Cloud Slicing
A new Architecture for Cloud Security Monitoring

Abdulaziz Aldribi, Issa Traore
The Department of Electrical and Computer Engineering
University of Victoria
Victoria, Canada
Email: aaldribi@uvic.ca , itraore@ece.uvic.ca
Website:http://www.uvic.ca/engineering/ece/isot/

Gabriel Letourneau
Department of Computer Science
University of Victoria
Victoria, Canada
Email: gletour85@gmail.com

***Abstract*—Cloud computing has become one of the popular terms in academia and IT industry. The security of the cloud computing infrastructure is the main concern to adopt it. Despite this concern, cloud providers do not disclose any information about their security precautions. Therefore, a provider's clients cannot be certain that their applications are protected while they are in cloud. Furthermore, clients are not granted access to the network level of the system to implement any of their own security features.**

In this paper we propose a new model we are naming Cloud Slicing. Cloud Slicing uses a technique called logical partitioning, to divide a cloud servers resources. By doing this division, clients on a server can safely implement their own network security features to reassure themselves, and their customers, that the applications are protected.

I. Introduction

Cloud computing solutions are rapidly increasing in popularity within the IT industry. After all, the cloud model offers such enticing features such as greater scalability, cost reduction, and flexibility. However, the concern of this paper is not about what cloud 'providers' (companies who sell cloud services) can offer, but what concerns 'clients' (companies that purchase cloud services) about moving to the cloud. An enterprise panel survey by IDC (International Data Corporation) in 2008, showed that security was the biggest challenge and issue with cloud computing [1]. This (security) concern can lead to hesitation in a cloud provider's potential clients, especially if a client is dealing with sensitive information.

If a provider is unwilling to disclose information about their security features, we propose allowing clients to deploy their own network level security. Unfortunately, the current cloud model does not allow clients any network level access [2] (see section 3). Therefore, there is a fundamental relationship between the decision making of stakeholders and security transparency problems. In order to facilitate security transparency, we introduce in this paper an alternative cloud model named *Cloud Slicing*. Cloud slicing uses hardware partitioning (sometimes named logical partitioning) to allow a client access to the network level of the server. With network level access, clients can implement further protection measures to give assurance to themselves and their customers.

978-1-4799-7492-4/15/$31.00 ©2015 IEEE

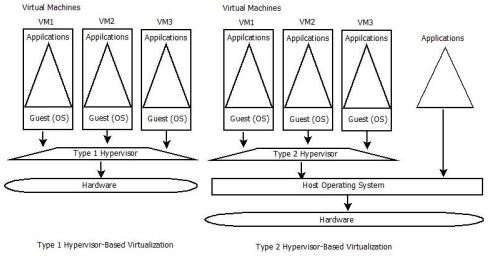

Fig. 1. The current model used in cloud computing today. Both Type 1 and Type 2 use hypervisor which spawns multiple VM instances.

The rest of the paper is sectioned as follows: Section II discusses the current cloud model in more detail, including the main motivations behind its design. The strengths and perceived limitations of our model are presented in Section III, along with a discussion of our approach to address these potential weaknesses. Finally, the paper's conclusion and future work are provided in Section IV.

II. The Current Cloud Computing Model

Cloud computing is typically offered in three different provision models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS).These three models are the basis of all services provided by cloud computing. We will be focusing on only one of these three models, namely, IaaS. When a provider offers a potential client IaaS, they are offering the client a VM (Virtual Machine) on one of their servers. IaaS is the most basic service model and allows a client to use whatever operating system and configuration they would like. Today IaaS is given through a technique called "multitenancy", shown in Fig. 1, which uses virtualization technology to deploy multiple virtual platforms for each client. Cloud computing acquires some of its key characteristics from multitenancy: scalability, location independence, and resource pooling. In terms of security, the most important feature gained from multitenancy is isolation of the VMs. Successful isolation will give each client the illusion that they are on their very own server, despite sharing the server with others [3]. Isolating clients via virtual machines (VMs) is key to protecting them from not only the outside Internet but also from other clients. If

a client's VM is infected by a virus, or the client is malicious, the other VMs should remain unaffected. Multitenancy makes use of a Virtual Machine Monitor (VMM), commonly referred to as a hypervisor [4]. The hypervisor is a piece of software that deploys and manages all of the VMs and ensures isolation. Hypervisors usually sit on either the host OS or directly on the hardware. By creating a barrier between the hardware and the VMs, the hypervisor is able to filter network information such that each VM receives only the traffic associated with its applications. In the current IaaS model, resources are shared (pooled resources). If clients were granted access to the network level it would mean compromising the privacy of the other clients. Just like any piece of software on a network, the hypervisor itself is vulnerable to attack. Code errors in the hypervisor could allow attackers to compromise all VMs on that server [2]. The hypervisor is especially vulnerable to any malicious clients since it is constantly exchanging information between the hardware and the VMs. Malicious clients are especially threatening because not only are they in constant contact with the hypervisor, but they typically have a greater understanding of the system than an outside attacker would.

Cloud providers are in a constant struggle against new forms of attack from both outside and inside their servers [5]. This struggle makes the isolation of each VM a necessity for a provider and not a preference. However, isolating each VM comes at a price for the clients. In an isolated VM, a client cannot take precautions on a network level to safeguard their own applications. Our model, Cloud Slicing, maintains isolation while still allowing clients to implement their own, software-based, network security features.

III. OUR MODEL: CLOUD SLICING

Allowing clients to monitor the network level while maintaining each client's isolation cannot realistically be accomplished with shared resources. This conundrum has been the motivation for us to develop our new cloud model: ***Cloud Slicing***. Cloud slicing is a new cloud model that gives the client opportunity to fully monitor his cloud resources and traffics. Cloud slicing does not require VMs or a hypervisor; the clients can work directly off of the server's hardware or firmware. Despite doing away with VMs and a hypervisor, Cloud Slicing still maintains isolation between all clients. To accomplish this each client gets a finite portion of the server's hardware resources (processors, memory, etc) exclusively. This requires a technique called Logical Partitioning (LP), or Dynamic Hardware Partitioning (DHP) (Fig. 2).

A. Partitioning a Server

Logical partitioning is the ability to make a single server run as two or more independent servers by dividing its resources into subsets. This concept and technology is not new to the industry. In 1976 in New York, IBM began researching logical partitioning and in 1984 launched the IBM S/370, the first partitionable server [7]. Today, all major mainframe producers (including Microsoft, Intel, IBM, etc) produce and sell partitionable servers.

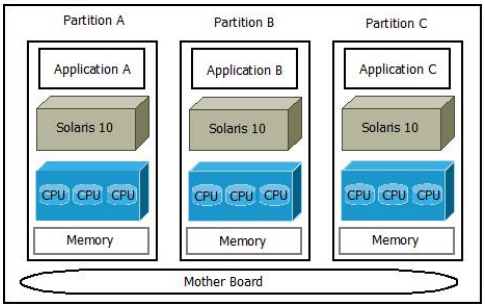

Fig. 2. Logical partitioning on a SPARC Enterprise M5000 [6].

1) The Servers: Currently, logical partitioning is used by organizations that wish to work on multiple projects and/or run multiple test environments all on the same server. To actually partition a server you must have the required hardware, which differs depending on which manufacturer the server is purchased from. For instance, IBM, currently has an entire series of servers that are partitionable known as z9 series.

An alternative to logically partitioning a big tower server, like the z9s, would be to buy multiple blade or rack servers. For blade servers, to achieve isolation similar to using logical partitions, a provider must allocate each client to a separate blade server [8]. This causes two obvious problems:

1) clients are restricted to the hardware specifications of the blade server they are hosted on and,
2) a z9 server can have up to 60 logical partitions: to achieve the same, a provider would need to purchase 60 blade servers.

A z9 server is not cheap, ranging anywhere from a hundred thousand dollars to a couple million [9]. In contrast, it would cost more to purchase 60 blade servers, ranging anywhere from fifteen thousand to a hundred and fifty thousand dollars for each blade server. Additionally, working with 60 blade servers is much more cumbersome than a single z9 server. Rack servers suffer from the same drawbacks except they are even harder to setup than blade servers [8]. Following the reasons above, we believe performing a logical partition on a single large server, like the ones in the z9 series, is the best choice for a provider. Nonetheless, as great as the z9 servers are by themselves, they only really show their versatility when coupled with a Hardware Management Console (HMC) [10].

2) The Hardware Management Console: : The HMC is the standard interface for configuring and operating all logical partitions on one or more z9 servers. IBM's HMC runs as an embedded application on an Intel based workstation that can be either a desktop or a rack mounted system (Fig. 3). This embedded application can be thought of as a conductor of multiple orchestras (servers) each performing multiple symphonies (logical partitions). From an HMC, a provider personnel is able to create and maintain logical partitions, display each partitions resources and status, power servers on and off, remote login, manage platform firmware, and act as a service focal point [11]. Some of the more advanced abilities include

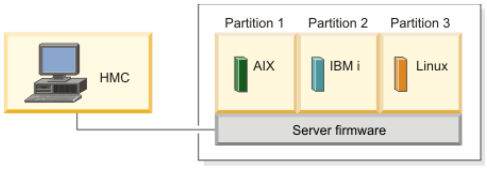

Fig. 3. Architecture of an HMC connected to a partitioned server which has a different operating system with each partition [12].

Capacity on Demand (CoD) and Dynamic Logical Partitioning (DLPAR). CoD allows CPU and memory resources to remain dormant and, if needed, activated without shutting the server down. This ability couples nicely with DLPAR, which allows resources to be added or removed from any logical partition in the server. The partitions of a particular server are stored on a system profile, which consists of multiple partition profiles (each one represents a client's partition). The system profile can be configured prior to booting a server and dynamically configured while a server is running. As the server boots the HMC will begin partitioning off the server's resources according to the system profile. If the server were to crash, or lose power for any reason, the HMC will maintain the system profile [11]. If a provider's HMC fails, the servers connected to it will maintain their current system profile but the provider will be unable to make any changes until a functional HMC is reconnected.

B. Strengths and Challenges

Like any model, Cloud Slicing has strengths and potential limitations. We will cover both starting with a summary of Cloud Slicing's strengths

1) Strengths: As mentioned previously, the biggest fundamental strength of Cloud Slicing is that it achieves software isolation without using a hypervisor. This allows clients to build their applications right off the firmware using whatever operating system and configuration they like. Without a hypervisor blocking access to the network layer, clients can implement security features on a network level. The ability to maintain isolation while allowing clients to access the network is what makes Cloud Slicing so special. We believe this strength is worth all the weaknesses.

Another strength of Cloud Slicing is the control given to both the clients and the provider. While clients gain control in terms of security, the provider gains control over resource allocation. With the HMC, providers can reserve, allocate, and remove resources with any client. This allows providers to guarantee a client a certain amount of exclusive resources from the server. Providers can also accommodate clients who want their own partition and clients who prefer the traditional multitenancy model, all on the same server; (Fig. 4) depicts what this setup may look like. To accomplish this setup, a provider can create logical partitions for clients who wish to use Cloud Slicing, and one big partition with a hypervisor for clients who wish to operate inside a VM.

2) Potential Challenges and Remedies: Since logical partitioning resources are allocated indefinitely, Cloud Slicing

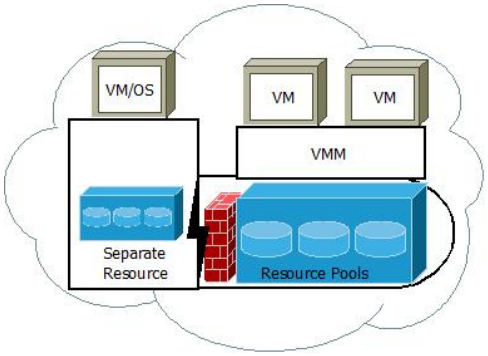

Fig. 4. Architecture of a server where part has the traditional VM/VMM setup and the other part uses Cloud Slicing.

face potential challenges that do not exist when resources are pooled. These challenges become readily apparent if a server is completely partitioned among clients; in other words, there are no unpartitioned resources on the server. In this case, if a client's application is using close to all of its partition's resources it will not receives additional resources even if other clients have resources to spare. This can be avoided by the idea of Capacity on Demand (CoD) Fig. 5. With CoD, the provider reserve some extra resources on the client's partition and offers these resources to the client when all available resources are used. The provider could offer two options for the client when extra resources are needed to complete the client's tasks:

1) Provide the client with more resources temporarily to complete the tasks with certain amount of charges.
2) If the client wants to upgrade and obtains more resources the provider could provide these reserved resources until the process of migrating client's server to a new logical partition is done.

Migrating client's server to a new logical partition with more resources without suspending his services is a challenge as well. Since suspending these services could cause a huge loss for the client.This problem could be mitigated by cloning that partition's image to the new partition and running all the services into this new one before the migration is done. An idea of "a migration tower" could be a solution as well and used during the migration process between two logical partitions. Clustering the logical partitions is another solution since it offers increased availability of applications and data. Another challenge is that, with Cloud Slicing, there is a higher risk for disrupting the server when adding or removing clients than there is with a hypervisor. A provider needs to have correctly configured a module called the Advanced System Management Interface (ASMI) [13]. If configured correctly, the ASMI will allow the personnel to add or remove logical partitions without disrupting the server. Throughout all these operations more planning and management is required from the provider than needed using the current multitenancy model.

C. Logical Partitions Security

Cloud Slicing isolates clients using logical partitioning which allows them to have different security setups. In the

Fig. 5. Capacity on Demand

current cloud model, security incidents occur in cloud environments and typically get fixed by the provider quickly and silently. During these incidents, the provider does not usually alert clients of the security issue. If clients are unaware of security incidents, it is impossible for them to evaluate the risk and take their own countermeasures [14]. Clients who are unaware of security incidents, system vulnerabilities, and malware reports, is evidence of the lack of security transparency between the provider and clients. The client does not get a fully transparent view of what is happening on a network level with a hypervisor filtering network traffic. The provider/client relationship consists of clients being forced to rely on the provider fully to handle security.

In Cloud Slicing, since logical partitions act as independent servers, network monitoring is completely possible for each partition using network monitoring tools such as Zabbix [15]. Clients get direct access to their partition's firmware and are then able to closely monitor the network traffic the hardware is in contact with. One might question what prevents two clients on the same server from reading each other's network traffic. The answer to the question is using the HMC to establish virtual networks. The HMC can create virtual Ethernet adapters and assign each partition a Virtual Local Area Network ID (VLAN ID) [11]. If this is done correctly, each client will have their own broadcast domain; meaning network packets could only pass between clients via one or more routers. From a network perspective, this setup is equivalent to having each client deploy on completely separate servers. The HMC's ability to deploy virtual networks is also convenient if a client has two partitions, on one server, for different applications. If the client wants these applications to be able to communicate, the provider can use the HMC to assign one VLAN ID to both partitions.

IV. Conclusion

Security is one of the primary concerns with cloud computing. The provider that successfully assures clients that their applications are safe will be the provider that gains more clients. This makes Cloud Slicing a wise model for a provider to adopt. Cloud Slicing assures clients of security by giving them control over the network level security. Logical partitioning is an established technology and has been thoroughly tested in the industry. Since this technology is the basis of our Cloud Slicing model, the majority of the model has already been proven.

For future work, more research and testing would be required before the Cloud Slicing model could be fully adopted in the industry. The course of the path we would suggest is to start by testing Cloud Slice's isolation on a partitionable server. Once the basic isolation is confirmed, next we suggest network monitoring software be deployed onto each of the partitions. This should be done to confirm that no network traffic from one partition can be read by any other partition. If the isolation features of Cloud Slicing are verified, additional software-based network security options could be found and tested. More security features would mean more options for clients who choose Cloud Slicing. Once complete isolation has been verified, real world trials could start.

Software based firewalls and IP control lists are two examples of security features that require network level access; these may work very well with Cloud Slicing. Our vision for the future is customisable security in cloud services. We intend for Cloud Slicing to balance the current multi-tenancy model, not replace it. If a company wishes to purchase cloud services, they should simply open a pamphlet and choose the best security options for their company.

From our research, it is clear that providers want to gain client's trust. If clients trust their provider, they will recommend their provider. Cloud Slicing lobbies for the client's trust by giving them the freedom to secure their own applications. This conveys a sense of trust from the provider to the client. Building strong relationships between the clients and the provider is vital to cloud computing, and we believe this relationship can be achieved by giving clients a bigger sandbox to play in.

References

[1] F. Gens, "It cloud services user survey, pt. 2: Top benefits & challenges," *IDC eXchange*, 2008.

[2] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, "A survey on security issues and solutions at different layers of cloud computing," *The Journal of Supercomputing*, vol. 63, no. 2, pp. 561–592, 2013. [Online]. Available: http://dx.doi.org/10.1007/s11227-012-0831-5

[3] P. You, Y. Peng, W. Liu, and S. Xue, "Security issues and solutions in cloud computing," in *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on*, June 2012, pp. 573–577.

[4] M. K. Srinivasan, K. Sarukesi, P. Rodrigues, M. S. Manoj, and P. Revathy, "State-of-the-art cloud computing security taxonomies: A classification of security challenges in the present cloud computing environment," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, ser. ICACCI '12. New York, NY, USA: ACM, 2012, pp. 470–476. [Online]. Available: http://doi.acm.org/10.1145/2345396.2345474

[5] H.-Y. Tsai, M. Siebenhaar, A. Miede, Y.-L. Huang, and R. Steinmetz, "Threat as a service?: Virtualization's impact on cloud security," *IT Professional*, vol. 14, no. 1, pp. 32–37, 2012.

[6] Hardware partitioning of the highest reliability : Fujitsu global. [Online]. Available: http://www.fujitsu.com/global/services/computing/server/sparcenterprise/technology/virtual/hard-partitioning.html

[7] G. Schimunek, D. Dupuche, T. Fung, P. Kirkdale, E. Myhra, and H. Stein, *Slicing the AS/400 with Logical Partitioning: A how to Guide*. Poughkeepsie, NY: IBM Corporation, 1999.

[8] S. Lowe. (2011, Jul.) The pros and cons of tower, rack, and blade servers. [Online]. Available: http://www.techrepublic.com/blog/the-enterprise-cloud/the-pros-and-cons-of-tower-rack-and-blade-servers/

[9] Ibm z9 ec, technical specifications from top gun technology. [Online]. Available: http://www.hds.com/products/compute-blade/compute-blade-500.html
[10] Ibm system z9 enterprise class. [Online]. Available: http://www-03.ibm.com/systems/z/z9ec/
[11] A. S. Dharma, M. Hais, M. Kang, S. Vetter, and Y. Wakayama, *IBM Power Systems HMC Implementation and Usage Guide*. Poughkeepsie, NY: IBM Redbooks, 2013.
[12] Hardware management console. [Online]. Available: http://www-01.ibm.com/support/knowledgecenter/POWER7/p7hat/iphat507.gif
[13] A. Altmark and C. Laking, *z/VM Security and Integrity*. Poughkeepsie, NY: IBM Redbooks, 2005.
[14] F. Doelitzscher, C. Reich, M. Knahl, and N. Clarke, "An autonomous agent based incident detection system for cloud environments," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, Nov 2011, pp. 197–204.
[15] The enterprise-class monitoring solution for everyone. [Online]. Available: http://www.zabbix.com