Towards Performance Evaluation in Volunteer Clouds

Abdulelah Alwabel, Robert Walters and Gary Wills

Electronics and Computer Science School,
University of Southampton

{aa1a10,rjw1,gbw}@ecs.soton.ac.uk

Abstract. Cloud computing has emerged as a new paradigm that promises to reduce costs of IT by allowing customers to harness computing resources and pay for their usage only. However, cloud computing can still be costly for some projects such as scientific projects. Therefore, volunteer cloud model appears with a goal to provide cloud services at a little cost, if not free. Volunteer clouds aim at providing cloud capabilities out of non-dedicated resources such as normal PCs based on the cloud business model. However, volunteer clouds present numerous challenges that need to be tackled before it can be seen as a viable solution. The performance of services provided by volunteer clouds is a major issue in this context. This report presents our research problem, motivation and work progress in a PhD research.

Keywords: cloud, volunteer cloud computing, cloud architecture

1 Introduction

Cloud computing represents a shift away from computing being purchased as a product to be a service delivered over the Internet to customers. Economic benefits are the key role behind the appearance of cloud computing [1]. The Cloud transforms IT assets from being capital expenditure to be operational expenditure. Traditionally, small and medium enterprises obtain IT infrastructure by purchasing it. In the cloud, using a server for five hours costs the same as using five servers for an hour [2]. However, the cost of consuming services can be an obstacle against moving toward the cloud. For example, some organisations cannot afford the charge of cloud services in the long term. Therefore a new type of cloud has emerged recently to overcome this limitation. The new type provides clouds' capabilities based on non-dedicated resources. The new type can be called volunteer or non-dedicated clouds.

Volunteer cloud computing (VCC) can be an alternative choice to the current version of clouds which are offered by commercial companies such as Google and Amazon (to be called commercial clouds). VCCs aim at providing cloud capabilities at no or low costs by harnessing idle resources that are contributed by the public. However, VCC is in its infancy level and suffers from some issues that need attention before it can become a viable solution. This report addresses the evaluation of performance of

adfa, p. 1, 2011.

VCC service as being a research challenge. In addition, our motivation to participate in this context is discussed. Furthermore, the paper presents a brief background about volunteer clouds and related works. We present architecture for volunteer clouds that can help in solving some issues in volunteer clouds. Our conclusion and future work are presented at the end of this report.

1.1 Research Problem

The services provided by volunteer clouds is expected to be low [3] compared to that provided by commercial clouds due to the nature of the underline infrastructure. This requires a way to evaluate VCC performance in order to be able to enhance it in the future since it has been shown that performance is one of the major concerns in the cloud. However, the literature shows very little work has been conducted in this area. In our research, the intention is to find a technique to measure performance in VCCs. The contribution that we hope to produce by the end of this research as follows:

- Define performance metrics in volunteer clouds: The metrics can be gathered from related computing models such as cloud computing and Grid computing. For example, some performance metrics: response time, resource utilisation and scalability. However, the nature of the used resources to from a volunteer cloud may require new metrics.
- Evaluation of performance provided in volunteer clouds is quite vital. This can be implemented in a tool. The research can lead to present performance benchmarks in volunteer clouds.
- We will try to find a way to predict the performance for each task in a particular volunteer cloud. We will conduct a study to examine whether this tool can apply to commercial clouds.

1.2 Motivation

We are motivated by the fact that building a cloud out of non-dedicated resources can serve the research community in producing a better quality of research. Indeed this allows them to benefit from the cloud services with little or no cost. Furthermore, volunteer clouds can help in preserving and reducing gas emission because they can be formed without building new data centres which have a negative impact on the environment. In addition, volunteer clouds utilise idle resources which means they benefit from those idle resources, provided that approximately 80% of resources in organisations remains idle most of the time according to [3]. Finally, the rapid growth of devices connected to the Internet can constitute a viable way for cloud service providers to exploit them in case their data centres reach their limits.

2 Background

Volunteer cloud computing is based on merging two computing concepts: cloud computing and volunteer computing. Cloud computing is a new computing model that offers shared resources to be accessed online on a pay-as-you-go basis. Volunteer computing is a term which means offering computing resources, such as processing power, to be used by others on a voluntary basis [5]. Volunteer clouds, however, are not limited to resources that are denoted but rather it is more generic to involve any non-dedicated resources used to from a cloud. For example, an organisation may wish to use their local infrastructure to build a cloud.

Nebula [6] is a project aiming to exploit distributed resources in order to create a volunteer cloud which offers services free of charge. Weissman et al. evaluate the performance of Nebula vs. commercial clouds in dealing with highly distributed data-applications [7]. The paper starts by presenting the architecture of Nebula. The model consists of a master node, data nodes, execution nodes and a database. The master node forms the interface between Nebula and its clients and is responsible for managing and assigning tasks to nodes. Data nodes store the data that will be processed by execution nodes. Execution nodes process tasks from the master node and retrieve data from data nodes. The database maintains a list containing information about the master node, execution nodes and data nodes. The list is updated frequently in order to add new or remove unavailable nodes. Experiments were conducted to evaluate the performance of both Nebula-like (a prototype on Planet-Lab [8] simulates Nebula) and cloud-like (a prototype simulates commercial clouds such as Amazon EC2). They used a distributed blog analysis application for their experiment. Overall, the results show that Nebula was better in terms of performance, even given node failure in Nebula, which is highly likely to happen in reality.

Cloud@home is a project representing the @home philosophy in cloud computing [3]. The goal of Cloud@home is to form a new model of cloud computing contributed by individual users. That means using clouds' clients to participate in building clouds. However, some challenges remain to be solved. The first is that it requires a managing mechanism for services and resources. The management phase involves Quality of Service and Service Level Agreements (SLAs) for the provided services. Secondly, an interface is required between the participants and the underlying resources in the cloud. Furthermore, security is a major issue in cloud computing and it has more impact in volunteer clouds since it must prevent local access. In addition, volunteer clouds require means to interact with other clouds for data migration or to gain extra computing resources. Finally, it is necessary to implement a resource replication mechanism in order to maintain an acceptable level of reliability. The architecture of Cloud@home consists of frontend, virtual and physical layers [9]. The frontend layer represents the interface between clients and Cloud@home. The virtual layer is responsible for providing the frontend layer with virtual machines for execution and storage services by virtualised the heterogeneous physical resources. The physical layer is the group of available resources volunteered by contributors.

3 The Architecture

This section proposes our architecture for VCC. The architecture can be employed in order to overcome issues in VCCs. The most abstract level of the architecture, **Fig. 1**, divides VCC into three layers in a way similar to cloud@home: (i) a service layer; (ii) a middleware layer; and (iii) a physical layer. We argue that the cloud@home architecture is not detailed enough to cope with some issues in volunteer clouds, performance issues for example.

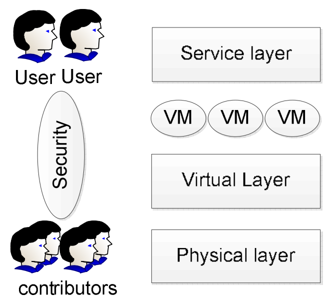

Fig. 1. A proposed Architecture

3.1 Service Layer

The service layer is a layer providing services via an interface to customers based on SOA approach. The business model in VCC is similar to that of commercial clouds, which aim to provide metered computing power as a customer desires. VCC's contributors volunteer their resources to form a VCC for a certain time, and they may be services consumers at the same time if they wish.

Fig. 2. Middleware Layer

3.2 Middleware Layer

This layer represents the core structure of the VCC architecture. The aim of the layer is to provide resources to the service layer as they would be provided by a commercial cloud. The layer, shown in **Fig. 2**, consists of *task management* and *QoS management*. *Task management* works with tasks received from the service layer. It involves *task scheduler*, *load balancing* and *self-automation*. The task *scheduler* organises tasks coming from the service layer by passing them to suitable resources. Resources are offered by the *resource manager* in the physical layer. *The load balancing* ensures that the load is distributed appropriately, thus minimizing the required time to process a task. *Self-automation* helps to provide the rapid elasticity in VCCs. It allows users to scale services up or down according to their needs. *Quality of service management* ensures that a minimum quality level is maintained. The *performance monitor* in *QoS management* ensures that the performance of each task is maintained at an acceptable level which is reported in the *service level agreement (SLA) reporting* component. Node volatility is quite high in VCCs, so the *performance monitor* must cooperate with the *resource management* to find reliable nodes among available resources that suite each task. The *fault recovery* component can be vital with regards to improving the performance of the overall VCCs. *Fault recovery* can employ a number of techniques to improve the availability level.

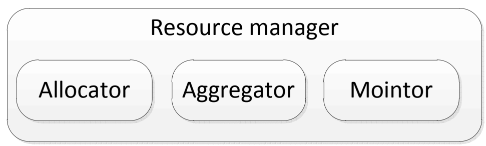

Fig. 3. Physical Layer

3.3 Physical Layer

The physical layer contains nodes contributed by the public. However, the contributors can only be from sources that are trustworthy; predefined organisations for example, according to administrator preferences. The physical layer, **Fig. 3**, is responsible of managing resources. *Resource manager* is responsible for resource aggregation, resource allocation and resource monitoring. Resource aggregation aggregates volunteer nodes from the public. An aggregation mechanism can classify resources according to a number of criteria with the aim of optimising the quality of service. For example, the history of each volunteer node can be useful in terms of recognising which node should be selected by the resource allocator for each task. The *resource allocator* receives tasks from tasks management and allocates them to the required resources. The *allocator* can decrease the interruption of services by assigning tasks to nodes with higher reliability. It can choose nodes with low reliability to be replicated nodes. However, this requires a technique to compute reliability for each volunteered node in volunteer clouds. The *monitor* component observes allocated resources regularly in case any of them becomes unavailable. In this case, the *monitor* informs the *fault recovery* in the middleware in order to recover the task from a replicated node.

4 Conclusion and Future Work

In conclusion, this report presented volunteer clouds as a new type of cloud computing based on infrastructure that is made out of non-dedicated resources. The infrastructure of volunteer clouds is made of unreliable nodes that can join or leave the cloud without prior notice. The intention is that this research will lead to present performance benchmarks which require a list of metrics in order to evaluate and predicate the performance of volunteer clouds. As a first step in this research, we proposed a VCC architecture which can help us in solving the research problem. Our future plan involves: Implement the architecture in order to evaluate it. Then compare the architecture with other available architecture used in commercial clouds. We will define a list of performance attributes for VCCs and implement them in a VCC evaluation system. Finally, based on our results, we will design an algorithm to predict the performance of each task in volunteer clouds.

References

1. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems. 25, 599-616 (2009).
2. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., others: A view of cloud computing. Communications of the ACM. 53, 50–58 (2010).
3. Cunsolo, V.D., Distefano, S., Puliafito, A., Scarpa, M.: Volunteer computing and desktop cloud: The cloud@ home paradigm. Network Computing and Applications, 2009. NCA 2009. Eighth IEEE International Symposium on. pp. 134–139. IEEE (2009).
4. Arpaci, R.H., Dusseau, A.C., Vahdat, A.M., Liu, L.T., Anderson, T.E., Patterson, D.A.: The Interaction of Parallel and Sequential Workloads on a Network of Workstations. ACM (1995).
5. Anderson, D.P., Fedak, G.: The Computational and Storage Potential of Volunteer Computing. Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06). 73-80 (2006).
6. Chandra, A., Weissman, J.: Nebulas: Using distributed voluntary resources to build clouds. Proceedings of the 2009 conference on Hot topics in cloud computing. pp. 2–2. USENIX Association (2009).
7. Weissman, J.B., Sundarrajan, P., Gupta, A., Ryden, M., Nair, R., Chandra, A.: Early experience with the distributed nebula cloud. Proceedings of the fourth international workshop on Data-intensive distributed computing. pp. 17–26. ACM (2011).
8. Chun, B., Culler, D., Roscoe, T., Bavier, A., Peterson, L., Wawrzoniak, M., Bowman, M.: PlanetLab: An Overlay Testbed for. Broad-Coverage Services. ACM SIGCOMM Computer Communication Review. 33, 3–12 (2003).
9. Cunsolo, V.D., Distefano, S., Puliafito, A., Scarpa, M.: Applying Software Engineering Principles for Designing Cloud@Home. 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. 618-624 (2010).