ICT
SOFTWARE AND SYSTEMS
ENGINEERING

Introduction to Service Level Agreements

Ulrik Franke
Thomas Olsson

RISE Report 2019:23

Introduction to Service Level Agreements

Ulrik Franke
Thomas Olsson

Key words: Service Level Agreements; Cyber-physical systems; PIMM DMA

RISE Research Institutes of Sweden AB

RISE Report 2019:23
ISBN: 978-91-88907-49-3
Kista 2019

Content

1 Introduction

Modern industrial production environments are rapidly transforming. Concepts such as smart industry and Industry 4.0 encompass many expectations on how digital technology can improve industrial plants. Some strands are better algorithms for robotics, better situational awareness through ubiquitous RFID, fewer production interruptions through smarter predictive maintenance, and more agile production lines enabling greater customization of products. Many of these ideas depend on reliable access to IT services such computing power and data availability. If these falters, the benefits will not materialize. Therefore, it is crucial to study the *Service Level Agreements* (SLAs) that are used to regulate such services.

SLAs typically include requirements on performance (e.g. a response time of a maximum of 100 milliseconds), availability (e.g. 99.98%) and maintenance windows (e.g. scheduled maintenance between 3 and 4 a.m. on the first Tuesday of every month) for the services purchased. When IT is provided in-house, such SLAs may not be so important. There are other ways to govern quality of service, and incentives are more aligned. But for externally procured IT services, contracts such as SLAs can be the *only* avenue, short of changing supplier, to govern quality of service and manage the risks of service degradation.

For obvious reasons, such all-out service-orientation has first taken place in pure IT services such as digital payments, financial services, etc., where information only is created, transmitted, transformed, and analyzed. Therefore, incumbents in industries such as banking and insurance now worry that they will be outcompeted by disruptive, "born digital" startups.[1]

In various industrial production settings, it is less straightforward to outsource IT into the cloud. Indeed, digital equipment in these settings, such as PLCs, RTUs, SCADA systems, etc. are typically referred to as *operational technology*, OT, rather than IT, to distinguish their unique features such as tighter integration of hardware and software, longer lifecycles, special requirements on physical ruggedness, and performance designed for use in real time control loops. Nevertheless, transitions to "as-a-service" paradigms are bound to happen in industrial production as well. Better understanding of how such services should be governed is one of the building blocks of smarter industry.

2 The importance of understanding the business

As in all IT service management, the starting point for any SLA must be the business requirements. What is it that is required by the business? What are the consequences if these requirements are not met? How much is it worth to invest in order to avoid certain outcomes? If questions like these cannot be answered, it is meaningless to draw up an SLA.

[1] "The future of insurance: Counsel of protection", The Economist, March 11, 2017, 67–68.

To answer the relevant questions, a *risk analysis* should be conducted. Of course, there are many methods to conduct such analyses, but in general they all first (i) identify risks, then (ii) assess their impacts, and finally (iii) manage the risks accordingly. Some risks can be *avoided* by changing processes or technology. But some risks are inherent to the organization and thus very difficult to avoid, e.g., the risks of a business model, a jurisdiction, a customer base etc. However, some such risks can be *mitigated* by investments, or *transferred* to another party such as a service provider or an insurance company. Risks that cannot be avoided, mitigated, or transferred must simply be *accepted*.[2]

Once a risk analysis has been conducted, it is possible to address the contents of an SLA in a more meaningful way. The risk analysis can answer core SLA questions such as, e.g., the following:

- When must services be available? Are there any windows when they can be unavailable?
- Is there an order of prioritization of services that need to be available, due to different business requirements?
- Are the consequences of service outages always the same, or are there certain times (e.g., during Christmas shopping, when quarterly financial statements are produced, or when a particular artefact is assembled) when consequences are worse?
- What is the time acceptable before service is restored? Are there different acceptable times for different services?
- Are there any data that must not be lost? Are there data that can be unavailable for a while without major consequences?

An SLA should never just be based on gut feeling, tradition or a template that has not been actively deemed appropriate for the business that will use it. A good SLA is based on a good understanding of the business that needs the services procured, and on a risk analysis of what happens when those services become unavailable.

3 SLAs and definitions of terms

One important aspect of any contract is the definition of terms. It is prudent to define all important terms that will be used in a contract in a special section or as an appendix. This minimizes the risk that the parties interpret terms differently, or that a court will impose a meaning not intended by any of the parties.

The definition of an *error* is important. For example, has an error occurred if the service is only partially down, e.g., it works for some users but not all, or if some packages/jobs/transaction are lost? If the service is slower than usual? If the service is up, but only because a redundant fail-over has been activated? If the API has been changed, unannounced?

There are no general answers to these questions that apply to all kinds of services in all lines of business. However, if no other definition of error is given in the contract, under

[2] The classification of risk management responses is from D. Hillson. Extending the risk process to manage opportunities. *Int J Project Manage*, 20 (3) (2002), pp. 235-240, http://dx.doi.org/10.1016/S0263-7863(01)00074-6

Swedish law this means that the definition from the Buying Act (*Köplagen*) will apply, i.e. that "the product shall be suitable for the purpose for which similar products are generally used" (17§ KL). This is a vague wording, so the consequences of a court applying it in a technically advanced case are difficult to foresee, underscoring the importance of including applicable definitions in the contract.

Some examples of other terms that can be important to define include *service*, *support*, *maintenance*, *availability*, *time to restore service*, *response time*, *service window*, *upgrade*, *release*, *update*, etc.

4 SLAs, templates, standards and guidelines

While templates should never be used without making sure that they are suitable for the case at hand, they can nevertheless be very useful, both by reducing the amount of paperwork when entering into a contract, and by making different offerings more comparable. The Swedish IT and Telecom Industries (*IT&Telekomföretagen*) offer a range of standard templates for contracts, e.g., SLA templates for IT maintenance, IT infrastructure services, and cloud services.[3]

To get an understanding of the strengths and weaknesses of SLA templates, standards and guidelines, it is instructive to consider how availability and service interruptions are treated in some sources.

Unplanned service outages can occur at any time. While the SLA cannot govern *when* they occur, it can set requirements on *how* they are handled. Essentially, there are three such requirements that can be set:

- Requirements on the *number* of unplanned interruptions (e.g. 10 at most) during the contract period.
- Requirements on the (maximum or average) *time to restore* service (e.g. 2 hours) when an interruption occurs.
- Requirements on the *average availability* (e.g. 99.98 %) during the contract period.

It is prudent to set all three requirements to ensure that the required service level is achieved.[4] But this is not always the case in templates, standards and guidelines.

For example, renowned IT strategy consultancy Gartner sometimes stresses the importance of setting requirements on the time to restore service, for example in reports on SLAs for continuously available IT services[5] and SLAs for SaaS.[6] However, in other

[3] IT&Telekomföretagens Standardavtal, https://webshop.almega.se/it-telekomf%C3%B6retagen, accessed January 25, 2019.
[4] U. Franke, "Optimal IT service availability: Shorter outages, or fewer?". IEEE Transactions on Network and Service Management. 9(1):22-33, March 2012.
[5] B. Malik and D. Scott, "How to calculate the cost of continuously available IT services," Gartner, Inc., Tech. Rep., Nov. 2010.
[6] D. Williams, "The challenges and approaches of establishing IT infrastructure monitoring SLAs in IT operations," Gartner, Inc., Tech. Rep., September 2010.

SLA reports on clouds[7] and SaaS[8], requirements on time to restore are not mentioned. Similarly, the sample SLA found in the ITIL volume on service design suggests only to include the number of interruptions and the average availability, not the time to restore.[9]

Among the templates available from the Swedish IT and Telecom Industries there are three specific SLA appendices for IT maintenance, IT infrastructure services, and cloud services, respectively:

- In the SLA appendix for IT Maintenance, hours of service (e.g., 8 a.m. to 5 p.m.) and agreed response times (e.g., 2 within hours) for different categories of faults are to be defined. However, the number of interruptions or the overall average availability of the service being maintained are not mentioned. This may be reasonable if the SLA applies only to *the maintenance service*, not to *the IT service being maintained*, but this serves as a good example of the crucial importance of understanding and agreeing on all definitions in the contracts.
- In the SLA appendix for IT Infrastructure Services, an agreed availability (e.g. 99.98 %) is to be defined. Furthermore, *implicit* requirements on the time to restore service are set in the form of agreed hours of service and an agreed helpdesk response frequency (e.g. 75 %) within a certain time (e.g. 10 minutes). However, no hard requirements on the time to restore service are set. Also, no requirements are set on the *number of unplanned interruptions*, though the number of *planned* (maintenance) interruptions is set to once a month by default.
- The SLA appendix for Cloud Services is identical to that for IT Infrastructure Services with respect to the aspects described above.

To conclude, while templates can be very useful, it is important to always assess them from the point of view of what is needed by the business, as informed by the risk analysis. If may be that what the business really needs is not available on the market (at a reasonable price), but it should never be that what the business really needed was never on the negotiating table because it was not there in a template.

5 Example: translating business requirements into an SLA

The National Agency for Public Procurement (*Upphandlingsmyndigheten*) offers an instructive example of how to translate business requirements into SLA clauses. In this case, the objective is to procure an externally hosted electronic procurement system in the Swedish public sector. The guidelines published by the National Agency for Public

[7] D. O'Connell and D. Kraus, "Critical elements of cloud-based contact center services: pricing, service-level agreements and service integration," Gartner, Inc., Tech. Rep., June 2010.
[8] B. Pring, C. Ambrose, W. Maurer, and A. Bona, "Best practices for service-level agreements for software as a service," Gartner, Inc., Tech. Rep., November 2010.
[9] S. Taylor, V. Lloyd, and C. Rudd, *Service Design (ITIL)*, the Stationery Office, 2007. Cf. appendix F for the SLA.

Procurement offer some considerations on how the business requirements in this case should affect an SLA:[10]

- First, it is noted that imposing an *average availability* requirement (e.g. 99.98 %) during the contract period is not relevant for most smaller public sector actors. Instead, it is recommended to impose an availability requirement the time windows when it is foreseen that the system will be used, i.e., during office hours and some evening time.
- It is recommended to impose a requirement that planned service outages are announced well before they take place, e.g., one or two months. This relates to the particular procurement business requirement that a tender shall be announced for 52 days, and the risk that the system is down for maintenance on the last day should be avoided.
- It is also recommended to impose requirements on the support offered, e.g., its opening hours, the ways to contact it, and how the tickets are registered and followed-up. Specifying different categories of tickets with correspondingly different time requirements is also recommended. However, it is also noted that the support on offer is typically the same to all customers of the provider, and for a small public sector actor, it is not always relevant to impose particular requirements over and above those that are on offer.
- As a final good practice, it is recommended to have regular meetings with the supplier to discuss availability, support, and any errors that have occurred.

While the details of these considerations are only relevant to those actually procuring electronic procurement systems in the public sector, the kinds of reasoning and the deductions from business requirements to SLA clauses are of interest to a wider audience. Note that in some cases, it is considered reasonable to just go with whatever is on offer (e.g. support processes), in other cases it is crucial to the business that some particular requirements are met (e.g. announcement of planned service outages). Such a mix can be expected to occur in most lines of business. It is important to not over-reach but rather impose only the most relevant requirements.

6 Negotiable and non-negotiable SLAs

SLAs are often treated as objects of negotiation, just like any contract. Especially the scientific literature abounds with work aiming to enable and improve automatic SLA negotiation between intelligent autonomous agents in marketplaces.[11]

Nevertheless, SLA negotiations are prone to pitfalls. For example, there is often a gap between IT and operations, where IT departments find it difficult to formulate

[10] Vägledning: Att införa elektronisk upphandling, Kammarkollegiet, https://www.upphandlingsmyndigheten.se/globalassets/publikationer/kammarkollegiet/vagledning/2012-2.pdf, accessed January 28, 2019.

[11] Cf. e.g. G. C. Silaghi, L. D. Serban & C. M. Litan. ”A framework for building intelligent SLA negotiation strategies under time constraints”. In: *Economics of Grids, Clouds, Systems, and Services*, pp.48–61. Springer, 2010 and E. Yaqub, R. Yahyapour, p. Wieder, C. Kotsokalis, K. Lu, & A. i. Jehangiri. “Optimal negotiation of service level agreements for cloud-based services through autonomous agents”. In *2014 IEEE International Conference on Services Computing (SCC)* (pp. 59-66). IEEE, 2014.

availability requirements so that the business understands.[12] Research also suggests that IT decision makers find it difficult to make rational SLA accessibility decisions under risk, failing to maximize expected value,[13] and that the information expressed in the SLA can lead to sub-optimal decisions.[14]

Furthermore, all SLAs are not negotiable. Comuzzi et al. make the useful distinction between SLAs that are either unilateral (U-SLAs), typical of public cloud services where the offer is essentially *take it or leave it*, or bilateral (B-SLAs), typical of private cloud offerings and traditional IT outsourcing.[15] It is an increasingly important task for anyone procuring IT as a service to analyze and decide in which cases services with non-negotiable SLAs are acceptable, in which cases SLAs must be negotiable, and how to combine the two into an overall architecture aligned with the risk appetite of the enterprise.

7 Partitioning of risk in ecosystems

Such considerations about risk in an overall architecture take on a special significance in an ecosystem, such as the PIMM DMA context, where many different services from different actors are needed to provide certain functions. Such a function may be affected not only if the services become unavailable, but also if their quality of service is degraded, e.g., by longer response times or less throughput.

Such loss of functionality typically entails a cost, and the key question thus becomes who bears this cost. Is it the service provider of the service affected, is it the service recipient, or is it perhaps another service provider who has assumed the responsibility of an integrator? Do the same principles apply for all services, for all magnitudes of cost, and for all kinds of disruptions?

Following the reasoning above, these questions should be resolved by all the parties conducting their risk analyses and then negotiating corresponding SLAs. However, this may be difficult for many reasons. Parties used to offer only U-SLAs will be reluctant to offer B-SLAs. Very large parties may be reluctant to negotiate separately with (many) small parties. The fact that the number of possible bilateral interactions/SLAs grows quadratically with the number of involved parties (and even more if trilateral or higher order contracts are needed) can impede cooperation.

The concept of risk sharing related to IT services is, of course, not new. The standard template for agile projects available from the Swedish IT and Telecom Industries includes an appendix on risk- and profit-sharing, intended precisely to provide

[12] N. Rickard & D. Young. “Bandwidth Doesn’t Matter; Availability Drives Enterprise Network Costs.” Gartner, Inc., Tech. Rep., July 2013.
[13] U. Franke & M. Buschle. Experimental Evidence on Decision-Making in Availability Service Level Agreements. *IEEE Trans. Network and Service Management*, *13*(1), 58-70, 2016.
[14] A. Kieninger, D. Straeten, S. Orla Kimbrough, B. Schmitz & G. Satzger. Leveraging service incident analytics to determine cost-optimal service offers. In: *11th International Conference on Wirtschaftsinformatik*, pp. 1015–1029, 2013.
[15] M. Comuzzi, G. Jacobs & P. Grefen. Understanding SLA elements in cloud computing. In *Working Conference on Virtual Enterprises* (pp. 385-392). Springer, Berlin, Heidelberg. September 2013.

reasonable incentives to the parties in the face of uncertainty. However, this situation is arguably less complex than that of a larger ecosystem, such as that in PIMM DMA.

One solution, at least in the short term, is to cooperate anyway, *without* clear contractual underpinnings and thus without a clear partitioning of the risk. However, the disadvantages of this solution are obvious. Indeed, the Swedish Civil Contingencies Agency (MSB) has pointed out that while such cooperation without contracts is common in the public sector, it is not desirable, and the prospects for allowing public sector entities to enter into the equivalent of commercial agreements with each other should be investigated.[16] Cooperation without contracts makes it more difficult to achieve a mature partitioning of risk. Lack of proper incentives might be an explanation of why Swedish government agencies are immature when it comes to business continuity planning: In a 2014 survey, 65% of the responding government agencies did not have a business continuity plan at all.[17]

Another solution, if the parties in a complex ecosystem cannot agree on how to partition the risk among themselves, is to bring in an external insurer, i.e., someone who is neither a service provider nor a service recipient, who can assume risk for a fee. Any residual risk that cannot be assumed within the ecosystem could in principle be insured in this way – if an agreement can be reached with an insurer. The market for cyber insurance is growing, globally as well as in Sweden,[18] and it is not unreasonable that insurers could play a risk management role in complex ecosystems. However, it is also reasonable to expect that risks that are poorly understood and where historical incident data is scarce will *not* be insurable on the market.

[16] Outsourcing av it-tjänster i kommuner. Myndigheten för samhällsskydd och beredskap, 2014. Publication number MSB728. https://www.msb.se/sv/Produkter--tjanster/Publikationer/Publikationer-fran-MSB/Outsourcing-av-it-tjanster-i-kommunen/

[17] En bild av myndigheternas informationssäkerhetsarbete. Myndigheten för samhällsskydd och beredskap, 2014. Publication number MSB740. https://www.msb.se/sv/Produkter--tjanster/Publikationer/Publikationer-fran-MSB/En-bild-av-myndigheternas-informationssakerhetsarbete-2014/

[18] U. Franke. The cyber insurance market in Sweden. *Computers & Security*, 68, pp. 130–144, 2017.

Through our international collaboration programmes with academia, industry, and the public sector, we ensure the competitiveness of the Swedish business community on an international level and contribute to a sustainable society. Our 2,200 employees support and promote all manner of innovative processes, and our roughly 100 testbeds and demonstration facilities are instrumental in developing the future-proofing of products, technologies, and services. RISE Research Institutes of Sweden is fully owned by the Swedish state.

RISE Research Institutes of Sweden AB
Box 857, SE-501 15 BORÅS, Sweden
Telephone: +46 10 516 50 00
E-mail: info@ri.se, Internet: www.ri.se

Software and Systems Engineering
RISE Report 2019:23
ISBN: 978-91-88907-49-3