Sourcing Strategies for Virtualized Services: An Intuitionistic Fuzzy Risk-based Approach

Roland Schütze, roland.schuetze@unifr.ch, University of Fribourg, Switzerland

Abstract

An example for sourcing of virtualized services is “cloud bursting” as a deployment model in which applications run in a private cloud or data center and ‘burst’ into a non-private cloud when more computing capacity is needed [1],[2]. The availability of free burst capacity is not guaranteed, so ‘burst’ has a cheaper cost but comes with a higher risk that the Service Level Agreement (SLA) e.g. in terms of application response time are not met for peak-traffic events. To balance risk versus costs the impact of service levels (e.g. committed vs. burst capacity) of technical services on customers’ business processes need to be understood. But there are rarely accepted engineering methods available to recommend on cost-vs.-risk efficient service levels. In a SLA-aware service composition problem for minimizing cost without compromising the service quality, there is a great research interest in integrated management tools that automatically control the quality of multi-tier sourcing applications and autonomously warn for arising problem on frontend impacts to end-users or other business process implications. The proposed concept is providing a bridge between business impacts to distributed infrastructure systems to recommend for capacity and quality of technical components by defining fuzzy dependency couplings in a practical and feasible manner in order to satisfy aspects of the distributed nature of SLA dependencies.

1 General Approach

When considering the trade-off between costs and benefits in the service design phase, a key challenge is the derivation of business-relevant performance metrics and associated cost-efficient target values. We propose as the efficient Service Level Objective (SLO) to select the minimum sum of business opportunity cost based on the associated customer business risks and the cost of delivery/capacity for the service provider. This implies that neither the service provider nor the customer can find the efficient service level objective by their own as this is the aggregation of both views and interests. In the following we address the risks based on negative consequences associated with a service incident (Cost-of-Failure). Thus, the negative impact on process performance with regard to a chosen service level is translated into its monetary equivalent. This can be called an adverse business impact in service offers [3]. Considering in addition the coupling for each service to possible adverse business impacts, quantitative assessments of service value and monetary consequence can be elaborated. Thus chaining the critical relationships between business impact and IT assets allow service provider to balance between cost of service delivery or capacity for a service level to the associated business risks. This coupling approach can also be leveraged to justify costs for improvements to the backend IT infrastructure that improve the Quality of Services (QoS) on the front-stage or end-user level, by demonstrating how the proposed improvements deliver monetary benefits to the business.

2 Defining Dependence Couplings by applying Fuzzy Set Theory

2.1 Estimating the Degree of Coupling

Dependence Coupling is a measure that we propose to capture how dependent the component or service is on other services or resources for its delivery. Loose coupling describes an approach where integration interfaces are developed with minimum assumptions between the sending/receiving parties, thus reducing the risk that failure in one module will affect others. Tight coupling on the other hand indicates that successful delivery of other services or availability of resources is a prerequisite for the completion of a service. In very most cases the relationship between technical performances attributes and dependent business services cannot be exactly mathematically described [4]. When the dependency is between a service and some resource it uses, coupling will essentially be a function of how often the resource is used. For instance, the dependence of a service on the network layer might be measured by how often it is making a socket call, or how much data it is transferring [5]. For web-services we can examine environmental coupling which is caused by calling and being called. More advanced are coupling measures to evaluate the coupling level real-time by runtime monitoring, introduced as dynamic coupling metrics. Within an inductive approach to investigate for coupling dependencies between infrastructure components and their relying services, historical data is collected from the whole application environment network and the performance behavior of related components is analyzed.

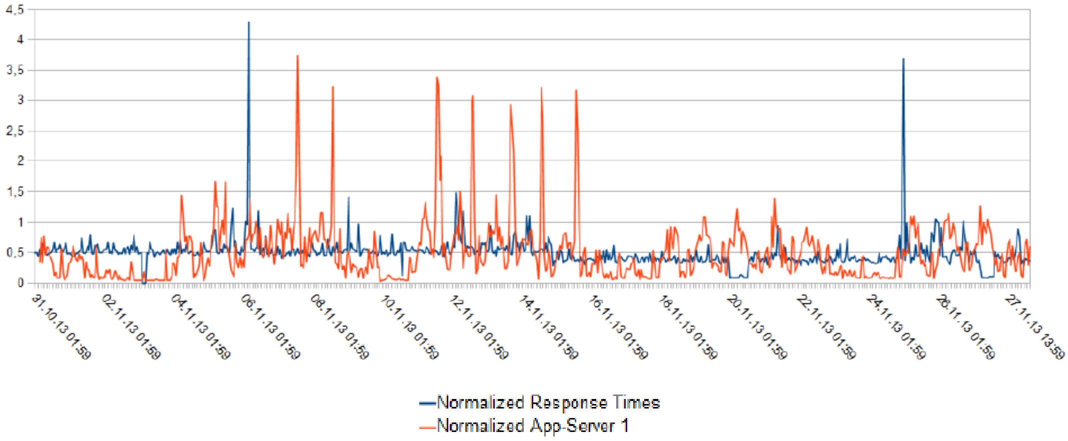

Figure 1: Inductive Comparison of User Response Time to Application Server Load

As opposite, using a deductive method dependencies are not calculated based on data the system produces, but rather the system itself [6],[7].

A key principle of realistic impact simulations is the idea of considering both, positive and negative aspects of dependency relations and simultaneous consideration by pulling these strengths together. The negative effects are the dependencies through transactional interactions (in case an incident happens), the controversy mitigation ability are the built-in systeme resilience capabilities. The simultaneous play of contrary forces, dependence and resilience together, considering direct and also indirect interactions, will define the overall system behaviour and expected impact to the end-user experience and business processes.

An initial gradual and bi-polar concept called IFSFIA (Intuitionistic Fuzzy Service Failure Impact Analysis) was first published in [8], which describes a seven step approach for assessing impacts by means of fuzzy-related components to a business service and has been developed at University Fribourg, Switzerland [9].

2.2 *Mathematical Background on Intuitionistic Fuzzy Sets (IFS)*

Let E be a fixed universe and A is a subset of E. The set $A^* = \{\langle x, \mu A(x), \nu A(x)\rangle | x \in E\}$ where $0 \leq \mu A(x) + \nu A(x) \leq 1$ is called Intuitionistic Fuzzy Set (IFS) [10]. Every element has a degree of membership (validity) $\mu A(x)$: $E \rightarrow [0,1]$ and a degree of non-membership $\nu A(x) : E \rightarrow [0,1]$. Intuitionistic Fuzzy Sets (IFS) have only loosely related membership and non-membership values unlike classical [Zadeh] fuzzy sets. An IFS is a generalization of the classical fuzzy set [11] which defines another degree of freedom into the set description, the independent judgment of validity and non-validity. For each IFS A in E, $\pi A(x) = 1 - \mu A(x) - \nu A(x)$ is called the intuitionistic index of x in A which represents the third aspect, the degree of uncertainty or limited knowledge.

Let now *a* be the intuitionistic fuzzy logical statement of coupling with membership and non-membership $\langle \mu_a, \nu_a\rangle$. The coupling degree of truth is $\langle \mu_a\rangle$ and degree of falsity $\langle \nu_a\rangle$ with possible values between zero and one omitting that the sum of both degrees of truth is equal or less than one. Pulling the positive and negative coupling aspects together in one fuzzy index is further called the Intuitionistic Fuzzy Direct Coupling Index (IFDCI) between two components.

2.3 *Calculation of Indirect Coupling Effects*

A property graph is an attributed, labelled, multi-relational graph which contains connected entities and which can hold attributes in the form of key-value pair [12]. This graph is applied to model multi-level business service dependencies to component services. Any Configuration Item (CI) can be presented as a node of the property graph which has a directed association to one ore more other nodes.

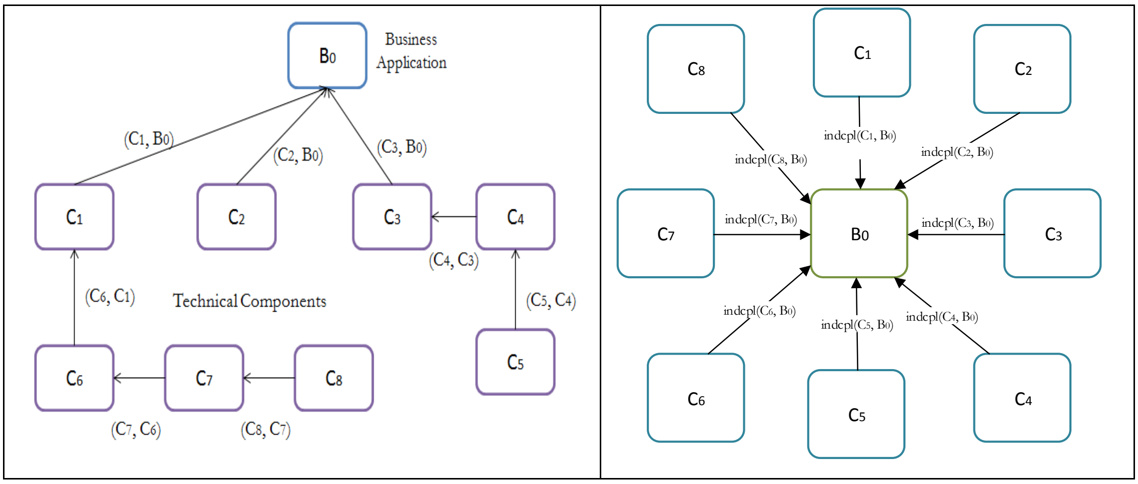

Figure 2: Directed Dependency Property Graph Model with Indirect Couplings

The edge reflects the relationship between two nodes which always has a direction, a dependent and an antecedent. The edge can be named and like nodes can have properties such as weights, distances, costs, ratings etc.

After definition of the direct impacts, the indirect coupling between components or services can be automatically calculated considering the degrees for direct coupling as invented by [13] within a Fault Tree Analysis concept and extended by [14]. The partial impact between components or services can now be expressed by means of intuitionistic fuzzy values carrying probabilistic information. Depending on which operations are applied, classical or probabilistic, the results will be greater or smaller (interpreted as classical, moderate, worst and best case).

Worst case impact analysis $V(p \wedge q) = \langle \min(\mu(p), \mu(q)), \max(\nu(p), \nu(q)) \rangle$ $V(a \vee b) = \langle \mu(a) + \mu(b) - \mu(a).\mu(b), \nu(a).\nu(b) \rangle$	**Best case impact analysis** $V(p \wedge q) = \langle \mu(p).\mu(q), \nu(p) + \nu(q) - \nu(p).\nu(q) \rangle$ $V(a \vee b) = \langle \max(\mu(a), \mu(b)), \min(\nu(a), \nu(b)) \rangle$
Moderate impact analysis $V(p \wedge q) = \langle \mu(p).\mu(q), \nu(p) + \nu(q) - \nu(p).\nu(q) \rangle$ $V(a \vee b) = \langle \mu(a) + \mu(b) - \mu(a).\mu(b), \nu(a).\nu(b) \rangle$	**Classical fuzzy impact analysis** $V(p \wedge q) = \langle \min(\mu(p), \mu(q)), \max(\nu(p), \nu(q)) \rangle$ $V(a \vee b) = \langle \max(\mu(a), \mu(b)), \min(\nu(a), \nu(b)) \rangle$

Figure 3: Calculation on Intuitionistic Fuzzy Indirect Couplings

The intuitionistic fuzzy dependencies may have different kinds of semantics (functional and probabilistic) depending on the type of information they represent.

Example: Considering the best case impact scenario between component C2 and service B0, the coupling relation is calculated as $indcpl_{best}(C2,B0) = (0.36,0.51)$. Using probabilstic semantics it means that in case the component C2 fails, the expected probability that business service B0 breaches the SLA is 36% and 51% that the performance of B0 stays within the tolerated thresholds. An uncertainty of 13% is estimated which means this coupling relation is seen as quiet mature.

As an example for a functional semantical interpretation using best an ordinary measurable coupling relationship this statement would mean that the service B0 is expected to be functional degraded or partly available (e.g. response time goes down by 36%) in case the component C2 performance fails. This allows a notion of having the business service still usable with some kind of degradation.

3 Cost Models and Cost-of-Failure

Virtual Capacity Units (VCU), which may represent several technical elements in the IT system that are related to the virtualization services offered to the clients.

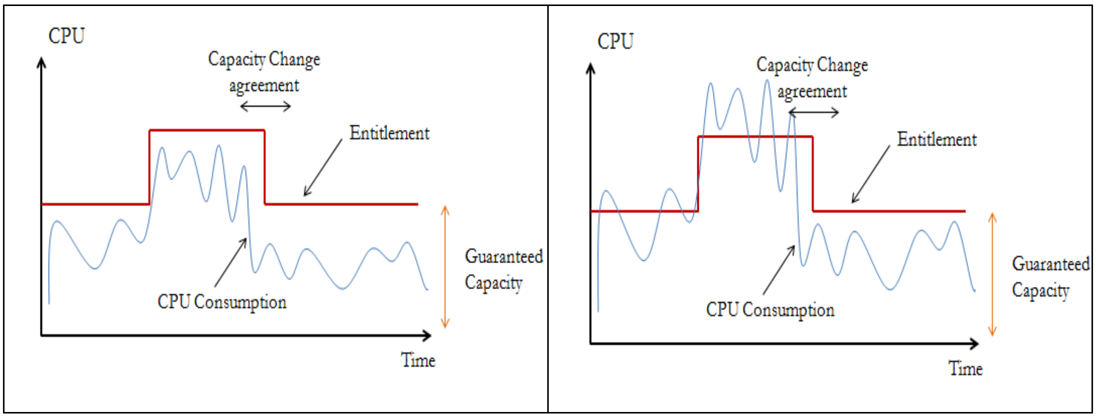

Figure 4: Baseline Model versus Consumption Based Model

Within Baseline Model the customer books a fixed capacity in VCU on a shared system. He is able to change this capacity by notifying a change request to the service provider. The customer pays the booked capacity independent of the actual usage. Despite in a Consumption Based Model (CBM), the consumer books an entitlement in VCU, the logical partition is allowed to burst over that entitlement and this burst is not guaranteed. At the end of the billing period Bursting is measured and charged. The catalog service products offer a different booked and committed capacity that the clients can choose depending of their needs. As closer the entitlement is to 100%, which is the case only for the Platinum Service Level, the better availability and performance conditions apply. But the higher the entitlement, also the higher the cost to pay for the service usage, here the consumed VCU units. Each Service Level has an associated negative Bussiness Impact or Cost-of-Failure based on the expected numbers of incidents (like too slow responses) which is contrary to the capacity entitlement of the chosen Service Level.

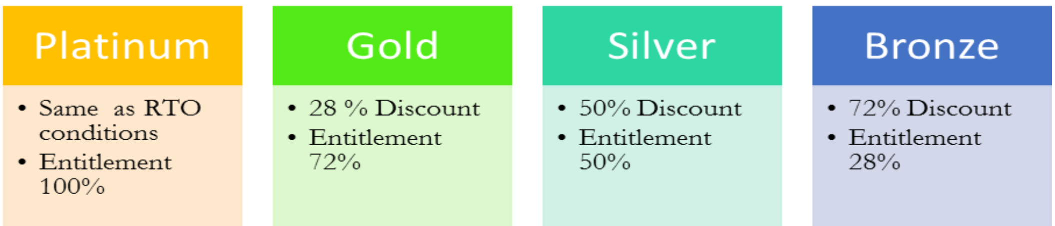

Figure 5: Service Levels with Capacity Entitlements and corresponding Discounts

The expected total Cost-of-Failure is related to the booked Service Level, we summarize this cost of lost business further as Business Opportunity Cost.

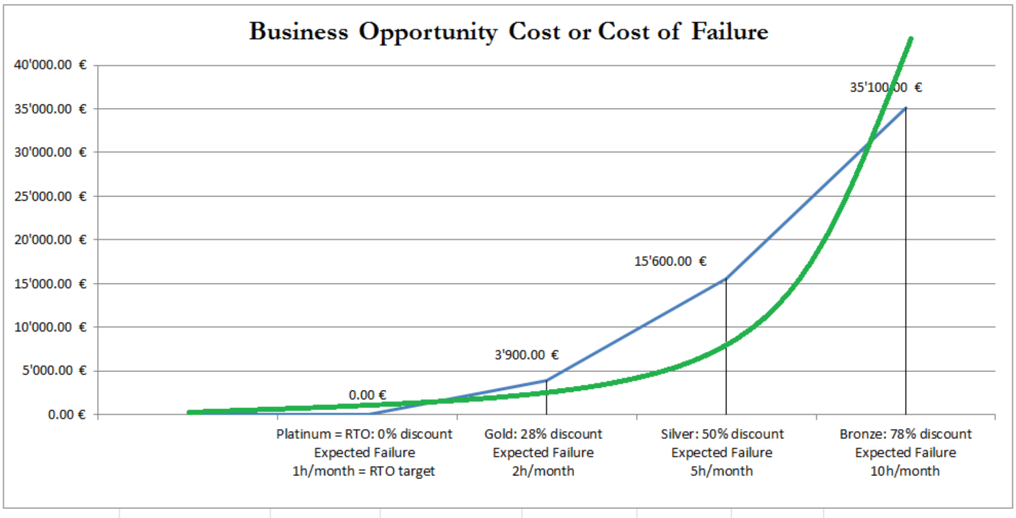

Figure 6: Business Opportunity Cost Curves: Real (Blue) and Ideal (Green)

4 Intuitionistic Fuzzy Service Failure Impact Analysis (IFSFIA)

A complete methodical assessment approach was developed, which is practically usable in datacentre environments, with several sequential steps to be processed. It starts from automated discovery of the details of the managed resources and backend components, the grouping of components to impacted frontend services and the enrichment in several tasks and calculation steps up to the gradual business impact assessments, including monetary cost-of-failure information and business objectives [14]. The overall frame for incorporating all data is a grid with service components on one axis and impacted business services on the other. This matrix can be freely extended with different kind of variables showing failure modes, reliability parameters, financial data or operational capabilities and extends the pure system view to include also the processes, tools and people that are necessary for functioning of a distributed information system.

The result of the IFSFIA method is a sorted intuitionistic fuzzy distribution of dependencies, providing an ordered set by the probability of incident root causes.

Extended IFCFIA Grid with indirect Couplings and Cost of Failure			Cost of Failure per Hour 10000 Euros / RTO and RPO for 1 hour / B0 / FCC Coupling to B0		Total Failure Cost Per hour		
Component	**Failure Mode and Effect**	**Direct Impact (IFS)**	**Prob. Failure**	**Prob. Resilience**	Failure	Uncertainty	Resilience
C1	Outage	(0.4,0.4)	0.47	0.47	4'700.00 €	600.00 €	4'700.00 €
C2	Outage	(0.8,0.1)	0.39	0.54	3'900.00 €	700.00 €	5'400.00 €
C3	Outage	(0.6,0.3)	0.63	0.33	6'300.00 €	400.00 €	3'300.00 €
C4	Limited Function	(0.5,0.5)	0.47	0.47	4'700.00 €	600.00 €	4'700.00 €
C5	Slow Response	(0.4,0.5)	0.39	0.54	3'900.00 €	700.00 €	5'400.00 €
C6	Limited Function	(0.3,0.4)	0.63	0.33	6'300.00 €	400.00 €	3'300.00 €
C7	Slow Response	(0.5,0.3)	0.63	0.33	6'300.00 €	400.00 €	3'300.00 €
C8	Slow Response	(0.4,0.4)	0.47	0.47	4'700.00 €	600.00 €	4'700.00 €
Technical Support	Slow Repair	(0.8,0.1)	0.39	0.54	3'900.00 €	700.00 €	5'400.00 €
User Experience	Quality Issue	(0.6,0.3)	0.63	0.33	6'300.00 €	400.00 €	3'300.00 €

Figure 7: IFSFIA Matrix with Couplings and Financial Impacts per Component

Here Repair Time Objective (RTO) and Recovery Point Objective (RPO) are roughly speaking based on setting the time objective for "the amount of time the business can be without the service”. This means in case the defined service level enables a full restoration within these time limits no severe effect on the business will be assumed and the Cost-of-Failure calculation starts after this grace period.

5 Recommendation on efficient Service Levels

Each service linked to the business impacts in terms of Business Opportunity- or Failure Costs and can be matched against the Cost of Capacity or Delivery Cost.

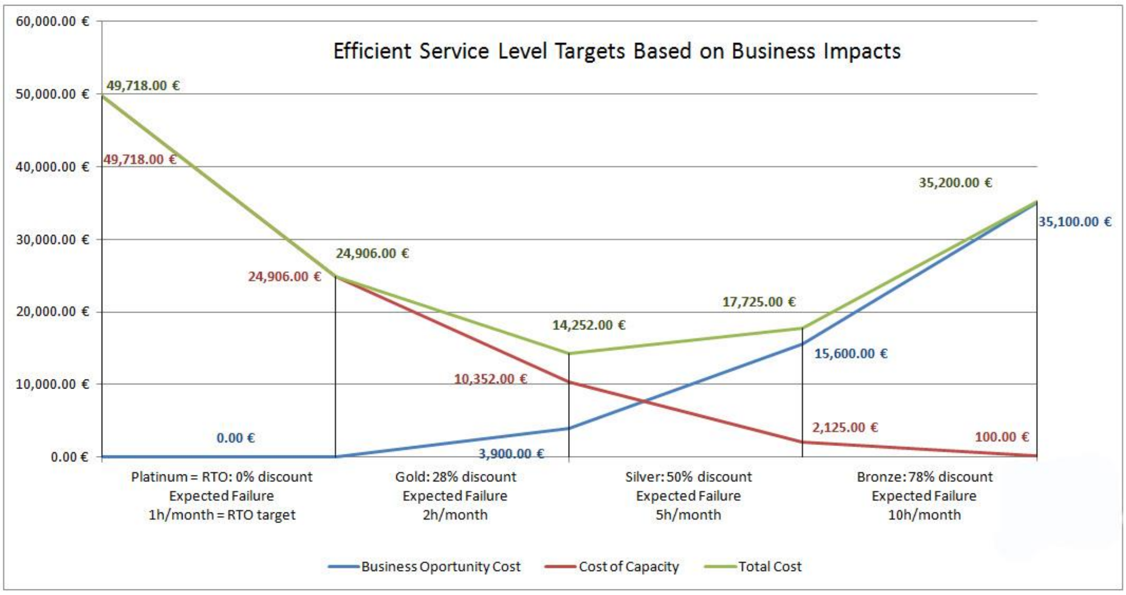

Figure 8: Efficient Service Level Targets Based on the Business Impacts

The expected business impact can be compared against the discounts indicating the cost benefits for chosing a lower SLO capacity target. The minimum point of a total cost (not be necessarily the intersection between the Business Opportunity Cost and Cost of Capacity) is below be determined as an approximation.

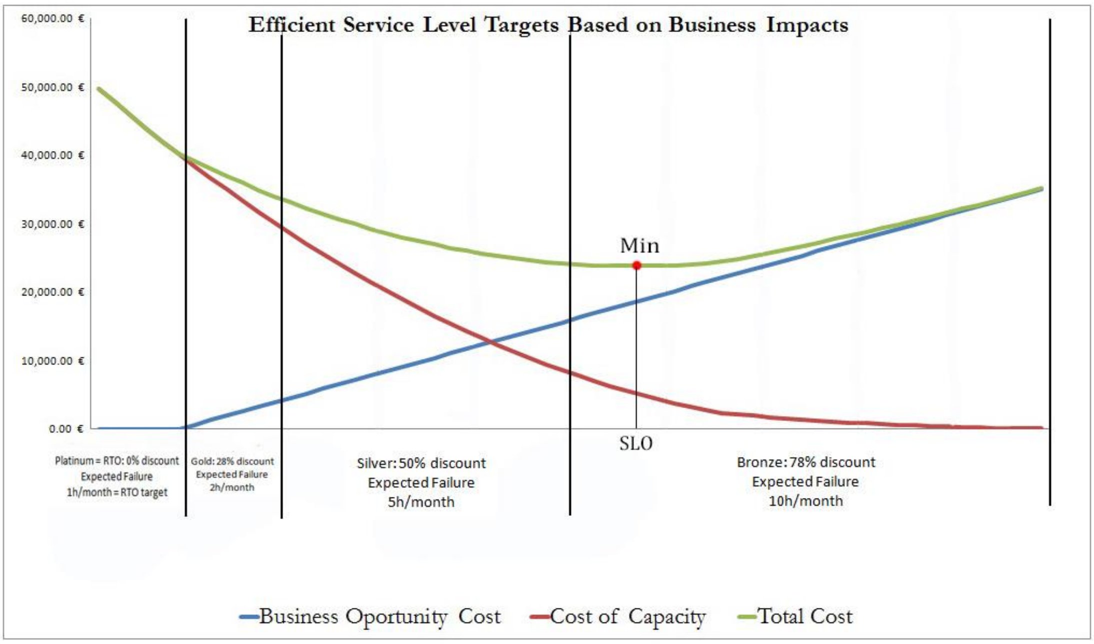

Figure 9: Efficient SLO targets with minimized overall total cost.

The cost-optimized Service Level Objective (SLO) has a total cost of 23813.06 € as a value and the point is located on the Bronze Category. Thus the best product to select minimizing total costs and better overall business opportunity for client and service provider is the bronze category. The proposed method can be taylored to choose services depending on the individual customers' business model.

Conclusion: Combining well-grounded academic research for Service Level Engineering, as also contributed in related research work [5],[15],[16], with practice oriented scenarios provides high value to the service business. By expanding IT reliability engineering with fuzzy mathematical models the framework can transform performance data into knowledge that allows understanding the impact of incidents on the business and recommend on efficient service level objectives. Instead of tightening SLAs across the board, which is a costly approach, individual service levels should be directly driven by business needs where the organization benefits by projecting and paying for only what is required.

References

[1] Guo T et al.:Seagull: Intelligent Cloud Bursting for Enterprise Applications: https://www.usenix.org/system/files/conference/atc12/atc12-final57.pdf

[2] Mattes, M et al.: Cloud Bursting: Managing Peak Loads by Leasing Public Cloud Services http://www.cloudbus.org/papers/CloudBurst-BC-2011.pdf

[3] Barroero et al.: Aligning IT service levels and business performance: *Proceedings of IEEE Computer Society, Miami, Florida (2010), pp. 570–577*

[4] Hui Li (2009) Challenges in SLA Translation – *SLA@SOI European Commission Seventh Framework Programme SAP Research, Dec. 2009*

[5] Joshi, K. P., Joshi, A., & Yesha, Y. (2011) Managing the Quality of Virtualized Services, *Proceedings of the SRII Service Research Conference.*

[6] Alghamdi, J. S. (2007). Measuring software coupling. *Proceedings of the 6th WSEAS International Conference on Software Engineering.*

[7] Quynh, P. T. (2009) Dynamic Coupling Metrics for Service–Oriented Software, *International Journal of Electrical Engineering*, 3(5), p. 282.

[8] Schütze, R. (2013) Intuitionistic Component Failure Impact Analysis, *Notes on Intuitionistic Fuzzy Sets*, 19(3), 62–72.

[9] Andrade, J.A. M. Sourcing Strategies for Virtualized Services: A Fuzzy Risk-based Approach, *Seminar Paper Univ. Fribourg, Dept. Inf., Dec 14*

[10] Atanassov, K. (1999) Intuitionistic Fuzzy Sets: *(Studies in Fuzziness and Soft Computing), Springer, Heidelberg.*

[11] Zadeh, L. (1994) Soft Computing and Fuzzy Logic, *IEEE Software*, 11(6)

[12] M. A. Rodriguez and P. Neubauer, "Constructions from dots and lines," *Bulletin of the American Society for Information Science, vol. 36, 2010*

[13] Kolev, B., & Ivanov, I. (2009) Fault Tree Analysis in an Intuitionistic Fuzzy Configuration Database, *Notes on Intuitionistic Fuzzy Sets*, 15(2), 10–17.

[14] Schütze, R.: IFCFIA, a gradual method for SLA dependency mapping. *Technical Report 07/13, Univ. of Fribourg, Switzerland, Dept. Informatics*

[15] Kieninger, A., Berghoff, F., Fromm, H., & Satzger, G. (2013) Simulation-based Quantification of Business Impacts, Pro*c. of the 4th Int. Conf. on Exploring Service Science. (Portugal) 2013 LNBIP, No.143. Springer*.

[16] Schmitz, B., Kieninger,A., Satzger G., Fromm H. (2014) Towards the Consideration of Performance Risks for the Design of Service Offers, *IESS 2014, LNBIP 169, pp. 108–123, 2014*.