Streamlining Management of Multiple Cloud Services

Aditi Gupta, Sudip Mittal, Karuna P. Joshi, Claudia Pearce and Anupam Joshi
University of Maryland, Baltimore County, Baltimore, MD 21250, USA
Email: {adigupta, smittal1, kjoshi1, cpearce, joshi}@umbc.edu

Abstract—With the increase in the number of cloud services and service providers, manual analysis of Service Level Agreements (SLA), comparison between different service offerings and conformance regulation has become a difficult task for customers. Cloud SLAs are policy documents describing the legal agreement between cloud providers and customers. SLA specifies the commitment of availability, performance of services, penalties associated with violations and procedure for customers to receive compensations in case of service disruptions. The aim of our research is to develop technology solutions for automated cloud service management using Semantic Web and Text Mining techniques. In this paper we discuss in detail the challenges in automating cloud services management and present our preliminary work in extraction of knowledge from SLAs of different cloud services. We extracted two types of information from the SLA documents which can be useful for end users. First, the relationship between the service commitment and financial credit. We represented this information by enhancing the existing Cloud service ontology proposed by us in our previous research. Second, we extracted rules in the form of obligations and permissions from SLAs using modal and deontic logic formalizations. For our analysis, we considered six publicly available SLA documents from different cloud computing service providers.

Keywords—Cloud computing, Service Level Agreements, Modal Logic, Cloud Management and Operations, Service Management Automation, Knowledge Discovery.

I. INTRODUCTION

With the evolution of cloud computing platforms, a plethora of services are now being offered to customers. But the process of discovering and negotiating for these cloud services is still a very manual process. Consumers often select cloud providers based on the provider’s market share, reputation and service cost rather than on their enterprise needs and constraints. Legal contracts between the cloud consumer and providers, like Service Level Agreement (SLA), Terms of Service (TOS) descriptions and privacy policies, are used to define the details of service data, delivery mode, performance metrics and costs and penalty terms. These contracts are currently managed as text documents and require extensive manual effort to understand their conditions and to extract information from them, for example, key service performance indicators. Hence another challenge faced by cloud consumers is how to efficiently manage the various cloud services, that they have acquired, through a continuous monitoring system that automatically compares service performance with acceptable SLA benchmarks.

We have analyzed many publicly available SLAs from various cloud service providers and our preliminary observation has been that each service provider has a separate SLA document with custom definitions, rules and policies for each of the service offerings. While understanding of the SLA terms is essential for customers to make right decisions when choosing a service provider, consumers seldom familiarize themselves with the complete SLA document and its details that define the quality of services they can expect from the service provider. The complexity of this situation is further increased when complex legal contracts are written between third party service providers and customers based on the publicly available SLA documents of service providers. In these scenarios, managing and monitoring the services along with understanding nuances of the legal documents manually, becomes quite tedious. Hence, the need for automated management of cloud services and SLA documents is a need of the hour.

The aim of our research is to build a framework to represent cloud legal documents in a machine-actionable form so that automated techniques can be developed to understand, analyze and make decisions based on the knowledge extracted from documents such as service level agreements, terms of service and privacy policies. We have utilized a combination of Semantic Web and Text Mining techniques for building our proposed framework. In this paper, we present our analysis of automating the extraction of knowledge and policy rules from Cloud SLA documents. We used a cloud computing ontology built by us in our previous work and extended the same to incorporate classes to represent the service commitment and other cloud SLA specific metrics. The main contributions of this paper are:

- We extracted the relationship between the service commitment and financial credit information, and populated this data in the knowledgebase created over a cloud ontology.
- We extracted rules, permissions and obligations from SLA documents using the grammatical patterns based on Modal Logic. We analyzed SLA documents by different cloud computing service providers. We further analyzed the rules and policies extracted using Deontic Logic formalizations.

The next section contains the literature review for our work. Section IV describes the methodology of work done
in this paper. Section V and VI discuss the analysis performed for various SLA documents. Section VII presents the conclusions and future work for this research.

II. LITERATURE REVIEW

In a virtualized service-oriented environment, consumers and service providers need to exchange information, queries, and requests with some assurance that they share a common meaning. This is critical not only for the data but also for the policies followed by service consumers or providers. The handling of heterogeneous policies is usually not present in a closed and/or centralized environment, but is an issue in the open cloud. The interoperability requirement is not just for the data itself, but even for describing services, their service level agreements, quality related measures, and their policies for sharing data. We propose the use of Semantic Web techniques for modeling and reasoning about services related information. We have used this approach for automating management and analysis of Cloud service level agreements. The Semantic Web deals primarily with data instead of documents. It enables data to be annotated with machine understandable meta-data, allowing the automation of their retrieval and their usage in correct contexts. Semantic Web technologies include languages such as Resource Description Framework (RDF) [1] and Web Ontology Language (OWL) [2] for defining ontologies and describing meta-data using these ontologies as well as tools for reasoning over these descriptions.

In recent years, international bodies have been working towards defining standard definitions and terminologies for a Cloud SLA. The International Organization for Standardization’s ISO/IEC DIS 19086 [3] standard, European Commission’s “Cloud Service Level Agreement Standardization Guidelines” [4] and NIST’s Special Publication 500-307 on Cloud Computing Service Metrics Description [5], are some of the publications that when finalized will help consumers mandate a fixed SLA format from various cloud providers. That will also help consumers compare and contrast the various cloud services based on their terms of service/SLA metrics. Our proposed ontology is flexible enough to capture these standardization efforts as they mature, and our system will be able to express SLAs in terms of these proposed standards.

For our cloud services lifecycle framework [6], we have used semantic web technologies like OWL instead of WS-Agreement and WS-Negotiation protocol as we were able to more richly define the cloud SLA ontologies, thereby allowing us to incorporate different descriptions of the same SLA measure.

Researchers have applied Natural Language Processing (NLP) techniques to extract information from text documents. In Rusu et. al. [7] the authors suggest an approach to extract subject-predicate-object triplets. They generate Parse Trees from English sentences and extract triplets from the parse trees. Etzioni et. al. [8] developed the KNOWITALL system to automate the process of extracting large collections of facts from the Web in an unsupervised, domain-independent, and scalable manner. Etzioni et. al used Pattern Learning to address this challenge. Various textual information extraction and retrieval systems have been proposed in [9], [10], [11], [12].

Another important natural language technique used for information extraction from unstructured text is ‘Noun Phrase Extraction’. Rusu et. al. in [7] show how to create triplets by considering ‘Noun Phrases’ obtained by using various part-of-speech taggers. Barker et. al. [13] extract key-phrases from documents and show that noun phrase-based system performs roughly as well as a state-of-the-art, corpus-trained key-phrase extractor. Similar techniques have also been suggested in [8].

Documents consist not only of a huge number of unstructured texts, but also a vast amount of valuable structured data in the form of tables. Extracting knowledge from structured tables is an ongoing research problem with multiple solutions proposed to handle both general and domain specific tables. Mulwad et al. [14] proposed a framework which assigns a class to table columns and links table cells to entities, and inferred relationships between columns to properties. Bhagavatula et al. [15] created a system called TabEL which works by extracting content using entity linking. Various other solutions have been proposed in [16] [17] [18] [19]. We use a modified version of [14] to improve our knowledge extraction system.

Researchers have explored the automated techniques for extracting permissions and obligations from legal documents using text mining and semantic techniques [20] [21] [22]. Kagal et al. formed [23] [24] proposed a semantic web based policy framework to model conversation specifications and policies using obligations and permissions.

In one of our prior works, we described a new integrated methodology for the lifecycle of IT services delivered on the cloud, and demonstrate how it can be used to represent and reason about services and service requirements and to automate service acquisition and consumption from the cloud. We have divided the IT service lifecycle into five phases of requirements, discovery, negotiation, composition, and consumption. We detail each phase and describe the ontologies that we have developed to represent the concepts and relationships for each phase. We have described the five phases in detail along with the associated metrics in [25].

We have also described a preliminary knowledge extraction system for cloud SLA documents in [26]. In this paper we extend our framework to include extraction of information from tables and rules in the form of obligations and permissions from cloud SLA documents.
III. CLOUD LEGAL DOCUMENTS

A. Types of Legal Documents

We conducted a comprehensive survey of publicly available cloud legal documents by reviewing over 100 cloud providers. Details of results of that study have been published in [27]. The cloud legal documents fall into the following broad categories -

- Service Contract documents. Cloud legal documents in this category lay down rules and clauses that specify service functionality, quality and performance metrics. They also specify user access policies and service availability. Documents in this category include Service Level Agreement, Terms of Service, Customer Agreement, Acceptable Use Policy, etc. A key role of these contracts is to enable efficient management of the cloud service. Metrics listed in these contracts are used by the vendors to assure the consumer of high performance of their services. NIST’s special publication 500-307 [28] defines ‘metrics’ as - ‘provides knowledge about characteristics of a cloud property through both its definition (e.g. expression, unit, rules) and the values resulting from the observation of the property. For instance, a customer response time metric can be used to estimate a specific response time property (i.e. response time from customer to customer) of a cloud email service search feature. It also provides the necessary information that is needed for to reproduce and verify observations and measurement results.’ However, one glaring observation after reviewing these contracts was the sheer variety of formats of these documents and the metrics used to track performance of the same type of service. Due to lack of a standard cloud contract format, we came across SLA documents that were 3 pages long to ones that were over 25 pages in length even though they were offering the same type of computing service.

- Privacy and Security Data Documents. Cloud legal documents in this category lay down the data security and privacy policies for the service. These documents are usually required by federal and state regulations. Cloud providers either create a single privacy policy for all their services or create multiple privacy policies depending on the service data / category. Privacy policy documents were observed to be similar in content and format across most of the cloud vendors.

- Regulatory Compliance documents. Compliances are set of rules, policies or standards formulated by regulatory agencies or standards organizations [29]. Compliance models implement rules and regulations across various components of Information Technology (IT) to make them work harmoniously. Security and privacy compliance models have been proposed for Cloud computing security to ensure data protection and user privacy. Some of the proposed models include ISO 27001, COBIT etc. The features of all regulations that are relevant to cloud data security and privacy is discussed in detail in [27].

B. Research Challenge

1) Lack of standardization: Due to the lack of standards for cloud service performance, providers often construct their own rules for the performance measures and metrics and define them as ‘clauses’ in the cloud legal documents, such as TOS, SLAs or privacy policy documents, that are part of the cloud contract. In addition, the regulatory and compliance bodies have also developed rules and policies that affect the way cloud services can be provided or consumed [27]. Reviewing all these cloud legal documents to ensure the cloud service is meeting the organizational requirements is a labor and time intensive endeavor for consumers and is often an afterthought when a cloud service fails to live up to its expectations. Hence, continuous cloud service monitoring has been identified as a key open issue by consumers [3]. A critical step in automating cloud service management is to make the cloud SLAs machine understandable so that monitoring tools can interpret the policy rules and metrics defined in the service contracts.

2) Co-referencing/cross referencing: Many legal documents tend to reference other documents or different sections within the document. There is a need to develop techniques to represent and reason over multiple legal documents that co-reference / cross reference another set of documents and/or sections within the document.

C. Components of SLA

In our research we focus on SLA documents for various cloud services. The main components of a SLA document are:

- Service Commitment: The SLA document specifies the service commitments made by the service provider to their customers.
- Definitions and metrics: The SLA documents also contain definitions and metrics such as “Availability”, “Region” and “Uptime/Downtime” which are used to compute the service commitments.
- Rules for computation of service credit corresponding to the commitments by the provider.
- Procedure to request for service credit by the user and payment method by the service provider. These describe the specifications about how a user can apply for a service credit, for example, the request for service credit must include the account information and exact dates of outages of the service.
- Conditions under which the service provider is not liable to its service commitment (exclusions) to the users.
IV. METHODOLOGY

The aim of our research is to develop a system that will enable the end user to manage multiple cloud services. Figure 1 describes the overall architecture diagram of such a system. In this paper we describe our framework to automatically extract knowledge (facts and rules from cloud computing SLAs) using text mining and semantic web techniques from SLAs. We envision a system which will maintain knowledge about various terms and rules contained in SLAs, compliance and regulatory policies, contracts, privacy documents, etc. We build upon our work in [25], [26].

A. Creating a Cloud SLA ontology

A critical step in automating cloud service management is to make the cloud SLAs machine readable so that software agents can be used for service monitoring and interpreting the policy rules. As part of our previous work, we had developed a semantically rich ontology to capture key elements of cloud SLAs [6].

B. Populating the Cloud Services Knowledge Base

We perform unsupervised knowledge extraction from unstructured textual documents in order to store the information in a machine understandable format. In one of our prior works, we created a system to automatically extract various metrics, important terms-of-art and their definitions. We did that by iterating over all sentences and each sentence is passed to the Stanford POS Tagger [30] and CMU Link Parser [31] to generate a Parse Tree. After generating the Parse Tree we created triples (Subject-Predicate-Object) where important keywords from the ontology are found [26]. In this paper, we extend our knowledge extraction system by automated extraction of information from SLA documents.

We extract the relationship between the service commitment and financial credit defined in tables as well as rules that are present in SLA documents in the form of obligations and permissions. The next two sections describe this in detail.

C. Data Description

Cloud service providers cover multiple services such as compute, DNS, online data and file storage services. In our paper, we analyze publicly available SLA documents of the cloud computing platforms such as Amazon Web Services EC2, Google Compute Engine, HP Cloud Compute, Microsoft Azure Cloud, VMware vCloud Air and IBM SoftLayer [32]–[37].

V. EXTRACTION OF SEMANTIC INFORMATION FROM TABLES

Often, legal documents such as SLA, Terms of service and contracts contain data in the form of tables. Tables contain high-quality facts present in a form that is easy for humans to understand and comprehend, however making it machine computable is a challenge. Extracting the semantics of Web tables to produce machine-understandable knowledge has become an active area of research. Figure 2a gives examples of tables present in a Cloud SLA to describe the service credit. In our system we have automated the extraction of data in tables present in different cloud SLAs. After extracting the data from SLA we encode it as RDF statements [38]. These RDF statements are then added to our knowledge base [26]. Once added to the knowledge base, the information may be used for computing and reasoning to answers for complex queries by the user.

We extracted text data from HTML versions of the SLA documents using Beautiful Soup Library1, which is a python library. After parsing the table using this library we get a python dictionary with the content of various table cells. Beautiful soup helps us to do text retrieval from structured HTML documents. After getting the python dictionary with the data we use the table headers to associate semantics to various cell data using our ontology discussed in [25], [26]. This data can now be stored as RDF statements in our knowledge base. RDF statements generated for Figure 2a:

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>:uptimel a owl:DatatypeProperty; rdfs:domain :Service_Availability; rdfs:range &quot;Less than 99.95% but equal to or greater than 99.0%&quot;; sla:hasServiceCredit '10\%'.</td>
<td></td>
</tr>
<tr>
<td>:uptimel a owl:DatatypeProperty; rdfs:domain :Service_Availability; rdfs:range &quot;Less than 99.0%&quot;; sla:hasServiceCredit '30\%'.</td>
<td></td>
</tr>
</tbody>
</table>

VI. EXTRACTING RULES AND OBLIGATIONS FROM SLA USING MODAL LOGIC

Next, we propose using Modal logic formalizations to extract rules, permissions and obligations specified in cloud SLAs [39]. This way of analyzing the rules, permissions and obligations can help customers to understand their obligations as well as their rights while using cloud services. This analyses will also help customers to make decisions regarding which service provider to use for cloud computing services. One of the major challenges in extracting rules and polices from SLA documents is that most documents use vendor specific terminologies. Therefore, we propose a generic framework based on grammatical structure of sentences to extract rules and policies from SLA.

Modal logic is a broad term used to cover various other forms of logic such as temporal logic and deontic logic [39]. Deontic logic describes statements about permissions and obligations, and temporal logic describes the temporal expressions. We first extracted statements from

1https://pypi.python.org/pypi/beautifulsoup4
SLA documents which contained modal expressions and then we further selected statements containing deontic expressions such as permissions, obligations, dispensations and prohibitions. Researchers have used deontic principles in the past to analyze policy and legal documents. Work done by Travis D. Breaux utilized semantic web and text mining approaches to extract obligations and permissions from privacy policies [20] [21]. We also used similar techniques to extract deontic rules from cloud SLA documents.

We considered SLA documents of six cloud compute services for analysis in this section: Amazon Web Services, Microsoft Azure, IBM SoftLayer, HP Cloud Compute, Google Compute Engine and VMware vCloud Air services.

A. Using part-of-speech tags for extracting Modalities

In order to extract rules in the form of obligations and permissions from statements in the cloud SLAs we used a text mining approach based on the part-of-speech tags. We extracted the modal expressions from the documents using grammatical rules based on modal verbs. We first used the Stanford Lexicalized Parser to analyze the sentences extracted from the SLAs [40] in our dataset. Figure 3 shows the output of the parser for each one of the statements from the SLAs. Each word of the sentence is tagged with its corresponding part-of-speech such as Noun (NN), Verb (VB) and Personal pronoun (PRP).

We extracted statements from the SLA containing the modal verbs by the grammatical rules defined below. From
the output of the parser obtained (as shown in Figure 3), we formed grammar rules to represent the expressions containing modal verbs. The basic format of the rules are given below.

Positive Modalities:
\[ <\text{Noun} / \text{Pronoun}> <\text{modal verb}> <\text{verb}> \]

Negative Modalities:
\[ <\text{Noun} / \text{Pronoun}> <\text{modal verb}> <\text{negation}> <\text{verb}> \]

Table I below represents the frequency distributions of modal verbs extracted from the SLA documents using the above grammatical rules. The noun and pronoun part of the modalities are used to define the actors associated with the particular modality. We found that there were two types of actors present in the SLAs: the ‘Service Provider/Service’ and the ‘Customer’.

Most of these statements represented the rules regarding the services provided by the cloud platforms in the form of permissions and obligations. In Figure 4, we analyzed the verbs associated with the modalities in the statements extracted above. Verbs such as ‘eligible’ accompany the rights/permissions of customers and ‘provide’ usually indicate statements about the service provider’s commitment.

<table>
<thead>
<tr>
<th>modal verb</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>will</td>
<td>56</td>
</tr>
<tr>
<td>may</td>
<td>22</td>
</tr>
<tr>
<td>must</td>
<td>16</td>
</tr>
<tr>
<td>can</td>
<td>5</td>
</tr>
<tr>
<td>could</td>
<td>3</td>
</tr>
<tr>
<td>shall</td>
<td>2</td>
</tr>
<tr>
<td>should</td>
<td>1</td>
</tr>
</tbody>
</table>

Table I: Frequency of occurrence modal verbs in cloud service level agreements.

Figure 3: Part-of-speech tags obtained for each statement in the SLA documents using the Stanford Lexicalized Parser.

B. Comparison across service providers

Figure 5 shows the comparison of number of statements containing rules extracted using Modalities versus total number of statements extracted (excluding tables and headings) for each of the service provider. Identifying statements containing modalities can be useful for users to understand their obligations and rights regarding the services provided by the vendor. This comparison may also help the customers in selecting the appropriate services for their requirements.

C. Extracting Permissions and Obligations from SLAs

Next, we extracted rules such as permissions and obligations (Deontic expressions) from statements containing modal expressions. There are four basic types of deontic expressions:

- Permissions/Rights: Permissions are expressions that describe rights or authorizations for an entity/actor.
- Dispensations: Dispensations describe optional or non-mandatory statements.
- Obligations: Obligations define the responsibilities that an entity/actor must perform.
- Prohibitions: Prohibitions specify the conditions or actions which an entity is not permitted to perform.

We categorized the statements extracted above as permissions/obligations depending upon the modal verbs present in them. For example, statements with modal verbs: shall, should, must were categorized as obligations/dispensations and those containing modal verbs as may, can, could were categorized as permissions/dispensations. Following are examples of deontic rules extracted by our analysis:

- Obligation (Actor: Customer)
  “You must follow the procedure described herein within seven days of the end of the Claimed Outage.”

- Obligation (Actor: Service Provider)
  “For each 30 continuous minute period of Qualifying Outage Minutes for a Service in a Measurement Period, [Service Provider] shall provide an SLA Credit of ...”
VII. Conclusion & Future Work

There are multiple cloud services offered by various vendors, each having their own service level agreements. Customers such as organizations, universities etc. often use multiple cloud providers as specialized services provided by a cloud platform may not be available on other platforms. Hence, a centralized cloud management system is useful to manage multiple cloud providers. The aim of our research is to transform long legal documents such as SLAs, terms of service, etc. into machine-actionable format. We present part of our work done by using semantic web and text mining techniques in automating cloud service management. We extracted facts and rules present in SLA documents as tables, permissions and obligations. The relationship between the service commitment and financial credit was extracted from tables present in SLA and included as a separate class in existing Cloud service ontology proposed by us in our previous research. Also, we extracted various rules in the form of permissions and obligations from SLAs using modal and deontic logic formalizations. In this paper, we considered six publicly available SLA documents from different cloud computing service providers. In future, we would aim to extend our dataset to more publicly available SLAs as well as other legal documents such as Terms of Service and privacy policy documents. Also, we would aim to build a web based interface for users to analyze and manage multiple cloud services.

ACKNOWLEDGMENT

This research was partially supported by a DoD supplement to the NSF award #1439663 : NSF I/UCRC Center for Hybrid Multicore Productivity Research (CHMPR).

REFERENCES


