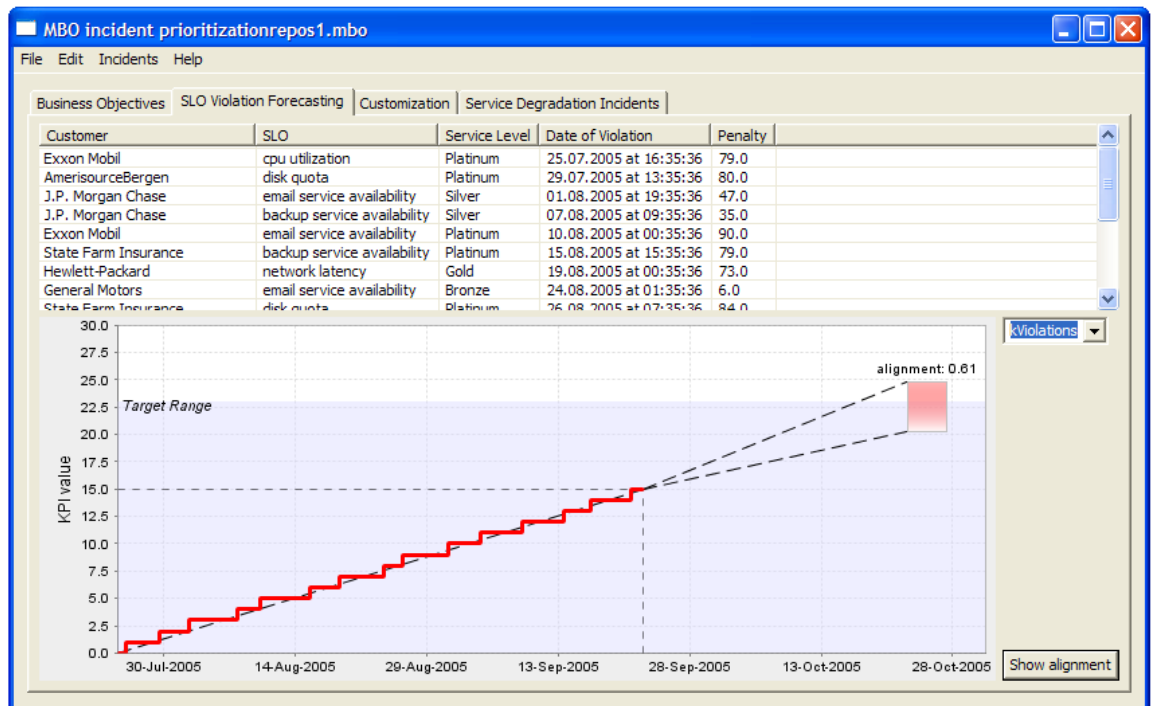


Figure 11: Representation of current alignment

If business objectives are defined, current alignment may be shown on “SLO Violation Forecasting” tab by pressing “Show alignment” button.

5.2.2 Part II: Configuration and use of the incident prioritization solution

Define episodes that have impact on the key performance indicators

The solution architect – incidentally to the flow of this demo – defined the impact that *episodes* (defined section 4.2.4) have on the KPIs. In our example we will consider business objectives defined over key performance indicators that are impacted by SLO violations. Therefore our episodes will be SLO violations. For each episode, the impact on the relevant key performance indicators is expressed through a formula as in the GUI screen simulated by the following spreadsheet.

5.2.3 Part III: Configuration and use of the incident prioritization solution

Once objectives, KPIs and their effect have been defined, we take a look at the current set of incidents to be prioritized. Suppose that the situation that is presented to the IT manager is as in Figure 12.

The screenshot shows a software window titled "MBO Incident Prioritizationrepos1.mbo". The window has a menu bar with "File", "Edit", "Incidents", and "Help". Below the menu bar are several tabs: "Business Objectives", "SLO Violation Forecasting", "Customization", and "Service Degradation Incidents". The "Service Degradation Incidents" tab is active, displaying a table with the following columns: "Customer", "Service Level", "SLO", "Arrival Time", "Time to violation", and "Priority". The table contains 25 rows of incident data. At the bottom right of the window, there is a button labeled "Prioritize incidents".

Customer	Service Level	SLO	Arrival Time	Time to violation	Priority
Exxon Mobil	Platinum	cpu utilization	21.09.2005 at 21:14:27	16:04	
J.P. Morgan Chase	Silver	email service availability	22.09.2005 at 09:22:16	18:52	
Exxon Mobil	Platinum	email service availability	22.09.2005 at 08:55:06	14:46	
Hewlett-Packard	Gold	network latency	22.09.2005 at 03:08:15	11:32	
State Farm Insurance	Platinum	disk quota	22.09.2005 at 00:31:22	18:08	
Verizon Communications	Platinum	network latency	22.09.2005 at 14:35:20	17:53	
International Business Machines	Silver	network latency	22.09.2005 at 06:20:32	03:39	
Verizon Communications	Platinum	email service availability	22.09.2005 at 02:03:00	19:15	
AmerisourceBergen	Platinum	cpu utilization	22.09.2005 at 08:53:00	14:45	
General Motors	Bronze	network latency	22.09.2005 at 04:47:33	18:09	
ConocoPhillips	Silver	backup service availability	21.09.2005 at 21:00:48	03:35	
International Business Machines	Silver	cpu utilization	22.09.2005 at 10:49:49	14:36	
Boeing	Platinum	disk quota	22.09.2005 at 14:57:14	06:43	
Pfizer	Gold	disk quota	21.09.2005 at 22:53:52	19:27	
International Business Machines	Silver	cpu utilization	22.09.2005 at 11:16:02	14:36	
Pfizer	Gold	backup service availability	22.09.2005 at 14:34:04	14:53	
Bank of America	Silver	email service availability	22.09.2005 at 04:04:12	13:51	
Wal-Mart Stores	Gold	email service availability	22.09.2005 at 05:59:15	13:50	
Costco Wholesale	Bronze	network latency	21.09.2005 at 21:40:11	07:53	
American International Group	Gold	backup service availability	22.09.2005 at 12:06:48	12:20	
Ford Motor	Silver	cpu utilization	21.09.2005 at 20:39:40	03:48	
McKesson	Bronze	email service availability	22.09.2005 at 05:36:31	12:35	
International Business Machines	Silver	network latency	22.09.2005 at 06:14:04	03:39	
Berkshire Hathaway	Platinum	backup service availability	22.09.2005 at 03:42:53	11:46	
J.P. Morgan Chase	Silver	network latency	22.09.2005 at 02:02:20	04:06	

Figure 12: List of incidents to be prioritized

For each incident, the relevant customer, service level, SLO is presented together with the arrival time of the incident (time when the occurrence of the incident was first recorded) and time to violation (expected time from jeopardy to violation of SLO for the incident).

Configuration of the solution: definition of the priority levels and relative parameters

In order to prioritize incidents however, we need first to define priority levels through the definition of *priority buckets*. Each priority bucket is defined through its capacity (the maximum numbers of incidents that can be assigned a given priority

level) and the maximum expected closing time for all the incidents assigned to the bucket (in hours). An example of bucketization is given in Figure 13.

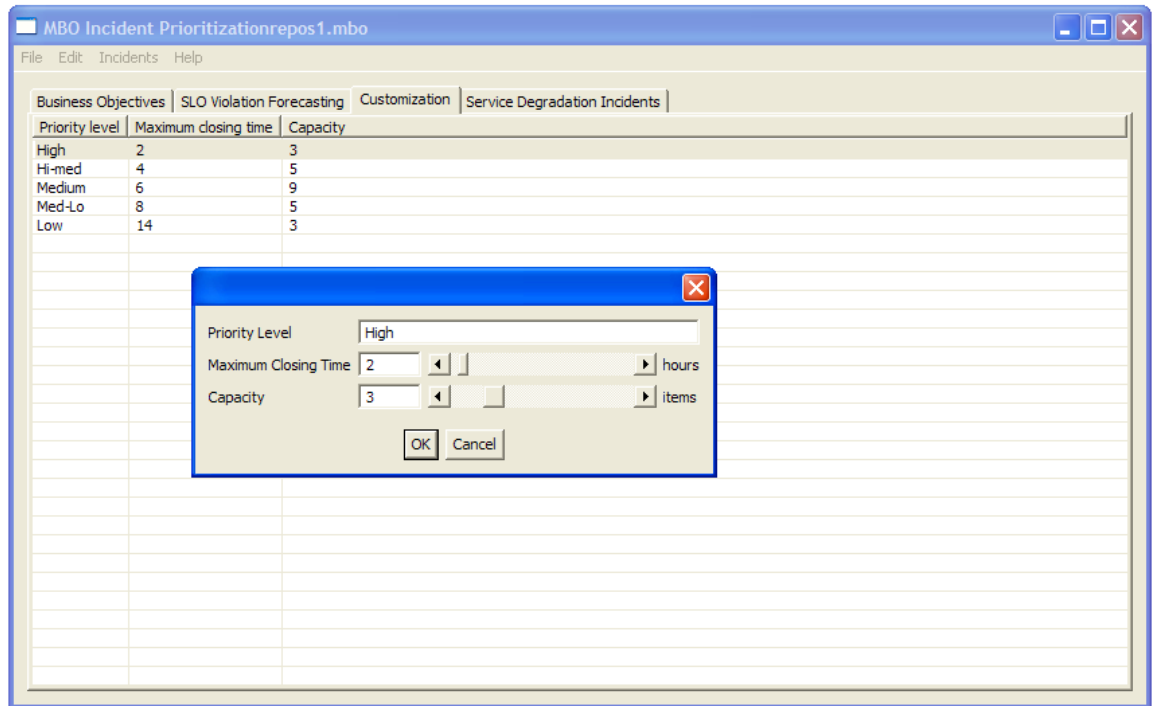


Figure 13: Customization of priority buckets

Prioritization of the service degradation incidents

At this point, the system executes the algorithm described in section 5.1.3. On completion, the demonstrator will show the result of prioritizing the incidents, as in Figure 14.

Customer	Service Level	SLO	Arrival Time	Time to violation	Priority
Verizon Communications	Platinum	network latency	22.09.2005 at 15:25:57	17:53	High
Boeing	Platinum	disk quota	22.09.2005 at 15:47:51	06:43	High
Pfizer	Gold	backup service availability	22.09.2005 at 15:24:41	14:53	High
Exxon Mobil	Platinum	email service availability	22.09.2005 at 09:45:43	14:46	Hi-med
AmerisourceBergen	Platinum	cpu utilization	22.09.2005 at 09:43:37	14:45	Hi-med
International Business Machines	Silver	cpu utilization	22.09.2005 at 11:40:26	14:36	Hi-med
International Business Machines	Silver	cpu utilization	22.09.2005 at 12:06:39	14:36	Hi-med
American International Group	Gold	backup service availability	22.09.2005 at 12:57:25	12:20	Hi-med
J.P. Morgan Chase	Silver	email service availability	22.09.2005 at 10:12:53	18:52	Medium
Hewlett-Packard	Gold	network latency	22.09.2005 at 03:58:52	11:32	Medium
State Farm Insurance	Platinum	disk quota	22.09.2005 at 01:21:59	18:08	Medium
Verizon Communications	Platinum	email service availability	22.09.2005 at 02:53:37	19:15	Medium
General Motors	Bronze	network latency	22.09.2005 at 05:38:11	18:09	Medium
Bank of America	Silver	email service availability	22.09.2005 at 04:54:49	13:51	Medium
Wal-Mart Stores	Gold	email service availability	22.09.2005 at 06:49:52	13:50	Medium
McKesson	Bronze	email service availability	22.09.2005 at 06:27:08	12:35	Medium
Berkshire Hathaway	Platinum	backup service availability	22.09.2005 at 04:33:30	11:46	Medium
Exxon Mobil	Platinum	cpu utilization	21.09.2005 at 22:05:04	16:04	Med-Lo
International Business Machines	Silver	network latency	22.09.2005 at 07:11:09	03:39	Med-Lo
Pfizer	Gold	disk quota	21.09.2005 at 23:44:29	19:27	Med-Lo
Costco Wholesale	Bronze	network latency	21.09.2005 at 22:30:48	07:53	Med-Lo
International Business Machines	Silver	network latency	22.09.2005 at 07:04:41	03:39	Med-Lo
ConocoPhillips	Silver	backup service availability	21.09.2005 at 21:51:26	03:35	Low
Ford Motor	Silver	cpu utilization	21.09.2005 at 21:30:17	03:48	Low
J.P. Morgan Chase	Silver	network latency	22.09.2005 at 02:52:57	04:06	Low

Figure 14: Prioritized service degradation incidents

Justification of the prioritization levels

In order to give some justification of the prioritization levels chosen for the incidents, on highlighting and double-clicking on one of the incidents, we show the likelihood of violation of the relevant SLO in function of the priority level assigned to it. A pop-window appears (Figure 15) that offers a view of what would be the consequences on the alignment calculation of assigning the incident to any priority level. The top-right portion of the pane reports information on the incident details, such as the customer, the service level, the SLO and the penalty. The top left portion of the pane shows an *alignment report*, which is the alignment with respect to the objectives in the default case (as if the SLO violation never occurred). Finally the bottom part of the pane gives a view of the consequences of prioritizing the incident. For each priority level, the expected time of closure of the incident is represented, along with the consequent likelihood of violation. Then for each KPI a short report is

given on what would be the expected value at period, and the consequent measure of alignment and loss of alignment. The graph offers a visual representation of the same information, you may choose KPIs and priority levels to show. The dotted rectangle presents the “default alignment” (forecasted measure of alignment if no service degradation incidents occur).

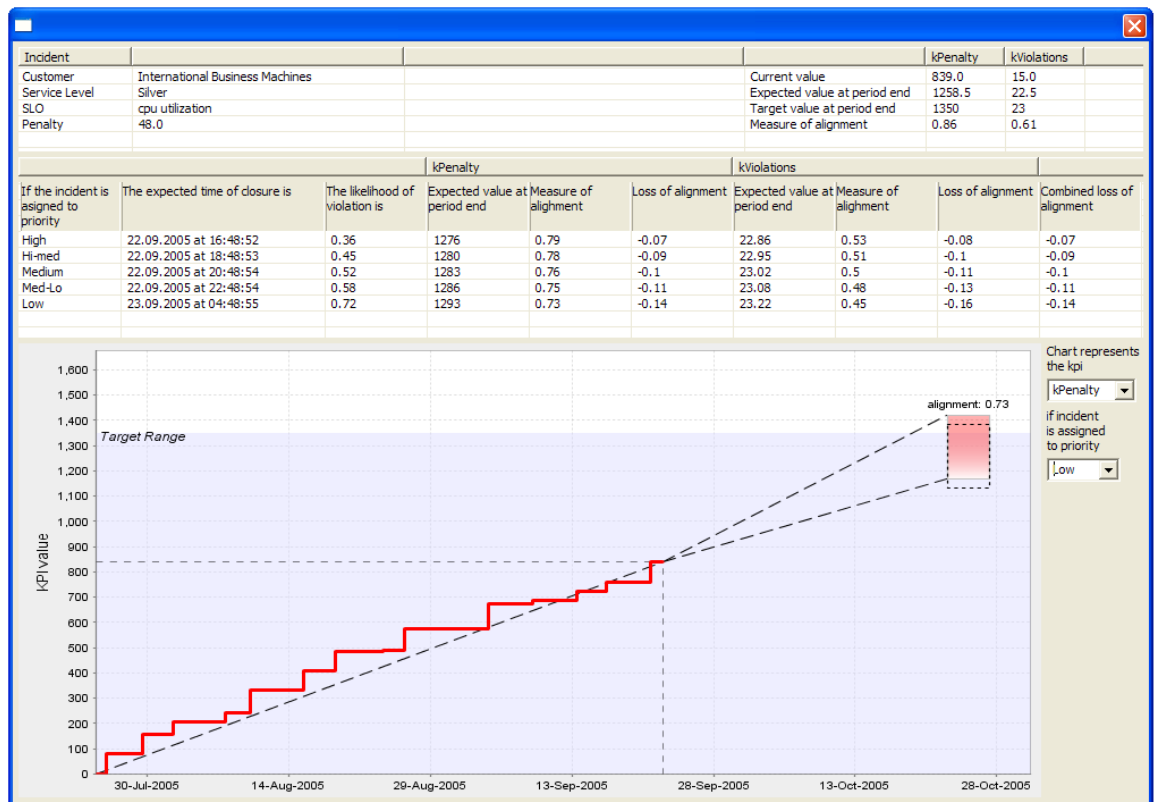


Figure 15: Service degradation incident profile

5.3 Applicability to other decision problems in IT service management

Incident prioritization is representative of a class of decision problems that face IT managers, for which having visibility into the business impact of alternative choices can help make better informed decisions. We have shown how one can build a BDIM solution to such self-contained decision problems. Other such examples have been proposed in the academic literature and we reviewed some of them in section 2.3. An example of such decision problems is change scheduling. In [44] we provided a problem formulation mapping the change scheduling to a mixed-integer optimization problem similarly with what explained above for the incident prioritization problem. Even though a complete BDIM solution was never built on top of that algorithm, the template solution we present here could equally well be applied on top of that formulation once a set of business objectives are given. Resulting from a collaboration led by the Bottom Line project at Federal University of Campina Grande, Brazil, we had also shown in [43] an alternative business-driven formulation of the change scheduling problem, this time using utility-as-cost rather than utility-as-business-impact (refer back to discussion in section 3.1.3)

6 A complete business-driven solution: organizational re-design for optimizing the performance of an IT support organization

*In this chapter we demonstrate the full usefulness of business-driven IT management, showing an application of it to architect, design and implement a comprehensive **business-driven solution for organizational re-design of an IT support organization** in order to optimize its performance with respect to its help desk function and incident management process.*

We first take the reader through a thorough analysis of the help desk function and corresponding incident management process. We show how the performance of the organization in terms of its contribution to the business can first be measured at the workgroup level (rather than through obvious organizational wide metrics that are not actionable). Then we present a what-if analysis tools that is able to show through simulation how the performance of the IT organization can be improved through organizational re-design (merging, or splitting workgroups, changing staffing levels, appropriately selecting prioritization policies at the workgroup level).

ITIL ([3]) defines Incident Management as the process through which IT support organizations manage to restore normal service operation as quickly as possible and with minimum disruption to the business. In order to be most effective, the IT organizations need to measure their own performance in dealing with service disruptions. Frameworks and guidelines such as ITIL and COBIT [1] define objectives for incident management, but link them to simple high-level organization-wide metrics such as the average duration of incidents. However, the reality of IT support organization is much more complicated, with staff working around the clock in the most disparate geographies. To have a clear picture of the organization's performance one must really define metrics that look inside the support organization and describe its inner working at a much finer grain level of detail.

Building on this first level of performance assessment, it soon becomes clear that the best way to assess performance is in terms of what the businesses that the organization support experience as a result of its dealing with service disruptions. In the point above, we address the performance in dealing with incidents. But not all incidents are created equal. Some have a higher penalty cost than other if they are not closed by a specified deadline. Other conditions being even, some incidents may affect more strategically important partners than other, and this must be taken into account when defining the impact of the service disruptions on the business that the organization supports.

Last but not least, having assessed the historical performance, and identified areas for improvements, it is usually not trivial to understand the improvements brought about by restructuring the support organization by increasing or cutting staffing levels, moving operators around support groups (possibly on retraining), and even

implement different prioritization policies for the technician when dealing with queues of incidents.

The objective of our BDIM solution is to improve the performance of the IT support organization in dealing with incidents. First, such performance needs to be measured through the impact that unresolved incidents have on the business because of service disruption.

Here we face two main challenges: i) the complexity and lack of transparency of the IT support organization, and ii) the extremely high volume of incidents and service calls that an IT support organization with global presence experiences.

COBIT ([2]) does provide a comprehensive set of indicators for assessing incident management. However, these indicators are quite high-level and designed to provide “black-box” measures of how the organization as a whole is performing. Their usefulness is therefore very limited when investigating causes of poor performance. The complexity and lack of transparency of the organization makes it hard to measure performance of the different support groups, and hard to predict the variation on business impact due to corrective actions to be taken (such as re-staffing for example).

6.1 Analysis of the incident management process

Our solution is based on an in-depth analysis of the incident management process in order to realize an accurate model of the system. The analysis was first presented in [76].

In general, an IT support organization (Figure 16) consists of a network of support groups, each comprising of a set of skilled technicians, with their work schedule. Support groups are divided into support levels (usually three to five), with lower level groups dealing with generic issues such as “user forgot password” and higher level groups handling specialized and time-consuming tasks.

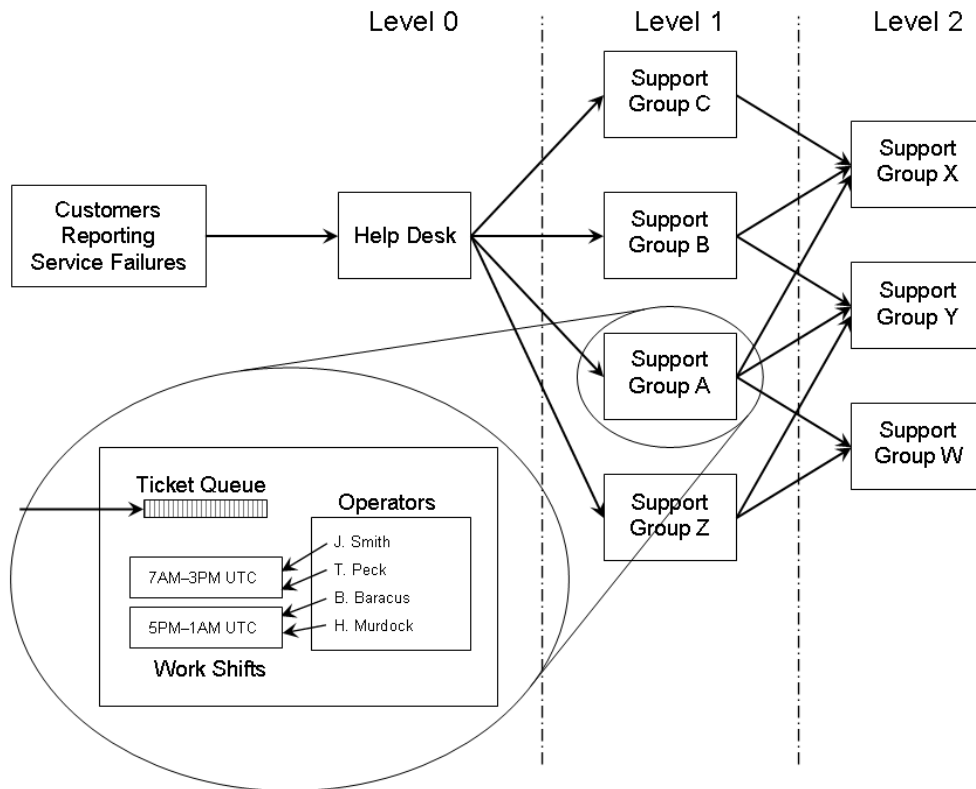


Figure 16: Conceptual model of the IT support organization for incident management

Support groups are further specialized by category of incidents that they treat (network, server, etc...) and divided into geographies, to ensure a more prompt handling of incidents especially at lower support levels. The Help Desk represents the interface for customers reporting an IT service disruption. In response to a customer request, the Help Desk opens an incident, (sometimes called trouble-ticket or ticket), and assigns a priority level to it. The incident will be managed by different support groups throughout its lifetime, as technicians in a given support group can either resolve the incident or pass it on to a different support group (usually escalating to a higher level of support). Support groups deal with the incident according to its internal policy, taking its priority level into account.

We model the support organization as an open queuing network [77]. In particular, in our model the organization is composed by a set of hierarchically organized

support groups. Our abstraction of the incident management process is illustrated in Figure 17. Each support group has a number of operators and a queue of incoming tickets. When an operator is idle, he/she picks the ticket on top of the queue (following the group's priority policy) and starts working on it, until the incident is solved or cannot be further processed and needs to be escalated to a higher level support group. Upon closure or escalation of an incident, the operator takes another incident from the incoming queue or becomes idle if no more incidents exist.

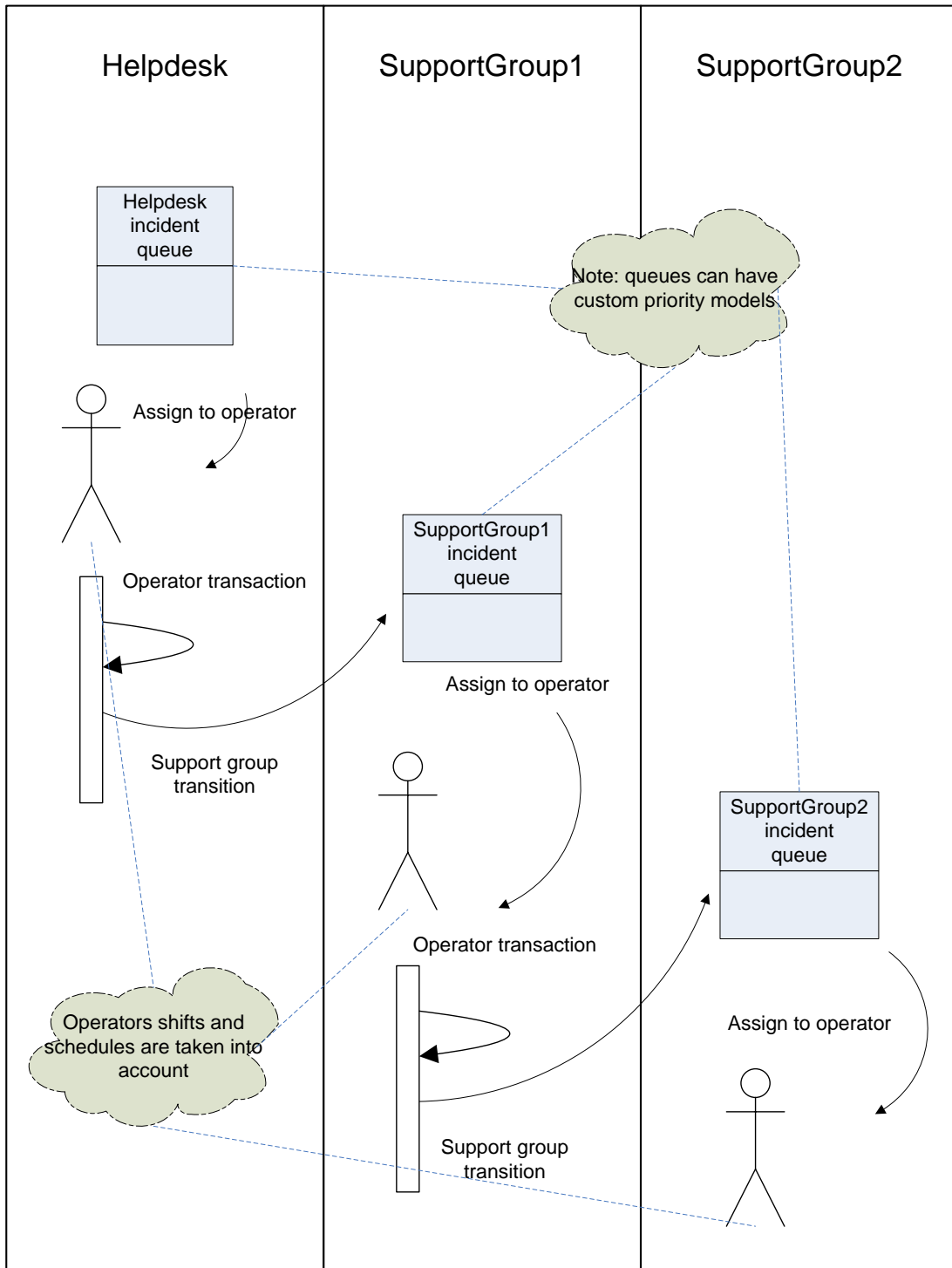


Figure 17: Abstraction of the incident management process

The incident lifecycle is described as the progression of an incident from its initial occurrence to detection, then to the diagnosis of the cause of failure, the repair of the issue, recovery of the component that was at fault, and finally restoration of service.

Throughout the lifetime of an incident, the incident gets "opened" by the helpdesk, then assigned to a support group which is going to work on it and either mark it "closed" or "reassign" it to a different support group. At each of these states, the incident record is updated with the pertinent information, such as the troubleshooting that has taken place, which team is responsible for action, which engineer on the team, and what the current status of the incident is. Figure 18 demonstrates the progression between these different states: An incident is opened when a call comes in to the call center after an issue is detected.

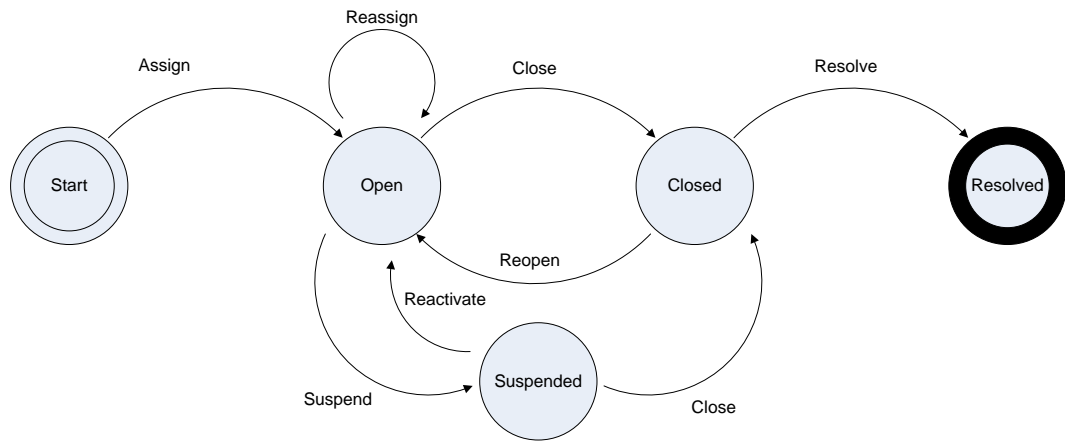


Figure 18: Incident lifecycle

Once the incident is assigned to the appropriate team that will handle the issue, it is updated numerous times to reflect the progression of troubleshooting that the engineers attempt. If, for some reason, the end-users requests that the engineer stops working on the issue, it is placed in a "suspended" state so as not to incur SLO penalty while not being worked on. Once the disruption is repaired, the incident is placed in "closed" state until the end-user confirms that service has been restored and all is well. Once the restoration is confirmed, the incident is "resolved" and its lifecycle ends.

6.2 A reformulation in BDIM terms of the IT support organizational re-design problem

Let's set up a BDIM problem for our IT manager. Consistently with our BDIM framework described in chapters 3 and 4, assume that she has a set of business objectives for which she wants to maximize alignment (or minimize business impact). Her management options are the following: re-architect the organization, by adding, removing, merging and splitting support groups; change staffing levels in support groups; and change prioritization policies for dealing with incidents in support groups. The problem the IT manager is facing is to implement the options that she has available and obtain better alignment with her objectives (maximize utility, minimize business impact etc...).

It is important to note that this is a real-life problem that has occurred to many support organization face. The complexity of IT support organizations hinders the verification of the alignment of current organizational, structural, and behavioral processes with the strategic objectives defined at the business management level. In fact, the performance assessment of the incident management function is a very complex procedure which involves the business impact evaluation of the available incident management strategic options, through the definition of a set of metrics that allow the objective measurement of performance indicators, as we saw in chapter 4. Performance analysis and optimization are also organization-specific procedures, since the business impact of service disruptions, and consequently the metrics to consider, vary with the nature of the services and the types of disruptions that occur.

This problem is further complicated by the fact that putting in place processes and changing staffing to address performance issues and then measuring and evaluating the effects of those changes is a long and expensive process. Moreover, even just in

order to assess the results of the measures taken the manager will have to wait and monitor the performance of the IT organization for a long time.

Finally, even though the types of options available to our IT manager is small, the complexity of the organization is such (typical IT support organizations consist of thousand of technicians organized into hundreds of support groups) that the optimization process will have to be guided somehow.

We will help our IT manager solve her BDIM problem with a what-if scenario analysis tool – SYMIAN (fully described in section 6.4, first introduced in [78]) - that allows a user to play out what-if scenarios such as adding technicians to a given support group, merging support groups together, experimenting with alternative incident prioritization policies or other such actions. At the core of the tool is a discrete event simulator that re-enacts the lifecycle of incidents, allowing the user to assess likely improvements in performance without having to go through the expensive and time consuming process of implementing the corrective measures.

Suppose now that the IT manager has the ability to choose what what-if scenarios to play out. Where should she start from? What are the spots in the organization where improvements in terms of re-architecting support groups, changing staffing levels and prioritization policies are liable to yield the best results in terms of performance improvements? In order to answer these questions we need to first define some metrics for the performance evaluation of the IT support organization in dealing with incidents, and then show how measuring their values can help the IT manager identify bottlenecks and other such spots in the IT support organization.

We therefore defined metrics assessing effectiveness in routing incidents between support groups as well as metrics assessing efficiency within support groups (see also [76]). In the next section we define such metrics and describe how we were able to

obtain values to them by performing data mining on logs of commercial software tool for helpdesk.

6.3 Metrics for performance evaluation of the IT support organization in dealing with incidents

“You can’t manage what you can’t measure”

From the analysis of the incident management process described in section 6.1, there are two orthogonal dimensions along which the performance of the IT organization can be assessed. One dimension is represented by the **effectiveness** of the routing of incidents: are incidents being dispatched to the right support groups? “Right” is defined here as the support group best equipped with the knowledge for dealing with the incidents. The other dimension is the **efficiency** of every single support group in dealing with the incidents: are support groups staffed at the right level? Are there inefficiencies in the way a support group manages the schedule of its technicians? This observation proved to be extremely useful to enforce separation of concerns in the performance metric definition.

6.3.1 Incident Routing Between Assignment Groups

The metrics we defined to measure effectiveness of routing were:

- Number of reassignments per incident
- Number of assignment cycles; Number of incidents seen twice or more:
- Number of cross-level reassignments
- Number of updates (operator transactions) between (onwards + back) reassignments
- Number of updates (operator transactions) before incident bounced back
- Time to closure after reassignments

- Number of incidents that had a large processing time and were then passed on

As it can be seen, most of the metrics represent quite intuitively the aspects of the incident management process that can highlight ineffective routing of incidents. For example, the number of reassignments per incident (Figure 19) gives quite a clear picture of how well routing is functioning. The support groups that tend to treat more “long-lived” tickets than average can be pinpointed as a good starting point for troubleshooting performance issues in the IT organization.

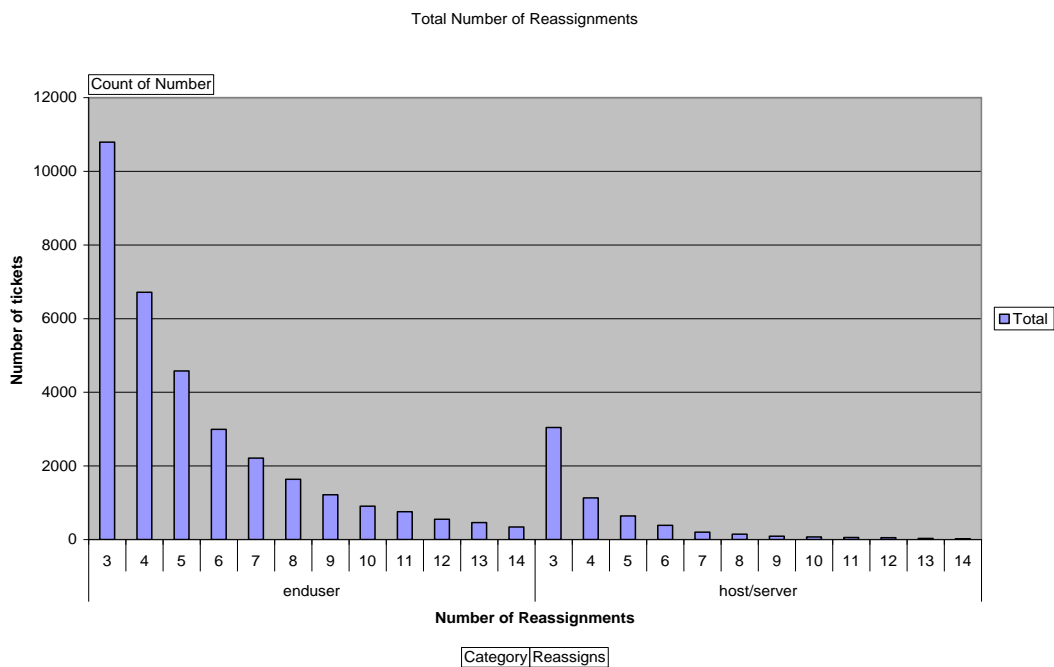


Figure 19: Incidents by number of reassignments

Another example of an information-rich metric is the number of assignment cycles. The rationale behind it is that an incident loops between a few support group before resolution, then most probably something wrong is going on in those groups. The same goes for tickets that are bounced back and forward by a pair of support groups.

6.3.2 Performance within Assignment Groups

The following were the metrics that we defined aimed at measuring the efficiency of support groups in dealing with incidents:

- Fan-in and fan-out of the support group
- Time spent in support group
- Number of incidents received vs. number of incidents resolved
- Number of incidents treated
- Number of operators that looked at ticket in support group

Again, most of the metrics are quite intuitive. *Fan-in* and *fan-out* are intended to represent the *centrality* of the support group in the organization Figure 20. They measure the number of support groups that this group receives incidents from (fan-in) and forwards incidents onto (fan-out). The idea behind it being that the organization is more sensitive to performance issues in central support group versus peripheral ones.

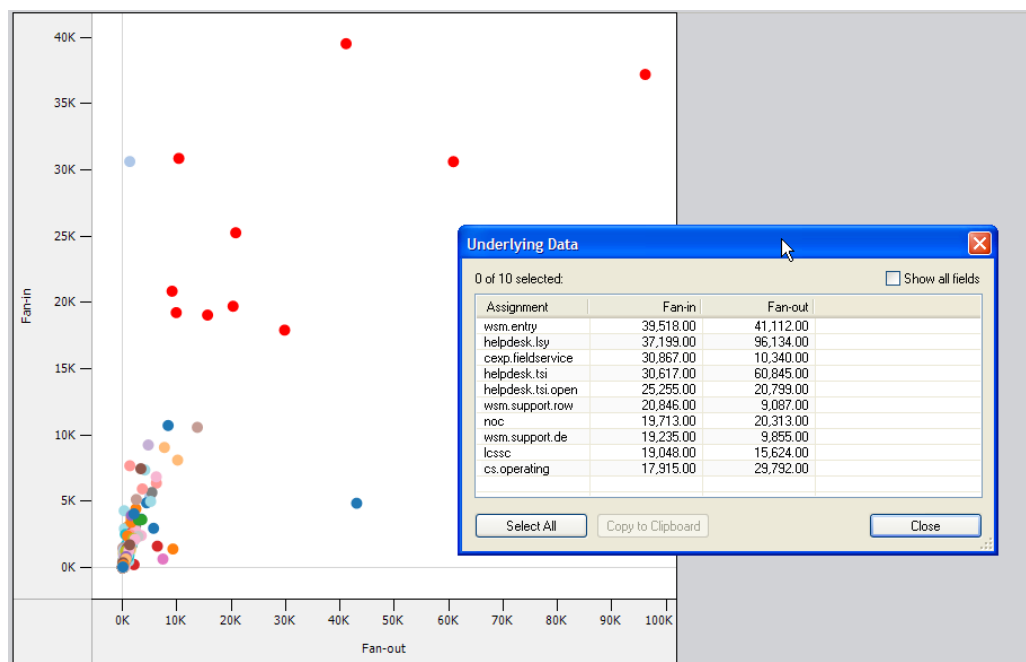


Figure 20: Fan-in and fan-out of assignment groups

Support groups in which incidents seemed to spend a much longer time on average were marked as *bottlenecks*, thereby possibly highlighting serious performance issues (Figure 21). To compute the supporting metric, we profiled the lifecycle of each incident, flagging the support group at which it spent the most time (summing queue waiting time and work time). The support groups that were flagged the most on average were the bottlenecks.

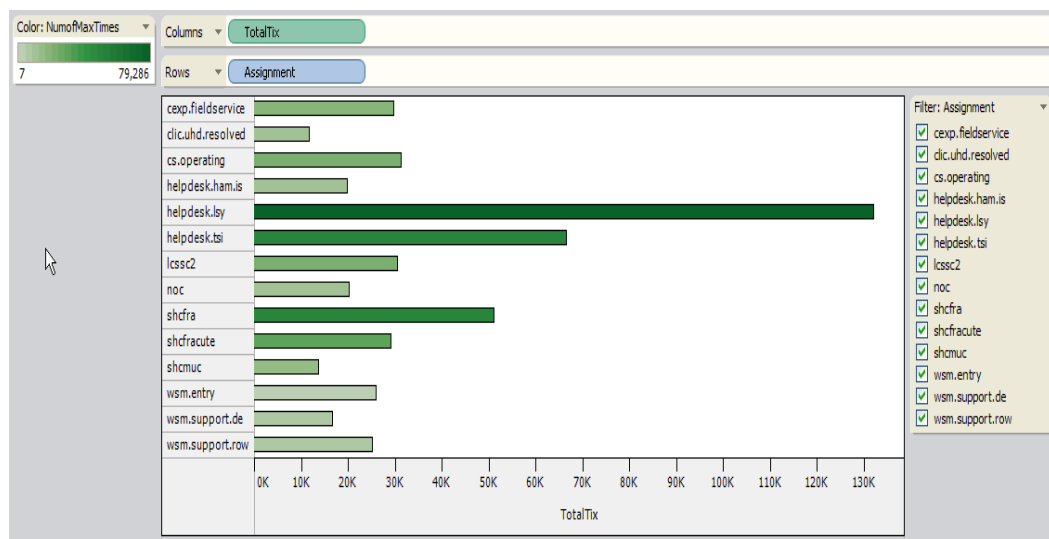


Figure 21: Bottleneck assignment groups

6.3.3 Validation of the metrics in a real-life situation

We applied our metrics to a real life case study, involving an HP customer worldwide leader in the provision of IT services to the airline industry (the Skyes-IT (not the organization’s real name) case study, described in [76]). Having a global 24/7 presence, supporting multiple environments for multiple customers is a challenge that Skyes-IT faces daily. 700 assignment groups utilizing over 1600 engineers provide support to end users at ticketing desks, check-in counters and service centers all around the world. Since Skyes-IT were managing all their incidents through an installation of HP ServiceCenter™ [79], we were able to obtain a one-year worth

incident management database, consisting of data for over 600,000 incidents. The data carried information on what incidents were opened, when, by whom, on what customer's account, what support groups dealt with them, what transactions on them were made by which technician, all time-stamped. Mining that data, we determined values for the metrics defined in the previous section.

6.4 Optimizing the incident management performance through simulation techniques

Let's now get back to our IT manager looking to improve the performance of her support organization. Following the analysis described in the previous section, she has a starting point for **what-if scenario analysis**, in terms of workgroups that have been identified as performance bottlenecks. In the previous section we have shown how to define some metrics for the performance evaluation of real IT support organization. In this section we will show how to evaluate the impact of changes in the organization based on these metrics, and more importantly the business impact of these changes.

The operations available to the IT manager for the optimization of IT organization are:

- **increasing or cutting staffing levels,**
- **transferring operators around support groups and**
- **implementing different prioritization policies for incident queues.**

In this section, we introduce a simulation-based tool that can perform the necessary what-if scenario analysis. A user of the tool is able to explore the options that are available to an IT manager wanting to optimize the organization, and analyze the output of the simulation in terms of a) the same performance metrics for intra-workgroup and inter-workgroup performance as seen in 6.3; b) the two ITIL-recommended IT metrics of *Mean Time To (incident) Resolution (MTTR)* and *Mean*

Incidents Closed Daily (MICD) and c) business impact in terms of alignment with the business objectives, as seen in chapter 3 and 4.

MTTR and *MICD* are organization-wide metrics, and as such they provide little insight on the internal dynamics of the organization. For a comprehensive performance analysis of the incident management process, we have to delve into a deeper level of detail. More specifically, we need to consider both inter- and intra-support groups dynamics, along two orthogonal dimensions: the *effectiveness* of incident routing and the *efficiency* of every single support group in dealing with the incidents. This requires us to take into consideration other performance metrics which can evaluate the organization capability to directly forward incidents to the best equipped support groups and the optimality of staff allocation and operator work shift scheduling.

6.4.1 The SYMIAN Decision Support Tool

SYMIAN is a decision support tool for the performance analysis and optimization of the incident management function in IT support organizations. In particular, SYMIAN exploits a *discrete event simulator* to reproduce the behavior of IT organizations and to evaluate their incident management performance.

SYMIAN enables its users to play out *what-if scenarios*, allowing them to assess likely improvements in performance without having to go through the expensive and time-consuming process of implementing organizational, structural and behavioral changes. More specifically, SYMIAN allows users to incrementally specify the set of changes to apply to the current organization in order to define an alternative organization configuration that will be tested on a set of performance metrics. For instance, SYMIAN allows modifications such as re-staffing support groups, merging support groups together, experimenting with alternative work shifts, incident routing

and/or prioritization policies, or other such actions. SYMIAN guides users all along the optimization process, providing ad hoc visualization of simulation results and, in case a limited set of predefined metrics such as MTTR is considered, explicit tips for the modification of some organization parameters such as the staff allocation.

SYMIAN models the IT support organization (in terms of the number of support groups, the support level, the set of operators, the operator work shifts, the relationships with other support groups, etc.) and permits to define the set of performance metrics to consider for the optimization. SYMIAN then simulates the organization behavior considering a user specified set of incidents, evaluating the desired performance metrics.

6.4.2 How SYMIAN works

At the core of SYMIAN is a discrete event simulator that re-enacts the lifecycle of incidents, allowing the user to assess likely improvements in performance without having to go through the expensive and time consuming process of implementing the corrective measures.

From the abstraction of the incident management process we observed that the three main phases for a simulation are: incident generation, incident transition and operation transaction.

Incident generation

The architecture of our tool allows for the co-existence of multiple incident generators that can be used one at a time or concurrently. When used concurrently, they behave like multiple sources of incidents feeding the system. What characterizes an incident generator is the *incident arrival distribution* that it uses to generate incidents. This is often assumed to be exponential, and therefore its main parameter is

the time to next incident (*TTNI*) factor, expressing the inverse of number of incidents in the unit of simulated time. Throttling the *TTNI*, our simulator can cope with what-if question such as what happens if the frequency of incident arrival increases by 10%. The default incident generator for our system is one that re-samples the population of incidents observed historically and re-creates them using the category and first assignment information from the historical logs.

Incident transition

For simulating incident transitions, we make the assumption that the process of forwarding incidents is **memory-less**⁸: a support group tends to forward incidents to other groups regardless of where they come from. Therefore our simulation re-enacts a Markov process. We build a **transition matrix** from the historical data that contains the frequencies of transition of an incident from support group to support group, and we use those values as the probabilities that an equivalent transition would happen in the simulated process. The matrix is further indexed by incident category.

Operator transactions

We represent each support group as a **queuing system** with multiple processors, (IT operators). Each time the simulator fetches an incident from the queue - according to the group's prioritization policy – it makes an operator work on it for a period of time consistent with a given distribution, made to fit historical observation for incident worktime in the given support group.

⁸ We show how we validated this assumption at the end of this section.

6.4.3 The SYMIAN Simulation process

Because of the stochastic nature of the process being simulated, the tool averages the IT performance and business impact metrics collected using a Monte Carlo approach over multiple runs.

Each simulation run works as follows:

1. Each of the incident generators start generating incidents according to the parameters that were programmed into them (*TTNI*, category, first assignment group information)
2. Each of the simulated incidents created in the previous step is forwarded to the appropriate support group (first assignment group)
3. On receiving a new incident, the support group will *either* select at random an *available* operator to start working on it immediately, *or* put it in its queue if the group has no available operator
4. When an operator starts working on a ticket, the simulator computes the duration of the operator transaction according to the given distribution and marks the operator as *busy* for said duration
5. When the simulated time reaches the end of an operator transaction, the simulator marks the operator as *available* and the incident as *transition-ready*.
6. On becoming available, the operator checks if there are ticket in the queue. If so, it extracts a ticket from the queue according to the group's *prioritization policy* and starts work on it as in step 4.
7. When an incident at a workgroup i is marked *transition-ready* (step 5), the simulator closes it with probability c_i . $(c_i) \forall i = 1..N$ s.t. $\sum_{i=1}^N c_i = 1$ is the *incident closing vector*. Each element c_i of the vector represents the

probability that the incident will be closed on terminating a transaction at group i . If the incident is closed, the simulator collects the necessary information about the incident. Else, the incident is transitioned to a workgroup j with probability t_{ij} . $(t_{ij}) \forall i, j = 1..N$, s.t. $\forall i = 1..N, \sum_{j=1}^N t_{ij} = 1$ is the *transition matrix* described above. Each element t_{ij} represents the probability that – unless it's closed – on completing an operation transaction at the workgroup i , a ticket will be forwarded from i to the workgroup j .

6.4.4 What-if scenario analysis

SYMIAN performs what if scenario analysis on creation, removal, merging or splitting of support groups by way of transformations on the original transition matrix. For example, the merging of support groups is done by amending the transition matrix by replacing the two original groups with the merged group. For the merged group, on the source side we use a weighted average of the frequencies observed historically in the original groups and on the target side we use the sum of the frequencies observed historically. SYMIAN can also enact scenarios such as re-staffing of support groups (done by changing the number of operators available to fetch tickets from the groups' queues), and experimenting with alternative prioritization policies (changing the order in which tickets are fetched from the groups), besides obvious use cases to do with throttling *TTNI* in incident generation.

In the following few sub-sections we look at these operations in more details. For simplicity reasons we omit to deal with incident categories, and will suppose that all the incidents are created in the same category. The extension to the multi-category case should be obvious to the reader. We assume an organization composed of N

workgroups, with *incident closing vector* $(c_i), i = 1..N$ and transition matrix $(t_{ij}) \forall i, j = 1..N$.

Removing support groups

When removing an assignment group \bar{i} , the closing vector and transition matrix will be updated to reflect deletion of the group, and subsequently renormalized to satisfy the invariants $\sum_{i=1}^N c_i = 1$ and $\forall i = 1..N, \sum_{j=1}^N t_{ij} = 1$.

Supposing without loss of generality that \bar{i} is the N -th group (which can be obtained simply through re-indexing), the new closing vector is given by $(c'_i) = (c_i) * 1/(1 - c_N) \forall i = 1..N - 1$ (excluding the trivial case where $c_N = 1$). The new transition matrix is $(t'_{ij}) = (t_{ij}) * 1/(1 - t_{Nj}) \forall i, j = 1..N - 1$ (again excluding trivial cases).

Creating support groups

When creating a new support group (without loss of generality indexed $N + 1$), the user will be required to provide a scalar c_{N+1} , representing the closing probability at the group; a vector $(t_{0j}), \forall j = 1..N$ s. t. $\sum_{j=1}^N (t_{0j}) = 1$ representing the transition probability from the group to each of the other groups; and a vector $(from_i), \forall i = 1..N$, representing the transition probability from each of the other groups. The closing vector get extended with c_{N+1} and re-normalized as above. The rows of the transition matrix are first updated according to $(t'_{ij}) = (t_{ij}) * (1 - from_i) \forall i, j = 1..N$ so that $\forall i = 1..N, \sum_{j=1}^N t'_{ij} = 1 - from_i$. At this point the matrix gets extended with the row $(t'_{N+1,j}) = (t_{0j})$ and the column $(t'_{i,N+1}) = (from_i)$. The reader can verify that the invariants are now satisfied for (t'_{ij}) so extended.

Merging support groups

When merging two support groups g_1 and g_2 , SYMIAN requires information on the volume of incidents processed at each group⁹. It is expected that SYMIAN will have that information available because of historical computation or previous simulations (note that these parameters represents metrics discussed in section 6.3). If so, it will suggest those values to the user letting the user override them. Else the user will be required to input their estimate on them.

The merging operation is equivalent to the removal of each group, followed the creation of a new group that will have closing and transition probabilities calculated as follows. If $r = v_1/v_2$ is the ratio between the volume of incidents processed at each group, the closing probability of the addendum group will be $r * c_{g_1} + (1 - r) * c_{g_2}$. The (to_j) and $(from_i)$ vectors for the addendum group will be respectively $(to_j) = r * (t_{g_1,j}) + (1 - r) * (t_{g_2,j})$ and $(from_i) = r * (t_{i,g_1}) + (1 - r) * (t_{i,g_2})$.

Splitting support groups

When splitting an existing support group g into two, SYMIAN will require the user to state the ratio r of the incident volume that each new group is expected to have. SYMIAN will suggest to set by default this ratio at $1/2$. The splitting operation is equivalent to the removal of the old group, followed by the addition of two new groups that will have the same closing probability as the original group c_g ; (to_j) vectors that are identical to the original group's transition matrix column (t_{gj}) ; and $(from_i)$ vectors that are given by $r * (t_{ig})$ and $(1 - r) * (t_{ig})$, respectively.

⁹ Strictly speaking, it only requires the ratio of the volumes, r

Changing staffing level

SYMIAN will require the user to state the new value of staffing level for a given support group or set thereof, and use those new levels in its simulations.

Changing incident generation properties

Incident generators can be configured with all the parameters described in 6.4.2 under *incident generation* (e.g. *TTNI*, categories, initial support groups, etc.)

6.4.5 Implementation of SYMIAN

SYMIAN is implemented in the Ruby (<http://www.ruby-lang.org/>) programming language. Ruby was chosen for its remarkable extensibility and its support for meta-programming. The capability to easily redefine the behavior of time-handling classes in the Ruby standard library allowed the implementation of a simulated clock which models the flow of simulation-time in a very similar way to what happens in real life. In addition, Ruby's meta-programming enabled the definition of domain-specific languages and their use in the realization of several simulator components. These have proved to be particularly effective development techniques.

The availability of a wide range of high-quality scientific libraries was also a major reason behind the adoption of Ruby. In particular, SYMIAN exploits the GNU Scientific Library (GSL), via the Ruby/GSL bindings, for high-quality random number generation, and it integrates with the Gnuplot data visualization tool to plot some of the simulation results. Finally, SYMIAN exploits Ruby facilities to import configuration parameters and export simulation results in the XML, YAML, and CSV formats, in order to ease integration with external software for the automation of multiple simulation runs and with scientific tools for post processing of simulation results.

6.4.6 Interpretation of SYMIAN simulations results

This section presents an example of how an IT manager should go about interpreting the output of the SYMIAN simulation. Here, SYMIAN is applied to minimize the service disruption time in the context of a case study IT support organization, with the constraint of preserving the current number of operators. As a result, the objectives of the performance improvement process are the maximization of the mean incidents closed daily (*MICD*) metric, as well as the minimization of the mean time to resolution (*MTTR*) metric.

The subject of this experimental evaluation is the fictitious incident management organization *INCS'R'US*, which is composed of 3 support levels (0-2), 31 support groups, and 348 operators. The complete characterization of the 31 support groups is presented in Table 3. To limit the complexity of the case study, the routing of incidents in the *INCS'R'US* organization is assumed to be unidirectional, that is support groups of level *N* can only receive incidents from support groups of level *N-1* and escalate incidents to support groups of level *N+1*. In addition, an equal probability of incident escalation to each of the support groups of immediately higher level is assumed.

INCS'R'US deals with incidents modeled according to the characterization provided in Table 3. Incidents have 4 categories (*A – D*) and 3 severity levels (1 – 3). For every specific combination of incident category and severity, the amount of work that incidents require for service restoration, at every support level, follows a uniform random probability distribution. In Table 4, the abbreviated notation $U(\alpha)$, where $\alpha > 0$, represents the uniform random variable distribution in the $[0, \alpha]$ interval.

Support Level	Support Group (Number of Operators)	Work Shift
0	Help Desk (75)	(25 operators) 7AM-3PM UTC
		(25 operators) 4AM-12PM UTC
		(25 operators) 12PM-8PM UTC
		(10 operators) 5PM-1AM UTC
1	SG1 (15), SG9 (12), SG15 (13), SG18 (5)	7AM-3PM UTC
	SG2 (7), SG10 (7), SG13 (7)	8AM-4PM UTC
	SG3 (15), SG19 (12)	12PM-8PM UTC
	SG4 (4), SG11 (6)	2PM-10PM UTC
	SG5 (14), SG16 (12), SG20 (6)	4AM-12PM UTC
	SG6(12), SG17 (9)	3AM-11AM UTC
	SG7 (5), SG14 (5)	5PM-1AM UTC
	SG8 (6), SG12 (8)	9AM-5PM UTC
2	SG21 (9), SG25 (10)	2PM-10PM UTC
	SG22 (8), SG26 (8)	9AM-5PM UTC
	SG23 (7), SG27 (7)	8AM-4PM UTC
	SG24 (9), SG28 (10)	5PM-1AM UTC
	SG29 (9)	3AM-11AM UTC
	SG30 (6)	4AM-12PM UTC

Table 3: Support group characterization in the *Incs'R'Us* incident management organization.

	Severity Level 1	Severity Level 2	Severity Level 3
Category A	L0: U(300) L1: 0 L2: 0	L0: U(900) L1: U(240) L2: 0	L0: U(1800) L1: U(900) L2: U(120)
Category B	L0: U(300) L1: U(1200) L2: U(120)	L0: U(600) L1: U(2400) L2: U(240)	L0: U(900) L1: U(3600) L2: U(480)
Category C	L0: U(600) L1: U(150) L2: U(1200)	L0: U(900) L1: U(300) L2: U(2400)	L0: U(1200) L1: U(450) L2: U(3600)
Category D	L0: U(900) L1: U(1200) L2: U(1200)	L0: U(1800) L1: U(4800) L2: U(4800)	L0: U(2400) L1: U(6000) L2: U(6000)

Table 4: Stochastic characterization of the amount of work time (in seconds) required for incident closure

Category *A* models incidents which mostly require work at support level 0, and a limited amount of work at higher support levels. Category *B* and *C* model incidents which require work at every support level, but mostly at support level 1 and 2 respectively. Category *D* models incidents which require a significant amount of work at every support level. For every incident, category and severity level are randomly chosen, with uniform probability, at generation time. Incident inter-arrival times follow a random exponential probability distribution with an average of 30 seconds.

We run a first simulation to evaluate the performance of the current organization. The simulation covered three whole days of simulated time, starting from 2PM UTC¹⁰. We show the values for the *MICD* and *MTTR* performance metrics obtained from the simulation in Table 5 (first column). The table also shows the Mean Work Time (*MWT*) metric, defined as the mean work time per closed incident, as an indication on the amount of work spent on service restoration.

By analyzing the variation of the incident queue size at every support group using both SYMIAN graphical visualization and time series analysis functions, it was easy to realize that support groups *SG1*, *SG4*, *SG7*, *SG8* and *SG14* at support level 1 and support group *SG30* at support level 2 were a major performance bottleneck, while the Help Desk and support groups *SG3* and *SG17* were oversized. As an example of the effectiveness of visual analysis to locate performance bottlenecks, Figure 22 plots the variation of incident queue size at support group *SG30*.

¹⁰ We did not consider the first 24 hours of simulated time for the evaluation of the performance metrics, as they were introduced only to prime the simulation environment to avoid taking measurements on a cold start.

	First simulation	Second simulation
Total incidents generated	8609	8609
Incidents generated after warm-up	5728	5728
MICD	1811	2002
MTTR (in seconds)	53423	47047
MWT (in seconds)	L0: 508, L1: 809, L2: 784	L0: 506, L1: 811, L2: 773

Table 5: Performance metrics from the first and second simulation

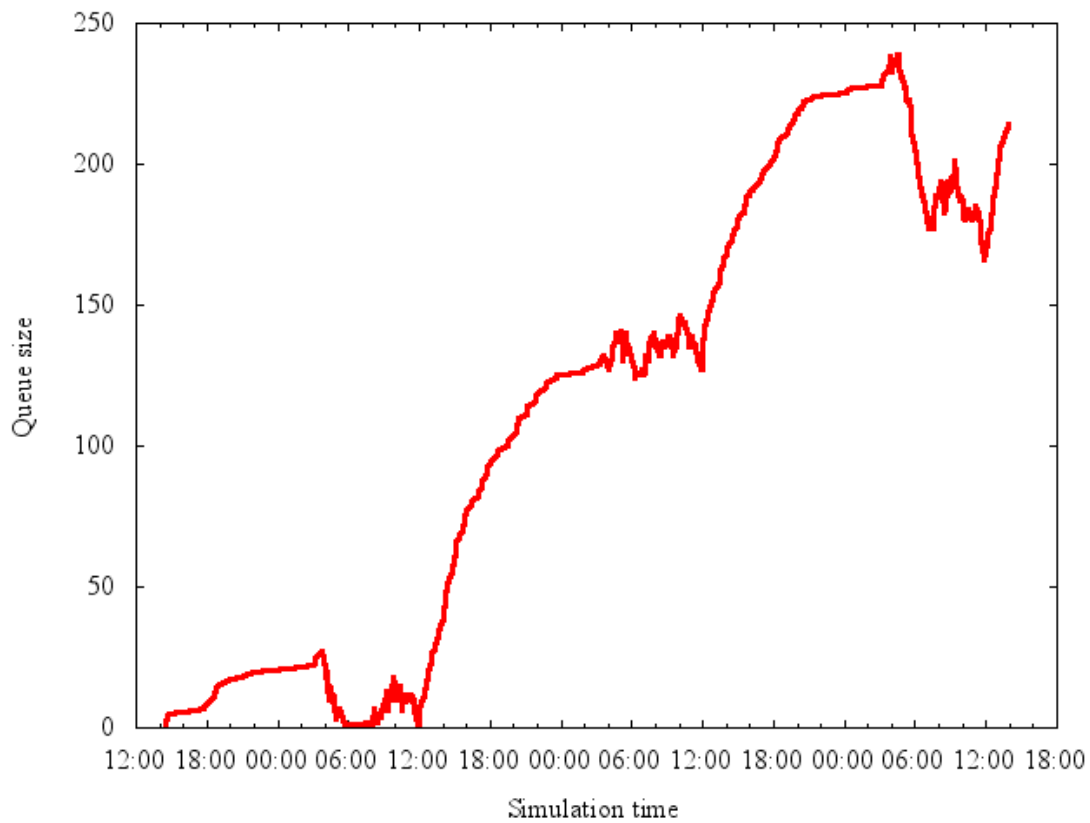


Figure 22: Incident queue size at support group SG30 during the first simulation

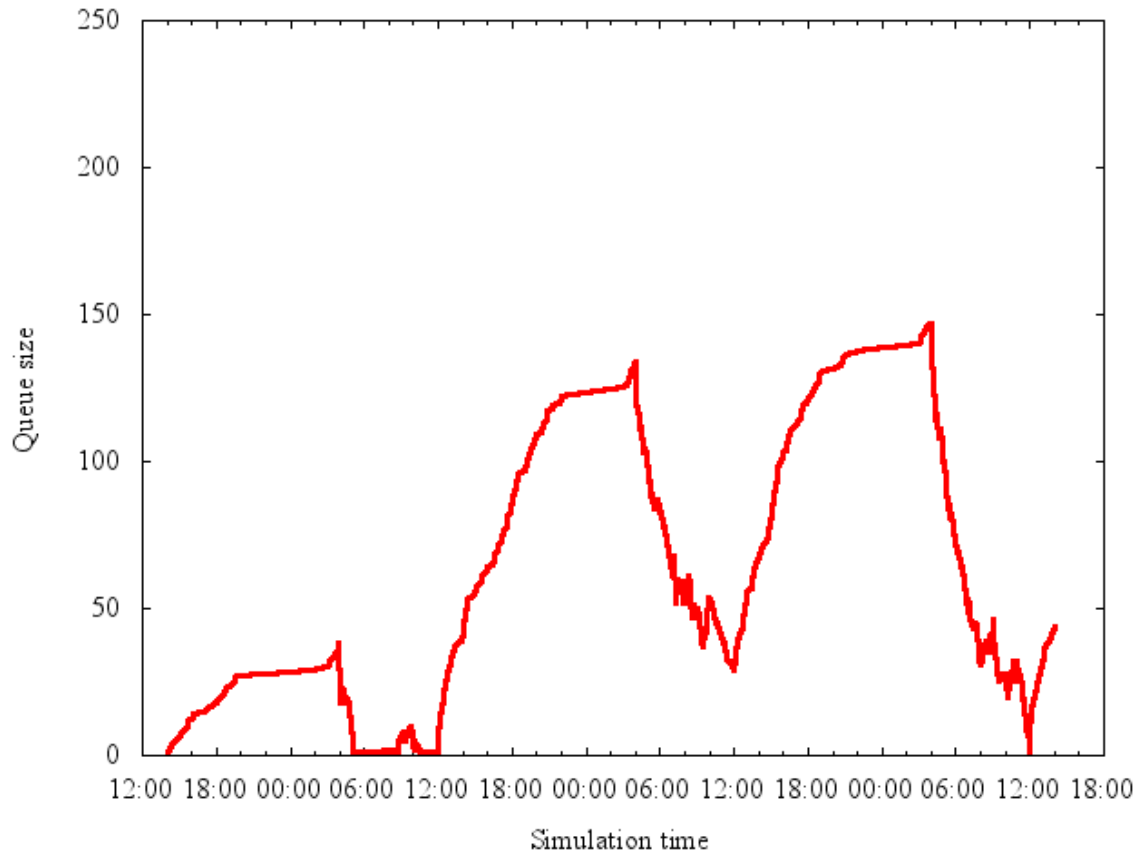


Figure 23: Incident queue size at support group SG30 during the second simulation

To improve the organization performance, we transferred 8 operators from the Help Desk to support groups *SG1*, *SG4*, *SG7*, and *SG8* (2 operators for each group), 3 operators were transferred from support group *SG3* to support group *SG14*, and 2 operators were transferred from support group *SG17* to support group *SG30*. We then launched a new simulation to assess the performance of the new organization. Table 5 (second column) and Figure 23 provide respectively the performance metrics and the variation of incident queue size at support group *SG30* for the new simulation.

The results of the second simulation proved that the reallocation of operators was very effective in improving the whole system performance. In particular, the INCS' R' US organization exhibited a 10.5% improvement of the *MICD* and a 11.9% decrease of the *MTTR*.

Although the target of the previous performance optimization experiment is a fictitious organization, the case study was carefully designed to be representative of the complexity of real-life IT organizations. Therefore, the simulation results demonstrate the effectiveness of the SYMIAN tool for the performance optimization of the incident management function in IT support organizations.

6.4.7 Validating assumption of memory-less process

We validated the behavior of the simulator against organization-wide and group metrics that were observed in the historical data described in section 6.3. For validating the assumption that the process is memory-less, we showed that the historical number of incidents processed by the support groups (Figure 24), and the number of ticket re-assignment were recreated with excellent fidelity using the Markov methodology based on the incident transition matrix (see Figure 25). For operation transactions, the two principal such measures are the mean time to resolution of incidents (*MTTR*), and the lengths of the group queues. For these we had somehow less accurate predictions. However, given that the errors tended to go in the direction of showing a *MTTR* consistently lower than what observed, we were able to introduce satisfactory corrective measures.

6.5 Advanced visualization techniques for guided performance analysis

One of the most successful outcomes of the Skyes-IT pilot project, described in 6.3, was the ability to provide insight into the incident management process flow. We were able to provide data for the statistical distribution of the incident categories and subcategories at the time they are opened.

Source	Category	Target	Total	
aaa.operation	enduser	wwa.operation	26.19%	
		wwa.planning	16.67%	
		ras.engineering	7.14%	
		helpdesk.lsy	4.76%	
		multihost.sales	4.76%	
		tcxp.fieldservice	4.76%	
		ads.support	2.38%	
		firewall	2.38%	
		ham.is.tc.entry	2.38%	
		mes.switch.center	2.38%	
	enduser Total		73.81%	
	host/server	helpdesk.lsy	helpdesk.lsy	4.76%
			firewall	4.76%
cs.maintunixlinux			2.38%	
helpdesk.ham.is			2.38%	
svm.support.row			2.38%	
host/server Total		16.67%		
network	ras.engineering	ras.engineering	2.38%	
		helpdesk.lsy	2.38%	
		noc	2.38%	
network Total		7.14%		
pcicm	clic.server.row	2.38%		
pcicm Total		2.38%		
aaa.operation Total			100.00%	
ads.support	enduser	helpdesk.lsy	68.75%	
		internet	12.50%	
		noc	6.25%	
	enduser Total		87.50%	
	host/server	helpdesk.lsy	helpdesk.lsy	6.25%
cs.win.engineering			6.25%	
host/server Total		12.50%		

Table 6: Incident transition frequency pivot table

What Skyes-IT found most useful in providing insight in the functioning of the support organization was the information on historical frequency of transition of incidents between support group, with the ability to further break results down by category and subcategory, a minimal portion of which is shown in Table 6. In this example above the transitions are grouped by category information per assignment group. As an aside, we observe that this table is equivalent to the transition matrix required by SYMIAN for its simulations.

What the IT managers at Skyes-IT found extremely valuable though, was the graph representation of the incident transitions that highlights what nodes (support groups) see the highest number of incidents and what edges (transition probability) carry the most incidents between assignment groups. The graph in Figure 26 represents a graphic rendition of some of the information in the incident transition frequency previous table. The frequency information is represented through the thickness of the arrows, and the traffic (number of incidents per time period treated by an assignment group) through the size of the node.

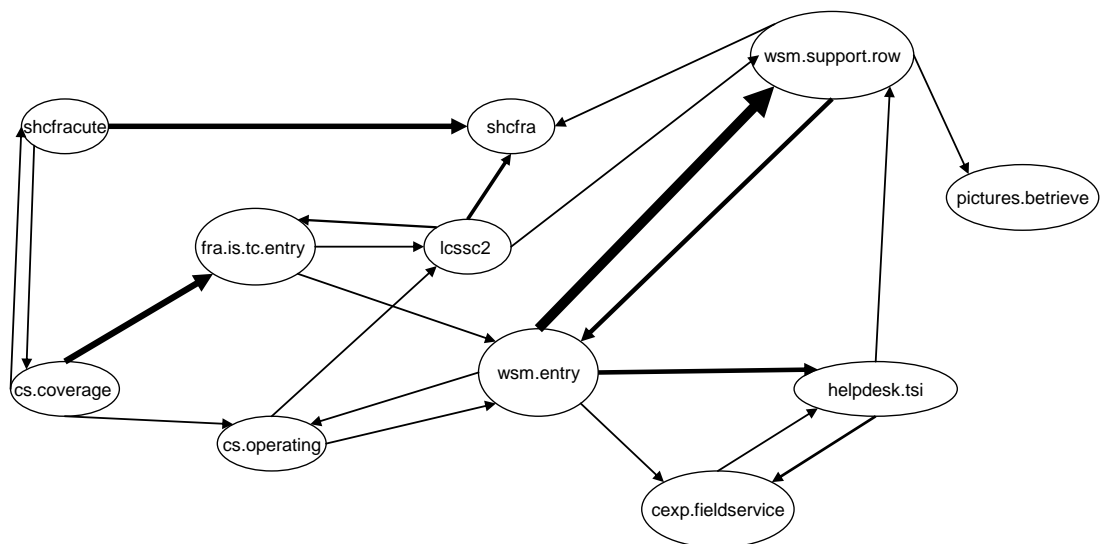


Figure 26: Graphical representation of transition matrix

Besides being a contribution of its own, the graphical representation of the incident transition matrix is at the heart of the visualization tool that we used to drive SYMIAN simulations and that we describe in the next section.

6.5.1 ITSupportster

Given the complexity of real life IT support organization, the sheer volume of information on the values of the performance metrics makes it difficult for an IT manager to understand the structure and the dynamics of an IT support organization than is currently available through state-of-the-art software. We addressed this problem through advanced visualization techniques, inspired to social networks. In this section we describe ITSupportster [80], a visually rich decision-support system applied to incident management performance evaluation and optimization. ITSupportster provides insight into the structure and the dynamics of an IT support organization.

The main view of ITSupportster is a map of the workgroups that compose the IT support organization, as shown in Figure 27. Workgroups are laid out on the map according to inter-group communication dynamics: workgroups that communicate often – i.e. redirect tickets from one another – are near to each other on the map. This is similar to the concept of social networks where entities that are marked as “friend” are drawn near to one another. All of the data necessary to populate the ITsupportster social network map is collected through off-the-shelf IT help desk software products. To develop our prototype, we gathered data from the same installations of HP Service Manager that we used for our Skyes-IT pilot, as described in section 6.3.

Various pictorial features on the map are used to convey information. A prominent feature of the visualization is that the size of the node is proportional to the number of tickets that the workgroups process. In this way it is immediate to get an impression

of what workgroups see the most traffic. The thickness of the directed edges connecting nodes is used to represent the ticket throughput between the workgroups. A thicker edge represents a higher number of tickets being passed from one group to the next. In this edges come to represent “information highways”. The thickness of the node border is proportional to the number of tickets resolved by the support group represented by the node. The coloring can be used to represent the level of support (helpdesk, first line of support, etc.)

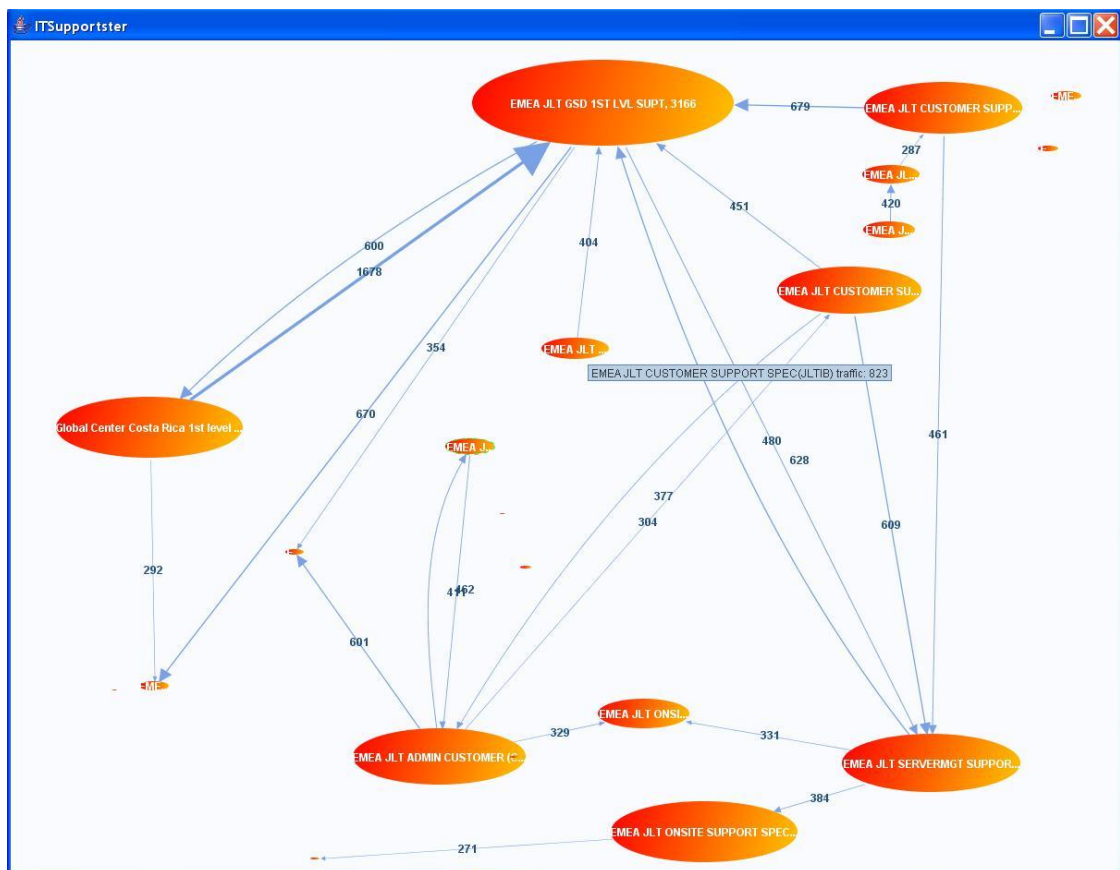


Figure 27: ITSupportster main view

It has to be noted that the mapping of visual features to metrics for support group dynamics is highly customizable. For example a user may want the dimension of the node to indicate the staffing level of the workgroup, or the coloring to indicate the

geography that the workgroup belongs to. Finally, node vicinity (the map layout) could also be customized to represent other features. For example, the nodes could be layered as to represent to support level hierarchy or even drawn on a geographic map.

Given the elevated number of nodes on the graph (the number of workgroups in an IT large support organization can be up to thousands), the map is represented through a hyperbolic graph [81]. The main feature of a hyperbolic graph is that the region of the map that is in focus is rendered with greater detail, while still giving a picture comprising of the whole map. Zoom-in and zoom-out are supported.

6.5.2 Driving IT organization optimization through a visual tool

Because of the extremely high density of information that it conveys, the main ITSupportster map is the ideal starting point for IT performance analysis, business impact analysis and assisted optimization (re-design) of the IT support organization.

Highlighting a node representing a workgroup on the map, a side pane (workgroup view) shows structural information about the workgroup (name, manager, contacts, location, technicians, etc...), and collates information useful for IT performance analysis that is collected through the SYMIAN tool (section 6.4.1).

The architecture of ITSupportster follows the Model-View-Controller architectural pattern. The social network map that is used as a view in the IT performance analysis and business impact analysis modalities, is also used as controller in the guided optimization (or assisted organization design) that is available through SYMIAN. Using the social network map as a controller, the user is enabled to hand it out SYMIAN to carry out what-if scenario analyses that guides them in the assisted organization design step, as described in section 6.4.2.

Pictorially, the SYMIAN use cases such as merging support groups are enacted by dragging workgroups onto one another (Figure 28). Splitting support groups is done

by highlighting a group and selecting the split option from a menu. By default, the newly created group is supposed to receive half of the ticket traffic of the original workgroup, and have a fan-out that is identical to the original group.

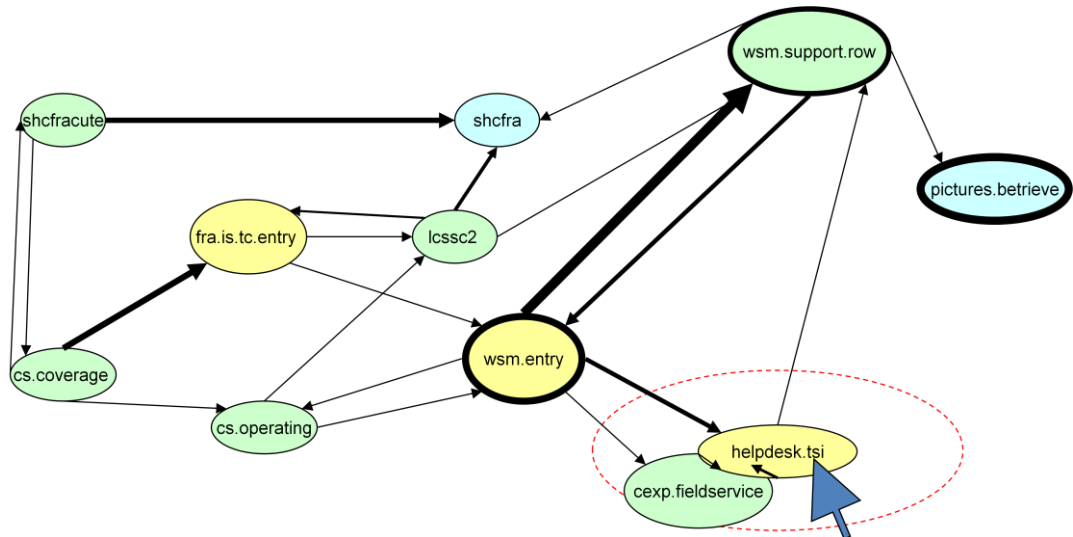



Figure 28: ITSupportster use case: merging workgroups

However, as seen in section 6.4.2, SYMIAN allows a user some customization, and for example, a slider-like control can be used to determine what share of the tickets the newly created group is expected to see. What-if analyses where staffing levels are modified are driven through the map view by just dragging technician icons in, out and between workgroups, as seen in the mock-up represented in Figure 29.

Move staff between groups by drag-and-drop
 Add staff by dragging from  to support group
 Remove staff by dragging from support group to 

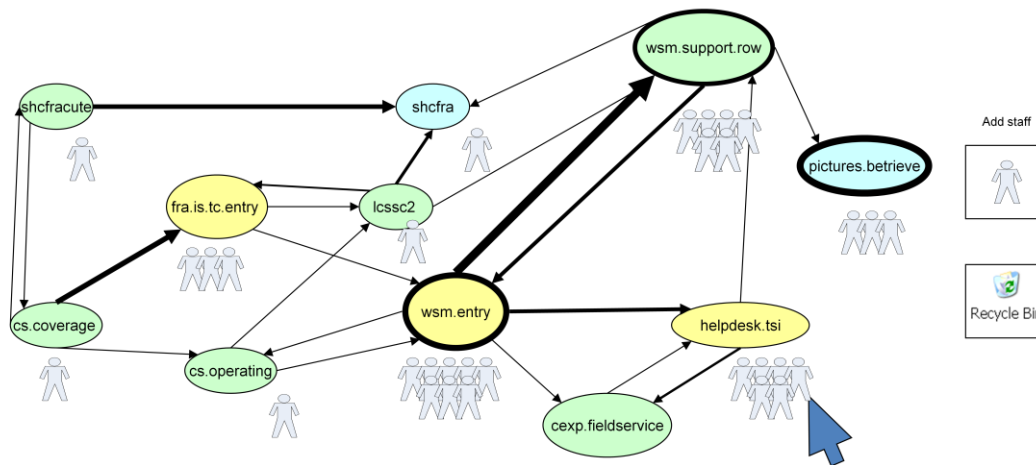


Figure 29: ITSupportster use case: changing staff levels

The usefulness of ITSupportster in these use cases is in that the user interfacing with SYMIAN has a far greater insight into the current state of support organization when experimenting with organization designs. Moreover, ITSupportster allows saving and retrieving of partial organization designs so as making the assisted design process smoother and more effective.

6.6 Putting it all together: a complete BDIM solution for organizational re-design aimed at optimizing help desk and incident management performance

In this section we describe a complete business-driven IT management solution for optimizing the performance of an IT support organization in incident management built using the components described in the previous sub-sections.

Starting from a set of critical service incidents, our solution, (introduced as HANNIBAL in [82]) proposes a set of candidate strategies for organizational re-

design, and policies for incident prioritization and routing¹¹. It then evaluates the performance of the incident management process in the context of each candidate strategy using SYMIAN (6.4.1) over the metrics described in 6.3. It calculates the cost of each candidate strategy, and finally reports which strategy has the least business impact using Aline (described in chapter 4)

6.6.1 HANNIBAL

Incidents that can bring to SLO violations, or *critical incidents*, are of particular interest for IT support organizations. In fact, while they are not as frequent as normal (non-critical) incidents, critical incidents have the most significant impact on business operations.

This suggests the implementation of special strategies for the management of critical incidents, optimized to reduce their business impact. In fact, the performance of an IT support organization in the incident management process is subject to dramatic variation depending on the effectiveness of critical incident routing and on the efficiency of each single support group in dealing with tickets.

As a result, the optimization of critical incident management should consider both **strategies** for organizational re-design (merging, splitting, creation and elimination of support group, changes in staffing levels) and **policies** increasing the *effectiveness* of incident routing (comparing alternate support group policies for forwarding and escalating incidents) and the *efficiency* in dealing with the most important incidents

¹¹ Future versions of the tool will be integrated with the ITSupportster model-view-controller (section 6.5.1) and will include options for organization re-design as described in section 6.4. This has not yet been done at the moment to wrap up this thesis, and the version that we used for experimental evaluation only uses strategies as policies for incident prioritization and routing. The skilled reader should not have difficulty envisaging how to extend the solution to cater for the SYMIAN use cases for organizational re-design.

first (comparing alternate prioritization schemata for extracting incidents from support group queues).

However, the assessment of both the *tangibles* (immediately visible costs due to SLO penalties, hardware and personnel) and the *intangibles* (“hidden” costs due to increase/reduction of incident response efficiency) business impact factors of various strategies is a very complex process, and calls for support tools to enable informed and accurate decision making.

HANNIBAL is a business-driven decision support tool for the selection of strategies in critical incident management. HANNIBAL enables business managers to make well informed decisions about the critical incident management processes, at the organizational, structural, and behavioral level.

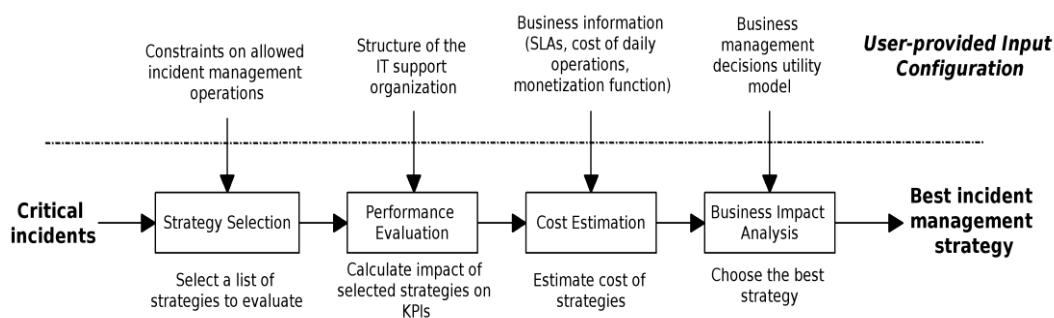


Figure 30: HANNIBAL decision process

The HANNIBAL decision support tool embodies a 4-phase process, as described in Figure 30. Starting from a set of critical incidents, HANNIBAL proposes a set of candidate strategies for organization optimization (**Strategy Selection** phase), evaluates the performance of the incident management process in the context of each candidate strategy (**Performance Evaluation** phase), calculates the cost of each candidate strategy (**Cost Estimation** phase), and finally reports which strategy has the least business impact (**Business Impact Analysis** phase).

HANNIBAL also integrates reporting functions, which provide a detailed analysis of strategy evaluation process, thereby offering a comprehensive set of information to support business decisions.

The rest of this section discusses in details the 4 phases of the HANNIBAL decision support process: *Strategy Selection*, *Performance Evaluation*, *Cost Estimation*, and *Decision Making*.

Strategy Selection

In the **strategy selection phase**, HANNIBAL considers several strategies for performance optimization. The strategies taken into account cover the whole space of allowed options, according to the user-specified constraints.

Before proceeding to the next phase, a user can modify the set of selected strategies for evaluation. For instance, users may want to further refine or restrict subset of the proposed incident management strategies, and/or add new candidate strategies to be considered in the evaluation process. Besides exposing the SYMIAN use cases for organizational re-design (section 6.4), there are two main performance aspects to take into consideration in policy design: *support level efficiency* and *incident routing effectiveness*.

In order to address *support group efficiency*, HANNIBAL considers different policies for critical incident prioritization. The policies are represented by criteria over fundamental attributes of the incidents, in particular incident priority levels. Example of such criteria are: “*if the priority level is low priority, when the incident is escalated to another support group, it is put at the end of the incoming incident queue*”; “*if the priority level is high priority, when the incident is escalated to another support group, it is put at the front of the incoming incident queue*”; “*if the priority level is very high*

priority, when the incident is escalated to another support group it is immediately assigned an operator, thereby preempting other incidents”, and so on.

In order to address *incident routing effectiveness*, HANNIBAL also considers different incident routing strategies, such as dictating assignment to specific support groups for incidents of a given category, or criteria to select between equally plausible forwarding alternatives. More specifically, the support group routing strategies for choosing between plausible alternatives include assignment of critical incidents to the support group with the *shortest incident response* (shortest average time spent by incidents waiting on incoming incident queue), to the support group with the *lowest load* (largest operator idle/busy time ratio), to the support group which provides the *most appropriate skills set* for incident resolution, to the *largest* support group (largest operator set), to the most appropriate support group taking into account *geographies and time zones*, or to a *random* support group. In the context of the selected support group, different operator assignment strategies are considered, such as assignment to the *best skilled* operator, and assignment to the *first available* (random) operator.

Performance Evaluation

In the **performance evaluation** phase, HANNIBAL estimates the impact of the strategies selected at the previous phase on the IT support organization performance. In particular, this phase evaluates the impact of strategies on the KPIs.

The performance evaluation task is performed via *what-if scenario analysis*. To this end, HANNIBAL leverages on the SYMIAN tool to reenact IT support organization processes, as fully described in section 6.4.

Cost Estimation

In the **cost estimation** phase, HANNIBAL calculates the cost of implementing the strategies under evaluation in the IT support organization. The calculation has to consider several factors: the costs for the strategies implementation itself, the costs related to SLO violations occurred in the context of the strategies, and variations of normal operations costs due to strategy implementation.

The costs directly related to the strategies implementation depend from the specific strategies considered. For instance, the cost of adding new operators to a specific support group must consider the costs for operator training, equipment, and salary. The cost of implementing software/hardware replacement and/or upgrades, instead, must consider the cost for buying new software/hardware and the cost for installation, configuration, and training. As a result, HANNIBAL requires the user to provide specific implementation costs for each strategy to be evaluated.

HANNIBAL also considers SLO violation penalties due to strategies implementation. To this end, HANNIBAL requires users to define the conditions in which SLO violations occur and their penalty amount. The tool then uses the values of service level indicators obtained from the previous phase to find whether SLO violations occur in the context of the strategies under evaluation, and calculates their costs.

Business Impact Analysis

In the Business Impact Analysis phase, HANNIBAL calculates the business impacts of strategies and ranks them based on their alignment to a given set of business objectives (from which a business impact function is calculated).

HANNIBAL models business objectives following the IT-MBO information model described in chapter 4.1. We briefly recall here that a business objective is defined by

defining a *target region* (usually expressed through a constraint) over a *KPI* (key performance indicator). A business objective is met when the specific KPI value lies within the target region at the end of the evaluation period. Each business objective has an associated *weight* - which is configurable by the HANNIBAL user - and expresses user preferences in terms of relative importance of business objectives. The IT-MBO information model for business objectives used by HANNIBAL models business management preferences using weights based on the Balanced Scorecard concept [36]. Importance weights are associated to perspectives of the business scorecard, and in turn these are further modified by weights associated to the objective proper. Examples of objectives used in the validation of HANNIBAL are shown in Table 7.

HANNIBAL then uses Aline (chapter 4) to compute the *alignment* of all the evaluated strategies with the user-provided business objectives to find the strategies with the minimum business impact.

6.6.2 HANNIBAL Architecture and Implementation

HANNIBAL implements each of the four phases of the decision support pipeline through a specific component. These are the *Strategy Selector* component, the *SYMIAN* tool (6.4.1) for what-if scenario analysis, the *Cost Analyzer* component, and the *Aline* alignment engine for business impact calculation.

The Strategy Selector component implements the selection of strategies according to user-provided configuration and constraints. The output of this phase represents the whole space of candidate strategies to consider for the evaluation.

Business objective	KPI	Target region	Obj. Wt.	Final Wt.
<i>External Perspective – Importance weight: 0.3</i>				
<i>Customer Satisfaction</i>	Total Number of SLO violations	Less than 10 violations	1.0	0.30
<i>Financial Perspective – Importance weight: 0.7</i>				
<i>Cost of implementing new strategies</i>	Total cost of implementing new strategies	lower than 50,000 \$ per three month period	0.6	0.42
<i>Aggregated cost for SLO penalties</i>	Total Cost of SLO penalties	lower than 10,000 \$ per month	0.4	0.28

Table 7: Example business objectives

SYMIAN (section 6.4) is the component realizing the performance evaluation phase. SYMIAN implements an accurate model of IT support organizations which allows, via discrete event simulation, to reproduce their behaviour and to evaluate their KPIs in the context of each candidate strategy for critical incident management.

The *Cost Analyzer* component implements both performance analysis and the cost analysis of strategies on the incident management process.

Aline (section 4.2) calculates the alignment of strategies with business objectives, and compares them to find out which one has the minimum impact on business.

The components implementing the decision support process are supported by other components, implementing coordination and auxiliary functions.

The *Coordinator* component directs and supervises the decision support process. It connects the decision support components together, performing ad-hoc transformation and processing of data when necessary.

The *User Interface* component allows users to define configuration parameters, to launch the decision making process, and to save its outcomes to file. The User Interface component provides both an interactive textual and a non-interactive command-line interface.

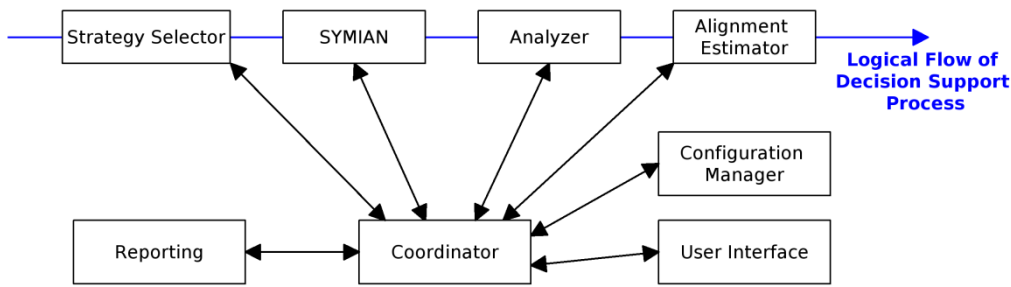


Figure 31: Architecture of HANNIBAL

The *Configuration Manager* takes care of all the aspects regarding HANNIBAL configuration. Among the functions provided by the Configuration Manager are configuration file parsing and validation of configuration parameters.

The *Reporting* component provides reporting functions in order to offer a comprehensive set of information to support business decisions. The Reporting component implements statistical analysis functions and integrates with the Gnuplot data visualization tool (<http://www.gnuplot.info/>) to plot time-varying values such as incoming incident queue sizes at the support groups.

HANNIBAL is implemented in the Ruby programming language (<http://www.ruby-lang.org/>). Ruby was chosen for its excellent support for rapid application development, by means of its remarkable extensibility and its capabilities to define domain-specific languages, and for the availability of a wide range of high-quality scientific libraries and tools.

The Ruby language is particularly well suited for the implementation of discrete event simulation-based tools like HANNIBAL, and offers satisfying performance levels which allow HANNIBAL simulations with a volume of incidents up to several tens of thousand to run in a few minutes.

6.6.3 Experimental Evaluation

This section presents an experimental evaluation of the HANNIBAL effectiveness in the analysis and optimization of the incident management process. Here, HANNIBAL is applied to optimize a case study IT support organization modeled according to real-life experiences. The experiments compare gains in performance when optimization is driven by business impact consideration rather than by IT level metrics.

Experiment Configuration

The IT organization subject of this experimental evaluation, *WOLFE INC.*, consists of a help-desk plus 2 support levels (*levels 0 – 2*), and 31 support groups, comprising of 348 technicians.

To limit the complexity of the case study, the organization model assumes the routing of incidents in the *WOLFE INC.* organization to be unidirectional, meaning that support groups of level N only receive incidents from support groups of level $N - 1$ and escalate incidents to support groups of level $N + 1$.

The experiments covered one month (31 days) of simulated time¹², starting from *August 12th 2008, 11:10AM UTC*.

WOLFE INC. deals with both non-critical and critical incidents. Non-critical incidents arrive continuously at a swift pace and need only a limited amount of work time before they can be closed, while critical incident arrive rarely and require a significant amount of work time for service restoration.

¹² The first day of simulated time was used exclusively to prime the simulation environment in order to prevent “cold start” measurements from affecting the simulation accuracy, and as such was not considered for the performance evaluation.

In all the experiments, non-critical incidents have random arrival times following an *exponential probability distribution* with a mean arrival rate of 30 minutes. The exponential probability distribution models non-critical incident arrival as a *memory-less* process. Non-critical incidents require a random amount of work at every support level, modeled according to a uniform distribution probability in the (0, 30 minutes] time interval.

The 3 critical incidents considered in the experiments were instead deterministically modeled, with regards to their arrival time and time to resolution at each support level. Table II provides their detailed characterization.

Critical Incident	Arrival time (from simulation start time)	Required work time for resolution
1	After 12 days and 5 hours	L0: 30 minutes L1: 1 day L2: 2 days and 12 hours
2	After 16 days and 9 hours	L0: 45 minutes L1: 1 day and 6 hours L2: 2 days
3	After 20 days and 11 hours	L0: 50 minutes L1: 16 hours L2: 1 day and 12 hours

Table 8: Characterization of critical incidents considered in simulation

The set of possible strategies¹³ considered in the optimization process was intentionally kept small, to demonstrate more clearly that the performance-driven and business impact-driven optimization processes lead to very different results.

¹³ Note: the current version of the HANNIBAL tool that we describe here is limited to the selection of policies for incident prioritization and routing. Using “*Strategy*” in our explanation below may therefore sound like a bit of a misnomer and “*Policy*” might be better instead. However in future version of the tools, strategies will also include options for organizational re-design such as merging, splitting support group and staff levels modifications. Because of this we keep with our use of “*Strategy*” in our explanation and we occasionally interchange it with “*Policy*” when the latter feels more appropriate.

More specifically, two policies were considered for incident prioritization: *high priority* and *normal priority*, and two policies for support group assignment: *random* and *largest*. For operator assignment, only the *first available* policy was considered.

Table 9 shows the four SLOs considered in these experiments. The first SLO imposes a maximum *MTTR* value of 10 hours. The second SLO states that no more than 530 incidents per month shall be closed in more than 2 hours from the time of their arrival. The third SLO states that all critical incidents shall be closed within one month (which coincides with the end of the simulation period). If any one of these SLOs is not met, the *WOLFE INC.* organization is charged a penalty of 10,000 \$ and a SLO violation is reported. The fourth SLO established an extra penalty of 10,000 \$ if more than two violations occur.

Condition	Penalty	Triggers violation
<i>MTTR</i> metric greater than 10 hours	10,000 \$	Yes
Time to closure greater than 2 hours for more than 530 incidents per month	10,000 \$	Yes
All critical incidents should be closed in the simulation period	10,000 \$	Yes
Violations > 2	10,000 \$	N/A

Table 9: Characterization of service level objectives

The business objectives considered for the optimization are presented in Table 10. Only two objectives were taken into account. The most important objective - with a weight of 0.65 - aims at keeping the total cost of monthly SLO penalties under 10,000 \$. The secondary objective - with a weight of 0.35 - is a measure of customer satisfaction, limiting the number of allowed SLO violations to two.

Before the optimization process, a performance evaluation of the *WOLFE INC.* organization in case of no critical incident occurrence was conducted. The purpose of

this evaluation is to obtain a benchmark serving as a comparison for the optimization outcome.

Business Objective	KPI	Target region	Wt
<i>Customer Satisfaction</i>	Total Number of SLO violations	At most 2 violations	.35
<i>Aggregated cost for SLO penalties</i>	Total Cost of SLO penalties	At most 10,000 \$ per month	.65

Table 10: Characterization of business objectives

Table 11 provides the values for the Mean Incidents Closed Daily (*MICD*) and Mean Time To (incident) Resolution (*MTTR*) metrics obtained from the simulation. The Mean Wait Time (*MWT*) metric, defined as the mean time spent in queues waiting for an available operator per closed incident, and the *MWT/MTTR* ratio are also provided as an indication on the efficiency of service restoration operations.

Total incidents arrived	661
Incidents arrived after warm-up threshold	645
Closed incidents	632
<i>MICD</i>	21.07
<i>MTTR</i>	10 hours, 21 minutes, and 41 seconds
<i>MWT</i>	9 hours, 37 minutes, and 37 seconds
<i>MWR/MTTR</i> ratio	0.93

Table 11: Performance analysis without critical incidents occurring

IT-driven Optimization

The first experiment optimizes performance using minimization of service disruption time as objective. Here HANNIBAL is configured to ignore business objectives and select the strategy which minimizes the *MTTR* metric.

Table 12 shows the HANNIBAL outcomes for the performance-driven optimization process: the number of incidents arrived, considered, and closed, the *MTTR* and *MWT* metrics, the characterization of the selected strategy for critical

incidents, the SLO penalties and violations occurred, and finally the strategy alignment with business objectives.

Total incidents arrived	664
Incidents arrived after warm-up threshold	648
Closed incidents	635
MICD	21.16
MTTR	10 hours, 59 minutes, and 35 seconds
MWT	9 hours, 55 minutes, and 16 seconds
MWT/MTTR ratio	0.90
Selected strategy for critical incidents	incident prioritization => high, supp. group assignment => largest, oper. assignment => first available
SLO penalties	20,000 \$
SLO violations	2
Alignment	35%

Table 12: Results of optimization driven by maximizing performance with respect to IT metrics

Analyzing the data presented in Table 12, it is possible to notice that the 3 critical incidents do not significantly impact the *MTTR* metric (only a 6% increase compared with the reference value in case of no critical incident occurrence), and that the selected strategy satisfies the first business objective, as only 2 SLO violations occurred. However, the total amount of SLO penalties was well above the 10,000 \$ threshold set by the second business objective. As a result, the strategy selected by HANNIBAL has only a rather low (35%) value of alignment with the given business objectives.

Business impact-driven Optimization

The objective of the business impact-driven optimization process is the selection of the strategy for critical incident management scoring the highest level of alignment with the given business objectives. Table 13 shows the HANNIBAL outcomes for the business impact-driven optimization process.

Total incidents arrived	664
Incidents arrived after warm-up threshold	648
Closed incidents	636
MICD	21.2
MTTR	11 hours, 22 minutes, and 28 seconds
MWT	10 hours, 18 minutes, and 10 seconds
MWT/MTTR ratio	0.91
Adopted strategy for critical incidents	incident prioritization => high, supp. group assignment => random, oper. assignment => first available
SLO penalties	10,000 \$
SLO violations	1
Alignment	100%

Table 13: Results of business-driven optimization

The analysis of the data in Table VII shows that the selected strategy has a significant impact on the *MTTR* metric (a 9,8% increase compared with the reference value in case of no critical incident occurrence). However, both the SLO violations and penalties are below the thresholds set by business objectives. As a result, the strategy selected by HANNIBAL has the maximum (100%) value of alignment with the given business objectives. This case is representative of the fact that optimizing for IT metrics does not necessarily result in the best possible business performance.

6.7 Discussion

In this chapter we demonstrated the full usefulness of business-driven IT management, showing an application of it to architect, design and implement a comprehensive **business-driven solution for organizational re-design of an IT support organization** in order to optimize its performance with respect to its help desk function and incident management process.

The thorough analysis of the help desk function and corresponding incident management process is in itself a contribution. We built on that to show how the

performance of the organization in terms of its contribution to the business can first be measured at the workgroup level rather than through obvious organizational wide metrics that are not actionable, such as the ones that are suggested by ITIL and COBIT. We then presented a what-if analysis tools that is able to show through simulation how the performance of the IT organization can be improved through organizational re-design (merging, or splitting workgroups, changing staffing levels, appropriately selecting prioritization policies at the workgroup level). This is possibly the main contribution of this chapter. As our experimental simulations showed, the simulation tool brings benefits already when applied to IT metric. The performance gains are even more meaningful when we combine our tool the other component of our BDIM methodology that we described in previous chapters and tackle the problem of optimizing for business results rather than for performance measured by IT metrics alone.

7 Impact of the research and conclusions

Research in Business-driven IT management (BDIM) aims at ensuring successful alignment of business and IT through thorough understanding of the impact of IT on business processes and business results, and vice versa. As we showed in our review of the state of the art of BDIM research and the open challenges, the field is young and exciting and very promising. On the other hand it is way too vast for any researcher or small group to tackle it all. We therefore decided to limit the scope of our contribution to IT service management (ITSM), and within ITSM we tackle problems that are amenable to decision support rather than automation.

Our first contribution is therefore a decision theoretical framework for BDIM that moves from a constructive and quantitative re-definition of some terms (such as business impact, risk and urgency) that are very widely used in ITSM but for which there's want of precise definitions. Our decision theoretical framework and models for BDIM bring the concepts of business impact and risk to the fore, and are able to cope with both monetizable and intangible aspects of business impact.

Our second contribution is a methodology for IT-business linkage that builds on the decision theoretical framework and exploits a re-definition of business impact through (mis-) alignment with business objectives, which in turn is defined as the likelihood – to the best of one's knowledge – that the objectives will be met. The methodology naturally entails the definition of a tool for computation of business impact, the Aline alignment computation engine. We put it all together by showing a sample BDIM solution for incident prioritization that can be used as a blueprint for other decision problems to do with other ITSM processes, such as change management for example.

Our third and possibly most important contribution is a second fully fledged application that shows the full power of BDIM. While incident management is used as a scenario for this second application as well, this BDIM application is really a comprehensive solution for business-driven organizational redesign to optimize the performance of an IT support organization. Each of the components developed to create the solutions are meaningful contributions in their own rights: the SYMIAN simulator, is built on a thorough analysis of the incident management process. It also exploits our work in defining metrics for assessing the performance of the IT support organization along the dimension of effectiveness (routing of incidents) and efficiency (speed of dealing with incidents within one support group), as well as our work in data mining applied to log of real life support organizations. As we demonstrate that the process of routing incidents is memory-less we can derive a transition matrix that represents the probability of re-directing incidents from group to group. And again, that matrix becomes a contribution in itself when we use a graphical representation derived from it and inspired to the study of social network (ITSupportster) to be used as the controller (in a model-view-controller pattern sense) of our comprehensive BDIM solution for organizational redesign. We show that the techniques we use - in particular the simulation of an IT organization enacting the incident management process – bring considerable benefits both when the performance is measured in terms of traditional IT metrics (mean time to resolution of incidents), and even more so when business impact metrics are brought into the picture, thereby providing a justification for investing time and effort in creating BDIM solutions.

Finally, in terms of impact measures, the work presented in this thesis produced about twenty conference and journal publications, and resulted so far in three patent applications. Moreover it has greatly influenced the design and implementation of

Business Impact Optimization module of HP DecisionCenter™, a leading commercial software product for IT optimization, whose core has been re-designed to work as described here.

Acknowledgement

This is the only section in this thesis where I use the “I” pronoun, rather than “we”. The main reason for my using “we” is that the work behind this dissertation owes an immense amount to the help and dedication of several colleagues, many of which appears as co-authors of the publications that first introduced some of the work that I presented here.

This dissertation and all it represents for me wouldn’t have been possible without the help and support of Cesare Stefanelli and Mauro Tortonesi, of the University of Ferrara, Italy. In particular, Mauro singlehandedly realized SYMIAN and HANNIBAL and contributed to a lot of the ideas presented here.

I’m extremely grateful to Mathias Sallé, David Trastour, Athena Christodoulou, Chris Preist, Maher Rahmouni, Kirill Sotnikov (who implemented the Aline incident management demonstrator), Issam Aib, Andrew Farrell, Silvia Vecchi, Rodrigo Rebouças, Weverton Cordeiro, Guilherme Machado, Andrew Byde, Marianne Hickey, Fabio Casati, Vijay Machiraju, Akhil Sahai, John Wilkes, Sven Graupner, Sharad Singhal, Bryan Stephenson, Mehmet Sayal, Malu Castellanos, Hamid Motahari, Kay-Yut Chen, Marco Casassa-Mont, Simon Shiu, Adrian Baldwin, Yolanta Beres, Patrick Goldsack, Dejan Milojicic, Chris Whitney, Mike Yearworth, Harumi Kuno, Lucy Cherkasova, Phil Day, Chris Peltz, Keith Jahn and Katia Saikoski, all of whom at various stages were part of the Business-driven IT Management project at HP Labs or other closely related projects. Most of the ideas presented here came from my continued interaction with them, and most of them duly figure as co-authors of the works behind this dissertation. I thank my managers at HP Labs for their support and flexibility during this time, in particular Abdel

Boulmakoul, who's been also been a close collaborator, a co-author and most importantly a friend, Giuliano Di Vitantonio, Martin Sadler and Martin Merry.

The Bottom Line team at the Federal University of Campina Grande, Brazil, can be called a center of excellence in BDIM. I have enjoyed a long standing collaboration with Antão Moura, Jacques Sauv , Rodrigo Rebouças and Guga de Oliveira Neto. The review of the literature of BDIM presented here borrows much from their thorough study and classification of the field. Their ideas have been a great source of inspiration for me and they have provided me with wonderful sparring partners in many a debate.

Along with some of the colleagues mentioned above, Mark Burgess, Hanan Luftiyya, Kamal Bhattacharya, Heiko Ludwig, Vladimir Tomic, Luciano Paschoal Gaspar, Lisandro Zambenedetti Granville, Jeff Kephart, Alex Keller, Yixin Diao, Chris Ward, Raouf Boutaba, Heinz-Gerd Hegering, Thomas Schaaf, Andreas Hanemann, Arkady Zaslavsky, Amir Padovitz and many others researchers in industry and academy have been instrumental in establishing BDIM as an exciting and prolific research field. I'm particularly grateful to Akhil, Jacques and Kamal for sharing with me the burden and the joy of organizing the series of IEEE BDIM workshops.

The HP (formerly Peregrine) IT Analytics (formerly DecisionCenter) team has provided me with the opportunity of trying out some of the ideas presented here in a real-life software development environment. A lot the work on incident management and IT optimization presented in this dissertation is the brainchild of Mike Lazarus, Lionel Fradin, Bud Farnham, Gilad Barash, Liya Wu, Barbara Newton, Tim Knight, Myles Suer, Todd Bashor and many other colleagues. My interaction with Gilad and Lionel in particular has been fruitful and very fulfilling. Gilad has run many of the

experiments the results of which are presented as validation of the ideas behind this thesis. Along with Liya, he's also appeared as a co-author of pieces of work behind this dissertation.

Finally, I owe a lot to Suan for her unconditioned and unwavering support; and I'm thankful to her again and (indirectly) to Sri Tim Miller for providing me with the opportunity of writing up in an extraordinary setting, Maya Tulum, Mexico.

This work is dedicated to Suan, Silvia and Diego and most importantly to my parents, without whom I wouldn't have been possible.

References

- [1] IT Governance Institute, “COBIT 3rd Edition”, 2000, www.isaca.org/COBIT.htm
- [2] Public Company Accounting Reform and Investor Protection Act of 2002 – “The Sarbanes-Oxley Act 2002”, www.legalarchiver.org/soa.htm
- [3] Office of Government Commerce, “IT Infrastructure Library (ITIL) Lifecycle Publication Suite, Version 3: Continual Service Improvement, Service Operation, Service Strategy, Service Transition, Service Design”, Stationery Office, 2007
- [4] Van Bon, J., Chief Editor, “IT Service Management, an introduction based on ITIL”, itSMF Library, Van Haren Publishing, 2004
- [5] Bartolini, C.; Salle, M.; Trastour, D., "IT service management driven by business objectives An application to incident management," Network Operations and Management Symposium, 2006. NOMS 2006. 10th IEEE/IFIP , vol., no., pp.45-55, 3-7 April 2006
- [6] Aiber, S.; Gilat, D.; Landau, A.; Razinkov, N.; Sela, A.; Wasserkrug, S., "Autonomic self-optimization according to business objectives," Autonomic Computing, 2004. Proceedings. International Conference on , vol., no., pp. 206-213, 17-18 May 2004
- [7] Bartolini, C., Sauv , J. and Sahai, A. (eds) “Information Technology Management from a Business Perspective”. Proc. BDIM 2006
- [8] Bartolini, C., Sauv , J. and Sahai, A. (eds) “Information Technology Management from a Business Perspective”. Proc. BDIM 2007
- [9] Bartolini, C., Sauv , J. and Sahai, A. (eds) “Information Technology Management from a Business Perspective”. Proc. BDIM 2008

- [10] Bartolini, C., Bhattacharya, K. and Sauvé, J. (eds) "Information Technology Management from a Business Perspective". Proc. of BDIM 2009 – to appear
- [11] Machiraju, V., Bartolini, C. and Casati, F., "Technologies for Business-Driven IT Management", in "Extending Web Services Technologies: the Use of Multi-Agent Approaches", edited by Cavedon, L., Maamar, Z., Martin, D. and Benatallah, B., Kluwer Academic.
- [12] HP, "DecisionCenter Optimization Module white paper", available from www.hp.com
- [13] Moura, A.; Sauve, J.; Bartolini, C., "Research Challenges of Business-Driven IT Management," in Proc.2nd IEEE/IFIP International Workshop on Business-Driven IT Management, 2007. BDIM '07, vol., no., pp.19-28, 21-21 May 2007, IEEE Press
- [14] Moura, A; Bartolini, C; Sauvé, J.; Business-Driven IT Management - Upping the Ante of IT, IEEE Communication Magazine, Vol 46, No. 10, Oct 2008, pp 148-153.
- [15] Keller, A.; Diao, Y.; Eskesen, F.; Froehlich, S.; Hellerstein, J.I.; Surendra, M.; Spainhower, L.F., "Generic On-Line Discovery of Quantitative Models," Network and Service Management, IEEE Transactions on , vol.1, no.1, pp.39-48, April 2004
- [16] Menascé, D., Almeida, V.A.F., Fonseca, R. and Mendes, M.A., "Business-Oriented Resource Management Policies for e-Commerce Servers", Performance Evaluation 42, Elsevier Science, 2000, pp. 223-239.
- [17] Liu, Z., Squillante, M. S., and Wolf, J. L. 2001. On maximizing service-level-agreement profits. In Proceedings of the 3rd ACM Conference on Electronic

- Commerce (Tampa, Florida, USA, October 14 - 17, 2001). EC '01. ACM, New York, NY, 213-223.
- [18] Keller, A., Hellerstein, J., Wolf, J.L., Wu, K. and Krishnan, V., "The CHAMPS System: Change Management with Planning and Scheduling", In Proc. NOMS 2004
- [19] Bartolini, C., and Sallé, M., "Business Driven Prioritization of Service Incidents", In Utility Computing: 15th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2004, Davis, CA, USA, November 15-17, 2004. Proceedings, Springer, 2004.
- [20] Sauv e, J., Marques, F., Moura, A., Sampaio, M., Jornada, J. and Radziuk, E., "SLA Design from a Business Perspective", In: Ambient Networks, 16th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2005, Barcelona, Spain, October 24-26, 2005. Proceedings
- [21] Moura, A.; Sauve, J.; Jornada, J.; Radziuk, E., "A quantitative approach to IT investment allocation to improve business results," Policies for Distributed Systems and Networks, 2006. Policy 2006. Seventh IEEE International Workshop on , vol., no., pp.9 pp.-95, 5-7 June 2006
- [22] Abrahao, B.; Almeida, V.; Almeida, J.; Zhang, A.; Beyer, D.; Safai, F., "Self-Adaptive SLA-Driven Capacity Management for Internet Services," Network Operations and Management Symposium, 2006. NOMS 2006. 10th IEEE/IFIP , vol., no., pp.557-568, 3-7 April 2006
- [23] Schaaf, T., "Frameworks for Business-driven Service Level Management: A Criteria-based Comparison of ITIL and NGOSS," Business-Driven IT Management, 2007. BDIM '07. 2nd IEEE/IFIP International Workshop on , vol., no., pp.65-74, 21-21 May 2007

- [24] Schaaf, T.; Brenner, M., "On tool support for Service Level Management: From requirements to system specifications," Business-driven IT Management, 2008. BDIM 2008. 3rd IEEE/IFIP International Workshop on , vol., no., pp.71-80, 7-7 April 2008
- [25] Diao, Y.; Keller, A.; Parekh, S.; Marinov, V.V., "Predicting Labor Cost through IT Management Complexity Metrics," Integrated Network Management, 2007. IM '07. 10th IFIP/IEEE International Symposium on , vol., no., pp.274-283, May 21 2007-Yearly 25 2007
- [26] Diao, Y.; Bhattacharya, K., "Estimating business value of IT services through process complexity analysis," Network Operations and Management Symposium, 2008. NOMS 2008. IEEE , vol., no., pp.208-215, 7-11 April 2008
- [27] Tasic, V.; Pagurek, B.; Patel, K.; Esfandiari, B.; Ma, W., "Management applications of the Web Service Offerings Language (WSOL)", Information Systems, Volume 30, Issue 7, The 15th International Conference on Advanced Information Systems Engineering (CAiSE 2003), November 2005, Pages 564-586
- [28] Anderson, E.; Fornell, C.; Rust, R., "Customer Satisfaction, Productivity, and Profitability: Differences Between Goods and Services", MARKETING SCIENCE 1997 16: 129-145
- [29] Keeney, R. L.; Raiffa, H.; Rajala, David W., "Decisions with Multiple Objectives: Preferences and Value Trade-Offs," Systems, Man and Cybernetics, IEEE Transactions on , vol.9, no.7, pp.403-403, July 1979
- [30] Linstone. H.A., and Turoff, M., "The Delphi Method: Techniques and Applications", Addison Wesley, 2002.
- [31] Barrett, R.; Maglio, P.P.; Kandogan, E.; Bailey, J., "Usable autonomic computing systems: the administrator's perspective," Autonomic Computing,

2004. Proceedings. International Conference on , vol., no., pp. 18-25, 17-18 May 2004
- [32] Sauv , J. P. ; Santos, R. A.; Almeida, R. R. ; Moura, J. A. B. On the Risk Exposure and Priority Determination of Changes in IT Service Management. In: 18th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2007, 2007, San Jos , United States. Large Scale Management of Distributed Systems. Heidelberg : Springer Berlin, 2007.
- [33] Sauve, J., "Risk-based service testing," Business-driven IT Management, 2008. BDIM 2008. 3rd IEEE/IFIP International Workshop on , vol., no., pp.106-109, 7-7 April 2008
- [34] Sauv , J.; Santos, R.; Rebou as, R.; Moura, A.; Bartolini, C.; Change Priority Determination in IT Service Management Based on Risk Exposure, IEEE Transactions on Network and Service Management, Vol. 5, No. 3, Sept. 2008, pp. x-y – to appear
- [35] Salle, M.; Rosenthal, S., "Formulating and Implementing an HP IT Program Strategy using CobiT and HP ITSM," System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on , vol., no., pp. 236c-236c, 03-06 Jan. 2005
- [36] Kaplan, R. S. Kaplan and Norton, D. P., "The Balanced Scorecard: Translating Strategy into Action" Harvard Business School Press, 1996.
- [37] Diao, Y., Hellerstein, J.L., and Parekh, S., "A Business-Oriented Approach to the Design of Feedback Loops for Performance Management" in Proceedings of the 12th DSOM 2001, pp. 11-22.

- [38] Kephart, J. O. 2005. Research challenges of autonomic computing. In Proceedings of the 27th international Conference on Software Engineering (St. Louis, MO, USA, May 15 - 21, 2005). ICSE '05. ACM, New York, NY, 15-22.
- [39] Sahai, A.; Singhal, S.; Machiraju, V.; Joshi, R., "Automated policy-based resource construction in utility computing environments," Network Operations and Management Symposium, 2004. NOMS 2004. IEEE/IFIP , vol.1, no., pp.381-393 Vol.1, 23-23 April 2004
- [40] Aib, I.; Salle, M.; Bartolini, C.; Boulmakoul, A.; Boutaba, R.; Pujolle, G., "Business aware Policy-based Management," Business-Driven IT Management, 2006. BDIM '06. The First IEEE/IFIP International Workshop on , vol., no., pp. 55-62, 07-07 April 2006
- [41] Bucu, M.J., et al., "Managing eBusiness on Demand SLA Con-tracts in Business Terms Using the Cross-SLA Execution Manager SAM", International Symposium on Autonomous Decentralized Systems, April, 2002.
- [42] Sauv e, J.; Reboucas, R.; Moura, A.; Bartolini, C.; Boulmakoul, A.; Trastour, D., "Business-Driven Decision Support for Change Management: Planning and Scheduling of Changes", In Large Scale Management of Distributed Systems 17th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2006, Dublin, Ireland, October 23-25, 2006. Proceedings
- [43] Reboucas, R.; Sauve, J.; Moura, A.; Bartolini, C.; Trastour, D., "A decision support tool to optimize scheduling of IT changes," Integrated Network Management, 2007. IM '07. 10th IFIP/IEEE International Symposium on , vol., no., pp.343-352, May 21 2007-Yearly 25 2007
- [44] Trastour, D.; Rahmouni, M.; Bartolini, C., "Activity-based Scheduling of IT Changes", In Inter-Domain Management First International Conference on

- Autonomous Infrastructure, Management and Security, AIMS 2007, Oslo, Norway, June 21-22, 2007. Proceedings
- [45] Wei, H., Frinke, D., Carter, O., "Cost-benefit Analysis for Network Intrusion Detection Systems", In Proceedings of the 28th Annual Computer Security Conference, October 2001
- [46] Beulen, E., and Ribbers, P., "IT Outsourcing Contracts: Practical Implications of the Incomplete Contract Theory", in Proceedings of the 36th Hawaii International Conference on System Sciences Systems, 8(5), 1992, pp.365-387.
- [47] Schmidt, H., "Service Contracts Based on Workflow Modeling", in Services Management in Intelligent Networks 11th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2000 Austin, TX, USA, December 4–6, 2000 Proceedings
- [48] Alves, C.H.F.; Alencar, A.J.; Schmitz, E.A.; Ferreira, A.F., "A quantitative risk model-building method for offshoring IT applications," Business-driven IT Management, 2008. BDIM 2008. 3rd IEEE/IFIP International Workshop on , vol., no., pp.41-50, 7-7 April 2008
- [49] van Grembergen, W., "Strategies for Information Technology Governance", Idea Group Publishing, 2004.
- [50] Mingay, S., and Bittinger, S., "Combine COBIT and ITIL for Powerful IT Governance", in Research Note, TG-16-1849, Gartner Group, 2002.
- [51] Salle, M., "IT Service Management and IT Governance: Review, Comparative Analysis and their Impact on Utility Computing", HP Labs Technical Report HPL-2004-98, Palo Alto, CA, USA, June 2004.
- [52] Salle, M.; Di-Vitantonio, G., "Business Service Management: The Impact of IT Governance Models on IT Management Policies," Services Computing, 2006.

SCC '06. IEEE International Conference on , vol., no., pp.373-380, 18-22 Sept. 2006

- [53] Brynjolfsson, E., and Hitt, L., "Beyond Computation: Information Technology, Organization Transformation and Business Performance", In Journal of Economic Perspectives, 2000, 14(4), pp. 23-48
- [54] Carr, D., "IT doesn't matter", In Harvard Business Review, 81(5), 2003
- [55] Reed, K., "EQUITY and the Problem of Return on IT Investment", IEEE Software, Jan/Feb 2006, pp. 114-115
- [56] Van Der Zee, J.T.M. and De Jong, B., "Alignment is Not Enough: Integrating Business and Information Technology Management with the Balanced Scorecard", In Journal of Management Information Systems, Fall 1999, 16(2), pp. 137-156.
- [57] Reich, B.H. and Benzabat, I., "Measuring the Linkage between Business and Information Technology Objectives", Management Information Systems (MIS) quarterly, 1996, vol. 20, pp 55-81.
- [58] Pisello, T. and Strassmann, "IT Value Chain Management – Maximizing the ROI from IT Investments", Digital Publications from the Information Economics Press, 2004.
- [59] Sichel, D.E., "Computers and Aggregate Economic Growth: An Update ", Business Economics, 34(2), 1999.
- [60] Menon, N., Lee, B. and Eldenburg, L., "Productivity of Information Systems in the Healthcare Industry", Information Systems Re-search, 11(1), 2000, pp.83-92.

- [61] Inn, K.S., Dow, K.E., Grover, V., “A Reexamination of IT Investments and the Market value of the Firm – An Event Study Methodology”, *Information Systems Research*, 12(1), 2001, pp. 103-117.
- [62] Dutta, A., and Roy, R., “A Process-Oriented Framework for Justifying Information Technology Projects in e-Business Environments”, *International Journal of Electronic Commerce*, 9(1), 2004.
- [63] Morgan, J. N., “A Roadmap of Financial Measures for IT Project ROI”, *IT Pro Magazine*, IEEE Computer Society, January/February 2005, pp.52–57.
- [64] Morgan, J. N., “A Roadmap of Financial Management for IT Projects ROI”, *IT Pro*, January/February 2005, pp; 52-57.
- [65] Bakos, Y.J. and Kemerer, C.F., “Recent Applications of Economic Theory in Information Technology Research”, *Decision Support Systems*, 8(5), 1992, pp.365-387.
- [66] Peebles, C.S., Stewart, C.A., Voss, B.D. and Workman, S.B., “Measuring Quality Cost and Value of IT Services”, In *Proceedings of the Quality Congress*, 2001, pp.468-493.
- [67] Ki, Y.J. and Sanders, G.L., “Strategic Actions in Information Technology Investment based on Real Option Theory”, In *Decision Support Systems*, 33(1), 2000, pp.1-11.
- [68] Kohli, R. and Sherer, S., “Measuring Payoff of Information Technology Investments: Research Issues and Guidelines”, In *Communications of the Association for Information Systems*, Vol. 9, pp.241-268, 2002
- [69] Raiffa, H., *Decision Analysis: Introductory Readings on Choices Under Uncertainty*. McGraw Hill. 1997.

- [70] Schreiber, G.; Akkermans, H.; Anjewierden, A.; de Hoog, R.; Shadbolt, N.; Van de Velde, W.; Wielinga, B., "Knowledge Engineering and Management: The CommonKADS Methodology", The MIT Press, 1999.
- [71] Markowitz, Harry M. (1952). "Portfolio Selection". *Journal of Finance* 7 (1): 77–91.
- [72] Box, G.; Jenkins, G. (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day
- [73] Wolfram," *Mathematica*". Available at www.wolfram.com
- [74] Wolfram, "Mathematica Java Toolkit", Available at www.wolfram.com/solutions/mathlink/jlink/
- [75] Martello, S.; Toth, P., "Generalized Assignment Problem", in Ibaraki, T. et al. (Eds.): *Algorithms and Computation, Third International Symposium, ISAAC '92*, Nagoya, Japan, December 16-18, 1992, Proceedings. *Lecture Notes in Computer Science* 650 Springer 1992
- [76] Barash., G.; Bartolini, C.; Liya Wu, "Measuring and Improving the Performance of an IT Support Organization in Managing Service Incidents," *Business-Driven IT Management, 2007. BDIM '07. 2nd IEEE/IFIP International Workshop on* , vol., no., pp.11-18, 21-21 May 2007
- [77] Lemoine, A.J., *Network of queues - A survey of equilibrium analysis*, *Manag. Sci.* 24 (1977), 464-481.
- [78] Bartolini, C., Stefanelli, C. and Tortonesi, M., "SYMIAN: a Simulation Tool for the Optimization of the IT Incident Management Process", In *Managing Large-Scale Service Deployment, 19th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2008*, Samos Island, Greece, September 22-26, 2008. Proceedings

- [79] HP, “ServiceCenter”, available at www.hp.com
- [80] Bartolini, C., “IT Incident Management as a Collaborative Process: a Visualization Tool inspired to Social Networks”, In Proc. of the 4th International Conference on Collaborative Computing: Networking, Applications and Worksharing, Orlando, Florida, USA, November 13-16, 2008 CollaborateCom 2008
- [81] Herman, I., Melançon, G., Marshall, M.S.: Graph Visualization and Navigation in Information Visualization: A Survey. In: IEEE Transactions on Visualization and Computer Graphics, 2000
- [82] Bartolini, C., Stefanelli, C. and Tortonesi, M., “Business-impact analysis and simulation of critical incidents in IT service management”, In proc. IEEE Integrated Management, IM 2009, *to appear*

Table of Contents

Business-driven IT Management

Part I: The Problem of Business-driven IT Management

1	What is business-driven IT management (BDIM)?	11
1.1	BDIM applications	14
2	State of the art and research challenges in Business-driven IT management ...	17
2.1	Research on BDIM models.....	19
2.1.1	Modeling (legacy) IT infrastructure	20
2.1.2	Modeling of IT and business processes.....	21
2.1.3	Modeling human behavior:	22
2.1.4	Modeling of financial aspects: monetary costs and benefits.....	22
2.1.5	Modeling of financial aspects: intangible costs and benefits.....	23
2.1.6	Modeling risk	23
2.1.7	Modeling viewpoints of multiple stakeholders:	24
2.1.8	Extensive validation:	24
2.2	Research on BDIM and Autonomic Computing	25
2.3	Research on BDIM over IT service management processes	27
2.4	Research on BDIM and IT Governance	30
2.5	Discussion.....	33

3	Enabling Decision Support in IT Service Management	35
3.1	Enabling Decision Support in ITSM: a decision theoretical framework based on a constructive and quantitative re-definition of some terms.....	40
3.1.1	Cost, income, net cost	41
3.1.2	Business impact.....	42
3.1.3	Utility functions harmonizing tangibles and intangibles (cost and business impact).....	43
3.1.4	Risk.....	45
3.1.5	Urgency	48
3.2	Mathematics of risk and urgency.....	49
3.2.1	Risk is the variance of utility.....	49
3.2.2	Urgency is the first derivative on time of utility	50
3.3	Worked examples	50
3.3.1	Risk.....	52
3.3.2	Urgency	53
3.4	Discussion.....	54
3.4.1	Risk.....	55
3.4.2	Urgency	57
4	An approach to business-IT linkage	59
4.1	The IT Management by business objectives (IT-MBO) information model.	61
4.1.1	Business Objectives.....	62

4.1.2	Key Performance Indicators.....	63
4.1.3	Perspectives.....	63
4.1.4	Examples.....	63
4.2	Aline: deriving Business Impact from Alignment with Business Objectives	64
4.2.1	Alignment with business objectives as the likelihood that objectives will be met.....	65
4.2.2	Using alignment with business objectives to derive a business impact function.....	69
4.2.3	Modeling cost as one of the dimensions of alignment.....	73
4.2.4	Modeling the impact of courses of action on the business objectives	74
4.2.5	Implementation of Aline using the Mathematica™ kernel and J/Link toolkit.....	78
4.2.6	Building BDIM solutions for IT service management using IT-MBO and Aline.....	79
4.3	Discussion.....	81

[Part II: Solutions for Business-driven IT Management](#)

5	A representative BDIM solution for decision problems in IT service manager: business-driven prioritization of incidents.....	85
5.1	Proof of concept BDIM solution for IT incident prioritization using IT- MBO and Aline.....	87
5.1.1	Likelihood of SLO violation as a function of the time of closure of the service degradation incident.....	88

5.1.2	Alignment with the business objectives in function of the likelihood of violation.....	89
5.1.3	Incident prioritization to maximize alignment with business objectives	92
5.2	Demonstrator of the BDIM solution for Incident prioritization	95
5.2.1	Part I: Business view of the solution.....	96
5.2.2	Part II: Configuration and use of the incident prioritization solution	99
5.2.3	Part III: Configuration and use of the incident prioritization solution... ..	100
5.3	Applicability to other decision problems in IT service management.....	104
6	A complete business-driven solution: organizational re-design for optimizing the performance of an IT support organization.....	105
6.1	Analysis of the incident management process.....	107
6.2	A reformulation in BDIM terms of the IT support organizational re-design problem	112
6.3	Metrics for performance evaluation of the IT support organization in dealing with incidents	114
6.3.1	Incident Routing Between Assignment Groups	114
6.3.2	Performance within Assignment Groups	116
6.3.3	Validation of the metrics in a real-life situation.....	117
6.4	Optimizing the incident management performance through simulation techniques	118
6.4.1	The SYMIAN Decision Support Tool	119

6.4.2	How SYMIAN works.....	120
6.4.3	The SYMIAN Simulation process	122
6.4.4	What-if scenario analysis	123
6.4.5	Implementation of SYMIAN	126
6.4.6	Interpretation of SYMIAN simulations results	127
6.4.7	Validating assumption of memory-less process.....	132
6.5	Advanced visualization techniques for guided performance analysis.....	134
6.5.1	ITSupportster.....	136
6.5.2	Driving IT organization optimization through a visual tool	138
6.6	Putting it all together: a complete BDIM solution for organizational re- design aimed at optimizing help desk and incident management performance.....	140
6.6.1	HANNIBAL	141
6.6.2	HANNIBAL Architecture and Implementation.....	146
6.6.3	Experimental Evaluation	149
6.7	Discussion.....	154
7	Impact of the research and conclusions	157
	Acknowledgement	161
	References.....	165

