SLA-aware operational efficiency in AI-enabled service chains: challenges ahead

Robert Engel[1] · Pablo Fernandez[2] · Antonio Ruiz-Cortes[2] · Aly Megahed[1] · Juan Ojeda-Perez[2]

Abstract

Service providers compose services in *service chains* that require deep integration of core operational information systems across organizations. Additionally, advanced analytics inform data-driven decision-making in corresponding AI-ena-bled business processes in today's complex environments. However, individual partner engagements with service consumers and providers often entail individu-ally negotiated, highly customized Service Level Agreements (SLAs) comprising engagement-specific metrics that semantically differ from general KPIs utilized on a broader operational (i.e., cross-client) level. Furthermore, the number of unique SLAs to be managed increases with the size of such service chains. The resulting complexity pushes large organizations to employ dedicated SLA management sys-tems, but such 'siloed' approaches make it difficult to leverage insights from SLA evaluations and predictions for decision-making in core business processes, and vice versa. Consequently, *simultaneous* optimization for both global operational process efficiency *and* engagement-specific SLA compliance is hampered. To address these shortcomings, we propose our vision of supplying online, AI-supported SLA analytics to data-driven, intelligent core workflows of the enterprise and discuss current research challenges arising from this vision. Exemplified by two scenarios derived from real use cases in industry and public administration, we demonstrate the need for improved semantic alignment of heavily customized SLAs with AI-enabled operational systems. Moreover, we discuss specific challenges of prescriptive SLA analytics under multi-engagement SLA awareness and how the dual role of AI in such scenarios demands bidirectional data exchange between operational processes and SLA management. Finally, we discuss the implications of federating AI-supported SLA analytics across organizations.

Contributions to this article by R. Engel and A. Megahed have been carried out during employment at IBM Research - Almaden.

Extended author information available on the last page of the article

Keywords Service level agreement · SLA · Service analytics · AIOps · AI · Machine learning · Service chains · Optimization · Prescriptive analytics · Operations research · Analytics · KPI · Key performance indicators

1 Introduction

Information systems in organizations face a new reality where the traditional boundaries of applications are blurred to rather create ecosystems of services that are orchestrated in highly data-driven, intelligent workflows. Much like supply chains, these *service chains* (a.k.a. service supply chains; Baltacioglu et al. 2007) can transcend organizational boundaries and require deep integration of business processes and information systems across organizations. As the complexity of such engagements increase, decision making to drive core operational processes becomes ever more challenging: Decision factors originating from across the service chain—typically represented by (intra- and inter-organizational) operational key performance indicators (KPIs) – need to be taken into account in order to improve *operational efficiency* (cf. Wang et al. 2015). Here, operational efficiency is assessed/measured based on the values of the realizations of the different aforementioned KPIs related to the operations of the underlying systems.

Additionally, advanced analytics and artificial intelligence (AI) based on detailed monitored data from core operational information systems oftentimes provide crucial information for controlling data-driven decision-making in corresponding business processes in real-time (cf. Forrester Research 2017). By *AI-enabled service chains* we refer to service chains infused with AI technologies that take advantage of the significant amount of data associated with the myriad of above-described decision factors that affect operational efficiency.

However, large engagements between consumers and providers of services that entail unique, highly customized service level agreements (SLAs) require organizations to employ dedicated SLA management systems to account for the specific metrics and measurement rules specified in those SLAs (cf. Sfondrini et al. 2015; Mubeen et al. 2017). The highly engagement-specific and inconsistent semantics of these SLAs leads to a 'disconnect' between the domains of SLA management on the one side and operational monitoring and decision making with standardized KPIs on the other side. In other words, informing AI-enabled business processes with data points from either domain is challenging. As a result, two equally important aspects of high-quality service delivery—namely, dynamic customization/optimization of processes and on-point SLA compliance - are not jointly considered, despite the potential synergies arising from their combination.

In such a context, our main aim is to pave the way to a new generation of Information Systems that provide support to engagement-specific, highly customized SLAs within AI-enabled service chains. Specifically, in this vision paper, our guiding questions are twofold: (i) Which are the main elements and perspectives to be taken into account when building such Information Systems? and (ii) Which are the research directions that could establish a potential roadmap to solve the problems involved in building those Information Systems?

It is important to note that this aspect of considering engagement-specific, highly customized SLAs within AI-enabled service chains has not been a primary focus in current literature. To the best of our knowledge, this aspect and its associated problems have not been stated before. Thus, there is no chance to find a specific baseline against which to compare our proposal. Furthermore, closer proposals to inspire a solution approach, such as trying to adapt negotiation or provide specific templates to harmonize the outcomes, helps to solve the problem's symptoms or effects, but not the problem itself. Consequently, we address the root of the problem by extending the notion of operational efficiency by *multi-engagement SLA-awareness*: We define it as the joint consideration of the aforementioned general factors for operational decision making together with engagement-specific, potentially heavily customized SLAs such that the global outcome - i.e., the result on organizational and/or supra-organizational levels across service chains—is optimal. Following this notion, we present our vision of providing detailed insights from online, AI-supported SLA analytics to data-driven, intelligent core workflows of the enterprise. The goal is to enable core operational (AI) systems to reason with a more comprehensive view of the state of business processes that extends beyond the typically collected cross-client data points to operational KPIs at any point in time. In particular, we suggest to integrate the real-time state and predictions from detailed SLA analytics across multiple *individual* partner engagements with heavily customized SLAs into operational decision-making algorithms. We envision that this will enable optimization algorithms that focus on overall (i.e., cross-client) operational efficiency to pursue *simultaneous* optimization of both global operational process efficiency *and* engagement-specific SLA compliance across multiple individual partner engagements with highly individualized SLAs. Furthermore, we suggest to establish bi-directional information exchange between AI models employed in SLA management systems and operational processes, respectively, in order to enable truly prescriptive (cf. Bertsimas and Kallus 2020) SLA analytics. Finally, we discuss the potential benefits and implications of the federation of SLA analytics across organizational and departmental boundaries across highly integrated service chains.

The remainder of this paper is organized as follows: In Sect. 2, we elaborate on current trends in industry and academia that motivate the vision presented in this paper. In Sect. 3, we present two scenarios derived from real use cases that will serve as running examples throughout the rest of the paper. This is followed by a description of research challenges that arise from and in our vision in Sect. 4. We conclude the paper in Sect. 5.

2 Background and state of the art

In the following, we elaborate on three current trends in industry and academia that motivate the vision presented in this paper: (i) The permeation of AI technologies into the core operational processes of the enterprise and the corresponding need for (big) data; (ii) The complexity of customer-based SLAs frequently encountered in large business engagements between service providers and consumers; and (iii) The increasingly complex composition and orchestration of services in inter-organizational service chains.

2.1 AI technologies and core operational efficiency

The role of AI to drive better and deeper customer experiences within digital processes is widely recognized today (cf. Martorelli and Stroud 2017), and AI technologies are used to actively optimize business processes in real-time (e.g., Veit et al. 2017). For instance, AI-generated insights may be used to dynamically assign human jobs through algorithms and tracking data (Lee et al. 2015); for enabling interactive visual exploratory data analysis applications that can help service managers reveal the low-level root causes of high-level business phenomena (Caron and Daniels 2008); to control automated industrial processes (Pasic et al. 2019); or to dynamically route ships for fuel efficiency on a real-time basis (Beşikçi et al. 2016), to just give a few examples.

In typical enterprise settings, core operational IT systems include business process management (BPM) and workflow management systems (WfMS), enterprise resource planning (ERP) systems, application performance monitoring (APM) solutions, and custom-built applications, to name a few. Such core operational IT systems provide the abundance of data that is needed to train machine learning (ML) models and prescriptive analytics (optimization and recommendation models) that control decision points and customization options in intelligent workflows. The data feeding into ML models may be structured or unstructured, and often includes KPIs that monitor business processes in the enterprise (cf. e.g., Márquez-Chamorro et al. 2018; Pérez-Álvarez et al. 2018). In some cases, KPIs represent cross-organizational (Wetzstein et al. 2009) or inter-organizational (Krathu et al. 2013) measures. In the remainder of this paper, we refer to KPIs that inform such AI-enabled business processes and whose semantics are defined on an (intra-, inter- or cross-)organizational or departmental level as *operational KPIs*. Note that even though operational KPIs may be evaluated for individual clients, by definition their *semantics are consistent across different clients*, and therefore not client-specific. This contrasts with individually negotiated service level indicators (SLIs) in customer-based SLAs, as described in the following section.

2.2 Complexity of bilateral SLAs in large business engagements

Service level management (SLM) is a critical part of service delivery and modern frameworks for SLM emphasize the need to customize SLAs for individual clients. For example, the ITIL 4 framework for IT service management (ITSM) emphasizes that SLM should *"[f]ocus on outcomes for the service consumer organization and on user experience more than on technical details and associated metrics"* (Agutter 2020). (Conger et al. 2008) add that ITSM *"focuses on defining, managing, and* *delivering IT services to support business goals and customer needs, usually in IT operations"*. Given this need for customization, SLAs of large service consumer/provider engagements are typically extensively negotiated between the collaborating parties on a per-deal basis (MarketsAndMarkets.com 2017, p. 18) and require extensive bilateral human interaction (cf. Butler et al. 2011; Wieder 2006; Megahed et al. 2020). The resulting complex, natural-language *bilateral SLAs* (a.k.a. customer-based SLAs Comuzzi et al. 2013) are highly customized and stand in stark contrast to the universality of well-known examples of general, published SLAs for public cloud services which are essentially unilaterally imposed (*"take it or leave it"*).

Service level indicators (SLIs) are carefully defined quantitative measures of some aspect of the level of service that is provided (Beyer et al. 2016, pp. 37–40). It is understood that any pair of similar, but customer-specific SLIs of different engagements may differ with respect to targeted service levels. However, an often overlooked, but significant property of bilateral SLAs is that otherwise similar SLIs often differ in their *semantics* in the context of different engagements. In particular, customer-specific SLIs and their constituent metrics may semantically differ from the 'bulk' monitoring data that is collected on the organizational (i.e., cross-client) level and aggregated into operational KPIs (cf. Sect. 2.1). Such semantic differences may include:

- Customer-specific rules for measurement, (algebraic) computation, time windows for validity or guarantee intervals, or units of measurement (cf. Longo et al. 2018)
- Customer-specific categorization/classification systems for datapoints (cf. Engel et al. 2018)
- Customer-specific inclusion and exclusion rules
- Customer-specific SLIs that are truly unique to a customer's SLA (e.g., "percentage of minutes of meetings emailed within 3 business days")

For instance, a frequently cited case study conducted at the Eindhoven University of Technology on the elicitation and specification of SLA terms regarding IT services for student notebooks, including incident management, defines *incident priorities* based on a distinction between the different types and different periods of usage, such as lectures, tutorials, self-study and examinations (Trienekens et al. 2004). This specific definition of incident priorities then becomes part of the terms of the overall SLA by its employment in corresponding definitions of metrics and SLIs/SLOs related to incident resolution times. In other words, the SLA terms are highly specific to both customer and the type of market in this example, and their definitions most likely semantically differ from typical generic operational KPIs used by e.g., an IT outsourcing provider. Another example for similar, but semantically different SLIs for *service availability* is given in Longo et al. (2018) and reproduced in Table 1. Further examples are given in the scenarios depicted in Sect. 3.

Organizations which sign into significant numbers of such complex, bilateral SLAs employ dedicated SLA management approaches to deal with the resulting management complexity (Sfondrini et al. 2015). Current SLA management approaches, which have

Table 1 Exemplary semantic differences in service availability SLIs (reproduced from Longo et al. 2018, p. 180)

Service	SLI	SLO	Unit	Guarantee	Validity	Observation	Measurement	Reporting
Service #1	Service availability	< 99.9	%	8:30–13:30 14:30–18:30	01/01/2017 31/12/2017	Yearly	30s	By the 15th of the following Measurement interval
Service #2	Service uptime	< 98	%	24 × 7	01/01/2017 30/06/2017	Monthly	30s	By the 15th of the following Measurement interval
Service #3	Service availability	< 98	%	8:30–13:30 14:30–18:30	01/01/2017 30/06/2017	Monthly	30s	By the 15th of the following Measurement interval

been well studied in the current literature (cf. Mubeen et al. 2017), provide formalisms to define generic constructs like metrics, SLIs and SLOs, classification systems (cf. Engel et al. 2018), inclusion/exclusion rules, penalties/rewards (cf. García et al. 2017; Muller et al. 2018), etc. to model the natural language SLAs. The objective is to represent the SLAs as concisely and accurately as possible while enabling their utmost automation in terms of monitoring and management. We refer to such efforts of managing bilateral SLAs across multiple partner engagements as *multi-engagement* SLA management. In many cases, organizations not only aim at achieving and monitoring SLA compliance (cf. Müller et al. 2014), but also at avoiding over-fulfillment of SLOs to proactively contain costs in the context of overall quality management strategies. Moreover, in some cases different conflicting strategies with regard to quality management are weighed against each other, e.g., to minimize or maximize (intra-engagement or cross-engagement) compensations (i.e., penalties or rewards; Muller et al. 2018).

2.3 Inter-organizational service chains

Orthogonal to the trend towards using AI technologies for informing and controlling core business processes, organizations employ increasingly complex orchestrations of both internal and external (e.g., outsourced) services in *service chains* (cf. Cho et al. 2012; Baltacioglu et al. 2007) to provide higher-level services to their customers. In the traditional industrial landscape, by effectively managing a supply chain, firms can benefit from reduced costs, boosted revenues, increased customer satisfaction, improvements in delivery and product or service quality (Baltacioglu et al. 2007). However, because of a higher degree of substitution, perishability and non-trivial over-capacity cost, service chains behave substantially differently than physical goods supply chains (Prasad and Shankar 2018). The notion of a service chain has been analyzed mainly from business-oriented perspectives and has been specifically developed in the field of management for logistics (Zhong et al. 2020), goods (Fernández et al. 2015), healthcare (Baltacioglu et al. 2007) and finance (Hofmann et al. 2017) domains. However, most existing approaches do not deal with the intricacies of the integration challenges that information systems face to orchestrate the behavior of the chain participants. Those aspects become crucial in software intensive domains. For example, in the well-known cloud computing domain, the software as a service (SaaS) paradigm relies on a supporting service chain composed by two other levels defined in the cloud computing model: platform as a service (PaaS) and infrastructure as a service (IaaS), whereas these different levels are often, if not typically, provided by different organizations. Given the reliance on real-time monitoring data to drive corresponding AI-enabled business processes in the back-end, such service chains require deep integration of business processes and information systems across organizational boundaries. For example, predictions regarding anticipated usage spikes generated by an application on the SaaS level managed by a particular organization may be used for pro-active elastic (cf. Muñoz-Escoí and Bernabéu-Aubán 2017) (and potentially SLA-aware) resource provisioning tasks on the PaaS and IaaS levels managed by another organization.

3 Motivating scenarios

In this section, we present two scenarios derived from real use cases in industry and public administration. These scenarios serve as running examples and for motivating the vision presented in this paper. Scenario 1, presented in Sect. 3.1, is derived from a typical large-scale client engagement of a leading global IT service provider (cf. Megahed et al. 2020) in combination with modern AI-enabled approaches to incident management (cf. Lerner 2017; Masood and Hashmi 2019; Levin et al. 2019; Chen et al. 2020). Scenario 2, depicted in Sect. 3.2, is derived from a real-world deployment of the *Governify* framework described in Gamez-Diaz et al. (2019) within a European public administration entity. Both scenarios have been specifically selected to highlight and exemplify the key aspects of the research challenges presented in this paper. Finally, Sect. 3.3 contrasts the two scenarios regarding the implications on SLA management complexity in their corresponding service chains.

3.1 Scenario 1: large outsourced IT service management provider

Company A, a large IT Service Management (ITSM) provider, delivers IT services to a number of large enterprise clients (i.e., its customers). We denote the set of customers by I and each customer would be $C_i \in I$, where $i \in \{1, \dots, |I|\}$ and $|I|$ is the cardinality of set I. Services rendered include provisioning of IT resources, incident management, change management, etc. The size of the majority of the contracts is large (i.e., multi-million dollar range). For the purposes of this scenario, in the following we focus on the aspect of *incident management* at A—both with respect to incidents occurring within A and subsequently affecting one or more customers C_i (e.g., cloud service outage) as well as incidents whose origin may lie within operations of a customer C_i (e.g., network router outage at a client's site).

When new incidents occur, they are generally assigned to, and handled by, internal incident response teams (1st level support or 'helpdesk' services); however, in order to handle seasonal peaks in incident prevalence across its customer base, A outsources part of its 1st level support services to incident response teams at outsourcing provider P. This is done in a transparent manner regarding customers C_i. In other words, the service provided by the external incident response teams is essentially the same as the service that is provided by the internal teams. However, various performance-related KPIs for the outsourced teams may differ from the internal ones (e.g., average *incident resolution time*).

For improving operational efficiency, A has adopted an *AIOps* approach for its operations: *"AIOps platforms utilize big data, modern machine learning and other advanced analytics technologies to directly and indirectly enhance IT operations (monitoring, automation and service desk) functions with proactive, personal and dynamic insight."* (cf. Lerner 2017; Levin et al. 2019). In particular, A's incident management processes have been heavily automated using AI techniques such that initial incident categorization, triaging and assignment *("routing")* to either internal teams at A or outsourced teams at P is largely automated based on evidence found in structured and unstructured data from corresponding tickets (cf. Frick et al.

Table 2 Example of an incident severity classification matrix used by some customer C_i

Incident priority classes matrix	Urgency			
	1	2	3	4
Impact				
A	P1	P1	P1	P1
B	P1	P1	P2	P3
C	P3	P3	P3	P3
D	P3	P3	P3	P3

2019). The AIOps solution at A takes many factors into account for making routing decisions. Among these factors is A's *internal* notion of incident severity that defines four different severity levels ("P1", "P2", "P3", or "P4"). All corresponding internal KPIs (e.g., *incident resolution time*) are organized around the notion of these four categories. However, due to the large sizes of the engagements between companies A and C_i as well as A and P, respectively, the corresponding SLAs in all of the aforementioned business engagements are the result of extensive negotiations during contract closure, and are correspondingly complex. Consequently, A employs a dedicated SLA management system using some modern SLA formalism from current literature (e.g., L-USDL or ysla; García et al. 2017; Engel et al. 2018) for management and monitoring of these SLAs. In particular, the custom SLAs reflect the specific notions of *incident severity*[1] for each of the customers C_i and at P:

- Customers C_i all have their own notion of incident priority. For instance, a given customer uses an arbitrary matrix relating notions of *urgency* and *impact* of an incident to an assigned severity, as shown in Table 2. For example, high-urgency *(1)* and medium-impact *(B)* incidents are assigned to high incident severity *(P1)*. Another customer only distinguishes two urgency levels ("Urgent", "Normal") instead of four different levels (1, 2, 3, 4); yet another customer uses another system for incident severity classification that factors in *risk of event re-occurrence* in addition to urgency and impact.
- Outsourcing provider P manages incidents by classifying them into one of "low severity", "medium severity", and "high severity".

Since A and any of C_i have different notions of incident severity and different classification criteria, the customer-specific SLIs for *incident resolution time* defined in the corresponding formal SLA documents are based on custom metrics using the arbitrary priority classes of each customer C_i rather than the four severity classes used by the internal monitoring systems of A.[2] Moreover, some of the constituent

[1] The reader is referred to Sect. 2.2 for real-world examples of arbitrary notions of incident severity documented in current literature.

[2] See e.g., Engel et al. (2018) for an example of how different incident classes can be modeled in a formal SLA representation.

metrics of these customer-specific SLIs involve notions of arbitrary concepts such as *urgency* or *risk of event re-occurrence*. As a result, at *A* the SLIs for *incident resolution time* all have different semantics than the internal KPI for *incident resolution time*; in other words, the semantics depend on whether this metric is regarded in the context of *A*'s internal monitoring efforts or in the context of SLA compliance regarding any particular customer C_i. In the case of SLO misses (i.e., SLA non-compliance), *A* is responsible for paying penalties to its customers.

3.2 Scenario 2: large service aggregator

A public administration wishes to implement and evolve a citizen application (app) to act as a single entry point aggregating various different services provided by multiple and diverse departments. In such a context, a service governance committee is established to design and operate an application programmable interface (API) gateway that would act as a mediator between the citizen app and the different micro-services deployed in the departments' infrastructures with different scalability models.

In order to provide the desired service levels to their citizens, the governance committee wishes to establish a minimal SLA that all provided services should fulfill. The minimal SLA is composed of two SLIs: *monthly availability* and *daily average response time*. The SLIs are characterized by two activity periods during the day: *high activity (9:00-22:00)* and *low activity (22:00-9:00)*. Consequently, each department should establish a specific SLA with the governance committee. In such a context, multiple complications can arise:

- The need for a common understanding of *monthly availability* and *daily response time* and how to measure it. Since multiple infrastructures come into play, the monitoring should be integrated and homogenized with an accurate definition.
- As is common in micro-service architectures, a significant set of services provided rely on other services such that a graph of dependencies is established. In such a context, the capability for analyzing and reasoning about the root cause of a given behavior observed could be crucial.
- The global scalability of the system as a whole should be harmonized and consistent among the different departments' infrastructures. To address this challenge, it would require a shared predictive model joint with an explicit and dynamic set of expectations for each infrastructure that would drive distributed scalability operations.

3.3 Service chain complexity

The scenarios presented above correspond to focused views of potentially large service chains that span across organizations and/or departments and highlight two different types of complexity growth (as depicted in Fig. 1):

On the one hand, Scenario 1 motivates a dynamic growth of the chain towards the service consumer based on the different customer needs that define different customer-based SLAs, while having a common supporting SLA from an outsourcing

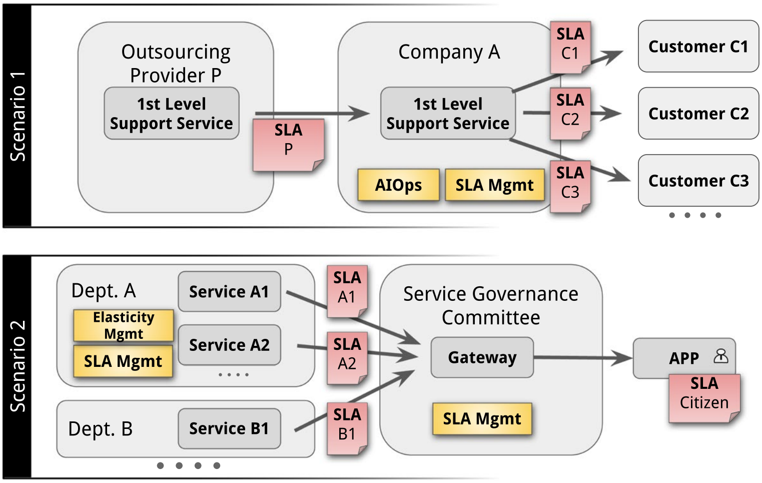

Fig. 1 Service chains represented by our scenarios

service provider. In contrast, Scenario 2, motivates the need of growth towards the service providers having a common (i.e., service-based) SLA on the side of the service consumer (i.e., the citizens). Moreover, the two scenarios exemplify different organizational boundaries that can appear. In Scenario 1, all organizations correspond to different stakeholders in a potential industrial market. In Scenario 2, a single organization (the public administration) structures its services in different internal departments and exposes those services externally as a *one-stop shop* (Wimmer and Tambouris 2002) to service consumers. In this context, it is important to highlight that real service chains could have a combination of both types of complexity growth and, hence, the need for advanced SLA analytics would be further amplified.

4 Research challenges

The two scenarios described in Sect. 3 highlight, motivate, and inspire the need to provide operational processes with sufficient information from SLA management: For achieving optimal operational efficiency, decision making in intelligent workflows must not only account for the optimization of internal processes with respect to operational KPIs, but also take into account downstream implications regarding actual or predicted SLA compliance and/or over-fulfillment under multi-engagement SLA awareness.

In the following, we identify four research challenges that are motivated specifically by the *dual role of AI* in the context of the problem at hand: On the one hand, increasingly data-driven, AI-enabled service chains represent a significant reason why as much relevant information as possible needs to be provided from multi-engagement SLA management systems to core operational information systems (cf. Research Challenge 1). On the other hand, AI technologies are a necessary component of modern SLA management systems to provide predictive and prescriptive information for forward-looking operational decision-making. However, this comes with its own set of challenges regarding multi-engagement SLA awareness (cf. Research Challenge 2). Because of this dual role of AI, achieving optimal operational efficiency requires deep integration and bidirectional flow of information between AI models in both domains (cf. Research Challenge 3). Finally, we discuss the benefits and implications of applying pervasive AI-enabled SLA management across inter-organizational service chains, including data confidentiality concerns (cf. Research Challenge 4).

4.1 Research challenge 1: fine-granular semantic alignment between SLIs and operational KPIs

To inform decision making in core operational processes, actual insights from SLA analytics need to be made accessible to the systems managing those processes. Rather than (only) top-level SLA evaluation results, these insights need to include detailed intermediate (historic or real-time) computation results (e.g., low-level metric values feeding into some aggregated SLI) to reason on the level of root causes, rather than symptoms, of high-level phenomena on the SLI/KPI level. Because such intermediate computation results can vary in semantics on all aggregation levels (e.g., employed classification systems, measurement periods, units of measurement, inclusion/exclusion rules, etc.; cf. Sect. 2.2), this underscores the need to close the semantic gaps between the specific SLIs from SLAs and the operational KPIs in a fine-granular and non-trivial manner. Specifically, given such fine-grained semantic differences between KPIs and SLIs as described above, it is necessary to devise methods that enable the translation of these different semantics into a single global reference space (e.g., SLIs need to be aligned with a 'reference KPI') such that they can be subsequently used to inform the numerous constraints and objective functions in complex multi-objective optimization problems that model the notion of operational efficiency on a global level. Notably, the challenge extends beyond that of 'traditional' schema matching and mapping: Differing semantics on different aggregation levels need not only be 'matched', but reconciled in a way that allows to effectively reason on how different values compare in the context of different time frames, units of measurement, scopes, etc. While the problem of data availability in a broader sense has been extensively studied in the current literature on schema matching and ontology matching (e.g., Shvaiko and Euzenat 2005; Noy 2004; Bernstein et al. 2011), the more specific problem of mapping SLIs to KPIs in a fine-granular manner and including the ability to reason on the differences between intermediate computation results in different stages of aggregation has been largely overlooked so far, except for some works in early stages (cf. Longo et al. 2018) or in the context of specific use cases (e.g., Bellini et al. 2018).

For instance, in Scenario 1, the differences in classification systems for incident severity that exist between *A* and its customers and providers regarding the semantic definition for SLI *incident resolution time* poses a major challenge for the AIOps solution at *A* that attempts to proactively manage incidents according

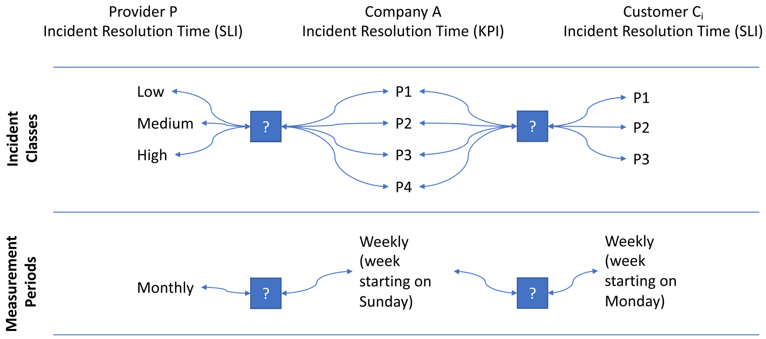
Provider P
Incident Resolution Time (SLI)
Company A
Incident Resolution Time (KPI)
Customer C_i
Incident Resolution Time (SLI)
Incident Classes
Low
Medium
High
?
P1
P2
P3
P4
?
P1
P2
P3
Measurement Periods
Monthly
?
Weekly (week starting on Sunday)
?
Weekly (week starting on Monday)

Fig. 2 Challenges arising in Scenario 1 regarding the alignment of SLIs and KPIs used by A, P and an exemplary customer C_i

to A's own notion of incident severity. When it comes to *precisely meeting* (i.e., neither failing nor over-fulfilling) service level objectives for *incident resolution time* defined in the SLA documents between A and any customer C_i, the employed AIOps approach is unable to resolve the semantics of that SLI, and therefore unable to proactively manage SLA compliance (e.g., when triaging and ranking incident response resources across the entire customer base to minimize global SLA penalties). Analogously, due to the different notion of incident severity at P, assignment of incidents to either internal (A) or external resources (P) requires custom coding and mapping of incidents due to the semantic heterogeneity. To better enable the AIOps solution to efficiently manage incidents, it would need to 'understand' how to map the classification systems for incidents from the SLI definitions to the general KPI definition. In addition to the different classification systems used, other semantic differences may add difficulty to the task, such as unaligned measurement periods. These KPI/SLI alignment challenges arising from Scenario 1 are illustrated in Fig. 2.

For the challenge at hand, we suggest to represent the formal semantics of both SLIs and KPIs as directed relational networks where the entities are constituent metrics, classification systems, algebraic expressions, thresholds, etc. Hence, we hypothesize that the mapping problem may be approached by regarding it as an instance of the class of *entity alignment for knowledge graphs* (e.g., ontology matching; Euzenat et al. 2007) problems. As such, it may be potentially solved with techniques involving varying degrees of automation:

- Regarding mostly automated approaches, a recent benchmarking study underscores the high efficacy of *embedding-based entity alignment* techniques (e.g., Sun et al. 2020), and corresponding techniques have been developed to map relationships between entities from one domain to another (e.g., Chen et al. 2017). Other works have employed traditional network analysis techniques to relate different KPIs and their constituent components to each other (e.g., Krathu et al. 2015). Similar approaches may prove useful for mapping SLIs to KPIs, and vice versa.

- Regarding semi-automated approaches, prior works have focused on modeling SLAs in OWL and applying Logic Programming (i.e., Prolog) to reason about different semantic definitions of SLIs among different SLAs with the goal of facilitating their comparison (Longo et al. 2018). Such approaches may be potentially extended to inform and aid semi-automated methods for mapping SLIs to KPIs.
- Regarding manual approaches, extending current SLA formalisms with new constructs (e.g., annotations) that allow for manually provided mapping 'hints' may provide a certain degree of inter-operability between SLA analytics and operational systems. The approach followed in Estrada-Torres et al. (2019) for measuring performance in knowledge-intensive processes may be a suitable starting point for such an effort. However, the usual drawbacks of manual methods apply, such as the challenge that dynamic changes to either SLIs or KPIs would require manual reconfiguration.

4.2 Research challenge 2: integration of predictive SLA analytics with AI-enabled operational decision-making

In addition to historic and real-time fine-granular monitoring data, it is necessary to provide operational processes with predictions of future values, (non-)compliance events, looming penalties, etc., to be able to prevent undesired outcomes by taking appropriate actions. Several approaches for predictive SLA analytics and/or corresponding SLA violation prevention have been proposed in the current literature (e.g., Leitner et al. 2013; Márquez-Chamorro et al. 2017; Nawaz et al. 2018). In terms of generic approaches that are reusable across different domains, typically there is some mechanism for predicting the future state of SLIs/SLOs (e.g., time series analysis, machine learning, hand-written rules by domain experts, etc.) that is used in conjunction with some formalisms for specifying remedial or preventive actions to be taken when such a prediction suggests that an SLA violation is imminent or at least probable. However, such *policy-based approaches* (cf. Sloman 1994) exhibit two limitations that we explain below.

Firstly, as has been recognized in the current literature (e.g., Faniyi and Bahsoon 2015), they become infeasible when the inter-dependencies between actions and other actions (such as common in multi-engagement, multi-SLA settings), as well as global (optimal) outcomes compared to local (optimal) outcomes become too complex. In other words, in complex settings with multiple engagements and corresponding independent SLAs, multiple 'independent' control loops on the basis of SLI predictions may interfere with each other in an uncontrolled manner. The underlying optimization problem(s) we refer to here is related to the decision making of the operational aspects of the underlying system; in particular, to optimize the adherence to SLA requirements under the constraints of feasibility of such actions and the seamless orchestration of the different parts of the system. Solving the whole problem as one optimization problem to find the 'global' optimal solution is often computationally intractable for realistic-sized instances. Nevertheless, our problem here has a particular structure that has been studied in the operations research literature and efficient solution procedures/algorithms have been successfully proposed for solving it. The structure we are referring to can be described as follows: We have multiple optimization problems that are independent/separable except for linking variable(s), i.e., decision variables/actions that exist in many/all problems and cause them to have some overlap, leading to the aforementioned problem; a local optimal solution for each problem separately is not necessarily the global optimal solution for the whole problem. Rather than solving one considerable intractable global problem, some approaches have been proposed to use this structure and solve the problem more efficiently. The most notable of these approaches is the 'Dantzig-Wolf decomposition approach/algorithm (cf. Vanderbeck and Savelsbergh 2006). It has been successfully applied to multiple domains and applications, such as production planning (e.g., Wu et al. 2020), stochastic capacity planning problems (e.g., Singh et al. 2009), and logistics/routing problems (e.g., Petersen et al. 2008). To the best of our knowledge, it has not been applied to a SLA context like our research challenge here, and thus, there is a clear interesting research direction here of applying it to this challenge.

Secondly, with the currently available SLA formalisms, it is difficult to model feedback loops, i.e., a situation where the outcome of an action can indirectly influence a corresponding predictor. This problem, again, resembles a class of problems in the stochastic operations research literature, sometimes referred to as 'stochastic programs with decision-dependent uncertainty' (e.g., Goel and Grossmann 2006). Another opportunity here, therefore, is applying some of the advances in this area to solve this second limitation.

Consequently, assuming that the semantic heterogeneity between different SLIs and KPIs can be overcome as described in Research Challenge 1 (cf. Sect. 4.1), the question arises as to how one can provide a general framework and/or formalism for SLA management that facilitates the identification and provision of relevant predictions for SLIs from all available data (i.e., including both SLIs and metrics local to the SLA at hand, but also SLIs from other SLAs as well as operational KPIs) under consideration of the expected downstream impact of subsequent actions on the operational layer reflected in operational KPIs and the predictions of the same, or other, SLIs. Moreover, such a general framework or formalism for SLI prediction should provide a means for integrating contextual information in model learning (e.g., Márquez-Chamorro et al. 2020). This holds especially true in service chains, where each 'link/service' has its own context.

For instance, in the context of Scenario 2 it may be desirable to establish SLA-aware coordinated elasticity management (cf. Müller et al. 2016; Muñoz-Escoí and Bernabéu-Aubán 2017) such that the elastic models and rules of each service could be aligned to optimize a more fluid global operation. We could assume an elastic behavior in each component that analyzes the overall operation and allocates the optimal amount of resources to provide a given service at a point in time. For example, to drive SLA-driven elasticity of service infrastructures, a rule-based framework to orchestrate the service delivery as proposed in Müller et al. (2016) could be employed. In this case, predictions on SLA compliance regarding service availability inform the elasticity management component. However, for the case where resources are shared among different departments and providing different services (such as, for instance, a centralized database, storage system or network backbone), a more sophisticated strategy for resolving these inter-dependencies across resources and engagement-specific SLAs may need to be modeled due to the possible interference of prescribed actions stemming from different engagement-specific policies.

Analogous to the above described example regarding service elasticity, in Scenario 1, the AIOps solution at *A* makes decisions about triaging and routing of incidents. In this case, human-staffed incident response teams represent the resources that are to be 'elastically' managed. In the context of such human-supported services, an elastic and dynamic model that takes into account the SLAs to create and manage the team commitments that will deliver the service as proposed in Fernández et al. (2015), could be employed. If the deployment of team commitments, however, has side effects such as altering the predictions of outcomes for different tasks of involved team members, then the currently available formalisms for modeling those inter-dependencies quickly reach their limits.

4.3 Research challenge 3: end-to-end prescriptive SLA analytics

First, we note that given the resemblance between service chains and physical supply chains, there is a research direction/opportunity to leverage the comprehensive literature on physical supply chain planning and optimization when prescribing actions in service chains. Optimization and operations research work has been done for physical supply chains management and planning for decades (cf. Huan et al. 2004; Ayyildiz and Gumus 2021; Kouvelis et al. 2006; Hassini 2008; Stadtler 2008; Chopra et al. 2013). Additionally, relevant work done on such systems under uncertainty has picked attention more recently (cf. Peidro et al. 2009; Morteza and Kuan 2012; Lee 2014).

In addition, recent multidisciplinary research results at the interface of the fields of operations research and machine learning strongly suggest that the flow of information between SLA management systems and operational processes needs to be bidirectional: When predicting SLA compliance and/or related time series, it is important to train the corresponding ML models for prediction in a way that takes into account the characteristics of the actual task control and planning problem that is subsequently being addressed on the operational level (cf. Donti et al. 2017). In other words, instead of solving the prediction problem separately and using its output independently in a task optimization model that recommends the optimal actions, the two problems should be solved simultaneously with the goal of finding and focusing on the relevant data to a task optimization model that does the recommendation of the optimal actions.

For instance, let us consider how different routing decisions by the AIOps solution in Scenario 1 may affect the outcome of SLA compliance and the corresponding penalties to be paid by *A*. The 'traditional' approach to optimizing those decisions would be to use predictions regarding SLA compliance stemming from ML models that have been simply trained for overall prediction accuracy. In other words, the training objective of these models would be to maximize the number of accurately predicted SLA compliance events. Let us assume that the actual optimization objective of the AIOps solution at A is to minimize overall operational cost for A, including (but not limited to) the accumulated cost from penalties. Let us further assume that there are ten different potential SLA non-compliance events with associated penalties, where the penalty of one (the first) of them is more expensive than the penalties of the other nine non-compliance events combined. After model training *solely* with the objective of maximizing the average prediction accuracy, a trained ML model $Model_1$ might—by chance - not correctly predict the first (expensive) event, but correctly predict the remaining nine (less expensive) events. Such $Model_1$ would show high overall prediction accuracy (90%). Now let us consider another model $Model_2$ that predicts the first event correctly, but fails to predict the other nine events. While $Model_2$ has only 10% overall prediction accuracy, due to the uneven distribution of penalties among the non-compliance events, it would actually perform better in the optimization context where the ultimate goal is to keep total costs low. Clearly, $Model_2$ should be used instead of $Model_1$, but it can only be selected as the 'better model' by the SLA analytics system if it 'knew' that the goal is to keep the cost down, and how much cost the different SLA non-compliance events would cause, during the training phase of the prediction model. In other words, the SLA analytics system needs to receive information from the AIOps solution about the goals of the underlying optimization task, as well as any data points that are necessary to appropriately weigh the non-compliance events for model training (e.g., current exchange rates if some penalties are paid in foreign currency).

We argue that to enable such end-to-end learning between operational processes and SLA analytics, there is a need to devise general frameworks that facilitate the bidirectional exchange of the information necessary to train predictive models on the SLA analytics level under consideration of the specifics of the subsequent optimization tasks on the operational level. For example, a corresponding framework could provide a means for supplying parameters and data points to SLA analytics systems for adjusting loss functions used during the training of predictive models. However, it should be noted that the above-described approach of end-to-end learning is an active area of research, and significantly more sophisticated approaches can and should be considered (e.g., Elmachtoub et al. 2020; El Balghiti et al. 2019).

4.4 Research challenge 4: federation of SLA analytics across service chains

We envision that hypothetical solutions to Research Challenges 2 and 3 (cf. Sects. 4.2 and 4.3) would allow organizations to jointly optimize local (or locally controlled inter-organizational) operational processes while considering overall *organization-local* multi-engagement SLA compliance. However, in today's complex inter-organizational service chains, the scope of such optimization efforts may be expanded across organizational boundaries with the goal of optimizing the entire service chain with these objectives. While the topic of digital supply chain integration has been well studied in current literature (e.g., Korpela et al. 2017), to the best of our knowledge prior works in this field have not investigated deep integration of digital supply chains with current SLA management approaches (as well as prescriptive SLA analytics according to Research Challenge 3), and in particular their federation. This has the potential to allow for the simultaneous optimization of inter-organizational operational processes considering the overall multi-engagement SLA compliance across the service chain. Hence, we suggest that suppliers who are situated 'deeper' in a service chain could help enable better prescriptive SLA analytics in the subsequent stages of the chain by providing them with insights from their own predictive SLA analytics models. For instance, if we consider again the example of service elasticity in the context of Scenario 2, as service chains can grow and span departments and even organizations, it may be necessary to coordinate elasticity across the entire service chain while considering the set of SLAs involved. Since SLA management systems may be distributed across the service chain, their respective insights from SLA analytics need to be shared across the service chain as well.

This proposed federation model for SLA analytics may be particularly feasible and useful in situations where multiple departments of a larger, decentralized organization operate independent SLA management systems, but work towards a common goal (e.g., in the context of *one-stop shop* system designs). For example, in Scenario 2, the SLA management system of *Dept. A* may realize that *Service A*2 is about to miss the *monthly availability* SLO at a particular point in time in the future. By propagating this insight to the SLA management system at the service governance committee, the ability of the respective predictive SLA analytics models to predict compliance levels for the citizen SLA at that point in the service chain might increase significantly.

Taking this idea even further, it may prove useful to not only share operational data and SLA analytics *insights* across organizations, but also predictive/prescriptive ML *models* (or aspects thereof) themselves. In other words, sharing of 'experience' on the level of ML models regarding the ability to predict SLA outcomes could further be leveraged to improve the overall business outcome of a service chain. For instance, some operational KPIs relate to objects or processes that are shared across entire supply chains (cf. Wetzstein et al. 2009) rather than local operational processes (e.g., time series representing the trend of the temperature of refrigerated goods over time). In these settings, SLA management systems at any point in the service chain may benefit from knowledge from their collaboration partners in the form of ML models on how to predict future values of such time series such that their own predictive/prescriptive analytics models can be improved. In situations where either actual data points or ML models[3] should not be shared across organizations for confidentiality reasons, privacy-preserving federated learning technologies as proposed in Truex et al. (2019) can help protect confidential business information while using data from multiple different organizations for model training.

As an example, in the context of Scenario 2 a predictive model could be trained in each device running the citizen app, so the system could predict the expected behavior of a given user; those models could be then aggregated in an anonymized fashion to produce a general predictive model of the expected load for the whole set of citizens in a certain period. Consequently, the load prediction model could then be shared across the chain using federated learning technologies for guiding

[3] ML models may inadvertently leak detailed information about training data (cf. Nasr et al. 2019).

the elasticity rules that will be used to govern the service deployments so that they can be ready to attend and react to the expected loads—without compromising the privacy of individual citizens. Note that the aforementioned example in the context of Scenario 2 can be regarded as a variation of the IaaS/PaaS/SaaS coordination problem for SLA-aware elasticity management introduced in Sect. 2.3: The different layers of the service chain are managed by different organizations and require shared knowledge about anticipated usage spikes on the SaaS level to enable efficient and effective SLA-aware resource elasticity management on the PaaS and IaaS levels.

5 Conclusion

In this paper, we presented our vision of supplying online AI-supported SLA analytics to the core workflows of organizations involved in complex AI-enabled service chains with highly customized, bilateral SLAs. We exemplified the need for improving operational efficiency under multi-engagement SLA awareness by describing two scenarios derived from real use cases in industry and public administration corresponding to both human-supported services and software services. We presented four specific research challenges, motivated by the dual role of AI technologies in the problem at hand, and grounded in the current state of the art of AI techniques and SLA management: (i) Devising methods for fine-grained semantic alignment between SLIs and operational KPIs to have a consistent description of metrics across different SLAs and operational systems in complex multi-engagement scenarios; (ii) Developing general frameworks for providing predictive insights from SLA analytics to core operational processes under multi-engagement SLA conditions; (iii) Establishing bi-directional flow of information for enabling truly prescriptive SLA analytics based on ML models that are trained with the actual planning and control tasks in mind; and (iv) Enabling the federation of SLA analytics across inter-organizational service chains to enable optimization on a supply chain-level rather than on an intra-organizational level. These four research challenges represent increasingly comprehensive approaches to AI-enabled operational process optimization under multi-engagement SLA awareness. While the latter three of these research challenges can be tackled independently, the first (i.e., semantic alignment of SLIs with operational KPIs) is foundational to the latter three, and, hence, needs to be the starting point for realizing the vision presented in this paper. In addition, it could be interesting to analyze how the potential environmental conditions in a real setting (different regulations, cultural challenges, etc.) have an impact on the vision; as an example, privacy regulations across regions would affect how an international service chain would implement the flow of sensitive data across organizational boundaries.

It is important to highlight an intrinsic limitation derived from the visionary nature of the current work: the presented scenarios and challenges are derived from an argumentative analysis and our past subjective experience over real scenarios. From this perspective, there is no certainty that our arguments prove to be valid in all practical situations or scenarios. In contrast, those challenges are to be seen as exploration areas and our discussion over them as potential hints to find appropriate solutions or frameworks to address the challenges. Nevertheless, we envision that novel technical and organizational approaches stemming from the aforementioned research challenges can potentially enable a comprehensive global view of the trade-offs between overall operational efficiency across inter-organizational service chains and on-point SLA compliance across multiple individual partner engagements with heavily customized SLAs. Such a comprehensive global view could in turn lead to unprecedented levels of operational efficiency in such dynamically managed, complex business environments.

References

Agutter C (2020) ITIL® 4 Essentials: your essential guide for the ITIL 4 Foundation exam and beyond. IT Governance Ltd

Ayyildiz E, Gumus AT (2021) Interval-valued pythagorean fuzzy ahp method-based supply chain performance evaluation by a new extension of scor model: Scor 4.0. Complex Intell Syst 7(1):559–576

Baltacioglu T, Ada E, Kaplan MD, Yurt O, Kaplan YC (2007) A new framework for service supply chains. Serv Ind J 27(2):105–124

Bellini P, Bruno I, Cenni D, Nesi P (2018) Managing cloud via smart cloud engine and knowledge base. Future Gener Comput Syst 78:142–154

Bernstein PA, Madhavan J, Rahm E (2011) Generic schema matching, ten years later. Proc VLDB Endow 4(11):695–701

Bertsimas D, Kallus N (2020) From predictive to prescriptive analytics. Manag Sci 66(3):1025–1044

Beşikçi EB, Arslan O, Turan O, Ölçer A (2016) An artificial neural network based decision support system for energy efficient ship operations. Comput Oper Res 66:393–401

Beyer B, Jones C, Petoff J, Murphy NR (2016) Site reliability engineering: how google runs production systems. O'Reilly, http://landing.google.com/sre/book.html

Butler J, Lambea J, Nolan M, Theilmann W, Torelli F, Yahyapour R, Chiasera A, Pistore M (2011) SLAs empowering services in the future internet. In: The future internet assembly. Springer, pp 327–338

Caron E, Daniels HA (2008) Explanation of exceptional values in multi-dimensional business databases. Eur J Oper Res 188(3):884–897

Chen M, Tian Y, Yang M, Zaniolo C (2017) Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In: Proceedings of the 26th international joint conference on artificial intelligence. AAAI Press, pp 1511–1517

Chen Z, Kang Y, Li L, Zhang X, Zhang H, Xu H, Zhou Y, Yang L, Sun J, Xu Z, Dang Y, Gao F, Zhao P, Qiao B, Lin Q, Zhang D, Lyu MR (2020) Towards intelligent incident management: Why we need it and how we make it. In: Proceedings of the 28th ACM Joint Meeting on European software engineering conference and symposium on the foundations of software engineering, Association for Computing Machinery, New York, NY, USA, ESEC/FSE 2020, pp 1487–1497. https://doi.org/10.1145/3368089.3417055

Cho DW, Lee YH, Ahn SH, Hwang MK (2012) A framework for measuring the performance of service supply chain management. Comput Ind Eng 62(3):801–818 (**soft Computing for Management Systems**)

Chopra S, Meindl P, Kalra DV (2013) Supply chain management: strategy, planning, and operation, vol 232. Pearson, Boston

Comuzzi M, Jacobs G, Grefen P (2013) Understanding SLA elements in cloud computing. In: Working conference on virtual enterprises. Springer, pp 385–392

Conger S, Winniford M, Erickson-Harris L (2008) Service management in operations. In: AMCIS 2008, p 362. https://aisel.aisnet.org/amcis2008/362

Donti P, Amos B, Kolter JZ (2017) Task-based end-to-end model learning in stochastic optimization. In: Advances in neural information processing systems, pp 5484–5494

El Balghiti O, Elmachtoub A, Grigas P, Tewari A (2019) Generalization bounds in the predict-then-optimize framework. In: Advances in neural information processing systems, pp 14389–14398

Elmachtoub AN, Liang JCN, McNellis R (2020) Decision trees for decision-making under the predict-then-optimize framework. arXiv:200300360

Engel R, Rajamoni S, Chen B, Ludwig H, Keller A (2018) ysla: reusable and configurable SLAs for large-scale SLA management. In: 2018 IEEE 4th international conference on collaboration and internet computing (CIC), pp 317–325

Estrada-Torres B, Richetti PHP, Del-Río-Ortega A, Baião FA, Resinas M, Santoro FM, Ruiz-Cortés A (2019) Measuring performance in knowledge-intensive processes. ACM Trans Internet Technol 19(1)

Euzenat J, Shvaiko P et al (2007) Ontology matching, vol 18. Springer

Faniyi F, Bahsoon R (2015) A systematic review of service level management in the cloud. ACM Comput Surv (CSUR) 48(3):1–27

Fernández E, Toledo CM, Galli MR, Salomone E, Chiotti O (2015) Agent-based monitoring service for management of disruptive events in supply chains. Comput Ind 70:89–101

Fernández P, Truong H, Dustdar S, Ruiz-Cortés A (2015) Programming elasticity and commitment in dynamic processes. IEEE Internet Comp 19(2):68–74

Forrester Research (2017) Artificial Intelligence Revitalizes BPM

Frick N, Brünker F, Ross B, Stieglitz S (2019) Der einsatz von künstlicher intelligenz zur verbesserung des incident managements (the utilization of artificial intelligence for improving incident management). HMD Praxis der Wirtschaftsinformatik 56(2):357–369

Gamez-Diaz A, Fernandez P, Ruiz-Cortés A (2019) Governify for apis: Sla-driven ecosystem for api governance. In: ESEC/FSE 2019, association for computing machinery, New York, NY, USA, pp 1120-1123. https://doi.org/10.1145/3338906.3341176

García JM, Fernández P, Pedrinaci C, Resinas M, Cardoso J, Ruiz-Cortés A (2017) Modeling service level agreements with linked USDL agreement. IEEE Trans Serv Comput 10(1):52–65

Goel V, Grossmann IE (2006) A class of stochastic programs with decision dependent uncertainty. Math program 108(2–3):355–394

Hassini E (2008) Supply chain optimization: current practices and overview of emerging research opportunities

Hofmann E, Strewe UM, Bosia N (2017) Supply chain finance and blockchain technology: the case of reverse securitisation. Springer

Huan SH, Sheoran SK, Wang G (2004) A review and analysis of supply chain operations reference (scor) model. Supply chain management: An international Journal

Korpela K, Hallikas J, Dahlberg T (2017) Digital supply chain transformation toward blockchain integration. In: 50th Hawaii international conference on system sciences

Kouvelis P, Chambers C, Wang H (2006) Supply chain management research and production and operations management: review, trends, and opportunities. Prod Oper Manag 15(3):449–469

Krathu W, Engel R, Pichler C, Zapletal M, Werthner H (2013) Identifying inter-organizational key performance indicators from edifact messages. In: 2013 IEEE 15th conference on business informatics. IEEE, pp 276–283

Krathu W, Pichler C, Xiao G, Werthner H, Neidhardt J, Zapletal M, Huemer C (2015) Inter-organizational success factors: a cause and effect model. Inf Syst e-Bus Manag 13(3):553–593

Lee JH (2014) Energy supply planning and supply chain optimization under uncertainty. J Process Control 24(2):323–331

Lee MK, Kusbit D, Metsky E, Dabbish L (2015) Working with machines: the impact of algorithmic and data-driven management on human workers. In: 33rd Annual ACM conference on human factors in computing systems, pp 1603–1612

Leitner P, Ferner J, Hummer W, Dustdar S (2013) Data-driven and automated prediction of service level agreement violations in service compositions. Distrib Parallel Databases 31(3):447–470

Lerner A (2017) AIOps Platforms. Available at https://blogs.gartner.com/andrew-lerner/2017/08/09/aiops-platforms/

Levin A, Garion S, Kolodner EK, Lorenz DH, Barabash K, Kugler M, McShane N (2019) Aiops for a cloud object storage service. In: 2019 IEEE international congress on big data (BigDataCongress). IEEE, pp 165–169

Longo A, Potena D, Storti E, Zappatore M, De Matteis A (2018) Comparing SLAs for cloud services: A model for reasoning. In: European conference on advances in databases and information systems. Springer, pp 178–190

MarketsAndMarketscom (2017) Service integration and management (SIAM) market—global forecast to 2021

Martorelli B, Stroud R (2017) Comprehensive services integration needs more than just conventional siam. Forrester Research Inc, Tech. rep

Masood A, Hashmi A (2019) AIOps: predictive analytics & machine learning in operations. Apress, Berkeley, pp 359–382

Megahed A, Nakamura T, Smith M, Asthana S, Rose M, Daczkowska M, Gopisetty S (2020) Analytics and operations research increases win rates for ibm's information technology service deals. INFORMS J Appl Anal 50(1):50–63

Morteza L, Kuan YW (2012) A review of modelling approaches for supply chain planning under uncertainty. In: International conference on services systems and services management ICSSSM12. IEEE, pp 197–203

Mubeen S, Asadollah SA, Papadopoulos AV, Ashjaei M, Pei-Breivold H, Behnam M (2017) Management of service level agreements for cloud services in iot: a systematic mapping study. IEEE Access 6:30184–30207

Muller C, Fernandez AMG, Fernandez P, Martín-Díaz O, Resinas M, Ruiz-Cortés A (2018) Automated validation of compensable SLAs. IEEE Transactions on Services Computing

Muñoz-Escoí FD, Bernabéu-Aubán JM (2017) A survey on elasticity management in paas systems. Computing 99(7):617–656

Márquez-Chamorro AE, Resinas M, Ruiz-Cortés A, Toro M (2017) Run-time prediction of business process indicators using evolutionary decision rules. Expert Syst Appl 87:1–14

Márquez-Chamorro AE, Resinas M, Ruiz-Cortés A (2018) Predictive monitoring of business processes: A survey. IEEE Trans Serv Comput 11(6):962–977

Márquez-Chamorro AE, Revoredo K, Resinas M, Del-Río-Ortega A, Santoro FM, Ruiz-Cortés A (2020) Context-aware process performance indicator prediction. IEEE Access 8:222050–222063. https://doi.org/10.1109/ACCESS.2020.3044670

Müller C, Oriol M, Franch X, Marco J, Resinas M, Ruiz-Cortés A, Rodríguez M (2014) Comprehensive explanation of SLA violations at runtime. IEEE Trans Serv Comput 7(2):168–183

Müller C, Truong H, Fernandez P, Copil G, Ruiz-Cortés A, Dustdar S (2016) An elasticity-aware governance platform for cloud service delivery. In: 2016 IEEE international conference on services computing (SCC), pp 74–81

Nasr M, Shokri R, Houmansadr A (2019) Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. In: IEEE Symp. on security and privacy, pp 739–753

Nawaz F, Janjua NK, Hussain OK, Hussain FK, Chang E, Saberi M (2018) Event-driven approach for predictive and proactive management of SLA violations in the cloud of things. Future Gener Comput Syst 84:78–97

Noy NF (2004) Semantic integration: a survey of ontology-based approaches. ACM Sigmod Record 33(4):65–70

Pasic F, Wohlers B, Becker M (2019) Towards a KPI-based ontology for condition monitoring of automation systems. In: 24th IEEE international conference on emerging technologies and factory automation (ETFA), pp 1282–1285

Peidro D, Mula J, Poler R, Lario FC (2009) Quantitative models for supply chain planning under uncertainty: a review. Int J Adv Manuf Technol 43(3–4):400–420

Pérez-Álvarez JM, Maté A, Gómez-López MT, Trujillo J (2018) Tactical business-process-decision support based on KPIs monitoring and validation. Comput Ind 102:23–39

Petersen B, Pisinger D, Spoorendonk S (2008) Chvátal-gomory rank-1 cuts used in a dantzig-wolfe decomposition of the vehicle routing problem with time windows. In: The vehicle routing problem: latest advances and new challenges. Springer, pp 397–419

Prasad SK, Shankar R (2018) Service capacity coordination in it services supply chain. J Model Manag

Sfondrini N, Motta G, You L (2015) Service level agreement (SLA) in public cloud environments: a survey on the current enterprises adoption. In: 5th international conference on information science and technology. IEEE, pp 181–185

Shvaiko P, Euzenat J (2005) A survey of schema-based matching approaches. In: Journal on data semantics IV. Springer, pp 146–171

Singh KJ, Philpott AB, Wood RK (2009) Dantzig-wolfe decomposition for solving multistage stochastic capacity-planning problems. Oper Res 57(5):1271–1286

Sloman M (1994) Policy driven management for distributed systems. J Netw Syst Manag 2(4):333–360

Stadtler H (2008) Supply chain management-an overview. In: Supply chain management and advanced planning. Springer, pp 9–36

Sun Z, Zhang Q, Hu W, Wang C, Chen M, Akrami F, Li C (2020) A benchmarking study of embedding-based entity alignment for knowledge graphs. Preprint arXiv:200307743

Trienekens JJ, Bouman JJ, Van Der Zwan M (2004) Specification of service level agreements: problems, principles and practices. Softw Qual J 12(1):43–57

Truex S, Baracaldo N, Anwar A, Steinke T, Ludwig H, Zhang R, Zhou Y (2019) A hybrid approach to privacy-preserving federated learning. In: Proceedings of the 12th ACM workshop on artificial intelligence and security, pp 1–11

Vanderbeck F, Savelsbergh MW (2006) A generic view of dantzig-wolfe decomposition in mixed integer programming. Oper Res Lett 34(3):296–306

Veit F, Geyer-Klingeberg J, Madrzak J, Haug M, Thomson J (2017) The proactive insights engine: process mining meets machine learning and artificial intelligence. In: 2016 international conference on business process management (BPM), Demo Sessions

Wang Y, Wallace SW, Shen B, Choi TM (2015) Service supply chain management: a review of operational models. Eur J Oper Res 247(3):685–698

Wetzstein B, Danylevych O, Leymann F, Bitsaki M, et al. (2009) Towards monitoring of key performance indicators across partners in service networks. In: Workshop on service monitoring, adaptation and beyond, p 7

Wieder P (2006) SLA negotiation. Exploit Knowl Econ Issues Appl Case Stud 3:44

Wimmer MA, Tambouris E (2002) Online one-stop government. In: IFIP World computer congress, TC 8, Springer. pp 117–130

Wu T, Shi Z, Liang Z, Zhang X, Zhang C (2020) Dantzig-wolfe decomposition for the facility location and production planning problem. Comput Oper Res 124:105068

Zhong Y, Guo F, Tang H, Chen X, Xin B (2020) Research on coordination complexity of e-commerce logistics service supply chain. Complexity 2020

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Robert Engel[1] · Pablo Fernandez[2] · Antonio Ruiz-Cortes[2] · Aly Megahed[1] · Juan Ojeda-Perez[2]

✉ Robert Engel
robert.m.engel@gmail.com

[1] IBM Research - Almaden, San Jose, CA, USA

[2] SCORE Lab., Universidad de Sevilla, Seville, Spain