Storm Clouds Rising:
Security Challenges for IaaS Cloud Computing

Brian Hay
Dept. of Computer Science
University of Alaska Fairbanks
brian.hay@alaska.edu

Kara Nance
Dept. of Computer Science
University of Alaska Fairbanks
klnance@alaska.edu

Matt Bishop
Dept. of Computer Science
University of California at Davis
bishop@cs.ucdavis.edu

Abstract

Securing our digital assets has become increasingly challenging as our reliance on rapidly evolving technologies continues to grow. The security perimeter in computing has changed from a well-defined boundary that was relatively easy to identify and defend, to an elastic boundary that is constantly changing and for which the threats are constantly evolving. This paper investigates the complex security challenges that are introduced by the trend towards Infrastructure as a Service (IaaS)-based cloud computing. While not exhaustive, it identifies some technological and legal issues and concerns from the perspectives of identified stakeholders, and suggests some future directions for security research and development to help advance the security posture of this technology.

1. Introduction

The term cloud computing means many things to many people and the definitions will no doubt continue to evolve as new technologies and services enabling this model of computing are developed. For the purpose of this paper, the scope will be limited to IaaS cloud computing: a model in which units of computation (in the form of virtual machines (VMs)) and/or storage are allocated to consumers, who then access their assigned resources via some Wide Area Network. The cloud system consumers are granted complete control of any resources assigned to them (e.g., VMs or storage volumes), but have no control of the underlying virtualization or partitioning layer, the physical host(s) on which it executes, or the mapping of virtual resources to physical devices. While the security concerns for this realm are largely applicable to any external handling and processing of an individual or organization's digital assets, we will focus on security concerns for computational cloud computing from the perspectives of cloud service users, cloud service providers, and general security practitioners in both the technical and legal realms.

2. Background

The rapid evolution of technology coupled with our increased dependence on the same has melded together to make securing our digital assets an challenging problem. Originally computer systems were physically isolated and a data-centric approach to securing systems was accomplished largely through perimeter security. Physical security generally provided a means to isolate and secure the systems from malicious outsiders.

As technologies continued to evolve, and connectivity and mobility increased, it became difficult to secure an increasingly fluid perimeter. The focus of security began to shift from physical security and securing the data centers to protecting the endpoints themselves. This was accomplished through many mechanisms including firewalls, confining the end point services, changing configurations to restrict access, and similar measures. As connectivity increased, the security focus again had to shift, to protect the plethora of applications that depended on the network. These included applications that request and provide data, distributed components, and other virtual workgroups. This focus shift required that protection at the application layer become a concern.

The new evolution towards cloud computing, both IaaS-based and (as data and services are "outsourced" to the cloud) Software as a Service (SaaS)-based, again demands a reconsideration of methods used to provide security. The new critical point is that the changing perimeter that extends further into realms that are controlled by others. The concern is how to data in transit, in storage, and also from the service providers. The roles of the traditional stakeholders in such a system are changing and the distinction between insider and outsider has become increasingly blurred. In some cases, the degree of "insiderness" associated with an stakeholder needs to be identified in order to effectively assess risks as the traditional binary definition that requires a clear insider/outsider boundary is longer an appropriate model [1].

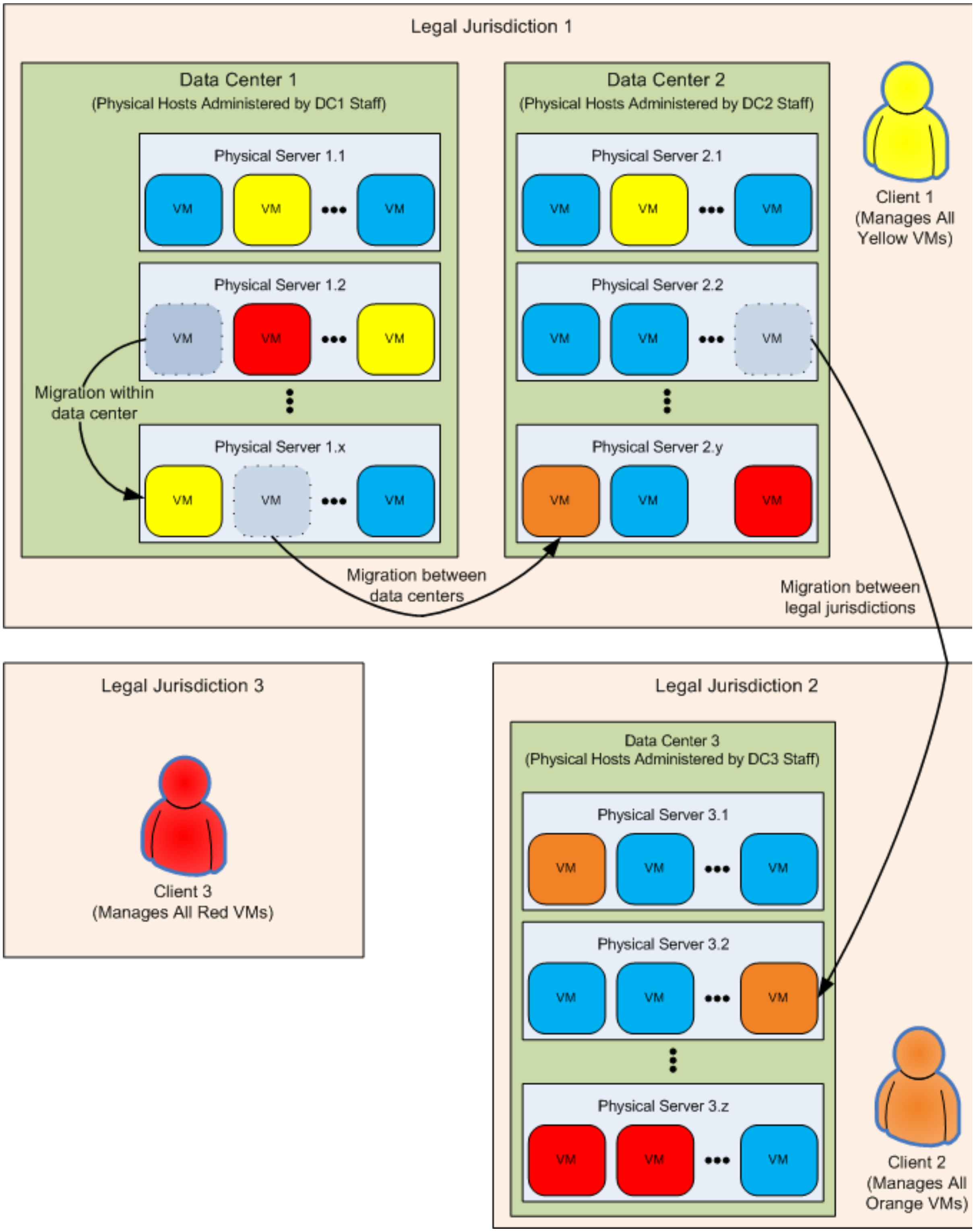

Figure 1: A generic model of a cloud computing provider

Figure 1 provides an example of a generic cloud computing scenario which demonstrates the complexity of the evolving security perimeter. In this example, the cloud provider has three data centers in two legal jurisdictions. Each data center has multiple physical servers, which are managed by the local data center staff (i.e., full control of those servers is in the hands of the local data center staff, and the cloud computing clients have no control of those physical hosts). Each physical host concurrently executes multiple virtual machines, which are assigned to clients. The clients have full control over their VMs, but no visibility into the virtualization layer (the hypervisor) or the physical hosts on which they execute. Decisions about the allocation of VMs to physical hosts are made entirely by the data center staff, and migration can occur between physical hosts in the same data center, between data centers in same legal jurisdiction, and even between different legal jurisdictions.

In order to effectively address the these and other new evolving security issues that IaaS-based cloud computing brings to the table and to enable stakeholders to provide security in this new and continually evolving environments, it is important to identify the technical and legal challenges that are facing cloud security providers.

3. Technical Challenges

While IaaS-based cloud computing brings many advantages, this model also raises significant technical security considerations and questions. Among these issues are operational trust modes, resource sharing, new attack strategies, and digital forensics. These are important areas of concern and are further complicated by the issues associated with giving up control in any environment. As you increasingly rely on others to provide you with functionality, you correspondingly give them control. Further this control is given to someone who most likely does not share your vested interest in your mission. As you relinquish control, you lose access to information and as a result, give up the ability to answer some of the important questions regarding technical and jurisdictional issues. In some cases, the information required to make informed security decisions is no longer available, and that data which is available, may no longer be as trustworthy as it would be in a system over which you have complete control. As a result it is vital that the security implications of cloud computing be carefully considered and factored into a decision about the appropriateness of a cloud-based solution to a given set of IT requirements.

3.1 Operational Trust Modes

A primary technical challenge that results from the decision to use IaaS-based cloud computing is the level of trust accorded the resource provider. These providers are part of your system and their roles as stakeholders in your business are complicated by the blurring of the insider/outsider line. Thus, there are two aspects here. The first is to determine what operational levels of trust are needed to capture the trust in the resource providers. The second is to use those levels in the risk assessment.

In our cloud computing scenario above, the cloud resource provider maintains sole access to the underlying physical components of the system, and provides the cloud consumer with full control over some portion of it (usually by means of an abstraction layer like virtualization). In practice this means that the provider has access to all of the consumer's operations and data in the cloud environment. Much of this may be business sensitive, and the security plan must take this into consideration. Among the approaches that can be used are:

- If the cloud is used only to store data, well-established cryptographic algorithms allow a cloud storage consumer to encrypt the data prior to insertion in the cloud, and decrypt it after moving it back to her own systems. It should be noted, however, that such algorithms are not effective if the data is encrypted for use *within* the cloud (as the computation device in the cloud would require access to the decryption key, which would then also make the key, and consequently the data, available to the cloud resource provider.
- In some cases it may be possible for a computational engine (such as a virtual machine) to perform operations on encrypted data, producing results without exposing the input data or the result in plaintext form. Such schemes are known as *homomorphic encryption*. While this is an area of active research, current results tend to be applicable to fairly narrow operations [2, 3]. However, even in this mode of operation, a malicious cloud provider may be able to infer useful information from the operations being performed, and the characteristics of the results (e.g., the size of the result).

We must also consider the extension of the network in the cloud environment, because that portion of the network provides the connection between the consumers and the cloud resources. In a more traditional environment, it may be reasonable for a corporation to control that entire pathway (on an internal LAN), or at least have some well-understood pathway between their offices and the corporate data center. In the cloud model, the WAN connections are likely to be far more complicated and dynamic. While encryption of network traffic on such links can provide some level of protection, we must again consider that even encrypted content can provide interesting information that might be exploitable [4].

Can we run in a mode in which we do not need to trust the cloud providers and hosts? Do we need to trust the networks and the network providers? Not trusting the latter may be simpler than not trusting the former, but is non-trivial if attacks against the core network infrastructure and routing algorithms are considered.

3.2 Resource Sharing

In the current corporate computing model, resources such as storage and hosts tend to be used exclusively by a single corporate entity. However, in the cloud model it is entirely reasonable that a resource allocated to one corporation may be instantiated on some physical infrastructure that also hosts resources allocated to other corporate users. For example, a virtual machine may be instantiated on a physical server hosting several virtual machines, each allocated to a different corporation. In this case, it is quite possible that two competitors may be allocated resources on the same physical infrastructure. Then security policies and procedures must consider the possibility that data may leak between competing corporations, or that the actions of one corporation could impact the ability of a competitor to conduct business. Some work has already demonstrated this potential [5].

Given this concern, we should consider whether a cloud provider can provide some minimal level of assurance that such conflict will not arise. Obviously the easiest way to do this is to allocate any physical component in the cloud to a single consumer, but this would significantly impact the flexibility of the cloud. A more practical approach may be to implement an analog of the Brewer and Nash model [6] (also known as the "Chinese Wall Model"), in which corporate cloud consumers are grouped into conflict of interest classes. The cloud provider would then be free to allocate resources in the cloud with the limitation that no two corporations in the same conflict of interest class can share the same *physical* resources.

3.3 New Attack Strategies

The cloud model makes new attack strategies possible. These strategies must be considered when assessing and attempting to secure cloud computing resources. One example of such a strategy is the ability of an attacker to co-locate its resources with that of their target, as Ristenpart and his colleagues demonstrated [5]. Once the attacker has gained such a foothold, subsequent efforts to attack the targets may be possible, both by using current attacks and by attacking the virtualization layer and/or physical hardware directly.

3.4 Digital Forensics

Several issues make digital forensics in the cloud more challenging than with ordinary systems and networks. These issues include:

- *The ephemeral nature of cloud resources.* Hosts in the cloud may be instantiated for the duration of some processing event (e.g., monthly accounting), after which they are decommissioned and their resources (CPU, RAM, storage, etc) are returned to the cloud for use by other hosts. Unless an event worthy of forensic investigation is discovered during the lifetime of the virtual host, or *possibly* very soon after that virtual host is decommissioned, it is likely that the "system" to be examined is lost. Once the resources are used by some other host in the cloud, recovering any useful data from the original virtual host is probably impossible, as would be attributing any data actually recovered to the original virtual host with any degree of confidence.
- *Seizing a "system" for examination.* In traditional digital forensics, it is relatively easy to seize an entire system, including all processing and storage components, for examination offline. We can also feel relatively confident that the data written by the system was written to these devices, and that data found on these storage devices was written by the host being examined. However, in cloud systems, what comprises the system to be examined is far less straightforward. At a logical level, we could attempt to seize the virtual disks associated with the cloud host in question, but this may consist of virtual disk file that was written to many areas of multiple physical storage devices, by many different physical hosts, as the cloud resource was migrated. While finding the current version of the file is probably quite easy, finding

historical artifacts is not, because they may be found on other areas of the physical disk, no longer be associated with this virtual disk file, or be on other physical disks no longer associated with this virtual host.

Some other operations currently used by digital forensics investigators (such as live analysis) can be performed more easily in virtual environments such as those commonly found in the cloud.

A full examination of digital forensics in a cloud computing environment is beyond the scope of this paper. But it is clear that digital forensics in the cloud presents challenges, and provides opportunities, not found in digital forensics as it is more commonly practiced today.

4. Legal Issues

In addition to the identified technical challenges, legal issues associated with IaaS-based cloud computing need to be incorporated into a risk analysis plan. That way, potential consumers can make informed decisions about the appropriateness of utilizing this potentially valuable resource. These legal issues include consideration of jurisdictional issues, an understanding and the evolution of cloud stakeholder rights, and technical approaches to addressing the associated legal and jurisdictional issues.

4.1 Jurisdictional Issues

Resources in the cloud are often not fixed to any geographical location such as a specific data center. They may migrate between physical locations during their lifetime. The decision as where a resource is instantiated or migrated may be based on a variety of factors, including load balancing by the cloud provider, network and datacenter performance and availability, and even the characteristics of the current clients. The result is that the resource may exist in multiple legal jurisdictions, each of which may have different, and even conflicting, rules about important security issues such as intrusions and data protection. In such a dynamic environment, it may not be possible for a cloud resource consumer (e.g., a corporate customer of a cloud computing provider) to remain in compliance with the legal requirements of the jurisdictions in which their assigned resources may operate.

For example, consider a resource which is instantiated in a jurisdiction that does not require personally identifiable information (PII) to be encrypted. The data is subsequently migrated (without the knowledge of the cloud resource consumer) to a jurisdiction for which such protection is mandatory. Should the cloud resource consumer configure her system to comply with the most stringent legal requirements in any jurisdiction in which the cloud has nodes—indeed, would the consumer even know what jurisdiction this was? Would it even be possible—for example, the cloud may span two jurisdictions with conflicting legal requirements, in which case it would not be possible to configure a system to meet the most restrictive case.

This issue is certainly not new. There are many examples of companies with datacenters in multiple jurisdictions. Companies with multiple datacenters have at least some known set of locations in which the resources reside, and the resource consumer and resource provider are often functionally equivalent entities (namely the same corporation or part of the same corporate hierarchy). Even in cases where one corporate entity is purchasing resources in a data center belong to another entity, there is still generally a contract describing the location of service, and any co-location services that will be provided.

But the cloud model is somewhat different. In that model, the cloud resource consumer and cloud resource provider are seldom the same entity, and the contract between cloud consumer and provider tends to describe the resource that will be provided (CPU cycles, RAM, storage capacity, network bandwidth, minimum uptime, and network characteristics, for example) rather than the locations at which these services will be provided. The cloud provider is then free to provision resources to meet these requirements at their discretion, and subsequently migrate them to address their continually changing system status.

4.2 Cloud Stakeholder Rights

Given that a resource logically located "in the cloud" can be instantiated in and migrated among multiple physical locations, we consider the implications this has for the stakeholders in this process.

- *Cloud Provider:* Given that the act of migrating a (virtual) host may change the legality of the activity taking place on that host, what restrictions should be placed on the provider, and to what extent are they liable for illegal activity that results in such a move? For example, consider a host that provides a social networking site. If this host is migrated from jurisdiction A, which has no cyber-bullying legislation, to jurisdiction B, which prohibits cyber bullying, is activity that was legal yesterday now illegal based *only* on this migration within the cloud? From another point of view, can the cloud provider migrate hosts into a jurisdiction if they know that the activity already occurring on the host is illegal in that jurisdiction? Is such a migration legal to aid law enforcement, and if so under what conditions?

- *Cloud Resource End Users:* Can a user of some resource in a cloud-based system be expected to know when their activities are illegal? Revisiting our cyber bullying example, if jurisdiction A has no cyber bullying provisions, and jurisdiction B does, a user could post three identical messages from the same location on the same service, and that same action could be legal on the first and last days, and illegal on the second day, because the server was migrated from jurisdiction A to B and back again.

4.3 Technical Solutions

One solution to this issue would be for the cloud resource consumer to “tag” their resources in a manner that would indicate which components could be migrated, to where, and under what conditions. The cloud provider could make informed decisions about migration based not only on the available resources in the cloud, but also on the requirements of the consumers as indicated by the tags. This would simplify the legal issues, particularly for the consumers. But it would also quite likely limit the cloud provider’s ability to efficiently manage their resources, resulting in a model that looks far more like the current “rackspace in a datacenter” approach than the free-flowing and flexible cloud many currently envision.

5. Future Considerations

While not insurmountable, the security challenges associated with IaaS-based cloud computing need to investigated in order to protect our digital assets. An increased understanding of cloud computing and the roles of various stakeholders in this realm are important, as is more research into the technical and legal issues that resource-based cloud computing introduce to the threat horizon. This requires cloud-oriented research into identification of technological issues including trust modes, resource sharing, attack strategies, and digital forensics implications. Also, legal issues such as jurisdictional issues, cloud stakeholder roles and rights, and technological approaches to solving these problems should be paramount in resource-based cloud computing research and development.

6. References

[1] Bishop, M. and Gates, C. 2008. Defining the insider threat. In Proceedings of the 4th Annual Workshop on Cyber Security and information intelligence Research: Developing Strategies To Meet the Cyber Security and information intelligence Challenges Ahead (Oak Ridge, Tennessee, May 12 - 14, 2008). F. Sheldon, A. Krings, R. Abercrombie, and A. Mili, Eds. CSIIRW '08, vol. 288. ACM, New York, NY, 1-3. DOI= http://doi.acm.org/10.1145/1413140.1413158

[2] Change, C, and S. Tsu. Arithmetic operations on encrypted data. International Journal of Computer Mathematics. International Journal of Computer Mathematics, Volume 56, Issue 1 & 2 1995 , pages 1 - 10

[3] Micciancio, D. 2010. A first glimpse of cryptography's Holy Grail. Commun. ACM 53, 3 (Mar. 2010), 96-96. DOI= http://doi.acm.org/10.1145/1666420.1666445

[4] Chen, S, R. Wang, X. Wang, and K. Zhang. Side-Channel Leaks in Web Applications: a Reality Today, a Challenge Tomorrow. RetrievedJune 11, 2010 from http://www.informatics.indiana.edu/xw7/WebAppSideChannel-final.pdf

[5] Ristenpart, T., Tromer, E., Shacham, H., and Savage, S. 2009. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In Proceedings of the 16th ACM Conference on Computer and Communications Security (Chicago, Illinois, USA, November 09 - 13, 2009). CCS '09. ACM, New York, NY, 199-212. DOI= http://doi.acm.org/10.1145/1653662.1653687

[6] Brewer, D.F.C., and M. Nash. "The Chinese Wall Security Policy," pp. 206, 1989 IEEE Symposium on Security and Privacy, 1989.