BOINC from the view point of Cloud computing

Albertas Jurgeleviius
albertas.jurgelevicius@mii.vu.lt

prof. habil. dr. Leonidas Sakalauskas
leonidas.sakalauskas@mii.vu.lt

Vilnius University Institute of Mathematics and Informatics
Akademijos str. 4, LT-08663 Vilnius, Lithuania

Abstract

Public distributed computing is a well-researched and applied computational model for organizations to adopt IT services with little cost. However, businesses and public institutions are not interested in adopting this model due to security, reliability concerns and challenges associated with it. While these issues are researched, and solved well for cloud computing applications, these problems remain open in distributed public computing models. In this paper, we have presented a review of such computational model adoption issues from business stand point perspective. The paper attempts to identify the fundamental security and reliability issues of distributed public computing model, preventing businesses and organizations from adopting the model and making them use less affordable and legal issues involving cloud computing model. The outcome of this paper provides a foundation for future analysis and review regarding the impact of public distributed computing service security and reliability increase for IT solutions in business environment.

1 Introduction

The term big data refers to the task of capturing, storing or processing of huge data sets. Due to the large amounts of data, traditional data management and data analysis approaches, like internal high-performance computing environments, are no longer feasible [1]. As business companies and organizations gather and connect to more and more data, demand for computational resources are starting to outgrow the capabilities companies internal IT infrastructures can handle. In such cases companies tend to either invest more into building their IT infrastructures further or look for third-party solutions. Since buying or upgrading servers is an expensive solution, this often leads to searching for companies that do have the required infrastructure and offer their solutions at affordable price. Cloud computing is growing continuously because it keeps providing high-performance computational services at cheaper and cheaper rates. Famous IT companies have provided their cloud service on the internet, such as Microsoft (http://azure.microsoft.com/), Amazon (http://aws.amazon.com/), Google (https://cloud.google.com/) and Rackspace (http://www.rackspace.com/) [2].

Such services are often used by companies to process enormous data sets they have. Gathered knowledge can help make right business decisions and provide value data to clients. Since this is a computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems, high-performance grid computing infrastructures may exceed the

In: E. Ivashko, A. Rumyantsev (eds.): Proceedings of the Third International Conference BOINC:FAST 2017, Petrozavodsk, Russia, August 28 - September 01, 2017, published at http://ceur-ws.org

financial capacities organization may afford [1]. An interesting research was done by [3]. It revealed that small and medium enterprises (SMEs) feel the service provided by cloud computing companies is more secure than the service that is provided in house. These factors altogether make cloud and public distributed computing a very interesting solution for SMEs.

Although, there are many available public distributed computing projects allowing to dedicate computational resources, there are no feasible applications allowing for SMEs to utilize their internal IT resources for their business needs. This paper will review the status of the field and will try to determine the work and research required to make public distributed computing a desired platform for business computational needs.

The rest of this paper is structured as follow. Section 2 presents a review of adoption issues in cloud computing, outlining the need for alternative solutions. Section 3 presents public distributed computing as an alternative solution to cloud computing. Section 4 explores BOINC framework for business applications to replace cloud computing. Finally, section 5 concludes this paper by summarizing the findings and presents directions for future work.

2 Cloud computing adoption issues

Data and system security has consistently been a major issue in IT. It becomes particularly serious in the cloud computing environment, since data is distributed among different machines including servers, PCs, various mobile devices such as wireless sensor networks and smart phones. Data security in cloud computing is more complicated than data security in the traditional information systems [2]. Traditional security mechanisms such as identity, authentication, and authorization are no longer enough for securing clouds in their current form. Research show that cloud computing presents an added level of risk because essential services are often outsourced to a third party, which makes it harder to maintain data security and privacy, support data and service availability, and demonstrate compliance. Although there are many benefits to adopting cloud computing, there are also some significant barriers to adoption. One of the most significant barriers is security, followed by issues regarding compliance, privacy and legal matters. Storage, virtualization, and networks are the biggest security concerns in cloud computing [4]. Emerging technologies like cloud of things (CoT) will create more business opportunities, making it bigger threat from the attackers [5], it is essential to secure data from any illegitimate user access or attack such as denial of service, modification and forgery of document [6]. Security, privacy and specially, identity protection becomes very important in hybrid clouds, where there is an essence of private and public clouds, used by businesses. To make the cloud computing be adopted by users and enterprise, the security concerns of users should be rectified first to make cloud environment trustworthy, which is the basic prerequisite to win confidence of users to adopt such technology [2] [4] [6]. Privacy issues differ according to different cloud scenarios and they can be divided as follows [2]:

1. issue controlling the data to avoid nefarious use and an unauthorized resale of the data;
2. issue avoiding data loss, leakage, and unauthorized modification or fabrication when replicating the data;
3. issue deciding which party is responsible for ensuring legal requirements for personal information;
4. issue deciding to what extent cloud subcontractors are involved in processing which can be properly identified, checked and ascertained.

The major issues in the cloud computing also include resource security, resource management, and resource monitoring. Due to the dynamics of the cloud environment there are currently no standards and regulations to deploy applications in the cloud, leading to a lack of standardization control in the cloud. Cloud computing can save an organizations time and money, but trusting the system is more important because the real asset of any organization is the data which they share in the cloud to use the needed services by putting it either directly in the relational database or eventually in a relational database through an application [2]. From the cloud consumers perspective, security is the major concern that hampers the adoption of the cloud computing model. Enterprises not only lose control of their IT assets by outsourcing security management to third parties, but also increase the probability of attacks by hosting their data on publicly available infrastructures. Vendor lock-in, availability of the systems, bandwidth, privacy and legal consequences are some other issues that are mentioned by cloud computing opponents [7].

This overview clearly shows, that companies and organizations tend to use cloud computing services if the data security issues poses no threat. However, even then organizations cannot be sure their data is not being misused. As cited in [5], on Feb 01, 2013, it was read on The Independent, stating, British internet users' personal information on major 'cloud' storage services can be spied upon routinely by US authorities. This suggests that sensitive or private data should be stored in a virtual storage server located inside the users country or trusted geographical domain, which can be a friendly country as well [5].

3 Public distributed computing

Public distributed computing is a computational method that uses multiple public computers in parallel. The communications are carried over the network using client-server architecture, where client nodes offer their resources by requesting tasks from server. Computations are usually done in parallel without affecting one another. The main purpose of distributed computing system is to join the computational resources into a dynamic open network. This is an efficient way of using computational resources that potential otherwise would not be fully utilized. One of the advantages of such approach over cloud computing is that the infrastructure required for public distributed computing may be fully stored inside the bounds of organizations IT infrastructure and be fully managed by it. This way it could potentially lower the expenses by liberating the companies from the need to buy cloud computing services. Furthermore, it saves from some of the data security issues that occur in cloud computing approach.

The Berkeley Open Infrastructure for Network Computing (BOINC) is an open-source middleware system that supports volunteer and grid computing and is a good example of such computational method. Resource demanding tasks are divided into smaller ones that are then distributed among many computers. The results are then aggregated into the main solution. Client-server based architecture is used for this purpose where client nodes requests server for data. The computational results are then sent back to server. Computing is performed without inner process communications and only common database is used. Couple of more well-known projects are:

1. CERN + KC Gigabit Computing Challenge (https://cernkcchallenge.github.io/CernKCChallenge/);

2. Gridcoin (http://gridcoin.us/);

3. SETI@home (http://setiathome.ssl.berkeley.edu/).

Next, we will explore existing BOINC framework solutions as alternatives to cloud computing solutions from performance and quality of service stand point perspective.

4 Cloud computing and big data mining using BOINC

Approach utilizing BOINC for big data mining has been well researched in [8]. Ad hoc cloud computing platform has been developed that deploys a cloud service upon an end-users existing infrastructure where member hosts are sporadically available and used for some other primary purpose. The ad hoc cloud concept is useful for those who wish to improve their infrastructure efficiency and utilization as well as reduce costs by improving their return on IT investments. Furthermore, those who are not able to or do not wish to migrate to the commercial or private cloud models can experiment and explore the potential of ad hoc clouds before adopting either of the commercial or private models. Ad hoc clouds harvest resources from existing sporadically available, nonexclusive (i.e. primarily used for some other purpose) and unreliable infrastructures. Examples of such infrastructures range from personal infrastructure users with several underutilized computers, to startup companies through to large-scale organizational infrastructures. Research results from [8] show that the concept of ad hoc cloud computing is feasible and based on their initial evaluation, can be reliable and offer comparable performance to Amazon EC2.

Furthermore, [1] integrated tools like RapidMiner for data mining tasks to BOINC. Each node uses the BOINC client to pull computing tasks and data packages from a central server. Then, the BOINC client starts an instance of the RapidMiner framework to process the data mining tasks. The data mining results are sent back to a central project server, which gathers the information and provides it to researchers for further analysis.

This clearly shows that public distributed computing approaches may compete with existing cloud computing solutions. In the next two subsections, we are going to review the advantages and issues in adopting such approach.

4.1 Advantages

The distributed computing platform utilizing BOINC framework allows for any organization to interconnect owned computational resources into dynamic size computation platform that would perform given computational tasks not disturbing the ongoing work for employees. Such solution solves not only the computational resource demand problem, but also solves data confidentiality problem, since all the computations are done within organization. Organizations would be able to set desired data access levels if needed. Furthermore, such platform has a potential to increase quality of service the organization provides by allowing to run solutions adapted to business needs. Finally, it can reduce the service costs.

To support resource availability and cost reduction claims, a short 28-day experiment was run on two randomly picked computers, each residing in different organization (organization A and B). Each computer had BOINC client installed and ran SETI@home project for two weeks. During this time, power consumption and CPU idle time measurements were taken using Performance monitor (application available on Microsoft Windows) and electronic energy meter. Then the process was repeated without any BOINC projects running. During the experiments, both computers were used by employees for work related tasks. As shown in Table 1, Fig. 1 and Fig. 2, computers are performing very little computations and are wasting resources that they have.

Table 1: Resource and power consumption statistics.

Organization	BOINC project	CPU average idle time	Power consumption
A	Not running	98,77%	16,61 kWh
A	SETI@home	65,23%	22,03 kWh
B	Not running	83,49%	1 kWh
B	SETI@home	26,86%	2,09 kWh

Since only a small set of computers was used to gather data, a follow-up research involving more computers should be conducted. Despite this, gathered results already suggests that the infrastructure can be used for additional computations at a very low cost without interrupting the ongoing work processes. Since resources available from employee work computers are used only in small scale, such infrastructure can be used for additional tasks like big data analysis using public distributed computing platform.

Next, we are going to review the problems adopting this technology.

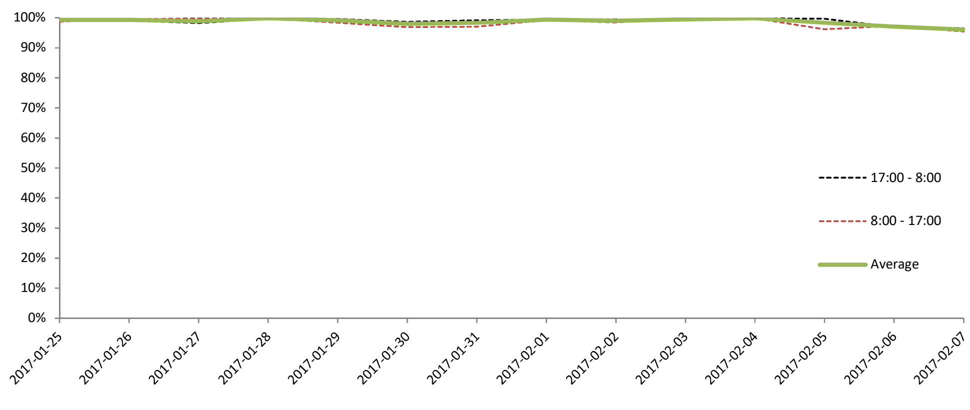

Figure 1: Processor idle time with no BOINC projects running during work and after work hours in Organization A.

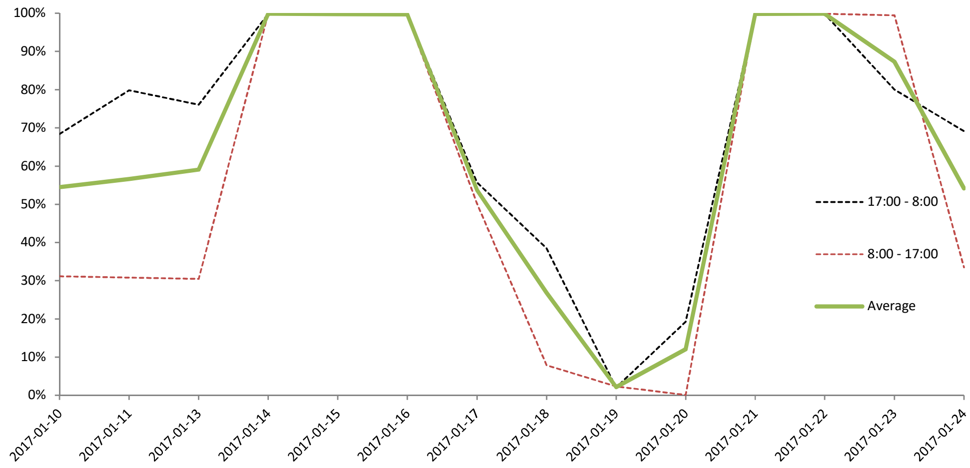

Figure 2: Processor idle time with SETI@home project running during work and after work hours in Organization A.

4.1.1 BOINC adoption issues and future work

Unlike cloud computing, service the public distributed computing provides is not as reliable. Clearly this is due to the nature of the model, since nodes can become unavailable at any moment in time and there is no way of knowing how many computational resources will be available. At any moment, any type and amount of data can be triggered. It may also be an emergency data as well. As a result, there is no way of determining how long it will take for a given task to complete. It is also unknown if the resources available at a time is even enough for a job to complete in a reasonable amount of time. This approach in strict business environment where project timelines are set is not acceptable. Depending upon the type of data and its urgency to be sent to the node, QoS (Quality of Service) must be supported. One way to solve this issue is to create a mathematical model that would determine the time and the costs required for the job. This can be an interesting case for future research.

Methods to improve QoS has been researched in [8]. Proposed method takes advantage of Virtual Machines. Their solution is called V-BOINC. This approach solves the task continuity problem. The V-BOINC server sends the virtual machine image and a script that configures it (e.g. sets CPUs, memory and disk space limits) to the V-BOINC client. The virtual machine is then configured and started to allow it to request, receive BOINC jobs and return job results. However, this raises a security concern. According to [7], VMs are still under risk even when they are offline. VM images can be compromised by injecting malicious codes in the VM. Another issue related to VM templates is that such templates may retain the original owner information which may be used by a new consumer.

To make distributed public computing adopted by users and enterprises, the security concerns should be rectified first to make the environment trustworthy same as in cloud computing. Trustworthy environment is the basic prerequisite to win confidence of users to adopt such technology. It is essential to secure data from any illegitimate user access or any other attack. According to [6], there are many cryptographic algorithms that can be deployed over the cloud to provide the security. We believe same approach can be applied to distributed public computing as well. Essential characteristics of cloud computing are: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service [6]. All of this is already made available by BOINC, except for measured service. This is one of the areas we believe requires additional research.

From [6] and other related researches we clearly see that cloud and distributed public computing share a lot of security issues. Basic security issues in cloud computing are availability, data and system integrity, authentication, data recovery, data confidentiality, privacy and access control. Same issues apply and needs to be solved in distributed public computing models. This would make distributed public computing a feasible alternative to cloud computing in business environment.

Furthermore, according to [9], there are many other factors influencing the decision to adopt the technology: environmental, organizational, managerial and technological. This paper only reviewed the technological factors influencing the decision to adopt distributed public computing. Other adoption issues are also important research topics.

5 Conclusion

In this paper, we have reviewed a use case of distributed public computing model use in business environment. We have examined the impact of security and reliability issues remaining open in public distributed computing models and in contrast showed how these problems have been solved in case of cloud computing solutions. Solving these issues would open new business and research opportunities making public distributed computing solutions like BOINC a great asset to any organization wanting to process large amounts of data. Our research showed that the solutions for high computational resources demanding tasks are already available, however, we conclude that until data security and service reliability issues are solved, no business will consider taking this approach. While we think that reliability issue cannot be completely eradicated, we believe new research of stochastic models minimizing unexpected costs can mitigate its impact.

The focus of our ongoing research is to tackle the challenge of distributed public computing model adoption in business environment. We consider investigating mathematical approaches that would help increase service reliability and make BOINC like solutions suitable for business.

References

[1] N. Schlitter, J. Laessig. Distributed Data Analytics using RapidMiner and BOINC RapidMiner. *RapidMiner Community Meeting and Conference*, RCOMM 2013, pp. 81-95, Porto, Portugal (2013).

[2] Y. Sun, J. Zhang, Y. Xiong, G. Zhu. Data Security and Privacy in Cloud Computing. *International Journal of Distributed Sensor Networks 10(7)*, 1-9 (2014).

[3] Y. Lu, X. Xu, J. Xu. Development of a Hybrid Manufacturing Cloud. *Journal of Manufacturing Systems 33(4)*, 551566 (2014).

[4] K. Hashizume, D. Rosado, E. Fernandez-Medina, E. Fernandez. An analysis of security issues for cloud computing. *Journal of Internet Services and Applications 4(1)*, 38-46 (2013).

[5] M. Aazam, P. Hung, E. Huh. Cloud of Things: Integrating Internet of Things with Cloud Computing and the Issues Involved. *11th International Bhurban Conference on Applied Sciences and Technology*, IBCAST, pp. 414-419. Islamabad, Pakistan (2014).

[6] N. Mishra, S. Siddiqui, J. Tripathi. A Compendium Over Cloud Computing Cryptographic Algorithms and Security Issues. *International Journal of Information Technology 7(1)*, 810-814 (2015).

[7] M. Morsky, J. Grundy, I. Mller. An Analysis of the Cloud Computing Security Problem. *17th Asia-Pacific Software Engineering Conference*, APSEC. Sydney, Australia (2010).

[8] G. McGilvary, A. Barker, M. Atkinson. Ad hoc Cloud Computing: From Concept to Realization. *IEEE 8th International Conference*, CLOUD. (2015).

[9] S. Tehrani, F. Shirazi. Factors Influencing the Adoption of Cloud Computing by Small and Medium-Sized Enterprises (SMEs). *International Conference on Human Interface and the Management of Informatio*, vol. 8522, pp. 631-642 (2014).