Managed Control of Composite Cloud Systems

Christopher Lamb **Pramod Jamkhedkar** **Greg Heileman**
Department of Electrical and Computer Engineering
University of New Mexico
{cclamb, pramod54, heileman}@ece.unm.edu

Abstract - *Cloud providers have just begun to provide primitive functionality enabling users to configure and easily provision resources, and primarily in the Infrastructure as a service domain at that. In order to effectively manage cloud resources in an automated fashion, systems must automate quality of service metric measurement as a part of a larger usage management strategy. Collected metrics can then be used within control loops to manage and provision cloud resources when needed.*

Keywords: Usage Management, Cloud Computing, System of Systems.

1 Introduction

Cloud computing services as a computational paradigm are more market oriented than previous attempts at commodity computing. Furthermore, they are in many cases designed to be composed into larger, more powerful customer facing systems. These kinds of aggregate systems fit neatly into one of the more commonly used definitions of a system of systems as well [1], [2]. With so much data in the hands of different providers in an aggregate system, system developers and users are hard-pressed to effectively monitor and control the use of sensitive content by various composite systems. Some of this information can be contained in Service Level Agreements (SLAs), but they have thus far been focused on quality-of-service (QoS) metrics rather than addressing issues like data flow or physical application residency. For the most part SLAs are simply not sufficient for addressing usage management concerns [3], [4], [5], [6].

Effective usage management monitoring coupled with feedback processing creates an event loop suitable for applying control theoretic concepts to cloud infrastructures.

The ability to express fine-grained usage policies will provide cloud users a greater trust regarding the usage of their data within clouds, and will provide them with confidence to employ cloud services. A mechanism to interpret, reason over, and enforce usage policies within clouds in an automated manner will allow cloud users to optimize resource allocation while ensuring safety and security of users data. A correct estimation of the cost involved in enforcing usage policies will further enable cloud providers business opportunities such as yield management and price differentiation.

This paper introduces the notion of usage management in cloud computing and provides an in-depth analysis of the challenges and principles involved in the design of an open, interoperable usage management framework that operates over a distributed cloud computing environment. The analysis includes application of well-known principles of system design and standards [7, 8, 9], research developments in the areas of usage control [10, 11], policy languages design principles [12], digital rights management (DRM) systems [13], and interoperability [14, 15, 16, 17, 18] towards the development of such a framework.

The paper delves into how such a framework will improve upon the status quo by leveraging existing security mechanisms, and enable automated reasoning of policies with respect to the underlying cloud security infrastructure. Based on these analyses, a system for expression and reasoning of usage management within a distributed cloud environment is proposed that makes use of a common cloud ontology shared by different cloud services. The proposed system architecture enables usage policies to be expressed in different policy languages, and interpreted across various cloud environments.

1.1 Previous Work

In the recent years, cloud computing has managed to emerge as a computing platform that allows computing services to be consumed as a utility by consumers. In cloud computing, applications, systems software, and hardware are offered as utility services to consumers over the Internet. In service-based architectures, service consumers need to be provided with highly reliable services that meets their expectations. The consumers indicate these expectations in terms of QoS parameters that are expressed in the form of

an SLA negotiated with the service provider.

In the realm of computing, there exist multiple service-based paradigms such as web-services, cluster computing, grid computing and cloud computing [19]. Cloud computing is set apart from the other forms of service-based computing paradigms by the collective set of distinguishing characteristics such as market orientation, visualization, dynamic provisioning of resources and service composition via multiple service providers [20]. These characteristics imply that in cloud computing, data and application of a cloud-service consumer reside inside the cloud for a finite amount of time; fractions of this data are handled by multiple cloud services; and these data fractions may be stored, processed and routed across a geographically distributed cloud infrastructure. These activities occur "behind-the-scenes", within the cloud, while giving the cloud consumer an impression of a single virtual machine. These operational characteristics of cloud computing raise serious concerns regarding the manner in which data and application of cloud consumers are handled or used within the cloud. Unlike other computing paradigms that focus on specific computing tasks, cloud computing enables cloud consumers to host entire applications on the cloud. As consumers aggressively start exploiting this advantage to "outsource" their IT services to the cloud, the manner in which data and application is handled within the cloud by various cloud services will become a matter of serious concern.

The handling or use of consumer data within the cloud by different services refers to the policies regarding constraints under which different actions may be carried out on the data. A cloud consumer might want to limit the way in which data is stored, routed, or processed, and specify who is authorized to carry out these activities and under what conditions. As an example, a government agency might want to prevent the data from being stored in one particular country, or prevent the routing of its data via a particular set of networks it considers unreliable. Similarly, a financial company might want to prevent its data from being processed by a particular cloud service, or may want it to be encrypted before being stored by an untrusted cloud storage service. Usage policies typically consists of a range of semantics such as restrictions on the manner in which data is used, temporal restrictions on usage, spacial or attribute-based restrictions, permissions, obligations, penalties, count-based limits on usage and partial dependencies to name a few. Hence, as cloud services become pervasive, cloud consumers will want to dictate the terms of usage for their data and applications within the cloud in a manner that is expressive enough to capture the concerns of cloud users. At present, cloud providers enable these features via rudimentary techniques by offering a one-size-fits-all options to consumers. For example, Amazon S3 storage services allow a region-based facility where data stored in one region is guaranteed not to leave that particular region [21]. Future cloud computing services will need to enable these terms to be expressed in SLAs in a sophisticated, fine-grained manner that will allow the cloud consumers to express the usage terms explicitly.

As mentioned earlier, service expectations in service-based computing paradigms are expressed in terms of SLAs negotiated with the service provider. For different types of service-based computing paradigms such as web services, clusters, and grid computing, there exist well established SLA frameworks that enable expression, interpretation, monitoring, control and enforcement of SLA terms [3, 4, 5, 6]. However, the SLAs supported by these frameworks focus on performance metrics such as availability, reliability, bandwidth, response times, instructions per second, etc. [22]. Also, the privacy and security metrics supported by these frameworks focus primarily on encryption of the data. It must be noted that usage policies are significantly different than these metrics and present a new dimension that is orthogonal to the manner in which the other metrics are managed. It is therefore necessary to have a separate mechanism for usage management in cloud computing environments. The next section discusses the characteristic features of such a framework.

2 Conclusion

In this paper we introduced the notion of usage management in cloud computing environment. Cloud computing exhibits a unique set of characteristics that will require usage management of users' data according to user concerns and expectations. We analyzed the challenges involved in the design and development of a framework for usage management in cloud environments. We showed that such a framework needs to be open to leverage existing security technologies and SLA frameworks. The framework needs to exhibit features such as support for multiple policy languages, existence of a common cloud ontology, dynamic interpretation and data transformations. Finally a preliminary framework that supports multiple policy languages was introduced that will provide a platform upon which such a framework can be built.

Future works involves efforts towards the development of a common extensible cloud ontology that will provide the vocabulary underlying this framework and enable interoperability. Such an ontology needs be developed by taking into consideration the requirements of different cloud services. In addition, it is necessary to standardize interfaces for other security mechanisms mentioned in the paper that can be incorporated within the framework. Finally, the framework allows use of multiple policy languages, to this effect, existing policy languages can be modified to be incorporated within the framework, and news ones need to designed to

address the specific needs of different cloud services.

References

[1] A. P. Sage and C. D. Cuppan, "On the systems engineering and management of systems of systems and federations of systems," *Inf. Knowl. Syst. Manag.*, vol. 2, pp. 325–345, December 2001. [Online]. Available: http://portal.acm.org/citation.cfm?id=1234195.1234200

[2] "IEEE SMC 2005 Conference Theme," http://ieeesmc2005.unm.edu/SoSE_Defn.htm, March 2011.

[3] *Web Services Agreement Specification*, Open Grid Forum, Mar. 2007. [Online]. Available: http://www.ogf.org/documents/GFD.107.pdf

[4] *Web Service Level Agreement (WSLA) Language Specification*, IBM, Jan. 2003. [Online]. Available: http://www.research.ibm.com/wsla/WSLASpecV1-20030128.pdf

[5] *Web Services Policy 1.5 Framework*, World Wide Web Consortium, Sep. 2007. [Online]. Available: http://www.w3.org/TR/2007/REC-ws-policy-20070904/

[6] P. Patel, A. Ranabahu, and A. Sheth, "Service Level Agreement in Cloud Computing," in *Proceedings of the Workshop on Best Practices in Cloud Computing: Implementation and Operational Implications for the Cloud at ACM International Conference on Object-Oriented Programming, Systems, Languages, and Applications*, Orlando, FL, Oct. 2009.

[7] M. S. Blumenthal and D. D. Clark, "Rethinking the design of the Internet: The end-to-end arguments vs. the brave new world," *ACM Transactions on Internet Technology*, vol. 1, no. 1, pp. 70–109, Aug. 2001.

[8] D. D. Clark, "The design philosophy of the DARPA internet protocols," in *ACM SIGCOMM*, Stanford, CA, Aug. 1988, pp. 106–114.

[9] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden, "Tussle in cyberspace: Defining tomorrow's internet," in *SIGCOMM*, Pittsburg, Pennsylvania, USA, Aug. 2002, pp. 347–356.

[10] J. Park and R. Sandhu, "The UCON$_{ABC}$ usage control model," *ACM Trans. Inf. Syst. Secur.*, vol. 7, no. 1, pp. 128–174, 2004.

[11] P. A. Jamkhedkar, G. L. Heileman, and C. Lamb, "An Interoperable Usage Management Framework," in *Proceedings of the Tenth ACM Workshop on Digital Rights Management*, Chicago, Oct. 2010.

[12] P. A. Jamkhedkar, G. L. Heileman, and I. Martinez-Ortiz, "The problem with rights expression languages," in *Proceedings of the Sixth ACM Workshop on Digital Rights Management*, Alexandria, VA, Nov. 2006, pp. 59–67.

[13] P. A. Jamkhedkar and G. L. Heileman, "Digital Rights Management Architectures," *Computers Electrical Engineering*, vol. 35, no. 2, pp. 376–394, 2009.

[14] ——, "DRM as a Layered System," in *Proceedings of the Fourth ACM Workshop on Digital Rights Management*, Washington, DC, Oct. 2004, pp. 11–21.

[15] G. L. Heileman and P. A. Jamkhedkar, "DRM interoperability analysis from the perspective of a layered framework," in *Proceedings of the Fifth ACM Workshop on Digital Rights Management*, Alexandria, VA, Nov. 2005, pp. 17–26.

[16] R. H. Koenen, J. Lacy, M. MacKay, and S. Mitchell, "The long march to interoperable digital rights management." *Proceedings of the IEEE*, vol. 92, no. 6, pp. 883–897, 2004.

[17] "Coral consortium whitepaper," Tech. Rep., Feb. 2006. [Online]. Available: www.coral-interop.org/main/news/Coral.whitepaper.pdf

[18] "Marlin architecture overview," Tech. Rep., 2006. [Online]. Available: www.marlin-community.com

[19] R. Buyya, "Market-oriented cloud computing: Vision, hype, and reality of delivering computing as the 5th utility," in *Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, ser. CCGRID '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 1–. [Online]. Available: http://dx.doi.org/10.1109/CCGRID.2009.97

[20] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009.

[21] "Amazon Web Services: Overview of Security Processes," Amazon, Sep. 2008. [Online]. Available: http://s3.amazonaws.com/aws_blog/AWS_Security_Whitepaper_2008_0

[22] A. Paschke and E. Schnappinger-Gerull, “A Categorization Scheme for SLA Metrics,” in *Proceedings of Multi-Conference Information Systems (MKWI06)*, Feb. 2006.