Karlsruher Institut für Technologie

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

SLA Establishment Decisions:
Minimizing the Risk of SLA Violations

von

Dipl.-Inform.Wirt Wibke Anna Michalk

Tag der mündlichen Prüfung: 16.12.2011
Referent: Prof. Dr. Christof Weinhardt
Korreferent: Prof. Dr. Rudi Studer
Karlsruhe, 2011

Abstract

Traditional products in many industries recently experienced *servicification*. This phenomenon, that has already been described by Vargo and Lusch in 2004, describes changes in society and markets that foster a shift towards a service-centered view. In this vein, the focus of trade is on services while the importance of goods decreases. The emergence of advanced Web technologies leverages the provisioning and consumption of services over the Internet. Additionally, the presence of the Internet and the availability of Web-related technologies has influenced the offering and provisioning of services. The existence of platforms like Google's AppExchange or Amazon's Web Services manifest this trend. The legal framework for service outsourcing and thus, service provisioning and consumption is stipulated in service level agreements (SLAs). SLAs determine the objectives for service quality through service level objectives (SLOs), contain a price for service provisioning and a penalty in case of SLA violation. This way, SLAs set incentives for providers to adhere to SLAs. The provisioning of services underlies an inherent risk of service failure caused for instance by power outages, hardware malfunction or human failures, which leads to uncertainty concerning SLA violations that manifest in due penalties. Consequently, for a service provider it is of major interest, which SLAs should be established in order to minimize the risk of SLA violation.

This thesis presents a novel approach that enables service providers to select a particular combination of SLAs that minimizes the risk of SLA violation. Furthermore, the approach takes constraints on expected profit and available resources into account. This problem is addressed by applying methodologies from decision theory and approaches for measuring risk. In particular, the concept of portfolio selection by Harry Markowitz is adapted and extended in order to formulate the objective function of a service provider's decision for SLA establishment. In order to capture a decision maker's attitude towards risk, utility theory and the concept of risk aversion are employed to express a decision maker's preferences.

In this thesis, the risk of SLA violation is calculated from monitoring data of SLAs that were established in the past that comprises information on the degree of violation of an SLA. Therefore, the methods that solve the decision problem of SLA establishment are evaluated with respect to the amount of past observations that is required for calculating the risk of SLA violation. The results of the evaluation imply that between 10 and 100 observations suffice, depending on the employed method. This low number showcases the applicability of the approaches presented in this thesis for real-world scenarios, as the observations can be collected in reasonable time.

Acknowledgements

The completion of a work like this is only possible with the support of and collaboration with many people. Foremost I would like to thank my advisor, Prof. Dr. Christof Weinhardt, who gave me the great chance to gain a lot of experience in the ValueGrids project and the manyfold tasks at the Institute of Information Systems and Management (IISM) and the Karlsruhe Service Research Insitute (KSRI). He gave me the freedom to pursue this research and encouraged me in my ideas. I would like to show my gratitude to my co-advisor Prof. Dr. Rudi Studer for his guidance and fruitful discussions especially on the semantic aspects and possible extensions to this work. I would particularly like to thank the other members of the committee, Prof. Dr. Orestis Terzidis for the interesting discussions on practical implications of my thesis and Prof. Dr. Andreas Geyer-Schulz for the discussions on statistical aspects.

My sincere thanks go the teams of IISM and KSRI who improved my work through constant discussions and ensured a vivid social life with a large variety of social events. It is a pleasure to thank Dr. Simon Caton for his guidance, help, long discussions and for making me look beyond my own nose. Additionally, I want to thank Dr. Lilia Filipova-Neumann who has made available her support in numerous ways. I would like to express my gratitude to Dr. Sudhir Agarwal, Dr. Arun Anandasivam, Dr. Tobias Conte, Christian Haas, Rico Knapper, Jochen Martin, Thomas Meinl and Dr. Frank Schulz for proof-reading this thesis and for giving valuable comments that improved this work a lot. It is a pleasure to thank Annette Bannwarth, Christian Haas and Mercè Müller-Gorchs for their time and comments that helped to improve the presentation of this work.

I am also very grateful to Dr. Yudistira Dwi Wardhana Asnar of the University of Trento and to Dr. Patrick Hung of the University of Ontario Institute of Technology. They gave me the opportunity to shed light on new aspects of my work and included in me in their research groups. I gratefully acknowledge that the research stay was supported by a scholarship of the Karlsruhe House of Young Scientists (KHYS).

Finally, I would like to thank those who have affected me most. I owe my deepest gratitude to my parents, Karin and Detlev Michalk, for their support and unconditioned love, to my brother, Eike Felix Michalk, for reminding me that there is life beyond work and to Cornelius for his love, support and his caring encouragement. Without you, this work would not have been possible.

Wibke A. Michalk

Contents

List of Figures

List of Tables

Part I

Foundations & Preliminaries

Chapter 1

Introduction

"International trade in services is growing in importance both among OECD countries and with the rest of the world. Traditional services - transport, insurance on merchandise trade, and travel - account for about half of total international trade in services, but trade in newer types of services, particularly those that can be conducted via the Internet, is growing rapidly."

(OECD 2010)

WITHOUT doubts, services have become a major driver of value creation in the last decades (OECD 2010). This manifests itself in official statistics showing that services make up the largest part of the gross domestic product (GDP) in industrialized countries. In 2010, the share of the GDP within the European Union amounted to 73.2% and in the United States to 76.7%[1] increasing steadily over last years. The increasing share that service provisioning makes up of the GDP demonstrates a shift of focus away from value creation by producing goods towards revenue generation through service provisioning. According to Vargo and Lusch (2004b), goods are merely considered vehicles for the transportation of value that was created by service provisioning and the dominant logic that underlies economies experienced a *servicification*. Consider the case of a person that wants to build a house for their family. In accordance with Vargo and Lusch (2004b), the person procures the service of building the house rather than the accomplished house. More precisely, value is generated by the process of building the house and is carried by the tangible output, the house.

[1] `https://www.cia.gov/library/publications/the-world-factbook/fields/2012.html` last accessed: October 8, 2011

The kind of services and their way of provisioning is manifold. Traditional types of service include for instance logistics, that is efficient transportation and storage of people or goods, healthcare, that is in- or outpatient services, ambulances, supply of medication, and many more, or the provision of a large variation of insurances (OECD 2010). Newer types of services have emerged with the Internet since the end of the 1990s especially in the software branch. Software was traditionally developed as a monolithic product tailored to a customer's needs. However, after the delivery to the customer and installation on the customer's infrastructure, the software was out of the software developer's control and the customer could use the software as often and intensely as they wished. The rise of the Internet and the emergence of new (Web) technologies brought a change to the software industry.

The availability of the Internet enabled software developers to run software on their own infrastructure and to allow customers to access the software over the Internet. This new type of accessing software was leveraged by the emergence of technologies like SOAP, JSON (Crockford 2006), WSDL and RESTful architectures (Fielding 2000), which specified and standardized the access and communication to software services. This way, software was no longer required to be installed on customer's infrastructure, which lead to enormous reductions in hardware costs for the customer. Additionally, by running the software on the software developer's resources the control over usage duration and intensity of the software was back in the developer's hands. Awareness of the intensity and duration of usage by particular users enables the software producer to change software pricing. The installation on the customer's infrastructure and with it, the lack of control about usage, only allowed the selling of a perpetual-use license to the customer. With the knowledge about intensity and duration of usage, software that is run on the producer's infrastructure is eligible for price discrimination. Nowadays, usage- and subscription-based pricing models can be observed, which offer a monetary advantage to the customer (Choudhary 2007; Dubey and Wagle 2007). The customer is now able to request software *as a service* (SaaS) from a service provider, which reduces licensing and infrastructure costs as the customer is only billed for the actual consumed usage of the software and is not required to procure hardware.

Besides this change in pricing, the provisioning of software *as a Service* via the Internet brought about changes in the scope of software functionality. In order to enable the access to software services flexibly and instantly, providers offer standardized service modules, which provide a certain functionality but are not tailored to specific customer demands. Furthermore, by offering standardized services, the variety of application scenarios and with it, the potential number of customers, increases. Addressing potential customers experienced a shift in recent years. Before the advent of the Internet, service providers and producers of goods reached potential customers by advertizing in local newspapers, radio-broadcasting or television. Additionally, personal recommendations by former customers played an important role. By offering services via the Internet, a provider of specialized services is able to reach more customers and hence, is able to exploit even the small parts of overall demand, the so called *Long Tail* of demand (Anderson 2006). Being aware of this property of trading on the Internet, a provider is able to focus on their core-competencies and increase their demand.

However, offering standardized service modules brought about a change for customers. As software was no longer offered tailored to a customer's needs, customers are required to combine service modules in order to reach the required functionality. As service providers focus on their competencies service modules are provided by a diverse pool of service providers.

Analogously to contracts of purchase, which regulate the kind and specification of delivered goods, agreements about the delivered services are stipulated among service providers. Agreements serve to ensure a proper provisioning of services according to the customer's needs and the providers' abilities. These agreements are employed for specifying the functionality of the provided services, as well as non-functional aspects of service provisioning that specify aspects of the service that are not quality-related, for example the provision of temperature information in degrees Celsius for a weather forecast service. Additionally, quality aspects of the service are determined. Quality aspects specify *how* the service is provided, e. g. the percentaged availability of a weather forecast service. Furthermore, a price for service execution is agreed on. In order to enforce the adherence to the agreement, a penalty is stipulated, which is applied in the case of a violation of one or more of the quality goals. The penalty itself can take multiple forms, e. g. direct monetary payments, or a credit for the customer's next service request. It serves to reimburse the customer for the inaccurate provisioning of service and the resulting loss in utility and to incentivize providers to stick to the agreement (Becker et al. 2008).

For the joint provisioning of services, different aspects have to be considered. On the one hand, the technical composition of services is carried out with the help of common interfaces and protocols. On the other hand, aspects of service provisioning like quality, price and penalties are defined in service level agreements (SLAs). Especially in the context of on-demand provisioning and composition of services, a (semi-) automation of establishing SLAs is required. An approach, which defines a common structure, building parts and description language of SLAs is presented in Andrieux et al. (2007).

However, the composition of software service modules is only one example of joint service provisioning. Examples like Amazon Web Services (AWS) demonstrate that besides software, infrastructure like storage space[2], databases[3] and computational entities[4] are offered via the Internet. Furthermore, development platforms provide development tools for particular development languages to users in a flexible way without a need for the user to procure hardware and to install development software on their own infrastructure.[5] In the context of Cloud computing, the flexible provisioning of infrastructure, platforms and software is described by a layered model (Baun et al. 2009), which comprises infrastructure as a service (IaaS) as the downmost layer, platform as a service (PaaS) as middle layer and software as a service (SaaS) as topmost layer. Besides the flexible and instant provisioning of services, a major driver of Cloud computing is the virtualization of physical resources, which promises nearly unlimited scalability of the offered services (Baun et al. 2009). In this context, services can be composed horizontally, that is, on the same layer (Baun et al. 2009, p. 28), or vertically, that is, across layers. All of the above mentioned trends lead to the fact that services, which can be provided via the Internet have experienced an increasing momentum (OECD 2010).

Obviously, joint provisioning of services by a multitude of service providers is not only required in the case of services that are provided via the Internet. For the building of a house, numerous different services have to be carried out including the actual building of the walls, adding the roof and windows, tiling, and many more. Consequently, the cooperation and joint service provisioning and with it, SLAs are required for non-Internet services as well. However, according to the Long Tail phenomenon (Anderson 2006) the number of services that are requested from a provider via the Internet and with it, the number of SLAs that can be established exceed the demand for *physical* services by far. Thus, the decision on the establishment of SLAs for services that are provided via the Internet has to cope with a higher number of alternatives and consequently is more complex.

There are many aspects of joint service provisioning that have been subject to research in recent years. In analogy to the multi-step production of a good under participation of a multitude of producers, joint service provisioning was described in service supply chains (Prasad and Kalai Selvan 2009). In this context, traditional services like financial, healthcare, postal and courier, retail, entertainment and tourism services were examined and the difference of service supply chains from goods supply chains was

[2] `http://aws.amazon.com/s3/` last accessed: October 8, 2011
[3] `http://aws.amazon.com/simpledb/` last accessed: October 8, 2011
[4] `http://aws.amazon.com/de/ec2/` last accessed: October 8, 2011
[5] `http://www.salesforce.com` last accessed: October 8, 2011

elaborated. However, the decision about providing services or not and the inherently coupled decision about the establishment of SLAs was not covered explicitely. Furthermore, the impact that the rise of the Internet has on SLA establishment was not considered.

The technological aspects of joint service provisioning via the Internet gained enormous research interest. Approaches for service engineering, service description, service discovery and the description of service level agreements were and still are investigated (Scheithauer et al. 2009; Barros and Dumas 2006; Christensen et al. 2001; Toch et al. 2007; Küster et al. 2007; Andrieux et al. 2007). However, these works lack specific economic considerations of joint service provisioning like provider's preferences towards profit or the utility resulting from service provisioning. Nevertheless, economic aspects of joint service provisioning were in the focus of researchers. For instance, approaches that maximize total welfare that consists of a customer's and providers' utility by selecting service offers from a set of available alternatives were developped (Blau 2009; Conte 2010). Whereas these approaches consider technical requirements that arise with the provisioning of services via the Internet, the focus of the methods that are provided lays on the joint provision of a single service and the development of a method for the efficient choice of service modules for this particular complex service. However, the decision that a provider makes about the establishment of SLAs in the context of joint service provisioning of a multitude of different complex services has not yet been investigated.

The decision about the establishment of SLAs is the focus of this thesis. The following section derives Research Questions that enable the modeling of the scenario for decision making, the definition of a method for solving the decision problem and the evaluation of the method.

1.1 Research Outline

The core research activity in this thesis is dedicated to the decision on the establishment of SLAs. In order to develop and evaluate a method for decision making about the establishment of SLAs, a thorough understanding of the situation in which the decision is made and the factors that influence the decision maker is required. Therefore, the first research question of this thesis addresses *the stipulation of Service Level Agreements*. In more detail, it explores the terms that regulate service provisioning in an SLA and which resources that are available to the provider of a service along with economic expectations impact a provider in their decision about the establishment of an SLA.

Research Question 1 ≺Stipulation of Service Level Agreements≻. *What aspects of service provisioning do a provider and a customer of a service stipulate in a Service Level Agreement and which technical properties and economic preferences of a service provider influence the decision about establishing an SLA?*

Different aspects of joint service provisioning via the Internet were due to research in the recent time. From a technical perspective, the description of services (Mika et al. 2004; Oberle et al. 2009; Speiser et al. 2008) as well as the (semantic) discovery of services (Küster et al. 2007; Toch et al. 2007) have been well-discussed and described in literature. Together with the emergence of (semantically annotated) service repositories (Junghans et al. 2010), the composition of (Web) services becomes feasible for service providers from a technical point of view. The decision for a particular service offer includes functional aspects, non-functional properties of service provisioning that specify aspects of the service that are not quality-related like the provision of temperature information in degrees Celsius for a weather forecast service and quality aspects of service provisioning. Methods that include the aforementioned factors have been subject to research, which focuses on the customer's preferences by means of policies (Lamparter 2007), or on the maximization of the overall welfare (Blau 2009), or on increasing the variety of alternatives of available services (Conte 2010).

However, in the existing literature, the aspects of decision making on the establishment of SLAs are not considered abstracted from single complex services, nor is the risk of violating SLAs and the resulting implications on profit taken into consideration. For this reason, Research Question 1 is addressed from a broad viewpoint on joint service provisioning without focusing on single complex services, thereby considering joint service provisioning by investigating the establishment of SLAs between single service providers. Therefore, in a first step, the trading object *service* is defined and an interdisciplinary delineation among types of services is given. The description and formalization of an SLA, which is established between the provider and the customer of a service are discussed in the next step. Constraints and characteristics, which influence service providers in the decision making process about the establishment of SLAs, are illustrated in the third step. These characteristics are split up into technical properties of the service provider and economic preferences. Two different types of service providers are distinguished, *service providers* that provide all requested functionality by themselves and a particular instantiation of a provider, a *service intermediary* that needs to procure service functionality from providers other than themselves. With the joint provisioning of services and under consideration of the different types of service providers, a networked structure of SLAs arises. This structure is denoted by agreement networks (ANs) in the remainder of this thesis. Thus, in a final step a formal description of an agreement network is given in order to model the scenario in which decisions on the establishment of SLAs are made.

The solution to Research Question 1 lays the foundation for defining the decision problems of two types of service provider in the newly arising, highly interconnected field of co-opetitive service provisioning, which is regulated by SLAs. With the establishment of SLAs, service providers and intermediaries experience uncertainty about future SLA violations. Such uncertainty translates into a risk for the decision maker, which has to be considered in the SLA establishment decision making process. There are well-established definitions and measures for risk (Markowitz 1959; Arrow 1971), which are applied in most distinctive domains (Bonini 1975; Filipova 2009; Kauffman and Sougstad 2010; Buhl and Fridgen 2011). However, there has been no approach, which applies risk measures in the context of SLA violations and which takes account of risk in the decision about SLA establishment.

This gap in current research directly leads to Research Questions 2 and 3, which are concerned with the development of a method for decision making for the establishment of SLAs from the perspective of a service provider and the special case of an intermediary. Classic measures of risk stem from the finance or insurance sectors (Markowitz 1959; Lintner 1965; Sharpe 1964; Jorion 2007) and illustrate uncertainty about expected profit, which plays an important role in service provisioning via the Internet (Berger 2005). In order exploit the Long Tail phenomenon for the maximization of the market share and with it profit, service providers and intermediaries exploit the possibilities of offering and providing services via the Internet. However, with a growing number of requests for service provisioning, the number of SLAs increases and with it, the complexity of the decision about SLA establishment.

Besides the sheer number of requests for service provisioning, there are particular factors that a service provider needs to consider in their decision on SLA establishment. For instance power outages, breaking hardware, human failure could impact the performance of a service in a way that leads to a violation of the corresponding SLA. Each of these causes for SLA violation is not under direct control of the service provider but lead to a loss of profit. So, even if a provider knows about this risk of SLA violation they cannot directly influence it.

Research Question 2 ≺Provider's risk-minimizing Decision≻. *How can a provider select the SLAs that minimize the risk of SLA violation while taking their technical properties and economic preferences into account?*

This question will be addressed by reviewing the technical properties and economic preferences of a service provider that were identified in context with Research Question 1. The decision problem about the establishment of a set of SLAs is determined and constraints, which influence the decision, are formulated and included in the decision problem. In order to include the risk of SLA violation in the service provider's decision problem, a suitable measure of risk is identified. In analogy to the selection of an efficient portfolio of shares, the theory of porfolio selection (Markowitz 1959; Markowitz 1991) is reviewed and adapted to the case of the selection of an SLA portfolio. This measure of risk allows the provider to select the portfolio of SLAs which bears the lowest possible risk of violation while achieving a given profit. In this context, past observations of SLA violations serve as a proxy for future SLA violations and thus enable the expression of the risk of SLA violation.

In a second step, the decision problem is extended in order to explicitly include the provider's attitude towards risk in the sense of uncertainy of profit. This is achieved by applying the theory of risk-aversion (Arrow 1971; Pratt 1964), where the decision maker's preferences disclose the amount of money which the decision maker requires for an additional marginal unit of uncertainty of profit. Applying the theory of risk-aversion to the decision problem enables the provider to express their degree of risk-aversion directly and to consider a trade-off between profit and uncertainty of SLA violation in their decision. As before, the uncertainty of future SLA violations is approximated with the help of past observations of SLA violations.

As not all service providers can provide all the requested functionality by themselves, service intermediaries were identified to play an important role in joint service provisioning. Service intermediaries offer services to customers that are complex and value-enhanced and comprise of functionality that is provided by service providers other than the intermediary. Thus, the decision on the establishment of SLAs that an intermediary faces concerns SLAs with customers as well as providers.

Research Question 3 ≺Intermediary's risk-minimizing Decision≻. *How can an intermediary select the SLAs with customers and supplying providers that minimize the risk of SLA violation while taking their technical properties and economic preferences into account?*

The intermediary's decision problem is approached similarly to the provider's decision problem. In a first step, the technical properties and economic preferences that were identified in conjunction with Research Question 1 are reviewed and the impact of the procurement of service functionality from other providers is highlighted. Based on the results, the decision problem is specified. Analogously to the risk-minimizing decision of the provider, the measure of SLA violation risk is adapted from the theory of portfolio selection (Markowitz 1991). In order to account for the risk of SLA violation by supplying providers, an extension of the model that was presented in the context of Research Question 2 is provided.

In order to allow the intermediary to express their preferences by means of a trade-off between expected profit and risk of SLA violation, the decision problem is extended in a second step. Similar to the case of the provider's decision about the establishment of a portfolio of SLAs, the theory of risk-aversion is applied to the intermediary's selection of SLAs with customers and supplying providers.

Yet, the specification of methods that provide solutions to the decision problems of service providers and intermediaries does not suffice for completing this work. In order to show the applicability in real-world agreement networks, it is crucial to evaluate the presented methods. As the risk of SLA violations is calculated from past monitoring observations, the most intriguing question for providers as well as intermediaries concerns the number of observations needed for an *optimal* decision. This leads to Research Question 4:

RESEARCH QUESTION 4 ≺REQUIRED AMOUNT OF MONITORING DATA FOR DECISION MAKING≻. *How many monitoring observations are required for the calculation of SLA violation risk in a) a provider's and b) an intermediary's decision on SLA establishment?*

The monitoring of services has been well discussed and described in theory (Baresi et al. 2006; Ghezzi and Guinea 2007) and practice (Barth 2008; Badger 2008) with the aim of achieving performance traceability. With the help of a monitoring system, it is possible to determine if an SLA was adhered to or violated in the case that its state is determinable. Consequently, service providers are able to discover problems in service execution and solve them. Furthermore, a monitoring system can assist a customer to claim penalties. On top of monitoring systems, business analytics systems are employed to generate reports on the observed adherence or violation to the terms that are stipulated in SLAs. These reports contain information about violations in different descriptions that range from textual descriptions to boolean values that state if a term was adhered to or not. In order to express future risk of SLA violation, past observations of SLA violations are required to be numerical. This requirement is not met by current business analytics systems and thus, no real-world data is available for the evaluation of the method for solving the risk-minimizing decision problem.

Consequently, Research Question 4 is addressed by a simulation-based evaluation, in which monitoring observations are created artificially from known probability distributions. The knowledge of distributions allows the application of the method for solving the provider's decision problem to the most precise specification of the distribution of SLA violations and thus, to identify the risk-minimizing choice. By comparing this choice to the choices that result from increasing numbers of observations facilitates the identification of the number of observations that the model requires for making the actually risk-minimizing choice.

1.2 Structure

The research outline that was presented in the previous section reflects the structure of this thesis, which is divided into four parts. Part I lays the foundations for the decision problems, which are introduced in Part II. The evaluation of the methods that solve the decision problems is described in Part III. Part IV concludes the thesis and highlights future research directions.

A high-level illustration of this work's structure is shown in Figure 1.1. Chapter 1 gave an insight into joint service provisioning and the agreements, SLAs, that are established in order to regulate cooperation among providers and intermediaries. From the identified research gaps, research questions were deduced. Chapter 2 lays the foundations for the following sections by giving definitions of services and service level agreements. The joint provision of services fosters the rise of a networked structure of SLAs. This structure is defined and formalized in order to define the scenario in which decisions on the establishment of SLAs are made. Additionally, the difference between service providers and intermediaries is highlighted. Chapter 3 highlights the economic foundations, which are the basis for the decision problems in the main part of this work. Concepts of decision theory (Section 3.1), implementations of decisions under uncertainty (Section 3.2) by measuring risk (Sections 3.2.1 and 3.2.2) and utility theory (Section 3.2.3) are discussed in order to lay the groundwork for the following sections. Current literature on joint service provisioning and decision making about SLAs is discussed in Section 3.3 in order to highlight the challenges that are addressed

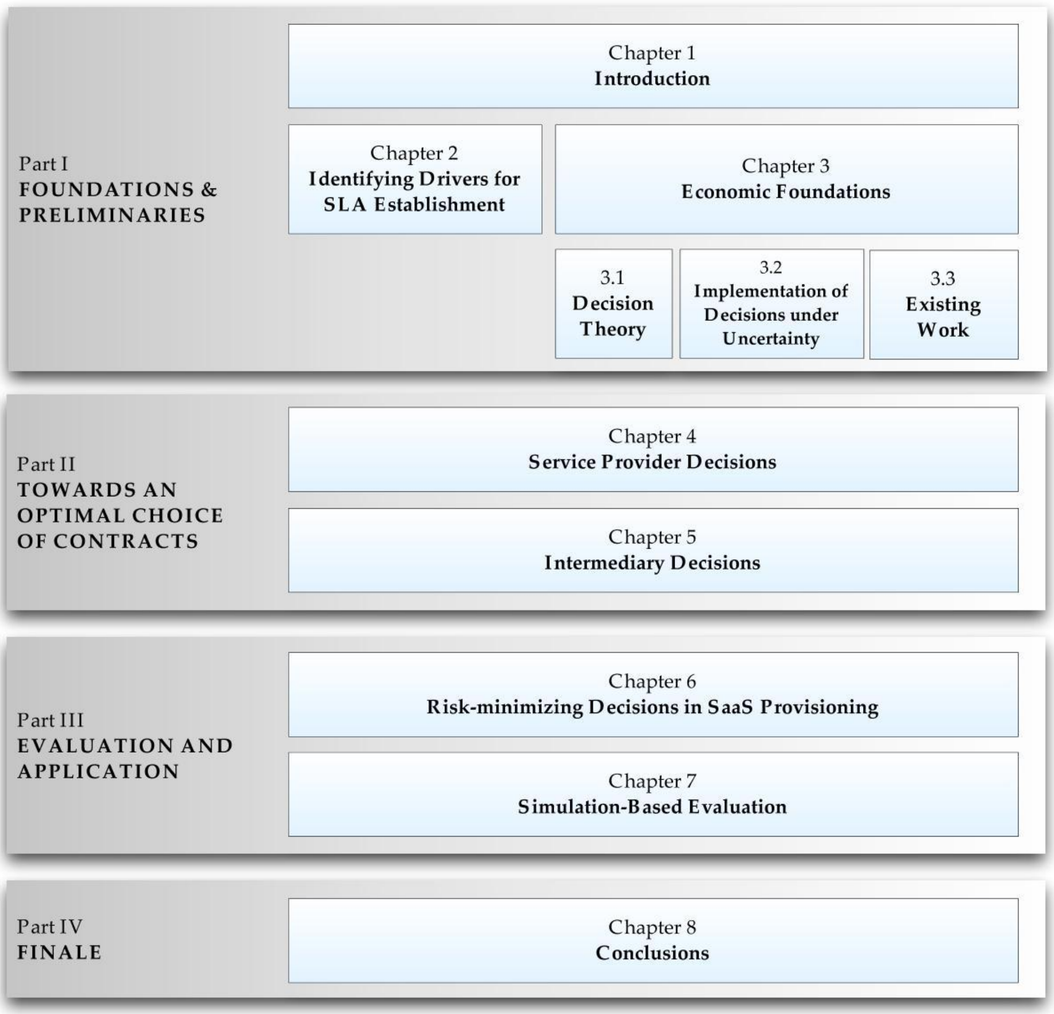

FIGURE 1.1: *Structure of this Thesis*

in this thesis.

These foundations lead in to Section 4, which focuses on the decision problem of a provider on the establishment of SLAs. The foundations from Part I are merged and a method for solving the decision problem is derived that copes with technical and economic constraints of a provider's decision. In a second part, the model is extended in order to account directly for the provider's degree of risk-aversion and to allow for a decision that is based on a trade-off between profit and risk.

In Chapter 5, the viewpoint of an intermediary is taken, which extends the provider's decision problem. In contrast to a provider's decision problem, which is concerned with the selection of a portfolio of SLAs with customers, an intermediary needs to decide about the portfolio of SLAs, which will be established with supplying providers, additionally. Analogously to Chapter 4, a risk-based decision model is defined, which is extended in a second step in order to allow for a trade-off between profit and risk.

Chapter 6 emphasizes the applicability of the introduced models by illustrating the implementation in the project ValueGrids in the area of software as a service (SaaS) composition. In Chapter 7 the introduced approaches are evaluated by means of a simulation-based evaluation. Chapter 8 summarizes the key contributions of this work and points to future research, limitations, and complementary topics.

1.3 Research Development

Excerpts of this work have been accepted for publication and presentation at european and international conferences. Additionally, parts have been published as journal articles. This section serves to give an overview which parts have been published in which research community. Furthermore, this section highlights the development of the work at hand by means of focusing on the steps of refinement and extension.

The basic principals of the provider's decision process that takes the risk of SLA violation into account was presented at the 43^{rd} Hawaii International Conference on System Sciences (HICSS '43) (Michalk et al. 2010). This work was adapted due to applicability reasons in order to account for service monitoring rather than resource monitoring and extended by an objective function, which allows for a trade-off between risk and profit. This revised and extended version was published in the Journal of Service Science (Michalk et al. 2011) and corresponds to the work presented in Chapter 4.

As an extension to the provider's decision process, a first version of the intermediary's decision problem that is based on the risk of SLA violation was published in the Journal of Information Systems and e-Business Management (ISeBM) (Michalk and Blau 2010). This publication serves as groundwork for the decision process that is presented in Section 5.1.

First evaluation results about the required number of monitoring observations were been presented at the European Conference on Information Systems (ECIS 2011) (Michalk 2011), along with the latest version of the provider's risk-based decision problem. This work corresponds to the results in Section 7.2.

In addition, between 2009 and 2011, research in the context of service level management and risk-based decisions was contributed to the ValueGrids[6] project funded by the German Federal Ministry of Research and Education. In the context of this project, the overall project goal of a holistic SLA management along service value chains and across different levels of service composition was published at the Cracow Grid Workshop (CGW 2009) (Schulz et al. 2010). The design and conceptualization of the ValueGrids component is described in Section 6. It facilitates risk-based decisions in the context of service composition, was presented at the annual meeting of the *Gesellschaft für Informatik* (Informatik 2010) (Michalk and Caton 2010).

[6] `http://www.valuegrids.de/` last accessed: October 8, 2011

Chapter 2

Identifying Drivers for Service Level Agreement Establishment

"In a distributed service-oriented computing environment, service consumers like to obtain guarantees related to services they use, often related to quality of a service."

(Andrieux, Czajkowski, and Keahey 2007)

THE goal of this chapter is to illustrate the concept of service level agreements (SLAs) and the factors that drive their establishment from a decision maker's point of view. Furthermore, agreement networks (ANs) are introduced, which are the application scenario of this thesis and form the environment for the decision problems that are the main concern of this work. As preparational steps, first the concept of a *service* is introduced in Section 2.1 in order to specify the trading object under consideration. Therefore, an extensive literature review gives insight into definitions that are employed in different research disciplines. This enables a definition of *service* to be identified, which captures the information systems (IS) perspective, yet is rich enough to also account for the increasing IT (Information Technology) orientation in the service sector. It is set into the context of Cloud Computing.

In Section 2.2, the emergence of, and need for, SLAs in business contexts is discussed. The building blocks of SLAs are introduced and a formal description is given that serves as a basis for the decision problems that will be introduced in Sections 4 and 5. In the context of joint service provisioning that is regulated by means of SLAs, two types of providers are distinguished and the drivers that influence their decision on SLA establishment are identified. The network of providers and consumers of services that arises with the establishment of SLAs is discussed and set into context with related concepts like business networks and service value networks in Section 2.3. Besides a formal description of ANs and the definition of roles that exist in ANs in Section 2.3, a hands-on example is given to illustrate the scenario in Section 2.3.4.

This Section is dedicated to address Research Question 1 from Section 1.1:

RESEARCH QUESTION 1 ≺STIPULATION OF SERVICE LEVEL AGREEMENTS≻. *What aspects of service provisioning do a provider and a customer of a service stipulate in a Service Level Agreement and which technical properties and economic preferences of a service provider influence the decision about establishing an SLA?*

2.1 Service Concepts and Definitions

In accordance with Vargo and Lusch (2004b), changes in society and markets yield the major shift towards a service-centered view that leads to an exchange of services rather than goods (see Section 1). This manifests across industries, where traditional *manufacturers* tend to integrate services in their core offerings. This service-orientation is often motivated by a change in customer demand towards a higher degree of customization. The *servicification* in the software industry is leveraged by ubiquitous advanced Web service technologies and leads to a fundamental change in company strategies and business models: software vendors become service providers (Dubey and Wagle 2007) (see Section 1). The increasing momentum of on-demand applications that is driven by the availability of the Internet and technologies that facilitate service provisioning via the Internet is underpinned by a series of Gartner studies (Mertz et al. 2007; Mertz et al. 2008).

The goal of Section 2.1 is to review service concepts from disciplines that are concerned with decision making about service provisioning and the impact of technological changes on service provisioning. Therefore, service concepts from business economics, economics, both for decision making, and computer science, for the technological impact, are explored in order to identify a definition that is suitable for research on the intersection of these disciplines in the discipline of IS. This lays the groundwork for the following sections by defining services in general and delineating electronic services and Web services as well as explaining the connection to the *as-a-service* paradigm highlighting the increasing impact of the Internet and related technologies for service provisioning. This definition is essential to properly describe SLAs and the building parts that influence the decision about their establishment.

2.1.1 Related Work

To date, a magnitude of different definitions of a service exist. This is because each discipline concerned with a particular aspect of a service determined a definition that best suited its needs. In the area of computer science a service is defined based on the requirements that arise in the context of practical implementations with a focus on technical properties of a service. Contrarily, business economists consider the general properties of a service, the prerequisites for service provisioning and the value creation. To understand decisions in service-oriented economies, it is not sufficient to focus on just one discipline. A definition is required that allows for a systemic view on services that includes technical as well as business and economic properties.

The following sections introduce different types of definitions and give an overview on business-related and technical service definitions.

Business-related Service Definitions

The number of contributions in the context of *service* is enormous and in equal measure heterogeneous. There is no common definition of a service. The large variety of existing definitions reflects the application of different approaches and even completely different views on ecosystems.

Most business economic definitions of service have one element in common: a focus on value creation. According to Engelhardt et al. (1993), there are three perspectives on value creation: *potential-*, *process-*, and *outcome*-orientation. Potential-orientation is concerned with the *preparation* of service provisioning, that is, the allocation of production factors. Since the allocation of factors is necessary in almost any value creation activity, a potential-oriented definition of service is not sufficiently distinctive for the definition of a service. The activation and integration of allocated resources is considered in process-orientation, where services are described by *activities* that employ resources through either the customer and the provider. Outcome-orientation denotes a view that directly relates to the *result* of this process. Outcome-orientation concentrates on the nature of the outcome being material or immaterial and hence, cannot describe the nature of service adequately. It is rather the delivery phase, that is, in the words of Engelhardt et al. (1993), the process-oriented definitions, that define the heart of a service by adequately and equally awarding importance to both, the service provider and the service consumer. Consequently, the process of providing a service and the corresponding activities are considered in detail in the following.

The *preparation* of a service as a first step, followed by its *delivery* need to be considered seperately (Engelhardt et al. 1993). Provider and consumer set up readiness for service provisioning in the first step. In the case of a hair cutting service, such preparation would include, for instance, the education or hiring of staff, procurement of equipment, leasing of a premise, and so forth. For a Web-based service, the preparation phase would, inter alia, include programming efforts and the allocation of sufficient computational resources.

After the general preparation as described above, additional individual preparation is required in order to provide the tailored service that meets the customer's demand. Before commencing the service provisioning, an agreement is stipulated between customer and provider that regulates the terms of service provisioning. Provisioning and simulataneous consumption of the service are included in the delivery phase. While the preparation phase is universal, i.e. represents the basis for every concrete service delivery, the outcome relates to a distinct delivery process. Figure 2.1 illustrates the phases of service provisioning.

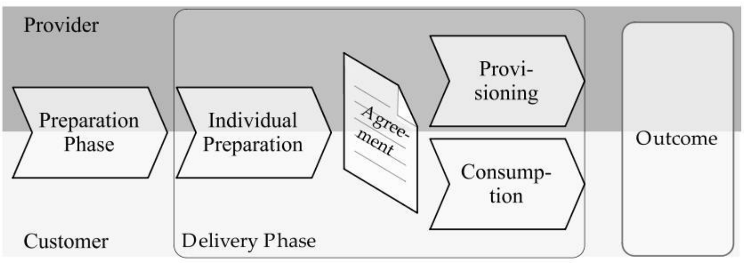

FIGURE 2.1: *The Interrelation of Service Preparation, Provisioning, and Consumption*

Besides taking the phases of service preparation and delivery into account, definitions of service in the literature focus on characteristics of service as compared to goods. These are **i**ntangibility, **h**eterogeneity, **i**nseparability of production and consumption and **p**erishability (IHIP criteria) (Edgett and Parkinson 1993; Rathmell 1966; Regan 1963; Shostack 1977; Zeithaml et al. 1985).

Flipo (1988) as well as Kotler and Connor Jr (1977) describe a service as *intangible*, that is, it cannot be perceived before it is bought. This property is perceived as the major delineating factor between products and services by a majority of scholars (Shostack 1977). The large variety in provided services that results from an ever changing demand from customers is described by *heterogeneity* (Zeithaml et al. 1985). The uno actu principle states the simulatenous consumption and delivery of services is captured in the *inseperability* property (Regan 1963; Wyckham et al. 1975) (cp. also last part of the delivery phase in Fig. 2.1). Finally, the fact that services cannot be stored or transported is described by *perishability* (Donnelly Jr 1976; Rathmell 1966; Zeithaml et al. 1985). In relation with the IHIP criteria, Lovelock et al. (1999) defines services as an act or performance offered by one party to another. In summary, the definition of a service that is provided is rather vague and the identified criteria, the IHIP criteria, and their applicability have been subject to discussion (e.g. Edvardsson et al. (2005), Lovelock and Gummesson (2004), Vargo and Lusch (2004a)). While recognizing that there are numerous services which take some form of tangible representation (e.g. car repair, programming, taking photos, etc.) services cannot be defined to be completely intangible. Analogously, there are products which are highly adapted to customers' requests (e.g. Dell notebooks) and hence cannot be differentiated from goods by their heterogeneity.

Based on many possible exceptions that arise from the IHIP criteria, continuum-based approaches like the product-service-continuum gained momentum (Shostack 1977). In these approaches of which most stem from marketing research, dimensions are defined which characterize services. For example services are characterized along their "degree of intangibility" (outcome-orientation) and "degree of customer integration" (Engelhardt et al. 1993), "degree of individuality". This enables a distinction between individualized and standardized services (Meffert and Bruhn 2008), or simply a differentiation between tangible- and intangible-dominant entities (Shostack 1977). A major implication of these approaches is the neglection of a dichotomy between products and services. Specifying a service along different shades of dimensions does not allow for a clear-cut separation between services and products. This results in the emergence of a continuum, where pure products resemble one end and pure services the other (Berry and Pasuraman 2004; Chase 1981; Shostack 1977).

Defining a service by identifying features that enable the distinction from a product does not prevent an ambiguous solution, as discussed above. Even the application of continuum-based approaches does not allow for a unique definition of what a service is. A different view on services is taken in Vargo and Lusch (2004b). In context with the *service dominant logic* (Vargo and Lusch 2004b), a good can be seen as an appliance, or medium, for service delivery. Vargo and Lusch (2004b) postulate a philosophy rather than a service definition, which states that there is a major shift towards a service-centered view that is driven by changes in society and markets that leads to the exchange of services rather than goods. According to Vargo and Lusch (2004b), a service is "*the application of specialized competencies (knowledge and skills) through deeds, processes, and performances for the benefit of another entity or the entity itself*". This paradigm is underpinned by a service prosumer that co-creates value with a service provider or service producer by "applying specialized skills and knowledge". This specification of a service subsumes a large variety of concepts.

A definition of a service that is more specific was introduced by Hill (1977). In this notation, a service denotes an activity that is performed by an economic unit B for an economic unit A, where the result of this activity is the change in condition of an economic unit C that either is or belongs to economic unit A. Note that the prior agreement of economic unit A is required. Hill (1977) defines that it is not the readiness for service provisioning, but rather the "production of a service" that constitutes value creation, where "production" refers to the aforementioned activity. With respect to Figure 2.1, the definition provided in Hill (1977) is concerned with the delivery phase.

The formalization of a service according to Hill (1977) specifies standard services like hair-cutting, car repairs, or post shipment very accurately. However, the definition does not allow for a distinction of services performed on electronic data or provided via electronic networks which is required for an interdisciplinary understanding of services that allows for the impact that the Internet and related technologies have on provisioning. Gadrey (2000) formulated several extensions to the definition provided by Hill (1977). Three necessary extensions are defined in order to include assistance and intervention, provision of technical capacities, and live performances in Hill's definition. However, the definitions of Hill (1977) and Gadrey (2000) do not include the concept of value co-creation by consumer and provider.

Hence, currently there is no approach that enables an unambiguous definition of service for IS research and that captures the impact that the Internet and related technologies have on service provisioning.

Information and Communications Technology View on Services

Value creation has been identified as one of the major properties of service provisioning from a business economic viewpoint. Nevertheless, the definition of a service that is provided by ITIL (IT Infrastructure Library) includes this aspect of a service by defining it as "*a means of delivering value to customers by facilitating outcomes customers want to achieve without the ownership of specific costs and risks*" (Bon 2005). Thus, ITIL offers a high-level definition of service that abstracts from potential- and process-orientation but focuses on the outcome, i.e. value creation. Furthermore, the cost and risk of a service are considered to be essential characteristics.

With the advent of service-oriented computing (SOC), Web services have been established as technical services that do not only affect persons or goods, but may also affect other electronic resources such as addressable data sets. From a technology viewpoint, Web services cater for an abstraction layer over different network protocols, operating systems and programming languages. Thus, Web services provide possibilities to expose the functionality of an application system by means of Web technologies (Alonso et al. 2004).

Similar to the field of business economics, there is a variety of contributions and apart from the very vague definition above, there is no commonly applied definition of services in computer science.

A Web service, however, is specified as "*a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP-messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards*" (W3C 2004). In Papazoglou (2008) two kinds of topological Web service are distinguished: informational and complex Web services. While the former simply supports request/response patterns, the latter implements coordination functionality.

The description of a service by means of WSDL or SOAP-messages is a very common approach, which allows for the specification of interfaces and communication protocols. Nevertheless, it lacks the ability to describe service functionality. In order to include common wording that induces a broad understanding, semantic annotations can be, and typically are, employed.

One approach that allows the specification of a service in an understandable way is provided in O'Sullivan et al. (2002) by listing non-functional properties of a service that form constraints on the service functionality. A service is defined along the following characteristics: *activities performed by one entity for another*, *inherent value that is transfered from provider to customer*, *service composition or aggregation*.

The definitions sketched above are tailored to computer scientists' needs, but do not include important economic aspects of a service like participants in service provisioning, the change of state, and in most parts: value creation.

System View on Services

Both, Business and ICT (Information and Communication Technology) views on services suggest definitions of *service* that meet the respective discipline's requirements. A definition that unites the diverse approaches and suggests a systemic view that considers (business) economic and computer science related aspects of service provisioning is missing.

A first approach to define a view on a service world as a system is given in Spohrer et al. (2008) and Maglio and Spohrer (2008). In their definition, service systems revolve as "*dynamic value co-creation configuration of resources, including people, organizations, shared information, and technology, all connected internally and externally to other service systems by value propositions*". This view provides insights on concerned parties, resources and the aim of service provisioning and consequently highlights the environment of service provisioning rather than a definition of the process of service provisioning.

Even though the definitions provided by Vargo and Lusch (2004b), Spohrer et al. (2008), Maglio and Spohrer (2008) are very vague and do not provide a means to describe a service unambiguously and consequently are not suitable for application in IS research, they provide a basis for further steps.

Rather than providing a service definition, Baida et al. (2004) presents an empirical analysis of the application of the term service. In this work, three major terms were identified: services, e-services and Web services. All of these terms are applied in all of the examined disciplines with differing meanings. Therefore, an interpretation of each of the terms is suggested. Services are defined in accordance with business science, as business activities which create value, mostly in intangible forms. E-services are interpreted as services that are delivered over the Internet, accoring to information science. Computer science regards Web services as software with certain technical properties that realizes e-services.

In the following Section, a definition of service is given that represents the essence of the discussed service definitions while coping with the requirements in IS research. Furthermore, a distinction of service types along criteria is suggested that allows for the impact of the Internet and related technologies.

2.1.2 Service Definition

This section serves to give a definition of *service* that fuses the technical service definitions, which focus on exchanged messages, protocols, and interfaces and business economic specifications with a focus on value creation. The given definition is designed to be applicable in interdisciplinary research like IS research. First, a generic service definition is given in accordance with Blau (2009) and Conte (2010). This definition is then specialized analogously to Baida et al. (2004) in order to account for electronic and Web services. Based on the service definition by Hill (1977) and its extension by Gadrey (2000), a service is defined as follows.

Definition 2.1 (Service)**.** *A service is a set of activities that are performed and intended to bring about a change of state to either an entity that is owned or used by a consumer or to the consumer itself. The set of activities is performed by a provider or jointly by the provider and consumer. The outcome or resulting change of state is based upon a prior agreement between the consumer and provider, which aims at the co-creation of value.*

In contrast to *service* as defined in Vargo and Lusch (2004b), the above definition emphasizes the intention to change the state of an entity, e.g cut hair or increase the degree of information of students. Furthermore, it requires a prior agreement on the provision of the service in order to exclude unrequested provisioning. By co-creation of value, contribution of both parties, consumer and provider, is understood. The consumer's contribution may range from the provisioning of the entity that is changed through contributing to the execution of the set of activities. The provider mainly contributes by performing the activities.

Building on this basic definition of *service*, properties are identified and specific service types are derived. With the rise of information and communication technology and rapid growth of the Internet, the environment of service provision changed completely. In this context, a special kind of *service* emerged that is defined as *Electronic Service*.

Definition 2.2 (Electronic Service)**.** *An electronic service (e-Service) is a service where the input and outcome are provided via an electronic network.*

In more detail, a service can be characterized as an electronic service, in the case that the input provided to the service and/or the outcome of the service are distributed by means of an electronic network like the telecommunication network or the Internet.

The ongoing paradigm shift in service provision is constituted most evidently by the growing importance of providers like Amazon Web Services[1] and Salesforce[2]. There, offers comprise services that are interoperable and hence may be composed to meet consumers' special demand. This specific kind of e-Service is denoted as a Web service.

Definition 2.3 (Web Service)**.** *A Web service is an e-service identified by a URI (Uniform Resource Identifier) that exposes a public, well-defined interface. The input and output are provided via a Web protocol.*

There are two attributes that allow to specify a service as a Web service. First, the existence of a URI that discloses a public, well-defined interface to the service is required. Second, a Web service involves the communication by means of a special Web protocol (such as HTTP). Both of these aspects can be regarded as key enablers of automatic service composition, and hence, cooperation among service providers, which is the driving factor for the application scenario of this thesis.

The criteria for discriminating the service types are presented in Table 2.1. From this disambiguation, the typology as depicted in Figure 2.2 is derived.

Real-world examples for Web services include i. e. Google's search service, which exposes a public interface, and can be composed with other services. But additionally, services like salesforce.com's Sales Cloud or Amazon Web services are comprised that include for instance customer relationship management (CRM) systems, databases, computing instances or storage space. All of these services are hosted on the provider's resources and can be requested and used on-demand without the need for the customer to procure infrastructure or install monolithic software systems. The provisioning of these services is influenced by a technology that emerged in the context of service provisioning via the Internet and gained momentum in recent years. With the help of virtualization techniques it is possible to provide a magnitude of types of Cloud services (Baun et al. 2009) comprising infrastructure services, e. g. Amazon's Elastic Compute Cloud (EC2), platform services like salesforce.com's Sales Cloud, and software as a service like Google's AppExchange. Therefore, Cloud services denote a further type of service that is influenced by the availability of the Internet and related technologies.

[1] `http://aws.amazon.com` last accessed: October 8,2011
[2] `http://www.salesforce.com/` last accessed: October 8, 2011

TABLE 2.1: *Criteria of Service Categories*

Attribute	**Service**	**e-Service**	**Web service**
Input/Output/ Outcome Transmission	electronically/ personally	electronically	electronically
Interface	(not) well defined	(not) well defined	well defined
Communication Protocol	any	any	Web Protocol

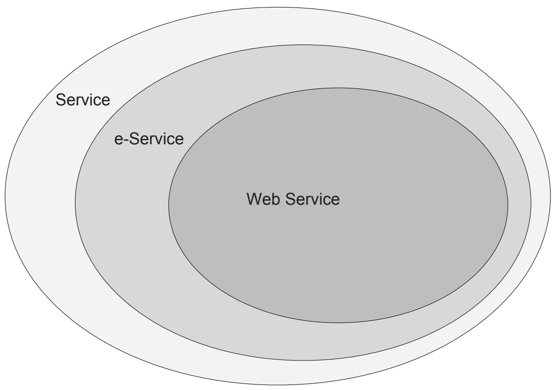

FIGURE 2.2: *Service Categories and their Interrelation*

2.1.3 Conclusion

This section served to give a comprehensive definition of a service that is relevant and fitting for the context of this thesis. Therefore, literature was reviewed that defines services from different research disciplines by applying different approaches and even different philosophies. In accordance with Blau (2009) and Conte (2010), a service definition that can be applied in information systems research was presented. Properties were identified that allow the differentiation of e-services and Web services.

These properties can be found in Table 2.1. An electronic service is distinguished from a service by the transmission of input and outcome. Analogously, a Web service is an e-service with a well-defined interface and to which communication is performed via a Web protocol. In a final step, Cloud services were presented as a special kind of services that are provided by means of virtualization techniques.

Thus, this section served to establish a common understanding of the trading object *service* and to provide a definition that considers aspects for (business) economic decision making as well as technological properties of service provisioning. This definition denotes the starting point for further investigations on joint service provisioning of different kinds of service. The agreements that are stipulated between the provider and a consumer of a service and that serve to regulate service provisioning and consumption are explored in the following section.

2.2 Governing Service Provision and Consumption

The joint provisioning of services is facilitated by a multitude of service providers (Blau et al. 2009). Similar to contracts of purchase, which rule properties and condition of the procured good, a concept of agreements exists that regulates the provisioning and consumption of a service. Especially in business-related scenarios, SLAs gain increasing momentum. SLAs define functional as well as non-functional aspects of service provision, along with duties and rights of the concerned parties. This section serves to give an introduction to the building parts and representation of SLAs.

In a first step, literature on SLAs is reviewed and discussed. A definition of what an SLA is and what it contains is introduced and special charachteristics of SLAs for Web services are identified. A representation of SLAs is given that illustrates the building parts of an SLA which are relevant to the decision problems in Chapters 4 and 5.

2.2.1 Related Literature

In recent years, scholars paid much attention to SLAs. This results in an enormous body of definitions of what the goal of an SLA is and what it should comprise. Most of the existing definitions are based on a similar essence but exhibit slightly different aspects. Nevertheless, there is no commonly agreed definition of what an SLA represents. This section serves to give an overview on a share of SLA definitions that emerged in the area of IT-related services.

According to Berger (2005), an SLA is an agreement that stipulates levels and standards of a service. This statement does directly reflect the components of an SLA: the *service*, its *level* and the *agreement* on them. Hence, this definition states the core essence of what an SLA comprises.

The concerned parties of an SLA are highlighted in the definition provided by Hiles (2002), where an SLA is defined as "*an agreement between the computing service provider and the user*" which quantifies the minimum acceptable service to the user. Hence, besides the definition by Berger (2005), the aspect of particpating actors is added in Hiles (2002).

The aim of an SLA is considered to materialize in incentives that are set to providers as well as customers. While providers are encouraged to adhere to the agreed acceptable level(s) of service, that is, quality of service (QoS) by a monetary reward and a penalty in case of SLA violation that are stipulated in the SLA, consumers can reduce the uncertainty of the outcome of service provisioning by the liabilities that are defined in the SLA (Brandic et al. 2008; Lewis and Ray 1999). This way, providers and customers are incentivized to adhere to their duties. Having identified expectations about service provisioning and by specifying actions that need to be taken in specific circumstances, i. e. service failure, provider and customer of a service have a clear means of communication and action (Karten 2001; Lewis and Ray 1999). In order to reach this goal, an SLA is required to comprise the business partners, a desription of provided services, their parameters and acceptable service levels in addition to liabilities (Lewis and Ray 1999).

With the advent of Web service technologies as already stated in Chapter 1 and Section 2.1, SLA definitions with a special focus on Web services are required. Besides being able to cope with electronic interactions, which are inherent to Web services, and hence, being able to bear descriptions of Web services, SLAs need to reflect QoS attributes of Web services (Keller and Ludwig 2003). A special challenge in this respect is that the agreement still needs to be understood by the concerned parties, that is, customers and providers, without requiring specific knowledge about the provider (Czajkowski et al. 2002). Consequently, an SLA for Web services needs to specify one or more service level objectives (SLOs) in order to identify requirements on the service quality of the customer and to stipulate guarantees by the provider (Andrieux et al. 2007).

These goals do not differ significantly from the aim of general SLAs. However, it is inherently important that SLAs for Web services take a form that is comprehensible but at the same time machine-readible. Furthermore, SLAs need to reflect the customer's requirements and the provider's guarantees towards service quality and to allow for a specificity that is measurable and can be monitored. SLAs for Web services are required to follow a specification that allows for (semi-)automatic instantiation in order to cope with the requirements of Web service composition. In order to consider these requirements, the following section provides a general definition of an SLA and derives a special kind of SLA that is suitable for Web services. Finally, a formal notation of SLAs that suits as a basis for the decision problems in Sections 4 and 5 is given.

2.2.2 Service Level Agreements

The goal of this section is to provide a definition of what an SLA is by means of specifying its purpose and its required constituents. Therefore, existing definitions are merged and their essence is presented. In a second step, a definition of an SLA that is particularly suitable for Web services is given as an extension of the general definition.

Based on the definitions of Lewis and Ray (1999), Hiles (2002), Brandic et al. (2008), an SLA is defined as follows:

Definition 2.4 (Service Level Agreement)**.** *A service level agreement is an agreement between a provider and a customer of a service that identifies the business partners, the functionality of the provided service, the non-functional and quality parameters for the service and corresponding service levels, as well as a monetary reward for service provisioning and liabilities when service levels are not met.*

This definition does not only include the parties that stipulate the agreement and the service that is provided. Besides the functionaliy that is provided, non-functional aspects of the service that do not relate to quality are defined as well as requirements that a customer has towards the provided service and the guarantees that the provider agrees to. These guarantees are specified by means of assertions to service attributes.

In other words, in the agreement, target values for quality paramaters, so called key performance indicators (KPIs), are determined in order to establish QoS objectives, service level objectives (SLOs). Additionally, the monetary incentives that are stipulated in order to foster the adherence to the agreement are considered by means of a monetary reward, that is the price for service provisioning and liabilities for the case of SLA violation that are commonly called penalties.

Generally, penalties enforce the adherence to SLAs by means of a credible threat (Becker et al. 2008). In the case of the violation of an SLA, a penalty is applied by means of a monetary amount on behalf of the provider or a credit for the customer. One example of a penalty is a direct monetary payment from the provider to the customer or the reduction of the agreed price to pay for service execution. Another example for a penalty is exerted by Amazon and Google Apps Engine, where the customer is credited free usage time or service credits if the service is not provided as agreed upon. Other examples are implicite impacts on future agreements or the enforced re-execution of the service after failure (Rana et al. 2008). However, no matter which approach might be applied, it is possible to translate the penalty to a monetary amount. Therefore, in the remainder of this work, a monetary amount will denote the penalty.

The rise of information and communication technology and especially the increasing application of Web service technologies require matching SLAs that consider special characteristics of Web services. Based on the definition of an SLA that is provided above, a Web service agreement is defined as follows:

Definition 2.5 (Web service Agreement)**.** *A Web service Agreement is an SLA, in which the provisioning of a Web service is agreed. Web service Agreements expose a machine readible representation that does allow for human understanding.*

Following this definition, an SLA is of the special type Web service Agreement if the service about which the agreement is stipulated is a Web service and can be described by description languages like WSDL (Christensen et al. 2001), JSON (Crockford 2006) or SOAP (Fielding 2000). Additionally, the SLA is required to expose a machine readible representation that can be achieved, for instance, by applying SNAP (Czajkowski et al. 2002), WSLA (Keller and Ludwig 2003) or WS Agreement (Andrieux et al. 2007). In this work, the SLA specification is chosen according to WS Agreement due to its degree of distribution and its status as a recommendation of OGF[3].

According to WS Agreement, an SLA basically contains three parts: its identifier, the

[3] `http://www.gridforum.org/` last accessed: October 8, 2011

context and the terms. Where the context comprises of information about the agreement initiator, that is, who requested to establish the agreement, the agreement responder, and which role the service provider takes. Additionally, a validity period for the SLA is given along with information about the template on which the SLA was built. In conclusion, the context contains information about the concerned parties, provider and customer, as well as the validity period and emergence of the SLA.

The last big building part of an SLA, the *terms*, consists of service terms and guarantee terms. Service terms contain either a service description or a reference to a service endpoint, that is, a URL where the Web service can be contacted. In the guarantee terms, the assurance on service quality is specified in terms of SLOs, that is, KPIs and target values, along with a reward, the price for service provisioning and a penalty in case of SLA violation.

In order to provide a formal representation of SLAs that can be used in the following chapters, an SLA can be summarized to contain an identifier for provider, customer and service as well as a set of SLOs, a price for service provisioning and a penalty in case of SLA violation. Consequently, an SLA α can be formalized as

$$\alpha = (p, s, \theta, L_\alpha, f_\alpha(s), \mu_\alpha) \, ,$$

where p denotes the provider, θ the customer and s the service. Additionally, L_α describes the set of SLOs that is stipulated in the SLA, $f_\alpha(s)$ is the price for service provisioning and μ_α is the agreed penalty.

2.2.3 Conclusion

The aim of this section was to give a general definition of the scope and goal of an SLA as well as its constituents. Therefore, existing definitions from literature were reviewed and condensed and a general definition of an SLA was derived from Lewis and Ray (1999), Hiles (2002), Brandic et al. (2008). In a second step, the definition was specified in order to account for a particular type of service in the SLA, Web services, as defined in Section 2.1. By considering these Web services along with exhibiting a machine-readible representation, Web service agreements were defined as a special type of SLA.

SLAs of each of the above specified types form the basis of cooperation among providers and consequently, joint value creation. Especially in the context of services that are provided via the Internet, the demand for customized services increases and with it the need for cooperation among service providers. This cooperation fosters the emergence of different types of service providers and networks of agreements. This phenomenon is discussed in detail in Section 2.3.

2.3 Service Provider Types & Agreement Networks

Ever since industrialization revolutionized the creation of goods, a focus of manufacturers on their core competencies could be observed. Production was tailored to create standardized goods in order to reduce costs and producers learnt to cooperate in order to leverage economies of scale. The same phenomenon has been observed in the services sector for several years now (Prasad and Kalai Selvan 2009). Whereas the formation of business networks for the joint provisioning of services has been subject to a large variety of research, a special focus on Web service networks emerged recently (Blau et al. 2009). However, existing views on Web service networks focus on the (efficient) composition of Web services in order to create a particular complex service (Blau 2009; Conte 2010) rather than the network of agreements and cooperations that arises between providers and customers.

This section serves to establish a common understanding for Agreement Networks. Therefore, in a first step, a thorough literature review discusses existing perceptions of networks that arise with the joint provisioning of services. In a constructive approach, ANs are differentiated from existing approaches and a definition is given in Section 2.3.1. From this definition, different types of service providers are identified and defined. Furthermore, a formal specification of ANs is given in Section 2.3.3 that can be applied in the following, and the scenario under consideration is illustrated (Section 2.3.4).

2.3.1 Definition & Related Concepts

This section serves to give an overview on existing perceptions of networks that facilitate business cooperation. Besides providing a brief insight into concepts that are related to ANs, this section sketches the development of business cooperation with an increasing degree of IT application.

The most general description of cooperation among different firms or other legal entities that is motivated by economic reasons is given by the term *business networks* (Holm et al. 1996; Steiner 2004; Tapscott et al. 2000; Zerdick et al. 2000). A collective reward is regarded to be an incentive for participation in a network that yields joint value creation. Network partners in business networks can be of either homogeneous (i. e. competing) and heterogeneous (i. e. complemetary) nature that cooperate in a temporary manner in project-driven or goal-oriented partnerships (Bengtsson and Kock 2000). However, agreements in general are not considered explicitly in business networks.

With the increasing momentum of information and communication technology (ICT), business networks evolved to a new generation that is called smart business networks (SBNs). In this kind of network, cooperation is facilitated by the application of emerging technologies. The increasing effectiveness and corresponding advantage that is achieved by the use of ICT lead to appending the characteristic *smart* (Van Heck and Vervest 2009). ICT is also seen as an enabler of network agility, i.e. the network's ability to "rapidly pick, plug, and play" business processes (Heck and Vervest 2007). The trend from "mass customization" to "mass individualization" is tightly coupled with SBNs that allow for dynamic on-demand adaptation to customers' needs and requirements (Busquets et al. 2009).

The focus of business networks and SBNs lies on the cooperation aspect of joint value creation. However, SLAs and agreements in general are not considered explicitly. Besides focusing on the area of Web services, service value networks (SVNs) highlight the dynamic and on-demand composition of service modules (Blau et al. 2009). Service value networks are considered to reflect a specialization of SBNs, in which value is created by the on-demand composition of services that are provided in a steady but open pool of complementary as well as substitutive Web service modules. This perspective concentrates on a single complex service, the added value that is created by it and the efficient selection of service modules for the formation of the complex service. In a broader view, service providers offer their services in a multitude of SVNs and take part in a large variety of value creation processes. This view on joint value creation across the boundaries of a diversity of complex services is not comprised in SVNs. Therefore, a definition is required that allows for a focus on the network of agreements that arises with joint service provisioning.

Definition 2.6 (Agreement Network)**.** *Agreement networks are business networks that provide added value by facilitating the dynamic cooperation of service providers with the help of SLAs in order to foster joint service provisioning.*

In more detail, an AN is characterized by the dynamic cooperation of service providers, which is facilitated by agreements about service provisioning, SLAs, as defined in Section 2.2. Furthermore, if emerging technologies are employed for service provisioning, ANs form a specialization of SBNs. This reflects the impact of ICT in the field of service provisioning, which leads to on-demand customization of services that requires a dynamic cooperation among parties. In order to facilitate joint service provisioning, different types of service providers are required. These are defined and differentiated in the following section.

2.3.2 Service Provider Types

Cooperation among service providers is required if a customer requests particular functionality that cannot be provided by one provider alone. In order to meet the customer's request, the provider needs to cooperate with other providers. This in turn leads to the emergence of different types of service providers, which are introduced below.

First, consider a provider that is able to provide the requested functionality.

Definition 2.7 (Service Provider)**.** *A service provider offers stand-alone services along with corresponding SLAs. A service provider is characterized by their services and their resource constraints as well as an amount of profit that they want to gain.*

According to this definition, a service provider can provide all the requested functionality by themselves. A service provider is constrained in service provisioning, and consequently the establishment of SLAs, by the their available resources and economic preferences. Whenever the provider stipulates an SLA, they agree on SLOs that determine the quality of the delivered service (see Section 2.2). SLOs can be translated to the provider's resources, like available bandwidth or storage space. Furthermore, in each SLA, a price for service provisioning and a penalty for the case of SLA violation are determined (see Section 2.2), which influence the expected profit of the provider. Consequently, these are the factors that need to be considered in the provider's decision on SLA establishment.

Consider the case where a provider cannot deliver all the requested functionality by themselves. In this case, the provider needs to cooperate with other providers in order to deliver the requested service. This fosters the rise of a special type of service provider, the intermediary.

Definition 2.8 (Intermediary)**.** *An intermediary is a service provider that provides a combination or composition of services from a multitude of providers to a customer. An intermediary* *is characterized by the functionality that they can provide by themselves, the functionality that they can provide jointly with others, resource constraints and the amount of profit that the intermediary wants to achieve. Intermediaries establish SLAs with customers and other providers.*

This way, the intermediary is a special kind of service provider that fosters joint service provisioning by contracting with other providers. The intermediary's decision problem concerning the establishment of SLAs is constrained by resource and economic constraints like the provider's problem. However, the intermediary is required to concurrently consider SLAs that are established with customers as well as SLAs that are established with other providers.

2.3.3 Formalizing Agreement Networks

The creation of value that is enabled by joint service provisioning fosters the rise of agreement networks. In order to enter an SLA, the customer has to be able to discover available SLAs from providers. Therefore, a customer is assumed to operate or have access to a service and SLA repository that can be queried for specific service functionality as well as corresponding SLAs. With respect to an AN, customers $\theta \in \Theta$ denote the headnode(s) of the AN.

Providers. The set of nodes $p \in P$ denotes the set of service providers that take part in the AN, either actively by providing a service or passively by offering services and SLAs. Service providers are characterized by their supply of services offers $(s \in S)$ as well as their resource constraints L_p that denote the maximum amount of available technical resources such as disk space, bandwidth or availability. Service providers, according to Definition 2.7, constitute leaf nodes in an AN, as their exclusive aim is to provide services. Intermediaries, which are a special type of service provider, according to Definition 2.8, resemble internal nodes that exhibit connections between customer and providers.

Service level agreements. An edge $\alpha_i(p,\theta) \in A$ denotes an SLA between a provider p and a customer θ. An SLA is specified according to Section 2.2 and comprises of a provider p, the consumer θ, the service s as well as multiple Service Level Objectives (SLOs) $l_\alpha \in L_\alpha$. Furthermore, an SLA α comprises of a price $f_\alpha(s)$ and a penalty μ_α. The SLA can take the states *inactive*, as long as there is knowledge about its existence,

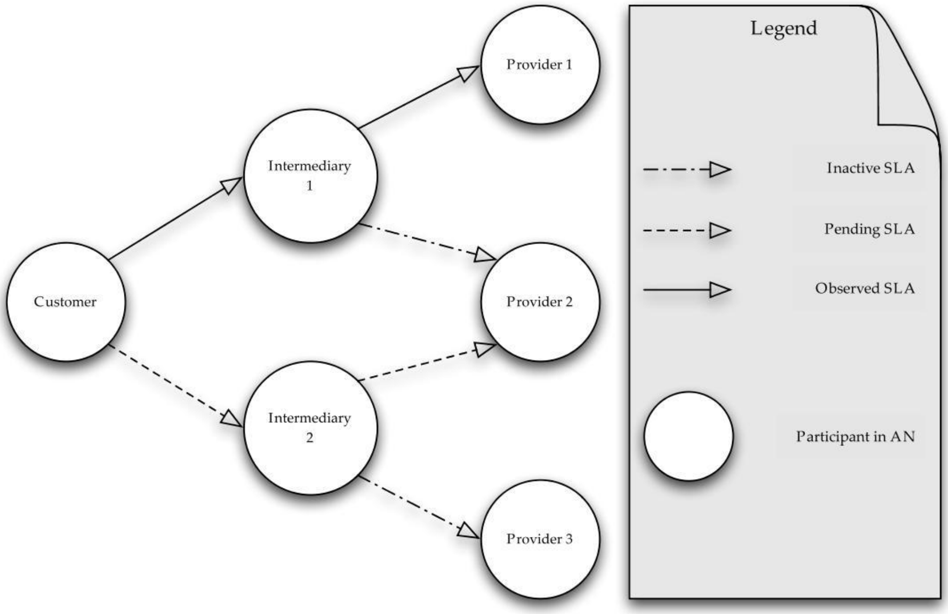

FIGURE 2.3: *Agreement Network Model*

pending if a request for service provisioning has been issued and *observed* if the SLA was established (see Andrieux et al. (2007)).

The formalized agreement network. From the above notation, a specific AN can be formalized as follows: $G := (P, A(P))$.

Figure 2.3 displays an exemplary formalization of an agreement network with three providers, two intermediaries and one customer. There are pending SLAs between the customer, intermediary ι_2 and the provider p_2 as well as observed SLAs between the customer, intermediary ι_1 and the provider p_1. The following section illustrates an example AN in order to highlight the scenario, in which the decision problems that will be introduced in Sections 4 and 5 arise.

2.3.4 Scenario

The purpose of this section is to illustrate the scenario of the thesis. Therefore, in a first step an example for the emergence of an agreement network is given. Second, the viewpoints of an intermediary and a provider are highlighted and the decisions each of them has to face are derived along with the important assumptions required for reasons of feasibility.

Scenario. *Assume that a bank wants to achieve 'better' prices in institutional trading on stock markets. The prices that have to be paid on the market depend on network latency (Wagener and Riordan 2009) giving rise to so called proximity services.*[4] *Proximity services enable the procurement of complete systems that are co-located with a stock market and therefore result in extremely low latency. In the example, these services are offered by a provider that is called "Stock Market".*

Additionally, the bank could opt for the installation of the exchange software on Cloud instances and connect those via an extreme low-latency connection to the stock market. In the example, the provider of Cloud instances is embodied by "Amazon Web Services" with their offer that is called "Elastic Compute Cloud"[5] *by which infrastructure can be used on-demand. The low-latency conncection is offered by "Cisco Systems"*[6]*, a provider of high-performance network connections. A third possibility would result in a combination of co-locating the access point to the stock market, and procure the IT infrastructure by means of a Cloud instance with installed exchange software.*

Now, assume that the bank additionally wants to outsource backoffice processes like accounting and data storage in order to reduce costs. This can be facilitated by employing an accounting software, e. g. "SAP"[7]*, on infrastructure in the Cloud, that is in this case embodied by "CBTS"*[8]*. The resulting agreement network is depicted in Figure 2.4.*

The scenario outlined above illustrates the emergence of an agreement network. In a real-world scenario, analogously to the bank, each of the participants of the AN is not only confronted with this small amount of inactive SLAs but rather faces a big variety of possible cooperations with other providers. The provision of services via the Internet leverages the amount of available alternatives for joint service provisioning. In Figure 2.5, this is indicated by inactive SLAs that the participants face. For simplicity reasons and without loss of generality, the corresponding business partners were omitted.

As already mentioned, to facilitate the emergence of an AN, customers as well as intermediaries have to be able to discover services and corresponding SLAs. Theoretically,

[4] `http://deutsche-boerse.com/dbag/dispatch/de/listcontent/gdb_navigation/technology/30_Access_Products/30_Proximity_Services/page0_ts_sp_proximity_`

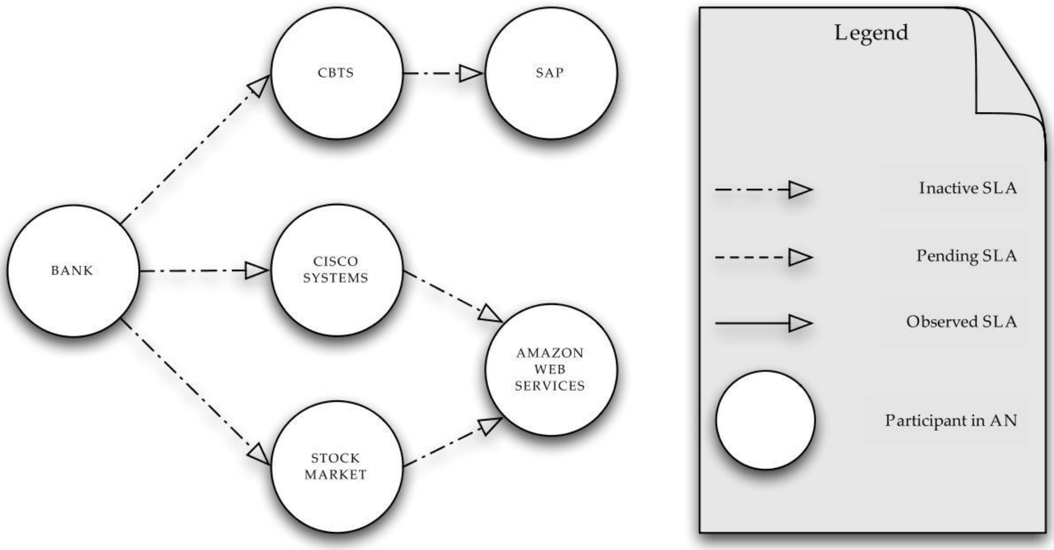

FIGURE 2.4: *AN: Co-location of trading infrastructure*

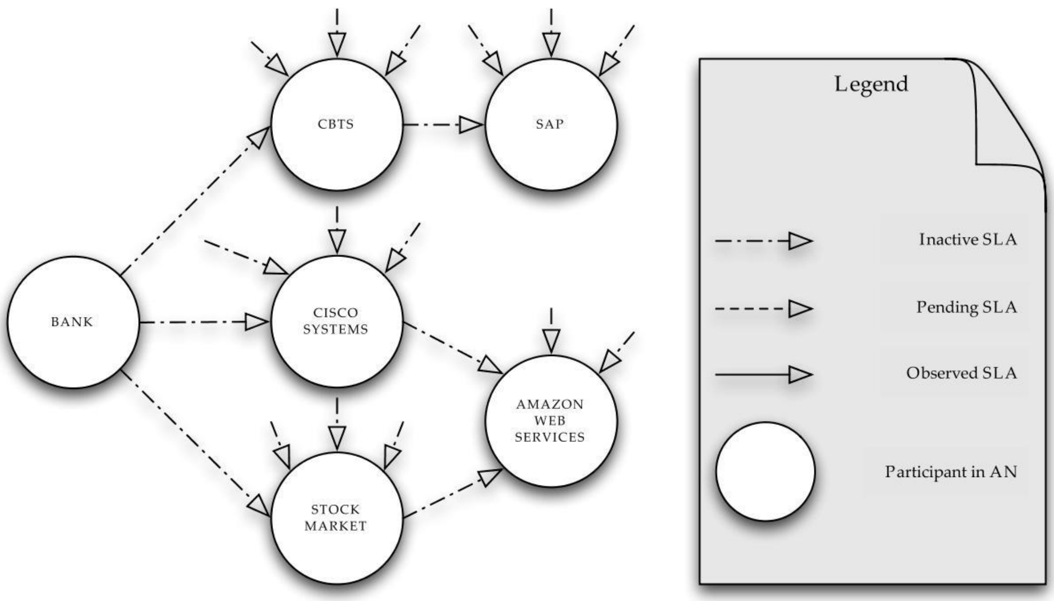

FIGURE 2.5: *AN: Co-location of trading infrastructure - general case*

service providers are able to offer their services in a magnitude of versions with respect to their delivered quality of service. For the customer (or an intermediary that wants to procure a service module), this would lead to a high complexity of choosing the correct service and corresponding SLA. Therefore, for the remainder of this thesis, providers and intermediaries face certain constraints.

Assumption 1 (Pre-defined set of SLAs)**.** *Providers and intermediaries offer a pre-defined set of SLAs for each service. This allows to constrain the number of available SLAs, that is quality classes, for a service.*

For simplicity and without loss of generality, one SLA is assumed to relate to exactly one service, as SLA representations can include multiple services within a single document.

Assumption 2 (Relationship of SLAs and services)**.** *Each SLA regulates the provisioning of one service.*

The role of a provider according to Definition 2.7 in an AN implies the decision problem that arises for the provider. As providers (Definition 2.7) constitute the leaf nodes in ANs, it is obvious that their role does exclusively include the provisioning of services and with it, the decision about the establishment of SLAs with customers. Consequently, the provider's decision problem can be formulated as follows:

Which SLA(s) with customers should be active in the next period?

As there is a multitude of possible agreements for the provider (see Figure 2.5), the provider has to decide for a set of SLAs or in other words, a portfolio of SLAs that they want to activate in the following period. This translates into the decision for a provider that divides all known SLAs into a set of inactive, pending and observed SLAs. This is necessary when the provider is confronted with a request for service provisioning. They need to decide whether to enter pending SLAs or not and on continuing observed SLAs. Consequently, the provider decides about the set or portfolio γ of SLAs $\alpha \in A$ that will be active in the following period.

Analogously, the intermediary's decision is:

services.htm last accessed: October 8, 2011
[5] http://aws.amazon.com/ec2/ last accessed: October 8, 2011
[6] http://www.cisco.com/ last accessed: October 8, 2011
[7] http://www.sap.com/ last accessed: October 8, 2011
[8] http://cbts.cinbell.com/ last accessed: October 8,2011

Which SLA(s) with customers and supplying providers should be active in the next period?

Again as for the provider's case, the intermediary faces a large variety of possible cooperations. As the intermediary does provide services and procure service modules, their decision about active sets, or portfolios, of SLAs is twofold. They need to decide which combinations of two SLA portfolios will be active in the following period: the portfolio γ of SLAs offered to customers and the portfolio δ of SLAs $\beta \in B$ with supplying providers.

2.4 Summary

In this chapter, the first research question of this thesis, that deals with the emergence of agreement networks and their definition and participants, was addressed in detail. A definition of *service* was provided and a delineation of service types that emerged with the availability of the Internet and related technologies was introduced. With the growing momentum of services in business environments, customers require guarantees on the provided quality of service. This leads to an increasing number of existing service level agreements. In order to establish a common understanding of what an SLA is, what it comprises of and what its aim is, a general definition of SLA was given and a special kind of Web service agreements was derived that is applied to regulating the provisioning of Web services.

The joint provisioning of services is leveraged by establishing SLAs. With this trend, a network of agreements emerges among different types of service providers: agreement networks. A definition of ANs was given and backed up by a real-world example that constitutes the scenario of this work. This way, in this section the groundwork was laid for the decision problems of providers and intermediaries that form the main contribution of this thesis.

Chapter 3

Economic Foundations

" 'Would you tell me, please, which way I ought to go from here?' 'That depends a good deal on where you want to get to,' said the Cat."

(Carroll 1865)

JOINT service provisioning that manifests in the emergence of agreement networks was introduced in the previous section.

While it is a networked structure that arises with an increasing degree of joint service provisioning, entering SLAs and providing services still remains the decision of the concerned parties, that is provider and customer of a service. Therefore, this thesis focuses on providers' and intermediaries' roles and responsibilities rather than on the network perspective. The foundations for the decisions of providers and intermediaries that are presented in Section 4 and 5 and their evaluation in Section 7 are laid in this Section.

In Section 3.1, an introduction to decision theory is given. The concepts of *normative* and *descriptive* decision theory are differentiated and the general concept of decisions is illustrated in Section 3.1.1. In a further step, the impact of incomplete information on decision making is highlighted in Section 3.1.2.

As was introduced in Section 2.2, one of the building parts of SLAs is a penalty that represents a liability that becomes active in case of an SLA violation in order to reimburse the customer for perceived service failure. In conjunction with uncertainty about future SLA violations, the payment of penalties becomes an inherent risk for the decision maker. In Section 3.2, foundations for the measurement of risk are laid and different concepts of risk perception are introduced. In Section 3.2.1, approaches for the expression of risk are introduced and the special case of the semi-variance of a portfolio is discussed in Section 3.2.2. Section 3.2.3 introduces utility theory, that enables to express a trade-off between risk and expected profit.

After having introduced the foundations that are relevant to decisions about the establishment of SLAs in which the risk of SLA violation is minimized, current litetature in the area of decisions about SLAs is presented and discussed. Research gaps are identified that highlight the challenges that will be addressed by this thesis.

3.1 Decision Theory

The matter of making decisions has been the subject of many research disciplines in the past. Mathematics, philosophy, logic and especially economics paid special interest in decision making (Resnik 1987). Consequently, a collection of theories exists in the context of decisions, comprising utility theory, social choice theory, and relating to game theory, just to name a few.

In order to shed light on the different areas and topics of decision theory, this section first introduces the types of decisions that exist. Generally, decisions can be classified into group decisions and individual decisions. Considering the scenario at hand, where the joint provisioning of services fosters the emergence of ANs, a certain cooperative element in the decisions cannot be neglected. Without an agreement on the content and terms of a service provision, no SLA could be established. Consequently, the joint provisioning of a service would be the result of a group decision of the participating providers and intermediaries. Nevertheless, the decision about the establishment of SLAs that is to be made by individuals, providers and customers, is not part of a joint decision, like in elections (Black 1948; DeSanctis and Gallupe 1987). It is rather the decision on bilateral cooperation with one business partner respectively. Therefore, the decision is considered to be an individual one.

Furthermore, decision theory exhibits different directions of research that can be distinguished by their aim. On the one hand, descriptive decision theory investigates the question of decision making from an observational point of view (Resnik 1987), that is, it answers the question:

"How are decisions made?"

With respect to decisions in agreement networks descriptive decision theory would observe service providers and intermediaries in reality with the aim to explain how decisions are made and what the determinants of decisions are. In this respect, aspects of empirical economic research and experimental economics offer a toolset for the investigation of making decisions (Kagel and Roth 1995).

A different goal is pursued by normative and prescriptive decision theory (Resnik 1987). Under the assumption of rational decision makers that seek to come to an optimal decision, advice is sought for answering the question:

"How should decisions be made?"

In the case of service providers and intermediaries in ANs, support for the decision on whether to enter SLAs (or not) is required. In order to provide this support, the decision problem needs to be formulated while considering the constraints of providers and intermediaries as introduced in Section 2.3.4. The decision problem considers the guarantee terms of SLAs, agreed prices and penalties as well as the provider's (or intermediary's) constraints on available resources and expectations towards profit. In addition, the chance of violating SLAs is of importance to the decision maker as it has a major impact on profit, or in the case of the intermediary, on performance. The following section introduces the general concept of fomulating decisions and illustrates the aspects of the provider's and intermediary's decision that are covered. The aspect of uncertainty about future SLA violations is introduced in Section 3.1.2 with the concept of decisions under uncertainty.

3.1.1 Decisions in General

A decision involves the choice for one of two or more alternatives (Laux 2005; Resnik 1987). This choice or act by the decision maker produces an outcome, that is, the result of the act and its impact. Translated to the case of agreement networks, providers choose from a set of SLAs that can be entered. The provider chooses to enter these SLAs and to provide the corresponding services. This way, the provider incurs costs for service provisioning and gains a reward, the agreed price.

As an extension to the general setting of selecting an act and perceiving its outcome, decisions need to account for different states of nature (Laux 2005; Resnik 1987) or observable properties that influence the outcome of the decision. Consider the example in the following case: a person starts their car and the check-engine light flashes. Now, they've got two options to act: first, leave the car and call the repair service or second, drive regardless. If the person doesn't enter the car, no negative outcome can be expected. On the other hand, if the person starts to drive, there are several potential states of nature and resulting outcomes. First, the engine is heavily damaged which causes a severe accident and the person is injured badly. Second, the engine is defect and an accident happens, the car is broken but the person isn't injured and last, nothing happens at all. In case of perfect information, the person decides on their action (start driving or not) depending on the event that impacts the outcome.

In the case of a provider's decision on establishing SLAs the problem is sketched as follows. The provider decides either to enter SLAs or not. In the case that they enter SLAs, costs for providing the respective services arise and prices for service provisioning are collected. Additionally, depending on the degree of SLA violation by the provider, they may have to pay a penalty to the customer. Consequently, the provider includes the degree of SLA violation and resulting penalty in the decision about entering SLAs as this influences the profit from service provisioning. This decision process is only applicable if there is perfect information about the events that influence the outcome. In the case of a service provider and the degree of SLA violation, the information about future SLA violations is not available, therefore a different approach has to be applied, which is presented in the following section.

3.1.2 Decisions under Uncertainty

There are many situations, in which decisions have to be made without complete information about the possible states of nature or events that influence the outcome of a decision. Consider again the example from the previous section about starting to drive the car or not. The person that has to decide about driving most likely does not know the actual severity of the damage of the engine when the decision is made.

Analogously, a provider in an AN that has to decide about the active portfolio of SLAs for the next period has to consider achieved prices, costs for service provisioning, resource constraints and penalties that have to be paid in case of SLA violation. The degree of SLA violation for each of the SLAs is subject to the effort that the provider exerts in order to avoid SLA violations like providing redundant systems or service personnel. However, there are factors that the provider cannot influence like hardware failures, power outages, or human error. Therefore, there is uncertainty with respect to the degree of SLA violations in the future.

Decisions under uncertainty can be classified into decisions for which there is a belief for the probability distributions for events and those, for which there is no such distribution available (Laux 2005; Resnik 1987). For service providers according to Definitions 2.7 and 2.8 that decide on the establishment of SLAs, there is no information about the likelihood of realizations of degrees of violation in the future. A common approach to solve this problem is to approximate the probability distribution from past observations, e. g. monitoring observations of SLAs that were established in the past and corresponding degrees of violation. An application of this approach is applied in Chapters 4 and 5. This allows a provider to include the probability for a particular degree of violation with the resulting penalty in their decision about the portfolio of SLAs that is to be active in the following period.

3.1.3 Conclusion

This section briefly introduced decision theory as a research method for economic decision making. Different branches of decision theory were introduced and the approach of a prescriptive model for individual decisions was identified as the method of choice for the scenario that was presented in Section 2.3.

The basic building parts of a decision were identified as *acts*, *states*, and *outcome*, where each of the states can be known or unknown ex ante. Decisions under uncertainty, in which a probability distribution for the states is known in advance, are presented as a potential solution to the decision problems that arise in the scenario of the work at hand.

Having presented the basic methodology with which the decision problem is going to be solved, a means to evaluate the outcome of each decision that is taken is required. Therefore, the following sections introduce measures of risk that lay the basis for the inclusion of the risk of SLA violation in the provider's and intermediary's decision about future active SLAs.

3.2 Implementation of Decisions under Uncertainty

The previous section described decisions under uncertainty as a method that allows choices to be made among alternatives even if there is uncertainty about the state of nature that influences the outcome of a choice. In this context, the concept of risk plays an important role, where risk is understood as the uncertainty of a future event.

The decision for one of the available alternatives includes the outcome as well as its probability of occurence. In order to be able to choose *the optimal* alternative, the uncertainty of the occurence of an event, that is the risk, needs to be formulated. In this respect, decision support often includes the formulation of an objective function (Laux 2005) that is to be optimized, that is, minimized or maximized with respect to the semantic meaning of the objective function.

Formulating an objective function as a means to decide for one of the available alternatives includes the specification of risk. In the case of a provider's choice for a portfolio of active SLAs in the next period, the risk of SLA violation has to be formulated. A probability for a particular degree of SLA violation is also required to express the penalty that can be expected. The resulting function that expresses the risk of SLA violation has to be minimized. This is the objective function, by which the desirable portfolio of SLAs is identified. The following sections introduce several established measures of risk from different research directions and discuss their applicability in an objective function for a provider's and an intermediary's decision problem. In a further step, the concept of utility functions is introduced in order to include multiple objectives like maximizing profit and minimizing the risk of SLA violation in the objective function.

3.2.1 Measuring Risk

The concept of risk affects a variety of research disciplines like finance, insurance, and business administration, but does as well affect completely different areas like technology acceptance (Pavlou 2003) or research in nuclear power. Consequently, many different approaches for measuring risk exist. This section serves to introduce the most common methods to measure risk and discuss their applicability.

Stochastic Dominance

In order to select one of the available alternatives, an ordering has to be established among the options. Stochastic dominance is a concept that allows the sorting of alternative outcomes that underly an uncertainty of occuring by means of their probability distributions.

Below, the ordering according to first and second order stochastic dominance is explained. Assume that there are two random variables, X and Y, which have the same domain. X is distributed according to a cumulative distribution function (CDF) F and Y's CDF is G. With respect to first order stochastic dominance (FSD), F dominates G if the following holds (Levy 1992):

$$F(x) \leq G(x), \ \forall x. \tag{3.1}$$

This implies that distributions with a more homogeneous slope are preferred to those with a rather unstable slope. The heterogeneity of the slope is therefore perceived as more risky and hence, the riskier alternative is rejected. The sorting of alternatives according to FSD is only possible, if one of the CDFs dominates the other one in the whole domain. As soon as equation (3.1) does not hold for all x, no ordering is possible. Therefore, the concept of second order stochastic dominance (SSD) was introduced (Hadar and Russell 1969; Hanoch and Levy 1969; Hanoch and Levy 1970; Rothschild and Stiglitz 1970). According to SSD, G is dominated by F in case that

$$\int_{-\infty}^{x} G(t) - F(t)dt \geq 0, \ \forall x \tag{3.2}$$

holds. Equation (3.2) implies that F does not need to be more stable over all of the domain. However, SSD requires that in total, F is less risky than G by means of the amount of instability in the distribution. In order to evaluate available alternatives with respect to SSD, the left-hand-side of Equation (3.2) is not permitted a change of sign. As soon as there is a change of sign, no ordering among alternatives F and G is possible.

The application of stochastic dominance to decision theory has been a matter of research for many decades (Allais 1953; Fishburn 1964). In order to apply stochastic dominance to the provider's and intermediary's decision, the CDFs of the degrees of SLA violation in conjunction with the resulting penalties would be compared according to Equations

(3.1) and (3.2) and sorted with respect to the outcome of dominance. However, with a growing number of alternatives, the pairwise comparison of each alternative with each other that is required to achieve a ranking among all alternatives, becomes increasingly infeasible. Even more, if neither FSD nor SSD can be shown, no ordering is possible among alternatives and consequently, no choice for one alternative is possible. Therefore, stochastic dominance is not feasible as a measure of risk in the scenario of the work at hand.

Value at Risk

Measuring the risk of a loss that a particular portfolio of securities incurs gained much attention in financial mathematics and financial risk management in recent years. A concept that is frequently employed is Value at Risk (VaR) which is defined as a threshold value that is not exceeded with a certain probability in a specific period. More detailed, if a portfolio exhibits a 2.5% VaR of 100,000 € for one day, then there is a 2.5% probability that the portfolio will incur a 100,000 € loss within the next day.

According to Jorion (2007) the VaR of a portfolio is defined as "*the worst loss over a target horizon such that there is a low, prespecified probability that the actual loss will be larger*". In order to calculate it, define ψ as the confidence level and *Loss* as the loss, which leads to

$$Prob(Loss > VaR) \leq 1 - \psi. \tag{3.3}$$

The VaR is found by identifying the smallest value for which Equation (3.3) holds. For example, if $\psi = 0.99$, then VaR is the smallest value for which the probability of incuring a higher loss is smaller than 1%. In order to include the period of time for which the decision is made, and to explicitly calculate the amount of money that could potentially be lost, VaR is expressed as follows (Jorion 2007):

$$VaR = MV(\gamma) \cdot \sigma(\gamma) \cdot t_{disc} \cdot \phi(\psi), \tag{3.4}$$

where $MV(\gamma)$ denotes the current market value of the portfolio γ, $\sigma(\gamma)$ is the standard deviation of profits of γ in the past, t_{disc} states the disconted period of time relative to the trading year and $\phi(\psi)$ resembles the absolute confidence interval that results from the confidence level ψ under the assumption of a normal distribution of profits.

By applying VaR as a measure of risk, a decision maker is able to estimate the loss that will not be exceeded with high probability, which depends on the chosen confidence level. VaR is applied in many decisions on the financial market, especially in institutional trading (Holton 2003). In the context of decision making in ANs, service providers and intermediaries would be able to calculate the loss of a portfolio that will not be exceeded with a certain probability if VaR was applied as measure of risk. The portfolios would then be sorted according to the amount of money that results from VaR calculations. Even though this approach exhibits a basic feasibility for the application in ANs, there are some drawbacks that prevent its application. First, VaR focuses on the tails of the distributions and ignores the center of distributions. This leads to the neglection of the big mass of losses each of which might be lower than the amount that is identified by VaR but accumulates a higher probability of occurence (Einhorn and Brown 2008). Second, VaR does only point to an amount of loss that will not be exceeded in a particular period of time, that is, it highlights merely a worst case and hence could provide false confidence (Taleb 1997). Therefore, VaR is not feasible for the application of the scenario that was introduced in Section 2.3.4.

GINI's mean difference

is also known as absolute mean difference. It describes an overall degree of dispersion of a random variable. For a discrete random variable, GINI's mean difference is based on the absolute deviation of every pair of realizations of the random variable (Shalit and Yitzhaki 1984; Yitzhaki 1982). For a continuous random variable X with probability density function g, the GINI coefficient is calculated as follows:

$$GINI = \frac{1}{2} \int_a^b \int_a^b |x_i - x_j| \cdot g(x_i) \cdot g(x_j) \mathrm{d}x_i \, \mathrm{d}x_j, \tag{3.5}$$

where x_i and x_j denote realizations of the random variable X, and a and b denote the limits of the domain of the probability distribution. GINI's mean difference is bounded from above by $\overline{X} - a$. The lower bound is zero. Values closer to zero constitute a low degree of dispersion and hence, low risk, whereas values closer to $\overline{X} - a$ denote a high degree of dispersion. Due to the fact that a uniform distribution does not expose any dispersion, GINI's mean difference for a uniform distribution is equal to zero. Consequently, GINI's mean difference can as well be interpreted as a proxy for the deviation from a uniform distribution.

In order to include GINI's mean difference in the objective function in the provider's and intermediary's decision in an AN, the absolute deviations of each observed degree of failure including the resulting penalty would have to be calculated. By calculating GINI's mean difference for each portfolio, an ordering of portfolios according to their degree of dispersion and hence, risk, is possible. One value of GINI's mean difference can relate to a large variety of distributions. This is basically caused by the application of the absolute difference between values that is taken into account in Equation (3.5). This implies that GINI's mean difference is only able to account for dispersion but not for the quality of the unequal distribution. Employing GINI's mean difference in ANs would consequently only indicate unequal distributions of degrees of violation but not in which area the center of the distribution, that is, the majority of violations lies. This could lead to a situation in which two SLAs are rated equally with respect to GINI's mean difference, but one of the distributions exhibits the center in lower degrees of violation and one exhibits the center in higher degrees of violation. Consequently, GINI's mean difference can only give a hint on unequal distributions but does not differentiate with respect to the severity of the violation. Consquently, it does not suffice as a measure for risk.

Mean absolute deviation (MAD)

Similar to GINI's mean difference, MAD is a measure for inequality, whereas MAD employs the absolute deviations from the expected value of a distribution (Konno and Yamazaki 1991). This way, MAD expresses risk as the deviation from a target value, which in this case is the expected value. In general, for a sample of size N MAD is calculated as follows:

$$MAD = \frac{1}{N}\sum_{i}^{N}|x_i - \overline{x}|\,, \tag{3.6}$$

where x_i denotes the observations and $\overline{x}$ states their expected value. For security portfolios, this translates to the sum of expected deviations from the expected profit of each security. Translated to the case of the provider's decision about the establishment of SLAs, as a first step, the expected degree of SLA violation would be calculated. The target value is retrieved by multiplying it with the corresponding penalty. In a second step, for each observation, the absolute deviation from the expected penalty is calculated and the mean absolute deviation is computed. After calculating MAD for each available portfolio, an ordering of portfolios is feasible.

Similar to GINI's mean difference, MAD resembles a proxy for dispersion. In contrast to GINI's mean difference, dispersion is expressed by the deviation from a target value rather than from a uniform distribution. Nevertheless, applying MAD as a measure of risk implies that all deviations are perceived equally bad by the decision maker. For a service provider or intermediary this implies that deviations from the expected value are considered risky, no matter if the deviation expresses less violation than expected or more violation. Consequently, even though MAD exhibits properties that suggest is as suitable for the application for a measure of risk it lacks the ability to evaluate deviations with respect to their implication, that is, if the deviation indicates gain or loss to the decision maker.

Lower Partial Moment (LPM)

Measuring risk by means of the deviation from a target value has already been introduced in the case of MAD. Similar, LPM denotes a measure for risk in which only those deviations are taken into account that are "*worse*" than a target value. This way, LPM expresses the risk of shortfall (Bawa 1975; Libby and Fishburn 1977), which is calculated as follows

$$LPM = E((\max\{\overline{x} - x_i, 0\})^k), \tag{3.7}$$

where x_i are the observations and $\overline{x}$ their expected value as before, k expresses the *order* of the LPM, and $E(\cdot)$ denotes the expected value.

According to Equation (3.7), a provider in an AN would resort to calculating the expected degree of SLA violation and multiply it by the corresponding penalty. This way, the target value is retrieved. Having retrieved the target value, the deviations that are "*worse*" than average need to be identified. In the case of a provider, loss by means of penalties is considered rather than profit. Therefore, Equation (3.7) needs to be adapted in order to account for the risk that higher penalties than expected incur:

$$LPM' = E((\max\{x_i - \overline{x}, 0\})^k). \tag{3.8}$$

Consequently, by applying Equation (3.8), for each observed degree of violation, the deviation from the target value can be calculated. For those observations that are higher than the expected penalty, the deviation is taken to the k^{th} power. This way, only those deviations from the expected degree of violation are taken into account that impose the risk of loss to the decision maker. Furthermore, the factor k can increase the weight of bigger deviations, implying that larger deviations from the expected value impose a higher risk to the decision maker.

A special instance of the LPM is retrieved in the case of $k = 2$. This instance is also known as the semi-variance , which is a special case of the variance in which only those deviations that are "*worse*" than average are considered. A more detailed discussion of the semi-variance is provided in the following section.

3.2.2 The Semi-Variance as a Proxy for Risk

As was argued at the beginning of Section 3.2, risk is understood as the uncertainty of a future event. In agreement networks, service providers and intermediaries perceive risk as the uncertainty of future SLA violations and the resulting penalties. In Section 3.2.1, different approaches for expressing risk have been introduced and discussed. As a result, LPM has been identified as a feasible measure of risk in the context of ANs as it accounts for deviations from the expected degree of SLA violation that are "*worse*" than average and enables the weighting of deviations by means of the factor k (see Equation (3.8)).

The semi-variance as a special instantiation of LPM has been employed in decisions in finance since it has been suggested in Markowitz (1959) and is as follows:

$$SV = E((\max\{\overline{x} - x_i, 0\})^2), \tag{3.9}$$

where x_i states the observed profits of the portfolio, $\overline{x}$ their mean, and $E(\cdot)$ the expected value. In this regard, the semi-variance expresses risk as the deviation from expected profit, where larger deviations are weighted a higher risk, which is achieved by squaring the deviations. By employing the LPM methodology, the semi-variance manifests itself as an adaptation of the variance (and with it, the standard deviation), which was applied as a measure for risk in finance and especially, in evaluating the risk that is incured by a share for a long time (Elton et al. 2003; Merton 1972). In current research, computational properties of the semi-variance approach are investigated to evaluate the performance of heuristics that approximate the semi-variance for the selection of portfolios in order to improve the computational time that is required to identify the solution (Buhl and Fridgen 2011).

Besides its application in portfolio theory, semi-variance found applications in many other branches. Exemplarily, the management of project portfolios, that is the selection of projects that promise success, can be facilitated by employing the semi-variance (Hubbard 2007). Furthermore, areas that are completely unrelated to the selection of stock portfolios were shown to be feasible application areas of the semi-variance. This includes the modeling of labor force in the economy, which allows the investigation into properties of the labor force like growth and variance (Chandra 2003). Portfolio theory has been employed in psychology in order to express the stability of the "*self-concept*". In cases, in which the *self-attributes*, that comprise the *self-concept*, manifest in a diversified portfolio, self-esteem is assumed to be more stable than in undiversified cases (Chandra and Shadel 2007).

In order to apply the semi-variance as a proxy for the risk of SLA violation, an adaptation of Equation (3.9) is required. Risk in the context of SLAs arises due to the uncertainty of SLA violations. Consequently, the deviations that a service provider and an intermediary perceive as "*worse*" than average are those, that are higher than expected. Therefore, Equation (3.9) is adapted as follows

$$SV' = E((\max\{x_i - \overline{x}, 0\})^2), \quad (3.10)$$

where x_i denotes the degree of violation and $\overline{x}$ the expected degree of violation. This adaptation considers only observed SLA violations that are worse than average. The formulation of the semi-variance in Equation (3.10) resembles a proxy for the risk of SLA violation that can be employed in the objective function of the decisions about the establishment of SLAs.

3.2.3 Utility Theory

The decision problem on the establishment of SLAs that a provider (according to Definitions 2.7 and 2.8) faces was described in Section 2.3. Besides the risk of SLA violation, factors that constrain a provider's decision about SLA establishment were identified. Resources that are available to a provider constitute one of the constraints. The expected profit that can be achieved by service provisioning (and consumption) resembles another driving factor.

For each of the alternative SLA portfolios, a particular resource consumption, expected profit and risk of SLA violation exists. Objective functions that employ one criterium express the preference relation of the decision maker about this criterium. In the case of a provider that makes decisions on SLA establishment, the minimization of SLA violation risk constitutes the objective function that is evaluated subject to resource and profit constraints. In order to include the expected profit in the objective function rather than as a constraint and to be able to express the trade-off between risk and profit that the provider is willing to accept, an objective function that includes multiple criteria is required. For the ranking of available options in the presence of multiple criteria different approaches exist.

First, approaches that do not formalize preference structures are available (Keeney and Raiffa 1993). In this kind of approach, *dominant* alternatives are identified. This is achieved by comparing each criterium of two alternatives, e. g. profit and risk of two SLAs. One SLA dominates the other one, if it is equal or better in all criteria and better in at least one criterium, that is, for instance, the expect profit is equal for both SLAs but one SLA has a lower risk of SLA violation (Keeney and Raiffa 1993, p. 69). Employing this approach allows the identification of the *efficient frontier* among the available options (Keeney and Raiffa 1993, p. 70), which denotes the set of options that is not dominated by other options (Markowitz 1959). This approach does only lead to an unambiguous ranking among alternatives, if dominant alternatives can be identified. As there are situations, in which dominance is not given,[1] decisions in which preference structures are not formalized are not feasible for the application in the decision on SLA establishment.

Second, there are approaches that formalize preference structures by determining a trade-off between criteria. Therefore, for each criterium the respective *valuation* is determined. These valuations are then associated with a *weight* in order to express relative importance for the decision maker. For a provider that wants to decide between two available SLAs, the valuations for the expected profit and risk of SLA violation are determined and weighted. By adding the weighted valuations of profit and risk for each of the SLAs, one value results for each SLA and the provider is able to select the SLA with the higher valuation (Keeney and Raiffa 1993, p. 75).

An alternative approach that expresses a trade-off between risk and expected profit is the definition of a utility function (Savage 1972), i.e., a decision maker's attitude towards risk is reflected in a utility function by illustrating the gain in utility from addi-

[1]For instance, if the expected profit of SLA α_1 is higher than the profit of SLA α_2 but the SLA violation risk of SLA α_1 is higher, and hence worse, than the SLA violation risk of SLA α_2.

tional marginal units of profit. In this context, risk-averse, risk-neutral and risk-loving individuals can be observed. Risk-averse individuals expose a decreasing marginal utility, thereby stating that the additional gain from higher profit decreases the higher the profit is. Risk-neutral individuals have a constant marginal utility, whereas risk-loving individuals exhibit an increasing marginal utility (Friedman and Savage 1948). In this context, risk-aversion states that the expected utility of a gamble is lower than the utility of the expected value of the gamble. This implies that risk-averse individuals prefer safe payoffs that are lower to risky payoffs that might be higher. The lowest amount of money that the risk-averse individual is willing to accept instead of deciding for the gamble is called certainty equivalent (Keeney and Raiffa 1993). The more risk-averse an individual is, the lower is the amount of money that the indivual is willing to accept as a safe payment as opposed to a gamble. The degree of risk-aversion manifests in the slope of the utility function[2] (Von Neumann et al. 1947) and can be expressed with the help of relative and absolute coefficients of risk-aversion (Arrow 1971; Pratt 1964). This means that an individual's preferences can be expressed by means of a utility function. The slope of the utility function provides information on the individual's attitude towards risk. By applying a utility function, a provider is able to rank SLA portfolios with respect to the expected profit and the uncertainty (risk) that the portfolios incur. After calculating the *expected utility* for each of the alternative SLA portfolios, the provider selects the portfolio that maximizes expected utility, that expresses a trade-off between profit and risk.

3.3 Existing Work on SLA Decisions and its Implications: A Critical Analysis

Many aspects of joint service provisioning have recently been subject to research. Technical preliminaries that have to be coped with in order to facilitate the provisioning of complex higher-valued services to a customer were investigated. Different approaches that solve, for instance, the technical composition of Web service modules have been suggested (Berardi et al. 2005; Rao and Su 2005; Zeng et al. 2003). The composition of Web services by means of platforms like Salesforce's composition platform force.com[3] has gained momentum over recent years. The cases of SAP's service marketplace[4] or

[2]Risk-averse individuals exhibit a concave utility function.
[3]`http://www.salesforce.com/platform/` last accessed: October 8, 2011
[4]`https://service.sap.com` last accessed: October 8, 2011

IBM's Cognos Mashup Service[5] document the rise of branded service communities, or service parks, where a provider of a core service attracts providers of add-ons. In service parks (Petrie and Bussler 2008) the provider of a core service is the single point of contact to the customer and selects the matching complementary services from supplying providers, analogously to the intermediary's decision that was introduced in Section 2.3.4. For the joint provisioning of services, providers have to be aware of available services and need to be able to discover services. To solve this challenge, service repositories, which allow for the semantic querying of services by means of functionality and quality aspects have been researched (Agarwal et al. 2008; Junghans and Agarwal 2010). This work is considered complementary to the work at hand and ensures the availability of required information as input to the decision making process about the establishment of SLAs.

First approaches towards a decision concerning the establishment of SLAs have been described in Lang et al. (2008). A scenario, in which a service intermediary offers services of which the intermediary is not able to provide all of the functionality by themselves. Thus, an intermediary needs to procure services from other providers. In analogy to the decision problem of an intermediary (see Definition 2.8), an intermediary needs to decide which services will be procured. For this scenario, an optimization model is introduced that allows a service intermediary to maximize profit by selecting services on a functional level in a service oriented architecture (SOA). The profit can be increased by offering functionality that exceeds the requested one. Quality aspects or penalties in SLAs are not taken into account, nor is the risk of SLA violation.

Including the risk of SLA violation in the decision on SLA establishment is crucial for a service provider. SLA violation results in the payment of penalties. Furthermore, a provider's reputation decreases if the violation of SLA becomes public, which may lead to a lower future demand. A model that describes governance, risk and compliance for service outsourcing is introduced in Asnar et al. (2010). Challenges in the context of Service Oriented Architectures (SOAs) and service composition are highlighted and the paradigm of compliance by control is applied for enforcing compliance based on monitoring results. The model allows for the qualitative identification of risk, that is, what the risk is, and the definition of counter-measures in a very technical way. However, a quantitative measurement of risk that defines how severe the impact of the risk is, is not supported, nor is decision support that minimizes the risk of SLA violation. In Asnar et al. (2007), a risk management model is introduced. The identification of risk is possible, but quantitative risk measurement is not feasible with this approach.

[5] `http://www-01.ibm.com/software/data/cognos/products/cognos-8-business-intelligence/cognos-8-mashup-service.html` last accessed: October 8, 2011

Similarly, Kompella et al. (2010) introduce an approach for risk modeling for localizing fault in IP networks. Risk is applied for the analysis of monitoring data to identify common failure patterns. By applying this approach error propagation along links in IP networks can be detected and common risks can be identified. Analogously to Asnar et al. (2007) and Asnar et al. (2010), a qualitative description of risk is given, but no quantification or metric is introduced that would allow the evaluation of the severity of a failure.

Recently, the joint provisioning of services and related economic considerations have drawn the attention of researchers. There is a major body of work that concentrates on the coordination of value creation in joint service provisioning. This includes setting prices for the complex service and the single contributed services, and in the meantime incentivizing service providers to provide services as agreed.

In Blau et al. (2008), an approach is introduced that identifies compatible service modules for the formation of a complex service that provides requested fucntionality. A mechanism-design based approach allows the selection of the set of service offers that fits the customer's needs best and incentivizes providers to submit truthful bids with respect to their costs. The approach is extended in Blau (2009) in order to account for service failure by reducing a provider's payoff in the case of service failure. This way, the adherence to an agreed quality of service is taken into account in the payoff structure. However, the risk of violating an SLA is not considered in the selection of services. Furthermore, the approach focuses on the efficient choice of services for the composition of one complex service. Decisions on the establishment of SLAs with a multitude of customers and hence, the provisioning of more than one service, are not considered.

A similar approach for the selection of services is introduced in Conte et al. (2009). This approach employs the same selection of service offers as (Blau 2009). However, providers' payoffs are determined to reward contribution to the variety of possible compositions. This approach rewards a service provider in a service composition scenario for holding their service ready and hence, increasing the possibilities of recovery in case of service failure. Alike Blau (2009), Conte et al. (2009) focuses on the provisioning of a single complex service to a single customer. A multitude of customers or provided services and the resulting choices for SLAs in joint service provisioning are not taken into account. Thus, strategic decisions are only possible for an intermediary that selects services that are offered from providers. Providers' decisions for offering services are only considered by means of a bid price determination but not with respect to offering a service and a corresponding SLA or not. Furthermore, the risk of SLA violation by providers is not taken into account in the selection of services.

An economic view on service composition and incentivation of participants is shown in Filipova-Neumann et al. (2010). In this setting, an intermediary with complete market power determines prices and penalties for customers, to which they offer complex services. Additionally, the intermediary sets prices and penalties for services that they procure from providers. The exact choice for a particular price and penalty incentivizes service providers and customers to participate in the market. The intermediary is able to set incentives for service providers to invest in the effort that they exert in order to adhere to the established SLAs. In the decision making process on prices and penalties, the uncertainty of future SLA violation is taken into account. However, the focus of the approach lies on setting incentives to providers in a pre-defined set of SLAs by defining prices and penalties. The selection of service providers to cooperate with or the corresponding SLAs is not considered. An adaptation of the approach in Filipova-Neumann et al. (2010) would allow a service intermediary to select a portfolio of SLAs and to set prices and penalties in SLAs, in order to incentivize providers to adhere to the active SLAs in a further step.

The tasks of an intermediary comprise of selling services to customers and buying services from supplying providers. As was argued in Section 2.3.2, the profit of service intermediaries is influenced by the establishment of SLAs. In this vein, an intermediary can be compared to a market maker in financial markets that buys and sells stocks in order to make profit. Hart and Jaffee (1974) discuss the application of portfolio theory as introduced by Markowitz (1952) to the behaviour of a depository financial intermediary that holds assets and liabilities. A unique solution for an intermediary's utility-maximization problem is retrieved that is determined by an intermediary's attitude towards risk. The approach allows for the selection of a combination of assets and liabilities, whereas risk and profit can always be measured directly as profit resulting from buying and selling stocks. In the decision of a service provider or an intermediary on the establishment of SLAs, profit cannot be measured directly from selling goods or shares. The profit of a service intermediary employs a definition of profit that reflects the intermediary's profit from providing services to customers and from procuring services from supplying providers. Thus, the expected profit reflects the uncertainty of SLA violations.

O'Hara and Oldfield (1986) examined the decisions of a risk-averse market maker in stock markets. Alike Hart and Jaffee (1974), a situation is investigated where profit directly results from buying and selling stocks.

In the context of decisions under uncertainty, measures of risk play an important role in the formulation of the objective function that allows the selection of one of the available alternatives. As was already argued in Section 3.2, different measures of risk are employed in most diverse branches. Approaches for expressing risk by deviations from a target value include the mean absolute deviation (MAD) (Konno and Yamazaki 1991), Gini's risk measure (GINI) (Shalit and Yitzhaki 1984; Yitzhaki 1982), and lower partial moments (LPM) (Bawa 1975; Libby and Fishburn 1977).

In general, MAD states the expected deviations from an expected value. For security portfolios, this translates to the absolute expected deviations from expected profits of each security (Konno and Yamazaki 1991). Applying this measure of risk allows a decision maker the determination of the dispersion of a set of observations. In the calculation of MAD, all of the deviations from the expected value are included. MAD considers a deviation from a target value as risk, no matter what the semantic meaning of the deviation is. In the context of SLA violation uncertainty MAD would consider each deviation from expected violation as a risk, no matter if the deviation corresponds to better or worse performance than expected.

The GINI coefficient is a proxy for statistical dispersion like MAD. It can be considered a proxy for the deviation from an equal distribution, whereas a coefficient of 0 denotes an equal distribution and higher values of the GINI coefficient state a more unequally distribution of values (Shalit and Yitzhaki 1984; Yitzhaki 1982). The optimization problem in this case would result in the minimization of the GINI coefficient while taking constraints on expected profit into account. Analogously to MAD, GINI treats the unequality of the distribution of profit as a whole as risk. The separation of deviations that are worse than expected from those that are better than expected is not possible.

LPM denotes a measure for the risk of shortfall, that is, only those deviations that are "worse" than a reference point Z are taken into account (Bawa 1975; Libby and Fishburn 1977). In order to calculate the risk of shortfall, the expected value of a function of those deviations is applied. One example for this function $g(\cdot)$ is $g(\cdot) = (\cdot)^2$. In this case, shortfall denotes the semi-variance of profits/penalties and enables to include the semantic meaning of a deviation.

The decision on SLA establishment that considers the risk of SLA violation constitutes a provider's and an intermediary's business concept. Defining the business concept and with it the future strategy of an enterprise by means of the products that will be offered as well as respective prices has been in the focus of marketing research and specifically product portfolio management for many years. Product portfolio management includes methods for the evaluation, selection, and priorization of new projects, as well as the assessment of existing projects, which leads to their acceleration, deprioritization or the end of the project. Finally, resources are allocated and reallocated to the active projects. Cooper et al. (1999) gives an overview of current approaches for product portfolio management, none of which is applicable to the determination of the risk of Web service SLAs as concrete specifications of risk are missing. A method that allows to apply product portfolio management to the establishment of SLAs especially in the context of service provisioning over the Internet is still missing.

In summary, in current literature, there has been no approach that allows for all of the challenges that arise in the context of decision making on SLA establishment for the provisioning of services, especially via the Internet. The challenges that need to be considered are the technical composition of services and the impact of the Internet on service provisioning. Furthermore, provisioning of more than one service to more than one customer is important for a provider's decision on SLA establishment, as well as the possibility of quantifying risk, thus measuring risk, and including risk in the selection of SLAs to establish is. Table 3.1 summarizes the challenges that arise and indicate which of the challenges are already considered in current literature, where ● denotes that a challenge is considered and ○ where it is not. From Table 3.1 it becomes obvious that a decision support that considers all of the challenges is still missing.

This section served to give a thorough introduction to the economic foundations of this thesis. As a first step, an overview over decision theory was given and a distinction of decisions by the respective application scenario was given: decisions that are made by individuals were differentated from group decisions and the degree of information was introduced as a factor for identifying decisions under uncertainty. Due to the fact that there is uncertainty in future SLA violations, providers and intermediaries in ANs are confronted with decisions under uncertainty.

In order to facilitate decisions under uncertainty, the available alternatives, different SLA portfolios, that can be active in the next period, need to be evaluated. Therefore, a measure for the risk of SLA violation that can be employed in the objective function of the decision problem is required. In Section 3.1, a multitude of different measures of risk was introduced, explained and discussed. Finally, a special instantiation of LPM, the semi-variance , was identified as the measure of risk that fits best to the case of decision making in ANs as it enables the consideration of only those deviations from a benchmark that are perceived as "*worse*" than expected. Additionally, a means to express a trade-off between risk and expected profit with the help of utility functions is introduced in Section 3.2.3.

Chapters 2 and 3 have jointly laid the foundations of this thesis. Whereas Section 2

TABLE 3.1: *Critical Evaluation of Related Work*

Literature	technical composition of services	measurement of risk, quantification	impact of Internet on service provisioning considered	more than one customer	risk in selection considered
Rao and Su (2005), Berardi et al. (2005), Zeng et al. (2003),Petrie and Bussler (2008),(Agarwal et al. 2008; Junghans and Agarwal 2010)	●	○	●	●	○
Lang et al. (2008)	●	○	●	○	○
Asnar et al. (2007),Kompella et al. (2010),Asnar et al. (2010)	●	○	●	●	●
Blau et al. (2008),Blau (2009),Conte et al. (2009)	●	○	●	○	○
Filipova-Neumann et al. (2010)	○	○	●	○	○
Hart and Jaffee (1974),O'Hara and Oldfield (1986)	○	●	○	●	●
Bawa (1975), Libby and Fishburn (1977),Yitzhaki (1982), Shalit and Yitzhaki (1984),Konno and Yamazaki (1991)	○	●	○	○	n.a.
Cooper et al. (1999)	○	●	○	●	●

introduced the environment in which decisions are made, Section 3 introduced and discussed the means of decision making and measuring risk. Related work that has been carried out recently in the area of joint service provisioning and decision making about SLAs was presented and discussed and challenges that arise with respect to decisions about SLA establishment were highlighted. The following sections are dedicated to solving these challenges.

Part II

Towards an Optimal Choice of Contracts

Chapter 4

Service Provider Decisions

"If the SaaS provider is not up and running 24*x*7*x*365*, or if they're not fully secure, they're out of business."*

(Business Finance Magazine 2010)

THIS chapter merges the foundations on joint service provisioning and SLA establishment from Chapter 2 with the economic foundations from Chapter 3 in order to solve the challenges in decision making for SLA establishment that were identified in Section 3.3. A thorough literature review highlighted that currently no decision support for service providers exists that caters for the whole spectrum of requirements that arises in SLA establishment decision making for all services (including e-Services, Web services and Cloud services (see Section 2.1.2)). The decision problem of a service provider on the establishment of SLAs is influenced by the changes in service provisioning that result from the availability of the Internet. By offering and providing services via the Internet a service provider is able to exploit the *Long Tail* phenomenon (Anderson 2006), as was argued in Chapter 1. Even providers of *niche* services are able to reach more customers. Consequently, the number of requests that are issued to the provider (the demand) increases. For providers of standard services a high volume of requests for service provisioning can be assumed to be issued, as well, according to OECD (2010). With a high number of requests for service provision, the number of SLAs for which the establishment needs to be decided is high, and consequently, the decision problem of the service provider becomes more complex. This complexity has to be considered when designing the method that facilitates the provider's choice among different portfolios of SLAs. Along with the impact that the availability of the Internet has on service provisioning, more aspects were identified in Section 3.3 that need to be considered in the design of a decision support for SLA establishment. In total, the aspects that require consideration are:

1. the impact of the availability of the Internet on service provisioning
2. the technical composition of services
3. potentially more than one customer
4. the risk of SLA violation and
5. the quantitative measurement of the risk of SLA violation.

This chapter's aim is to consider these requirements in the development of decision support for SLA establishment. More detailed, this chapter focuses on the decision of a service provider as defined in Section 2.3.2, i.e. one is able to provide all of the requested functionality by themself. As the provider is able to provide all functionality, the decision on SLA establishment is concerned exclusively with customer SLAs, and was formulated in Section 2.3.4 as follows:

Which SLA(s) with customers should be active in the next period?

The decision on SLA establishment with customers is illustrated with the help of the scenario on IT outsourcing that was introduced in section 2.3.4. In this scenario, the focus on a provider's decision on SLA establishment is as follows:

Scenario. *[Provider Decisions on SLA Establishment] Consider again the scenario in which a bank wants to achieve "better" prices in institutional trading. Besides choosing a co-location offer from a trading facility (see Section 2.3.4), the bank could opt for installing the exchange software on a Cloud computing instance offered by Amazon Web Services. Now, assume that the bank decides for the latter alternative. In this case, Amazon receives a request for service provisioning from the bank. From Amazon's perspective, this request for service provisioning is only one of a multitude of requests. The outsourcing scenario with a focus on Amazon's situation is depicted in Figure 4.1. In this view of the scenario, Amazon needs to decide which of the service requests will be accepted and which will not.*

In order to identify the SLAs that minimize the risk of SLA violation and hence, which the provider should enter with customers, besides the aspects that were identified in Section 3.3 and named above, the provider's abilities have to be considered. On the

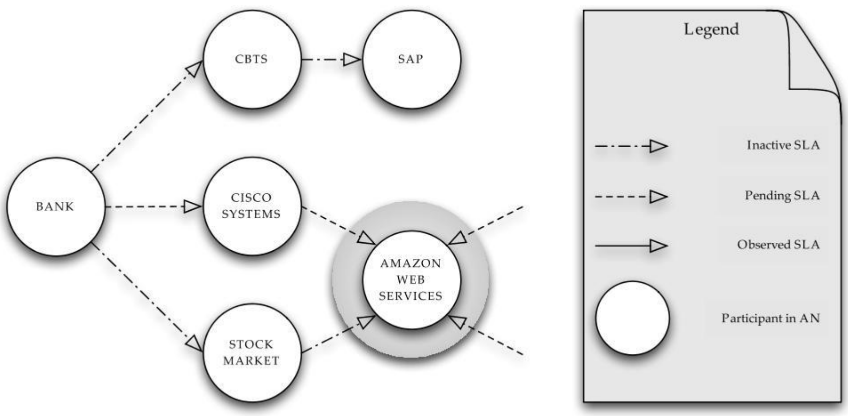

FIGURE 4.1: *Example Agreement Network - Focus on Provider*

one hand, these abilities comprise technical properties like resource availability. On the other hand, the provider's economic preferences of the provider regarding expectations towards profit and risk-attititude need to be included. Consequently, the provider's decision problem on the establishment of SLAs is in the focus of Research Question 2, that was formulated in Section 1.1:

RESEARCH QUESTION 2 ≺PROVIDER'S RISK-MINIMIZING DECISION≻. *How can a provider select the SLAs that minimize the risk of SLA violation while taking their technical properties and economic preferences into account?*

The following sections are dedicated to address Research Question 2. This is achieved by applying the foundations about joint service provisioning and SLA establishment from Chapter 2 along with the economic foundations about decisions under uncertainty and measuring risk from Chapter 3. Therefore, in a first step a method is introduced in Section 4.1 that minimizes the risk of SLA violation for a provider. This method incorporates an adaptation of portfolio selection by Markowitz (1959) and employs the semi-variance as a measure of risk. This approach is especially feasible for services that follow the *Long Tail* phenomenon, as it considers portfolios of SLAs. Furthermore, the application of the semi-variance allows the quantification of downside risk as discussed in Section 3.2.2. Downside risk expresses risk as deviations from an expected value, where only those deviations are taken into account that are *worse* than average. This way, the semi-variance considers only deviations from the expected SLA violation that are *worse* than expected. The exact specification of expected SLA violation is given below. The minimization of SLA violation risk is carried out under constraints that consider provider's resource availability and profit expectations. In a second step, this approach is extended in order to allow a provider to express their attitude towards risk explicitly. This is achieved by maximizing a provider's expected utility, which takes the trade-off between expected profit and the risk of SLA violation into account, which is individually weighted by the provider's coefficient of risk aversion. Section 4.2 provides a thorough introduction to the maximization of the provider's expected utility and discusses in which cases utility maximization is superior to risk minimization. Finally, the time complexity of the methods introduced in Sections 4.1 and 4.2 is analyzed in Section 4.3.

4.1 Provider's Constrained Minimization of SLA Violation Risk

This section introduces a method that addresses a provider's decision problem on the establishment of SLAs from Research Question 2. The method that is presented in this section considers the challenges identified in Section 3.3 regarding the expression of risk and the impact of the availability of the Internet as well as a provider's technical properties and economic preferences that were identified in Section 2.3.2 in the decision on the establishment of SLAs. The method that solves a provider's decision problem is derived along sub questions that are concerned with parts of the main research question.

The decision on the establishment of SLAs is driven by the need of service providers to avoid the violation of SLAs (Penna and Wandresen 2004; Spillner and Schill 2009), as the violation of an SLA leads to a penalty. As was argued in Section 2.2, penalties in their very nature do not have to be monetary but can be translated into a monetary amount, that is, a payment on behalf of the service provider that consequently reduces the achievable profit. Furthermore, the violation of an SLA affects a service provider's reputation negatively and perceived trustworthiness, which can lead to customers' lower willingness to request services from that provider, and results in lower future demand. Due penalties can be expressed in monetary terms that can be included in the provider's decision problem, whereas reputation is very hard to express objectively and the resulting decrease in demand for the service provider's services can barely be expressed explicitly. Therefore, reputation and future demand will be excluded from the optimization problem and the risk of SLA violation is expressed in terms of due penalties. The expression of the risk of SLA violation is the in the focus of sub question 2.1:

Research Question 2.1 ≺Measuring Risk≻. *How can the risk of SLA violation be expressed in order to reflect a service provider's aim to avoid the violation of SLAs?*

Risk in the focus of this work stems from the uncertainty of SLA violation as well as the resulting monetary outcomes and is to be minimized in the provider's decision problem. The expression of risk allows a provider to consider SLAs with more than one customer, and hence meets challenges 3-5 from the beginning of this section. In order to consider challenges 1 and 2, the expression of risk needs to be able to cope with the number of SLAs that result from the *Long Tail* phenomenon. An approach that satifies all of the challenges manifests in an adapatation of portfolio selection (Markowitz 1959). This approach does not consider risk-aversion as defined by von Neumann-Morgenstern utility functions (Von Neumann et al. 1947) as introduced in Section 3.2.3. However, it was shown that the methodology of porfolio selection as introduced by Markowitz (1959) results in most cases in the same choice as maximizing a utility function of a risk-averse decision maker (Kroll et al. 1984), when the number of assets is constrained but not necessarily low.

Besides the challenges from the beginning of this section, the provider's technical properties and economic preferences regarding available resources and expected profit were identified as drivers for the provider's decision about SLA establishment in Section 2.3.2. Technical properties influence the number and quality of provided services, which generate profit for the provider. These constraints are the topic of sub question 2.2.

Research Question 2.2 ≺Formulating Impacting Factors as Constraints≻. *How can the factors that constrain the provider's decision on SLA establishment be included in a provider's decision problem?*

Service providers are limited in service provision by their resource capacities (Kwok and Mohindra 2008). Especially in the context of Cloud computing, the assumption of resource constraints is under constant discussion. Whereas unlimited scalability may hold for computation, there are certain quality of service attributes like storage space that do not scale infinitely (Armbrust et al. 2009, p. 20). Additionally, scaling up and out of computational power is not possible instantly but takes some time for preparation (Armbrust et al. 2009, p. 20). In order to reduce the degrees of freedom in the provider's decision and in order to ensure the comparability of observed service performance, the provider's resources are assumed to stay constant over time.

Assumption 3 (Consistency of Provider's Resources)**.** *Provider's resources do not change over time.*

In case of a Web service provider, Assumption 3 ensures that for instance computational power is not increased or that storage space is not extended. Besides ensuring the comparability of collected data, the consistency of the provider's resources constrains the possibility of scaling up and hence, resources are scarce to a certain degree. A service provider can only offer as many services as their resources allow, such as storage space or bandwidth. The exact amount and specification of SLAs that can be established need to be included in the service provider's optimization problem in order to exclude impossible SLA sets. Another criterium that is to be considered is the profit that can be expected from the provision of services as agreed in an SLA portfolio. By employing SLAs, service providers want to ensure the profitability of service provisions (Berger 2005). Service providers are assumed to require a minimum expected profit (at least a positive profit) from an SLA portfolio in order to include it in the set of feasible solutions. The formulation of resource and profit constraints is introduced in Section 4.1.2.

Having identified the driving factors for the provider's decision and having given a definition of risk in the context of this work, the last sub question of research question 2 is formulated:

RESEARCH QUESTION 2.3 ≺FORMULATION OF THE PROVIDER'S RISK-MINIMIZING DECISION PROBLEM≻. *How is the provider's decision problem formulated in order to minimize the risk of SLA violation and to account for the challenges on decision making (from Section 3.3) as well as the provider's technical and economic properties?*

The method that serves to solve a service provider's decision problem is presented in Section 4.1.3. The method is developed analogously to the method of portfolio selection from Markowitz (1952), where the decision maker starts the identification of efficient portfolios with the portfolio with smallest risk. Analogously, the provider's objective function is set to minimizing the risk of SLA violation. Likewise, the constraints on expected profit and resources from Section 4.1.2 are taken into account. Considering the objective function and the constraints, the service provider's decision model is specified as illustrated in Figure 4.2.

When receiving a request for service provision from a customer, a provider first identifies which SLA portfolios possibly can be established. For each of these portfolios, the

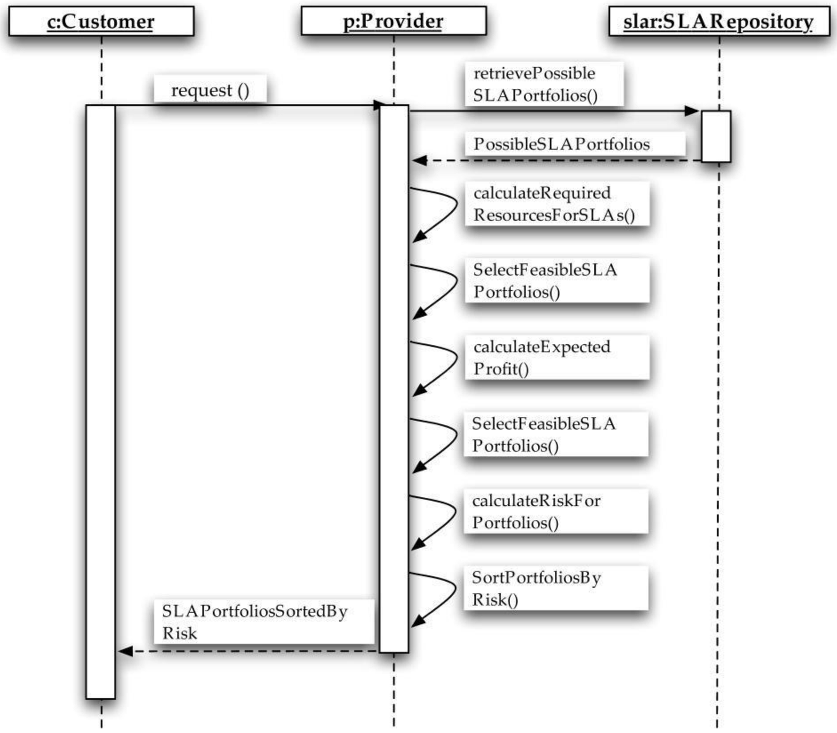

FIGURE 4.2: *Risk-minimizing Portfolio Selection*

resource constraints are verified and only those portfolios that adhere are taken further into consideration. Analogously, the set of SLA portfolios is adjusted after evaluating the expected profit constraint. Finally, for each of the remaining SLA portfolios the incurred risk is calculated and the SLA portfolios are ranked according to their risk.

In the following sections, Research Questions 2.1, 2.2 and 2.3 are answered in depth. Section 4.1.1 introduces a risk measure that is applied for optimizing the provider's SLA portfolio.

4.1.1 Measuring the Risk of SLA Violations

A service provider monitors their own adherence to single a SLA α_i in a portfolio γ of SLAs, where $\gamma = \{\alpha_i, \ldots, \alpha_j\}$, and i, j are identifying indices of services and stores the collected data. The reports that result from the monitoring data contain information on the agreements that were active in each reporting period and information on the provider's adherence to each of the SLAs $\alpha_i \in \gamma$. The information on the provider's adherence is expressed in a degree of violation of each SLA $\lambda^{\alpha_i}_{\gamma,t}$, where $\lambda^{\alpha_i}_{\gamma,t} = 0$ states that the SLA α_i was fully adhered to in period t and $\lambda^{\alpha_i}_{\gamma,t} = 1$ denotes a violation of 100% of the SLA.

Monitoring the adherence of an SLA means monitoring the KPIs that were defined in the SLA. The analytics of this monitoring data for measuring the risk of SLA violation is only possible if the following prerequisites are met:

Assumption 4 (Length of Monitoring Periods)**.** *Each monitoring period is assumed to have equal length.*

This is important as different lengths of monitoring periods would lead to an implicit weighting of monitoring results (longer periods would influence the risk in exactly the same way as shorter periods). Such a task must be controlled by the customer or an agent that represents them.

Assumption 5 (Aggregation of Service Monitoring)**.** *Monitoring reports contain a degree of violation of the monitored SLA, which represents an aggregated form of the degree of violation of the SLA's KPIs.*

The degree of violation is obtained by evaluating each single KPI in the SLA against the specified thresholds (Service Level Objectives, SLOs) and then aggregating the resulting values into one degree $\lambda^{\alpha_i}_{\gamma,t}$ for the whole SLA. The approach presented in this work relies on service monitoring, where each service relates to one SLA, and each SLA is monitored individually. The methodology for aggregating KPI monitoring results into one degree of violation for SLAs can be decided upon by the customer or the intermediary and is therefore not considered in this work. Nevertheless, in the following the most intriguing requirements to the aggregation of monitoring results are briefly introduced.

Monitoring Approach Aggregating monitoring results also needs to include a specification of the exact monitoring approach to be applied considering th SLOs to be observed, the monitoring intervals, reliability of provider's monitoring results, and many other functions (Berger 2005).

Aggregation Function Apart from the applied monitoring approach, the aggregation itself faces several challenges. First, there needs to be a common understanding between the provider and the customer on the meaning of a certain degree of SLA viola-

TABLE 4.1: *Extract of Service Provider's Failure Monitoring*

Time t	SLA α_1	SLA α_2	SLA α_3
1	0	0	
2	0	0	
3		0.5	0.5
4	1	1	
5		0.7	0.4
6	0.78	0.9	0.8
7	0.82	0.9	
8	0.67		0.5
9	?	?	?

tion and its consequences, for example by means of a penalty. Furthermore, the aggregation needs to reflect the customer's preferences for each of the KPIs, for example by assigning weights that state relative importance of each KPI.

Preference Elicitation Elicitation of preferences can be achieved for instance with methodologies like Analytic Hierarchy Process (AHP) (Saaty 2005) or conjoint analysis (Green and Rao 1971). The preferences have to be communicated to and agreed upon with the provider, as the monitoring party is required to perform the aggregation. Even in the rather unrealistic case of complete awareness of preferences, the concrete aggregation function has to be specified. Examples could be a weighted mean or another function that enables aggregation and allows for the reflection of dependencies between KPIs. A sample of a provider's monitoring data including the periods t, the respective agreements α_i for this period and degree of violation of an SLA $\lambda^{\alpha_i}_{\gamma,t}$ is shown in Table 4.1. Here, $\lambda^{\alpha_i}_{\gamma,t}$ corresponds to the entries in any of the cells. For instance, the entries in the first row show that in period $t = 1$ two SLAs, α_1 and α_2 were active ($\gamma' = \{\alpha_1, \alpha_2\}$), and violated both with a degree of 0 ($\lambda^{\alpha_1}_{\{\alpha_1,\alpha_2\},1} = 0, \lambda^{\alpha_2}_{\{\alpha_1,\alpha_2\},1} = 0$). The empty cell for SLA α_3 in period $t = 1$ indicates that α_3 was not active in $t = 1$.

As was argued in Section 4.1, providers' risk in the focus of this work stems from the uncertainty of SLA violation and the resulting monetary consequences. The specification of volatility of a security portfolio from Markowitz (1959), the semi-variance, expresses risk as the deviation from a target value and takes only those deviations into account that are *worse* than average. The literature review in section 3.2.1 identified the semi-variance as the most feasible measure of SLA violation risk due to the emphasis on higher deviations and the expression of a downside risk that considers only those deviations that are *worse* than expected. Therefore Definition 4.1 is deduced:

Definition 4.1 (SLA Portfolio Risk)**.** *The risk that a portfolio incurs on a provider is measured by the semi-variance of the applied penalties.*

The following scenario illustrates the analogy of SLA portfolios and security portfolios.

Scenario. *[Risk of Portfolios] Consider an owner of a portfolio of stocks. In order to calculate the risk associated with a stock, the owner takes the past profits, given a certain purchase price, of the stock into account and computes the average profit of the stock in the past. Especially in stock trading, not only the average profit is of importance but also the uncertainty of the expected profit. Uncertainty in this context is measured by the variance of profit, that is, the average of the squared deviations from the expected profit.*

Translating to Scenario 2, the provider Amazon calculates the risk for the establishment of a portfolio of SLAs. Therefore, the provider collects information on the past performance as measured by the respective degree of SLA violation. Analogously to the past profit of a security portfolio, Amazon calculates the average past degree of violation and weights it by the penalties to be paid. This results in an expected value of penalties. For a service provider, not only the expected degree of SLA violation, but as well the uncertainty of achieving this degree of violation is a major driver for the decision on SLA establishment. Uncertainty translates into the variance of SLA violation and resulting penalties, that is, the variance of expected penalties, which is calculated as the average of squared deviations from expected penalties.

Using the variance of due penalties as a measure of risk implies that a stronger deviation from average penalties is perceived as more uncertain. This definition of risk is refined by the concept of semi-variance as introduced by Markowitz (1959). Only those deviations that are "worse" than average are taken into account.

This measure of risk has been employed in portfolio optimization since the beginning of modern portfolio theory and is especially useful to account for asymmetric probabilistic distributions and downside risk, and was identified to be the most feasible measure of SLA violation risk in Section 3.2.1. For a service provider, this translates into taking only those realizations of SLA violation into account, which are higher than its mean.

The calculation of the semi-variance of penalties of a certain SLA combination γ is conducted in four steps:

1. the sum of penalties $\mu_{\gamma,t}$ for each period is calculated,
2. the mean degree of violation $\overline{\lambda}_{\gamma}^{\alpha_i}$ for a particular SLA α_i belonging to a specified SLA configuration γ is computed,
3. the average penalty of γ is calculated based on $\overline{\lambda}_{\gamma}^{\alpha_i}$ and the penalties μ_{α_i} from each respective SLA $\alpha_i \in \gamma$, and finally,
4. the semi-variance is calculated as the squared deviations of actual penalties from the mean penalties.

The sum of penalties $\mu_{\gamma,t}$ that a service provider would have had to pay in period t, in case the same penalties had been agreed on then, depends on the degree of violation of SLA α_i in the portfolio of SLAs γ in this period and is given by

$$\mu_{\gamma,t} = \sum_{\alpha_i \in \gamma} \lambda_{\gamma,t}^{\alpha_i} \cdot \mu_{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i}). \tag{4.1}$$

Define

$$X(\lambda_{\gamma,t}^{\alpha_i}) = \begin{cases} 1 & \text{if } \lambda_{\gamma,t}^{\alpha_i} \in [0,1], \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

This variable serves to indicate whether the SLA α_i was active ($X(\lambda_{\gamma,t}^{\alpha_i}) = 1$) or not ($X(\lambda_{\gamma,t}^{\alpha_i}) = 0$) in a particular period t. Equation (4.1) relates the degree of violation of an SLA α_i to the penalty μ_{α_i} that was agreed upon and enables the provider to express the amount of money that would have been due in a certain period, assuming that the penalties would have been the same in that period.[1]

The average degree of violation of a particular SLA α_i belonging to a specific SLA portfolio γ is calculated by summing the observed degrees of violation and dividing by the number of observations and results in

$$\overline{\lambda}_{\gamma}^{\alpha_i} = \frac{\sum_t \lambda_{\gamma,t}^{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i})}{|T_\gamma|} \text{ with } X(\lambda_{\gamma,t}^{A}) = 1 \wedge X(\lambda_{\gamma^c,t}^{A}) = 0, \tag{4.3}$$

where $X(\lambda_{\gamma,t}^{A}) = \prod_{\alpha_i \in \gamma} X(\lambda_{\gamma,t}^{\alpha_i})$ and $X(\lambda_{\gamma^c,t}^{A}) = \prod_{\alpha_i \in \gamma^c} X(\lambda_{\gamma^c,t}^{\alpha_i})$, with γ describing a particular SLA portfolio and γ^c comprehending all SLA portfolios but γ. The indicator variables $X(\lambda_{\gamma,t}^{A})$ and $X(\lambda_{\gamma^c,t}^{A})$ serve to select from the observations in the service

[1] Calculating the average degree of failure for each period and multiplying it by the sum of penalties in the considered portfolio would be an alternative approach. However, this would lead to an imprecise notation of actual penalties and would rather be an approximation.

provider's records only the subset of observations, which pertains to a particular portfolio of SLAs γ. Thus, $|T_\gamma|$ is the number of periods in that the particular SLA combination γ appears in the records of the service provider, where T_γ is defined as the set of periods in which the portfolio γ was active $T_\gamma = \left\{ t \mid X(\lambda^A_{\gamma,t}) = 1 \wedge X(\lambda^A_{\gamma^c,t}) = 0 \right\}$.

The average penalty, which the service provider paid in the past for the combination of SLAs γ, is obtained by multiplying the average degree of violation of an SLA α_i with the currently valid penalty μ_{α_i} and summing up for all SLAs $\alpha_i \in \gamma$

$$E(\mu_{\gamma,t}) = \sum_{\alpha_i \in \gamma} \overline{\lambda}^{\alpha_i}_{\gamma} \cdot \mu_{\alpha_i} \cdot X(\lambda^{\alpha_i}_{\gamma,t}). \tag{4.4}$$

The average degree of SLA violation is related to the currently valid penalty and the average penalties for an SLA portfolio can be expressed. An alternative representation of the average penalty would result in the calculation of the overall average degree of violation, that is, the average of the average violations, and multiply it by the sum of penalties. Although being a correct representation of average penalties, the impact of the degree of violation on the resulting penalty is not reflected correctly. Therefore, Equation 4.4 is employed for calculating the average penalty.

The semi-variance of due penalties for the portfolio of SLAs γ can be calculated. It is the risk measure based on which the service provider chooses the risk-minimal combination of SLAs and it is defined as follows:

$$S_E(\mu_{\gamma,t}) = \sum_t q_t \cdot \left(\max\{0; \mu_{\gamma,t} - E(\mu_{\gamma,t})\}\right)^2 , \tag{4.5}$$

where

$$q_t = \frac{1}{|T_\gamma|} . \tag{4.6}$$

The semi-variance of due penalties takes only those periods into account, in which the actual penalties were higher than the expected penalties and consequently, the provider's performance was worse than expected. Hence, the concept of semi-variance as introduced by Markowitz (1959) is reflected by Equation (4.5).

The selection of a security portfolio that was introduced by Markowitz (1959) and extended in Markowitz (1991) is based on the calculation of semi-variance. For a portfolio of stocks, the semi-variance is retrieved as a function of variances of single stocks and covariances of each pair of stocks in the portfolio. Each covariance of two stocks only has to be calculated once and can be reused in each portfolio that the two respective stocks are part of. This calculation is only possible as the profit of one stock does not impact the other stock's profit.[2]

Calculating the semi-variance of SLA violation based on covariances as presented in Markowitz (1959) would imply that the covariance of SLA violations of two SLAs are the same, no matter which other SLAs are active at the same point in time. In practice, however, the provider's performance in SLAs that is reflected in the degree of SLA violation heavily depends on the concurrently active SLAs and with them, concurrently executed services.

Assumption 6 (Dependencies in Service Execution)**.** *The execution of services is subject to dependencies and services affecting one another.*

Dependencies in service execution are reflected in a higher degree of SLA violation for each respective SLA in a certain SLA portfolio indicating that the service executions of the SLA portfolio impact one another in a negative way. Dependencies in service execution can stem from concurrent resource occupation of services, like concurrent database or processor access, interference of respective software executions, joint dependence on third party software or many other aspects[3], which are very special to service provisioning. Hence, calculating covariances of pairs of SLAs without taking other concurrently active SLAs into account, would lead to an imprecise measure of risk and would not reflect the dependencies between all of the service executions correctly. Therefore, this approach the semi-variance of SLA violation for each portfolio, that is all SLAs in a combination, seperately and consequently allows for a correct representation of dependencies among service executions.

The following section introduces the formulation of expected profit and resource contraints that are considered in a provider's decision problem as illustrated in Figure 4.2.

[2] There might be cases, in which the return of one stock, e. g. a big car brand's stock, has an impact on the return of the other stock, e. g. a supplier of building parts. This will only be the case in extreme scenarios and hence, can be omitted here.

[3] A very basic example is running programs on a laptop. When running a simple text-processing program in parallel with an Internet browser in which only one tab is open, it is very unlikely that the performance of either program is influenced. In case instead of the text-processing program a software that renders video data is run, which is very resource intensive, the Internet browser's performance as well as the video rendering software's performance might be influenced negatively.

4.1.2 Formulating Resource and Profit Constraints

In Section 2.3.2, technical properties and economic preferences of a provider were identified as influencing factors in a provider's decision on SLA establishment, which impose constraints on a provider's decision. Thus, besides considering the challenges from Section 3.3 in the objective function, constraints to the decision problem are formulated with respect to the available resources L_p of a provider and a minimum expected profit requirement Π_p.

Service provisioning is constrained by the resources that are available to a provider. For instance in the case of a car repair, the maximum number of simultaneously repaired cars equals the number of repair stations and is additionally constrained by the number of available personnel. The provision of e-Services, Web services or Cloud services is constrained by technical resources like available storage space, share of the CPU that is available, bandwidth and many more. The resources that are required and agreed on in a portfolio of SLAs may not exceed a provider's resources. This is expressed by summing all of the required resources in an SLA portfolio and comparing them to the available resources of the provider in Equation (4.7).

Service providers want to achieve profit from service provisioning (Berger 2005), which is at least non-negative. Therefore, the expected profit for each SLA in a portfolio is calculated in order to compute the portfolio's expected profit by adding the single SLA profits. This is shown in Equation (4.8).

$$\sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda^{\alpha_i}_{\gamma,t}) \leq L_p \tag{4.7}$$

$$E(\pi_{\gamma,t}) \geq \Pi_p, \tag{4.8}$$

where $L_{\alpha_i} = (l^1_{\alpha_i}, \dots, l^M_{\alpha_i})$ describes the set of KPIs $l^m_{\alpha_i}$ agreed upon in SLA α_i. Only those resources that can be mapped directly to the set of resources $L_p = (l^1_p, \dots, l^M_p)$ that are available to the provider are considered in the resource constraint. M denotes the number of KPIs that can be mapped to resources. This mapping is especially difficult in Cloud-based environments. One possible approach has been introduced in Kwok and Mohindra (2008).

The minimum expected profit that the provider wants to achieve is denoted by Π_p, and is set by definition of the provider's preferences. The expected profit from an SLA portfolio γ is denoted by

$$E(\pi_{\gamma,t}) = \sum_{\alpha_i \in A} (f_{\alpha_i}(s_j) - c(s_j)) - E(\mu_{\gamma,t}), \tag{4.9}$$

where $f_{\alpha_i}(s_j)$ denotes the current price that the provider charges for the provision of service s_j in SLA α_i and $c(s_j)$ denotes the provider-specific costs that arise with the execution of service s_j.

Having formulated a provider's technological properties and economic preferences as constraints to objective function from Section 4.1.1, the following section introduces the risk-minimizing decision problem about SLA establishment of a provider.

4.1.3 Provider's Risk-Minimizing Decision about SLA Establishment

Section 4.1.1 introduced a measure of risk that can express the risk of SLA violations. This measure of risk will be applied in the objective function of a provider's decision problem on SLA establishment. Constraints to the decision problem were introduced in Section 4.1.2 in order to include a provider's technical properties and economic preferences in the decision problem. By applying the risk measure from Section 4.1.1 and the constraints from Section 4.1.2 along with a provider's properties the challenges from Section 3.3 are met. In more detail, the decision problem considers the impact of the Internet and related technologies as well as service composition by employing a method that considers portfolios of SLAs, which relate to a multitude of customers. Furthermore, the risk of SLA violation is considered in the decision on SLA establishment by a quantitative expression of SLA violation risk. The risk-minimizing decision problem is formulated as follows:

$$\begin{aligned} \min_{\gamma} \; & S_E(\mu_{\gamma,t}) = \sum_t q_t \cdot (\max\{0; \mu_{\gamma,t} - E(\mu_{\gamma,t})\})^2 \\ \text{subject to} \; & \sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i}) \leq L_p \\ & E(\pi_{\gamma,t}) \geq \Pi_p \end{aligned}$$

In order to illustrate the risk-minimizing decision problem, Scenario 2 is extended by concrete SLA instantiations in order to give a hands-on example.

Example 4.1 (Provider's Risk-minimizing Decision Problem)**.** *Consider the example monitoring data in Table 4.1 to be service provider Amazon's data. For completeness of the numerical example further assume that the SLAs appearing in the table are defined as follows:*

$$\begin{aligned}\alpha_1 \in A :&(p_{AWS}, s_1, \theta_1, B_w = 5, f_{\alpha_1}(s_1) = 4, \mu_{\alpha_1} = 2),\\ \alpha_2 \in A :&(p_{AWS}, s_2, \theta_2, B_w = 10, f_{\alpha_2}(s_2) = 11, \mu_{\alpha_2} = 2),\\ \alpha_3 \in A :&(p_{AWS}, s_1, \theta_3, B_w = 6, f_{\alpha_3}(s_1) = 4, \mu_{\alpha_3} = 3),\end{aligned}$$

where p_{AWS} represents "Amazon Web Services", B_w denotes the bandwidth for the provision of the service. Let the provider p_{AWS} have a maximum bandwidth capacity of 15*, which constitutes the optimization constraint* (4.7)*, as well as internal costs for service execution of* $c(s_1) = 2$, $c(s_2) = 3$*. Further assume that p_{AWS} seeks to attain a miminum expected profit of* 5 *as expected profit requirement* (4.8) *and that penalties can be expressed as monetary amounts (see Section 2.2).*

Following the sequence depicted in Figure 4.2, the risk-minimal SLA portfolio is identified. Exemplarily, consider SLA portfolio $\gamma' = \{\alpha_1, \alpha_2\}$. It can be easily verified that this combination of SLAs satisfies the bandwidth constraint. In the past records (Table 4.1) it appears four times in periods $t \in \{1,2,4,7\}$ such that $q_t = \frac{1}{4}$ in (4.4) *and Equation* (4.5)*. The sum of penalties that the provider had to pay in each of these periods, $\mu_{\gamma',t} = \sum_{\alpha_i \in \gamma'} \lambda^{\alpha_i}_{\gamma',t} \cdot \mu_{\alpha_i}$, is*

$$\begin{aligned}\mu_{\gamma',1} &= 0 \cdot 2 + 0 \cdot 2 &&= 0,\\ \mu_{\gamma',2} &= 0 \cdot 2 + 0 \cdot 2 &&= 0,\\ \mu_{\gamma',4} &= 1 \cdot 2 + 1 \cdot 2 &&= 4,\\ \mu_{\gamma',7} &= 0.82 \cdot 2 + 0.9 \cdot 2 &&= 3.44.\end{aligned}$$

For the mean percentage of failure of the SLAs one obtains

$$\begin{aligned}\overline{\lambda}^{\alpha_1}_{\gamma'} &= \tfrac{0+0+1+0.82}{4} &&= 0.455,\\ \overline{\lambda}^{\alpha_2}_{\gamma'} &= \tfrac{0+0+1+0.9}{4} &&= 0.475.\end{aligned}$$

Hence, the average penalty paid in the past for the SLA combination γ' is

$$E(\mu_{\gamma',t}) = 0.455 \cdot 2 + 0.475 \cdot 2 = 1.86.$$

Expected profit from SLA combination γ', when substituting the above result into (4.9)*, is*

$$E(\pi_{\gamma',t}) = 4 + 11 - 2 - 3 - 1.86 = 8.14. \tag{4.10}$$

γ' meets the expected profit constraint (4.8)*. After substituting into* (4.5) *the semi-variance can be calculated*

$$S_E(\mu_{\gamma',t}) = \frac{1}{4}\left((max\{0;0-1.86\})^2 + (max\{0;0-1.86\})^2 \right.$$
$$\left. +(max\{0;4-1.86\})^2 + (max\{0;3.44-1.86\})^2\right) = 1.769$$

The same procedure is followed for each available SLA portfolio: resource and expected profit constraints are evaluated first and in case these constraints are met, the semi-variance is calculated. Finally, the SLA portfolio that incures the lowest risk is selected as it solves the optimization problem of the service provider.

In the numerical example from Table 4.1, besides the calculated SLA portfolio $\gamma' = \{\alpha_1, \alpha_2\}$, there is no further SLA portfolio meeting the above criteria. Note that for the SLA portfolios $\{\alpha_1\}$, $\{\alpha_2\}$ and $\{\alpha_3\}$ there are no reporting data in the extract in Table 4.1. The SLA portfolios $\{\alpha_2, \alpha_3\}$ and $\{\alpha_1, \alpha_2, \alpha_3\}$, for which data are avaibale, have a combined bandwidth of 16 *and* 21*, respectively, whereby they do not meet the bandwidth constraint and do not need to be considered. The SLA portfolio $\{\alpha_1, \alpha_3\}$ with combined bandwidth of* 11 *meets the bandwidth constraint. However, as one can verify by substituting the values into Equation* (4.9)*, it yields an expected profit of $E(\pi_{\{\alpha_1,\alpha_3\},t}) = 4+4-2-2-(0.67\cdot 2+0.5\cdot 3) = 1.16$ and thus fails the expected profit constraint.*

4.1.4 Summary

This section introduced a method that allows a provider to select the portfolio of SLAs that minimizes their risk of SLA violation and which considers a provider's technical properties and economic preferences as presented in Section 2.3 (Research Question 2.2) as well as the challenges in decision making for SLA establishment that were identified in Section 3.3. These challenges comprise of the impact of the availability of the Internet on service provisioning and service composition. Furthermore, considering and expressing the risk of SLA violations explicitly in the decision of a provider on SLA establishment with a multitude of customers are challenges that need to be addressed. The application of the semi-variance as a measure of risk in analogy to security portfolio selection (Markowitz 1959) allows the expression of downside risk as introduced in Section 3.2.1 and the consideration of the dependencies of service execution (Research Question 2.1). Dependencies in service execution may arise due to concurrent resource access, software intereference or many more aspects as argued before. Thus, the method presented in this section formulates a provider's decision problem by minimizing the risk of SLA violation, which considers the challenges from Section 3.3 under the constraints of a provider's technical properties and economic preferences (Research Question 2.3). In this section, a method that addresses Research Question 2 was presented.

Note that, by construction, the method introduced in Section 4.1 might exclude a portfolio of SLAs based on yielding insufficient expected profit, although this portfolio could incur a very low risk and hence might have been preferred by the service provider. Alternatively, if the minimum level of expected profit is set very low in order to include such portfolios in the feasible set, the portfolio of SLAs that is the result of the optimization might possess the desired low variance, but at the same time yield insufficiently low profit in terms of a provider's preferences.

Example 4.2 (Constraints in a Provider's Risk-minimizing Decision)**.** *To see this at the numerical example in Table 4.1, assume that the profit requirement in the expected profit constraint* (4.8) *was* $\Pi_p = 1$. *The SLA portfolio* $\gamma'' = \{\alpha_1, \alpha_3\}$ *would thus enter the feasible set, and, due to* $S_E(\mu_{\{\alpha_1,\alpha_3\},t}) = 0$ *based on the reporting data, it would be chosen instead of* $\gamma' = \{\alpha_1, \alpha_2\}$. *This example stresses the importance of the chosen profit constraint. In the above case, the provider would have to accept an extremely low profit in order to find an SLA portfolio that incurs lower risk than* γ'. *However, the profit achieved by* γ'' *could leave the provider unsatisfied.*

In general, the method might reject portfolios of SLAs, which combine above average expected profit with only slightly higher variance. To avoid this inefficiency of choice, a provider's preferences should be explicitely taken into account by reflecting the provider's attitude towards risk by means of a trade-off between risk and expected profit. Section 4.2 therefore introduces a method that employs a provider's utility function and thus includes both risk and expected profit in the objective function.

4.2 Provider's Maximization of Expected Utility

The method for solving a provider's decision problem in SLA establishment that was introduced in Section 4.1 enables a service provider to select the SLA portfolio γ^* that minimizes the risk of SLA violation, as measured by the semi-variance of due penalties, subject to attaining at least a certain pre-specified level of expected profit and adhering to resource constraints. The method from Section 4.1 meets the challenges on decision making for SLA establishment that were identified in Section 3.3 and considers a provider's technical properties and economic preferences. However, in order to solve the problem, a threshold for expected profit needs to be determined. SLA portfolios that do not exceed this threshold are rejected. Hence, applying the method from Section 4.1 does not allow for a provider's preferences with respect to a trade-off between risk and profit. In order to specify an approach to Research Question 2 that takes this trade-off into account, the decision problem that was introduced in Section 4.1 needs to be adapted. As was argued in Section 3.2.3, the formulation of a utility function allows to solve a multi-objective decision problem. As utility functions comprise of a trade-off between expected profit and risk by definition, this section will present a method that applies the maximization of expected utility as the objective function of a provider's decision problem about SLA establishment. The first sub question of research question 2 in this section is therefore concerned with a provider's utility-maximizing decision problem.

RESEARCH QUESTION 2.4 ≺PROVIDER'S UTILITY MAXIMIZING DECISION PROBLEM≻. *How is the provider's decision problem formulated in order to maximize the utility from SLA establishment and to account for the challenges on decision making (from Section 3.3) as well as the provider's technical properties?*

In this approach, a provider is assumed to maximize their utility that results from the establishment of SLAs. The utility function expresses a trade-off between expected profit from service provisioning and the risk of SLA violation. As the expected profit from service provisioning is already reflected in a provider's utility, the only remaining constraints consider resources that are available to a provider. Figure 4.3 illustrates the provider's utility-based decision process.

When receiving a request for service provision from a customer, a provider first identifies which SLA portfolios can be established. For each portfolio the resource constraints are verified and only those portfolios that meet the constraints are taken into further consideration. Afterwards, the utility of the respective SLA portfolio is calculated and the portfolios are ranked according to their resulting utility.

The formulation of a provider's utility is in the focus of Sub Question 2.5:

RESEARCH QUESTION 2.5 ≺PROVIDER'S UTILITY FUNCTION≻. *How can a provider's utility that results from service provisioning and the risk of SLA violation be expressed?*

A provider's utility here directly reflects a provider's attitude towards risk by means of

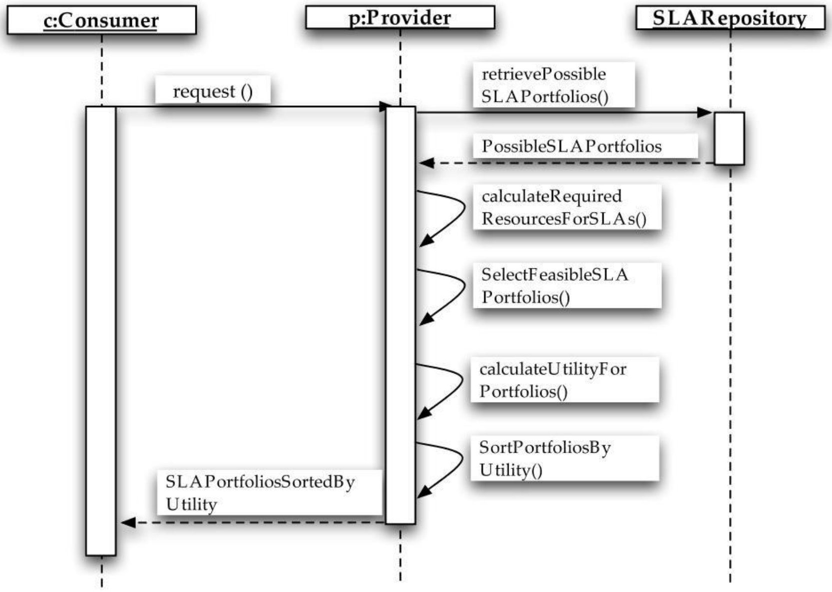

FIGURE 4.3: *Utility-maximizing Portfolio Selection*

a trade-off between marginal incurred risk and the required expected profit for bearing the additional risk. Whereas the approach in Section 4.1 does not specify the decision maker's attitude towards risk explicitely, this section allows providers to express their attitude towards risk directly and hence, focuses on risk-averse decision makers.

Assumption 7 (Risk-averse providers)**.** *Service Providers are assumed to be risk averse.*

The risk-aversion of a provider defines characteristics of the provider's utility function. The following section determines these characteristics and specifies a provider's utility function.

4.2.1 Expressing a Provider's Utility

A provider that wants to determine their utility from service provisioning needs to employ service monitoring in order to retrieve the required information for the calculation of expected profit and risk. The reports that result from the service monitoring are required to meet Assumptions 4 and 5 (equal lengths of all monitoring periods and the ability to aggregate monitoring data, respectively). From the monitoring reports, information about the degree of violation of SLAs can be extracted. Combined with information on prices $f_{\alpha_i}(s_j)$ and penalties μ_{α_i} from SLAs, and the provider's costs for service provision $c(s_j)$, the expected profit can be calculated according to Equation (4.9), where the profit in each period is defined as:

$$\pi_{\gamma,t} = \sum_{\alpha_i \in \gamma} \left(f_{\alpha_i}(s_j) - c(s_j) \right) - \mu_{\gamma,t} \tag{4.11}$$

where $\mu_{\gamma,t}$ denotes the actual penalties as defined in Equation (4.1).

Risk averse decision makers are assumed to possess a von Neumann-Morgenstern utility function $u(\pi_{\gamma,t})$, where $u'(\pi_{\gamma,t}) > 0, u''(\pi_{\gamma,t}) < 0$ (Von Neumann et al. 1947). The expected utility $E(u(\pi_{\gamma,t}))$ can be calculated from historical data and is defined as

$$E(u(\pi_{\gamma,t})) = \sum_t q_t \cdot u(\pi_{\gamma,t}), \tag{4.12}$$

where q_t is the probability for $u(\pi_{\gamma,t})$ and calculated as before (see Equation (4.6)). In theory, it should be employed as the objective function of the service provider. The expected utility is to be maximized under the resource constraints in Equation (4.7). In order to obtain an explicit solution for the optimal combination of SLAs, a particular specification of the utility function has to be applied. Common utility functions in the economic literature are, for example, the quadratic, $u(\pi_{\gamma,t}) = a\pi_{\gamma,t} - \frac{1}{2}(\pi_{\gamma,t})^2$, logarithmic $u(\pi_{\gamma,t}) = \ln \pi_{\gamma,t}$ as well as the power utility functions $u(\pi_{\gamma,t}) = (\pi_{\gamma,t})^{1-b}$ for $b \geq 0$, $b \neq 1$ (Eeckhoudt et al. 2005, p. 21). For $u(\pi_{\gamma,t}) = (\pi_{\gamma,t})^{1-b}$, the coefficient of relative risk aversion $R(w) = -\frac{u''(w)}{u'(w)}w$, where $w = \pi_{\gamma,t}$, is constant and equal to b.

Even with the (rather unrealistic) assumption that the designer of the decision problem is completely aware of the particular form of the service provider's utility, pursuing this approach has a major restriction: The objective function only takes the expected degree of SLA violation into account but does not reflect the volatility of SLA violation directly. Therefore, an alternative representation of expected utility is employed, which holds for the general specification of the utility function $u(\pi_{\gamma,t})$ and contains the first two moments of the profit distribution. For facilitating the representation of profit as a lottery $w + z$, assume that w is some certain income and z is a stochastic variable with a mean of 0 and Variance $Var(z)$ without any specific probability distribution. Furthermore, assume that expected utility of the lottery is $E(u(w + z))$. Applying a Taylor approximation on expected utility, one obtains (Eeckhoudt et al. 2005)

$$E(u(w+z)) \approx u\left(E(w) - \frac{1}{2} \cdot ARA(w) \cdot Var(z) \right). \tag{4.13}$$

$E(w) - \frac{1}{2} \cdot ARA(w) \cdot Var(z)$ denotes the certainty equivalent. The certainty equivalent's utility corresponds to the lottery's expected utility $E(u(w+z))$ (Friedman and Savage 1948) per definition and $ARA(w) = \frac{R(w)}{w} = -\frac{u''(w)}{u'(w)}$ is the coefficient of absolute risk aversion measured at w (Pratt 1964; Arrow 1971).

Setting $w = E(\pi_{\gamma,t})$ and $z = \pi_{\gamma,t} - E(\pi_{\gamma,t})$, z exhibits the following properties

$$E(z) = E(\pi_{\gamma,t} - E(\pi_{\gamma,t})) = 0 \tag{4.14}$$

and

$$\begin{aligned} Var(z) &= Var(\pi_{\gamma,t} - E(\pi_{\gamma,t})) \\ &= E((\pi_{\gamma,t} - E(\pi_{\gamma,t}))^2 - (E(\pi_{\gamma,t} - E(\pi_{\gamma,t})))^2) = Var(\pi_{\gamma,t}). \end{aligned} \tag{4.15}$$

$w + z$ constitutes the lottery of expected profit that meets the above required characteristics.

Substituting w and z into Equation (4.13) results in $E(u(\pi_{\gamma,t})) \approx u(E(\pi_{\gamma,t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma,t})) \cdot Var(\pi_{\gamma,t}))$, where the variance of profit with SLA combination γ equals

$$Var(\pi_{\gamma,t}) = \sum_t q_t \cdot ((\pi_{\gamma,t} - E(\pi_{\gamma,t})))^2. \tag{4.16}$$

Consequently, maximizing the expected utility of the service provider $E(u(\pi_{\gamma,t}))$ is equivalent to maximizing the certainty equivalent $E(\pi_{\gamma,t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma,t})) \cdot Var(\pi_{\gamma,t})$. The certainty equivalent represents the trade-off between risk ($Var(\pi_{\gamma,t})$) and expected profit ($E(\pi_{\gamma,t})$) that is individually weighted by the coefficient of absolute risk aversion ($ARA(E(\pi_{\gamma,t}))$). The value of the coefficient of absolute risk aversion that is usually decreasing in wealth (Eeckhoudt et al. 2005, p. 21) still needs to be identified for the practical application in decision making.

The certainty equivalent is the objective function to be employed in the utility-maximizing decision problem that is introduced in the following section according to Figure 4.3.

4.2.2 Provider's Utility-maximizing Decision Model

The previous section introduced a formulation of a provider's utility that does not require an explicit formulation of a specific utility function for expressing a trade-off between expected profit from service provisioning and the risk of SLA violation. The properties of a provider that were identified to affect a provider's decision on SLA establishment that were presented Section 2.3 need to be considered in a provider's utility-maximizing decision problem. Economic preferences of a provider capture the expected profit from service provisioning and are covered in the objective function that maximizes the utility. Technical properties relate to resources of a provider that constrain service provisioning, which are considered as constraints in the decision problem analogously to the constraints in the risk-minimizing decision problem in Section 4.1.3. Consequently, a provider's utility-maximizing decision problem is formulated as follows:

$$\max_{\gamma} C_E(\pi_{\gamma,t}) = E(\pi_{\gamma,t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma,t})) \cdot Var(\pi_{\gamma,t})$$
$$\text{subject to} \sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda^{\alpha_i}_{\gamma,t}) \leq L_p$$

By calculating the variance and profit for each portfolio, the method caters for the challenges that arise with the availability of the Internet and the composition of services, that is, a high number of SLAs that needs to be considered and the dependencies among service executions as argued in Assumption 6 in Section 4.1.1. Furthermore, the decision of SLA establishment considers a multitude of customers as potential contractual partners as well as the risk of SLA violation. This way, the method that was presented in this section meets the challenges that were raised in Section 3.3 and considers a provider's economic and technical properties. Furthermore, as an extension to the method presented in Section 4.1, it directly allows for a provider's attitude towards risk by formulating a trade-off between the expected profit and SLA violation risk.

In order to illustrate the method, Example 4.1 is adapted.

Example 4.3 (Provider's Utility-maximizing Decision Problem)**.** *Recall that p_{AWS} offers the following SLAs:*

$$\alpha_1 \in A :(p_{AWS}, s_1, \theta_1, B_w = 5, f_{\alpha_1}(s_1) = 4, \mu_{\alpha_1} = 2),$$
$$\alpha_2 \in A :(p_{AWS}, s_2, \theta_2, B_w = 10, f_{\alpha_2}(s_2) = 11, \mu_{\alpha_2} = 2),$$
$$\alpha_3 \in A :(p_{AWS}, s_1, \theta_3, B_w = 6, f_{\alpha_3}(s_1) = 4, \mu_{\alpha_3} = 3),$$

With costs for service execution of $c(s_1) = 2, c(s_2) = 3$, and that p_{AWS} can provide a maximum bandwidth of 15. As was argued in Example 4.1, the SLA combinations $\{\alpha_1,\alpha_2\}$ and $\{\alpha_1,\alpha_3\}$ exhibit reporting data in Table 4.1 and meet the bandwidth constraint. Consider SLA portfolio $\gamma' = \{\alpha_1,\alpha_2\}$ first.

In the past records the SLA portfolio appears four times in periods $t \in \{1,2,4,7\}$ such that $q_t = \frac{1}{4}$ in (4.9). The sum of penalties that the provider had to pay in each of these periods, results, as stated in Example 4.1, in $\mu_{\gamma',1} = 0$, $\mu_{\gamma',2} = 0$, $\mu_{\gamma',4} = 4$, and $\mu_{\gamma',7} = 3.44$. Expected profit from SLA combination γ' is thus substituted into (4.9) and gives

$$E(\pi_{\gamma',t}) = \sum_{\alpha_1,\alpha_2} (f_{\alpha_i}(s_j) - c(s_j)) - E(\mu_{\gamma',t}) \tag{4.17}$$
$$= 4 + 11 - 2 - 3 - 1.86 = 8.14.$$

Substituting the numerical values $E(\pi_{\gamma',t}) = 8.14$ into Equation (4.16) leads to

$$\begin{aligned} Var(\pi_{\gamma',t}) &= q_1 \cdot \left(\pi_{\gamma',1} - E(\pi_{\gamma',t})\right)^2 + q_2 \cdot \left(\pi_{\gamma',2} - E(\pi_{\gamma',t})\right)^2 \\ &+ q_4 \cdot \left(\pi_{\gamma',4} - E(\pi_{\gamma',t})\right)^2 + q_7 \cdot \left(\pi_{\gamma',7} - E(\pi_{\gamma',t})\right)^2 \\ &= \frac{1}{4} \cdot (10 - 8.14)^2 + \frac{1}{4} \cdot (10 - 8.14)^2 \\ &+ \frac{1}{4} \cdot (6 - 8.14)^2 + \frac{1}{4} \cdot (6.56 - 8.14)^2 \\ &= 3.4988. \end{aligned}$$

Empirical studies indicating constant relative risk aversion suggest that a value of 3 for the coefficient of relative risk aversion is reasonable (Szpiro 1988, p. 106). Since the coefficient of absolute risk aversion $ARA(w) = \frac{R(w)}{w}$ can be calculated for a certain value of expected profit $E(\pi_{\gamma,t})$, the following expression results

$$ARA(E(\pi_{\gamma',t})) = \frac{R(E(\pi_{\gamma',t}))}{E(\pi_{\gamma',t})} = \frac{3}{8.14} = 0.36855. \tag{4.18}$$

Thus, the certainty equivalent (CE) of the service provider equals

$$CE_{\gamma'} \approx E(\pi_{\gamma',t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma',t})) \cdot Var(\pi_{\gamma',t}) \qquad (4.19)$$
$$= 8.14 - \frac{1}{2} \cdot 0.36855 \cdot 3.4988 = 7.4953$$

For the second SLA combination $\gamma'' = \{\alpha_1, \alpha_3\}$, $Var(\pi_{\gamma'',t}) = 0$, *and hence, one obtains*

$$CE_{\gamma''} \approx 1.16 - \frac{1}{2} \cdot 2.5862 \cdot 0 = 1.16$$

Based on the comparison of Equations (4.19) *and* (4.20), *the service provider will still choose* $\gamma' = \{\alpha_1, \alpha_2\}$.

4.2.3 Summary

This section introduced the decision problem that takes a risk-averse provider's resource constraints into account while maximizing the provider's utility from providing services (Research Question 2.4). The provider's utility expresses a trade-off between expected profit and risk of SLA violation that is individually weighted by the coefficient of absolute risk aversion (Research Question 2.5). In this vein, the method meets the challenges from Section 3.3 as well as a provider's properties as was argued in Section 4.2.2.

The proposed representation of expected utility is only an approximation, except when the utility function is of the form $u(\pi_{\gamma,t}) = -\frac{e^{-a \cdot \pi_{\gamma,t}}}{a}$, with constant $ARA(\pi_{\gamma,t}) = a$, and profit is normally distributed (Eeckhoudt et al. 2005, pp. 20–21). Its drawback as of any approximation is that it is imprecise. However, in finance theory and practice this is still the most often employed approach.

Example 4.4 (Imprecision of Expected Utility Representation)**.** *To see the magnitude of the imprecision of representing the certainty equivalent in the proposed way, consider Example 4.3 with* $\gamma' = \{\alpha_1, \alpha_2\}$. *Constant relative risk aversion with coefficient* $R(w) = 3$ *is satisfied with the utility function* $u(\pi_{\gamma',t}) = (\pi_{\gamma',t})^{1-b}$ *with* $b = 3$. *The expected utility of the provider choosing SLA combination* γ' *equals*

$$\begin{aligned} E(u(\pi_{\gamma',t})) &= q_1 \cdot u(\pi_{\gamma',1}) + q_2 \cdot u(\pi_{\gamma',2}) + q_4 \cdot u(\pi_{\gamma',4}) + q_7 \cdot u(\pi_{\gamma',7}) \\ &= \frac{1}{4}10^{-2} + \frac{1}{4}10^{-2} + \frac{1}{4}6^{-2} + \frac{1}{4}6.56^{-2} \\ &= 0.017753858. \end{aligned} \qquad (4.20)$$

In contrast, since $E(u(\pi_{\gamma',t})) = u(CE_{\gamma'})$ *where* $CE_{\gamma'}$ *is the expression in* (4.19)

$$\begin{aligned} u(CE_{\gamma'}) &\approx u(E(\pi_{\gamma',t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma',t}) \cdot Var(\pi_{\gamma',t}))) \\ &= 0.017800276 \end{aligned} \tag{4.21}$$

A comparison of the numbers in Equations (4.21) *and* (4.20) *shows that the difference is negligible.*

The risk-minimizing method that was presented in Section 4.1 ignores the trade-off between risk and expected profit from a provider's perspective. Contrarily, the utility maximization in Section 4.2 that employs variance instead of semi-variance treats both, negative and positive deviations of profit from its mean as equally weighted. Therefore, the choice for an approach depends on the decision makers preferences. If the service provider has a strong aversion towards downside risk and the distribution of penalties is asymmetric, they should choose the procedure suggested in Section 4.1. If instead the distribution is symmetric and the provider considers the variation of profit around its mean as relatively small, they should resort to the procedure in Section 4.2. Another aspect that needs to be considered in the choice of the method to be applied. Approximating the maximization of a provider's utility by maximizing the certainty equivalent relies on standard economic literature (Arrow 1971; Pratt 1964; Friedman and Savage 1948; Savage 1972). The certainty equivalent in its very nature is defined exclusively for positive arguments, that is, positive values for profit and expected profit. Obviously, a provider can incur losses from providing services, especially in cases in which a penalty is applied that exceeds the difference between the agreed price for service provision and a provider's cost for service provision. In these cases, the certainty equivalent cannot be applied for the decision about SLA establishment. Different approaches, for instance an adaptation of the certainty equivalent that defines a piecewise defined function, are required to close this gap. Identifying an approach that closes this gap will be subject to future research.

4.3 Complexity Considerations and Implications

This section presents an analysis of the decision methods that were introduced in Sections 4.1 and 4.2. These methods enable a service provider to choose either a risk-minimal SLA portfolio that satisfies a minimal profit constraint or an SLA portfolio that maximizes expected utility expressed as a trade-off between risk and profit. In an environment of services that are provided electronically or even completely automated, e.g. Web services, the solution time of a method and hence its computational tractability plays an important role. In order to determine the runtime properties of the presented models, the complexity class that the models belong to is identified.

In order to find the risk-minimal/utility-maximizing SLA portfolio, a set of SLAs A that is of the size $|A|$, where SLAs include the respective prices, penalties and KPIs L_{α_i} that can be matched to the provider's resource constraints are handed over as an input to the model. Further, input is a matrix of monitoring data Λ of T periods for the $|A|$ SLAs (see Table 4.1). The optimization problems have a discrete set of feasible solutions and their goal is to find the "optimal" solution. One approach for solving the presented models is to calculate the risk/utility for each combination of SLAs and then to search for the risk-minimal/utility-maximizing solution. With a growing number of SLAs, exhaustive search becomes infeasible from a computational point of view. Solution concepts that are applied to linear optimization, e.g. the Dijkstra algorithm, cannot be applied to the presented models as the risk-/utility-values associated with combinations are not monotonouos in the number of SLAs.[4] Hence, the model presented in this thesis can be regarded to be in NP like many other optimization problems (Papadimitriou and Steiglitz 1998). A more detailed analysis of the computational complexity is carried out in the following. In order to determine the complexity in *Big-o-Notation*, a worst case analysis is carried out that analyzes the runtime properties of the method that solves a provider's decision problem as illustrated in Algorithm 4.1. This worst case analysis identifies the highest possible runtime and hence considers exhaustive search.

Algorithm 4.1 Identify the Risk-Minimal SLA portfolio

Require: A, L_{α_i} and Λ
$\Gamma \leftarrow$ retrievePossibleSLAPortfolios(A)
for all $\gamma \in \Gamma$ **do**
  $L_\gamma \leftarrow$ calculateRequiredResourcesForSLAs(γ, L_{α_i})
  if $L_\gamma \leq L_p$ **then**
    $\Pi_\gamma \leftarrow$ calculateExpectedProfit(γ, Λ)
    if $\Pi_\gamma \geq \Pi_p$ **then**
      $S_E(\lambda_{\gamma,t}) \leftarrow$ calculateRiskForPortfolios(γ, Λ)
      $\Gamma^R \leftarrow \Gamma^R + [\gamma, S_E(\lambda_{\gamma,t})]$
$\gamma^* \leftarrow \text{argmin}_{\gamma \in \Gamma^R} S_E(\lambda_{\gamma,t})$
return γ^*

For determining the computational complexity of Algorithm 4.1 based on the input

[4] Risk that is associated with the SLA portfolio α_1, α_2 might be higher than the risk of the SLA α_1 alone, whereas the risk of SLA portfolio $\alpha_1, \alpha_2, \alpha_3$ might be lower than the risk of α_1, α_2. Hence, risk can, but does not have to be monotonous.

specified above, each function is analyzed seperatly.

retrievePossibleSLAPortfolios(A) allows the provider to query storage, i. e. a repository or database, for every set of SLAs that has been established previously at least once. As this is a simple request, the time complexity is in O(| A |).

calculateRequiredResourcesForSLAs(γ, L_{α_i}) computes the resource requirements for a given SLA portfolio γ based on the agreed KPIs L_{α_i} as stated in Equation (4.7). The resulting required resources are then compared to those resources that are available to the provider. The time complexity of the evaluation of the resource constraints depends on the number of SLAs $\alpha_i \in A$, as well as on the number of KPIs $M = |L_{\alpha_i}|$. The computational complexity can be specified as $O(|A|) \cdot O(M) = O(|A| \cdot M)$.

calculateExpectedProfit(γ, Λ) specifies the computation of the expected profit constraint as stated in Equation (4.8). Therefore, the expected profit of an SLA portfolio γ is calculated with respect to the provider's monitoring data in Λ and compared to the profit Π_p that the provider expects to achieve. This computation depends on the number of SLAs $|A|$ and on the number of observed periods T in Λ. It has a time complexity of $O(|A|)$ for the calculation of price minus costs for each contract and a complexity of $O((|A|) \cdot (T))$ for the calculation of the expected penalties that depends on the number of observed periods as well as on the number of monitored SLAs. Hence, a time complexity of $O(|A|) + O(|A| \cdot T) = O(|A| \cdot T)$ results.

calculateRiskForPortfolios(γ, Λ) finally determines the risk that is associated with an SLA portfolio γ based on the provider's monitoring data in Λ as stated in Equation (4.5). Two factors influence its computational complexity; the first one is the number of periods T that have been observed. The second one is the number of SLAs $|A|$. The computation of the semi-variance has an exponential time complexity $O(T) \cdot (O(|A|) + O(|A| \cdot T)) = O(T^2 \cdot |A|)$.

Taking into consideration that this calculation has to be performed for each possible SLA portfolio $\gamma \in \Gamma$, the overall complexity of the model becomes apparent. Assuming that there are $|A|$ SLAs $\alpha \in A$ and that each of the SLAs can be established only once, then the power-set of A, $\Gamma = \mathcal{P}(A)$, would reflect the set of possible SLA portfolios and would have a size of $2^{|A|} - 1$ when the empty set is excluded. Relaxing the assumption that each SLA can only be established once would even increase the number of possible SLA portfolios and with it the number of calculations that have to be run. This leads to a time complexity for determining the risk of SLA portfolios of

$$\begin{aligned} O(\cdot) &= O(2^{|A|}) \cdot O(T^2 \cdot |A|) \cdot [O(|A| \cdot T) + O(|A| \cdot M)] \\ &= O(2^{|A|} \cdot T^3 \cdot |A|^2) + O(2^{|A|} \cdot T^2 \cdot |A|^2 \cdot M \\ &= O(2^{|A|} \cdot T^3 \cdot |A|^2 \cdot M) \\ &= O(2^{|A|} \cdot T^3 \cdot M). \end{aligned}$$

$\mathbf{argmin}_{\gamma \in \Gamma^R} S_E(\lambda_{\gamma,t})$ identifies the SLA portfolio γ^* that minimizes risk while adhering to the provider's constraints. An efficient search can be assumed to in $O(log|\Gamma^R|)$. The overall time complexity of Algorithm 4.1 hence results in $O(log|\Gamma^R| \cdot 2^{|A|} \cdot T^3 \cdot M)$.

The complexity of the model presented in Section 4.2 is determined analogously. As the result does not differ significantly, it will be omitted here.

The computational complexity of Algorithm 4.1 is mainly influenced by the size of Γ. Hence, the runtime properties of Algorithm 4.1 can be influenced positively by restricting the potential number of SLA portfolios. Three aspects support the feasibility of this approach. First, the provider itself restricts the number of different SLAs, e. g. a predefined number of quality classes per service, that they offer to customers in order to make sure that enough comparable monitoring data is available for each of the SLAs. Under the assumption that there is only one concurrent instance of each predefined SLA, the number of SLAs $|A|$ would be kept as small as possible that way. Relaxing the assumption of only one concurrent instance for each predefined SLA increases the number of SLAs $|A|$ and hence, the size of Γ. Second, a provider is likely to specialize on a certain kind of services to provide and thereby restricts the number of offered services. By restricting the number of offered services as well as the number of quality classes per SLA, the number of possible SLA portfolios can be kept to a minimum level. Third, the provider could refrain from solving the decision problem for all possible SLA portfolios. The provider's decision is triggered by customers' requests for service provision. Hence, the service provider's decision may be restricted to an evaluation of the current situation as compared to the situation when accepting the request for service provision. This approach would lead to only two calculations of semi-variance and consequently would reduce the complexity massively.

Applying the calculation of semi-variance as presented in Markowitz (1959) would result in a lower complexity as the pairwise covariances of SLA violations could be reused in each SLA portfolio that the respective pair of SLAs is part of. Nevertheless, the application of covariances as in Markowitz (1959) does not account for dependencies in service execution. In order to account for these dependencies, covariances would have to be calculated while considering the portfolio of SLAs. This would mean calculating the covariances for pairs of SLAs seperately for each portfolio of SLAs, which would result in the same complexity as the approach applied in this work.

Chapter 5

Intermediary Decisions

"Web service providers are interconnecting their offerings in unforeseen ways, giving rise to Web service ecosystems"

(Barros and Dumas 2006)

IN Section 2.3.2, two types of service providers were introduced: service providers (see Definition 2.7) that provide all requested functionality by themselves and service intermediaries (see Definition 2.8) that provide a combination or composition of services from a multitude of providers to a customer. While Chapter 4 focused on SLA establishment decisions that are made by a service provider, in this chapter a method for making decisions on SLA establishment by an intermediary will be presented.

In the design of the method that solves the intermediary's decision problem, the challenges on decision making for SLA establishment that were raised in Section 3.3 need to be considered. These challenges reflect the impact of the availability of the Internet on service provisioning and composition, the presence of a multitude of customers and the consideration and expression of SLA violation risk in the decision. Besides these challenges, a service intermediary incorporates particular properties which limit the provisioning of services (see Section 2.3.2) like a service provider. An intermediary faces resource constraints in a similar manner to a service provider that limit service provisioning and wants to make profit by providing services (Berger 2005), after the costs of procuring services from providers are considered. Consequently, an intermediary's decision problem resorts to the question that was raised in Section 2.3.2:

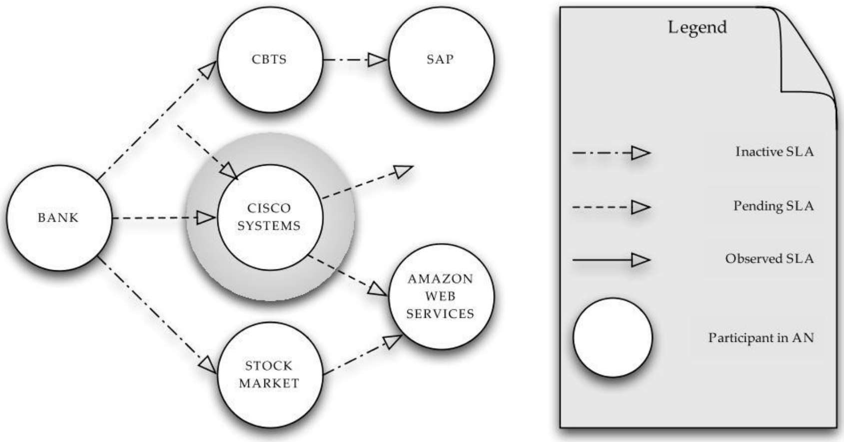

FIGURE 5.1: *Example Agreement Network - Focus on Intermediary*

Which SLA(s) with customers and supplying providers should be active in the next period?

This decision is illustrated with the help of an extension of the IT outsourcing scenario that was introduced in Section 2.3.4. In this chapter, a focus on an intermediary's decision fosters the following adaptation of the scenario:

Scenario. *[Intermediary Decisions in Agreement Networks] The provider Cisco Systems acts as an intermediary, which buys services from other providers, e. g. a compute instance from Amazon (see outgoing arrows in Figure 5.1). Cisco Systems offers the service that results from bundling the compute instance with Cisco Systems' network connection service to customers. As depicted in Figure 5.1, Cisco Systems receives requests from multiple customers (see incoming arrows) and requests services from multiple providers (see outgoing arrows). In this situation, Cisco Systems needs to decide which requests from customers to accept and which SLAs to request from other providers.*

As argued previously, the method that solves an intermediary's decision problem on SLA establishment needs to consider the challenges that were raised in Section 3.3 as well as an intermediary's properties. In this vein, the intermediary's decision problem exhibits parallels to the provider's decision problem with the extension of taking the additional procurement of services into account. Thus, in analogy to a provider's decision problem that was covered by Research Question 2, the intermediary's decision problem is the focus of Research Question 3 that was formulated in Section 1.1 as follows:

Research Question 3 ≺Intermediary's risk-minimizing Decision≻. *How can an intermediary select the SLAs with customers and supplying providers that minimize the risk of SLA violation while taking their technical properties and economic preferences into account?*

In the following sections, solutions for Research Question 3 are presented by applying the economic foundations from Chapter 3 and the foundations of joint service provisioning and service composition from Chapter 2 with a special focus on an intermediary's role according to Definition 2.8. Therefore, in a first step, an extension of the risk-minimizing decision of a provider is presented which allows an intermediary to select an SLA portfolio with customers and an SLA portfolio with providers of services that minimize the risk of SLA violations. Like the method in Section 4.1, the method applies an adaptation of portfolio selection by Markowitz (1959) and employs the semi-variance. The semi-variance was already identified to be especially feasible for the decision on SLA establishment in Section 3.2.1 and 4.1.1 as it allows the consideration of portfolios of SLAs and the downside risk of SLA violation, which considers only deviations from the expected value that are considered *worse* than expected. In a second step, the approach is extended in order to allow for an intermediary's preferences by means of a trade-off between expected profit from service provisioning and the risk of SLA violation. The maximization of an intermediary's expected utility, that is the trade-off between risk and expected profit, is discussed in detail in Section 5.2.

5.1 Intermediary's Constrained Minimization of SLA Violation Risk

The solution to a provider's SLA establishment decision problem that minimizes the risk of SLA violation was presented in Section 4.1 will be extended in this section to capture an intermediary's decision on customer as well as supplier SLAs.

Analogously to service providers, intermediaries seek to adhere to the SLAs that they establish with customers in order to avoid the payment of penalties (Penna and Wandresen 2004; Spillner and Schill 2009). The intermediary's adherence to SLAs does not only depend on their own performance as in a provider's case, but also on the performance of supplying providers. This implies that the method from Section 4.1 needs to be extended to capture SLA violation risk for customer as well as provider SLAs. As a first part of the approach to address Research Question 3, the expression of SLA violation risk in an intermediary's decision problem is in the focus of Sub Question 3.1:

RESEARCH QUESTION 3.1 ≺MEASURING RISK≻. *In an intermediary's decision on SLA establishment, how can the risk of SLA violation for the portfolio of SLAs that are offered to customers be extended to also capture the risk of SLA violation of procured SLAs?*

The measure of SLA violation risk that addresses Sub Question 3.1 needs to meet the challenges that were raised in Section 3.3. Besides the consideration and expression of risk in the decision on SLA establishment, the challenges tackle the impact of the availability of the Internet on service provisioning, the presence of more than one customer and aspects of service composition.

Approaching Sub Question 3.1 results in the question of how to alter the expression of SLA violation risk from Section 4.1 so that the risk of violations of SLAs with providers is captured in the intermediary's decision problem. The decisions for a customer SLA portfolio and a supplied SLA portfolio are not necessarily made at the same time, as the decision for supplying services can relate to a multitude of different offered services, e. g. infrastructure services, where storage space may be used for different services. Therefore, the decision on SLA establishment with customers and providers is taken into account seperatly in the following.

In order to formulate an intermediary's decision problem on SLA establishment, besides considering the challenges from Section 3.3, an intermediary's properties that influence the decision about SLA establishment need to be considered. In Section 2.3.2 it was identified that the technical properties of an intermediary (i.e. resources available) affect the SLA establishment decision. Nevertheless, intermediaries incorporate the possibility of procuring services, which may result in extended resources. Consequently, the intermediary needs to consider all resources that are available to them including procured resources. In order to account for an intermediary's expenses, the expected profit of service provision captures both the expected profit of the SLA portfolio to be established with customers as well as the expected costs of the procured SLA portfolio. Including an intermediary's technical properties and economic preferences is in the focus of Sub Question 3.2:

RESEARCH QUESTION 3.2 ≺INFLUENCING FACTORS≻. *How can the factors that constrain an intermediary's decision on SLA establishment be included in the decision problem?*

With the identification of a suitable measure of the risk of SLA violation for customer SLAs as well as for SLAs with providers that answers the challenges in Section 3.3 and the specification of constraints that affect an intermediary's choice of SLAs, the foundations are laid to formulate an intermediary's decision problem for SLA establishment. The intermediary's decision on SLAs with customers and providers is the focus of Sub Question 3.3:

RESEARCH QUESTION 3.3 ≺FORMULATION OF THE INTERMEDIARY'S RISK-MINIMIZING DECISION PROBLEM≻. *How is an intermediary's decision problem formulated that minimizes the risk of violation of SLAs with customers and providers and that meets the challenges from Section 3.3 and considers an intermediary's technical and economic properties?*

The risk of SLA violation translates into uncertainty about expected profit (Bonini 1975). Therefore, the intermediary aims at minimizing the risk of customer SLA violation. Consequently, the objective function minimizes the intermediary's risk. Additionally, constraints on expected profit and resources as well as the risk of supplying providers violating SLAs have to be taken into account. Figure 5.2 illustrates the decision process for SLAs with customers and supplying providers if made at the same point in time.

Customers that want to purchase a customized service send an SLA offer to the intermediary. The intermediary, that offers services to customers, selects the feasible SLA portfolios with respect to resource requirements and afterwards calculates the optimal SLA portfolio that should be established with customers. After having decided on the risk-minimal SLA portfolio, the intermediary looks up the suppliers from which they can purchase the required services and checks which ones are available. Based on the expected profit that the intermediary wants to gain, feasible provider SLA portfolios are chosen and the corresponding risk is calculated. According to this risk, a ranked list of the provider SLA portfolios is compiled. If there is only one option for an offered SLA (as well as for supplied SLAs), this one is the risk-minimal one.

The intermediary sends an SLA offer to the set of service providers that incure the lowest risk of SLA violation. Note that these offers are merely regarded as a request for proposal, which the intermediary sends to the providers. This means that SLAs will only be established if all of the providers are willing to establish an SLA and the intermediary sends a final confirmation of SLA establishment. Each provider that is requested for service provisioning may in turn calculate the risk of establishing the SLA as introduced in Section 4.1 and notifies the intermediary about their willingness to establish an SLA. Concerning the intermediary's next action, two cases have to be regarded. First, if each of the requested providers accepts the respective SLA offer, the intermediary notifies the customer and the providers on the acceptance of the SLA and all of the SLAs, with customers and providers, are established, and the services are executed and monitored. In the second case, at least one of the requested providers denies establishment of an SLA with the intermediary. In this case, the intermediary notifies all of the requested providers about the cancellation of the requests and attempts to establish the next SLA portfolio in the list. This procedure is repeated until either a feasible SLA portfolio with supplying providers is found or an average violation threshold, that denotes the maximum average degree of SLA violation that the intermediary is willing to accept, is exceeded. In the case that no supplying SLA portfolio is found, the intermediary declines the customer's offer. Alternatively, if the intermediary makes the decisions for SLAs with customers seperately from the decisions with supplying providers the SLAs with customers are established in a first step, without sending offers to potential supplying providers.

Note that the establishment of an SLA in this context may be the establishment of an actually new instance of an SLA but can as well be the establishment of a second instance of an SLA. The methodology, which is presented afterwards can deal with both cases as the establishment of a second instance of an already established SLA can be handled exactly like the instantiation of a new SLA.

The following sections illustrate the building blocks of the intermediary's decision on SLAs. For the decision on customer SLAs, the expression of SLA violation risk from Section 4.1 is applied. In order to capture the risk of SLA violation of procured SLAs, an adaptation of this expression is employed, which is introduced in Section 5.1.2. The complete decision problem that considers technical and economic properties of an intermediary is discussed and illustrated in Section 5.1.3.

5.1.1 Risk of Violating SLAs with Customers

Service intermediaries monitor their adherence to each customer SLA in the context of the SLA portfolio γ. This monitoring and the resulting reports are assumed to adhere to Assumptions 4 and 5 from Section 4.1 that are concerned with an equal length of monitoring periods and the availability of an aggregration method for monitoring observations. These assumptions ensure the availability of monitoring reports that contain one degree of SLA violation $\lambda^{\alpha_i}_{\gamma,t}$ for each SLA as defined in Section 4.1.

Considering the intermediary's decision on SLA portfolios that are established with

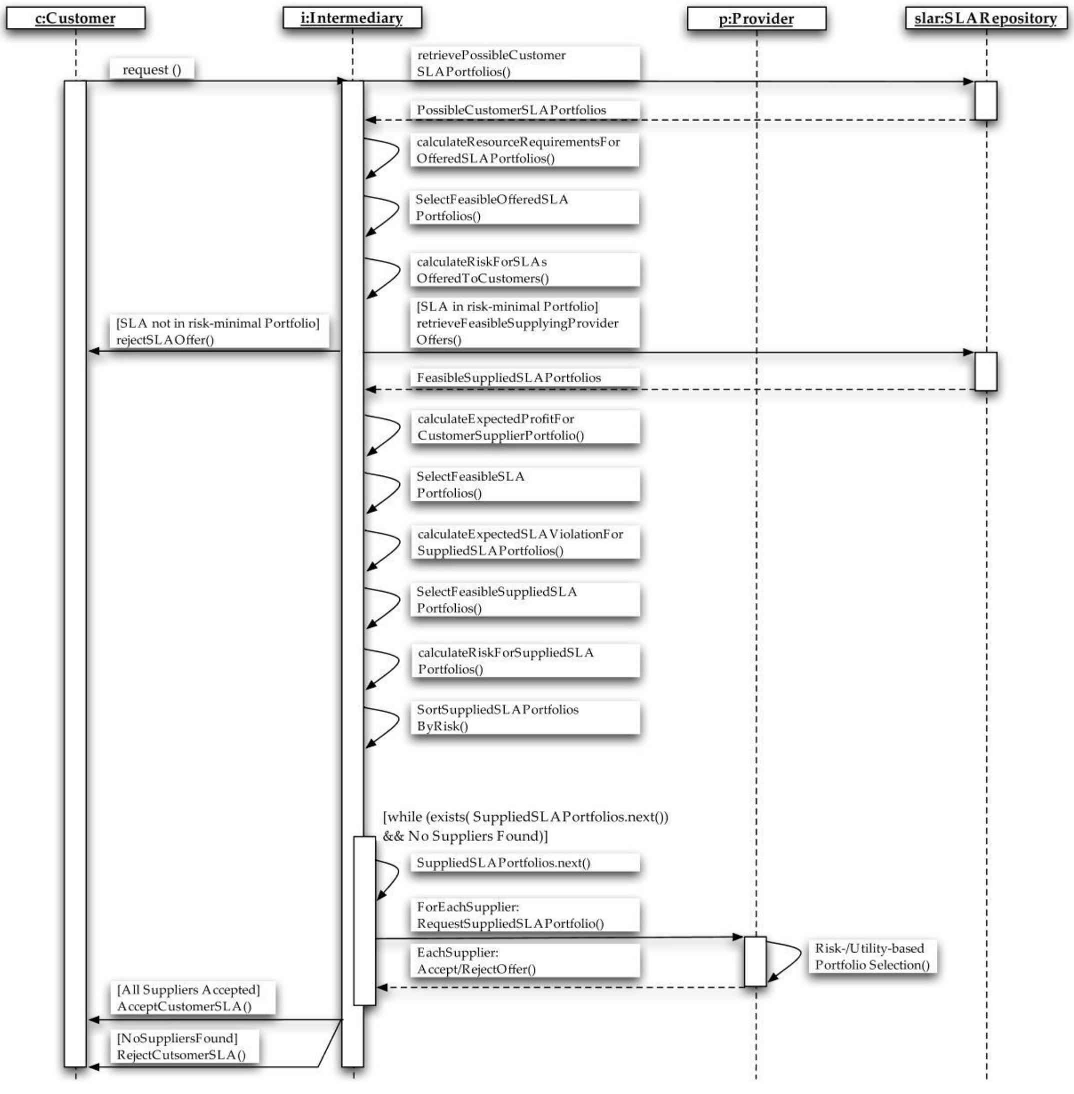

FIGURE 5.2: *Intermediary's Risk-minimizing Portfolio Selection*

customers, the intermediary is confronted with a situation that is similar to that of a provider as introduced in Section 4.1. The intermediary needs to decide on the SLA portfolio that is established with customers that minimizes the risk of SLA violation. Hence, the risk-measure that was presented in Section 4.1.1 is applied as a proxy for the risk of SLA violation in customer SLAs.

Consequently, the calculation of the semi-variance of penalties of a customer SLA portfolio γ is conducted in the same four steps:

1. the sum of penalties $\mu_{\gamma,t}$ for each period is calculated,
2. the mean degree of SLA violation $\overline{\lambda}_{\gamma}^{\alpha_i}$ for a particular SLA α_i belonging to a specified SLA portfolio γ is computed,
3. the average penalty of γ that ι was supposed to pay is calculated based on $\overline{\lambda}_{\gamma}^{\alpha_i}$ and, finally,
4. the semi-variance is calculated as the sum of squared deviations of actual penalties from the mean penalties.

Analogously to Section 4.1.1, these values are calculated as

$$\mu_{\gamma,t} = \sum_{\alpha_i \in \gamma} \lambda_{\gamma,t}^{\alpha_i} \cdot \mu_{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i}), \tag{5.1}$$

$$\overline{\lambda}_{\gamma}^{\alpha_i} = \frac{\sum_t \lambda_{\gamma,t}^{\alpha_i}}{|T_\gamma|} \cdot X(\lambda_{\gamma,t}^{\alpha_i}) \text{ with } X(\lambda_{\gamma,t}^{A}) = 1 \wedge X(\lambda_{\gamma^c,t}^{A}) = 0, \tag{5.2}$$

$$E(\mu_{\gamma,t}) = \sum_{\alpha_i \in \gamma} \overline{\lambda}_{\gamma}^{\alpha_i} \cdot \mu_{\alpha_i}, \tag{5.3}$$

$$S_E(\mu_{\gamma,t}) = \sum_t q_t \cdot \left(\max\{0; \mu_{\gamma,t} - E(\mu_{\gamma,t})\}\right)^2, \tag{5.4}$$

where $X(\lambda_{\gamma,t}^{A})$, q_t, $E(\mu_{\gamma,t})$, $|T_\gamma|$, and $S_E(\mu_{\gamma,t})$ are defined as in Section 4.1.1.

Unlike the provider's decision from Section 4.1, the intermediary's decision has to account for their own risk of SLA violation as well as for the risk of supplying providers that violate established SLAs. In order to account for this risk of suppliers' SLA violation, the intermediary monitors the supplying providers' performance. The following section introduces a measure of risk for SLA violation of supplied SLA portfolios.

TABLE 5.1: *Extract of ι's partner monitoring*

Time t	SLA β_1	SLA β_2	SLA β_3
1	0.5	0.3	
2	0.6		0
3	0.4	0.1	
4	0.5	0.3	
5	0.8		0.4

5.1.2 Risk of Provider SLA Violations

Having identified the risk-minimal SLA portfolio γ^* to establish with customers, an intermediary needs to find the risk-minimal portfolio of SLAs δ supplied by providers. The application of the semi-variance was identified to be the most feasible expression of risk in the context of SLA violation in Sections 3.2.2 and 4.1. In order to apply the semi-variance as a measure of SLA violation risk, observations of providers' performance in SLAs from the past are required. Table 5.1 exemplifies the structure of ι's monitoring. For each period t, provider p's degree of SLA violation $\lambda_{\delta,t}^{\beta_i}$ in SLA β_i is reported. It is to be read analogously to Table 4.1.

Based on this monitoring, the risk associated with a supplied SLA portfolio δ is calculated as an adaptation of the semi-variance of expected SLA violation similar to Equation (4.5). ι's risk of violating SLAs with a customer can be measured monetarily by means of penalties that ι is obliged to pay. Contrarily, it is the SLA violation of supplying providers that mainly influences an intermediary's performance. The penalties that result from supplying providers that violate SLAs exist in order to compensate the intermediary and hence, have a positive impact and do not constitute a risk in the context of the semi-variance.[1] Therefore, exclusively the degree of SLA violation is taken into account.

The calculation of the semi-variance of SLA violation with respect to the SLA portfolio δ is conducted in four steps:

1. the average degree of SLA violation $\overline{\lambda}_{\delta,t}$ for each period is calculated,
2. the mean degree of SLA violation $\overline{\lambda}_{\delta}^{\beta_i}$ for a particular SLA β_i that is part of SLA portfolio δ is computed,

[1]Nevertheless, the supplying providers' penalties are taken into account for the calculation of the total expected profit that constitutes an intermediary's profit constraint. It is introduced in Section 5.1.3 and includes costs for service consumption and expected penalties that ι receives from contracting partners.

3. the average degree of SLA violation of δ is calculated from all $\overline{\lambda}_{\delta}^{\beta_i}$ in the portfolio and,
4. the risk of δ is calculated as the squared deviations of the actual average degree of SLA violation from the mean degree of SLA violation, which are worse than average.

The average degree of SLA violation $\overline{\lambda}_{\delta,t}$ that a service intermediary experienced with portfolio δ in period t depends on the degree of violation of SLA β_i in SLA portfolio δ in this period and is given by the sum of observed SLA violations of SLAs β_i in period t divided by the number of SLAs in δ

$$\overline{\lambda}_{\delta,t} = \frac{\sum_{\beta_i \in \delta} \lambda_{\delta,t}^{\beta_i} \cdot X(\lambda_{\delta,t}^{\beta_i})}{|\delta|}, \tag{5.5}$$

where $|\delta|$ defines the number of SLAs in δ and $X(\lambda_{\delta,t}^{\beta_i})$ serves as indicator variable for SLA β_i being active in period t as part of portfolio δ.

The mean degree of SLA violation for a particular SLA β_i that is part of SLA portfolio δ is calculated as the sum of observed SLA violations of SLA β_i divided by the number of observations in the context of portfolio δ

$$\overline{\lambda}_{\delta}^{\beta_i} = \frac{\sum_t \lambda_{\delta,t}^{\beta_i}}{|T_\delta|}, \tag{5.6}$$

where $|T_\delta|$ is the number of periods in that the particular SLA portfolio δ appears in the records of the service intermediary analogously to the definition of $|T_\gamma|$ in Section 5.1.

For the average degree of SLA violation, which the service intermediary experienced in the past for the portfolio of SLAs δ, one obtains

$$E(\lambda_{\delta,t}) = \sum_{\beta_i \in \delta} \frac{\overline{\lambda}_{\delta}^{\beta_i}}{|\delta|}. \tag{5.7}$$

The average degree of SLA violation of portfolio δ describes the sum of average SLA violations for each of the SLAs divided by the number of SLAs in the portfolio, thus stating the total average of the degree of SLA violation for the portfolio. Based on this, the risk for the portfolio δ can be calculated.

Analogously to the approach that was presented in Section 4.1, the risk of SLA violation is calculated as the squared deviations of the actual average violation of the SLAs in δ from the total mean violation of the portfolio δ, which are worse than average. In contrast to the approach in Section 4.1, the deviations are not weighted by the penalties that are associated with the respective degree of violation. This is done in order to reflect the fact that it is the violation of the SLA with a provider itself that induces the risk on the intermediary, not the monetary outcome. On the contrary, considering the penalties that the provider was to pay in case of an SLA violation would distort the decision, as higher violations would lead to higher penalties, which in this case means higher profit.

In contrast to Michalk and Blau (2010), where a piece-wise defined exponential function is employed as a measure of risk, this approach uses squaring in order to weight deviations from the mean degree of violation. Both approaches are suited for the expression of risk of SLA violation in the intermediary's decision. The exponential function facilitates the differentiation of different shapes for the decision maker's attitude towards risk, but requires to identify the risk attitude beforehand. For reasons of comparability to the decision in the provider's case and for avoiding the assumption of a particular degree of risk aversion, in this work, the approach of squaring the deviations is chosen. Therefore, the risk incurred by different SLA portfolios δ is calculated as follows:

$$S_E(\lambda_{\delta,t}) = \sum_t \rho_t \cdot \left(\max\{0; \overline{\lambda}_{\delta,t} - E(\lambda_{\delta,t})\}\right)^2 , \tag{5.8}$$

where

$$\rho_t = \frac{1}{|T_\delta|}$$

states the relative frequency of occurence of the portfolio δ in the monitoring observations.

The formulation in Equation (5.8) allows the expression of the idea of semi-variance by taking only those deviations into account that are worse than average (Markowitz 1959) and accentuating larger deviations.

Having introduced a measure of risk for SLA portfolios that are established with customers as well as a risk measure for supplied SLA portfolios, the following section introduces the decision problem that takes total expected profit and resource contraints of an intermediary into account, as illustrated in Figure 5.2 and considers the challenges on decision making that were raised in Section 3.3.

5.1.3 Intermediary's Risk-Minimizing Decision about SLA Establishment

The previous sections introduced a risk measure for customer SLA portfolios and one for supplied SLA portfolios. This section presents the decision problem that allows an intermediary to first select the risk-minimal portfolio γ^* of customer SLAs and afterwards to identify the risk-minimal SLA portfolio offered by providers that support γ^*. This section illustrates the sequence of steps and calculations that are taken in an intermediary's decision on SLA establishment according to Figure 5.2. Furthermore a formulation of the decision problem is provided, which is illustrated by means of a numerical example.

In the first step of an intermediary's decision problem for SLA establishment, an intermediary evaluates if the available resources suffice for the provisioning of the requested service. This is achieved by evaluating the intermediary's resource constraint. Therefore, an intermediary takes information from SLAs $\alpha_i \in \gamma$, especially KPIs, into account. This information is translated to the resources that are available to the intermediary and may include those already provided by procured SLAs. On the one hand, there are KPIs like *available bandwidth* that translate directly to the intermediary's available resources L_ι. Even that this mapping is especially difficult in Cloud-based environments, where virtualized resources are present that make a mapping to physical resources difficult, an approach to this mapping has been introduced in Kwok and Mohindra (2008). According to these KPIs, the resource constraints of all customer SLA portfolios γ are evaluated analogously to the evaluation of provider's resource constraints in Section 4.1.2:

$$\sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i}) \leq L_\iota \,. \tag{5.9}$$

KPIs that can be extended, like storage space, are not taken into account here but are rather considered non-functional requirements for supplied services that need to be procured from providers.

With the help of the intermediary's resource constraints, the feasible SLA portfolios γ, which are offered to consumers, are identified. For each of the feasible portfolios γ, the risk is calculated according to Section 5.1.1 and the portfolio γ^* that incurs the lowest risk is chosen.

In the case that an intermediary's resources or functional capabilities do not suffice for providing γ^* to the customer the intermediary needs to detect SLAs supplied by providers that are able to meet functional and quality requirements for the provision of SLA portfolio γ^*. Therefore, a mapping from SLAs that are offered to required SLAs with providers is needed. This mapping indicates the minimum level of quality of a service from a provider, which needs to be procured in order to achieve a particular level of quality in a service offered to a customer. For feasibility reasons, the existence of a system that allows the intermediary to detect providers that are able to supply the required services is assumed. Similar approaches can be found in the project Value-Grids[2] and in Westermann et al. (2010). Furthermore, the system is assumed to support the mapping of provided services and quality levels to needed supplementary services and respective quality levels. The system is required to support service description and to enforce service composition, e.g. by means of common interfaces. Finally, the system enables the dynamic composition of services by checking the service providers' current availability status for further service executions. This system ensures the detection of technically and functionally required services. The design of a system that allows for the detection of functionally and quality-wise feasible SLAs δ is out of the scope of this work. Designing such a system would involve the description of services in a functional and quality-based way. Furthermore, a database for synonyms or a semantic annotation that enables querying of the knowledge-base is required as well as the possibility of mapping KPIs of offered SLAs to required SLAs with supplying providers. For research in this context, the reader is referred to Agarwal et al. (2008), Bodenstaff et al. (2008), Knapper et al. (2010) and Junghans and Agarwal (2010).

The service intermediary ι's choice of supplied SLAs, on the one hand, is based on functional requirements. On the other hand, the intermediary wants to achieve a particular expected profit from service provisioning taking into account the expected costs of procuring services. In order to compute the total expected profit that results from a portfolio of customer SLAs γ^* and a portfolio of supplied SLAs δ, ι considers the expected profit resulting from the SLA portfolio γ^* as well as the costs that ι has to expect from establishing the SLA portfolio δ. If the difference between these terms exceeds ι's profit threshold Π_ι, the configuration (γ^*, δ) is included in the set of feasible solutions. Assuming that ι is able to determine their preferences with respect to expected profit by means of a threshold for expected profit Π_ι.

From the SLA portfolio γ^* that was determined in the first step as depicted in Figure 5.2, ι calculates the expected profit as the sum of prices agreed in each $\alpha_i \in \gamma^*$ reduced by the sum of ι's incurred costs for service execution $c(s_j)$ and the expected penalties

[2] `http://www.valuegrids.de` last accessed: October 9, 2011

(see Equation 4.9):

$$E(\pi_{\gamma^*,t}) = \sum_{\alpha_i \in A} (f_{\alpha_i}(s_j) - c(s_j)) \cdot X(\lambda^{\alpha_i}_{\gamma^*,t}) - E(\mu_{\gamma^*,t}). \tag{5.10}$$

After ι has detected the functionally feasible services and has chosen the respective SLA portfolios δ that fulfill the quality requirements for the risk-minimal customer SLA portfolio γ^*, ι is able to calculate the expected costs χ resulting from each SLA portfolio δ. Expected costs of service procurement are computed as the sum of prices agreed in each $\beta \in \delta$ reduced by the expected due penalties that providers p have to pay ι:

$$E(\chi_\delta) = \sum_{\beta_i \in \delta} (f_{\beta_i}(s_j)) - E(\mu_{\delta,t}), \tag{5.11}$$

where $E(\mu_{\delta,t})$ is calculated based on ι's monitoring of their contracting partners' performance as the expected penalties per SLA β in a certain SLA portfolio δ. $E(\mu_{\delta,t})$ is calculated analogously to Equation 4.4 as the product of average degree of SLA violation and agreed penalties

$$E(\mu_{\delta,t}) = \sum_{\beta_i \in \delta} \overline{\lambda}^{\beta_i}_{\delta} \cdot \mu_{\beta_i} \cdot X(\lambda^{\beta_i}_{\delta,t}), \tag{5.12}$$

where the average degree of violation of $\beta_i \in \delta$, $\overline{\lambda}^{\beta_i}_{\delta}$, is computed as the sum of p's degree of violation of SLA $\beta_i \in \delta$, over all monitored periods t and divided by the number of periods, in which β was monitored

$$\overline{\lambda}^{\beta_i}_{\delta} = \frac{\sum_t \lambda^{\beta_i}_{\delta,t}}{|T_\delta|}. \tag{5.13}$$

Finally, ι calculates for each possible combination of γ^* and each δ the difference $E(\pi_{\gamma^*,t}) - E(\chi_{\delta,t}) = E(\pi_{\{\gamma^*,\delta\},t})$ and evaluates the expected profit constraint

$$E(\pi_{\{\gamma^*,\delta\},t}) = E(\pi_{\gamma^*,t}) - E(\chi_{\delta,t}) \geq \Pi_\iota. \tag{5.14}$$

If Equation 5.14 holds, the combination $\omega = \{\gamma, \delta\}$ is included in the set of acceptable solutions $\Omega = \{\Gamma, \Delta\}$.

Furthermore, ι specifies the threshold R_ι of average SLA violation that the portfolios δ under consideration may not exceed. Therefore, the average degree of violation is calculated for each portfolio δ as follows:

$$E(\lambda_{\delta,t}) = \sum_{\beta_i \in \delta} \frac{\overline{\lambda}_{\delta}^{\beta_i}}{|\delta|},$$

where

$$\overline{\lambda}_{\delta}^{\beta_i} = \frac{\sum_t \lambda_{\delta,t}^{\beta_i}}{|T_\delta|}.$$

If $E(\lambda_{\delta,t}) \leq R_\iota$, the combination $\omega = \{\gamma, \delta\}$ stays in the set of acceptable solutions. Otherwise, it will be removed.

Having identified the set of feasible solutions Ω, the intermediary calculates the risk associated with each supplied SLA portfolio δ and selects the risk-minimal one δ^*. The decision problem is then formulated in a two-step approach as follows:

$$\text{I} \qquad \min_{\gamma} \quad S_E(\mu_{\gamma,t}) = \sum_t q_t \cdot (\max\{0; \mu_{\gamma,t} - E(\mu_{\gamma,t})\})^2$$

$$\text{subject to} \quad \sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i}) \leq L_\iota$$

$$\text{II} \qquad \min_{\delta} \quad S_E(\lambda_{\delta,t}) = \sum_t \rho_t \cdot \left(\max\{0; \overline{\lambda}_{\delta,t} - E(\lambda_{\delta,t})\}\right)^2$$

$$\text{subject to} \quad E(\pi_{\{\gamma^*,\delta\},t}) \geq \Pi_\iota$$

$$E(\lambda_{\delta,t}) \leq R_\iota$$

The method that allows an intermediary to select the risk-minimal portfolio of customer and provider SLAs is illustrated with the help of a numerical example that is based on the scenario from the beginning of this chapter.

Example 5.1 (Intermediary's Risk-minimizing SLA Establishment Decision)**.** *Consider the scenario, where the intermediary Cisco Systems needs to decide on the SLAs to establish with customers and SLAs to establish with supplying providers. Assume that the intermediary's monitoring of their adherance to SLAs established with customers exposes the data illustrated in Table 4.1 and that Cisco Systems offers the following SLAs to customers:*

$$\alpha_1 \in A : (\iota_{Cisco}, s_1, \theta_1, B_w = 5, f_{\alpha_1}(s_1) = 4, \mu_{\alpha_1} = 2),$$
$$\alpha_2 \in A : (\iota_{Cisco}, s_2, \theta_2, B_w = 10, f_{\alpha_2}(s_2) = 11, \mu_{\alpha_2} = 2),$$
$$\alpha_3 \in A : (\iota_{Cisco}, s_1, \theta_3, B_w = 6, f_{\alpha_3}(s_1) = 4, \mu_{\alpha_3} = 3),$$

The selection of the risk-minimal SLA portfolio that is offered to customers is carried out as *argued in Section 5.1.1 according to Equation* (4.5). *The selection of a customer SLA portfolio corresponds to the risk-based SLA portfolio in Example 4.1 and results in the decision for* $\gamma' = \{\alpha_1, \alpha_2\}$ *that exposes an expected profit of* $E(\pi_{\{\alpha_1,\alpha_2\},t}) = 8.14$.

In order to decide on the supplied SLA portfolio, Cisco Systems monitors the supplying providers' adherence to established SLAs and stores the resulting monitoring reports (Table 5.1). Having identified the SLA portfolios that fulfill the functional and quality requirements, the intermediary evaluates for each portfolio the expected profit constraint from Equation (5.14), *by subtracting the expected costs for each supplied SLA portfolio from the expected profit of the customer SLA portfolio* $\gamma' = \{\alpha_1, \alpha_2\}$. *In this example, the intermediary's expected profit constraint is set to* $\Pi_\iota = 5$. *The intermediary's threshold value for the average degree of violation is* 0.5. *The providers offer the following SLAs to* ι*:*

$$\begin{aligned}
\beta_1 \in B &: (p_1,\, s_1,\, \iota_{Cisco},\, f_{\beta_1}(s_1) = 2,\, \mu_{\beta_1} = 2),\\
\beta_2 \in B &: (p_2,\, s_2,\, \iota_{Cisco},\, f_{\beta_2}(s_2) = 2,\, \mu_{\beta_2} = 2),\\
\beta_3 \in B &: (p_3,\, s_1,\, \iota_{Cisco},\, f_{\beta_3}(s_1) = 1,\, \mu_{\beta_3} = 1).
\end{aligned}$$

For calculating the expected costs of the SLA portfolios that expose monitoring data in Table 5.1, the average degree of violation of each SLA in a portfolio is computed. Consider SLA portfolio $\delta' = \{\beta_1, \beta_2\}$*, first.*

$$\overline{\lambda}^{\beta_1} = \frac{0.5 + 0.4 + 0.5}{3} \approx 0.4\overline{6} \tag{5.15}$$

$$\overline{\lambda}^{\beta_2} = \frac{0.3 + 0.1 + 0.3}{3} \approx 0.2\overline{3} \tag{5.16}$$

The expected costs $\delta' = \{\beta_1, \beta_2\}$ *result in*

$$\begin{aligned}
E(\chi_{\{\beta_1,\beta_2\}}) &= \sum_{\beta_i \in \delta'} (f_{\beta_i}(s_j)) - E(\mu_{\delta',t})\\
&\approx (2+2) - (0.4\overline{6} \cdot 2 + 0.2\overline{3} \cdot 2)\\
&= 2.6
\end{aligned}$$

The expected costs for $\delta'' = \{\beta_1, \beta_3\}$ *are calculated analogously and amount to* $E(\chi_{\{\beta_1,\beta_3\}}) = 1.4$. *As* $E(\pi_{\{\alpha_1,\alpha_2\},t}) = 8.14$, *both supplied SLA portfolios fulfill the intermediary's profit constraint:*

$$\begin{aligned}
E(\pi_{\{\alpha_1,\alpha_2\},t}) - E(\chi_{\{\beta_1,\beta_2\}}) &\approx 8.14 - 2.6 = 5.54 \geq 5\\
E(\pi_{\{\alpha_1,\alpha_2\},t}) - E(\chi_{\{\beta_1,\beta_3\}}) &\approx 8.14 - 1.4 = 6.74 \geq 5
\end{aligned}$$

In order to decide which SLA portfolio to establish, the intermediary calculates the risk associated with each of the portfolios. Consider SLA portfolio $\delta' = \{\beta_1, \beta_2\}$ for the time being. According to Section 5.1.1, first the actual degree of SLA violation for each period is calculated as follows:

$$\begin{aligned}
\lambda_{\delta',1} &= \frac{0.5+0.3}{2} = 0.4 \\
\lambda_{\delta',3} &= \frac{0.4+0.1}{2} = 0.25 \\
\lambda_{\delta',4} &= \frac{0.5+0.3}{2} = 0.4
\end{aligned}$$

Following the process described in Section 5.1.1, as next steps the mean degree of SLA violation for each SLA β_i and the average degree of violation of the SLA portfolio δ are calculated. Based on Equation 5.15, the mean degree of violation of $\delta' = (\beta_1, \beta_2)$ amounts to:

$$E(\lambda_{\delta',t}) \approx \frac{0.4\overline{6} + 0.2\overline{3}}{2} \approx 0.35.$$

Hence, δ' meets the constraint for the average degree of failure and stays in the set of feasible SLA portfolios.

Finally, the risk associated with $\delta' = (\beta_1, \beta_2)$ according to Equation (5.8) *results in*

$$\begin{aligned}
S_E(\lambda_{\delta',t}) &= \frac{1}{3} \cdot (\max\{0; 0.4 - 0.35\})^2 + \frac{1}{3} \cdot (\max\{0; 0.25 - 0.35\})^2 \\
&+ \frac{1}{3} \cdot (\max\{0; 0.4 - 0.35\})^2 \\
&= \frac{1}{3} \cdot 0.0025 + \frac{1}{3} \cdot 0.0025 \\
&= 0.00167
\end{aligned}$$

The average degree of violation for $\delta'' = \{\beta_1, \beta_3\}$ results in

$$\overline{\lambda}^{\{\beta_1,\beta_3\}} = \frac{0.7+0.2}{2} = 0.45\,. \tag{5.17}$$

Consequently, $\delta'' = \{\beta_1, \beta_3\}$ stays in the set of feasible SLA portfolios.

The risk associated with $\delta'' = \{\beta_1, \beta_3\}$ is calculated analogously an amounts to $S_E(\lambda_{\{\beta_1,\beta_3\},t}) =$ 0.01125. *Consequently, the supplied SLA portfolio $\delta' = (\beta_1, \beta_2)$ is requested first, as it has the lower risk.*

5.1.4 Summary

In this section, a method that supports an intermediary's decision about SLA establishment in order to facilitate the offering of composite services in dynamic, heterogenous environments was introduced. The intermediary's decision is separated into three steps:

1. Choose a customer SLA portfolio γ^*
2. Search for acceptable supplied SLA portfolios δ
3. Select a supplied SLA portfolio δ^*

For pursuing these steps, this section included the factors that constrain an intermediary's decision about the establishment of SLAs (see Sub Question 3.2) by means of resource and expected profit constraints. This way, an intermediary's properties are taken into account in the decision on SLA establishment. The risk of violating customer SLAs is measured by means of the semi-variance of expected penalties. This concept is adapted for calculating the risk of supplying providers that violate SLAs that are established with the intermediary (Sub Question 3.1). By minimizing the risk of SLA violation as measured by the semi-variance in the objective functions of the two-part decision problem, the challenges on decision making about SLA establishment that were raised in Section 3.3 are met. A decision on SLAs with a multitude of customers and providers is facilitated and the risk of SLA violation is expressed and taken into account. Furthermore, by applying an adaptation of portfolio selection the impact of the availability of the Internet and hence, a potentially high number of SLA portfolios is considered.

A critical point in the presented model is the construction of total profit from providing an SLA portfolio γ^* while procuring services in the SLA portfolio δ. The specification of total expected profit and with it, expected costs of procuring services reflects the monetary outcome of service consumption, that is, profit from offering services decreased by costs from procuring services. The total profit relies on the formulation of costs $\chi_{\delta,t}$ which arise with the procurement of services. The costs $\chi_{\delta,t}$ decrease with an increasing degree of violation by providers, as the prices for supplied services are reduced by expected penalties, whereas penalties are designed to reimburse the intermediary for an experienced SLA violation. From a monetary perspective, this gives the impression that a higher degree of violation can be preffered by the intermediary due to lower expected costs, for which reason the threshold for the average degree of violation was introduced. Furthermore, the impact of the violation of a supplied SLA to the adherence to offered SLAs is not taken into consideration in this work. In order to correctly represent these dependencies between violations of supplied SLAs and violations of offered SLAs, the calculation of risk would have to be done for each combination of offered and supplied SLA portfolios seperately, which could result in a huge amount of possible combinations and calculations. Additionally, this would require the intermediary to consider offered and supplied SLAs in the same point in time, which they may not wish to do.

The proposed risk-measure enables an intermediary to select the exact SLA combination γ^*, δ^* that imposes the lowest risk on them while considering resource, profit and violation constraints. Similar to the approach presented in Section 4.1, the model of an intermediary's risk-based SLA portfolio selection might exclude SLA portfolios based on the expected profit constraint, which might have been preferred by the intermediary because of a very low risk associated with the SLA portfolio. A trade-off between risk and expected profit that reflects the intermediary's preferences can help to overcome this inefficiency and needs to be employed as the objective function of the optimization. The following section introduces a further extension that applies a utility function to capture the provider's preferences towards expected profit in combination with risk.

5.2 Intermediary's Maximization of Expected Utility

The previous section introduced a method that solves an intermediary's decision problem by selecting the risk-minimizing portfolio of SLAs γ^* to offer to customers and the portfolio δ^* to purchase from providers, which incurs the lowest risk as measured by the semi-variance. An intermediary monitors providers' violation of established SLAs and consequently perceives risk of violation of supplied SLAs. An intermediary is also aware of the risk of violating customer SLAs and the resulting uncertainty of profit from service provisioning. In order express an intermediary's utility as required in Sub Question 3.4, both types of risk and the related expected profit (from a combination of customer and supplied SLAs) have to be taken into account. However, the decisions for customer SLAs and SLAs with supplying providers are not necessarily made concurrently. The decision for SLAs with supplying providers can be ommited if the required functionality is already covered by another SLA and resources are available. This could be the case if storage space is a required service and the intermediary already has a relevant SLA (e. g. an Amazon S3 instance[3]) where enough space is available. In contrast, if the intermediary requires more storage space, a new SLA for storage is required.

In order to reflect the possibility of seperate decisions, an intermediary's decision problem is split into the following steps analogously to Section 5.1:

1. Choose an SLA portfolio for provided complex services.
2. Select an SLA portfolio for supplied services.

Thus, the decision on the establishment of provided and required SLA portfolios is split in two steps. First, the risk-minimizing portfolio that is offered to customers is chosen. Afterwards, the risk-minimizing portfolio with supplying providers is selected taking into account the offered SLA portfolio γ^*. This method facilitates the selection of a risk-minimal portfolio of supplied SLAs for a given, risk-minimal, portfolio of customer SLAs. However, it minimizes risk and takes expected profit from a combination of customer and supplied SLA portfolios into account as a constraint. Yet, an intermediary cannot express their preferences by means of a trade-off between risk and expected profit explicitly. As was argued in Section 3.2.3, solving a multi-objective decision problem especially in the context of expected profit and related uncertainty that implies risk can be facilitated by the application of a utility function that reflects an intermediary's preferences. The expression of an intermediary's utility is in the focus of Sub Question 3.4:

RESEARCH QUESTION 3.4 $\prec$INTERMEDIARY'S UTILITY FUNCTION$\succ$. *How can an intermediary's utility from providing and procuring services and the corresponding risk of SLA violations be formulated?*

An intermediary's utility directly reflects their attitude towards risk by means of a trade-off between marginal incurred risk and the required additional profit for bearing any additional risk. The expression of utility that will be presented in Section 5.2 is an adaptation of a provider's utility as presented in Section 4.2, which considered exclusively the provision of services. However, it is as well an extension of the method that allows an intermediary to select the risk-minimizing SLAs with customers as well as providers by explicitly including an intermediary's attitude towards risk.

An intermediary maximizes their expected utility from providing services to customers and from purchasing services from supplying providers. The utility function is de-

[3] `http://www.amazon.com/aws` last accessed: October 9, 2011

signed to state an intermediary's preferences with respect to a trade-off between expected profit and risk. Consequently, there is no need for the expected profit constraint from Equation (5.14). The formulation of an intermediary's decision problem is covered by Sub Question 3.5:

RESEARCH QUESTION 3.5 ≺INTERMEDIARY'S UTILITY MAXIMIZING DECISION≻.
How is the intermediary's decision problem formulated in order to maximize the expected utility from SLA establishment and to account for the challenges from Section 3.3?

Figure 5.3 illustrates an intermediary's decision that maximizes the expected utility from providing and procuring services. Having received a request for service provision from a customer, an intermediary first evaluates the resource constraints and afterwards selects the SLA portfolio that maximizes the expected utility of providing SLAs. After selecting the supplied SLA portfolios that meet functional and quality requirements, the intermediary calculates the expected utility for each of the supplied SLA portfolios. Finally, the portfolio that maximizes the utility of supplied SLAs is chosen.

The following Sections present approaches to the Sub Questions that were raised in order to address Research Question 3. The selection of the customer SLA portfolio is presented in Section 5.2.1 and is carried out analogously to the utility-maximizing selection of SLA portfolios from Section 4.2. Section 5.2.2 introduces the selection of supplied SLA portfolios based on an indermediary's utility from providing and procuring services, which extends the expression of utility from Section 4.2. Based on the models from Sections 5.2.1 and 5.2.2, the intermediary's utility-maximizing decision problem is illustrated and discussed in Section 5.2.3.

5.2.1 Intermediary's Utility from Providing Services

Consider the service intermediary that was introduced in the Scenario. In order to decide on which SLA portfolio to establish with customers, the intermediary monitors their own adherence to previously established SLAs α in a portfolio of SLAs γ. An intermediary's monitoring is assumed to exhibit information on the degree of violation $\lambda_t^{\alpha_i}$ of SLA α_i for each monitored period. This information is retrieved from service monitoring that meets Assumptions 4 and 5 from Section 4.1.

The decision on the establishment of SLAs with customers is treated analogously to a

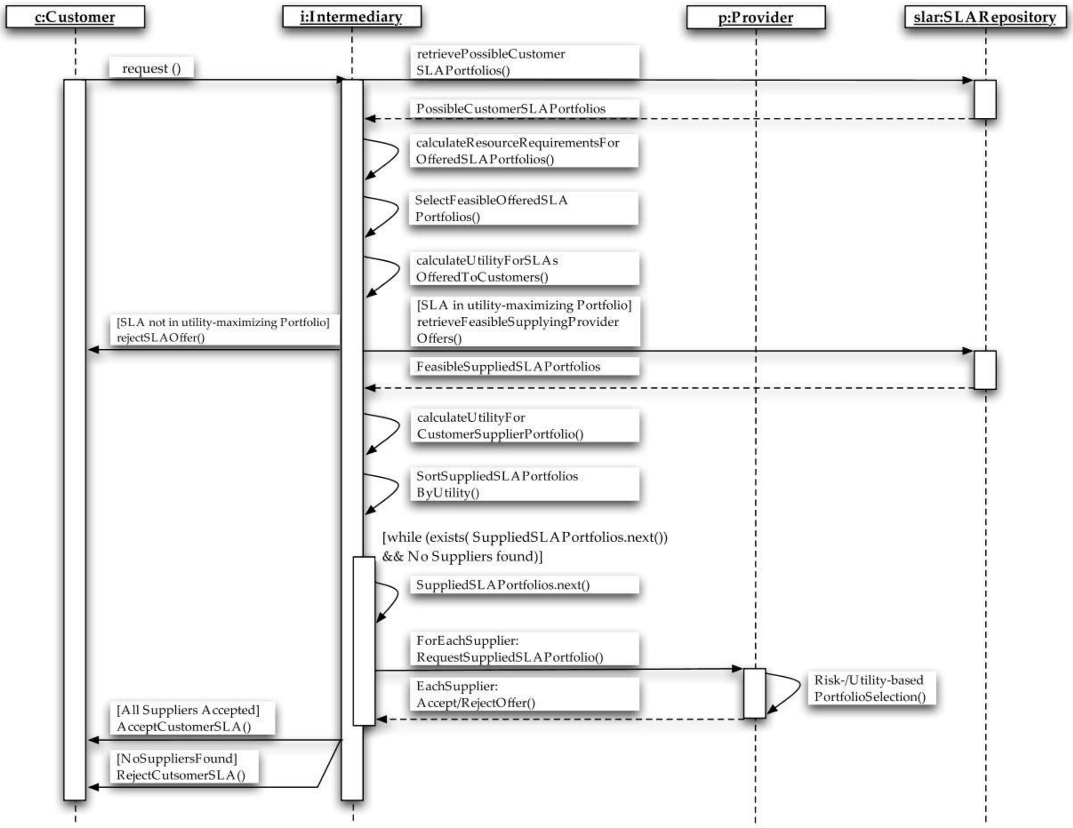

FIGURE 5.3: *Intermediary's Utility-maximizing Portfolio Selection*

provider's decision from Section 4.2. The intermediary's selection of the customer SLA portfolio γ is therefore based on the utility from providing services that is approximated by a trade-off between risk and expected profit individually weighted by the coefficient of absolute risk aversion.

As was shown in Section 4.2, maximizing the expected utility of profit $E(u(\pi_{\gamma,t}))$ is equivalent to maximizing the certainty equivalent

$$CE(\pi_{\gamma,t}) = E(\pi_{\gamma,t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma,t})) \cdot Var(\pi_{\gamma,t}), \tag{5.18}$$

where

$$\begin{aligned} \pi_{\gamma,t} &= \sum_{\alpha \in A} \left(f_{\alpha_i}(s_j) - c(s_j) \right) - \mu_{\gamma,t}, \\ ARA(E(\pi_{\gamma,t})) &= \frac{R(E(\pi_{\gamma,t}))}{E(\pi_{\gamma,t})}, \\ Var(\pi_{\gamma,t}) &= \sum_t q_t \cdot ((\pi_{\gamma,t} - E(\pi_{\gamma,t})))^2, \end{aligned}$$

with $R(E(\pi_{\gamma,t}))$ denoting the coefficient of relative risk aversion and $\mu_{\gamma,t}$ stating the actual penalties as before.

The SLA portfolio γ^* that maximizes the certainty equivalent is chosen. Subsequently, as illustrated in Figure 5.3, an intermediary selects those supplied SLA portfolios that meet functional and quality requirements. The calculation of an intermediary's utility of a supplied SLA portfolio δ is introduced in the following Section.

5.2.2 Intermediary's Utility from Procuring Services

After the selection of the utility-maximizing SLA portfolio with customers γ^*, and in the case that an intermediary's resources or functional capabilities are insufficient, an intermediary attempts to select the portfolio of supplied SLAs that maximizes the expected utility. An intermediary's performance in providing services to customers is not only influenced by their own failures but also by SLA violations from supplying providers. This makes the choice of a portfolio of supplied SLAs crucial. Expressing an intermediary's utility requires considering the risk that supplied SLAs are violated and the expected profit from procuring the SLA portfolio δ. Analogously to the approach in Section 5.1, the method aims to select the utility-maximizing supplied SLA portfolio for a given offered SLA portfolio.

In order to evaluate supplied SLA portfolios, an intermediary monitors how supplying providers perform in active SLAs in each reporting period. Monitoring reports contain information about active SLAs and the respective degree of SLA violation in each period. Table 5.1 showed an extract of an intermediary's supplied SLA monitoring. Using this monitoring data, an intermediary calculates the profit for each period as the difference between expected profit from the utility-maximizing offered SLA portfolio γ^* and the costs for the supplied SLA portfolio δ in the respective period

$$\pi_{\{\gamma^*,\delta\},t} = E(\pi_{\gamma^*,t}) - \chi_{\delta,t} \,, \tag{5.19}$$

where $E(\pi_{\gamma^*,t})$ denotes the expected profit from offering the utility maximizing SLA portfolio γ^* as before and $\chi_{\delta,t}$ states costs that result from procuring SLA portfolio δ in period t.

As a rational decision maker, the intermediary maximizes the expected utility from profit (Von Neumann et al. 1947), where the utility function $u(\pi_{\{\gamma^*,\delta\},t})$ is assumed to be a von Neumann-Morgenstern utility, where $u'(\pi_{\{\gamma^*,\delta\},t}) > 0$, $u''(\pi_{\{\gamma^*,\delta\},t}) < 0$. The expected utility that is to be maximized is calculated as

$$E(u(\pi_{\{\gamma^*,\delta\},t})) = \sum_t q_t \cdot u(\pi_{\{\gamma^*,\delta\},t}), \tag{5.20}$$

where q_t is the probability for $u(\pi_{\{\gamma^*,\delta\},t})$ as before. In order to identify the SLA portfolio δ^* that maximizes the intermediary's utility with respect to the offered SLA portfolio γ^*, a specific utility function has to be identified. As argued in Section 4.2, there are suitable utility functions that model risk aversion of decision maker. To apply these functions, decision maker's are required to be completely aware of their preferences, which is most often not the case. Furthermore, the presented utility functions do only include the risk of SLA violation implicitly. In order to include an explicit specification of risk, an alternative formulation of expected utility is derived that contains the first two moments of profit and hence, the variance of profit.

Analogously to the approach that was presented in Section 4.2, profit is represented as a lottery $w + z$, where w is some certain income and z is a stochastic variable with mean 0 and Variance $Var(z)$ without any specific probability distribution. The expected utility of the lottery is $E(u(w + z))$. According to Eeckhoudt et al. (2005), applying a Taylor approximation on expected utility results in

$$E(u(w+z)) \approx u\left(E(w) - \frac{1}{2} \cdot ARA(w) \cdot Var(z)\right). \tag{5.21}$$

$E(w) - \frac{1}{2} \cdot ARA(w) \cdot Var(z)$ denotes the certainty equivalent, as its utility corresponds to the lottery's expected utility $E(u(w+z))$. $ARA(w) = \frac{R(w)}{w} = -\frac{u''(w)}{u'(w)}$ is the coefficient of absolute risk aversion measured at w (see Pratt (1964) and Arrow (1971)).

Define

$$w = E(\pi_{\{\gamma^*,\delta\},t}) \,, \tag{5.22}$$

and

$$z = \pi_{\{\gamma^*,\delta\},t} - E(\pi_{\{\gamma^*,\delta\},t}) \,, \tag{5.23}$$

where

$$E(\pi_{\{\gamma^*,\delta\},t}) = \sum_t q_t \cdot \pi_{\{\gamma^*,\delta\},t} \,. \tag{5.24}$$

Then, z exhibits an expected value

$$E(z) = E(\pi_{\{\gamma^*,\delta\},t} - E(\pi_{\{\gamma^*,\delta\},t})) = 0 \tag{5.25}$$

and a variance

$$\begin{aligned} Var(z) &= Var(\pi_{\{\gamma^*,\delta\},t} - E(\pi_{\{\gamma^*,\delta\},t})) \\ &= E((\pi_{\{\gamma^*,\delta\},t} - E(\pi_{\{\gamma^*,\delta\},t}))^2 - (E(\pi_{\{\gamma^*,\delta\},t} - E(\pi_{\{\gamma^*,\delta\},t})))^2) \\ &= Var(\pi_{\{\gamma^*,\delta\},t}). \end{aligned} \tag{5.26}$$

Consequently, $w + z$ constitutes the lottery of expected profit that meets the above required characteristics.

Consider again Equation (5.21) and substitute w and z as defined above into the certainty equivalent. Then, the following statement is achieved $E(u(\pi_{\{\gamma^*,\delta\},t})) \approx u(E(\pi_{\{\gamma^*,\delta\},t}) - \frac{1}{2} \cdot ARA(E(\pi_{\{\gamma^*,\delta\},t})) \cdot Var(\pi_{\{\gamma^*,\delta\},t}))$, where the variance of profit with SLA portfolios γ^* and δ equals

$$\begin{aligned} &Var(\pi_{\{\gamma^*,\delta\},t}) = \\ &\sum_t q_t \cdot ((\pi_{\{\gamma^*,\delta\},t} - E(\pi_{\{\gamma^*,\delta\},t})))^2 \,. \end{aligned} \tag{5.27}$$

This specification of the certainty equivalent allows for expected profit as well as the risk that is expressed by the variance of profit $Var(\pi_{\{\gamma^*,\delta\},t})$. The intermediary's preferences towards risk and expected profit are weighted individually by the coefficient of absolute risk aversion $ARA(E(\pi_{\{\gamma^*,\delta\},t}))$, which still needs to be identified for the practical implementation.

Maximizing the expected utility of the intermediary $E(u(\pi_{\{\gamma^*,\delta\},t}))$ is equivalent to maximizing the certainty equivalent

$$CE(\pi_{\{\gamma^*,\delta\},t}) = E(\pi_{\{\gamma^*,\delta\},t}) - \frac{1}{2} \cdot ARA(E(\pi_{\{\gamma^*,\delta\},t})) \cdot Var(\pi_{\{\gamma^*,\delta\},t}). \tag{5.28}$$

Consequently, the certainty equivalent is employed in the objective function in the utility-maximizing decision problem that is presented in the following section. This specification of the certainty equivalent extends the formulation from Section 4.2 by considering the risk that procured SLAs are violated.

5.2.3 Intermediary's Utility-Maximizing Decision on SLA Establishment

In the previous sections, the means to express the expected utility from the establishment of SLAs for provided as well as procured services have been introduced. In this section, the decision model that employs the expressions of expected utility from Sections 5.2.1 and 5.2.2 for the choice of SLA portfolios is introduced. Therefore, the steps that are taken in the decision process are described according to the sequence illustrated in Figure 5.3. Additionally, a formulation of the decision problem is given and an illustrating example concludes.

As a first step for the decision on SLA establishment, an intermediary evaluates their resource constraints. Therefore, an intermediary takes information from SLAs, especially KPIs into account. This information is set into context with resources that are available to an intermediary (L_ι) and constrain the decision on SLA establishment. According to these KPIs, the resource constraints of all customer SLA portfolios γ are evaluated before (see Sections 4.1.2, 4.2.2, 5.1.3):

$$\sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda^{\alpha_i}_{\gamma,t}) \leq L_\iota \,. \tag{5.29}$$

KPIs that translate to the intermediary's resources in conjunction with the supplying providers' resources are considered requirements for the selection of procured services.

Having identified the feasible customer SLA portfolios, the intermediary calculates the certainty equivalent for each of the SLA portfolios as introduced in Section 5.2.1. The SLA portfolio γ^* that maximizes the intermediary's expected utility is chosen.

As a next step, an intermediary selects those SLA portfolios δ that meet functional requirements and support the provision of the customer SLA portfolio γ^* as described at the beginning of this Section.

Having identified the feasible SLA portfolios that are offered by supplying providers, an intermediary calculates the expected degree of SLA violation for each of the portfolios and evaluates the resulting value against the threshold R_ι for the average degree of SLA violation. This threshold is the maximum average degree of violation of an SLA portfolio that an intermediary is willing to accept. For those SLA portfolios that meet the average violation constraint, the intermediary calculates the certainty equivalent according to Equation (5.28). The portfolio that maximizes the expression in Equation (5.28) is chosen as the best-suited SLA portfolio δ^* for the given offered SLA portfolio γ^*.

The decision problem is formulated as follows:

$$\text{I} \qquad \max_{\gamma} \quad CE(\pi_{\gamma,t}) = E(\pi_{\gamma,t}) - \frac{1}{2} \cdot ARA(E(\pi_{\gamma,t})) \cdot Var(\pi_{\gamma,t})$$
$$\text{subject to} \qquad \sum_{\alpha_i \in A} L_{\alpha_i} \cdot X(\lambda_{\gamma,t}^{\alpha_i}) \leq L_\iota$$

$$\text{II} \qquad \max_{\delta} \quad CE(\pi_{\{\gamma^*,\delta\},t}) = E(\pi_{\{\gamma^*,\delta\},t}) - \frac{1}{2} \cdot ARA(E(\pi_{\{\gamma^*,\delta\},t})) \cdot Var(\pi_{\{\gamma^*,\delta\},t})$$
$$\text{subject to} \qquad E(\lambda_{\delta,t}) \leq R_\iota$$

The application of the method for solving an intermediary's decision problem that includes an intermediary's attitude towards risk explicitly is illustrated in the following example.

Example 5.2 (Intermediary's Utility-maximizing Decision)**.** *For illustration purposes, the situation given in Example 5.1 is solved applying the utility-maximizing method that was presented throughout this chapter by first selecting the utility-maximizing customer SLA portfolio γ^* and afterwards choosing the utility-maximizing SLA portfolio δ^* with supplying providers that meets the functional, resources' and average degree of violation constraints.*

Consider again the intermediary Cisco Systems with the same monitoring data (Table 4.1) and the same SLAs to offer customers:

$$\begin{aligned}
\alpha_1 \in A :&(\iota_{Cisco}, s_1, \theta_1, B_w = 5, f_{\alpha_1}(s_1) = 4, \mu_{\alpha_1} = 2),\\
\alpha_2 \in A :&(\iota_{Cisco}, s_2, \theta_2, B_w = 10, f_{\alpha_2}(s_2) = 11, \mu_{\alpha_2} = 2),\\
\alpha_3 \in A :&(\iota_{Cisco}, s_1, \theta_3, B_w = 6, f_{\alpha_3}(s_1) = 4, \mu_{\alpha_3} = 3),
\end{aligned}$$

Analogously to Example 4.3, the utility-maximizing SLA portfolio γ^* that is chosen by the intermediary is $\gamma^* = \{\alpha_1, \alpha_2\}$ with a certainty equivalent of $CE_{\gamma^*} = 7.4953$ and an expected profit of $E(\pi_{\{\alpha_1,\alpha_2\},t}) = 8.14$.

For selecting the utility-maximizing supplied SLA portfolio that supports γ^*, the monitoring reports about supplying providers' performance as well as information from SLAs is taken into account. Analogously to Example 5.1, the intermediary's partner monitoring is given in Table 5.1 and the SLAs offered by supplying providers are the following:

$$\begin{aligned}
\beta_1 \in B :&(p_1, s_1, \iota_{Cisco}, f_{\beta_1}(s_1) = 2, \mu_{\beta_1} = 2),\\
\beta_2 \in B :&(p_2, s_2, \iota_{Cisco}, f_{\beta_2}(s_2) = 2, \mu_{\beta_2} = 2),\\
\beta_3 \in B :&(p_3, s_1, \iota_{Cisco}, f_{\beta_3}(s_1) = 1, \mu_{\beta_3} = 1).
\end{aligned}$$

The intermediary's threshold value for the average degree of violation of a portfolio is again set to 0.5*. This value states that the intermediary only accepts an SLA portfolio, which is violated by at most* 50% *on average. The exact value for the threshold of the average degree of violation is up to the intermediary and depends on the kind and relevance of the services that the intermediary procures.*

Assume that the intermediary conducted a mapping of properties of offered SLAs to properties that supplied services need to provide. Furthermore, assume that the intermediary has access to a service repository that allows for functional as well as non-functional properties and that the SLAs resulting from a query to the repository are $\beta_1, \beta_2, and \beta_3$.

Having identified the feasible SLAs and consequently the portfolios, the intermediary checks for the average degree of violation for each of the portfolios. According to the monitoring data in Table 5.1, the average degree of violation for each SLA in portfolio $\delta = \{\beta_1, \beta_2\}$, amounts to

$$\overline{\lambda}^{\beta_1} = \frac{0.5 + 0.4 + 0.5}{3} \approx 0.4\overline{6}$$
$$\overline{\lambda}^{\beta_2} = \frac{0.3 + 0.1 + 0.3}{3} \approx 0.2\overline{3}$$

Consequently, the average degree of violation results in

$$\overline{\lambda}^{\{\beta_1,\beta_2\}} = \frac{0.4\overline{6} + 0.2\overline{3}}{2} \approx 0.35\,. \tag{5.30}$$

Accordingly, the average degree of violation for $\delta'' = \{\beta_1, \beta_3\}$ results in

$$\overline{\lambda}^{\{\beta_1,\beta_3\}} = \frac{0.7 + 0.2}{2} = 0.45\,. \tag{5.31}$$

Hence, both portfolios meet the average degree of violation constraint and for both, the certainty equivalent needs to be calculated in order to identify the best-suited one. Consider $\delta' = \{\beta_1, \beta_2\}$ first.

The monitoring data in Table 5.1 exhibits observations for three active periods for the portfolio, hence $q_t = \frac{1}{3}$ in Equation (5.24).

The penalties that the intermediary received in each of the periods result in

$$\mu_{\delta',1} = 0.5 \cdot 2 + 0.3 \cdot 2 = 1.6, \tag{5.32}$$

$$\mu_{\delta',3} = 0.4 \cdot 2 + 0.1 \cdot 2 = 1.0, \tag{5.33}$$

$$\mu_{\delta',4} = 0.5 \cdot 2 + 0.3 \cdot 2 = 1.6. \tag{5.34}$$

Expected profit from SLA combination $\{\gamma^*, \delta'\}$ is thus substituted into (5.11)

$$\begin{aligned} E(\pi_{\{\gamma^*,\delta'\},t}) &= E(\pi_{\{\alpha_1,\alpha_2\},t}) - \sum_{\beta_1,\beta_2} (f_{\beta_i}(s_j)) + E(\mu_{\delta',t}) \\ &= 8.14 - (2+2) + 1.4 = 5.54. \end{aligned} \tag{5.35}$$

The variance of profit for $\{\gamma^*, \delta'\}$ results according to Equation 5.27 is

$$\begin{aligned} Var(\pi_{\{\gamma^*,\delta'\},t}) &= q_1 \cdot \left(\pi_{\{\gamma^*,\delta'\},1} - E(\pi_{\{\gamma^*,\delta'\},t})\right)^2 + q_3 \cdot \left(\pi_{\{\gamma^*,\delta'\},2} - E(\pi_{\{\gamma^*,\delta'\},t})\right)^2 \\ &+ q_4 \cdot \left(\pi_{\{\gamma^*,\delta'\},4} - E(\pi_{\{\gamma^*,\delta'\},t})\right)^2 \\ &= \frac{1}{3} \cdot (5.74 - 5.54)^2 + \frac{1}{3} \cdot (5.14 - 5.54)^2 \\ &+ \frac{1}{3} \cdot (5.74 - 5.54)^2 \\ &= 0.08. \end{aligned}$$

Analogously to Section 4.2, the coefficient of absolute risk aversion is calculated based on the coefficient of relative risk aversion that is set to 3 *(see Szpiro 1988, p. 106) as follows $ARA(w) =$*

$\frac{R(w)}{w}$. *Hence, the coefficient of absolute risk aversion amounts to*

(5.36)
$$ARA(E(\pi_{\{\gamma^*,\delta'\},t})) = \frac{R(E(\pi_{\{\gamma^*,\delta'\},t}))}{E(\pi_{\{\gamma^*,\delta'\},t})} = \frac{3}{5.54} = 0.5415.$$

Thus, the certainty equivalent (CE) of the intermediary is

(5.37)
$$\begin{aligned} CE_{\{\gamma^*,\delta'\}} &\approx E(\pi_{\{\gamma^*,\delta'\},t}) - \frac{1}{2} \cdot ARA(E(\pi_{\{\gamma^*,\delta'\},t})) \cdot Var(\pi_{\{\gamma^*,\delta'\},t}) \\ &= 5.54 - \frac{1}{2} \cdot 0.5415 \cdot 0.08 = 5.51834 \end{aligned}$$

Analogously, the expected profit for $\delta'' = \{\beta_1, \beta_3\}$ *is*

(5.38)
$$E(\pi_{\{\gamma^*,\delta''\},t}) = 6.54,$$

variance of profit results in

$$\begin{aligned} Var(\pi_{\{\gamma^*,\delta''\},t}) &= q_2 \cdot \left(\pi_{\{\gamma^*,\delta''\},1} - E(\pi_{\{\gamma^*,\delta''\},t})\right)^2 + q_5 \cdot \left(\pi_{\{\gamma^*,\delta''\},2} - E(\pi_{\{\gamma^*,\delta''\},t})\right)^2 \\ &= \frac{1}{2} \cdot (6.34 - 6.54)^2 + \frac{1}{2} \cdot (7.14 - 6.54)^2 \\ &= 0.2. \end{aligned}$$

Analogously to the calculation above, the coefficient of absolute risk aversion results in

$$\begin{aligned} ARA(E(\pi_{\{\gamma^*,\delta''\},t})) &= \tfrac{R(E(\pi_{\{\gamma^*,\delta''\},t}))}{E(\pi_{\{\gamma^*,\delta''\},t})} \\ &= \tfrac{3}{6.54} = 0.4587. \end{aligned}$$

The certainty equivalent for $\delta'' = \{\beta_1, \beta_3\}$ *amounts to*

$$\begin{aligned} CE_{\{\gamma^*,\delta''\}} &\approx E(\pi_{\{\gamma^*,\delta''\},t}) - \frac{1}{2} \cdot ARA(E(\pi_{\{\gamma^*,\delta''\},t})) \cdot Var(\pi_{\{\gamma^*,\delta''\},t}) \\ &= 6.54 - \frac{1}{2} \cdot 0.4587 \cdot 0.2 = 6.49413 \end{aligned}$$

As

$$CE_{\{\gamma^*,\delta''\}} > CE_{\{\gamma^*,\delta'\}},$$

$\delta'' = \{\beta_1, \beta_3\}$ *is chosen as the SLA portfolio supplied by providers.*

5.2.4 Summary

In this section, an intermediary's decision problem on SLA establishment is formulated that maximizes an intermediary's expected utility from providing services to customers in a first step. In a second step, an intermediary is able to identify the SLAs that are offered by supplying providers that meet functional and non-functional requirements for supporting the selected customer SLA portfolio. The presented approach allows an intermediary to select the supplied SLA portfolio that maximizes their utility (Research Question 3.5). The utility of offering SLAs to customers expresses a trade-off between expected profit from providing services and risk of intermediary's SLA violation that is individually weighted by the coefficient of absolute risk aversion. The utility of procuring services from supplying providers is specified analogously by a trade-off between expected profit from procuring services and the risk of supplying providers' SLA violation that is individually weighted by the coefficient of absolute risk aversion (Research Question 3.4).

The method that was presented meets the challenges on decision making about SLA establishment that were raised in Section 3.3. The approach includes the risk of SLA violation in the decision, which is expressed by means of the variance of profit. The impact of the availability of the Internet on service provisioning that results in a potentially high number of requests for service provisioning and with it SLAs, is taken into consideration by allowing for portfolios of SLAs in the decision on SLA establishment. The composition of services is covered this way as well. The approach also allows a multitude of customers to be considered in the decision model. Besides accounting for the challenges from Section 3.3, an intermediary's properties with respect to available resources and expected profit are considered in the decision problem. This way, the presented method for solving an intermediary's decision problem represents a solution to Research Question 3, which considers an intermediary's attitude towards risk explicitly in contrast to the method from Section 5.1.

However, the representations of expected utility that are employed in this section's approach are, just like the representation of expected utility from Section 4.2, an approximation.[4] Like any approximation, it is imprecise. The approach, however, remains the one most often employed in finance.

Furthermore, as was already argued in Section 4.2, the approximation of expected util-

[4]Except when the utility function is of the form $u(\pi_{\gamma,t}) = -\frac{e^{-a \cdot \pi_{\gamma,t}}}{a}$, with constant $ARA(\pi_{\gamma,t}) = a$, and profit is normally distributed (see Eeckhoudt et al. 2005, pp. 20–21).

ity that was employed throughout this section is only defined for positive values of profit. As the total profit that results from service provisioning, especially when services need to be procured in order to meet functional and quality expectations of the provided service, profit may take negative values. In these cases, the certainty equivalent cannot be applied for the decision about SLA establishment. Therefore, different representations of expected utility need to be investigated, which is a crucial topic for future research.

Another critical point is the formulation of expected profit in Equation (5.24). Expected profit in this case increases with an increasing degree of SLA violation by supplying providers. This gives the counterintuitive impression that a higher degree of SLA violation is valued higher by the intermediary, as the penalties to be paid increase with the degree of SLA violation. Penalties are designed to reimburse the customer, that is in this case, the intermediary, for insufficient performance. Hence, the specification of profit reflects the monetary outcomes of service consumption correctly. Nevertheless, the impact of SLA violation by a supplying provider on the intermediary's performance is not reflected in the expected profit. A higher degree of SLA violation from supplying providers might lead to a higher probability and a higher degree of SLA violation by the intermediary. In order to take these dependencies into account, an approach that distinguishes the monitoring data of the intermediary's SLA violation in customer SLAs by the respective active supplied SLAs would be required. Although this approach would result in an exact reflection of dependencies between supplied and customer SLAs, it would require an enormously higher amount of monitoring data as profit and variance would have to be calculated for each combination of customer and supplied SLA portfolios.

Part III

Evaluation and Application

Chapter 6

Risk-minimizing Decisions in SaaS Provisioning

This chapter describes the application of the methods that address a service provider's decision problem from Chapter 4 and the methods that address an intermediary's decision problem from Chapter 5 in the context of software as a service (SaaS) provisioning in the research project ValueGrids. ValueGrids is part of the third term of the German Grid initiative and is funded by the German Federal Ministry of Education and Research.

In this chapter, idea and goal of the project ValueGrids are sketched in Section 6.1. The section highlights the setting of SaaS provisioning and the decisions on SLA establishment that are made. Furthermore, requirements on tools that support decision making are derived. In Section 6.2, the particular use case of the risk-minimizing decision method is described and the single steps that are taken for risk-minimizing decisions are illustrated. Finally, the technical implementation is briefly explained in Section 6.3 in order to stress the feasibility of the application of the methods that were presented throughout this work.

6.1 ValueGrids

The aim of the ValueGrids project is to support a holistic management of value chains, which are created by the dynamic composition of service modules from heterogenous providers of Web services, Cloud services or SaaS. These value chains emerge upon the request for a specific service by a customer. One example for such a value chain is depicted in Figure 6.1, where a customer requests an enterprise resource planning

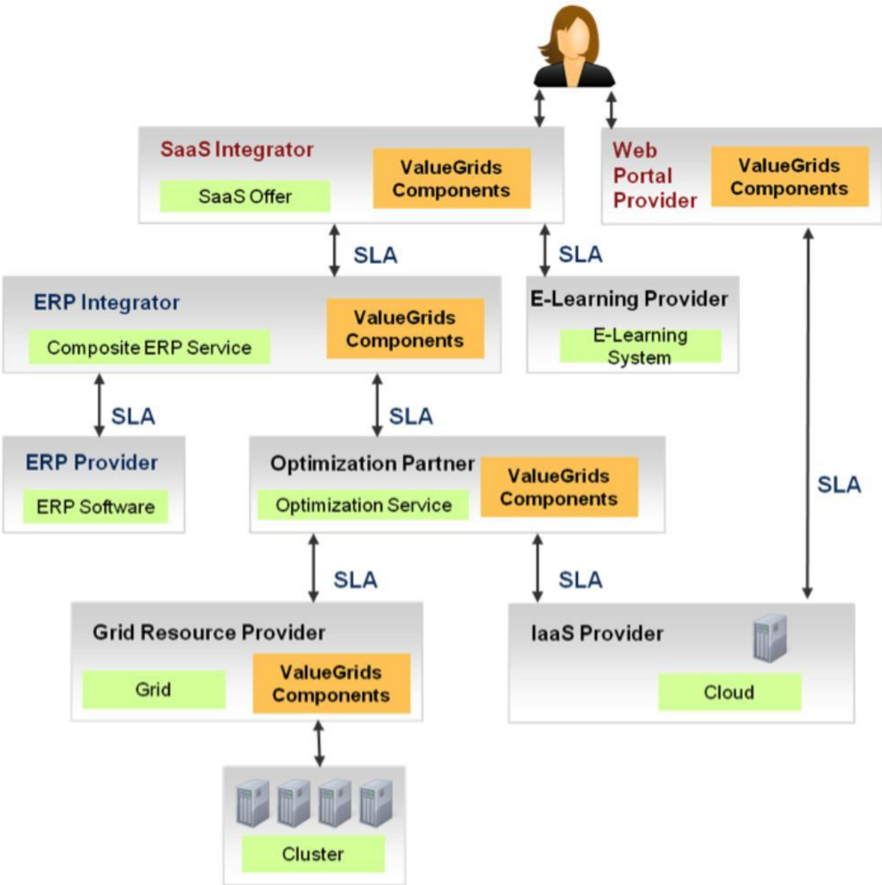

FIGURE 6.1: *ValueGrids Case Study Scenarios*

(ERP) service, that comprises e-learning functionality which supports the user if questions about the usage of the ERP system arise. In this scenario, the composition of a composite ERP service and an e-learning service are provided by an SaaS integrator to customers. The ERP service itself is contributed by an ERP provider. Additionally, an optimization service for optimization tasks in the ERP system is included in the composite ERP service. This optimzation service relies on infrastructure services, which are procured by the optimization partner in order to fulfill the optimization tasks. This process, however, is transparent to the end customer. The optimization partner and the ERP integrator are intermediaries according to Definition 2.8. Partners from industry that participate in the project ValueGrids are SAP,[1] IBM, [2] and Conemis,[3] which showcases the relevance of the project and the decision methods to industry and supports the applicability of the approach for practical use.

[1]`http://www.sap.com` last accessed: October 9, 2011
[2]`http://www.ibm.com` last accessed: October 9, 2011
[3]`http://www.conemis.com/` last accessed: October 9, 2011

Each of the compositions is governed by an SLA. SLAs are indicated by arrows in Figure 6.1 and regulate the functionality and quality aspects of the provided service. Additionally, a price for service provisioning and a penalty, which is applied in case of an SLA violation, are stipulated (Section 2.2).

Managing value chains does not only comprise of the functional composition of services in order to provide a particular functionality, but also of the management of quality aspects and with it service level management. For providers of services, the establishment of an SLA defines the possible monetary outcomes of service provisioning by specifying a price for service provisioning and a penalty that is applied in case of SLA violation. Consequently, service providers of both types (i. e. providers and intermediaries) are confronted with the decision of which SLAs should be established or active in the next period.

In the context of the ValueGrids project and in analogy to Research Questions 1 - 3, the following questions were identified to constitute the focus of a provider's decision:

1. Which services can be provided and in which quality categories? Which services have to be procured from other providers in order to provide the required functionality?
2. How do provided services perform in a regular setting?
3. Which service level is required from procured services in order achieve a particular provided quality?
4. What is the risk of violating a particular SLA? What impact do concurrently established SLAs have on the executed services? How can the risk of violating an SLA be expressed based on SLA monitoring data?

In order to find an answer to these questions, a set of tools is developed in the project according to the architecture depicted in Figure 6.2. The tools comprise of a Service Repository, a central storage for available services, SLAs and topologies, which can be queried according to the current customer request (Question 1). Topologies exhibit mappings from offered to required services and SLAs. In other words, a topology defines, which service and respective service level needs to be procured to achieve a certain service quality and functionality. The tools that provide the input for the mapping information that is saved in the Service Repository is created by the Performance Cockpit (Westermann and Momm 2010; Westermann et al. 2010), which tests regular behavior of services. On the basis of the regular performance of services that was identified with the help of the performance cockpit, SLA Translation contributes a matching for SLAs between provided and procured SLAs based on the results of the Performance

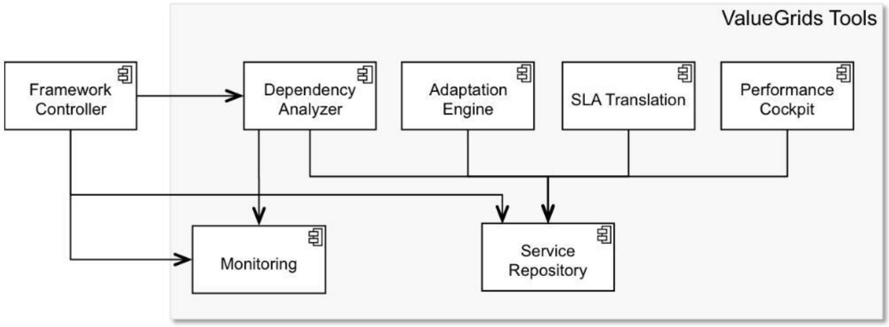

FIGURE 6.2: *ValueGrids Tools*

Cockpit. The Dependency Analyzer (Michalk and Caton 2010) is designed to identify dependencies between concurrently active SLAs and service executions and with these, the risk of an SLA violation. As was argued before, uncertainty about future events is often approximated from past observations (Section 3.2). In the case of the Dependency Analyzer, the risk of SLA violation is calculated from past monitoring observations, which are provided by the Monitoring tool, and states which SLAs have been active concurrently in the past and to which degree the respective SLAs have been violated. The monitoring observations meet Assumptions 3 - 5 from Section 4.1. Based on this information, the decision maker is supported in their decision for which SLAs to establish in the future.

6.2 Risk Analysis Use Case

This section describes the use case of risk analysis in the context of Web/Cloud service and SaaS provisioning in more detail in order to showcase the application scenarios of the Dependency Analyzer. The decision on SLA establishment can be made for different kinds of service. In the case of ValueGrids, the selection of a scenario in the area of Web services, Cloud services and SaaS offerings provides the benefits of existing monitoring systems and the principal availability of measures for the quality of a service that can be evaluated. The ability to measure the quality of a service and evaluate the adherence to SLAs is a driver for the application of the decision support as presented in this thesis. The Dependency Analyzer can be applied by each of the actors in Figure 6.1, denoted by "ValueGrids components", i. e. the SaaS integration, the ERP integration, optimization, and resource levels. This implies that each of the participants in a service composition scenario, i. e. each real-world service provider, that needs to make SLA establishment decisions can use the Dependency Analyzer as a tool for decision support.

By employing the Dependency Analyzer, a participant in a service economy can answer the following questions:

- Which service/SLA topologies are feasible for the provision of a service with respect to profit? (see Research Question 1)
- Which service/SLA topology is the one that incurs the lowest risk? (see Research Questions 2 and 3)

Decision support in highly flexible and dynamic real-world service economies is only possible if certain requirements are met, which relate to the properties of on-demand service provisioning and the availability of needed information. In the context of the project ValueGrids, the following requirements on service level management have been identified:

REQ 1. Service monitoring has to be possible and in action (Assumptions 4 and 5).

REQ 2. Monitoring reports have to be sent on a regular basis (Assumptions 4 and 5).

REQ 3. The length of all report periods has to be equally long (Assumption 4).

REQ 4. Reports have to include an aggregated value for all Service Level Objectives that reflects the overall degree of violation of the SLA (Assumption 5).

REQ 5. A service/SLA repository that is able to include relationships between services and SLAs has to be in place.

REQ 6. Each available SLA and service has to be included in the service/SLA repository.

These requirements were considered in the implementation of the ValueGrids tools, which are designed to solve the decision problems that were described in Sections 4.1.3, 4.2.2, 5.1.3 and 5.2.3. Figure 6.3 illustrates the steps that are taken for the decision on SLA establishment in a service provider's risk-minimizing decision, exemplarily. The steps are described more generally by considering the maximization of expected utility additionally, below.

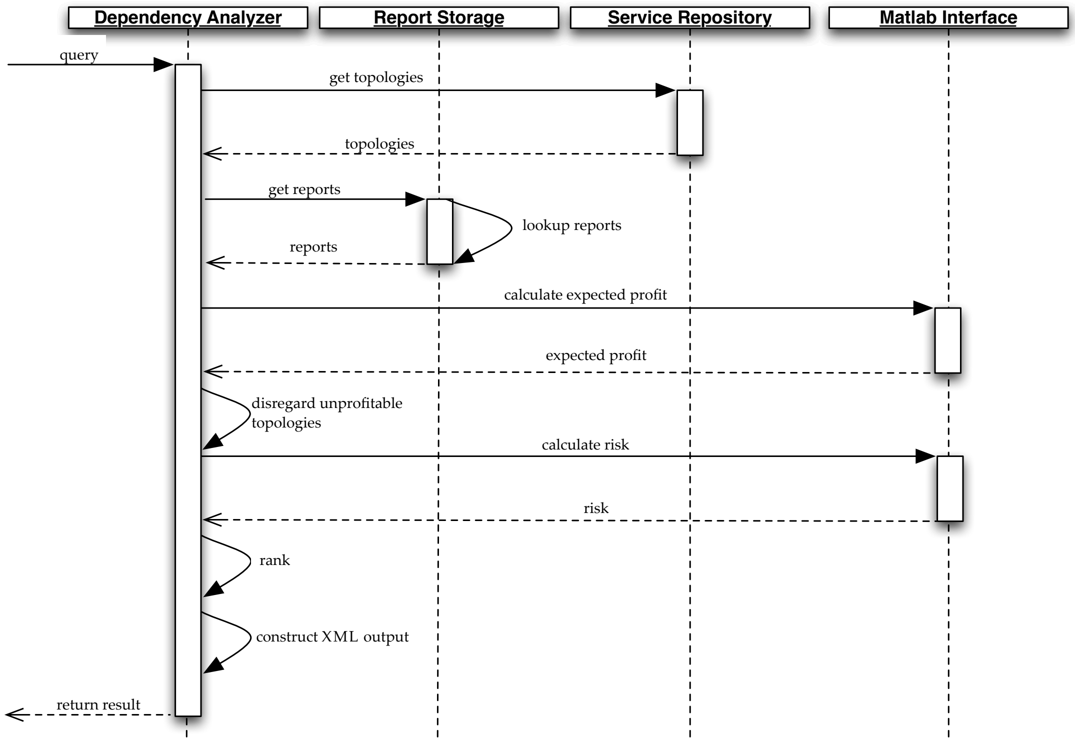

FIGURE 6.3: *Dependency Analyzer Decision Making*

query The decision itself is triggered by a user interface, the Framework Controller (see Figure 6.3), which is applied whenever a request for service provision is issued and a provider or intermediary is confronted with the decision whether to establish an SLA and if so, in which quality level.

get topologies Based on this request, the Dependency Analyzer (DA) retrieves the feasible SLA topologies for the requested service from the Service Repository (see REQ 6).

get reports The monitoring reports are retrieved from the DA's internal Report Storage. These have been received in the past from the monitoring tool (REQs 1 and 2) and stored internally by the DA for later use in decision support. Monitoring reports contain information about the past adherence of established SLAs, and each report comprises of a start and end time of the monitoring report, the ID of the monitored SLA(s) and the degree of violation (see REQ 2 - 4).

calculate expected profit From the reports that are stored in the ReportStorage, the expected profit is calculated by taking prices and penalties (from SLAs), costs for service execution as well as the expected violation as calculated from monitoring observations into account. The expected profit is calculated according to the expected profit constraint from Sections 4.1 and 5.1.

calculate risk The decision support that is provided by the Dependency Analyzer applies the concept of semi-variance as introduced by Markowitz (1959) as presented in Sections 4.1 and 5.1 or, according to the user's choice, maximizing the certainty equivalent of service provisioning as introduced in Sections 4.2 and 5.2. Employing one of these approaches allows the DA to rank the retrieved topologies according to the risk of SLA violation or the expected utility expressed by the certainty equivalent. The topologies and associated risk/certainty equivalent are returned as a sorted list to the Management Cockpit and subsequently displayed to the user, i. e. the service requester.

In the case of a service intermediary, the selection of suitable SLAs with supplying providers would follow. As these steps are carried out analogously to the choice of risk-minimal (utility maximizing) SLAs with customers, the description is omitted here.

6.3 Design and implementation of the Dependency Analyzer

This section covers the design of the ValueGrids Dependency Analyzer. It shows the interplay of components and their integration as well as a brief introduction to the Dependency Analyzer design.

Figure 6.4 depicts the Dependency Analyzer's components as well as those it interacts with. After being invoked, the Dependency Analyzer calls the Service Repository via a client/server interface to retrieve available services, SLAs and the respective topologies. A query as depicted in Figure 6.3 is passed to the Dependency Analyzer Web service, which wraps the implementation of the Dependency Analyzer. By querying the Service Repository as described in the previous section, the feasible SLA topologies are retrieved. For these topologies, monitoring reports are extracted from the ReportStorage, which is included in the Dependency Analyzer implementation. Based on the monitoring reports, profit-, risk-, or utility-calculations are carried out. This is facilitated through the Matlab Client. Compiled instances of the Matlab[4] code are dis-

[4] `http://www.mathworks.com/` last accessed: October 9, 2011

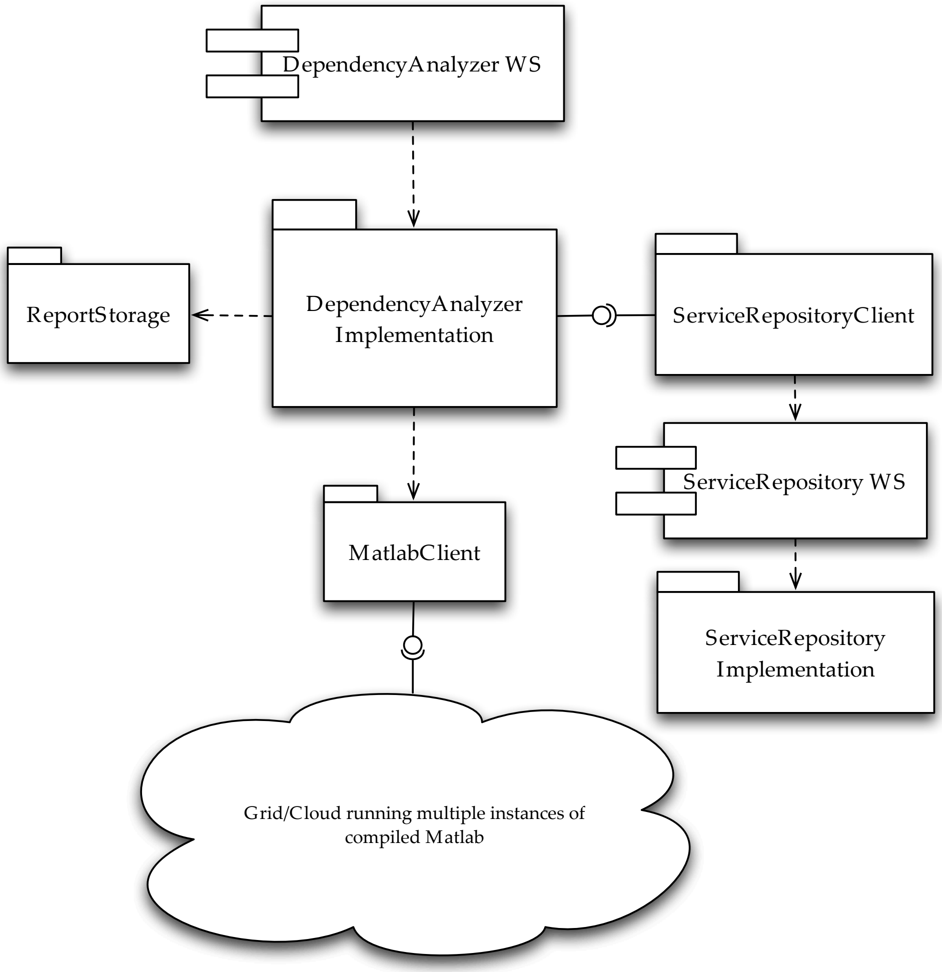

FIGURE 6.4: *Dependency Analyzer Class UML*

tributed and run in a Grid- or Cloud-environment. More details on the implementation of the parallel Matlab backend can be found in the Deliverables of the project ValueGrids, which are available on request[5], and in research (Caton 2009).

The results from the calculations using the MatlabClient are then passed back to the Dependency Analyzer, where they are ranked. The ranked list is then communicated to the requester via the Dependency Analyzer Web service (see Figure 6.3). The functions implemented in Matlab directly reflect the methodology, which was presented in Chapters 4 and 5 in the context of SaaS, Web or Cloud service provisioning as for these types of services, the measurability of QoS attributes of the service is given. Therefore, the ValueGrids implementation highlights the applicability of the approaches presented in this work, which support the decisions made by service providers in dynamic service economies.

The ValueGrids implementation is a proof-of-concept prototypical array of software artifacts that can be used individually or collectively to support decision making about the establishment of SLAs. However, as a proof-of-concept implementation the prototype cannot be simply deployed into any business scenario without extension and adaptation to the particular setting.

[5] `http://www.valuegrids.de` last accessed: October 9, 2011

Chapter 7

Simulation-Based Evaluation

"Always be suspicious of data collection that goes according to plan."

(Patton 2002)

IN Chapters 4 and 5 methods that solve a service provider's and an intermediary's decision problem on SLA establishment were presented. The methods were designed to consider a provider's and an intermediary's properties regarding available resources and expected profit from service provisioning. Furthermore, challenges on decision making for SLA establishment that were raised in Section 3.3 were addressed by the decision support methods. These aspects were covered by the application of an adaptation of portfolio selection (Markowitz 1959). Additionally, the challenges from Section 3.3 address the consideration and expression of risk in the decision on SLA establishment. In the methods that were introduced in Chapters 4 and 5, risk is expressed by the dispersion of expected penalties as measured by the (semi-) variance. The (semi-) variance of expected penalties was calculated from past monitoring observations that state the degree of SLA violation for past periods for each SLA. According to the *law of large numbers*, the more monitoring observations from the past that are available, the more precise is the risk that can be calculated from the observations. In order to select the portfolio of SLAs that incurs the lowest risk of SLA violation (or that exhibits the highest expected utility), a high number of monitoring observations is desirable for each SLA porfolio. The collection of monitoring observations requires that SLAs are established for periods of a certain length and can be monitored. Thus, the more monitoring reports are used in the calculation of the risk of SLA violation, the more time has elapsed for the monitoring of SLAs. For instance, if a provider offers three SLAs and all combinations of the three SLAs are feasible from a provider's technical resources and expected profit constraint, then a provider needs monitoring observations for all seven portfolios of SLAs that result from the three SLAs. If the provider needs to collect 50 monitoring observations for each portfolio, then $7 \cdot 50 = 350$ periods of time need to elapse before enough monitoring observations are available. Depending on the length of the periods, 350 periods can result in hours, days, or even years of monitoring.

As long phases of *exploration* in which portfolios of SLAs are established just for the reason of observing the degree of violation of SLAs are undesirable for decision makers, the question arises from which amount of monitoring observations the approximation of SLA violation risk cannot be improved any more. The approximation of SLA violation risk is the major driver of a provider's and an intermediary's decision on SLA establishment. Therefore, this chapter is dedicated to the evaluation of the applicability of the methods for solving a provider's and an intermediary's decision problem about SLA establishment. The applicability of the methods is evaluated by the identification of the amount of monitoring observations from which SLA violation risk is calculated from which on the approximation of SLA violation risk cannot be improved any further. This is expressed in Research Question 4:

RESEARCH QUESTION 4 ≺REQUIRED AMOUNT OF MONITORING DATA FOR DECISION MAKING≻. *How many monitoring observations are required for the calculation of SLA violation risk in a) a provider's and b) an intermediary's decision on SLA establishment?*

In order to address Research Question 4, the following sections present a simulation-based evaluation. Therefore, Section 7.1 describes the evaluation methodolgy in detail. In Section 7.2, the methods that solve a provider's decision on SLA establishment by selecting the risk-minimal SLA portfolio as introduced in Section 4.1 and by maximizing the expected utility from service provisioning as presented in Section 4.2 are evaluated. Section 7.3 describes the evaluation of the methods that address an intermediary's decision problem about SLA establishment with respect to risk-minimization and utility-maximization. Finally, Section 7.4 discusses the implications of the retrieved results.

7.1 Evaluation Methodology

A straightforward approach for the evaluation of the methods would involve real SLA monitoring observations where the degree of violation has been analyzed. The available monitoring observations could then be split up in parts of different lengths y and from the y observations, the risk of SLA violation as measured by the (semi-) variance could be calculated. Having calculated the risk of SLA violation for each available SLA portfolio, the portfolio incurring the lowest risk would be chosen. The decision on risk-minimizing (or utility-maximizing) SLA portfolios would be carried out for different amounts y of monitoring observations. Using the complete data set would result in the most precise approximation of the risk of SLA violation according to the *law of large numbers*. This decision can be considered as the *optimal* decision and hence, the benchmark. The portfolios selected based upon risk-calculations that employed lower amounts y of monitoring observations can be compared to the benchmark. This enables the derivation of the amount of monitoring observations to make a precise enough risk approximation for a correct choice, i. e. to select the risk-minimizing (or utility-maximizing) SLA portfolio. In order to substantiate the needed amount of observations and to exclude coincidentally correct selections, this method would need to be repeated, requiring lots of monitoring data.

However, monitoring data has to meet Assumptions 3 and 5. Assumption 3 requires constant resources over time in order to keep monitoring data comparable. Real-world scenarios, in which services are provided and monitored, usually are not specially designed to allow for the comparability of monitoring and analytics results. However, monitoring and analytics systems are designed to react on monitoring results and notifications of SLA violation, for instance by adapting the underlying infrastructure. Assumption 5 requires that monitoring results need to be available as one aggregated value that specifies the degree of SLA violation. The challenges of aggregating monitoring results for single SLOs into one value that states the degree of violation for an SLA have been discussed in Section 4.1.1 in depth and concern issues in the monitoring approach, frequency of reports, assessing the preferences of participants of an SLA and the aggregation method. In summary, at the moment, there are approaches for aggregating monitoring data in theory, but few are implemented in practice. In order to apply the methods that were presented in Chapters 4 and 5, a provider's and an intermediary's resources would have to be kept constant and a monitoring and analytics system that facilitates the aggregation of monitoring results into one value that states the degree of SLA violation is required. Currently, these assumptions cannot be met in real-world scenarios and thus, there are no monitoring observations which can be applied for the evaluation of the methods that solve a provider's and an intermediary's decision problem on SLA establishment. As was argued before, the higher the number of monitoring observations is, the better is the approximation of the distribution of SLA violation and with it, the calculated SLA violation risk. Consequently, the most precise representation of SLA violation is achieved if the distribution of SLA violations is known. The evaluation of the methods that solve either a provider's or an intermediary's decision problem is conducted on the basis of artificially created monitoring data. Hence, the simulation-based evaluation exposes an even more precise specification of risk than an evaluation that is based on real monitoring data. Furthermore, the gathering of monitoring observations is more convenient by applying a simulation.

In order to be able to evaluate the methods, even without the availability of real-world monitoring data, monitoring observations that meet Assumptions 3 to 5 have been created artificially with the help of a numerical simulation that was implemented in Java, of which the methodology is described below.

A pre-defined number of SLAs is created, where prices, penalties and costs are set fixed. This approach is chosen as prices and penalties are assumed to be given exogenously and not part of the decision. This supports the focus on the risk-minimizing (or utility-maximizing) selection of SLAs in contrast to pricing decisions. By setting fixed prices and penalties, the effect of varying prices can be neglected as the correctness of decisions only depends on the amount of monitoring observations and the distribution functions of SLA violations. Next, all available portfolios of SLAs are determined. As there is no service repository in place that allows the query for available SLA portfolios, the power-set determines all possible SLA portfolios. The power-set of SLAs reflects all possible combinations of SLAs under the assumption that only one instance of each SLA can be active per period.

For each SLA in each portfolio, a PDF of the degree of violation is assigned.[1] The degree of violation is measured in {0,1}, hence, a PDF with the domain {0,1} is required. The PDF is determined randomly for each problem set. Beta distribution functions, $B(d,e)$, exhibit a domain in {0,1} (Pham-Gia 1989), where d and e are parameters that define the shape of the Beta distribution function. Additionally, Beta distribution functions can be parametrized in a way that the probability of a higher degree of SLA violations increases with an increasing number of SLAs in a portfolio.[2] Without loss of generality,

[1] With the PDF, the CDF is determined.

[2] The absolute probability of SLA violations cannot be influenced, i. e. it is not possible that for one SLA the probability of SLA violations is higher than for another one. However, by altering d, the PDF will exhibit a shift from a high probability for a low degree of SLA violation ($d \leq e$) to a high probability for a high degree of SLA violation ($d \geq e$).

e is fixed to $e = 1$ in order to reduce the degrees of freedom. Setting e differently would have no impact on the shape of the distribution function as long as the manipulation of d was adapted proportionally. Having set $e = 1$, d is determined for each PDF, i. e. for each SLA in each portfolio, as follows: $d = |\gamma| \cdot x$, where x is a random draw from a uniform distribution $U(0, 1.0)$ and $|\gamma|$ is the number of SLAs in the portfolio. Artificial monitoring observations for each SLA in each portfolio are created by drawing y times randomly from the PDFs of SLA violations that were assigned to each SLA in each portfolio. The values for y were set to $5, 10, 20, 50, 100, 200, 500, 1000$, so that the amount of observations increases by a constant factor throughout.[3]

The artificial creation of random-numbers suffers from statistical errors. Therefore, the step of creating an amount of y artificial monitoring observations and calculating semi-variance or certainty equivalent from the observations is repeated $n = 1000$ times. After a repetition of $n = 1000$ times, the amount y of artificial monitoring observations that is created, is increased. To determine the amount of observations that is required to calculate the risk of SLA violation precise enough to make *correct* decisions in a general manner, the above described procedure is repeated for different distribution functions. Each instantiation of the procedure, which is characterized by the specification of particular PDFs for the degree of SLA violation is denoted as problem set. Thus, 150 problem sets are created in order to identify the amount of monitoring observations needed to reliably calculate semi-variance and certainty equivalent. The exact specification of input to the creation of artificial monitoring data is given in Sections 7.2.1 for the provider's case and 7.3.1 for the intermediary's case, respectively.

For the creation of artificial monitoring data, the distribution functions for the degree of SLA violations are determined and thus known in advance. Hence, the risk of SLA violation can be calculated directly from the distribution functions and the risk-minimizing or utility-maximizing SLA portfolio can be identified with respect to the most precise representation of SLA violation risk. This SLA portfolio is the *benchmark* and denotes the *correct choice* for the risk-minimizing (or utility-maximizing) SLA portfolio. The calculation of the SLA violation risk from distribution functions by means of semi-variance and certainty equivalent are illustrated in Sections 7.2.1 and 7.3.1.

For each amount y of monitoring observations, semi-variance and certainty equivalent are calculated from the observations. The SLA portfolio that minimizes the risk of SLA violation (or the one that maximizes the certainty equivalent) as calculated from the observations is chosen. The SLA portfolio that is selected in this step is compared to

[3] The number of observations approximately doubles in each step.

the benchmark portfolio for determining the *correctness* of the choice. Therefore, for each amount of observations y, there are $n = 1000$ values that state the correctness of the observation-based choice for each of the methods that solve a provider's or an intermediary's decision problem on SLA establishment. These values are employed in the evaluation that identifies the required amount of observations, which is described below.

7.1.1 Required Amount of Monitoring Observations

The values for correct decisions are employed that were retrieved from the comparison of the benchmark portfolio to portfolios that were chosen with respect to semi-variance and certainty equivalent that were calculated from artificially created monitoring observations as described before. This comparison created for each method for decision making for each of the 150 problem sets for each amount of monitoring observations under consideration 1000 values that state the correctness of the respective decision. As eight different amounts of observations were considered in the artificial creation of monitoring observations (see Table 7.1), for each problem set 8000 values for the correctness of decisions are available. From these values, for each amount of monitoring observations the fraction of correct decisions can be determined by dividing the number of correct decisions by 1000, the number of repetitions for this amount of observations. This way, a proxy for the probability of making a correct decision by employing this amount of observations in the calculation of the semi-variance or the certainty equivalent is retrieved.

After determining the fraction of correct choices for each amount of observations in each problem set, there are 150 observations for the probability of making a correct choice for each amount of observations. By comparing the probabilities for making correct decisions for two consecutive amounts of observations, the question if the probability of making a correct choice by increasing the number of observations can be answered. The answer to this question is retrieved by applying two-sided unpaired t-tests that evaluate the null-hypothesis of equal means of the probabilities of correct choices. The tests are conducted for consecutive amounts of observations to test the hypothesis that the mean difference between fractions of correct choices is equal to zero (Fay and Proschan 2010). As long as the null-hypothesis can be rejected, there is a statistically significant improvement in the probability of making a correct decision. For the particular amount of observations, from which on the null-hypothesis cannot be rejected any more, the probability of making a correct decision cannot be improved significantly any more and hence, the amount of observations which foster correct decisions is found. The high number of sampling repetition for each problem set and the high number of problem sets assures the robustness of the t-test to violations of the normality assumption (Bridge and Sawilowsky 1999; Ramsey 1980; Sawilowsky and Blair 1992).

7.1.2 Impact of Dispersion of Observations on the Required Amount of Monitoring Observations

The previous section enabled the identification of the amount of monitoring observations required for the decision on SLA establishment. The evaluation considered 150 problem sets and compared the probabilities of making a correct decision for different amounts of observations. The required amounts of monitoring observations give general advice of how many observations need to be collected before applying the methods that minimize the risk of SLA violation or maximize expected utility in the decisions on SLA establishment. This section is dedicated to extend the evaluation by analyzing the impact that dispersion of observations as measured by the variance has on the required amount of observations.

In analogy to the evaluation from the previous section, the values for correct decisions are applied. For each problem set 8000 values for the correctness of decisions are available.

In an approach that is similar to the evaluation from the previous section, an evaluation is carried out that analyzes if the probability of making a correct decision can be increased by the collection of more monitoring observations. In contrast to the evaluation from the previous section, the analysis is carried out for each problem set seperatly. Therefore, the values for correct decisions are not aggregated as before, but considered as they are. However, by comparing the values for correct decisions for consecutive amounts of observations for each problem set, it is possible to evaluate if a higher amount of monitoring observations increases the probability of making a correct decision, which is approximated by the distribution of values of correctness and its mean.

With a Wilcoxon rank sum test, the distributions of correct decisions for different amounts of observations are compared. The null-hypothesis of equal distributions, and hence, equal means, is evaluated in analogy to the pairwise t-tests from the previous section. The pairwise comparison of amounts of observations for which the null-hypothesis cannot be rejected any more identifies the required amount of observations for a particular problem set. Thus, 150 values that describe the required amount of monitoring observations are available.

In order to identify the impact that the dispersion of monitoring observations as measured by the variance of the corresponding PDF of the degree of violation has on the required amount of monitoring data, an ordinal regression is carried out. In this ordinal regression, the required amount of monitoring data that was identified for each problem set is set into context with the maximum variance of the degree of SLA violation from the benchmark portfolio of this problem set. The maximum variance of SLA violation is identified by calculating the variance of SLA violation for each SLA in the portfolio and choosing the maximum one.

In the following sections, the exact specification of the evaluation for the provider's and the intermediary's case, respectively are shown along with its results.

7.2 Provider Decisions

This section is dedicated to the evaluation of the methods that solve a service provider's decision problem in SLA establishment (Chapter 4). In Section 4.1, a method was introduced that facilitates a provider's decision by minimizing the risk of SLA violation as measured by the semi-variance of expected penalties under the consideration of resource and expected profit constraints. Section 4.2 presented an extension to this method that considers a provider's risk-aversion explicitly by employing the maximization of expected utility from service provisioning as objective function. The collection of monitoring data is time-consuming and implies an *exploration* phase to a provider, where different SLA portfolios are established for monitoring the resulting degrees of violation. Therefore, the amount of monitoring data from which on the approximation of risk is precise enough to select the *actually* risk-minimizing (or utility-maximizin) SLA portfolio is the factor that allows the evaluation of applicability of the presented methods. Thus, this section is dedicated to identifying the number of required monitoring observations for provider decisions. Below, first the specific input to the simulation is shown and the calculation of semi-variance and certainty equivalent from the PDFs is explained. Afterwards, the results of the evaluation are presented.

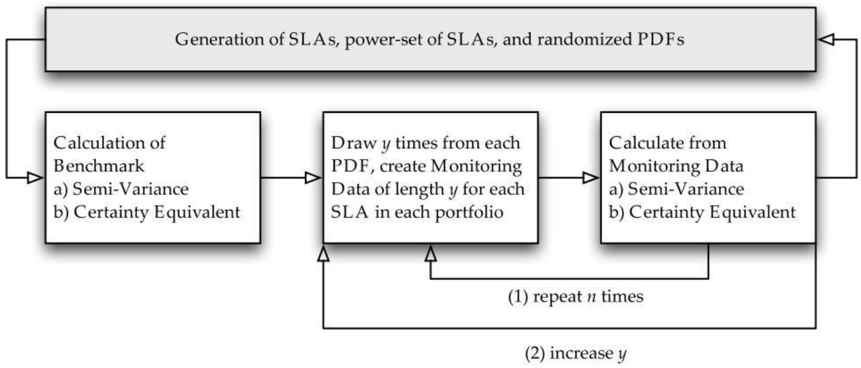

FIGURE 7.1: *Simulation Setup for Creation of Artificial Monitoring Observations*

TABLE 7.1: *Simulation Parameters for Evaluation of Provider's Decision*

Parameter	Value
Number of SLAs	3
$c(s_j)$	$c(s_1) = 1$, $c(s_2) = 1.5$, $c(s_3) = 2$
d	$\vert\gamma\vert \cdot x$
e	1
$f_{\alpha_i}(s_j)$	$f_{\alpha_1}(s_1) = 3.5$, $f_{\alpha_2}(s_2) = 3$, $f_{\alpha_3}(s_3) = 7.5$
μ_{α_i}	$\mu_{\alpha_1} = 2$, $\mu_{\alpha_2} = 1.3$, $\mu_{\alpha_5} = 5$
n	1000
x	$\mathcal{U}(0, 1.0)$
y	5,10,20,50,100,200,500,1000

7.2.1 Simulation Specification and Selection of the Benchmark Portfolio

The number of pre-defined SLAs that are created is set to 3. Without loss of generality, this approach serves to reduce the time complexity of choice (Section 4.3) and hence, the computation time that is required for carrying the evaluation out. Prices and penalties are set to $f_{\alpha_1}(s_1) = 3.5$, $f_{\alpha_2}(s_2) = 3$, and $f_{\alpha_3}(s_3) = 7.5$ and $\mu_{\alpha_1} = 2$, $\mu_{\alpha_2} = 1.3$, and $\mu_{\alpha_5} = 5$. Provider's costs for providing services are set to $c(s_1) = 1$, $c(s_2) = 1.5$, and $c(s_3) = 2$. Prices, penalties and costs are chosen arbitrarily and are considered exogeneously to the decision process. However, a further investigation on the impact of different prices and penalties is considered future work. Table 7.1 summarizes the input to the simulation.

The calculation of semi-variance and certainty equivalent from the PDFs is described below. In a first step, the calculation of expected profit from PDFs is described. This expression is then applied in the calculation of semi-variance and certainty equivalent.

The expected profit is calculated from the PDFs in order to be able to verify the provider's profit constraint by summing prices $f_{\alpha_i}(s_j)$ agreed in SLAs, reducing it by the sum of costs for service execution $c(s_j)$ analogously to Equation 4.9. Finally, the expected penalties are subtracted, which amount to the sum of expected penalties for each of the SLAs in the portfolio.

$$\hat{E}(\pi_\gamma) = \sum_{\alpha_i \in \gamma} \left(f_{\alpha_i}(s_j) - c(s_j)\right) - \sum_{\alpha_i \in \gamma} \hat{E}(\tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i}) \tag{7.1}$$

$$= \sum_{\alpha_i \in \gamma} \left(f_{\alpha_i}(s_j) - c(s_j) - \mu_{\alpha_i} \cdot \int_0^1 \tilde{\lambda}_\gamma^{\alpha_i} \cdot g_\gamma^{\alpha_i}(\tilde{\lambda}_\gamma^{\alpha_i}) \mathrm{d}\tilde{\lambda}_\gamma^{\alpha_i} \right), \tag{7.2}$$

where the expected penalty $\hat{E}(\tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i})$ for one SLA is calculated as the expected value for the degree of violations multiplied by the respective penalty. Let $\tilde{\lambda}_\gamma^{\alpha_i}$ be the degree of SLA violation of SLA α_i in portfolio γ, $g_\gamma^{\alpha_i}(\tilde{\lambda}_\gamma^{\alpha_i})$ the probability density function of $\tilde{\lambda}_\gamma^{\alpha_i}$, and μ_{α_i} the penalty in SLA α_i, as before. The expected penalty for SLA α_i is defined as the expected value of $\tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i}$:

$$\hat{E}(\tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i}) = \int_0^1 \tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i} \cdot g_\gamma^{\alpha_i}(\tilde{\lambda}_\gamma^{\alpha_i}) \mathrm{d}\tilde{\lambda}_\gamma^{\alpha_i}. \tag{7.3}$$

This expression of expected profit and expected penalties is employed in the calculation of semi-variance and certainty equivalent. The semi-variance of due penalties for a portfolio of SLAs is calculated as follows. The sum of penalties, of which each is multiplied by the respective degree of SLA violation $\tilde{\lambda}_\gamma^{\alpha_i}$, is compared to the sum of penalties from the expected penalties $\sum_{\alpha_i \in \gamma} \hat{E}(\tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i})$. This difference is squared for those cases, which are bigger than 0 and multiplied by the common density function of all of the degrees of violation. As the PDFs for the degree of violation are statistically independent, the common density function results in the product of the single density functions $\prod_{\alpha_i \in \gamma} g_\gamma^{\alpha_i}(\tilde{\lambda}_\gamma^{\alpha_i})$.

A multiple integral from 0 through 1 of the above term determines the semi-variance of due penalties, as only those deviations that are worse than expected are taken into account:

$$\hat{S}_E(\mu_\gamma) = \tag{7.4}$$

$$\int_0^1 \int_0^1 \cdots \int_0^1 \left(\max\left\{ 0; \sum_{\alpha_i \in \gamma} \mu_{\alpha_i} \cdot \tilde{\lambda}_\gamma^{\alpha_i} - \sum_{\alpha_i \in \gamma} \hat{E}(\tilde{\lambda}_\gamma^{\alpha_i} \cdot \mu_{\alpha_i}) \right\} \right)^2 \cdot \prod_{\alpha_i \in \gamma} g_\gamma^{\alpha_i}(\tilde{\lambda}_\gamma^{\alpha_i}) \mathrm{d}\tilde{\lambda}_\gamma^{\alpha_n} \cdots \mathrm{d}\tilde{\lambda}_\gamma^{\alpha_2} \mathrm{d}\tilde{\lambda}_\gamma^{\alpha_1},$$

where $g_{\gamma}^{\alpha_i}(\tilde{\lambda}_{\gamma}^{\alpha_i})$ denotes the PDF of the degree of violation for SLA α_i in portfolio γ. The semi-variance is computed from the PDF by adapting the calculation of the variance of a PDF (see Gujarati and Porter 2006) so that only those deviations that are worse than the expected value are taken into account. According to the lowest value for $\hat{S}_E(\mu_{\gamma,t})$, the benchmark portfolio is chosen.

The certainty equivalent that serves for the decision about SLA establishment by maximizing expected utility is calculated from the PDFs as follows:

$$CE(\pi_{\gamma}) = \hat{E}(\pi_{\gamma}) - \frac{1}{2} \cdot ARA(\pi_{\gamma}) \cdot \hat{Var}(\pi_{\gamma}) \tag{7.5}$$

denotes the certainty equivalent for portfolio γ, where $\hat{E}(\pi_{\gamma})$ is the expected profit from Equation (7.1), $ARA(\pi_{\gamma})$ denotes the provider's coefficient of absolute risk-aversion for the value of the expected profit and $\hat{Var}(\pi_{\gamma})$ denotes the variance of profit.

The variance of the portfolio γ's profit is calculated as the variance of the sum of single SLA profits:

$$Var(\pi_{\gamma}) = Var(\sum_{\alpha_i \in \gamma} \pi_{\alpha_i}) = \sum_{\alpha_i \in \gamma} Var(\pi_{\alpha_i}) + \sum_{\alpha_i \in \gamma} \sum_{\alpha_j \in \gamma} Cov(\pi_{\alpha_i}, \pi_{\alpha_j}). \tag{7.6}$$

Because all $g_{\gamma}^{\alpha_i}(\tilde{\lambda}_{\gamma}^{\alpha_i})$ are stochastically independent, the covariances of profit are zero, that is $Cov(\pi_{\alpha_i}, \pi_{\alpha_j}) = 0, \forall i, j,\ i \neq j$ and consequently, the variance of the sum of profits equals the sum of variances of profit.

The variance of profit for one SLA is calculated as the variance of the SLA violation PDF is weighted by prices, costs and penalties

$$\hat{Var}(\pi_{\alpha_i}) = Var(f_{\alpha_i}(s_j) - c(s_j) - \mu_{\alpha_i} \cdot \tilde{\lambda}_{\gamma}^{\alpha_i}), \tag{7.7}$$

$$= (\mu_{\alpha_i})^2 \cdot Var(\tilde{\lambda}_{\gamma}^{\alpha_i}).$$

$$= (\mu_{\alpha_i})^2 \cdot \int_0^1 \left(\tilde{\lambda}_{\gamma}^{\alpha_i} - E(\tilde{\lambda}_{\gamma}^{\alpha_i})\right)^2 g_{\gamma}^{\alpha_i}(\tilde{\lambda}_{\gamma}^{\alpha_i}) \mathrm{d}\tilde{\lambda}_{\gamma}^{\alpha_i}. \tag{7.8}$$

Using the characteristics of the Beta distribution function (see Papoulis 2002), the variance of the degree of violation can be described dependent on the parameter $d_{\gamma}^{\alpha_i}$ as follows

$$Var(\tilde{\lambda}_{\gamma}^{\alpha_i}) = \frac{d_{\gamma}^{\alpha_i}}{(d_{\gamma}^{\alpha_i} + 1)^2 (d_{\gamma}^{\alpha_i} + 2)}. \tag{7.9}$$

Substituting this result into Equations 7.6 and 7.7 one obtains

$$\begin{aligned} Var(\pi_\gamma) &= \sum_{\alpha_i \in \gamma} Var(\pi_{\alpha_i}), \\ &= \sum_{\alpha_i \in \gamma} (\mu_{\alpha_i})^2 \cdot Var(\tilde{\lambda}_\gamma^{\alpha_i}), \\ &= \sum_{\alpha_i \in \gamma} (\mu_{\alpha_i})^2 \cdot \left(\frac{d_\gamma^{\alpha_i}}{(d_\gamma^{\alpha_i}+1)^2 (d_\gamma^{\alpha_i}+2)} \right). \end{aligned} \tag{7.10}$$

By applying Equation 7.10 in Equation 7.5, the certainty equivalent is obtained from the PDFs of the degree of SLA violation and the benchmark SLA portfolio that maximizes the certainty equivalent can be chosen.

This section provided a thorough introduction to the creation of artificial monitoring observations and the selection of benchmark SLA portfolios, which are the first to steps in the evaluation as depicted in Figure 7.1. As next steps, the benchmark SLA portfolios are compared to those SLA portfolios that are chosen by minimizing the risk of SLA violation that is calculated from artificial monitoring observations and those SLA portfolios that are chosen by the maximization of the certainty equivalent that is computed from artificial monitoring observations, the *correctness* of choices is determined. The following sections describe the identification of the amount of observations that is required for making correct choices.

7.2.2 Required Amount of Monitoring Observations for Provider's Decision

This section illustrates the results that are obtained from the evaluation as described in Section 7.1.1. First, the results of the evaluation of the provider's risk-minimizing decision are presented and discussed. Afterwards, the evaluation of the utility-maximizing decision, which is based on the certainty equivalent, is shown.

The impact which the amount of available monitoring data has on the correctness of the provider's risk-minimizing decision is highlighted in Figure 7.2. With an increasing amount of observations, which is depicted on the x-Axis, the fraction of correct choices increases. The bars denote the mean fraction of correct choices, averaged over the 150 different problem sets. Figure 7.2 exhibits a decreasing marginal improvement of the fraction of correct choices with a growing number of observations. Table 7.2 shows the

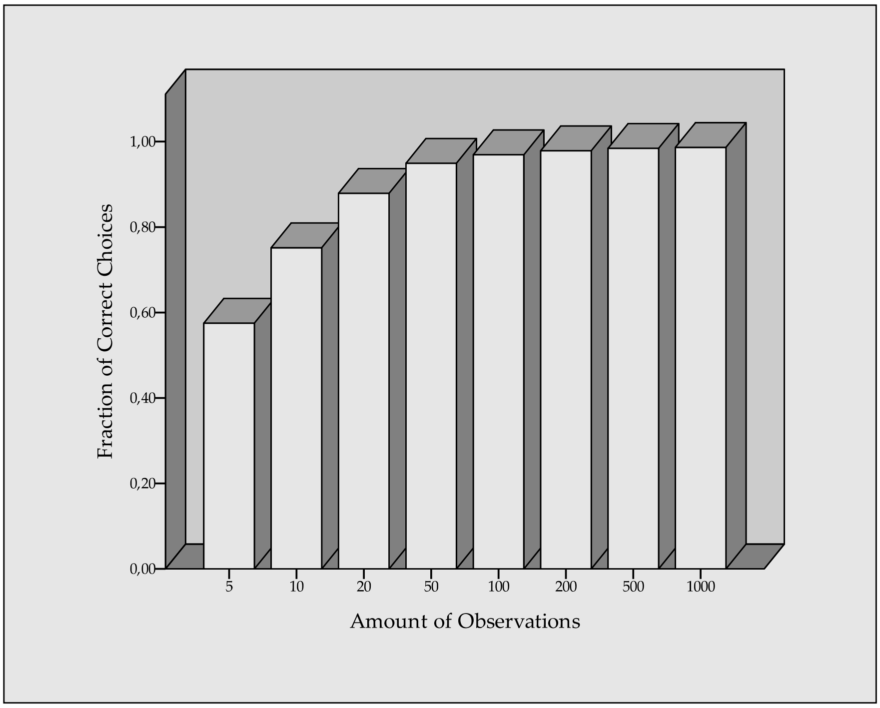

FIGURE 7.2: *Fraction of Correct Choices for Provider's semi-variance -minimizing Decision*

results of the pairwise t-tests.[4] The results indicate that the null-hypothesis of equal means can be rejected when comparing the amounts of observations (5,10), (10,20), and (20,50). For higher amounts of observations, the null-hypothesis cannot be rejected any more. This implies that the minimal amount of observations that a decision maker should have available for each of the portfolios is 50 in order to have a high probability of making correct decisions.

TABLE 7.2: *Evaluation of provider's risk-minimizing decision: Fraction of Correct Choices*

Compared amounts of observations	Mean Δ between fraction of correct choices
5,10	0.177^{***}
10,20	0.127^{***}
20,50	0.070^{***}
50,100	0.020
100,200	0.009
200,500	0.006
500,1000	0.002

The same procedure is applied to a provider's decision, which is made by maximizing the certainty equivalent. Figure 7.3 shows that the marginal improvement of the

[4]Note, that *** states significance to the 0.1% level.

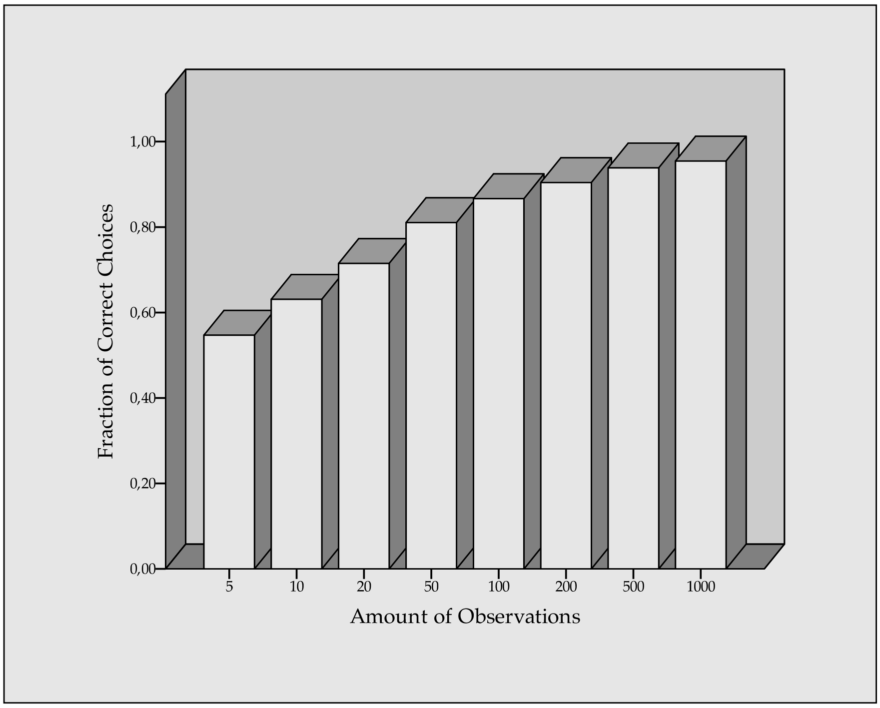

FIGURE 7.3: *Fraction of Correct Choices for Provider's Certainty Equivalent-maximizing Decision*

fraction of correct choices decreases with a growing number of observations. Analogously to the results of the risk-minimizing decision, the threshold value of the amount of observations is determined with the help of unpaired two-sided t-tests as described before. Table 7.3 shows the results, where *** denotes significance to the 0.1% level. The results show that the null-hypothesis of equal means can be rejected for the comparison of the amounts of observations 5 and 10. For higher amounts of observations, there is no statistically significant improvement. This implies that the decision maker is advised to collect at least 10 observations for each of the portfolios, as collecting more data does not improve the fraction of correct choices significantly.

7.2.3 Impact of Dispersion of Observations on the Required Amount of Monitoring Observations

This section presents the results that are obtained from the two-step evaluation that was described in Section 7.1.2. In the first step, the required amount of monitoring observations is determined seperatly for each problem set. Performing the statistical analysis with a Wilcoxon rank-sum test as described in Section 7.1.2 evaluates the null-

TABLE 7.3: *Evaluation of provider's utility-maximizing decision: Fraction of Correct Choices*

Compared amounts of observations	Mean Δ between fraction of correct choices
5,10	0.059***
10,20	0.050
20,50	0.052
50,100	0.020
100,200	0.013
200,500	0.008
500,1000	0.000

hypothesis of equal means of the difference in ranks for different amounts of monitoring observations. The results of the evaluation are for each problem set are shown in the Appendix in Tables A.1 and A.3.[5] If the null-hypothesis cannot be rejected any more, the threshold amount of monitoring observations is found. Below, the impact of the dispersion of monitoring observations on the required amount of monitoring observations is examined.

For this analysis, the maximum dispersion of monitoring observations as measured by the variance of the degree of SLA violation of the benchmark portfolio is identified for each problem set. Alternatively to selecting the maximum variance, the average variance of the degree of SLA violation of the benchmark portfolio could have been calculated. This approach would have smoothed the properties of a portfolio of SLAs and treated cases with similar variances for all SLAs equal to cases with very different variances but the same mean. Arguably, selecting the maximum variance of a portfolio does not take into account the properties of all PDFs, but it takes the property into account, which is assumed to have the strongest impact on the required amount of monitoring data: the variance of single PDFs of the degree of SLA violation. Having retrieved the maximum variance for each problem set, it can be set into context with the respective required amount of monitoring observations. This is illustrated in Table A.2 for the risk-minimizing decision and in Table A.4 for the utility-maximizing decision. The impact of the variance on the required amount of monitoring observations is evaluated with an ordinal regression. In the following, first the results for the risk-minimizing decision are described and afterwards, the results for the utility-maximizing decision are presented.

An ordinal regression that analyses the impact of the independent variable *variance* on

[5]For the results, *** denotes significance to the 0.001 level, ** states significance to the 0.01 level, and * shows significance to the 0.05 level, and hence, that the null hypothesis can be rejected to the respective level.

TABLE 7.4: *Impact of Variance on Required Amount of Observations for Provider's Risk-minimizing Decision*

Parameter Estimates						
					95% Confidence Interval	
		Estimate	Std. Error	Sig.	Lower Bd.	Upper Bd.
Threshold	Amount = 10	-3.747	1.091	.001	-5.885	-1.608
	Amount = 20	-1.691	.580	.004	-2.828	-.555
	Amount = 50	1.924	.565	.001	.817	3.032
	Amount = 100	2.858	.588	.000	1.706	4.010
	Amount = 200	3.536	.613	.000	2.335	4.737
	Amount = 500	4.660	.698	.000	3.292	6.029
Location	Variance	19.710	7.281	.007	5.440	33.980

the dependent variable *amount of observations* reveals that the maximum variance in the portfolio has a significant ($\alpha = 0.01$) impact on the required length of history. Coefficients and results can be found in Table 7.4, whereas the downmost row shows the coefficient (19.710) for the impact of the variance on the required amount of observations with a p-value of 0.07 (column "Sig.").

Analogoulsy, the impact of the maximum variance on the required amount observations for the utility-maximizing decision is identified by means of an ordinal regression, where the independent variable is *variance* and the dependent variable is the *amount of observations*. The ordinal regression reveals that the maximum variance in the portfolio has a positive impact (last row, Estimate = 45.101), which is significant to the 0.01 level on the required amount of observations (last row, column "Sig.").

7.3 Intermediary Decisions

In the previous section, the amount of monitoring observations was identified that is required for a service provider's decision on the establishment of SLAs by minimizing the risk of SLA violation or maximizing the expected utility from service provisioning. In this section, the methods that solve an intermediary's decision problem for the establishment of SLAs with customers and providers are evaluated. The methods that solve an intermediary's decision problem are an extension of the methods that solve a provider's decision problem. While in a provider's decision problem exclusively the establishment of SLAs with customers about the provisioning of services is considered,

TABLE 7.5: *Impact of Variance on Required Amount of Observations for Provider's Utility-maximizing Decision*

Parameter Estimates						
					95% Confidence Interval	
		Estimate	Std. Error	Sig.	Lower Bd.	Upper Bd.
Threshold	Amount = 5	-1.402	1.533	.360	-4.407	1.602
	Amount = 10	1.085	1.223	.375	-1.312	3.482
	Amount = 20	1.270	1.220	.298	-1.121	3.661
	Amount = 50	2.195	1.220	.072	-.195	4.586
	Amount = 100	3.212	1.235	.009	.791	5.633
	Amount = 200	4.004	1.250	.001	1.554	6.453
	Amount = 500	5.047	1.268	.000	2.563	7.532
Location	Variance	45.101	14.936	.003	15.827	74.375

an intermediary's decision concerns the establishment of both, SLAs with customers and SLAs with service providers. In more detail, the methods that were introduced in Chapter 5 are split up in two parts. Applying the method that minimizes an intermediary's risk (Section 5.1), first the SLA portfolio with customers that minimizes the risk of SLA violation is selected. In a second step, the portfolio of SLAs with providers that minimizes the risk of SLA violation by providers is chosen. Analogously, by employing the method that maximizes an intermediary's expected utility as presented in Section 5.2, first the SLA portfolio with customers is chosen which maximizes an intermediary's expected utility from service provisioning. In a second step, the portfolio of SLAs with providers that supply services is chosen that maximizes the expected utility from providing and procuring services.

The first step of the methods for decision making for an intermediary equals a provider's decision on SLA establishment. The methods that solve a provider's decision about SLA establishment have already been evaluated in the previous section. The amounts of observations that are required for a provider's decision about SLA establishment that result from the evaluation of the risk-minimizing and the utility-maximizing decision are applied in the evaluation of an intermediary's decision problem. Therefore, in this section the amount of monitoring observations that is required for an intermediary's decision about the SLA establishment with providers is evaluated. Below, the creation of artificial monitoring data for the intermediary's case and the selection of benchmark portfolios is described in detail followed by the results of the evaluation.

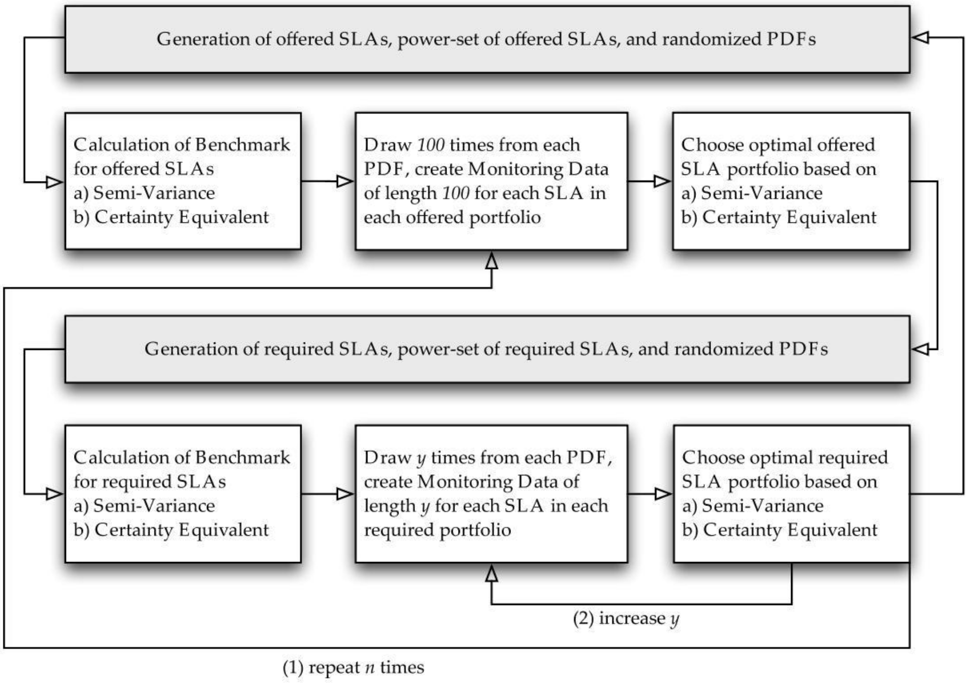

FIGURE 7.4: *Simulation Setup for Creation of Artificial Monitoring Observations for Intermediary's Decision*

7.3.1 Simulation Specification and Selection of the Benchmark Portfolio

Figure 7.4 depicts the process of creating artificial monitoring data and determining the correctness of an observation-based choice in the intermediary's case. Its building blocks are described in detail below.

In a first step, 100 monitoring observations for SLA violations in customer SLAs are created artificially. This amount is chosen as the evaluation from Section 7.2 suggested to collect 50 observations for the semi-variance-minimizing decision and to increase the number for higher variances. The number of observations is a conservative way to increase the probability of making a correct decision in the risk-minimizing as well as in the utility-maximizing decision. Prices, penalties and costs are set to $f_{\alpha_1}(s_1) = 8$, $f_{\alpha_2}(s_2) = 7.5$, and $f_{\alpha_3}(s_3) = 9.5$ and $\mu_{\alpha_1} = 2$, $\mu_{\alpha_2} = 1.3$, and $\mu_{\alpha_3} = 5$. The intermediary's costs for providing services are determined to be $c(s_1) = 1$, $c(s_2) = 1.5$, and $c(s_3) = 2$. The number of SLAs was 3, as before, in order to reduce the computational complexity. From the 100 artificially created observations, semi-variance and certainty equivalent for the decision about the establishment of customer SLAs are calculated as introduced in Sections 4.1.1 and 4.2.1. The portfolios that minimize the semi-variance and that maximize the certainty equivalent are chosen and compared to the benchmark *customer* SLA portfolio.[6]

In the second step, for the 3 procured SLAs,[7] an amount y of monitoring observations for SLA violations by providers is created artificially. Prices and penalties are set to $f_{\beta_1}(s_1) = 1$, $f_{\beta_2}(s_2) = 1.5$, and $f_{\beta_3}(s_3) = 2$ and $\mu_{\beta_1} = 0.5$, $\mu_{\beta_2} = 1$, and $\mu_{\beta_3} = 1.5$. Table 7.6 summarizes the input to the artificial creation of an intermediary's monitoring observations.

TABLE 7.6: *Simulation Parameters for Evaluation of Intermediary's Decision*

Parameter	Value
Number of offered SLAs	3
Number of required SLAs	3
$c(s_j)$	$c(s_1) = 1$, $c(s_2) = 1.5$, $c(s_3) = 2$
d	$\|\gamma\| \cdot x$
e	1
$f_{\alpha_i}(s_j)$	$f_{\alpha_1}(s_1) = 8$, $f_{\alpha_2}(s_2) = 7.5$, $f_{\alpha_3}(s_3) = 9.5$
$f_{\beta_i}(s_j)$	$f_{\beta_1}(s_1) = 1$, $f_{\beta_2}(s_2) = 1.5$, $f_{\beta_3}(s_3) = 2$
μ_{α_i}	$\mu_{\alpha_1} = 2$, $\mu_{\alpha_2} = 1.3$, $\mu_{\alpha_3} = 5$
μ_{β_i}	$\mu_{\beta_1} = 0.5$, $\mu_{\beta_2} = 1$, $\mu_{\beta_3} = 1.5$
n	1000
x	$\mathcal{U}\{0,1.0\}$
y	5,10,20,50,100,200,500,1000

The calculation of semi-variance and certainty equivalent from the PDFs that is required to choose the benchmark portfolios, is described below. For a provider's decision on SLA establishment with customers (equal to an intermediary's decision on SLA establishment with customers) the calculation of semi-variance and certainty equivalent from the distribution functions have been introduced in Section 7.2.1. In order to choose the benchmark portfolio of SLAs with providers, the calculation of the semi-variance of the degree of SLA violation and the certainty equivalent that relate to the expressions from Equations (5.8) and (5.28) needs to be introduced. Therefore, in a first step, the calculation of expected profit from PDFs is explained. This expression is employed to verify an intermediary's expected profit constraint and used in the calculation of the certainty equivalent of an intermediary's utility-maximizing decision.

The expected profit is calculated from the PDFs by calculating the expected profit of the optimal offered SLA portfolio according to Equation (7.1). From this, the sum of

[6] Due to the chosen amount of observations, the observation-based chosen SLA portfolio and the benchmark portfolio are the same.

[7] Again, for reasons of reducing the computational complexity.

prices for the required SLAs is substracted and the expected penalties for the supplied portfolio are added.

(7.11)

$$
\begin{aligned}
\hat{E}(\pi_{\gamma^*,\delta}) &= \sum_{\alpha_i \in \gamma^*} \left(f_{\alpha_i}(s_j) - c(s_j) \right) - \sum_{\alpha_i \in \gamma^*} \hat{E}(\tilde{\lambda}_{\gamma^*}^{\alpha_i} \cdot \mu_{\alpha_i}) - \sum_{\beta_i \in \delta} f_{\beta_i}(s_j) + \sum_{\beta_i \in \delta} \hat{E}(\tilde{\lambda}_{\delta}^{\beta_i} \cdot \mu_{\beta_i}), \\
&= \sum_{\alpha_i \in \gamma^*} \left(f_{\alpha_i}(s_j) - c(s_j) - \mu_{\alpha_i} \cdot \int_0^1 \tilde{\lambda}_{\gamma^*}^{\alpha_i} \cdot g_{\gamma^*}^{\alpha_i}(\tilde{\lambda}_{\gamma^*}^{\alpha_i}) \mathrm{d}\tilde{\lambda}_{\gamma^*}^{\alpha_i} \right) \\
&- \sum_{\beta_i \in \delta} \left(f_{\beta_i}(s_j) - \mu_{\alpha_i} \cdot \int_0^1 \tilde{\lambda}_{\delta}^{\beta_i} \cdot g_{\delta}^{\beta_i}(\tilde{\lambda}_{\delta}^{\beta_i}) \mathrm{d}\tilde{\lambda}_{\delta}^{\beta_i} \right), \\
&= \hat{E}(\pi_{\gamma^*}) - \sum_{\beta_i \in \delta} \left(f_{\beta_i}(s_j) - \mu_{\alpha_i} \cdot \int_0^1 \tilde{\lambda}_{\delta}^{\beta_i} \cdot g_{\delta}^{\beta_i}(\tilde{\lambda}_{\delta}^{\beta_i}) \mathrm{d}\tilde{\lambda}_{\delta}^{\beta_i} \right),
\end{aligned}
$$

where $\hat{E}(\tilde{\lambda}_{\delta}^{\beta_i} \cdot \mu_{\beta_i})$ denotes the expected penalty for one required SLA and is calculated as the expected degree of violation for an SLA multiplied by the corresponding penalty. The degree of violation $\tilde{\lambda}_{\delta}^{\beta_i}$ and the probability density function $g_{\delta}^{\beta_i}(\tilde{\lambda}_{\delta}^{\beta_i})$ are defined as before. The expected penalty for SLA β_i is defined as the expected value of $\tilde{\lambda}_{\delta}^{\beta_i} \cdot \mu_{\beta_i}$:

(7.12)
$$\hat{E}(\tilde{\lambda}_{\delta}^{\beta_i} \cdot \mu_{\beta_i}) = \int_0^1 \tilde{\lambda}_{\delta}^{\beta_i} \cdot \mu_{\beta_i} \cdot g_{\delta}^{\beta_i}(\tilde{\lambda}_{\delta}^{\beta_i}) \mathrm{d}\tilde{\lambda}_{\delta}^{\beta_i}.$$

For the second constraint in an intermediary's risk-based decision, the mean degree of SLA violation of an SLA portfolio is calculated. Therefore, the average of the expected degrees of violation for all SLAs in the *provider* SLA portfolio δ is retrieved. As before, the expected value of the degree of violation for one SLA amounts to $\int_0^1 \tilde{\lambda}_{\delta}^{\beta_i} \cdot g_{\delta}^{\beta_i}(\tilde{\lambda}_{\delta}^{\beta_i}) \mathrm{d}\tilde{\lambda}_{\delta}^{\beta_i}$. Consequently, the average degree of violation for a portfolio results in

(7.13)
$$\hat{E}(\tilde{\lambda}_{\delta}) = \frac{\sum_{\beta \in \delta} \int_0^1 \tilde{\lambda}_{\delta}^{\beta_i} \cdot g_{\delta}^{\beta_i}(\tilde{\lambda}_{\delta}^{\beta_i}) \mathrm{d}\tilde{\lambda}_{\delta}^{\beta_i}}{|\delta|}.$$

After introducing the calculation of the expected profit and the average degree of SLA violation constraint, the calculations of semi-variance and certainty equivalent from PDFs may be introduced. The semi-variance of *provider* SLA violations is computed differently than the semi-variance that was introduced in Section 7.3.1. In order to account for the deviations, which are worse than average, the calculation of the variance of a PDF (Gujarati and Porter 2006) is adapted as before, by integration over the piece-wise defined squaring function. However, the semi-variance is not weighted by the supplying providers' penalties, as these would negatively influence the intermediary's decision. The higher the penalties, the higher the intermediary's utility; such an approach would not capture the intermediary's aversion towards providers' SLA violation. As a consequence, the usual approach which is applied in the calculation of a variance, that is, squaring the deviations, would not lead to the required expression of putting an emphasis on higher deviations. The semi-variance of the degree of SLA violation is calculated as follows:

$$(7.14)\quad \hat{S}_E(\lambda_\delta) = \int_0^1 \int_0^1 \cdots \int_0^1 \left(\max\left\{ 0; \frac{\sum_{\beta_i \in \delta} \tilde{\lambda}_\delta^{\beta_i}}{|\delta|} - \hat{E}(\tilde{\lambda}_\delta) \right\} \right)^2 \cdot \prod_{\beta_i \in \delta} g_\delta^{\beta_i}(\tilde{\lambda}_\delta^{\beta_i})\, \mathrm{d}\tilde{\lambda}_\delta^{\beta_n} \cdots \mathrm{d}\tilde{\lambda}_\delta^{\beta_2}\, \mathrm{d}\tilde{\lambda}_\delta^{\beta_1},$$

where $g_\delta^{\beta_i}(\tilde{\lambda}_\delta^{\beta_i})$ denotes the PDF of the degree of violation for SLA β_i in portfolio δ and $\hat{E}(\tilde{\lambda}_\delta)$ denotes the expected degree of violation of the portfolio δ.

The average degree of violation deviation is compared to the expected degree of violation. This difference is then squared, for the values greater than zero and is finally multiplied by the common density function of all of the degrees of violation. As the PDFs for the degree of violation are statistically independent, the common density function results in the product of the single density functions. A multiple integral from 0 through 1 of the above term determines the semi-variance of SLA violations through supplying providers, as only those deviations that are worse than expected are taken into account. According to the lowest value for $\hat{S}_E(\lambda_\delta)$, the benchmark portfolio is chosen.

The benchmark for the utility-maximizing decision that was presented in Section 5.2 is calculated absolutely equivalent to the benchmark in the provider's case in Section 7.2.1 as the certainty equivalent from Equation (5.21) with respect to the PDFs for SLA violation.

$$(7.15)\qquad CE(\pi_{\gamma^*,\delta}) = \hat{E}(\pi_{\gamma^*,\delta}) - \frac{1}{2} \cdot ARA(\pi_{\gamma^*,\delta}) \cdot \hat{Var}(\pi_{\gamma^*,\delta})$$

is the certainty equivalent in the case of the benchmark *customer* SLA portfolio γ^*, which is combined with the provider SLA portfolio δ. In this case, the total expected profit $\hat{E}(\pi_{\gamma^*,\delta})$ is calculated according to Equation (7.11) and the intermediary's coefficient of absolute risk aversion is $ARA(\pi_{\gamma^*,\delta})$, whereas $\hat{Var}(\pi_{\gamma^*,\delta})$ denotes the variance of the total profit.

The variance of profit from offering and procuring services amounts to the variance of the sum of single SLA profits, that is, the profit from offering the benchmark *customer* SLA portfolio γ^* and procuring a single service with SLA β_i:

$$Var(\pi_{\gamma^*,\delta}) = Var(\sum_{\beta_i \in \delta} \pi_{\gamma^*,\beta_i}) = \sum_{\beta_i \in \delta} Var(\pi_{\gamma^*,\beta_i}) + \sum_{\beta_i \in \delta} \sum_{\beta_j \in \delta} Cov(\pi_{\gamma^*,\beta_i}, \pi_{\gamma^*,\beta_j}). \tag{7.16}$$

As in the case of the provider's decision, the variance of the sum of profits equals the sum of the variances of profit due to the stochastical independence of all $g_\delta^{\beta_i}(\tilde{\lambda}_\delta^{\beta_i})$. Applying the same steps as in Section 7.2.1 as well as the properties of the Beta distribution function (see Papoulis 2002), the variance of the degree of violation for one SLA is calculated with respect to the Beta PDF's factor $d_\delta^{\beta_i}$ as follows:

$$Var(\tilde{\lambda}_\delta^{\beta_i}) = \frac{d_\delta^{\beta_i}}{(d_\delta^{\beta_i}+1)^2(d_\delta^{\beta_i}+2)}. \tag{7.17}$$

Applying this result in the calculation of the variance of absolute profit when offering the benchmark *customer* SLA portfolio γ^* and establishing the provider SLA portfolio δ analogously to Section 7.2.1 results in

$$\begin{aligned} Var(\pi_{\gamma^*,\delta}) &= \sum_{\beta_i \in \delta} Var(\pi_{\gamma^*,\beta_i}), \\ &= \sum_{\beta_i \in \delta} (\mu_{\beta_i})^2 \cdot Var(\tilde{\lambda}_\delta^{\beta_i}), \\ &= \sum_{\beta_i \in \delta} (\mu_{\beta_i})^2 \cdot \left(\frac{d_\delta^{\beta_i}}{(d_\delta^{\beta_i}+1)^2(d_\delta^{\beta_i}+2)} \right). \end{aligned} \tag{7.18}$$

By applying Equation 7.18 to Equation 7.15, the certainty equivalent is obtained from the previously specified PDFs of the degree of SLA violation. The benchmark portfolio of *provider* SLAs is chosen as the portfolio that maximizes the certainty equivalent.

The benchmark portfolios are then compared to the portfolios that are chosen by the minimization of the risk of SLA violation that is calculated from artificial monitoring observations and those SLA portfolios that are chosen by the maximization of the certainty equivalent from artificial monitoring data (see Figure 7.4). This leads to the determination of the correctness of the observation-based choices. In this vein, for each of the 150 problem sets and each of the 8 different amounts of observations y, 1000 val-

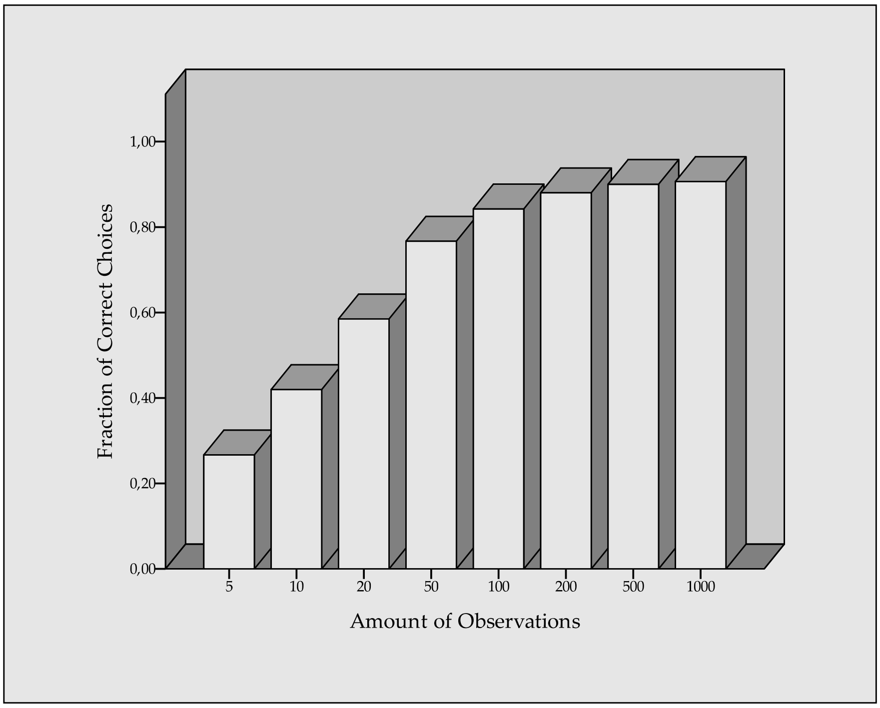

FIGURE 7.5: *Fraction of Correct Choices for Intermediary's semi-variance-minimizing Decision*

ues for the correctness of the observation-based choices are available. These values are employed in the evaluation that was described in Section 7.1, of which the results are described in the following sections.

7.3.2 Required Amount of Monitoring Observations for an Intermediary's Decision

This section presents the results that the evaluation as described in Section 7.1.1 produced. First, the results for the risk-minimizing decision are presented and discussed. Afterwards, the evaluation of the utility-maximizing decision, which is based on the certainty equivalent, is shown.

The impact that the available amount of monitoring observations has on the correctness of an intermediary's risk-minimizing decision on SLA establishment with providers is depicted in Figure 7.5. With an increasing amount of observations, which is shown on the x-Axis, the fraction of the intermediary's correct choices increases.

Table 7.7 shows the results of the pairwise t-tests.[8] The results indicate that the null-hypothesis of equal means can be rejected when comparing the amounts of observations (5,10), (10,20), (20,50), and (50,100). For higher amounts of observations, the null-hypothesis cannot be rejected any more. This implies that the minimal amount of observations that a decision maker should have available for each of the portfolios is 100 in order to have a high probability of making correct decisions.

TABLE 7.7: *Evaluation of intermediary's risk-minimizing decision: Fraction of Correct Choices*

Compared lengths of history	Mean Δ between fraction of correct choices
5,10	0.1532***
10,20	0.1651***
20,50	0.1821***
50,100	0.0754**
100,200	0.0378
200,500	0.0198
500,1000	0.0064

The method that solves an intermediary's decision problem on SLA establishment with providers by maximizing the certainty equivalent is evaluated analogously. Figure 7.6 shows that the marginal improvement of the fraction of correct choices decreases with an increasing amount of observations. Table 7.8 shows the results, where *** states significance to the 0.1% level and * denotes significance to the 5% level. The results show that the null-hypothesis of equal means can be rejected for the comparison of the amounts of observations 5 and 10, and 10 and 20. For higher amounts of observations, there is no statistically significant improvement. This implies that the decision maker is advised to collect at least 20 observations for each of the provider portfolios in order to have a high probability for a correct decision.

TABLE 7.8: *Evaluation of intermediary's utility-based decision: Fraction of Correct Choices*

Compared lengths of history	Mean Δ between fraction of correct choices
5,10	0.0272***
10,20	0.0101*
20,50	0.0040
50,100	0.0015
100,200	0.0004
200,500	0.0001
500,1000	0.0000

[8]Note, that *** states significance to the 0.1% level and ** denotes significance to the 1% level.

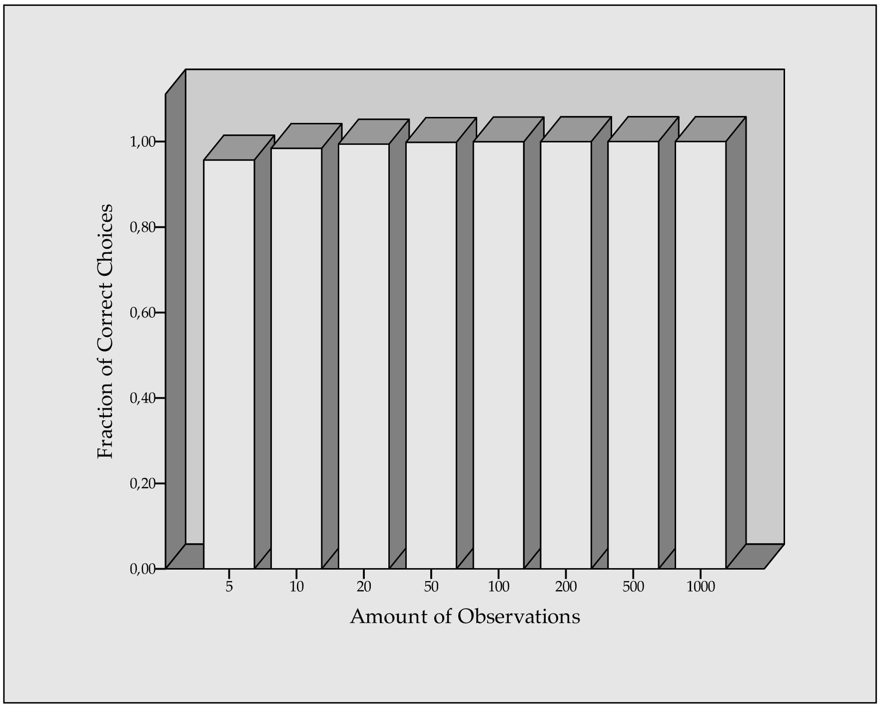

FIGURE 7.6: *Fraction of Correct Choices for Intermediary's Utility-maximizing Decision*

7.3.3 Impact of Dispersion of Observations on the Required Amount of Monitoring Observations

The previous section identified the amount of monitoring observations, which is required for an intermediary's decision on SLA establishment. The results that were retrieved hold for a general setting without the knowledge about particular distribution functions of the degrees of SLA violation. This section is dedicated to investigate the impact of dispersion of monitoring observations on the required amount of monitoring observations. This is achieved by extending the evaluation from the previous section as described in Section 7.1.2.

This section presents the results of the analysis of the impact of the variance of a PDF of the degree of SLA violation on the required amount of monitoring data. The analysis is carried out for the intermediary's semi-variance minimizing decision (Section 5.1) and their certainty equivalent maximizing decision (Section 5.2), respectively. Performing the statistical analysis with a Wilcoxon rank-sum test as described in Section 7.1.2 evaluates the null-hypothesis of equal means of the difference in ranks for different amounts of monitoring observations. The results of the evaluation for each problem set are shown in the Appendix in Tables A.5 and A.7.[9] If the null-hypothesis cannot be rejected any more, the threshold amount of monitoring observations is found. In the following, the impact of the dispersion of monitoring observations on the required amount of monitoring observations is examined.

For this analysis, the maximum dispersion of monitoring observations as measured by the variance of the degree of SLA violation of the benchmark *provider* SLA portfolio is determined for each problem set.[10] The maximum variance is then set into context with the required amount of monitoring observations for this problem set, which was identified in the first step. Variances and required amounts of monitoring observations for each problem set are shown in Table A.6 for an intermediary's risk-minimizing decision and in Table A.8 for an intermediary's utility-maximizing decision. The impact of the variance of monitoring observations on the required amount of monitoring observations is analyzed equally to Section 7.2.3 by means of an ordinal regression. In the following, first the results for the risk-minimizing decision of an intermediary and then, the results for the utility-maximizing decision are presented.

An ordinal regression is executed for identifying the impact of the independent variable *variance* on the dependent variable *amount of observations* for an intermediary's risk-minimizing decision about SLA establishment with providers. It reveals that the maximum variance in the benchmark *provider* SLA portfolio has a significant ($\alpha = 0.05$) impact on the required amount of observations. Coefficients and results can be found in Table 7.9, where the coefficient can be found in the last row, column *Estimate* and the corresponding p-Value is denoted in the column *Sig*.

The same analysis is carried out for an intermediary's utility-maximizing decision. The ordinal regression, which analyses the impact of the independent variable *variance* on the dependent variable *amount of observations* for an intermediary's utility-maximizing decision shows that the maximum variance of the benchmark *provider* portfolio has a negative impact on the required amount of monitoring data. This result is significant to the 0.001 level and can be found in Table 7.10.

The result of the regression implies that the higher the variance of observed degrees of violations is for SLAs, the less data is required. This result contradicts the assumption that the more dispersed data is, the more observations are required for a precise ap-

[9] For the results, *** denotes significance to the 0.001 level, ** states significance to the 0.01 level, and * shows significance to the 0.05 level, and hence, that the null hypothesis can be rejected to the respective level.

[10] For a discussion of the decision for the maximum variance as a proxy for dispersion please refer to Section 7.2.3.

TABLE 7.9: *Impact of Variance on Required Amount of Observations for Intermediary's Risk-minimizing Decision*

Parameter Estimates						
					95% Confidence Interval	
		Estimate	Std. Error	Sig.	Lower Bd.	Upper Bd.
Threshold	Amount = 10	-4.081	1.095	0	-6.227	-1.935
	Amount = 20	-2.418	0.635	0	-3.664	-1.172
	Amount = 50	-1.372	0.538	0.011	-2.426	-0.318
	Amount = 100	-0.255	0.509	0.616	-1.253	0.743
	Amount = 200	1.2	0.519	0.021	0.182	2.217
	Amount = 500	2.663	0.551	0	1.583	3.744
Location	Variance	13.066	6.563	0.047	0.202	25.93

TABLE 7.10: *Ordinal Regression: Impact of Variance on Required Amount of Observations for Intermediary's Utility-maximizing Decision*

Parameter Estimates						
					95% Confidence Interval	
		Estimate	Std. Error	Sig.	Lower Bd.	Upper Bd.
Threshold	Amount = 5	-3.989	0.727	0	-5.414	-2.565
	Amount = 10	-2.372	0.69	0.001	-3.724	-1.019
	Amount = 20	-0.736	0.657	0.263	-2.025	0.552
	Amount = 50	0.42	0.695	0.546	-0.942	1.782
	Amount = 100	0.874	0.742	0.239	-0.579	2.328
	Amount = 200	2.364	1.12	0.035	0.168	4.56
Location	Variance	-37.713	8.634	0.000	-54.635	-20.791

proximation of the distribution of the data. Therefore, the results that were retrieved throughout the evaluation of the utility-maximizing decision of an intermediary were thoroughly reviewed. Revisiting the benchmark portfolios that were determined for each of the problem sets disclosed that for each problem set, the same portfolio exhibited the highest expected utility in each problem sets and hence, was the benchmark portfolio. Furthermore, the maximum variances of the benchmark portfolio in each problem set as denoted in Table A.8 are very low (< 0.1). The interplay of the chosen values for prices, penalties and costs for service provisioning with the low variances of the distributions of degrees of SLA violation fosters the assumption that the degree of SLA violation and its dispersion has a negligible impact in this configuration of prices, penalties and costs. In order to support this statement, further evaluation runs that test the impact of special configurations of prices, penalties and costs on the utility-maximizing decision in conjunction with different specifications of probability distributions and their variance are required. Consequently, a sensitivity analysis that investigates the impact of different configurations of prices, penalties and costs is required in the future for the case of utility maximization in an intermediary's decision.

7.4 Summary

Section 7.2.2 identified the amount of monitoring observations that are required for the methods that solve provider's decision problem about SLA establishment by minimizing the risk of SLA violation, or maximizing expected utility, respectively. The statistical analysis identified that a provider needs to collect at least 50 observations for each portfolio in case of the semi-variance-minimizing decision on SLA establishment and at least 10 observations for the utility-maximizing decision. Consequently, the required amount of data is relatively low for both methods, allowing the data to be collected in a reasonable amount of time. This of course depends on the agreed monitoring intervals. The difference in the amount of required data could be caused by the semi-variance's property of only taking deviations worse than average into account. This leads to excluding some of the observations from the semi-variance calculation and hence, more observations are required in total.

The results that are presented in Section 7.2.2 advise providers on the amount of data to be collected for a correct decision. The amounts of observations that were determined are retrieved based on the average of 150 different problem sets. The influence that the dispersion of monitoring observations as measured by the variance of PDFs of the degree of SLA violation has on the required amount of monitoring data was identified in Section 7.2.3.

Therefore, for each of the problem sets, the benchmark portfolio was determined by means of semi-variance and certainty equivalent, respectively. The correctness of the sample based decision is determined by comparison to the benchmark portfolio, which was found based on semi-variance and certainty equivalent calculated from PDFs. Based on the observations of correctness, the required amount of observations was determined as the amount from which on the fraction of correct choices cannot be improved on significantly any more. Having identified the required amount of monitoring data for each of the problem sets, the impact of the dispersion of monitoring observations was evaluated. Thus, for each of the benchmark portfolios, the maximum variance of the degree of SLA violation for each of the SLAs in the optimal portfolio was selected. A regression that analyzed the impact of the maximum variance on the required amount of monitoring observations revealed that with an increasing variance, more monitoring data is required.

Consequently, these results imply that the amount of 50 observations, which was determined in Section 7.2.2 for the provider's semi-variance minimizing decision needs to be increased for high values of the variance and may be decreased for very low values of variance. The same applies to the previously determined 10 observations for the utility-maximizing decision. Thus, a service provider that monitors their SLAs needs to calculate the variance of the observed degrees of SLA violation. In case the variance is low (close to zero), the amounts of observations that were identified in Section 7.2.2 suffice to approximate the risk of SLA violation precise enough to make correct decisions. If the variance of observed degrees of SLA violation is significantly higher, more observations need to be collected. Advice on the additionally required number of observations can be derived from the coefficients resulting from the regression analysis. Exemplarily, the coefficient for the variance in Table 7.4 shows, that an additional unit of variance requires 19.710 additional observations. Table 7.5 is to be read analogously. This way, a solution to Research Question 4 concerning the provider's decision on SLA establishment is provided.

In Section 7.3.2, the amount of monitoring observations that is required for the methods that solve an intermediary's decision problem on establishing SLAs with providers was determined. The statistical analysis revealed that an intermediary is required to collect at least 100 observations for each provider SLA portfolio if the semi-variance -minimizing method is applied and to collect 20 observations if the utility-maximizing method is applied. The amount of required data is slightly higher than the required amount for the decision of entering SLAs with customers. Nevertheless, the data can still be collected in reasonable time, depending on the monitoring periods. By taking the results from Section 7.2 into account, an intermediary is required to gather at least 50 observations for the risk-minimizing decision about SLAs with customers and at least 100 observations for the risk-minimizing decision about SLAs with providers. If the utility-maximizing method is applied to solve an intermediary's decision problem, 10 observations for customer SLAs and 20 observations for provider SLAs have to be collected.

Similar to the decision on SLAs with customers (Section 7.2) for the semi-variance -minimizing decision requires to collect more monitoring observations than for the utility-maximizing decision. This can be explained by the semi-variance 's property of taking only those observations that are worse than average into account, which leads to a higher total of observations. The results from Section 7.3.2 give general advice on the minimum amount of data that needs to be collected based on an average of the 150 created problem sets. As the distribution of observations can differ very much in their properties, the following section investigates the impact of the PDFs' properties, measured by the variance, on the required amount of monitoring observations.

Section 7.3.3 explored the impact of dispersion of monitoring observations on the required amount of monitoring observations.

The evaluation results imply that the required amount of observations that were determined in Section 7.3.2, that is 100 observations for the semi-variance-minimizing approach, need to be increased for high variances. For low variances, the previously determined number of observations is sufficient. The result for an intermediary's decision was contradictory and requires further research.

An intermediary is required to collect monitoring observations for provider and for customer SLA portfolios. Thus, the results that were retrieved in Section 7.2 apply to an intermediary's decision problem. This implies that an intermediary is required to collect at least 50 observations for customer SLAs and at least 100 observations for provider SLAs if the risk-minimizing method is applied and at least 10 observations for customer SLAs and 20 observations for provider SLAs if the utility-maximizing method is applied. The amount of observations that is required for the decision about provider SLAs is slightly higher than the amount that is required for the decision about customer SLAs. However, it is only the next step in the chosen amounts y of monitoring observations that is required in addition for either approach and hence, is considered to be due to the differences in the calculation of semi-variance and certainty equivalent for the decision about provider SLAs. The amount of observations that an intermediary needs to collect depends on the variance of the degrees of SLA violation that an intermediary observes in reality. Once, an intermediary that wants to apply the risk-minimizing decision support has collected 50 observations of each portfolio of SLAs with customers and 100 observations for each SLA portfolio with providers, they need to determine the variance of the observed degrees of SLA violation. If this variance is very low, i. e. close to zero, the collected observations suffice and the risk-minimizing decision support leads to an optimal decision. Otherwise, the intermediary is advised to collect more monitoring observations.

Part IV

Finale

Chapter 8

Conclusions

THE *servicification* of value creation that fosters a shift away from the traditional trade of goods towards an exchange of services (Vargo and Lusch 2004b) in which goods are considered as vehicles for services was the motivating factor for this thesis. The rise of the Internet and related technologies in recent years brought about changes to the way in which services are provisioned. The Internet facilitates the offering and provisioning of services to an extended circle of potential customers. While potential customers were previously attracted by personal contacts, advertisements or in magazines of which the perception is limited by coverage, the Internet does not have these constraints. Nearly everybody can be reached by Internet advertisements, which increases the number of potential customers for service providers (Anderson 2006), the requests for service provisioning and the number of SLAs that regulate service provisioning. This results in the need for cooperation among service providers in order to meet the demand for customized services, which is facilitated by the joint provisioning of services. This thesis has provided a method that supports decisions on the establishment of SLAs.

The case of joint provisioning of Web or Cloud services is only one example, in which the cooperation of service providers is required. An example for the joint provisioning of a *general* service is the building of a house, where the cooperation of different craftsmen is required for the completion of a variety of tasks like building the walls, tiling floors, heating and ventilation installation. Thus, managing the cooperation of service providers is a problem that is inherent to all of the introduced types of services (Section 2.1). The joint provisioning of services is regulated by the stipulation of SLAs, in which the functionality and quality of provided services is determined along with a price for service provisioning and a penalty that is applied in case of SLA violation.

Thus, providers of different types of services need to make decisions concerning the establishment of SLAs. However, for the establishment of SLAs for services that are provided via the Internet, special challenges arise. These challenges were identified in Section 3.3 and concern the impact of the availability of the Internet and related technologies, which result in an increasing number of SLAs to decide on and the technical composition of services. Further challenges that need to be addressed concern the presence of a multitude of customers and the expression and consideration of the risk of SLA violation in the decision. Finding a solution to this decision problem that considers these challenges was the focus of this thesis.

8.1 Contributions

This thesis is concerned with making decisions on SLA establishment. In Section 3.3, five challenges that need to be addressed in decision support for the establishment of SLAs have been identified. The challenges concern the impact that the availability of the Internet has on the provisioning of services by fostering an increase in the demand for services and a change in the type of service provisioning and composition. Furthermore, the presence of a multitude of customers was identified as having an impact on making decisions about SLA establishment. Finally, in these decisions, uncertainty of future SLA violations needs to be taken into account. Thus, an expression for the risk of SLA violation was required that can be employed in the decision.

The decision on SLA establishment that considers these challenges was considered from the viewpoint of a service provider. In Section 2.3.2, two types of service providers were distinguished - *service providers* that can provide all functionality and quality by themselves and *intermediaries* who cannot provide all functionality by themselves and thus need to procure services from other providers. The first contribution of this thesis is the answer to Research Question 2:

RESEARCH QUESTION 2 ≺PROVIDER'S RISK-MINIMIZING DECISION≻. *How can a provider select the SLAs that minimize the risk of SLA violation while taking their technical properties and economic preferences into account?*

A method that solves a service provider's decision problem was introduced that applies the semi-variance of expected penalties as a measure of risk which is minimized in the objective function. In order to account for a provider's available resources and the profit that a provider wants to gain from service provisioning, constraints were included in the decision problem. By employing an adaptation of portfolio selection, a decision on a multitude of SLAs with customers is possible. Consequently, a provider's technical and economic properties and the five challenges in decision making for SLA establishment are met. However, this method did not include the possibility to specify a trade-off between the risk of SLA violation and expected profit. Thus, an extension to the risk-minimizing decision was presented that maximizes the expected utility from service provisioning as measured by the certainty equivalent. The certainty equivalent resembles a trade-off between expected profit and risk of SLA violation, expressed by the variance of SLA violation, individually weighted by a service provider's coefficient of absolute risk-aversion. Like the risk-minimizing decision, the utility-maximizing decision meets the challenges from Section 3.3. While the first method is suitable if the distributions of the degrees of violation are asymmetric and if a provider's main goal is to avoid the violation of SLAs, for instance in case of prohibitively high penalties. The second method is suitable for risk-averse decision makers and when the distributions of the degree of violation are symmetric.

The second contribution focuses on an *intermediary's* decision problem concerning the establishment of SLAs with customers and providers as an extension to the decision problem of a service provider. This was addressed by Research Question 3:

RESEARCH QUESTION 3 ≺INTERMEDIARY'S RISK-MINIMIZING DECISION≻. *How can an intermediary select the SLAs with customers and supplying providers that minimize the risk of SLA violation while taking their technical properties and economic preferences into account?*

In order to solve an intermediary's decision problem under consideration of the challenges that were described above, an intermediary's available resources and expected profit, the methods for a service provider's decision were extended. A method was presented that minimizes the risk of violating SLAs with customers and the risk of SLA violation by providers. Similar to the risk-minimizing decision of a provider, an intermediary's risk-minimizing decision with customers and providers does not account for a trade-off between the risk of SLA violation and expected profit. Therefore, an extension of the risk-minimizing method for decision making that maximizes an intermediary's expected utility from the service provision and consumption was expressed by the certainty equivalent. Both of the methods that solve an intermediary's decision problem consider the challenges on decision making that were raised in Section 3.3 and take an intermediary's available resources and expected profit into account. As before, the first method is suitable in cases of asymmetric distributions of the degrees of violation. The second method is suitable in situations with symmetric distributions of degree of SLA violation and risk-averse intermediaries.

However, in order to address both, Research Questions 2 and 3, several related questions had to be addressed. First, important groundwork was laid. Additionally, questions about the applicability in practice needed to be answered in order to supplement Research Questions 2 and 3. This thesis incorporates contributions, which complement the implementation of decision support for providers and intermediaries. The basis for the presented approaches was laid by defining what a service is and delineating service types that emerge with the availability of the Internet and related technologies. After having given a clear definition of the trading object, SLAs that regulate the provision and consumption of services were described and discussed. A thorough literature review gave an overview of the constituents of an SLA, which comprise a functional description of the provided service and quality objectives along with a price and a penalty. From the building parts of an SLA, technical and economic properties of a service provider and an intermediary that constrain the decision about SLA establishment were identified.

In a thorough literature review on economic principles of decision theory and the expression of risk, methods for decision making about SLA establishment that minimize the risk of SLA violation were discussed. Finally, from related literature in the field of SLA management and decision making about joint service provisioning, challenges that have not been considered yet were identified. The methods that address Research Questions 2 and 3 apply the foundations on decision making, expression of risk and SLA establishment under the consideration of the challenges that were identified from related work and a service provider's or intermediary's technical and economic properties. The aspects that drive a provider's and an intermediary's decision about SLA establishment are in the focus of Research Question 1:

RESEARCH QUESTION 1 ≺STIPULATION OF SERVICE LEVEL AGREEMENTS≻. *What do a provider and a customer of a service stipulate about service provisioning in a Service Level Agreement and which technical properties and economic preferences of a service provider influence the decision about establishing an SLA?*

The methods that address the decision problems of service providers and intermediaries, respectively, as the focal contribution of this thesis were formulated in order to account for providers' and intermediaries' constraints as identified in Research Question 1. The risk of SLA violation is calculated from monitoring observations about SLA violations in the past. Consequently, the methods that solve the problems depend extremely on the amount of available data. Therefore, the suggested approaches were evaluated with respect to the required amount of monitoring observations.

Research Question 4 ≺Required Amount of Monitoring Data for Decision making≻. *How many monitoring observations are required for the calculation of SLA violation risk in a) a provider's and b) an intermediary's decision on SLA establishment?*

It was shown that a provider needs to collect a relatively low amount of monitoring observations, no matter whether the risk-minimizing method or the utility-maximizing method are applied. For a provider's risk-minimizing decision, an amount of 50 observations was identified, whereas the utility-maximizing approach requires only 10 observations from the past. For the identification of the required amount of monitoring observations artificial monitoring observations were created as random draws from probability distributions. The knowledge of the distribution function enabled the calculation of semi-variance and certainty equivalent directly from the distribution function and hence, the most precise representation of semi-variance and certainty equivalent was retrieved. The SLA portfolio that minimized the semi-variance and the portfolio that maximized the certainty equivalent were chosen as the SLA portfolio that a provider should establish and thus, the portfolios are the benchmark portfolio for either method. By comparing the benchmark portfolios to the portfolios that were chosen by minimizing the semi-variance that was calculated from different amounts of monitoring observations, or maximizing the certainty equivalent that was computed from different amounts of monitoring observations, the correctness of the respective choice could be determined. The creation of a particular amount of monitoring observations and the choice of a portfolio from these monitoring observations was repeated so that for each amount of monitoring a fraction of correct choices could be calculated. The fraction of correct choices is a proxy for the probability of making a correct decision. In order to retrieve results that are independent from a particular distribution function of the degree of SLA violation, 150 problem sets with different probability distributions for the degree of SLA vioaltion are created.

By comparing the fraction of correct choices for different amounts of observations, it was possible to determine if the probability of makig a correct decision increases significantly with an increase in the amount of observations. The amount for which the probability can not be increased significantly is the amount which is required for a provider's decision about SLA establishment as the representation of SLA violation risk that is calculated from this amount of observations is as precise as it can become.

The results were derived on the basis of an average over the distribution functions from the 150 problem sets. Therefore, in an extended evaluation, the impact of the dispersion of monitoring observations as measured by the variance of the observed degrees of SLA violation on the required amount of monitoring observations was investigated. The results showed, that a higher variance in the observations leads to more observations that are required for a provider's decision. This implies, that for a low variance, 50 observations are sufficient for the risk-minimizing decision and 10 observations suffice for the utility-maximizing decision but that the number of observations needs to be increased the higher the variance in the observations. The required increase in the number of observations that depends on the increase of the variance is shown in Tables 7.4 and 7.5.

The amount of monitoring observations that is required for an intermediary's decision about SLA establishment was identified analogously to the evaluation of a provider's decision. For an intermediary's risk-minimizing decision, an amount of 100 observations is required, whereas the utility-maximizing approach requires only 20 observations from the past. Artificial monitoring data was created and the sample-based decisions were compared to the benchmark portfolios. This comparison was carried out for customer and provider SLA portfolios and resulted in observations if the decision that was made by taking artificial monitoring observations into account were correct. For each of the different amounts of monitoring observations, this approach was repeated 1000 times. Consequently, for each amount of observations, a fraction of correct choices could be determined.

By checking the fraction of correct choices for different amounts of observations against each other, the required amount of monitoring data could be identified as the amount of data from which on the fraction of correct choices could not be improved significantly any more.

As in the provider's case the impact of the variance of the observed degrees of SLA violation on the required amount of monitoring data was investigated. As before, the results imply that a higher variance required a higher amount of monitoring data in the risk-minimizing decision. Consequently, for a low variance, 100 observations are sufficient for the risk-minimizing decision. For high variances, this number of observations needs be increased. The results of the intermediary's utility maximizing decision imply that further investigation on the impact of prices, costs and penalties is required.

8.2 Open Research Questions and Future Work

This section critically discusses the assumptions that were raised in the work at hand and resulting restrictions. Possible solutions to these restrictions are outlined and future research directions are derived from the suggested solutions.

8.2.1 Existence of Monitoring Data

Each of the presented methods in this work that solve a provider's and an intermediary's decision problem on the establishment of SLAs, calculates the risk of SLA violation from monitoring observations from the past. First, this requires that monitoring data is available, or in other words that the decision maker had already entered into SLAs and monitored their adherence. Second, as it was shown in Chapter 7, relying on monitoring data from the past requires that a certain amount of observations are available for each portfolio of SLAs. Consequently, in order to ensure the applicability of the presented methods, in this work it was assumed that service providers and intermediaries restrict their offerings to a pre-specified set of SLAs (see Assumption 1). This constrains the number of available SLAs and the number of possible combinations as long as only one instance can be active in each period. Third, in order to facilitate comparability of monitoring results, the underlying infrastructure was assumed to remain constant over the monitored periods (see Assumption 3).

The availability of monitoring data can be assumed for service providers and intermediaries, which are experienced in providing services for some time. Nevertheless, in the case that a provider or intermediary intends to offer a new SLA for a known service, there is no information available about the degrees of violation that can be expected. For this case, an approach that estimates the degree of violation is required in order to include the new SLA in decisions. This could be facilitated by determining the degree of similarity of the new SLA to already known SLAs. Measures for similarity emerged with the field of multivariate statistics and cluster analysis (Johnson 1988; Kaufman et al. 1990; Steinhausen 1977). By applying similarity measures from the field of cluster analysis, the most similar SLA can be determined and the corresponding distributions of the degree of violation can be employed as a proxy for the new SLA's expected degree of violation.

In the case that a service provider or intermediary wants to expand their portfolio of offered services and include the new services, information about the performance of the service needs to be retrieved. Therefore, performance testing needs to be applied in order to retrieve an estimation of the service's behavior. Performance testing methodologies have been applied in the context of infrastructure planning (Happe et al. 2010; Kounev 2006) but can be applied for estimating service behavior as well.

Finally, if an intermediary is confronted with the case that a new service provider enters the market, information about the provider's adherence to SLAs needs to be retrieved in order to include the new SLAs in the decision problem. Therefore, an intermediary is required to spend money on information retrieval concerning the provider's properties in adhering to SLAs. The amount of money that should be spent on gathering information, as market analyses or profiling the new provider, can be determined by applying transaction cost theory (Coase 1937) or the theory of the expected value of perfect information (Felli and Hazen 1998).

Even though the assumption of a restricted number of pre-defined SLAs is realistic for cases of incumbent service providers like Amazon or Salesforce, the assumption that only one instance of each SLA is concurrently active is not (see Assumption 1). In order to relax the assumptions of a pre-defined set of SLAs and the number of concurrently active instances of one SLA, an approach that does not restrict the dynamic definition of SLAs needs to be sought. For the case of the extension of already known SLAs, measures of similarity and their application (Johnson 1988; Kaufman et al. 1990; Steinhausen 1977) require further investigation. By identifying similar portfolios, the behavior of the new portfolio can be estimated. Additionally, approaches for performance testing can be applied in cases, where no similar portfolios are available.

Assumption 3 required that the underlying infrastructure, that is, resources, does not change over time. This assumption aimed to achieve comparability of monitoring data. In order to apply the presented methods in settings, where the infrastructure is adapted in the case of SLA violation, monitoring data needs to be weighted according to its age. This way, newer monitoring data can be emphasized. Approaches that allow for the weighting of input data include exponential smoothing (Trigg and Leach 1967). Further assumptions served to ensure the applicability of the presented approaches, among them the restriction to include only one service in each SLA (Assumption 2) and equal length of monitoring periods (Assumption 4). Relaxing Assumption 2 requires an adaptation of the presented methods that captures a multitude of services per SLA by adapting the monitoring system. To consider monitoring periods that have different lengths, the formulation of semi-variance and certainty equivalent need to be adapted in order to allow a weighting of monitoring observations. Both of these extensions open perspectives for future research.

In summary, the existence of monitoring data and restrictions to the provider's and intermediary's freedom in specifying SLAs needs further consideration. Especially the development of a feasible measure of similarity for SLAs as well as for portfolios of SLAs in conjunction with the application of performance testing are considered as most important future steps. In order to complete the process of service level management according to ITIL[1] the suggested methods that solve providers' and intermediaries' decision problems for SLA establishment need to be extended in order to account for adapted infrastructure. Additionally, providers and intermediaries will include the experience from the past in their decisions on prices and penalties. This issue is discussed in the following section.

8.2.2 Seperation of SLA Portfolio Selection From Pricing Decisions

Based on the assumption of exogenously given prices and penalties, the methods that were introduced in Chapters 4 and 5 enable a service provider and an intermediary, respectively, to select the SLA portfolio that incurs the lowest risk (or the highest possible expected utility). The selection of an SLA portfolio is taken into account seperately from the decision on prices and penalties. This design decision serves to distinguish the effect of negotiation and strategic pricing decisions from SLA portfolio optimization. In the presented approach, the prices and penalties from the SLAs in the portfolio currently under consideration are taken into account. The uncertainty of an SLA violation, which is calculated from past SLA violation data, is put into context with the current specification of the SLAs, i. e. its prices and penalties.

In real-world service provisioning scenarios, there are cases in which the price (and penalty) for service provisioning is fixed and does not change with respect to the customer (e. g. Amazon Web Services pricing). Nevertheless, this does not mean that prices do not change over time. In contrast, it can be assumed that experience from the past influences pricing decisions. Consequently, taking the decision on prices and penalties explicitly into account would extend the methods for making decisions for SLA establishment, presented in this work, to allow for an endgenous and experience-based design of prices and penalties. The question that has to be answered in this context is in determining the most suitable pricing mechanism. This mechanism has to take into account the impact that the availability of the Internet and related technologies has on service provsioning and with it, for instance the specific characteristics of SaaS offers.

[1] `http://www.itil.org/de/vomkennen/iso20000/servicedeliveryprozesse/servicelevelmanagement.php` last accessed: October 9, 2011

Static pricing, as described in the Amazon Web services case above, can be employed for a big variety of goods as well as for services, even if it does not lead to the highest possible return in most cases. First degree price discrimination, which enables the setting of a specific price for each customer, exhibits challenges with respect to identifying a customer's preferences (Shapiro and Varian 1999; Varian 1997), which implies that it remains a merely theoretical approach. Versioning, bundling as well as amount-based pricing have already been observed in software and service trade (Krämer 2010; Lambrecht and Skiera 2006; Sundararajan 2004). Therefore, second degree price discrimination exhibits features, which makes it a feasible pricing mechanism for SaaS offers. For third degree price discrimination, challenges arise for group segmentation and the verification of customers' affiliation. Nevertheless, group pricing is a promising approach that can overcome problems of identifying a customer's type and is applicable for service pricing (Knapper et al. 2011).

Finally, dynamic pricing is considered. Services, including Web services and SaaS-offers, are often part of joint service provisioning scenarios and hence, experience (semi-) automatic service composition. This imposes time and complexity constraints on the pricing mechanism. Therefore, the time-consumption of auction mechanisms need further investigation. Dynamic pricing mechanisms can be observed in practice (Amazon EC2 Spot Instance[2]) and have been discussed in theory with different objectives like incentive compatibility (Blau 2009), fairness and network-growth (Conte 2010), or revenue management (Anandasivam 2010).

Despite the variety of the pricing mechanisms in place, none of the approaches considers penalties as a strategic means to incentivize the adherence to SLAs while considering past monitoring observations. First approaches in this direction have been introduced in Filipova-Neumann et al. (2010) and Michalk and Haas (2011) but still require further research.

8.2.3 Simulation-based Evaluation

Simulations often expose valuable results and give reasonable insights. Nevertheless, there are limitations of simulation-based evaluations, which have to be considered. In order to conduct a simulation-based evaluation, reality has to be abstracted from, which simplifies the real-world scenarios. In this thesis, the decision problems of providers and intermediaries have been constrained to small instances with only three available

[2] `http://aws.amazon.com/ec2/spot-instances/` last accessed: October 10, 2011

SLAs to offer (and to procure in the intermediary's case). In real-world scenarios, the decision maker will most-likely be confronted with a higher number of SLAs and possible combinations. Additionally, as no monitoring and analytics system that could meet the approach's requirements was in place, distributions for the degree of violation were chosen randomly as a basis for a provider's and an intermediary's decision. The validity of the chosen distribution could not be validated due to the unavailability of real monitoring data. In order to substantiate the approach of randomly chosen probability distributions, the simulation was run for different instantiations of distributions and results were deducted from a) the entirety of problem sets (see Sections 7.2.2 and 7.3.2) and b) each problem set seperatly (see Sections 7.2.3 and 7.3.3). Furthermore, an investigation on the impact of the number of SLAs that can be established simultaneously on the resulting decisions could shed light on the question of the optimal number of SLAs that a service provider, or intermediary, should hold available.

Even though the data gives valuable insights to the applicability of the presented methods, the suggested simulation-based evaluation includes simplifications, which prevent the generalization of the results in their totality. In order to allow for a more general validity of the results, further research should take up on the validity of the simulated probability distributions of the degrees of violation. Additionally, the implementation of a monitoring system, which delivers monitoring observations that meet the assumptions on monitoring data is suggested in order to underpin the approach's applicability.

8.2.4 Consumer Behaviour and Long-term SLA Design

The methods that address the decision on SLA establishment focussed on a provider's, or intermediary's, perspective. The impact that consumer behavior can have on the adherence to an SLA has not been considered, yet. Consumer behavior may include for instance the load that is put on a service as measured by the number of requests, the premature cancellation of an SLA because due to discontent or the violation of agreed terms. To include this behavior in the decision problem, the methods that address SLA establishment decisions need to be extended. Therefore, new methodologies need to be employed that allow to capture consumer behavior, which could be facilitated by the application of methodologies like Markov decision processes that model load curves, negotiation protocols that regulate bilateral agreement stipulation and their cancellation as well as incentive schemes that foster the adherence to agreed terms on the provider's as well as on the consumer's side. In this vein, considering consumer behaviour opens a field of intriguing and most interesting research.

The decision support for SLA establishment required a fixed set of SLAs that a provider offers to customers, in which prices and penalties are set exogenously and only once. In settings of service outsourcing it is often the case, that long term SLAs are stipulated and re-negotiated after certain periods of time. In this context, aspects of incentivation of providers and consumers gains an additional dimension. Besides the threat of a penalty, future negotiations need to be considered in decisions about effort that influences service provisioning and decisions that influence a consumer's behaviour. Consequently, infrastructure planning decisions (a provider's effort to increase service performance) may be affected. To investigate the impact of long term agreements opens a completely new perspective on SLAs and requires to capture negotiation protocols as well as infrastructure planning and many more. Hence, further research is required to address these fascinating questions.

8.3 Complementary Research

This section gives an overview about research directions, which are related to the work at hand but are not direct extensions. The topics point the reader to supplementary aspects of decision making on SLA establishment from an economic, technical and sociological viewpoint.

Intelligent Querying in Service Repositories The provisioning of services and the composition of service modules in order to create value and to meet customers' demand is increasingly important in today's business. With the fast growing market of Web and Cloud services and their compositions, (semi-)automated methods are required to find appropriate services. In order to facilitate service composition, especially in the context of service marketplaces, knowledge about services that can be offered, procured, and composed is required. As argued at the beginning of this work, service composition and with it, the cooperation of service providers, is regulated by means of SLAs. Additionally to the service description, qualitative attributes along with price and penalty are stipulated. Therefore, the corresponding SLAs have to be connected to the service desriptions.

In this context complementary research questions deal with formalisms for the description of functional and non-functional service properties and the corresponding quality of service attributes. In order to enable reusability of services in different contexts, the pure description with standards like WSDL (Christensen et al. 2001) and SOAP does not suffice, as these do only contain rudimentary descriptions of service functionality. Consequently, a more elaborate description of services is required, which allows for functional, non-functional and quality aspects. One approach that meets these requirements is presented in Agarwal et al. (2008). The discovery of services that match a user's request for functional, non-functional and quality aspects is described in Junghans and Agarwal (2010).

An extension to the approaches in Agarwal et al. 2008 and Junghans and Agarwal (2010) that allows for a specification of SLAs and that implements a ranking of SLAs with respect to their respective level of qualities would complement the work at hand in the stage of identifying feasible SLAs to offer (for the provider) and the combinations of offered and required SLAs (for the intermediary).

Aggregation of Monitoring Results The previously described complementary research about service discovery directly leads into the research area of service monitoring and analytics. Having identified feasible services in a business-relevant scenario of service provisioning, service level management gains momentum. Service level management does not only include the negotiation of service functionality and quality goals, that are stipulated in SLAs. It considers the monitoring of the active service and the communication of these to the customer (Bon 2005).

The standardized specification of SLAs and the included elements like SLOs, for specifying the stipulated quality goals as presented in Andrieux et al. (2007) lays the basis for service level monitoring. The technical implementation of monitoring systems like Nagios (Barth 2008) allows to observe the service performance and resulting status of the SLA. In Andrieux et al. (2007), a service in an SLA can take the states *not ready*, *completed*, or *ready* and *processing* or *ready* and *idle* (see Figure 8.1). If the service is processing, each guarantee, that is, each SLO can be monitored and determined to be either *fulfilled*, *violated*, or *not determined* (see Figure 8.2). In case one of the SLOs is found to be violated, the whole SLA is determined to be violated. As an extension to this rather strict way of judgement, in Schulz (2010) an approach is presented that allows the determination of the degree of violation of an SLA. This means that the severity of the SLA violation has an impact on the final degree of violation of an SLA and in last consequence on the penalty, which has to be payed. By implementing a measure for the degree of violation of an SLA, this research complements the work at hand by aggregating the input to the decision problems in a feasible way.

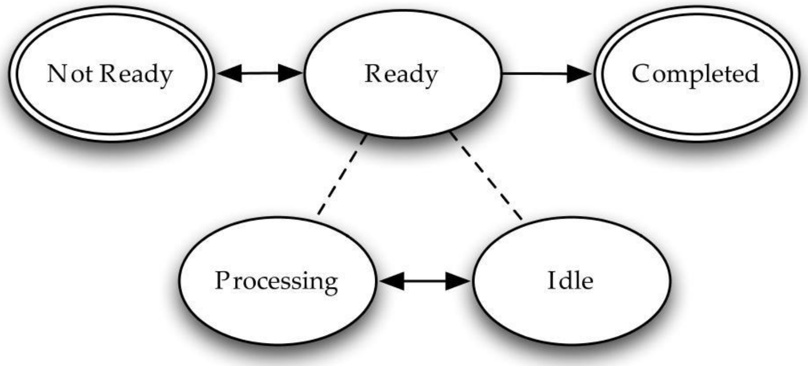

FIGURE 8.1: *Service States according to WS Agreement*

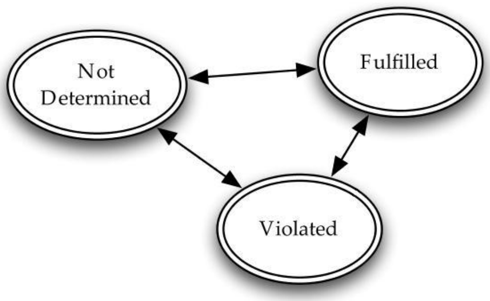

FIGURE 8.2: *Guarantee States according to WS Agreement*

Cross-layer Aggregation of QoS Metrics When requesting a service from a service provider or intermediary, a customer does have their requirements about functionality and a conception on quality goals and corresponding price and penalty in mind. For the customer, it does not matter whether the requested service is a composition of service modules or rather a bundle provided by one single provider. For a service provider and particularly an intermediary, it is of major importance if services need to be requested from other providers. However, besides the technical aspects of provisioning a certain functionality to the customer, the question of delivering the specified quality becomes more complicated. For an intermediary the question arises, how QoS attributes from supplied service modules relate to quality goals that will be stipulated in an SLA with a customer.

Similar considerations arise in the context of business processes, where the interplay of services and different patterns for the sequence of execution along with the aggregation of QoS attributes play an important role. More specifically, service attributes are subject to a different aggregation logic depending on the underlying process pattern. This implies that the intermediary has to be aware of the process pattern, that is, the exact sequence of service execution, in order to apply the correct aggregation of QoS attributes.

Approaches that map business process description and the aggregation of QoS attributes have been discussed in literature. van Der Aalst et al. (2003) introduced a small set of universally valid patterns to which every workflow can be reduced. Based on these workflow patterns, research in the area of aggregation of QoS attributes has tackled, besides others, the classification of attributes, automated aggregation with the help of ontologies (Blake and Cummings 2007; Unger et al. 2008; Knapper et al. 2010).

Non-monetary Penalties in Social Environments Apart from the increasing importance that service trading gained in recent years, another development could be observed. Especially the emergence and rapid growth social network platforms like Facebook[3], Xing[4], and LinkedIn[5] was detected. Social network platforms allow for the extension of usual Web technologies by a social context, which connects providers and consumers of information, goods and services.

Social aspects of online collaboration exhibit challenges for consumers, providers, and

[3] `https://www.facebook.com` last accessed: October 10, 2011
[4] `http://www.xing.com` last accessed: October 10, 2011
[5] `http://www.linkedin.com` last accessed: October 10, 2011

intermediating parties. The application of SLAs, which serve to stipulate service functionality, non-functional attributes, and quality goals is common in business relevant service provisioning scenarios. In this context, the adherence to SLAs is incentivized by specifying penalties by means of a monetary amount or sanctions that can be translated to a monetary amount on behalf of the provider. With the background of social networks, penalties need to be considered differently in order to take social rather than monetary aspects into consideration.

Standard economic assumptions state the maximization of expected utility as the main aim of an individual. However, even that there is no perceived utility gain from participating in systems like Peer-to-Peer networks or the community portal Wikipedia, vivid participation in these systems can be observed, because some users behave unselfishly to some degree (Hughes et al. 2005). Principles from the field of sociology or social economics that are subsumed by the term "*Social Preferences*" include concepts of altruism, reputation and reciprocation. These concepts enable the explicit investigation under which settings people have incentives to engage in certain types of interactions even if it would be optimal for rational individuals not to engage in the interactions (Berg et al. 1995; Charness and Rabin 2002; Trivers 1971). An adaptation of established ways to setting incentives in economic settings by including concepts from sociology will help to consider a social environment in the formulation of penalties and the establishment of agreements in social computing contexts.

Appendix A

Simulation Results

A.1 Evaluation for Provider's Decision Methods

A.1.1 Minimization of Semi-Variance

TABLE A.1: *Evaluation of Provider's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 1	272***	290***	147***	15***	0	0	0
Problem Set 2	228***	124***	59***	1	0	0	0
Problem Set 3	-18	-62**	71***	73***	74***	53***	33***
Problem Set 4	229***	101***	41***	0	0	0	0
Problem Set 5	215***	117***	59***	5*	0	0	0
Problem Set 6	248***	105***	13***	0	0	0	0
Problem Set 7	301***	124***	51***	2	0	0	0
Problem Set 8	264***	124***	23***	0	0	0	0
Problem Set 9	192***	79***	35***	0	0	0	0
Problem Set 10	18**	10**	0	0	0	0	0
Problem Set 11	203***	67***	8**	0	0	0	0
Problem Set 12	217***	150***	129***	48***	10**	0	0
Problem Set 13	155***	173***	138***	40***	4*	0	0
Problem Set 14	228***	159***	28***	4*	0	0	0
Problem Set 15	239***	217***	83***	1	0	0	0
Problem Set 16	145***	58***	43***	4*	0	0	0
Problem Set 17	223***	103***	18***	0	0	0	0
Problem Set 18	242***	249***	193***	49***	10**	0	0
Problem Set 19	-30	-41*	52**	76***	76***	52***	20***
Problem Set 20	170***	190***	213***	81***	51***	9**	0
Problem Set 21	208***	128***	187***	55***	25***	4*	0
Problem Set 22	83***	63***	11**	0	0	0	0
Problem Set 23	176***	140***	42***	1	0	0	0
Problem Set 24	62***	110***	43***	3	0	0	0
Problem Set 25	181***	158***	35***	4*	0	0	0
Problem Set 26	135***	102***	65***	3	1	0	0

TABLE A.1: *Evaluation of Provider's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 27	191***	116***	36***	1	0	0	0
Problem Set 28	94***	72***	21***	0	0	0	0
Problem Set 29	233***	112***	63***	2	0	0	0
Problem Set 30	154***	101***	25***	1	0	0	0
Problem Set 31	242***	124***	28***	0	0	0	0
Problem Set 32	148***	172***	130***	30***	1	0	0
Problem Set 33	133***	121***	25***	0	0	0	0
Problem Set 34	350***	159***	43***	0	0	0	0
Problem Set 35	139***	-133***	-5	5	-6*	0	0
Problem Set 36	100***	112***	75***	5*	0	0	0
Problem Set 37	121***	179***	210***	97***	91***	43***	5*
Problem Set 38	233***	153***	42***	0	0	0	0
Problem Set 39	197***	161***	48***	11**	0	0	0
Problem Set 40	211***	120***	33***	2	0	0	0
Problem Set 41	166***	114***	47***	4*	0	0	0
Problem Set 42	303***	168***	21***	0	0	0	0
Problem Set 43	212***	95***	16***	0	0	0	0
Problem Set 44	-106***	-10	-16	83***	57**	97***	67***
Problem Set 45	120***	85***	11**	0	0	0	0
Problem Set 46	46*	81***	115***	68**	57*	38	97***
Problem Set 47	104***	229***	277***	155***	81***	46***	2
Problem Set 48	217***	71***	8**	0	0	0	0
Problem Set 49	156***	115***	42***	2	0	0	0
Problem Set 50	142***	107***	74***	9**	0	0	0
Problem Set 51	213***	123***	15***	0	0	0	0
Problem Set 52	148***	107***	89***	108***	33***	11**	1
Problem Set 53	204***	125***	57***	2	0	0	0
Problem Set 54	147***	301***	314***	61***	9**	0	0
Problem Set 55	214***	120***	33***	0	0	0	0
Problem Set 56	247***	174***	33***	1	0	0	0
Problem Set 57	243***	232***	101***	3	0	0	0
Problem Set 58	22*	43***	8**	0	0	0	0
Problem Set 59	264***	170***	44***	3	0	0	0
Problem Set 60	111***	72***	25***	2	0	0	0
Problem Set 61	67**	254***	215***	79***	17***	2	0
Problem Set 62	217***	193***	136***	104***	43***	9**	0
Problem Set 63	305***	108***	22***	0	0	0	0
Problem Set 64	159***	99***	18***	0	0	0	0
Problem Set 65	133***	141***	60***	9**	0	0	0
Problem Set 66	109***	75***	44***	2	0	0	0
Problem Set 67	198***	127***	21***	0	0	0	0
Problem Set 68	138***	93***	15***	0	0	0	0
Problem Set 69	137***	131***	73***	8**	0	0	0
Problem Set 70	-23	-17	56***	67***	30***	8**	0
Problem Set 71	97***	85***	94***	5	12	-10	-32
Problem Set 72	190***	263***	277***	106***	27***	6*	0
Problem Set 73	39*	91***	37***	3	0	0	0
Problem Set 74	249***	108***	23***	0	0	0	0
Problem Set 75	159***	128***	36***	1	0	0	0
Problem Set 76	252***	112***	28***	1	0	0	0

TABLE A.1: *Evaluation of Provider's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 77	201***	94***	16***	1	0	0	0
Problem Set 78	206***	129***	63***	5*	0	0	0
Problem Set 79	242***	259***	154***	11**	1	0	0
Problem Set 80	260***	129***	30***	1	0	0	0
Problem Set 81	226***	110***	17***	0	0	0	0
Problem Set 82	132***	78***	65***	3	0	0	0
Problem Set 83	162***	116***	34***	1	0	0	0
Problem Set 84	222***	128***	37***	0	0	0	0
Problem Set 85	177***	133***	38***	4*	0	0	0
Problem Set 86	209***	183***	130***	35***	2	0	0
Problem Set 87	204***	119***	24***	1	0	0	0
Problem Set 88	168***	147***	170***	61***	93***	32***	2
Problem Set 89	107***	80***	11**	0	0	0	0
Problem Set 90	319***	305***	72***	2	0	0	0
Problem Set 91	192***	176***	103***	4	1	0	0
Problem Set 92	235***	76***	8**	0	0	0	0
Problem Set 93	138***	225***	312***	208***	54***	13***	0
Problem Set 94	127***	132***	151***	93***	79***	77***	8**
Problem Set 95	97***	138***	65***	2	0	0	0
Problem Set 96	90***	86***	16***	0	0	0	0
Problem Set 97	280***	188***	75***	0	2	0	0
Problem Set 98	190***	135***	50***	0	0	0	0
Problem Set 99	104***	179***	218***	138***	101***	65***	11**
Problem Set 100	155***	141***	148***	70***	16**	10**	0
Problem Set 101	144***	109***	117***	35***	4*	0	0
Problem Set 102	263***	228***	134***	14***	0	0	0
Problem Set 103	205***	70***	12**	0	0	0	0
Problem Set 104	-99***	-51*	-3	70**	65**	121***	50***
Problem Set 105	203***	169***	105***	9*	2	0	0
Problem Set 106	224***	89***	38***	2	0	0	0
Problem Set 107	215***	245***	239***	59***	12**	0	0
Problem Set 108	189***	100***	21***	0	0	0	0
Problem Set 109	274***	106***	10**	0	0	0	0
Problem Set 110	261***	320***	121***	8**	0	0	0
Problem Set 111	194***	121***	37***	2	0	0	0
Problem Set 112	28***	13**	1	0	0	0	0
Problem Set 113	268***	196***	173***	41***	15***	0	0
Problem Set 114	207***	122***	18***	0	0	0	0
Problem Set 115	156***	125***	42***	3	0	0	0
Problem Set 116	238***	105***	41***	1	0	0	0
Problem Set 117	201***	95***	17***	0	0	0	0
Problem Set 118	238***	130***	55***	1	0	0	0
Problem Set 119	254***	156***	78***	10**	0	0	0
Problem Set 120	253***	246***	110***	25***	1	0	0
Problem Set 121	238***	125***	27***	1	0	0	0
Problem Set 122	51**	84***	66***	27***	2	0	0
Problem Set 123	101***	74***	66***	6*	0	0	0
Problem Set 124	231***	259***	253***	71***	7**	0	0
Problem Set 125	298***	274***	106***	4*	0	0	0
Problem Set 126	224***	101***	49***	2	0	0	0

TABLE A.1: *Evaluation of Provider's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 127	-63***	-41*	20	88***	68***	45***	6*
Problem Set 128	208***	115***	111***	16***	2	0	0
Problem Set 129	206***	121***	84***	18***	0	0	0
Problem Set 130	243***	135***	86***	10**	1	0	0
Problem Set 131	222***	103***	37***	0	0	0	0
Problem Set 132	110***	146***	91***	49*	68**	103***	74***
Problem Set 133	156***	98***	39***	2	0	0	0
Problem Set 134	124***	97***	16***	1	0	0	0
Problem Set 135	204***	103***	28***	0	0	0	0
Problem Set 136	280***	179***	42***	0	0	0	0
Problem Set 137	60***	38***	18***	0	0	0	0
Problem Set 138	113***	106***	36***	0	0	0	0
Problem Set 139	195***	197***	120***	26***	9**	0	0
Problem Set 140	168***	250***	241***	60***	12**	1	0
Problem Set 141	217***	173***	102***	32***	3	0	0
Problem Set 142	272***	118***	18***	0	0	0	0
Problem Set 143	89***	37***	3	0	0	0	0
Problem Set 144	248***	143***	41***	0	0	0	0
Problem Set 145	207***	83***	31***	0	0	0	0
Problem Set 146	170***	165***	161***	46***	22***	2	0
Problem Set 147	168***	160***	146***	117***	45***	14***	0
Problem Set 148	167***	128***	42***	1	0	0	0
Problem Set 149	175***	127***	31***	1	0	0	0
Problem Set 150	312***	114***	25***	0	0	0	0

TABLE A.2: *Evaluation of Provider's Risk-minimizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 1	0.09015	100
Problem Set 2	0.08995	50
Problem Set 3	0.01771	1000
Problem Set 4	0.09013	50
Problem Set 5	0.08855	100
Problem Set 6	0.08634	50
Problem Set 7	0.08552	50
Problem Set 8	0.08718	50
Problem Set 9	0.07197	50
Problem Set 10	0.00596	20
Problem Set 11	0.08461	50
Problem Set 12	0.08679	200
Problem Set 13	0.07081	200
Problem Set 14	0.08405	100
Problem Set 15	0.06730	50
Problem Set 16	0.01351	100
Problem Set 17	0.08475	50
Problem Set 18	0.08992	200

TABLE A.2: *Evaluation of Provider's Risk-minimizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 19	0.01666	1000
Problem Set 20	0.08965	500
Problem Set 21	0.09017	500
Problem Set 22	0.02855	50
Problem Set 23	0.07607	50
Problem Set 24	0.04100	50
Problem Set 25	0.06530	100
Problem Set 26	0.04025	50
Problem Set 27	0.08371	50
Problem Set 28	0.03600	50
Problem Set 29	0.08672	50
Problem Set 30	0.05791	50
Problem Set 31	0.08917	50
Problem Set 32	0.06847	100
Problem Set 33	0.06269	50
Problem Set 34	0.08401	50
Problem Set 35	0.02896	20
Problem Set 36	0.04775	100
Problem Set 37	0.08540	1000
Problem Set 38	0.08948	50
Problem Set 39	0.05675	100
Problem Set 40	0.08993	50
Problem Set 41	0.07616	100
Problem Set 42	0.08959	50
Problem Set 43	0.08979	50
Problem Set 44	0.01700	10
Problem Set 45	0.04781	50
Problem Set 46	0.07936	200
Problem Set 47	0.08567	500
Problem Set 48	0.08396	50
Problem Set 49	0.07198	50
Problem Set 50	0.06149	100
Problem Set 51	0.09016	50
Problem Set 52	0.08795	500
Problem Set 53	0.08931	50
Problem Set 54	0.08946	200
Problem Set 55	0.08884	50
Problem Set 56	0.09017	50
Problem Set 57	0.08852	50
Problem Set 58	0.01467	50
Problem Set 59	0.08011	50
Problem Set 60	0.01640	50
Problem Set 61	0.04401	200
Problem Set 62	0.08741	500
Problem Set 63	0.08779	50
Problem Set 64	0.06829	50
Problem Set 65	0.07262	100
Problem Set 66	0.04295	50
Problem Set 67	0.06775	50

TABLE A.2: *Evaluation of Provider's Risk-minimizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 68	0.05871	50
Problem Set 69	0.06864	100
Problem Set 70	0.01041	500
Problem Set 71	0.08551	50
Problem Set 72	0.08883	500
Problem Set 73	0.03148	50
Problem Set 74	0.08920	50
Problem Set 75	0.06103	50
Problem Set 76	0.08933	50
Problem Set 77	0.08396	50
Problem Set 78	0.07759	100
Problem Set 79	0.08522	100
Problem Set 80	0.08696	50
Problem Set 81	0.08765	50
Problem Set 82	0.05383	50
Problem Set 83	0.06370	50
Problem Set 84	0.08611	50
Problem Set 85	0.08756	100
Problem Set 86	0.08881	100
Problem Set 87	0.08862	50
Problem Set 88	0.08898	500
Problem Set 89	0.04998	50
Problem Set 90	0.08372	50
Problem Set 91	0.08199	50
Problem Set 92	0.08647	50
Problem Set 93	0.08469	500
Problem Set 94	0.07732	1000
Problem Set 95	0.04257	50
Problem Set 96	0.03969	50
Problem Set 97	0.09015	50
Problem Set 98	0.08898	50
Problem Set 99	0.08475	1000
Problem Set 100	0.06081	500
Problem Set 101	0.01886	200
Problem Set 102	0.06553	100
Problem Set 103	0.08634	50
Problem Set 104	0.01838	20
Problem Set 105	0.07489	100
Problem Set 106	0.05989	50
Problem Set 107	0.07278	200
Problem Set 108	0.08989	50
Problem Set 109	0.08729	50
Problem Set 110	0.09013	100
Problem Set 111	0.08549	50
Problem Set 112	0.00920	20
Problem Set 113	0.08933	200
Problem Set 114	0.09017	50
Problem Set 115	0.07653	50
Problem Set 116	0.08443	50

TABLE A.2: *Evaluation of Provider's Risk-minimizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 117	0.08530	50
Problem Set 118	0.08924	50
Problem Set 119	0.08992	100
Problem Set 120	0.08985	100
Problem Set 121	0.08393	50
Problem Set 122	0.03794	100
Problem Set 123	0.01506	100
Problem Set 124	0.08944	200
Problem Set 125	0.08857	100
Problem Set 126	0.08796	50
Problem Set 127	0.01068	20
Problem Set 128	0.08706	100
Problem Set 129	0.08625	100
Problem Set 130	0.07635	100
Problem Set 131	0.08749	50
Problem Set 132	0.08238	1000
Problem Set 133	0.05142	50
Problem Set 134	0.04994	50
Problem Set 135	0.08302	50
Problem Set 136	0.08450	50
Problem Set 137	0.00881	50
Problem Set 138	0.05312	50
Problem Set 139	0.06742	200
Problem Set 140	0.07246	200
Problem Set 141	0.08649	100
Problem Set 142	0.08812	50
Problem Set 143	0.02400	20
Problem Set 144	0.07842	50
Problem Set 145	0.07767	50
Problem Set 146	0.08956	200
Problem Set 147	0.08299	500
Problem Set 148	0.06526	50
Problem Set 149	0.07832	50
Problem Set 150	0.08943	50

A.1.2 Maximization of Expected Utility

TABLE A.3: *Evaluation of Provider's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 1	94***	108***	87***	73***	32***	7**	0
Problem Set 2	77***	119***	123***	66***	22***	7**	0
Problem Set 3	36	69**	137***	75***	76***	41***	6*
Problem Set 4	-30	87***	27	117***	70**	157***	119***
Problem Set 5	146***	117***	68***	10**	0	0	0

TABLE A.3: *Evaluation of Provider's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 6	82***	102***	100***	82***	51***	4*	0
Problem Set 7	144***	128***	93***	32***	5*	0	0
Problem Set 8	140***	65**	128***	136***	69***	70***	0
Problem Set 9	181***	151***	135***	25***	2	0	0
Problem Set 10	154***	115***	90***	4*	0	0	0
Problem Set 11	56*	54*	60**	6	39	59**	91***
Problem Set 12	135***	88***	52***	6*	0	0	0
Problem Set 13	36	65**	75***	101***	85***	54***	6*
Problem Set 14	69**	65**	125***	85***	49***	8**	0
Problem Set 15	99***	104***	111***	16***	0	0	0
Problem Set 16	90***	102***	161***	100***	64***	26***	0
Problem Set 17	89***	94***	220***	55***	21***	0	0
Problem Set 18	25	61**	19	38	39	116***	91***
Problem Set 19	99***	95***	36***	8**	0	0	0
Problem Set 20	59**	27	73**	37	69**	101***	74***
Problem Set 21	48*	95***	111***	115***	58***	76***	11**
Problem Set 22	36	31	47*	16	-3	65**	-8
Problem Set 23	85***	58**	41*	86***	90***	85***	29***
Problem Set 24	136***	109***	29***	2	0	0	0
Problem Set 25	77***	102***	89***	88***	65***	36***	0
Problem Set 26	129***	89***	49***	6*	0	0	0
Problem Set 27	81***	107***	144***	80***	27***	2	0
Problem Set 28	72**	99***	102***	103***	47***	22***	2
Problem Set 29	31	87***	85***	74***	81***	123***	63***
Problem Set 30	43	19	100***	5	26	76***	105***
Problem Set 31	79***	90***	132***	99***	72***	38***	2
Problem Set 32	25	112***	86***	92***	66***	67***	16**
Problem Set 33	151***	130***	52***	9**	0	0	0
Problem Set 34	71**	93***	168***	63***	34***	2	0
Problem Set 35	44***	25***	8**	0	0	0	0
Problem Set 36	129***	94***	94***	7*	2	0	0
Problem Set 37	74**	140***	119***	58***	25**	20***	0
Problem Set 38	56*	85***	75**	149***	68***	96***	42***
Problem Set 39	116***	57**	97***	48***	12**	0	0
Problem Set 40	69**	59**	93***	82***	122***	145***	100***
Problem Set 41	134***	115***	62***	13***	0	0	0
Problem Set 42	21	35	55*	34	69**	111***	57***
Problem Set 43	51*	2	6	4	-2	26	23
Problem Set 44	113***	73***	75***	8*	3	0	0
Problem Set 45	65**	75**	137***	100***	61**	114***	37***
Problem Set 46	22	60**	102***	69***	78***	77***	17***
Problem Set 47	30	39	162***	51**	77***	33***	3
Problem Set 48	66**	107***	108***	87***	71***	54***	9**
Problem Set 49	116***	115***	85***	16***	2	0	0
Problem Set 50	80***	111***	145***	137***	101***	51***	3
Problem Set 51	105***	138***	122***	37***	6*	0	0
Problem Set 52	102***	136***	106***	31***	6*	0	0
Problem Set 53	147***	144***	113***	18***	1	0	0
Problem Set 54	-46*	-5	-2	6	57*	-10	-10
Problem Set 55	94***	121***	99***	80***	32**	31***	1

TABLE A.3: *Evaluation of Provider's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 56	120***	116***	90***	20***	4*	0	0
Problem Set 57	88***	-8	109***	37	77***	86***	72***
Problem Set 58	81***	58**	97***	87***	39*	110***	25***
Problem Set 59	120***	113***	75***	11**	0	0	0
Problem Set 60	97***	116***	97***	58***	31***	3	0
Problem Set 61	54**	102***	84***	27***	15***	0	0
Problem Set 62	127***	87***	81***	8*	1	0	0
Problem Set 63	119***	125***	102***	14***	1	0	0
Problem Set 64	87***	87***	163***	92***	63***	6*	0
Problem Set 65	98***	104***	136***	87***	70***	30***	1
Problem Set 66	148***	109***	81***	12**	1	0	0
Problem Set 67	112***	47*	156***	108***	106***	61***	1
Problem Set 68	78***	55*	54*	78***	35	59**	60**
Problem Set 69	135***	99***	105***	53***	14***	0	0
Problem Set 70	97***	71***	42***	2	0	0	0
Problem Set 71	109***	73***	96***	58***	12**	3	0
Problem Set 72	-17	24	5	10	-11	21	23
Problem Set 73	101***	109***	167***	60***	22***	4*	0
Problem Set 74	153***	40	166***	59***	50***	9**	0
Problem Set 75	95***	-12	7	-6	24	-13	22
Problem Set 76	86***	146***	150***	69***	37***	9**	0
Problem Set 77	117***	108***	144***	101***	41***	8**	0
Problem Set 78	42	84***	159***	129***	103***	59***	3
Problem Set 79	123***	44*	129***	75***	76***	45***	4*
Problem Set 80	128***	125***	50***	5*	0	0	0
Problem Set 81	8	69**	85***	48*	35	148***	114***
Problem Set 82	24	103***	124***	95***	102***	78***	23***
Problem Set 83	135***	105***	175***	91***	37***	1	0
Problem Set 84	76**	108***	123***	75***	41***	12**	1
Problem Set 85	101***	58**	110***	89***	78***	86***	13***
Problem Set 86	60**	6	77***	64**	85***	95***	54***
Problem Set 87	58**	32	139***	99***	102***	126***	73***
Problem Set 88	84***	79***	85***	20***	2	0	0
Problem Set 89	103***	126***	161***	95***	52***	16***	0
Problem Set 90	66**	128***	134***	39***	10**	2	0
Problem Set 91	63**	91***	71**	201***	108***	96***	11**
Problem Set 92	102***	147***	147***	68***	23***	1	0
Problem Set 93	-3	39	55*	63**	29	133***	60***
Problem Set 94	80***	45***	6*	0	0	0	0
Problem Set 95	107***	131***	156***	85***	30***	4*	0
Problem Set 96	70**	125***	131***	57***	53***	10**	1
Problem Set 97	42	62**	91***	45*	82***	107***	44***
Problem Set 98	86***	134***	127***	75***	21***	8**	0
Problem Set 99	98***	81***	51***	5*	0	0	0
Problem Set 100	21	-37	49*	-21	21	-40	-44*
Problem Set 101	40	92***	86***	77***	99***	80***	37***
Problem Set 102	11	81***	109***	55**	91***	54***	6*
Problem Set 103	86***	134***	105***	89***	41***	26***	1
Problem Set 104	46*	40	22	71**	72**	106***	89***
Problem Set 105	107***	121***	99***	41***	11**	0	0

TABLE A.3: *Evaluation of Provider's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 106	70**	52*	139***	70**	120***	97***	29***
Problem Set 107	7	88***	13	64**	52**	126***	58***
Problem Set 108	92***	67**	202***	162***	98***	23***	2
Problem Set 109	41	115***	124***	86***	65***	40***	4*
Problem Set 110	87***	77***	134***	52***	10**	2	0
Problem Set 111	126***	109***	133***	79***	29***	3	0
Problem Set 112	104***	95***	12**	0	0	0	0
Problem Set 113	68**	62**	93***	82***	52***	25***	0
Problem Set 114	52*	56*	45*	28	83***	60**	89***
Problem Set 115	121***	57*	215***	75***	91***	15***	1
Problem Set 116	100***	99***	165***	101***	86***	36***	4*
Problem Set 117	152***	106***	74***	21***	3	0	0
Problem Set 118	89***	24	56*	74**	69**	93***	91***
Problem Set 119	34	55*	56*	59**	60**	106***	78***
Problem Set 120	84***	62**	121***	92***	54***	62***	11**
Problem Set 121	78***	95***	178***	141***	95***	60***	4
Problem Set 122	123***	86***	44***	3	0	0	0
Problem Set 123	113***	73***	46***	7**	0	0	0
Problem Set 124	36	36	126***	54**	30	146***	53***
Problem Set 125	20	72**	134***	114***	77***	44***	3
Problem Set 126	66**	27	28	58**	29	37	59**
Problem Set 127	133***	81***	73***	15***	1	0	0
Problem Set 128	101***	124***	133***	66***	34***	3	0
Problem Set 129	126***	119***	59***	4*	0	0	0
Problem Set 130	85***	66**	94***	106***	58***	53***	6*
Problem Set 131	109***	110***	132***	19***	3	0	0
Problem Set 132	40	57*	19	80***	37	110***	83***
Problem Set 133	8	24	35	60**	53*	114***	126***
Problem Set 134	159***	130***	54***	9**	0	0	0
Problem Set 135	100***	102***	113***	59***	21***	1	0
Problem Set 136	53*	123***	115***	68***	19***	1	0
Problem Set 137	175***	64**	149***	88***	19*	19***	0
Problem Set 138	138***	158***	135***	35***	3	0	0
Problem Set 139	110***	76***	99***	73***	52***	17***	0
Problem Set 140	61***	107***	44***	3	0	0	0
Problem Set 141	88***	87***	96***	27***	7*	1	0
Problem Set 142	110***	126***	153***	86***	19***	2	0
Problem Set 143	62**	73**	104***	83***	90***	56***	14***
Problem Set 144	19	11	45*	66**	11	89***	62**
Problem Set 145	119***	140***	112***	29***	11**	0	0
Problem Set 146	120***	105***	129***	60***	21***	3	0
Problem Set 147	124***	88***	93***	52***	7**	0	0
Problem Set 148	159***	82***	106***	17**	5*	0	0
Problem Set 149	151***	93***	45***	3	0	0	0
Problem Set 150	-2	28	12	25	13	34	69**

TABLE A.4: *Evaluation of Provider's Utility-maximizing Decision: Variance and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 1	0.07243	500
Problem Set 2	0.08900	500
Problem Set 3	0.08873	1000
Problem Set 4	0.08611	1000
Problem Set 5	0.07048	100
Problem Set 6	0.09014	500
Problem Set 7	0.08807	200
Problem Set 8	0.09017	500
Problem Set 9	0.08991	200
Problem Set 10	0.08008	100
Problem Set 11	0.07580	1000
Problem Set 12	0.08314	100
Problem Set 13	0.08964	1000
Problem Set 14	0.08582	500
Problem Set 15	0.08905	100
Problem Set 16	0.08765	500
Problem Set 17	0.08249	200
Problem Set 18	0.07184	1000
Problem Set 19	0.06392	100
Problem Set 20	0.08121	1000
Problem Set 21	0.08411	1000
Problem Set 22	0.06554	500
Problem Set 23	0.09003	1000
Problem Set 24	0.07589	50
Problem Set 25	0.08828	500
Problem Set 26	0.08755	100
Problem Set 27	0.08883	200
Problem Set 28	0.08830	500
Problem Set 29	0.08563	1000
Problem Set 30	0.09001	1000
Problem Set 31	0.09007	500
Problem Set 32	0.08974	1000
Problem Set 33	0.08904	100
Problem Set 34	0.08898	200
Problem Set 35	0.03801	50
Problem Set 36	0.07555	100
Problem Set 37	0.06842	500
Problem Set 38	0.08996	1000
Problem Set 39	0.09003	200
Problem Set 40	0.06100	1000
Problem Set 41	0.08937	100
Problem Set 42	0.07233	1000
Problem Set 43	0.07692	10
Problem Set 44	0.08955	100
Problem Set 45	0.08813	1000
Problem Set 46	0.08001	1000
Problem Set 47	0.08423	500
Problem Set 48	0.07761	1000
Problem Set 49	0.08860	100

TABLE A.4: *Evaluation of Provider's Utility-maximizing Decision: Variance and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 50	0.08737	500
Problem Set 51	0.08893	200
Problem Set 52	0.08940	200
Problem Set 53	0.08103	100
Problem Set 54	0.05387	200
Problem Set 55	0.09005	500
Problem Set 56	0.08898	200
Problem Set 57	0.06765	1000
Problem Set 58	0.07560	1000
Problem Set 59	0.08978	100
Problem Set 60	0.09010	200
Problem Set 61	0.08942	200
Problem Set 62	0.07854	100
Problem Set 63	0.08273	100
Problem Set 64	0.09005	500
Problem Set 65	0.08947	500
Problem Set 66	0.07392	100
Problem Set 67	0.08975	500
Problem Set 68	0.09016	1000
Problem Set 69	0.08380	200
Problem Set 70	0.08372	50
Problem Set 71	0.08407	200
Problem Set 72	0.07045	5
Problem Set 73	0.08046	500
Problem Set 74	0.08470	500
Problem Set 75	0.08987	10
Problem Set 76	0.08641	500
Problem Set 77	0.07637	500
Problem Set 78	0.09004	500
Problem Set 79	0.08869	1000
Problem Set 80	0.08996	100
Problem Set 81	0.09017	1000
Problem Set 82	0.08776	1000
Problem Set 83	0.08909	200
Problem Set 84	0.08541	500
Problem Set 85	0.07212	1000
Problem Set 86	0.07445	1000
Problem Set 87	0.07052	1000
Problem Set 88	0.06718	100
Problem Set 89	0.08999	500
Problem Set 90	0.07462	200
Problem Set 91	0.08974	1000
Problem Set 92	0.08685	200
Problem Set 93	0.08448	1000
Problem Set 94	0.08982	50
Problem Set 95	0.09014	500
Problem Set 96	0.08368	500
Problem Set 97	0.08960	1000
Problem Set 98	0.08615	500

TABLE A.4: *Evaluation of Provider's Utility-maximizing Decision: Variance and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 99	0.07495	100
Problem Set 100	0.08808	1000
Problem Set 101	0.08864	1000
Problem Set 102	0.08976	1000
Problem Set 103	0.08044	500
Problem Set 104	0.08186	1000
Problem Set 105	0.08018	200
Problem Set 106	0.09014	1000
Problem Set 107	0.09015	1000
Problem Set 108	0.06814	500
Problem Set 109	0.07022	1000
Problem Set 110	0.05238	200
Problem Set 111	0.08989	200
Problem Set 112	0.05876	50
Problem Set 113	0.07894	500
Problem Set 114	0.07699	1000
Problem Set 115	0.08969	500
Problem Set 116	0.05210	1000
Problem Set 117	0.08898	100
Problem Set 118	0.08974	1000
Problem Set 119	0.08997	1000
Problem Set 120	0.08769	1000
Problem Set 121	0.05139	500
Problem Set 122	0.06379	50
Problem Set 123	0.08994	100
Problem Set 124	0.08031	1000
Problem Set 125	0.06565	500
Problem Set 126	0.08765	1000
Problem Set 127	0.08294	100
Problem Set 128	0.07362	200
Problem Set 129	0.06901	100
Problem Set 130	0.08981	1000
Problem Set 131	0.07137	100
Problem Set 132	0.08733	1000
Problem Set 133	0.08902	1000
Problem Set 134	0.07320	100
Problem Set 135	0.08890	200
Problem Set 136	0.07880	200
Problem Set 137	0.09016	500
Problem Set 138	0.09013	100
Problem Set 139	0.09012	500
Problem Set 140	0.09012	50
Problem Set 141	0.08522	200
Problem Set 142	0.09017	200
Problem Set 143	0.07429	1000
Problem Set 144	0.08779	1000
Problem Set 145	0.08979	200
Problem Set 146	0.08289	200
Problem Set 147	0.08973	200

TABLE A.4: *Evaluation of Provider's Utility-maximizing Decision: Variance and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 148	0.08975	200
Problem Set 149	0.08761	50
Problem Set 150	0.08990	1000

A.2 Evaluation of Intermediary's Decision Methods

A.2.1 Minimization of Semi-Variance

TABLE A.5: *Evaluation of Intermediary's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 1	207***	266***	163***	21***	-65	0	0
Problem Set 2	142***	109***	185***	76***	33***	104***	69***
Problem Set 3	181***	180***	249***	69***	-106**	0	0
Problem Set 4	173***	190***	260***	85***	2***	2	0
Problem Set 5	206***	275***	199***	34***	37	1	0
Problem Set 6	162***	187***	214***	109***	42***	1	0
Problem Set 7	195***	239***	205***	75***	18**	0	0
Problem Set 8	89***	125***	194***	158***	-147***	93***	56***
Problem Set 9	158***	166***	207***	122***	42***	8**	0
Problem Set 10	147***	142***	225***	77***	-127***	3	0
Problem Set 11	149***	214***	166***	64***	98***	19***	2
Problem Set 12	27	-13	-96***	-36***	-86***	0	0
Problem Set 13	133***	166***	204***	94***	88***	23***	2
Problem Set 14	170***	111***	169***	56**	-40***	45***	1
Problem Set 15	151***	170***	183***	106***	-161***	4	1
Problem Set 16	164***	135***	217***	107***	-49***	50***	2
Problem Set 17	206***	146***	207***	101***	-244***	3	0
Problem Set 18	212***	181***	200***	61***	71***	1	0
Problem Set 19	189***	155***	191***	86***	-61***	2	0
Problem Set 20	207***	217***	208***	33***	76*	0	0
Problem Set 21	167***	198***	244***	127***	104***	5*	0
Problem Set 22	221***	185***	193***	58***	-41**	1	0
Problem Set 23	74***	207***	251***	204***	105***	50***	4*
Problem Set 24	269***	260***	144***	27***	42	0	0
Problem Set 25	126***	156***	254***	179***	171***	30***	0
Problem Set 26	168***	151***	299***	107***	40***	3	0
Problem Set 27	193***	240***	148***	38***	6*	1	0
Problem Set 28	138***	184***	180***	128***	62***	11**	0
Problem Set 29	11	50***	9	6	-40**	-30**	-23***
Problem Set 30	65***	62***	165***	81***	103***	162***	73***
Problem Set 31	89***	206***	216***	146***	87***	24***	1

TABLE A.5: *Evaluation of Intermediary's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 32	187***	216***	231***	43***	4*	0	0
Problem Set 33	175***	229***	213***	64***	9*	2	0
Problem Set 34	198***	195***	226***	66***	26***	6*	0
Problem Set 35	112***	148***	155***	124***	74***	32***	1
Problem Set 36	203***	133***	133***	91***	43***	27***	0
Problem Set 37	88***	85***	164***	6	36	-7	2
Problem Set 38	176***	184***	211***	106***	11**	2	0
Problem Set 39	112***	165***	322***	109***	53***	14***	0
Problem Set 40	186***	233***	178***	56***	12**	0	0
Problem Set 41	146***	212***	206***	81***	31***	5*	0
Problem Set 42	56***	64***	89***	28	44*	21	11
Problem Set 43	231***	204***	164***	16**	4*	0	0
Problem Set 44	70***	65***	134***	87***	159***	150***	145***
Problem Set 45	87***	47*	37	-14	-77***	-98***	-99***
Problem Set 46	172***	230***	209***	42***	3	1	0
Problem Set 47	123***	105***	110***	42	93***	85***	69***
Problem Set 48	183***	196***	222***	118***	52***	9**	0
Problem Set 49	289***	229***	127***	11**	0	0	0
Problem Set 50	45**	54**	118***	47*	118***	140***	92***
Problem Set 51	162***	220***	199***	62***	7**	0	0
Problem Set 52	183***	188***	273***	108***	26***	4*	0
Problem Set 53	254***	236***	118***	9**	0	0	0
Problem Set 54	167***	140***	212***	136***	77***	16***	1
Problem Set 55	200***	163***	204***	76***	5*	0	0
Problem Set 56	139***	182***	264***	154***	94***	17***	1
Problem Set 57	141***	153***	198***	84***	27***	9**	1
Problem Set 58	193***	171***	289***	109***	32***	4*	0
Problem Set 59	269***	188***	155***	16***	0	0	0
Problem Set 60	78***	72***	163***	131***	81***	115***	94***
Problem Set 61	57***	68***	167***	111***	114***	126***	127***
Problem Set 62	161***	224***	233***	49***	5*	0	0
Problem Set 63	99***	212***	214***	124***	83***	39***	3
Problem Set 64	181***	199***	215***	82***	17***	1	0
Problem Set 65	140***	149***	293***	143***	31***	6*	0
Problem Set 66	203***	204***	160***	9*	3	0	0
Problem Set 67	131***	138***	155***	159***	88***	42***	0
Problem Set 68	236***	209***	153***	25***	0	0	0
Problem Set 69	193***	223***	243***	42***	7*	1	0
Problem Set 70	79***	112***	235***	186***	128***	116***	12**
Problem Set 71	121***	103***	184***	50*	96***	108***	31***
Problem Set 72	167***	179***	224***	75***	14***	0	0
Problem Set 73	116***	154***	275***	186***	58***	9**	0
Problem Set 74	287***	226***	146***	20***	0	0	0
Problem Set 75	124***	51*	-51*	-63**	-86***	-80***	-72***
Problem Set 76	186***	181***	187***	38***	7**	0	0
Problem Set 77	230***	223***	170***	23***	1	0	0
Problem Set 78	24***	23**	26*	-34**	-40***	-4*	0
Problem Set 79	163***	166***	247***	81***	62***	12**	0
Problem Set 80	177***	129***	203***	93***	26***	4*	0

TABLE A.5: *Evaluation of Intermediary's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 81	219***	193***	217***	43***	14***	0	0
Problem Set 82	249***	233***	142***	14***	0	0	0
Problem Set 83	121***	208***	246***	71***	13**	1	0
Problem Set 84	173***	118***	178***	108***	60***	25***	4*
Problem Set 85	160***	184***	218***	72***	35***	2	0
Problem Set 86	271***	237***	141***	19***	0	0	0
Problem Set 87	101***	164***	219***	160***	114***	62***	3
Problem Set 88	6	20**	18	-6	-38***	-12**	-1
Problem Set 89	171***	196***	247***	60***	14***	0	0
Problem Set 90	130***	164***	180***	113***	57***	30***	0
Problem Set 91	-25	-36**	-39***	-3	0	0	0
Problem Set 92	2	-35*	-81***	-39***	-24***	-4*	0
Problem Set 93	110***	134***	246***	184***	146***	53***	5*
Problem Set 94	223***	225***	198***	19***	2	0	0
Problem Set 95	271***	197***	154***	23***	1	0	0
Problem Set 96	204***	227***	188***	34***	5*	0	0
Problem Set 97	179***	169***	177***	58***	65***	53***	10**
Problem Set 98	267***	218***	157***	26***	1	0	0
Problem Set 99	-11	1	-30*	-34***	-21**	-9**	0
Problem Set 100	113***	102***	169***	157***	135***	88***	10**
Problem Set 101	137***	121***	220***	90***	92***	76***	4*
Problem Set 102	185***	222***	207***	73***	12**	2	0
Problem Set 103	113***	145***	179***	121***	86***	58***	5
Problem Set 104	125***	102***	257***	163***	141***	80***	2
Problem Set 105	136***	186***	226***	106***	53***	7**	0
Problem Set 106	215***	193***	230***	59***	8*	1	0
Problem Set 107	160***	180***	275***	91***	26***	4*	0
Problem Set 108	206***	228***	213***	61***	7*	1	0
Problem Set 109	145***	144***	200***	130***	140***	84***	36***
Problem Set 110	238***	229***	160***	3	0	0	0
Problem Set 111	199***	262***	202***	11**	0	0	0
Problem Set 112	110***	149***	296***	183***	69***	18***	0
Problem Set 113	178***	253***	130***	22***	0	0	0
Problem Set 114	88***	80***	155***	168***	92***	149***	84***
Problem Set 115	182***	222***	190***	57***	8*	1	0
Problem Set 116	119***	214***	286***	167***	59***	24***	0
Problem Set 117	96***	15	-26	-61**	-73***	-109***	-27***
Problem Set 118	174***	236***	194***	54***	9*	3	0
Problem Set 119	136***	144***	265***	186***	100***	33***	0
Problem Set 120	137***	214***	242***	59***	11**	1	0
Problem Set 121	124***	102***	154***	176***	107***	83***	14***
Problem Set 122	112***	158***	244***	127***	116***	46***	8**
Problem Set 123	126***	160***	279***	174***	111***	20***	0
Problem Set 124	128***	176***	276***	158***	110***	22***	0
Problem Set 125	237***	276***	122***	10**	0	0	0
Problem Set 126	176***	226***	183***	67***	9**	0	0
Problem Set 127	184***	178***	230***	105***	19***	1	0
Problem Set 128	244***	195***	171***	41***	4*	0	0
Problem Set 129	31***	81***	115***	31	127***	70**	56*

TABLE A.5: *Evaluation of Intermediary's Risk-minimizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 130	35	-27	-67**	37	9	73**	67**
Problem Set 131	155***	155***	293***	104***	56***	8*	1
Problem Set 132	162***	225***	202***	56***	7**	0	0
Problem Set 133	136***	153***	115***	26	-2	32	46*
Problem Set 134	193***	244***	187***	50***	7**	0	0
Problem Set 135	95***	128***	110***	91***	53**	116***	68***
Problem Set 136	122***	174***	194***	42**	54***	19***	0
Problem Set 137	177***	200***	174***	72***	19***	1	0
Problem Set 138	194***	267***	190***	52***	5*	0	0
Problem Set 139	156***	223***	219***	59***	6*	0	0
Problem Set 140	176***	168***	224***	120***	28***	0	0
Problem Set 141	117***	223***	215***	152***	57***	8**	0
Problem Set 142	65***	119***	247***	219***	145***	114***	28***
Problem Set 143	166***	128***	195***	151***	67***	63***	10**
Problem Set 144	212***	246***	180***	27***	3	0	0
Problem Set 145	161***	172***	128***	78***	24*	38***	7**
Problem Set 146	216***	226***	192***	34***	5*	0	0
Problem Set 147	158***	179***	221***	52***	23***	3	0
Problem Set 148	70**	64**	19	-37	-43*	-79***	-85***
Problem Set 149	152***	162***	211***	135***	90***	57***	2
Problem Set 150	193***	188***	189***	75***	38***	3	0

TABLE A.6: *Evaluation of Intermediary's Risk-minimizing Decision: Maximum Variance in Portfolio and required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 1	0.09013	100
Problem Set 2	0.05887	1000
Problem Set 3	0.08835	200
Problem Set 4	0.07133	200
Problem Set 5	0.09007	100
Problem Set 6	0.08546	200
Problem Set 7	0.08985	200
Problem Set 8	0.08970	1000
Problem Set 9	0.08982	500
Problem Set 10	0.07263	200
Problem Set 11	0.05876	500
Problem Set 12	0.00000	200
Problem Set 13	0.07125	500
Problem Set 14	0.09015	500
Problem Set 15	0.08932	200
Problem Set 16	0.09008	500
Problem Set 17	0.08451	200
Problem Set 18	0.08868	200
Problem Set 19	0.08921	200
Problem Set 20	0.08275	200

TABLE A.6: *Evaluation of Intermediary's Risk-minimizing Decision: Maximum Variance in Portfolio and required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 21	0.08523	500
Problem Set 22	0.08817	200
Problem Set 23	0.08938	1000
Problem Set 24	0.07671	100
Problem Set 25	0.06457	500
Problem Set 26	0.08826	200
Problem Set 27	0.06697	200
Problem Set 28	0.08771	500
Problem Set 29	0.00000	20
Problem Set 30	0.09015	1000
Problem Set 31	0.08399	500
Problem Set 32	0.06798	200
Problem Set 33	0.08744	200
Problem Set 34	0.08986	500
Problem Set 35	0.08362	500
Problem Set 36	0.08295	500
Problem Set 37	0.08807	50
Problem Set 38	0.06986	200
Problem Set 39	0.05466	500
Problem Set 40	0.07930	200
Problem Set 41	0.07543	500
Problem Set 42	0.05160	50
Problem Set 43	0.08510	200
Problem Set 44	0.09016	1000
Problem Set 45	0.06312	20
Problem Set 46	0.07380	100
Problem Set 47	0.06845	50
Problem Set 48	0.09006	500
Problem Set 49	0.07299	100
Problem Set 50	0.03277	1000
Problem Set 51	0.07295	200
Problem Set 52	0.04579	500
Problem Set 53	0.08032	100
Problem Set 54	0.09014	500
Problem Set 55	0.08176	200
Problem Set 56	0.08956	500
Problem Set 57	0.07493	500
Problem Set 58	0.09017	500
Problem Set 59	0.08822	100
Problem Set 60	0.06055	1000
Problem Set 61	0.05083	1000
Problem Set 62	0.08565	200
Problem Set 63	0.07313	500
Problem Set 64	0.08509	200
Problem Set 65	0.07423	500
Problem Set 66	0.08711	100
Problem Set 67	0.08788	500
Problem Set 68	0.06813	100
Problem Set 69	0.08295	200

TABLE A.6: *Evaluation of Intermediary's Risk-minimizing Decision: Maximum Variance in Portfolio and required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 70	0.08797	1000
Problem Set 71	0.09015	1000
Problem Set 72	0.08873	200
Problem Set 73	0.08724	500
Problem Set 74	0.07829	100
Problem Set 75	0.00000	1000
Problem Set 76	0.08906	200
Problem Set 77	0.07219	100
Problem Set 78	0.00000	500
Problem Set 79	0.08981	500
Problem Set 80	0.08480	500
Problem Set 81	0.08096	200
Problem Set 82	0.08182	100
Problem Set 83	0.06280	200
Problem Set 84	0.08724	1000
Problem Set 85	0.07926	200
Problem Set 86	0.07265	100
Problem Set 87	0.04699	500
Problem Set 88	0.00000	20
Problem Set 89	0.06483	200
Problem Set 90	0.08109	500
Problem Set 91	0.00000	50
Problem Set 92	0.00000	500
Problem Set 93	0.08301	1000
Problem Set 94	0.07289	100
Problem Set 95	0.08777	100
Problem Set 96	0.08497	200
Problem Set 97	0.08894	1000
Problem Set 98	0.07685	100
Problem Set 99	0.00000	500
Problem Set 100	0.09015	1000
Problem Set 101	0.09005	1000
Problem Set 102	0.08836	200
Problem Set 103	0.06302	500
Problem Set 104	0.07566	500
Problem Set 105	0.08948	500
Problem Set 106	0.08071	200
Problem Set 107	0.08033	500
Problem Set 108	0.08941	200
Problem Set 109	0.06156	1000
Problem Set 110	0.08471	50
Problem Set 111	0.08322	100
Problem Set 112	0.07882	500
Problem Set 113	0.08731	100
Problem Set 114	0.07920	1000
Problem Set 115	0.09012	200
Problem Set 116	0.07249	500
Problem Set 117	0.00000	10
Problem Set 118	0.08668	200

TABLE A.6: *Evaluation of Intermediary's Risk-minimizing Decision: Maximum Variance in Portfolio and required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 119	0.07067	500
Problem Set 120	0.06442	200
Problem Set 121	0.07842	1000
Problem Set 122	0.08809	1000
Problem Set 123	0.09008	500
Problem Set 124	0.08201	500
Problem Set 125	0.06498	100
Problem Set 126	0.08939	200
Problem Set 127	0.08533	200
Problem Set 128	0.08725	200
Problem Set 129	0.04688	50
Problem Set 130	0.00930	50
Problem Set 131	0.09015	500
Problem Set 132	0.08726	200
Problem Set 133	0.08384	50
Problem Set 134	0.08314	200
Problem Set 135	0.06093	1000
Problem Set 136	0.08099	500
Problem Set 137	0.07959	200
Problem Set 138	0.08975	200
Problem Set 139	0.08308	200
Problem Set 140	0.08799	200
Problem Set 141	0.07434	500
Problem Set 142	0.07583	1000
Problem Set 143	0.07061	1000
Problem Set 144	0.07611	100
Problem Set 145	0.07084	1000
Problem Set 146	0.08158	200
Problem Set 147	0.08736	200
Problem Set 148	0.06668	20
Problem Set 149	0.08915	500
Problem Set 150	0.08632	200

A.2.2 Maximization of Expected Utility

TABLE A.7: *Evaluation of Intermediary's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 1	60***	10**	0	0	0	0	0
Problem Set 2	0	0	0	0	0	0	0
Problem Set 3	21***	2	0	0	0	0	0
Problem Set 4	9**	1	0	0	0	0	0
Problem Set 5	0	0	0	0	0	0	0
Problem Set 6	7*	1	0	0	0	0	0

TABLE A.7: *Evaluation of Intermediary's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 7	55***	16***	0	0	0	0	0
Problem Set 8	3	0	0	0	0	0	0
Problem Set 9	9**	0	0	0	0	0	0
Problem Set 10	6*	0	0	0	0	0	0
Problem Set 11	119***	105***	39***	14***	0	0	0
Problem Set 12	18***	0	0	0	0	0	0
Problem Set 13	35***	4*	0	0	0	0	0
Problem Set 14	71***	16***	2	0	0	0	0
Problem Set 15	8**	0	0	0	0	0	0
Problem Set 16	6	2	0	0	0	0	0
Problem Set 17	21***	0	0	0	0	0	0
Problem Set 18	2	0	0	0	0	0	0
Problem Set 19	3	1	0	0	0	0	0
Problem Set 20	24***	3	0	0	0	0	0
Problem Set 21	1	0	0	0	0	0	0
Problem Set 22	27***	0	0	0	0	0	0
Problem Set 23	11**	1	0	0	0	0	0
Problem Set 24	3	0	0	0	0	0	0
Problem Set 25	0	0	0	0	0	0	0
Problem Set 26	57***	26***	1	0	0	0	0
Problem Set 27	23***	1	0	0	0	0	0
Problem Set 28	23***	2	0	0	0	0	0
Problem Set 29	52***	11**	3	0	0	0	0
Problem Set 30	0	0	0	0	0	0	0
Problem Set 31	4*	0	0	0	0	0	0
Problem Set 32	62***	24***	1	0	0	0	0
Problem Set 33	2	0	0	0	0	0	0
Problem Set 34	3	0	0	0	0	0	0
Problem Set 35	27***	2	0	0	0	0	0
Problem Set 36	94***	46***	6*	0	0	0	0
Problem Set 37	9**	0	0	0	0	0	0
Problem Set 38	19***	0	0	0	0	0	0
Problem Set 39	1	0	0	0	0	0	0
Problem Set 40	5*	0	0	0	0	0	0
Problem Set 41	35***	15***	0	0	0	0	0
Problem Set 42	0	0	0	0	0	0	0
Problem Set 43	20***	3	0	0	0	0	0
Problem Set 44	0	0	0	0	0	0	0
Problem Set 45	44***	17***	0	0	0	0	0
Problem Set 46	27***	8**	0	0	0	0	0
Problem Set 47	47***	8**	0	0	0	0	0
Problem Set 48	79***	15***	1	0	0	0	0
Problem Set 49	65***	15***	2	0	0	0	0
Problem Set 50	3	0	0	0	0	0	0
Problem Set 51	8*	1	0	0	0	0	0
Problem Set 52	102***	98***	76***	42***	10**	0	0
Problem Set 53	3	0	0	0	0	0	0
Problem Set 54	36***	5*	0	0	0	0	0
Problem Set 55	9**	0	0	0	0	0	0

TABLE A.7: *Evaluation of Intermediary's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 56	0	0	0	0	0	0	0
Problem Set 57	62***	9**	1	0	0	0	0
Problem Set 58	6*	0	0	0	0	0	0
Problem Set 59	0	0	0	0	0	0	0
Problem Set 60	7*	1	0	0	0	0	0
Problem Set 61	2	0	0	0	0	0	0
Problem Set 62	3	0	0	0	0	0	0
Problem Set 63	46***	6*	0	0	0	0	0
Problem Set 64	2	1	0	0	0	0	0
Problem Set 65	0	0	0	0	0	0	0
Problem Set 66	5*	0	0	0	0	0	0
Problem Set 67	49***	6*	0	0	0	0	0
Problem Set 68	8*	1	0	0	0	0	0
Problem Set 69	92***	64***	21***	2	0	0	0
Problem Set 70	10**	0	0	0	0	0	0
Problem Set 71	23***	5*	0	0	0	0	0
Problem Set 72	25***	1	0	0	0	0	0
Problem Set 73	11**	1	0	0	0	0	0
Problem Set 74	21***	0	0	0	0	0	0
Problem Set 75	30***	13***	1	0	0	0	0
Problem Set 76	30***	10**	0	0	0	0	0
Problem Set 77	54***	33***	3	0	0	0	0
Problem Set 78	4*	0	0	0	0	0	0
Problem Set 79	3	0	0	0	0	0	0
Problem Set 80	17***	0	0	0	0	0	0
Problem Set 81	67***	22***	7**	0	0	0	0
Problem Set 82	3	0	0	0	0	0	0
Problem Set 83	7*	1	0	0	0	0	0
Problem Set 84	20***	1	0	0	0	0	0
Problem Set 85	6	2	0	0	0	0	0
Problem Set 86	36***	6*	1	0	0	0	0
Problem Set 87	6*	1	0	0	0	0	0
Problem Set 88	0	0	0	0	0	0	0
Problem Set 89	24***	1	0	0	0	0	0
Problem Set 90	1	0	0	0	0	0	0
Problem Set 91	78***	28***	7**	0	0	0	0
Problem Set 92	31***	4	1	0	0	0	0
Problem Set 93	3	0	0	0	0	0	0
Problem Set 94	4*	0	0	0	0	0	0
Problem Set 95	1	0	0	0	0	0	0
Problem Set 96	87***	32***	2	0	0	0	0
Problem Set 97	26***	6*	0	0	0	0	0
Problem Set 98	1	0	0	0	0	0	0
Problem Set 99	48***	16***	0	0	0	0	0
Problem Set 100	34***	0	0	0	0	0	0
Problem Set 101	29***	4*	0	0	0	0	0
Problem Set 102	37***	2	0	0	0	0	0
Problem Set 103	44***	16***	0	0	0	0	0
Problem Set 104	16***	2	0	0	0	0	0

TABLE A.7: *Evaluation of Intermediary's Utility-maximizing Decision: Mean Difference in Ranks*

Observations	5,10	10,20	20,50	50,100	100,200	200,500	500,1000
Problem Set 105	0	0	0	0	0	0	0
Problem Set 106	36***	10**	0	0	0	0	0
Problem Set 107	1	0	0	0	0	0	0
Problem Set 108	99***	46***	23***	0	0	0	0
Problem Set 109	103***	45*	116***	82***	21***	8**	0
Problem Set 110	43***	27***	1	0	0	0	0
Problem Set 111	5*	0	0	0	0	0	0
Problem Set 112	45***	16***	1	0	0	0	0
Problem Set 113	37***	7**	0	0	0	0	0
Problem Set 114	0	0	0	0	0	0	0
Problem Set 115	52***	7*	1	0	0	0	0
Problem Set 116	0	0	0	0	0	0	0
Problem Set 117	94***	74***	20***	2	0	0	0
Problem Set 118	19***	5*	0	0	0	0	0
Problem Set 119	77***	67***	9**	0	0	0	0
Problem Set 120	91***	129***	78***	53***	18***	1	0
Problem Set 121	14***	1	0	0	0	0	0
Problem Set 122	2	0	0	0	0	0	0
Problem Set 123	1	0	0	0	0	0	0
Problem Set 124	1	0	0	0	0	0	0
Problem Set 125	15***	2	0	0	0	0	0
Problem Set 126	13***	1	0	0	0	0	0
Problem Set 127	22***	8**	0	0	0	0	0
Problem Set 128	2	0	0	0	0	0	0
Problem Set 129	57***	15***	2	0	0	0	0
Problem Set 130	57***	9*	4*	0	0	0	0
Problem Set 131	1	0	0	0	0	0	0
Problem Set 132	4	1	0	0	0	0	0
Problem Set 133	95***	89***	86***	17***	5*	0	0
Problem Set 134	106***	62***	41***	7**	0	0	0
Problem Set 135	21***	0	0	0	0	0	0
Problem Set 136	11***	0	0	0	0	0	0
Problem Set 137	13***	0	0	0	0	0	0
Problem Set 138	15***	2	0	0	0	0	0
Problem Set 139	5	1	0	0	0	0	0
Problem Set 140	19***	1	0	0	0	0	0
Problem Set 141	97***	37***	3	0	0	0	0
Problem Set 142	0	0	0	0	0	0	0
Problem Set 143	4*	0	0	0	0	0	0
Problem Set 144	16***	0	0	0	0	0	0
Problem Set 145	18***	1	0	0	0	0	0
Problem Set 146	9**	0	0	0	0	0	0
Problem Set 147	107***	45***	27***	0	0	0	0
Problem Set 148	18***	0	0	0	0	0	0
Problem Set 149	89***	47***	10**	0	0	0	0
Problem Set 150	54***	11***	0	0	0	0	0

TABLE A.8: *Evaluation of Intermediary's Utility-maximizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 1	0.08834	20
Problem Set 2	0.08902	0
Problem Set 3	0.09009	10
Problem Set 4	0.09000	10
Problem Set 5	0.08986	0
Problem Set 6	0.06808	10
Problem Set 7	0.06146	20
Problem Set 8	0.03250	0
Problem Set 9	0.06808	10
Problem Set 10	0.09001	10
Problem Set 11	0.04153	100
Problem Set 12	0.08831	10
Problem Set 13	0.08803	20
Problem Set 14	0.07928	20
Problem Set 15	0.09015	10
Problem Set 16	0.08969	0
Problem Set 17	0.09013	10
Problem Set 18	0.09016	0
Problem Set 19	0.09011	0
Problem Set 20	0.08954	10
Problem Set 21	0.08967	0
Problem Set 22	0.07954	10
Problem Set 23	0.04835	10
Problem Set 24	0.08994	0
Problem Set 25	0.08928	0
Problem Set 26	0.08772	20
Problem Set 27	0.08691	10
Problem Set 28	0.08280	10
Problem Set 29	0.08262	20
Problem Set 30	0.04423	0
Problem Set 31	0.08902	10
Problem Set 32	0.09008	20
Problem Set 33	0.08482	0
Problem Set 34	0.08758	0
Problem Set 35	0.08876	10
Problem Set 36	0.08394	50
Problem Set 37	0.08627	10
Problem Set 38	0.09012	10
Problem Set 39	0.08804	0
Problem Set 40	0.08570	10
Problem Set 41	0.08960	20
Problem Set 42	0.08771	0
Problem Set 43	0.08617	10
Problem Set 44	0.05736	0
Problem Set 45	0.08870	20
Problem Set 46	0.08468	20
Problem Set 47	0.09016	20
Problem Set 48	0.05429	20
Problem Set 49	0.08291	20

TABLE A.8: *Evaluation of Intermediary's Utility-maximizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 50	0.03176	0
Problem Set 51	0.08946	10
Problem Set 52	0.02418	200
Problem Set 53	0.08744	0
Problem Set 54	0.08950	20
Problem Set 55	0.08403	10
Problem Set 56	0.07165	0
Problem Set 57	0.08953	20
Problem Set 58	0.06582	10
Problem Set 59	0.08685	0
Problem Set 60	0.08997	10
Problem Set 61	0.08653	0
Problem Set 62	0.08972	0
Problem Set 63	0.03620	20
Problem Set 64	0.08708	0
Problem Set 65	0.08399	0
Problem Set 66	0.08982	10
Problem Set 67	0.06857	20
Problem Set 68	0.08991	10
Problem Set 69	0.06937	50
Problem Set 70	0.08499	10
Problem Set 71	0.08589	20
Problem Set 72	0.07549	10
Problem Set 73	0.08990	10
Problem Set 74	0.08791	10
Problem Set 75	0.08941	20
Problem Set 76	0.08813	20
Problem Set 77	0.08656	20
Problem Set 78	0.00241	10
Problem Set 79	0.08429	0
Problem Set 80	0.09016	10
Problem Set 81	0.08950	50
Problem Set 82	0.08835	0
Problem Set 83	0.08969	10
Problem Set 84	0.08770	10
Problem Set 85	0.08384	0
Problem Set 86	0.08251	20
Problem Set 87	0.09008	10
Problem Set 88	0.08159	0
Problem Set 89	0.09000	10
Problem Set 90	0.08618	0
Problem Set 91	0.08980	50
Problem Set 92	0.08667	10
Problem Set 93	0.05692	0
Problem Set 94	0.08695	10
Problem Set 95	0.08695	0
Problem Set 96	0.05504	20
Problem Set 97	0.08808	20
Problem Set 98	0.08758	0

TABLE A.8: *Evaluation of Intermediary's Utility-maximizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 99	0.08820	20
Problem Set 100	0.06492	10
Problem Set 101	0.07461	20
Problem Set 102	0.07461	10
Problem Set 103	0.08961	20
Problem Set 104	0.08883	10
Problem Set 105	0.08957	0
Problem Set 106	0.05822	10
Problem Set 107	0.08879	0
Problem Set 108	0.06210	50
Problem Set 109	0.01196	500
Problem Set 110	0.08854	20
Problem Set 111	0.08985	10
Problem Set 112	0.08823	20
Problem Set 113	0.08897	20
Problem Set 114	0.08595	0
Problem Set 115	0.07828	20
Problem Set 116	0.08506	0
Problem Set 117	0.08131	50
Problem Set 118	0.06597	20
Problem Set 119	0.02075	50
Problem Set 120	0.04012	200
Problem Set 121	0.08715	10
Problem Set 122	0.08964	0
Problem Set 123	0.05596	0
Problem Set 124	0.05596	0
Problem Set 125	0.08971	10
Problem Set 126	0.08189	10
Problem Set 127	0.09008	20
Problem Set 128	0.08819	0
Problem Set 129	0.02994	20
Problem Set 130	0.02994	50
Problem Set 131	0.08890	0
Problem Set 132	0.08943	0
Problem Set 133	0.05158	200
Problem Set 134	0.06748	100
Problem Set 135	0.08937	10
Problem Set 136	0.08669	10
Problem Set 137	0.09015	10
Problem Set 138	0.06627	10
Problem Set 139	0.08669	0
Problem Set 140	0.07046	10
Problem Set 141	0.04484	20
Problem Set 142	0.08417	0
Problem Set 143	0.08992	10
Problem Set 144	0.09014	10
Problem Set 145	0.08507	10
Problem Set 146	0.07759	10
Problem Set 147	0.07137	50

TABLE A.8: *Evaluation of Intermediary's Utility-maximizing Decision: Maximum Variance in Portfolio and Required Amount of Observations*

	Max Variance in Portfolio	Required Amount of Monitoring Observations
Problem Set 148	0.08677	10
Problem Set 149	0.06027	50
Problem Set 150	0.06277	20

Appendix B

List of Symbols and Abbreviations

B.1 List of Symbols

A	Set of all offered SLAs
$\|A\|$	Number of SLAs in A
α_i	Offered SLA
$ARA(w)$	Coefficient of absolute risk aversion
β_i	Supplied SLA
CE_γ	Provider's certainty equivalent for SLA portfolio γ
$c(s_j)$	Provider specific costs of service execution for service s_j
$\chi_{\delta,t}$	Actual costs from procuring services in portfolio δ in period t
d	adaptation factor in Beta Distribution Function
δ	Supplied SLA portfolio
$\|\delta\|$	Number of SLAs in δ
δ^*	Risk-minimal supplied SLA portfolio
e	adaptation factor in Beta Distribution Function
$E(\chi_\delta)$	Expected costs of portfolio δ
$E(\lambda_{\delta,t})$	Average degree of SLA violation of SLA portfolio δ
$E(\mu_{\gamma,t})$	Average penalty for SLA portfolio γ
$E(\mu_{\gamma^*,t})$	Average penalty for risk-minimal portfolio γ^*
$E(\pi_{\gamma,t})$	Expected profit of SLA portfolio γ
$E(\pi_{\gamma^*,t})$	Expected profit of risk-minimal offered portfolio γ^*
$E(\pi_{\{\gamma^*,\delta\},t})$	Intermediary's expected profit from combination of γ^* and δ
$E(u(\pi_{\gamma,t}))$	Expected utility of profit of SLA portfolio γ
$E(u(\pi_{\{\gamma^*,\delta\},t}))$	Expected utility of profit of SLA portfolio γ^* together with SLA portfolio δ
$f_{\alpha_i}(s_j)$	Price for service execution of service s_j agreed on in SLA α_i
$f_{\beta_i}(s_j)$	Price for service s_j in SLA β_i
$g_\gamma^{\alpha_i}(\cdot)$	Probability Density Function of $\lambda_\gamma^{\alpha_i}$
Γ	Set of offered SLA portfolios
Γ^R	Set of offered SLA portfolios that meets constraints
γ	Offered SLA portfolio
γ^*	Optimal offered SLA portfolio
ι	Intermediary
K	Number of KPIs in L_{α_i}
$l_{\alpha_i}^m$	one KPI in SLA α_i
L_{α_i}	Set of KPIs in SLA α_i that can be mapped to provider's resources
L_γ	Resource requirements for SLA portfolio γ
L_ι	Intermediary's resource constraints
L_p	Provider's resource constraints

Λ	Set of monitoring data represented by degree of violation of SLAs over periods
$\lambda_{\delta,t}^{\beta_i}$	Degree of violation of SLA β_i in period t in the context of portfolio δ
$\overline{\lambda}_{\delta,t}$	Actual average degree of SLA violation of portfolio δ in period t
$\overline{\lambda}_{\delta}^{\beta_i}$	Mean degree of violation of SLA β_i in the context of portfolio δ
$\lambda_{\gamma,t}^{\alpha_i}$	Degree of violation of SLA α_i in period t in the context of portfolio γ
$\overline{\lambda}_{\gamma}^{\alpha_i}$	Mean degree of violation of SLA α_i in portfolio γ
M	Number of KPIs of SLA α_i that can be mapped to provider's resources
μ_{α_i}	Penalty agreed in SLA α_i
$\mu_{\gamma,t}$	Actual sum of penalties of SLA portfolio γ in period t
n	Number of Repetitions
$O(\cdot)$	Time complexity in Big-O-Notation
ω	Feasible combination of offered portfolio γ^* and supplied portfolio δ
Ω	Set of feasible combinations ω
$\mathcal{P}(A)$	Power-set of A
P	Set of Service Providers
p	Service Provider
$\pi_{\gamma,t}$	Actual profit of SLA portfolio γ in period t
$\pi_{\{\gamma^*,\delta\},t}$	Actual profit from offering γ^* and supplied SLAs δ in period t
Π_ι	Intermediary's expected profit constraint
Π_p	Provider's expected profit constraint
q_t	Probability of occurence, relative frequency
$R(w)$	Coefficient of relative risk aversion
R_ι	Violation Threshold of ι
ρ_t	Probability of occurence, relative frequency
s_j	Service
$S_E(\lambda_{\delta,t})$	Risk of SLA violations for supplied SLAs
$S_E(\mu_{\gamma,t})$	Semi-variance of due penalties, Risk
$\hat{S}_E(\mu_{\gamma,t})$	Semi-variance of due penalties calculated from PDFs
T	Monitoring Periods
t	Period, Time
$\lvert T_\delta \rvert$	number of monitoring periods in which portfolio δ was active
$\lvert T_\gamma \rvert$	number of monitoring periods in which portfolio γ was active
θ_k	Customer
$u(\pi_{\gamma,t})$	Utility of actual profit of SLA portfolio γ in period t
$u(\pi_{\{\gamma^*,\delta\},t})$	Utility of actual profit of customer SLA portfolio γ^* and supplied portfolio δ in period t
$Var(z)$	Variance of z
w	Certain income
x	Random Variable

$X(\lambda_{\delta,t}^{\beta_i})$	Indicator variable for activity of SLA β_i in portfolio δ in period t
$X(\lambda_{\delta,t}^{B})$	Indicator variable for activity of SLA portfolio δ
$X(\lambda_{\delta^c,t}^{B})$	Indicator variable for activity of all SLA portfolios other than δ
$X(\lambda_{\gamma,t}^{\alpha_i})$	Indicator variable for activity of SLA α_i in portfolio γ in period t
$X(\lambda_{\gamma,t}^{A})$	Indicator variable for activity of SLA portfolio γ
$X(\lambda_{\gamma^c,t}^{A})$	Indicator variable for activity of all SLA portfolios other than γ
z	Stochastic variable with mean 0 and variance Var(z)

B.2 List of Abbreviations

AHP	Analytic Hierarchy Process	75
AN	Agreement Network	9
CDF	Cumulative Distribution Function	49
CE	Certainty Equivalent	91
DA	Dependency Analyzer	136
EC2	Elastic Compute Cloud	26
ERP	Enterprise Resource Planning	131
FSD	First Order Stochastic Dominance	49
GDP	Gross Domestic Product	3
ICT	Information and Communication Technology	24
IHIP	Inseperability, Heterogeneity, Intangibility, Perishability	20
IS	Information Systems	17
IT	Information Technology	17
ITIL	IT Infrastructure Library	23
KPI	Key Performance Indicator	31
LPM	Lower Partial Moment	53
MAD	Mean Absolute Deviation	52
OECD	Organisation for Economic Cooperation and Development	3
PDF	Probability Density Function	144
QoS	Quality of Service	29
SaaS	Software as a Service	4
SBN	Smart Business Network	34
SLA	Service Level Agreement	5
SLO	Service Level Objective	31
SSD	Second Order Stochastic Dominance	49
SOC	Service-oriented Computing	23
SVN	Service Value Network	34
URI	Uniform Resource Identifier	26
VaR	Value at Risk	50

Bibliography

Agarwal, S., S. Lamparter, and R. Studer (2008). Making Web services tradable - A policy-based approach for specifying preferences on Web service properties. *Web Semantics: Science, Services and Agents on the World Wide Web, Special Issue on Policies 7*(1), 11–20.

Agarwal, S., S. Rudolph, and A. Abecker (2008). Semantic description of distributed business processes. In *AAAI Spring Symposium-AI Meets Business Rules and Process Management, Stanford, USA*.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society 21*(4), 503–546.

Alonso, G., F. Casati, H. Kuno, and V. Machiraju (2004). *Web services : concepts, architectures and applications*. Data-centric systems and applications. Berlin: Springer.

Anandasivam, A. (2010). *Consumer preferences and bid-price control for cloud services*. Ph. D. thesis, Universität Karlsruhe (TH), Fakultät für Wirtschaftswissenschaften.

Anderson, C. (2006). *The Long Tail: How Endless Choice is Creating Unlimited Demand*. London, England: Random House Business Books.

Andrieux, A., K. Czajkowski, and K. Keahey (2007). Web Services Agreement Specification (WS-Agreement).

Armbrust, M., A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patteron, A. Rabkin, I. Stoica, and et al. (2009). Above the clouds : A berkeley view of cloud computing. Technical report, EECS Department, University of California, Berkeley.

Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Amsterdam: North-Holland Publishing.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*, 203–228.

Asnar, Y., P. Giorgini, and A. Mulyanto (2007). Risk in secure and dependable system: a survey. In *Proceedings of the 1st International Conference and Workshop on Risk Technology and Management*. Citeseer.

Asnar, Y., H. W. Lim, F. Massacci, and C. Worledge (2010). Realizing trustworthy business services by a new grc approach. Technical report, University of Trento, Ingegneria e Scienza dell'Informazione.

Badger, M. (2008). *Zenoss Core Network and System Monitoring*. Packt Publishing.

Baida, Z., J. Gordijn, and B. Omelayenko (2004). A shared service terminology for online service provisioning. In *Proceedings of the 6th international conference on Electronic commerce (ICEC '04)*, New York, New York, USA. ACM Press.

Baresi, L., S. Guinea, and P. Plebani (2006). Ws-policy for service monitoring. In C. Bussler and M.-C. Shan (Eds.), *Technologies for E-Services*, Volume 3811 of *Lecture Notes in Computer Science*, pp. 72–83. Springer Berlin / Heidelberg. 10.1007/11607380_7.

Barros, A. and M. Dumas (2006). The rise of web service ecosystems. *IT Professional 8*(5), 31 –37.

Barth, W. (2008). *Nagios: System and network monitoring*. No Starch Press.

Baun, C., M. Kunze, J. Nimis, and S. Tai (2009). *Cloud computing: Web-basierte dynamische IT-Services*. Heidelberg: Springer.

Bawa, V. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics 2*(1), 95–121.

Becker, M., N. Borrisov, V. Deora, O. Rana, and D. Neumann (2008). Using k-pricing for penalty calculation in grid markets. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*. IEEE.

Bengtsson, M. and S. Kock (2000). "coopetition" in business networks - to cooperate and compete simultaneously. *Industrial Marketing Management 29*(5), 411–426.

Berardi, D., D. Calvanese, G. D. Giacomo, M. Lenzerini, and M. Mecella (2005). Automatic service composition based on behavioral descriptions. *International Journal of Cooperative Information Systems 14*(4), 333–376.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*(1), 122–142.

Berger, T. G. (2005). *Konzeption und Management von Service-Level- Agreements für IT-Dienstleistungen*. Ph. D. thesis, Technische Universität Darmstadt, Fachbereich Rechts- und Wirtschaftswissenschaften.

Berry, L. and A. Pasuraman (2004). *Marketing services: Competing through quality*. Free Press.

Black, D. (1948). On the rationale of group decision-making. *The Journal of Political Economy 56*(1), 23–34.

Blake, M. and D. Cummings (2007). Workflow composition of service level agreements. In *Proceedings of the IEEE International Conference on Services Computing*. IEEE Computer Society.

Blau, B. (2009). *Coordination in Service Value Networks - A Mechanism Design Approach*. Ph. D. thesis, Universität Karlsruhe (TH), Fakultät für Wirtschaftswissenschaften.

Blau, B., J. Kraemer, T. Conte, and C. V. Dinther (2009). Service Value Networks. In *Proceedings of the 11th IEEE Conference on Commerce and Enterprise Computing*, Vienna.

Blau, B., D. Neumann, C. Weinhardt, and S. Lamparter (2008). Planning and pricing of service mashups. In *Proceedings of the 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services*, pp. 19–26. Ieee.

Bodenstaff, L., A. Wombacher, M. Reichert, and M. C. Jaeger (2008). Monitoring Dependencies for SLAs: The MoDe4SLA Approach. In *IEEE International Conference on Services Computing*, Honolulu, pp. 21–29.

Bon, J. v. (2005). *Foundations of IT Service Management*. Van Haren Publishing.

Bonini, C. P. (1975). Risk evaluation of investment projects. *Omega 3*(6), 137–139.

Brandic, I., S. Venugopal, M. Mattess, and R. Buyya (2008). Towards a meta-negotiation architecture for sla-aware grid services. In *Workshop on Service-Oriented Engineering and Optimizations*, pp. 17–20.

Bridge, P. D. and S. S. Sawilowsky (1999). Increasing physicians' awareness of the impact of statistics on research outcomes: Comparative power of the t-test and wilcoxon rank-sum test in small samples applied research. *Journal of Clinical Epidemiology 52*(3), 229 – 235.

Buhl, H. and G. Fridgen (2011, June). It-enabled risk/return management: service-oriented infrastructures vs. dedicated systems. In V. Tuunainen, J. Nandhakumar, M. Rossi, and W. Soliman (Eds.), *Proceedings of the 19th European Conference on Information Systems - ECIS 2011*, Helsinki, Finland. ISBN 978-952-60-3574-1.

Business Finance Magazine (2010). `http://businessfinancemag.com/article/saas-core-financials-anyone-0804`.

Busquets, J., J. Rodon, and J. Wareham (2009). Adaptability in smart business networks: An exploratory case in the insurance industry. *Decision Support Systems 47*(4), 287–296.

Carroll, L. (1865). *Alice's Adventures in Wonderland*. Macmillan and Co.

Caton, S. (2009). *On-Demand Distributed Image Processing Over An Adaptive Campus-Grid*. Ph. D. thesis, Cardiff University.

Chandra, S. (2003). Regional economy size and the growth–instability frontier: Evidence from europe. *Journal of Regional Science 43*(1), 95–122.

Chandra, S. and W. Shadel (2007). Crossing disciplinary boundaries: Applying financial portfolio theory to model the organization of the self-concept. *Journal of Research in Personality 41*(2), 346–373.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests*. *Quarterly Journal of Economics 117*(3), 817–869.

Chase, R. (1981). The customer contact approach to services: theoretical bases and practical extensions. *Operations Research 29*(4), 698–706.

Choudhary, V. (2007, January). Software as a service: Implications for investment in software development. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*. (CD-ROM).

Choy, R. and A. Edelman (2005). Parallel matlab: Doing it right. *Proceedings of the IEEE 93*(2), 331–341.

Christensen, E., F. Curbera, G. Meredith, and S. Weerawarana (2001). Web services description language (wsdl) 1.1.

Coase, R. (1937). The nature of the firm. *Economica 4*(16), 386–405.

Conte, T. (2010). *Value Creation through Co-Opetition in Service Networks*. Ph. D. thesis, Karlsruher Institut für Technologie, Fakultät für Wirtschaftswissenschaften.

Conte, T., B. Blau, G. Satzger, and C. van Dinther (2009, August). Enabling Service Networks Through Contribution-Based Value Distribution. In *Proceedings of the 15th Americas Conference on Information Systems*, Paper 764, San Francisco, USA.

Cooper, R., S. Edgett, and E. Kleinschmidt (1999). New product portfolio management: practices and performance. *Journal of Product Innovation Management 16*(4), 333–351.

Crockford, D. (2006). Json: The fat-free alternative to xml. In *Proceedings of XML*.

Czajkowski, K., I. Foster, C. Kesselman, V. Sander, and S. Tuecke (2002). Snap: A protocol for negotiating service level agreements and coordinating resource management in distributed systems. In D. Feitelson, L. Rudolph, and U. Schwiegelshohn (Eds.), *Job Scheduling Strategies for Parallel Processing*, Volume 2537 of *Lecture Notes in Computer Science*, pp. 153–183. Springer Berlin / Heidelberg. 10.1007/3-540-36180-4_9.

DeSanctis, G. and R. B. Gallupe (1987). A foundation for the study of group decision support systems. *Management Science 33*(5), pp. 589–609.

Donnelly Jr, J. (1976). Marketing Intermediaries in channels of distribution for services. *The Journal of Marketing 40*(1), 55–57.

Dubey, A. and D. Wagle (2007). Delivering software as a service. In *The McKinsey Quarterly, Web exclusive*. Online.

Edgett, S. and S. Parkinson (1993). Marketing for Service Industries - a Review. *The Service Industries Journal 13*(3), 19–39.

Edvardsson, B., A. Gustafsson, and I. Roos (2005). Service portraits in service research: a critical review. *International Journal of Service Industry Management 16*(1), 107–121.

Eeckhoudt, L., H. Schlesinger, and C. Gollier (2005). *Economic and financial decisions under risk*. Princeton: Princeton Univ. Press.

Einhorn, D. and A. Brown (2008). Private profits and socialized risk. In *Global Association of Risk Professionals Risk Review*.

Elton, E., M. Gruber, S. Brown, and W. Goetzmann (2003). *Modern Portfolio Theory and Investment Analysis* (6 ed.). New York, NY, USA: John Wiley & Sons.

Engelhardt, W., M. Kleinaltenkamp, and M. Reckenfelderbäumer (1993). Leistungsbündel als Absatzobjekte. *Zeitschrift für betriebswirtschaftliche Forschung 45*(5), 395–426.

Fay, M. P. and M. a. Proschan (2010). Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys 4*, 1–39.

Felli, J. and G. Hazen (1998). Sensitivity analysis and the expected value of perfect information. *Medical Decision Making 18*(1), 95.

Fielding, R. (2000). *Architectural styles and the design of network-based software architectures*. Ph. D. thesis, University of California, Irvine.

Filipova, L. (2009). *Anreize und Privatheit bei modernen Monitoring Technologien für das Beispiel der Kfz-Versicherungsverträge*. Ph. D. thesis, Universität Augsburg.

Filipova-Neumann, L., W. Michalk, B. Blau, and C. Weinhardt (2010, June). Incentivation of Service Providers for Efficient Service Composition Agreements. In *Proceedings of the International Conference on e-Commerce, e-Business and e-Service 2010 (ICCBS 2010)*, Paris, France, pp. 1569–1583.

Fishburn, P. (1964). *Decision and value theory*, Volume 10. Wiley New York.

Flipo, J. (1988). On the intangibility of services. *The Service Industries Journal 8*(3), 286–293.

Friedman, M. and L. Savage (1948). The utility analysis of choices involving risk. *The Journal of Political Economy 56*(4), 279–304.

Gadrey, J. (2000). The characterization of goods and services: an alternative approach. *Review of Income and Wealth 46*, 369–387.

Ghezzi, C. and S. Guinea (2007). *Run-time monitoring in service-oriented architectures*, Chapter 9, pp. 237–264. Springer.

Green, P. E. and V. R. Rao (1971, Aug). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research 8*(3), 355–363.

Gujarati, D. N. and D. C. Porter (2006). *Essentials of Econometrics* (3rd ed.). Mac-Graw Hill.

Hadar, J. and W. Russell (1969). Rules for ordering uncertain prospects. *The American Economic Review 59*(1), 25–34.

Hanoch, G. and H. Levy (1969). The efficiency analysis of choices involving risk. *The Review of Economic Studies 36*(3), 335–346.

Hanoch, G. and H. Levy (1970). Efficient portfolio selection with quadratic and cubic utility. *The Journal of Business 43*(2), 181–189.

Happe, J., D. Westermann, K. Sachs, and L. Kapova (2010). Statistical inference of software performance models for parametric performance completions. In G. Heineman, J. Kofron, and F. Plasil (Eds.), *Research into Practice â Reality and Gaps*, Volume 6093 of *Lecture Notes in Computer Science*, pp. 20–35. Springer Berlin / Heidelberg.

Hart, O. D. and D. M. Jaffee (1974). On the application of portfolio financial theory to depository. *The Review of Economic Studies 41*(1), 129–147.

Heck, E. and P. Vervest (2007). Smart business networks: how the network wins. *Communications of the ACM 50*(6), 28–37.

Hiles, A. (2002). *The Complete Guide to IT Service Level Agreements: Aligning IT Services to Business Needs*. Rothstein Associates Inc.

Hill, T. P. (1977). On goods and services. *Review of Income and Wealth 23*, 315–338.

Holm, D., K. Eriksson, and J. Johanson (1996). Business networks and cooperation in international business relationships. *Journal of International Business Studies 27*(5), 1033–1053.

Holton, G. (2003). *Value-at-risk: theory and practice*. Academic Press.

Hubbard, D. (2007). *How to measure anything*. John Wiley & Sons.

Hughes, D., G. Coulson, and J. Walkerdine (2005). Free riding on gnutella revisited: the bell tolls? *IEEE Distributed Systems Online 6*(6), 1.

Johnson, R. A. (1988). *Applied multivariate statistical analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Jorion, P. (2007). *Value at risk : the new benchmark for managing financial risk* (3 ed.). New-York: McGraw-Hill.

Junghans, M. and S. Agarwal (2010). Web Service Discovery Based on Unified View on Functional and Non-Functional Properties. In *Proceedings of the 4th IEEE International Conference on Semantic Computing (ICSC 2010)*. IEEE Computer Society.

Junghans, M., S. Agarwal, and R. Studer (2010). Towards practical semantic web service discovery. In L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, and T. Tudorache (Eds.), *The Semantic Web: Research and Applications*, Volume 6089 of *Lecture Notes in Computer Science*, pp. 15–29. Springer Berlin / Heidelberg. 10.1007/978-3-642-13489-0_2.

Kagel, J. H. H. and A. Roth (1995). *The handbook of experimental economics*. Princeton University Press.

Karten, N. (2001). How to establish service level agreements. *Naomi Karten Associates*.

Kauffman, R. and R. Sougstad (2010). Evaluating it services price benchmarks under uncertainty. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*. IEEE Computer Society.

Kaufman, L., P. Rousseeuw, and E. Corporation (1990). *Finding groups in data: an introduction to cluster analysis*, Volume 39. Wiley Online Library.

Keeney, R. and H. Raiffa (1993). *Decisions with multiple objectives: Preferences and value tradeoffs* (2 ed.). Cambridge University Press.

Keller, A. and H. Ludwig (2003). The wsla framework: Specifying and monitoring service level agreements for web services. *Journal of Network and Systems Management 11*, 57–81. 10.1023/A:1022445108617.

Knapper, R., B. Blau, T. Conte, A. Sailer, A. Kochut, and A. Mohindra (2011). Efficient Contracting in Cloud Service Markets with Asymmetric Information - A Screening Approach. In *Workshop Clouds for Enterprises (C4E) in proceedings of the 13th IEEE Conference on Commerce and Enterprise Computing*, Luxembourg.

Knapper, R., T. Conte, B. Blau, and H. Richter (2010). Prosa - process-oriented service level agreements for providing a single point of contract. In *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI)*, Goettingen, Germany, pp. 213–225.

Kompella, R. R., J. Yates, A. Greenberg, and A. C. Snoeren (2010, Oct). Fault localization via risk modeling. *IEEE Transactions on Dependable and Secure Computing 7*(4), 396–409.

Konno, H. and H. Yamazaki (1991). Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management science 37*(5), 519–531.

Kotler, P. and R. Connor Jr (1977). Marketing professional services. *The Journal of Marketing 41*(1), 71–76.

Kounev, S. (2006). Performance modeling and evaluation of distributed component-based systems using queueing petri nets. *IEEE Transactions on Software Engineering 32*(7), 486–502.

Krämer, J.; Wiewiorra, L. (2010, 7-9 June 2010). Consumer Perception of Mobile Telephony Tariffs with Cost Caps. In *Proceedings of 9th Conference on Telecommunications, Internet and Media Techno Economics (CTTE)*, Ghent, Belgium, pp. 1–8. IEEE Xplore.

Kroll, Y., H. Levy, and H. M. Markowitz (1984). Mean-variance versus direct utility maximization. *Journal of Finance 39*(1), 47–61.

Küster, U., B. König-Ries, M. Stern, and M. Klein (2007). Diane: an integrated approach to automated service discovery, matchmaking and composition. In *Proceedings of the 16th international conference on World Wide Web*, pp. 1033–1042. ACM.

Kwok, T. and A. Mohindra (2008). Resource calculations with constraints, and placement of tenants and instances for multi-tenant saas applications. In A. Bouguettaya, I. Krueger, and T. Margaria (Eds.), *Service-Oriented Computing - ICSOC 2008*, Volume 5364 of *Lecture Notes in Computer Science*, pp. 633–648. Springer Berlin / Heidelberg.

Lambrecht, A. and B. Skiera (2006). Paying too Much and being Happy About It: Existence, Causes, and Consequences of Tariff-choice Biases. *Journal of Marketing Research 43*(2), 212–223.

Lamparter, S. (2007). *Policy-based Contracting in Semantic Web Service Markets*. Phdthesis, Universität Karlsruhe (TH), Fakultät für Wirtschaftswissenschaften.

Lang, J. C., T. Widjaja, P. Buxmann, W. Domschke, and T. Hess (2008). Optimizing the supplier selection and service portfolio of a soa service integrator. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, pp. 89–89. Ieee.

Laux, H. (2005). *Entscheidungstheorie*. Springer Verlag.

Levy, H. (1992). Stochastic Dominance and expected utility: survey and analysis. *Management Science 38*, 555–593.

Lewis, L. and P. Ray (1999). Service level management definition, architecture, and research challenges. In *Global Telecommunications Conference, 1999. GLOBECOM 99*, Volume 3, pp. 1974–1978. IEEE.

Libby, R. and P. Fishburn (1977). Behavioral models of risk taking in business decisions: A survey and evaluation. *Journal of Accounting Research 15*(2), 272–292.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics 47*(1), pp. 13–37.

Lovelock, C. and E. Gummesson (2004). Whither Services Marketing?: In Search of a New Paradigm and Fresh Perspectives. *Journal of Service Research 7*(1), 20.

Lovelock, C., S. Vandermerve, and B. Lewis (1999). *Services Marketing: A European Perspective*. Prentice Hall.

Maglio, P. and J. Spohrer (2008). Fundamentals of service science. *Journal of the Academy of Marketing Science 36*(1), 18–20.

Markowitz, H. M. (1952). Portfolio Selection. *Journal of Finance 7*, 77–91.

Markowitz, H. M. (1959). *Portfolio selection: efficient diversification of investments* (1 ed.). Yale University Press.

Markowitz, H. M. (1991). *Portfolio Selection: Efficient Diversification of Investments*. Wiley.

Meffert, H. and M. Bruhn (2008). *Dienstleistungsmarketing* (6 ed.). Wiesbaden: Gabler GWV Fachverlage.

Merton, R. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis 7*(4), 1851–1872.

Mertz, S., T. Eid, C. Eschinger, H. Swinehart, C. Pang, and B. Pring (2008). Market trends: Software as a service, worldwide, 2007-2012. Technical report, Gartner report.

Mertz, S., C. Eschinger, T. Eid, and B. Pring (2007). Dataquest insight: Saas de- mand set to outpace enterprise application software market growth. Technical report, Gartner report.

Michalk, W. (2011). Risk-Aware SLA Establishment in Agreement Networks - an Evaluation. In V. Tuunainen, J. Nandhakumar, M. Rossi, and W. Soliman (Eds.), *Proceedings of the 19th European Conference on Information Systems - ECIS 2011*, Helsinki, Finland. ISBN 978-952-60-3574-1.

Michalk, W. and B. Blau (2010, March). Risk in Agreement Networks Decision Support for Service-Intermediaries. *Journal of Information Systems and e-Business Management (ISeBM) 9*(2), 247–266.

Michalk, W., B. Blau, J. Stosser, and C. Weinhardt (2010). Risk-based decision support in service value networks. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, HICSS '10, Washington, DC, USA, pp. 1–9. IEEE Computer Society.

Michalk, W. and S. Caton (2010). Service level management in dynamic value networks. In *INFORMATIK 2010: Service Science - Neue Perspektiven fuer die Informatik Band 1*, Bonn, Germany, pp. 126 – 131. Gesellschaft fuer Informatik.

Michalk, W., L. Filipova-Neumann, B. Blau, and C. Weinhardt (2011). Reducing Risk or Increasing Profit? Provider Decisions in Agreement Networks. *Service Science 3*(3), 206–222.

Michalk, W. and C. Haas (2011). Incentives in Service Level Agreement Establishment - The Case of Economic and Social Aspects. In *Proceedings of the 19th IEEE International Requirements Engineering Conference*, Trento, Italy.

Mika, P., D. Oberle, A. Gangemi, and M. Sabou (2004). Foundations for service ontologies: Aligning owl-s to dolce. In *Proceedings of the 13th international conference on World Wide Web*, pp. 563–572. ACM.

Oberle, D., N. Bhatti, S. Brockmans, M. Niemann, and C. Janiesch (2009). Countering service information challenges in the internet of services. *Business & Information Systems Engineering 1*(5), 370–390.

OECD (2010). Oecd factbook 2010. Technical report, Organisation for Economic Cooperation and Development - OECD.

O'Hara, M. and G. S. . Oldfield (1986). The microeconomics of market making. *The Journal of Financial and Quantitative Analysis 21*(4), 361–376.

O'Sullivan, J., D. Edmond, and A. ter Hofstede (2002). What's in a Service? Towards Accurate Description of Non-Functional. *Distributed and Parallel Databases 12*, 117–133.

Papadimitriou, C. H. and K. Steiglitz (1998). *Combinatorial Optimization: Algorithms and Complexity. Dover Publications, 1998.* Dover Publications.

Papazoglou, M. P. (2008). *Web services : principles and technology* (1. ed.). Pearson education. Harlow: Pearson Education.

Papoulis, Athanasios ; Pillai, S. U. (2002). *Probability, random variables, and stochastic processes* (4. ed., internat. ed. ed.). McGraw-Hill series in electrical and computer engineering. Boston [u.a.]: McGraw-Hill. Includes bibliographical references and index.

Patton, M. (2002). *Qualitative research and evaluation methods*. Sage Publications, Inc.

Pavlou, P. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce 7*(3), 101–134.

Penna, M. and R. Wandresen (2004). *On-line techniques for service quality management*, pp. 292–295. IEEE Comput. Soc.

Petrie, C. and C. Bussler (2008). The myth of open web services. *IEEE Internet Computing 12*(3), 93–95.

Pham-Gia, T. (1989). The generalized beta- and f-distributions in statistical modelling. *Mathematical and Computer Modelling 12*(12), 1613–1625.

Prasad, B. and N. Kalai Selvan (2009). *Supply Chain Management in Services Industry: An Introduction*. ICFAI Books.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica 32*(1), 122–136.

Ramsey, P. H. (1980). Choosing the most powerful pairwise multiple comparison procedure in multivariate analysis of variance. *Journal of Applied Psychology 65*(3), 317 – 326.

Rana, O., M. Warnier, T. Quillinan, and F. Brazier (2008). Monitoring and reputation mechanisms for service level agreements. In J. Altmann, D. Neumann, and T. Fahringer (Eds.), *Grid Economics and Business Models*, Volume 5206 of *Lecture Notes in Computer Science*, pp. 125–139. Springer Berlin / Heidelberg. 10.1007/978-3-540-85485-2_10.

Rao, J. and X. Su (2005). *A Survey of Automated Web Service Composition Methods*, pp. 43–54. Springer.

Rathmell, J. (1966). What is meant by services? *The Journal of Marketing 30*, 32–36.

Regan, W. (1963). The service revolution. *The Journal of Marketing 27*(3), 57–62.

Resnik, M. (1987). *Choices: an introduction to decision theory*. University Of Minnesota Press.

Rothschild, M. and J. Stiglitz (1970). Increasing risk: I. a definition. *Journal of Economic theory 2*(3), 225–243.

Saaty, T. L. (2005). *Analytic Hierarchy Process*. John Wiley & Sons, Ltd.

Savage, L. (1972). *The foundations of statistics*. Dover Pubns.

Sawilowsky, S. S. and R. Blair (1992). A more realistic look at the robustness and type ii error properties of the t test to departures from population normality. *Psychological Bulletin 111*(2), 352 – 360.

Scheithauer, G., K. Voigt, V. Bicer, M. Heinrich, A. Strunk, and M. Winkler (2009). Integrated service engineering workbench: service engineering for digital ecosystems. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, pp. 68. ACM.

Schulz, F. (2010). Towards measuring the degree of fulfillment of service level agreements. In *2010 Third International Conference on Information and Computing*, pp. 273–276. IEEE.

Schulz, F., C. Momm, D. Westermann, B. Blau, W. Michalk, M. Hedwig, D. Rolli, O. K. Afaghi, and A. Schmidt (2010). ValueGrids: Using Grids in Dynamic Service Value Networks. In M. Bubak, M. Turala, and K. Wiatr (Eds.), *Proceedings of the Cracow Grid Workshop 2009 (CGW'09)*, Cracow, Poland, pp. 204–211. ISBN 978-83-61433-01-9.

Shalit, H. and S. Yitzhaki (1984). Mean-gini, portfolio theory, and the pricing of risky assets. *The Journal of Finance 39*(5), 1449–1468.

Shapiro, C. and H. Varian (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business Press.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*(3), pp. 425–442.

Shostack, G. (1977). Breaking free from product marketing. *The Journal of Marketing 41*(2), 73–80.

Speiser, S., B. Blau, S. Lamparter, and S. Tai (2008, Dezember). Formation of service value networks for decentralized service provisioning. In *Proceedings of the 6th International Conference on Service Oriented Computing, 2008*.

Spillner, J. and A. Schill (2009). *Dynamic SLA Template Adjustments Based on Service Property Monitoring*, pp. 183–189. IEEE.

Spohrer, J., S. Vargo, N. Caswell, and P. Maglio (2008). The service system is the basic abstraction of service science. In *Hawaiian International Conference on Systems Sciences (HICSS), Jan*, pp. 7–10.

Steiner, F. (2004). *Formation and Early Growth of Business Webs: Modular Product Systems in Network Markets*. Physica-Verlag Heidelberg.

Steinhausen, Detlef ; Langer, K. (1977). *Clusteranalyse : Einführung in Methoden und Verfahren der automatischen Klassifikation; mit zahlreichen Algorithmen, FORTRAN-Programmen, Anwendungsbeispielen und einer Kurzdarstellung der multivariaten* *statistischen Verfahren; mit 63 Abb. und 4 Tabellen* (1. Aufl. ed.). De Gruyter Lehrbuch. Berlin [u.a.]: de Gruyter.

Sundararajan, A. (2004). Nonlinear Pricing of Information Goods. *Management Science 50*(12), 1660–1673.

Szpiro, G. G. (1988). Insurance, risk aversion and demand for insurance. *Studies in Banking and Finance 6*, 1–125.

Taleb, N. (1997). The world according to nassim taleb. *Derivatives Strategy 2*(1).

Tapscott, D., A. Lowy, and D. Ticoll (2000). *Digital Capital: Harnessing the Power of Business Webs*, Volume 1. Harvard Business School Press.

Toch, E., A. Gal, I. Reinhartz-Berger, and D. Dori (2007). A semantic approach to approximate service retrieval. *ACM Transactions on Internet Technology (TOIT) 8*(1), 2:1–2:29.

Trigg, D. and A. Leach (1967). Exponential Smoothing with an Adaptive Response Rate. *Journal of the Operational Research Society 18*(1), 53–59.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly review of biology 46*(1), 35–57.

Unger, T., F. Leymann, S. Mauchart, and T. Scheibler (2008). Aggregation of service level agreements in the context of business processes. In *Enterprise Distributed Object Computing Conference, 2008. EDOC'08. 12th International IEEE*, pp. 43–52. IEEE.

van Der Aalst, W., A. Ter Hofstede, B. Kiepuszewski, and A. Barros (2003). Workflow patterns. *Distributed and parallel databases 14*(1), 5–51.

Van Heck, E. and P. Vervest (2009). Smart business networks: Concepts and empirical evidence. *Decision Support Systems 47*(4), 275–276.

Vargo, S. and R. Lusch (2004a). The four service marketing myths: remnants of a goods-based, manufacturing model. *Journal of Service Research 6*(4), 324.

Vargo, S. L. and R. F. Lusch (2004b). Evolving to a new dominant logic for marketing. *Journal of Marketing 68*(1), 1–17.

Varian, H. (1997). Versioning Information Goods. University of California, Berkeley.

Von Neumann, J., O. Morgenstern, H. Kuhn, and A. Rubinstein (1947). *Theory of games and economic behavior* (1 ed.). Princeton university press Princeton, NJ.

W3C (2004). Web Services Glossary.

Wagener, M. and R. Riordan (2009). System Latency in Linked Spot and Futures Markets. *Value Creation in E-Business Management 36*(3), 231–245.

Westermann, D., J. Happe, M. Hauck, and C. Heupel (2010). The performance cockpit approach: A framework for systematic performance evaluations. In *Proceedings of the 36th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*.

Westermann, D. and C. Momm (2010). Using software performance curves for dependable and cost-efficient service hosting. In *2nd International Workshop on the Quality of Service-Oriented Software Systems (QUASOSS)*.

Wyckham, R., P. Fitzroy, and G. Mandry (1975). Marketing of services. *European Journal of Marketing 9*(1), 59–67.

Yitzhaki, S. (1982). Stochastiv dominance, mean variance, and gini's mean difference. *The American Economic Review 72*(1), 178–185.

Zeithaml, V., A. Parasuraman, and L. Berry (1985). Problems and strategies in services marketing. *The Journal of Marketing 49*(2), 33–46.

Zeng, L., B. Bentallah, M. Dumas, J. Kalagnanam, and Q. Z. Sheng (2003). Quality driven web services composition. In *Proceedings of the 12th international conference on World Wide Web (WWW '03)*, pp. 411–421. ACM New York, NY, USA.

Zerdick, A., A. Picot, K. Schrape, A. Artope, K. Goldhammer, U. T. Lange, E. Vierkant, L. E. Escobar, and R. Silvertone (2000). *E-economics. Strategies for the Digital Marketplace*. Berlin: Springer.