Agent based Aggregation of Cloud Services - A Research Agenda

Dr. Nandini Sidnal¹ and Sreedevi R. Nagarmunoli¹

¹KLEDRMSSCET, Belgaum, Karnataka, India
sidnal.nandini@gmail.com

Abstract—Cloud computing has come to the forefront as it overcomes some of the issues in computing such as storage space and processing power. It enables ubiquitous accessing and processing of information without the need of excessive computing facilities. In this work, we plan to brief some of the issues in aggregating the cloud services, discovering futuristic cloud service requests, develop a repository of the same and propose an agent based Quality of Service (QoS) provisioning system for cloud clients.

Index Terms—Aggregation, Futuristic cloud services, Repository.

I. INTRODUCTION

With the rapid development of processing and storage technologies and the success of the Internet, computing resources have become cheaper, more powerful and more ubiquitously available than ever before. This technological trend has enabled the realization of a new computing model called cloud computing, in which computing resources such as CPU, storage etc. are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion.

National Institute of Standards and Technology (NIST) defines cloud computing as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [29,30]. It enables ubiquitous accessing and processing of information without the need of excessive computing facilities. Compared to other distributed computing paradigms such as Grid computing and High Performance Computing (HPC), cloud computing provides broader interoperability over the world-wide web networks [6,9].

Some of the issues in cloud computing environment can be classified as Platform Management, Cloud-enabled Applications, challenges in Cloud Management, Cloud Enablement, Cloud Interoperability, elastic scalability, trust, security, privacy, data handling, programming models, resource control, systems development and systems management and Aggregation of Cloud Services.

Aggregation of Cloud Services - The research challenges in the aggregation of resources from diverse cloud providers adding additional layers of service management, novel architectural models for aggregation of cloud providers, brokering algorithms for high availability, performance, proximity, legal domains, price, or energy efficiency, sharing of resources between cloud providers, networking in the deployment of services across multiple cloud providers, Service Level Agreement (SLA) negotiation and management between cloud providers, additional privacy, security and trust management layers atop providers and support of context-aware applications. Works in [1-5, 15-27, 11-14] depict that not much attention is paid in monitoring and developing repository of cloud services, customized aggregation of services and distribution of services.

The field of software agent technology is a rapidly developing area of research which encompasses a diverse range of topics and interests [31-37]. Cognitive agents that mimic human thought process and represent the logical transition of research on human information processing to practical application are deployed to develop an autonomous aggregating system. Section 2 presents the literature survey and section 3 describes the objectives of the research to be carried out. Section 4 describes the proposed methodology and section 5 provides the possible outcome and conclusion.

II. LITERATURE SURVEY

The survey is based on the research works carried out in the universities, by the academicians and works in research laboratories such as HP, IBM etc. [13] discusses the concept of “cloud” computing, it tries to address some of the issues related to research topics, and the “cloud” implementation available today. [14] investigates the challenges of developing a Campus Cloud based on aggregating resources in multiple universities. [10] presents a policy-centered QoS meta-model which can be used by service providers and consumers alike to express capabilities, requirements, constraints, and general management characteristics relevant for SLA establishment in service aggregations. HP Lab is focusing on delivering the secure application and computing end-state of “Everything-as-a-Service”.

IBM researchers’ adopted cloud computing for faster turnaround times in provisioning of resources for specific research projects [14]. Google and IBM are jointly working on data centers in cloud. The Cloud Computing and Distributed Systems (CLOUDS) Laboratory is actively engaged in the design and development of next-generation computing systems and applications that aggregate or lease services of distributed resources depending on their availability, capability, performance, cost, and users’ QoS requirements [9].

Some of the ongoing research projects in cloud computing area by different universities are discussed in the following paragraphs. The Researchers at Boston University are...
exploring the merits of “Collocation Games” (CGs) as a novel, economically-sound framework upon which emerging cloud architectures could be implemented [38]. Research is going on at Duke University to explore and test Trustworthy Virtual Cloud Computing [38]. Florida International University (FIU) researchers are leveraging cloud computing to analyze aerial images and objects to help support disaster mitigation and environmental protection [38]. The researchers at Indiana University are working on Large-Scale Distributed Scientific Experiments on Shared Substrate, exploring the use of cloud techniques to overcome current medical computing obstacles [38]. The team at MIT is working in collaboration with Yale University and the University of Wisconsin at Madison on a comparative study of approaches to cluster-based, large-scale data analysis and cloud for education [38].

The aim of the project work at University of St. Andrew is to investigate how underused computing resources within an enterprise may be harvested and harnessed to improve return on IT investment [15], [16] aims at specifying, measuring and understanding high level cloud properties. The aim of the project [17] is to develop and evaluate techniques to allow desired high-level properties to be specified, mapped into appropriate low-level actions, and the results to be measured and reported in terms of the high-level properties. The goal of the research work defined in [18] is to make experiments better by using the Cloud in a number of ways. The project [19] would investigate a range of problems in the established area of computational abstract algebra in order to see whether, or how, they can be effectively parallelized using this framework.

The aim of the project [20] is to investigate the practical issues which affect data migration in the cloud and to propose mechanisms to specify policies on data migration and to use these as a basis for a data management system. The work defined in [21] aims to investigate how a migration of applications may result in changes to the way that work is actually done. [22] investigates the use of cloud computing for mobile network data archiving: there are varieties of topics in distributed systems including network measurement, privacy, sanitization, data protection and computation caching.

The research work in [23] discusses security issues. The topics of research within the topic of Cloud Verification, Validation and Testing (VVT) from formal verification through to empirical research and metric validation of multi part or parallel analysis are discussed in [24]. The aim of the project in [25] is to apply constraint programming techniques to solve issues in cloud environment efficiently.

The work in [26] explores the use of virtualization in system and network resources in order to minimize energy usage whilst still meeting the service requirements and operational constraints of a cloud. [27] discusses the consequence of dynamically provisioned resource allocation under denial of service attacks, in order to reduce the wasting of resources. [28] proposes to revise the analytical model to accommodate Cloud Computing and carry out experiments and measurements, to compare the responsiveness with earlier work done on Web and Grid computing.

III. OBJECTIVES

The objectives of the proposed research work that will be carried out are discussed in this section. The following issues in cloud computing are studied and some of the issues in aggregating the cloud services would be resolved. Cloud aggregator is a platform or service that combines multiple clouds with similar characteristics (geographic area, cost, technology, size, etc.) into a single point of access, format, and structure. Value is derived from cost savings and greater efficiency found from the ability to easily leverage multiple services providers.

As a cost-effective and time-efficient way to develop new applications and services, service aggregation in cloud computing empowers all service providers and consumers and creates tremendous opportunities in various industry sectors. However, it also poses various challenges in securing the information on cloud.

Some of the issues that need to be resolved in aggregating [7-8, 10] the cloud services are availability of services that may be hired in real time without conflicts, novel architectural models for aggregation of cloud providers, brokering algorithms for high availability, performance, proximity, legal domains, price, or energy efficiency, sharing of resources between cloud providers, networking in the deployment of services across multiple cloud providers, additional privacy, security and trust management layers atop providers, support of context-aware applications and automatic management of service elasticity.

Objectives of the research are to design an agent focused on aggregation of services for cloud clients. The issues considered in our research are:

- To design a cognitive agent based novel architecture/scheme for discovering futuristic cloud services (that may be in demand) and develop a repository of the same by networking multiple cloud providers.
- To design a scheme to autonomously and intelligently monitor, negotiate and aggregate the resources from the cloud repository based on the QoS (time, price, availability) defined in the cloud client’s requests. The scheme shall explore the use of virtualization in system and resources in order to minimize energy usage whilst still meeting the service requirements and operational constraints of a cloud.
- To design a scheme to dynamically and automatically schedule and deliver the services to the requested clients ensuring high availability of services and to develop billing and pricing model for measuring cloud services utility.

IV. PROPOSED METHODOLOGY

The research work aims at aggregating and providing customized set of services to the requesting clients in an efficient manner.
A. Discovery of cloud services

The cloud providers are networked by segmenting or clustering them based on type of services provided, geographical locations etc. Cognitive agents crawl blindly through the cloud to discover the cloud services fulfilling the futuristic requests and build the repository of services. Further request prediction may be done using log record, click stream record and user information or Markov model to anticipate futuristic requests for discovering the cloud services. Discovery process may be carried out in parallel using the concept of agent cloning. Repository shall be updated at regular frequency to eliminate the stale information using aging techniques. Multidimensional data structure shall be deployed to store the cloud services in the repository. Efficient indexing algorithms and meta-services (service cache) shall be adopted to retrieve the service information from the repository to improve the performance of the repository access. The repository shall store the services offered, vendor details, pricing, current status, QoS etc.

B. Aggregation of requested services

Based on the service requests from cloud clients cognitive agents monitor the status of the services, negotiate with the vendors, and aggregate them based on the specified QoS. Unsupervised learning mechanism may help the agents to negotiate intelligently for better prices to aggregate and distribute the cloud services. English auctions may be used to maximize the profits for vending the services. Multiple options of aggregated services are to be given to the clients in order to increase their satisfaction level.

C. Distribution of services

After the services are aggregated they are to be distributed in a customized way. Scheduling has to be done in an optimal way so as to maximize the availability and utility of services. Billing and pricing algorithms are to be developed for the delivered services. All the above objectives will be simulated under various scenarios to assess the performance and effectiveness of the proposed scheme. The simulation shall be carried out on IBM Blade Center HS22 using compatible programming language.

V. Possible outcome and conclusions

Services on the cloud are plenty but the clients are not able to get the required services. There is no common repository of availability of cloud services. The proposed work will develop a framework to overcome the above mentioned issues. Further the usage of cognitive agents offers several benefits in aggregating cloud services such as autonomy in discovering the cloud services, developing and updating the repository, embedding intelligence, flexibility in negotiation, adaptability to network environments, customization of QoS requirements etc. The research work may be enhanced in future by employing some agent based solutions to other issues such as cloud management, enablement, and interoperability and to develop some applications. Further we have planned to publish our research findings in referred journals and present in national/international conferences.

REFERENCES

[12] OxGrid http://www.oerc.ox.ac.uk/resources/oxgrid/oxgrid-concept