Cloud Environments and Security

Luke Bicklein, bicklein (at) umsl.edu (A paper written under the guidance of Prof. Jianli Pan)

Download 

Abstract:

Large virtual environments are a rapidly being deployed and utilized in many (if not most) large institutions. Private industries, public institutions, and individuals everywhere continue to migrate their data and shared resources to cloud environments. This paper details modern Cloud environments, security concerns involved in large environments, as well as current and possible future solutions to securing them.

Keywords:

Cloud, Cloud Computing, Cloud Stotage, Virtual Environment, Virtual Machine, data warehouse, virtual firewall

Table of Contents

1. Introduction

This section will provide a definition of Cloud Environments, and the justification for security concerns.

1.1 What is Cloud Computing?

Cloud computing is the practice of using shared computer resources over an internet-hosted network of remote severs. This virtual network can then be used for a large variety of things, but mostly commonly for data storage, management, and processing; this is an alternative to using local servers or Individual/personal computers.

1.2 Why Importance of a Secure cloud?

In today's technological environment, Cloud storage has become a cornerstone of personal and public data archiving. Users and virtual hosts alike expect their storage options to be accessible by multiple devices, and similarly (if not more) convenient as their previous local storage options. In order to justify migrating data to large virtual environments, security is a key concern. With mobile and local data demands constantly growing, the public needs to be able to have the same sense of trust and reliability for their private data in Cloud environments as they do with hard dish storage. Forbes reported earlier this year: **"**42% of IT decision makers are planning to increase spending on cloud computing in 2015, with the greatest growth in enterprises with over 1,000 employees (52%)". With this amount of industry growth completely reliant on virtualization, there is an exponentially growing concern for security.

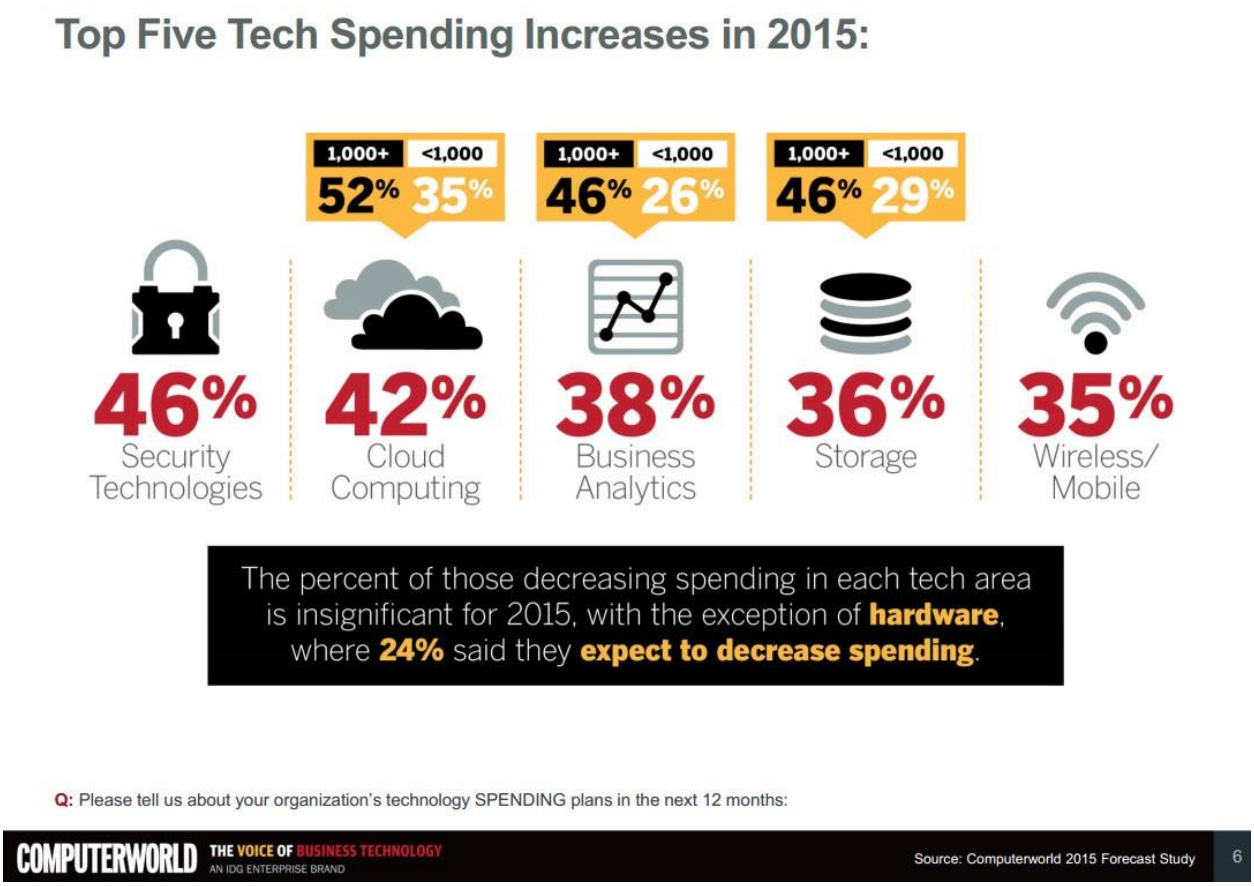

Figure 1: Projected Distribution of Tech Industry Spending in 2015 [10]

2. Large Scale Environments

This section provides a generalized look at what a large virtual environments and Clouds are, their capabilities and use in modern industry. It then provides a look into a modern Data Warehouse, and what a specialized third-party would have to offer businesses today. And lastly, the host-side hardware of a EDW (Enterprise Data Warehouse) to give a basic understanding of the physical data storage and processing.

2.1 Intentions of Large Scale Environments

A virtual machine (VM) is an efficient and isolated duplicate of a real machine [1]. Common uses for virtual machines are: development and testing of new operating systems, simultaneous running separate operating systems on the same hardware, and

server consolidation [2]. A virtual machine environment is created by a Virtual Machine Monitor (VMM), also called an “operating system for operating systems” [3]. The monitor creates multiple VMs on top of a single existing physical machine. Each VM provides facilities for an application or a “guest system” that believes to be executing on a normal hardware environment [4]. When a host computer runs an application known as a hypervisor, it creates one or more virtual machines that are capable of simulating physical computers so well, that the simulations can run any software, from operating systems, to end-user applications [5]. For many larger operations the hardware consists of a number of physical devices including at the very least: processors, hard drives and network devices. These are commonly located in independent datacenters which can be in a completely different location and are responsible for storage and processing needs. On top of all this, there can be multiple software, virtualization and management layers, which allow for the effective use and sharing and use of servers [6].

Virtual Environments have been rapidly replacing repetitive hardware deployment in business settings for years. This is a natural transition considering the expanding development of VM-friendly operating systems, the cost savings involved with resource sharing, and efficient integration of relevant data. Along many other added benefits, transitioning to large virtual environments allows administrators to preform simple and regular backups and disaster recovery that are not available when dealing with hardware failure. Cloud Computing allows dynamic allocation of resources without ever having to interact with the hardware itself. Developing large virtual environments intends to allow the consolidation of hardware resources, and efficient distribution and management of resources, as needed, to many users.

Cloud Storage

Cloud Storage, specifically, often refers to a system of remote databases allocated to digital data storage. They often consist of multiple servers that can also span multiple physical environments. This form of storage is kept accessible in logical pools controlled by the storage host and retrievable then by client devices anywhere with an internet connection.

2.2 Modern Deployment and Utilization

Microsoft: The Third Party Player

Deploying a data warehouse is quite an undertaking for many industries not previously holding large IT operations with qualified staff to back them up. For large companies such as Microsoft, this was an opportunity to offer services to industries not wanting to diversify into this complicated field. As Microsoft markets their

Modern Data Warehouse they talk about companies in the past: “The IT organization would purchase and install state-of-the-art hardware servers, optimally balanced and tuned for CPU, I/O, and storage. IT also would install the software and tune it for performance. Even before loading data, the company could spend months and hundreds of thousands of dollars on infrastructure with ongoing maintenance, support, and replacement.[7]” They then continue to propose their solution: “A [Microsoft] cloud deployment implements the same BI (Business Intelligence) strategy, but replaces on premises infrastructure with a Windows Azure cloud infrastructure maintained by Microsoft. Customers save the cost of maintaining on premises infrastructure in exchange for a low monthly service fee, lowering TCO. More importantly, high-value IT resources spend more time building the business and less time supporting infrastructure [7].” Third party data management is becoming more and more wide-spread due to the increasing complexity of data warehousing, management/processing, and analytics software. The cost of the technology and expertise involved in managing a large virtual data network has led to the birth of entire industries dedicated to deploying and simplifying them for clients. It is also important to note that services provided often range across 3 common categories which are defined in an article from the Department of Product and Systems Design Engineering at University of the Aegean as [6]:

Infrastructure as a Service (IaaS): Provides the consumer with the capability to provision processing, storage, networks, and other fundamental computing resources, and allow the consumer to deploy and run arbitrary software, which can include operating systems and applications. The consumer has control over operating systems, storage, deployed applications, and possibly limited control of select networking components.

Platform as a Service (PaaS): Provides the consumer with the capability to deploy onto the cloud infrastructure, consumer created or acquired applications, produced using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

Software as a Service (SaaS): Provides the consumer with the capability to use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices, through a thin client interface, such as a web browser (e.g. web-based e-mail). The consumer does not manage or control the underlying cloud infrastructure, including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

Organization

As shown in Figure 2, the entirety of a Microsoft virtual environment can be separated into layers which include [7]:

- Business Intelligence & Analytics
    - This layer would include business-specific software that would allow the user to see all the data in the contexts that complement their goals
- Data Enrichment and Federated Query
    - Responsible for ETL (Enrichment, Transformation, Load) and master data management providing consistent quality data
- Data Management & Processing
    - Includes handling data in real-time using event processing and allows predictive analysis and interactive perspectives of aggregated data
- Infrastructure
    - Covers client-side and host hardware spanning servers all the way to access and reading devices

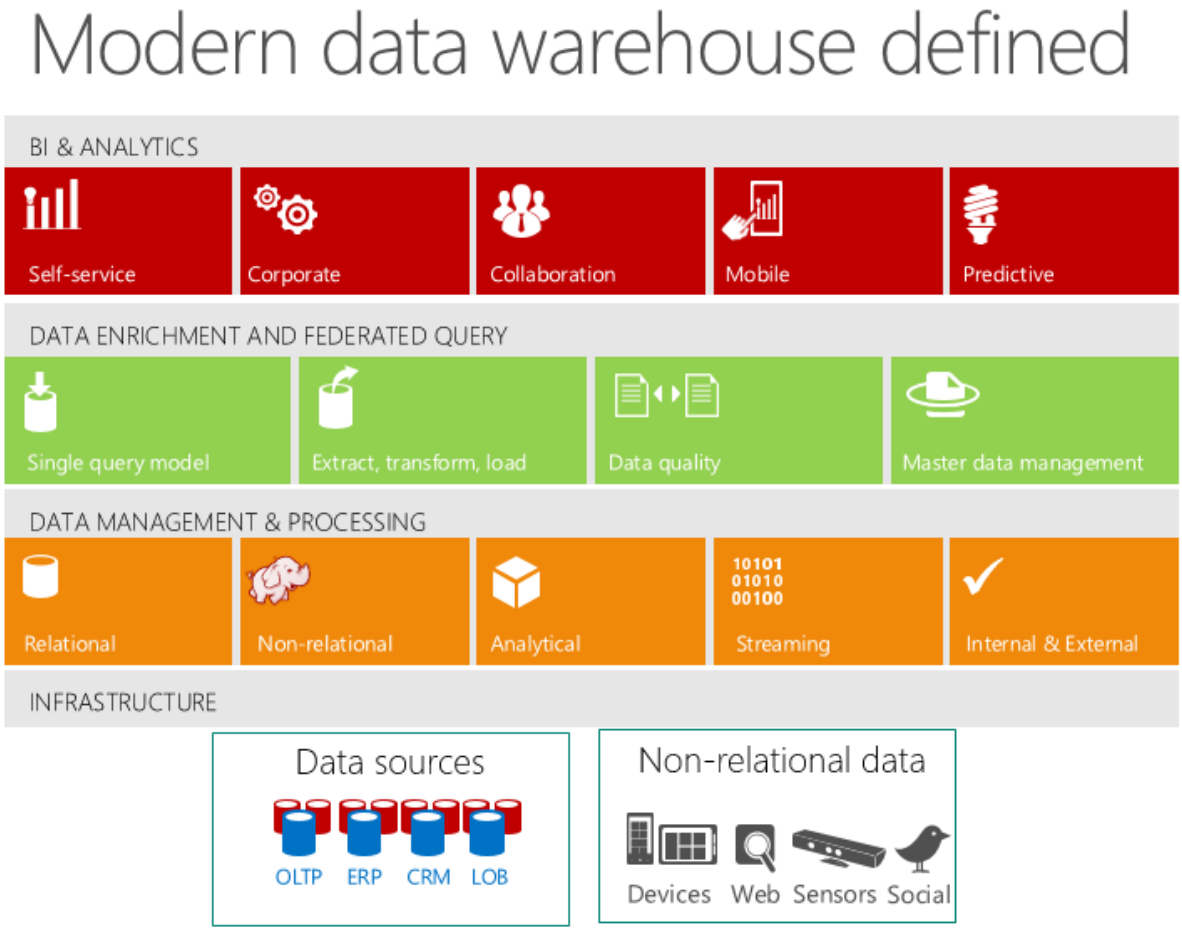

Figure 2: Framework of a Modern Microsoft Data Warehouse [7]

2.3 Infrastructure: Under the Hood

Components

Here is a list of hardware one might expect to find in a modern EDW today [8]:

- **Processors:** For each node, a simple two-way quad-core processor preferably clocked no slower than 2 GHz. Modern power-saving or economy processors sold have eight cores and are clocked around 2.16 GHz.
- **Memory:** Experts have come to an agreement that on-board memory per node can be about 24 GB, although some can be as high as 32 GB.

- **Storage:** Each node maintains some of the storage, and most experts agree on 4 TB per node. This can be a grouping of ordinary SATA hard drives, preferably 7200 RPM. RAID is allowed for master nodes.
- **Network connectivity:** Racks may be interconnected by ordinary 1GbE rack-level switches, which collectively connect to 10GbE cluster-level switches.
- **Operating System :** While the OS does not fall under the hardware category, it is important to realize that it exists server-side to talk to the hardware and conducts vital operations for controlling computing across all resources.

Data Storage Flow

A standard data warehouse is designed to provide client's with database services using Massively Parallel Processing (MPP) architecture. This entails multiple, tightly coupled computers with specific specialized functions. They also include at least one array of storage devices that are accessed in parallel. Specialized functions can include: system controller, database access, data load and data backup. [9].

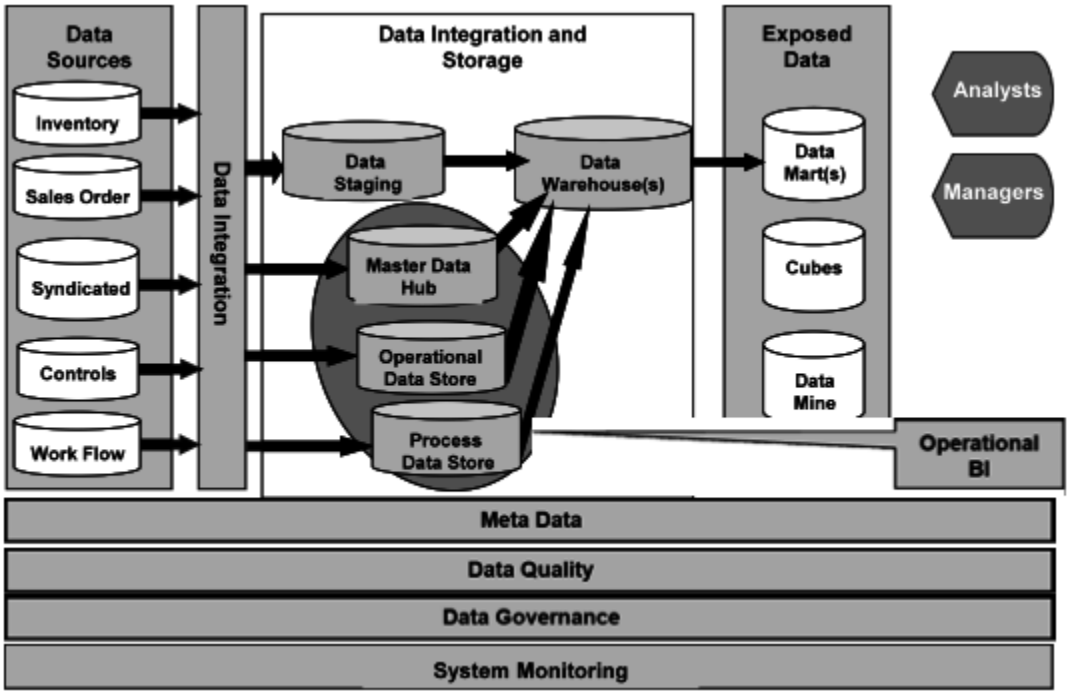

Figure 3: Example of Data Flow through a Warehouse System [9]

3. Security Concerns

This section discusses risk assessment and system engineering virtual environments to counteract security threats.

3.2 Identifying Threats

There are benefits to Cloud computing because of its design. These include: “centralization of security, data and process segmentation, redundancy and high availability [6].” However, while these characteristics provide security benefits, they also pose unique security threats not present in more primitive or less centralized systems.

The first step to understanding and countering intrusions in a virtual environment is knowing its vulnerabilities. Today, Intruder Detection Systems (IDS) are commonly deployed to collect and analyze data that points to intrusion. There are two forms [11]:

Network-Based IDS: most commonly sits on the ingress or egress point(s) of the network to monitor what's coming and going. Given that a network-based IDS sits further out on the network, it may not provide enough granular protection to keep everything in check -- especially for network traffic that's protected by SSL,TLS or SSH.

Host-Based IDS: protect just that: the host or endpoint. This includes workstations, servers, mobile devices and the like. Host-based IDS are one of the last layers of defense. They're also one of the best security controls because they can be fine-tuned to the specific workstation, application, user role or workflows required.

3.3 Modern Threats

There are countless vulnerabilities in Cloud environments that need to be considered, however, there are a few that are common threats that system developers must consider when designing security.

Malicious Insiders

Employees working at cloud service provider or responsible for administrating virtual environments can have very broad, if not unrestricted access to sensitive data. Companies must be aware of employee access to confidential information and enforce restrictions and privacy policy.

Secure Data Transmission

During data transfer from clients to the cloud, secure, encrypted communication like SSL/TLS should be used in order to avoid interception and malicious use of communicated information.

Shared Technology Issues

"The cloud service SaaS/PasS/IaaS providers use scalable infrastructure to support multiple tenants which share the underlying infrastructure. Directly on the hardware layer, there are hypervisors running multiple virtual machines, themselves running multiple applications.

On the highest layer, there are various attacks on the SaaS where an attacker is able to get access to the data of another application running in the same virtual machine. The same is true for the lowest layers, where hypervisors can be exploited from virtual machines to gain access to all VMs on the same server (example of such an attack is Red/Blue Pill). All layers of shared technology can be attacked to gain unauthorized access to data, like: CPU, RAM, hypervisors, applications, etc." [15]

Denial of Service

An attacker can issue a Denial of Service attack (commonly referred to as DOS attack), which is when multiple systems are used to overwhelm another system with traffic therefore denying service to anyone else who wishes to use it.

3.3 Virtual Firewalls

In virtual environments, system designers can no longer rely on security protocols and physical firewalls alone. Virtual firewalls are capable of adaptively locating themselves in or outside of the network in order to monitor specific ports. As opposed to physical firewalls, Virtual Firewalls allow logical separation of clients/accesses through their own firewall even if they are physically locating on the same machine. Cisco's documentation on Design Considerations representing their Nexus 1000V Service Virtual Machines (represented in Figure 4) [12] shows accesses to Tenant 1 all passing through their own virtual firewall. Notice each machine is connected to a VEM (Virtual Ethernet Module). Adaptive virtual firewalls are capable of logically separating each machine making them completely ignorant of the location of other machines even if they are accessing the same Tenant. This can be represented in Figure 5, and is referred to as a "fully collapsed trust zone" meaning each source access has passed through a virtual firewall [13].

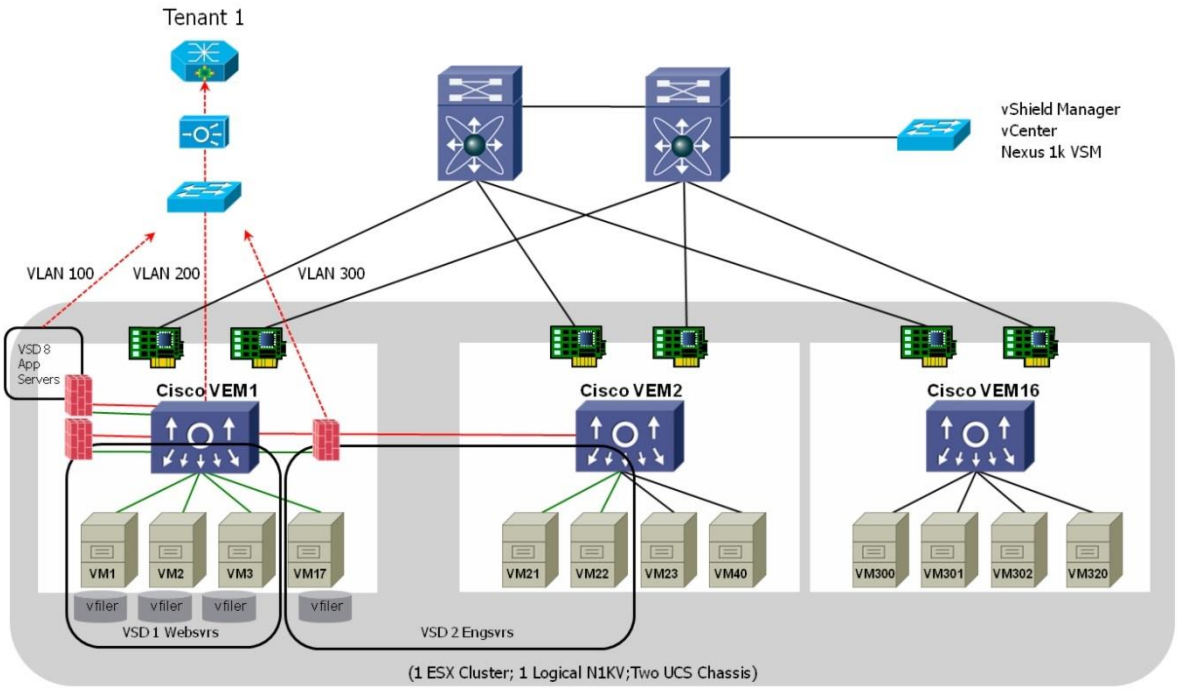

Figure 4: Cisco's representation of their Nexus 1000V Service Virtual Machines (SVMs) and VSDs with vShield virtual firewalls [12]

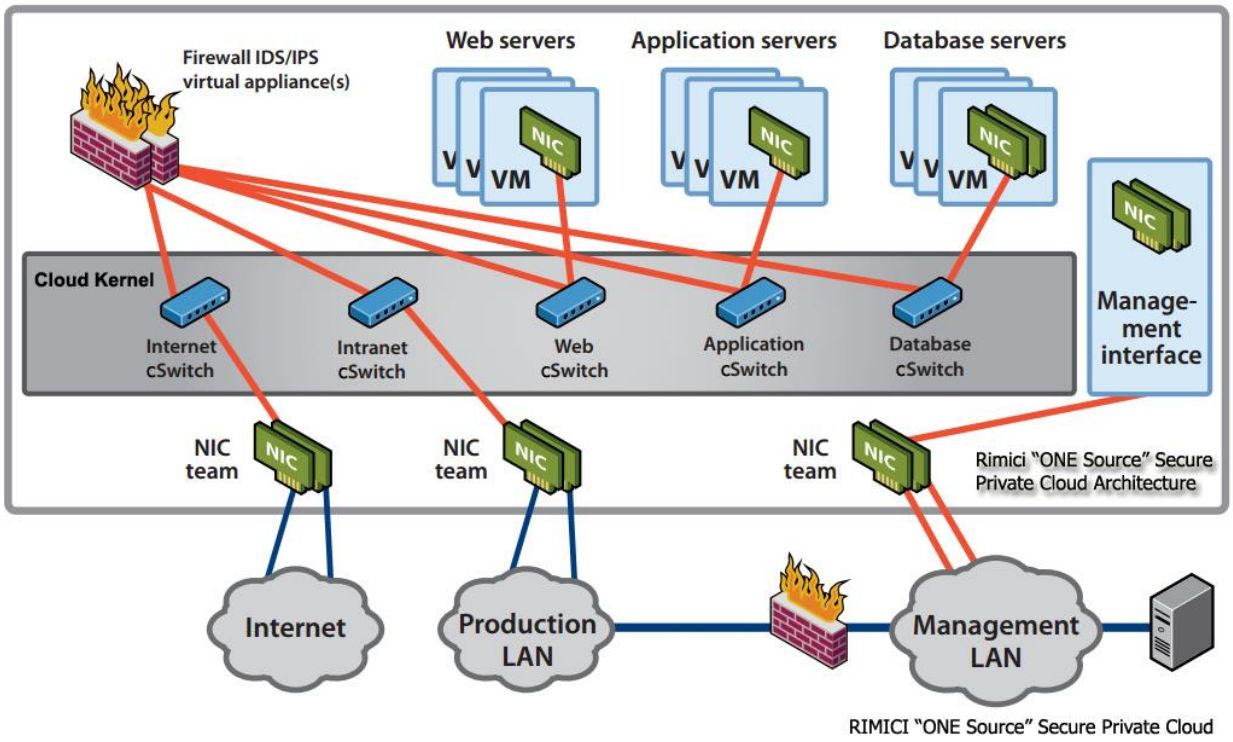

Figure 5: VMware representation of a fully collapsed trust zone [13]

This separation is vital for large environments to be practical, functioning and secure. Virtual firewalls allow system designers to set up virtual trust zones within a system and restrict access from unwanted segments of the environment. For example, a company may want to share data across specific applications between the accounting and production departments. However, they may want to restrict access to sensitive material to one or both departments by logically separating data zones within the system. This is a very common practice in modern systems and requires the use and strategic deployment of virtual firewalls.

Distributed Firewalls

Physical and virtual firewalls are nearly identical in function, just not in form. While we are capable of designing virtual firewalls to be dynamic, they still demand traffic be steered through it, processed instance by instance, and rule structured by the basics of IP addressed. This still creates problems with it becoming a "data chokepoint". Distributed Firewalls are different because they are not a form factor at all, but instead (Figure 6) "is now embedded as-a-service in the programmable hypervisor kernel networking stack. All participating hypervisors collectively become one "Firewall". Every virtual server is connected to a hypervisor. By consequence, in this model, every virtual server is directly connected to one omnipresent "Firewall". A firewall that knows everything about those virtual servers" [14]. This allows "Every packet sent or received flows through a stateful "Firewall", and the security policy follows the virtual machine when it migrates to another hypervisor. With a distributed firewall, traffic steering is completely removed from the process of implementing security policy – for the simple reason that it's directly connected to every virtual machine. In other words, security is omnipresent, at the very first and last hop" [14].

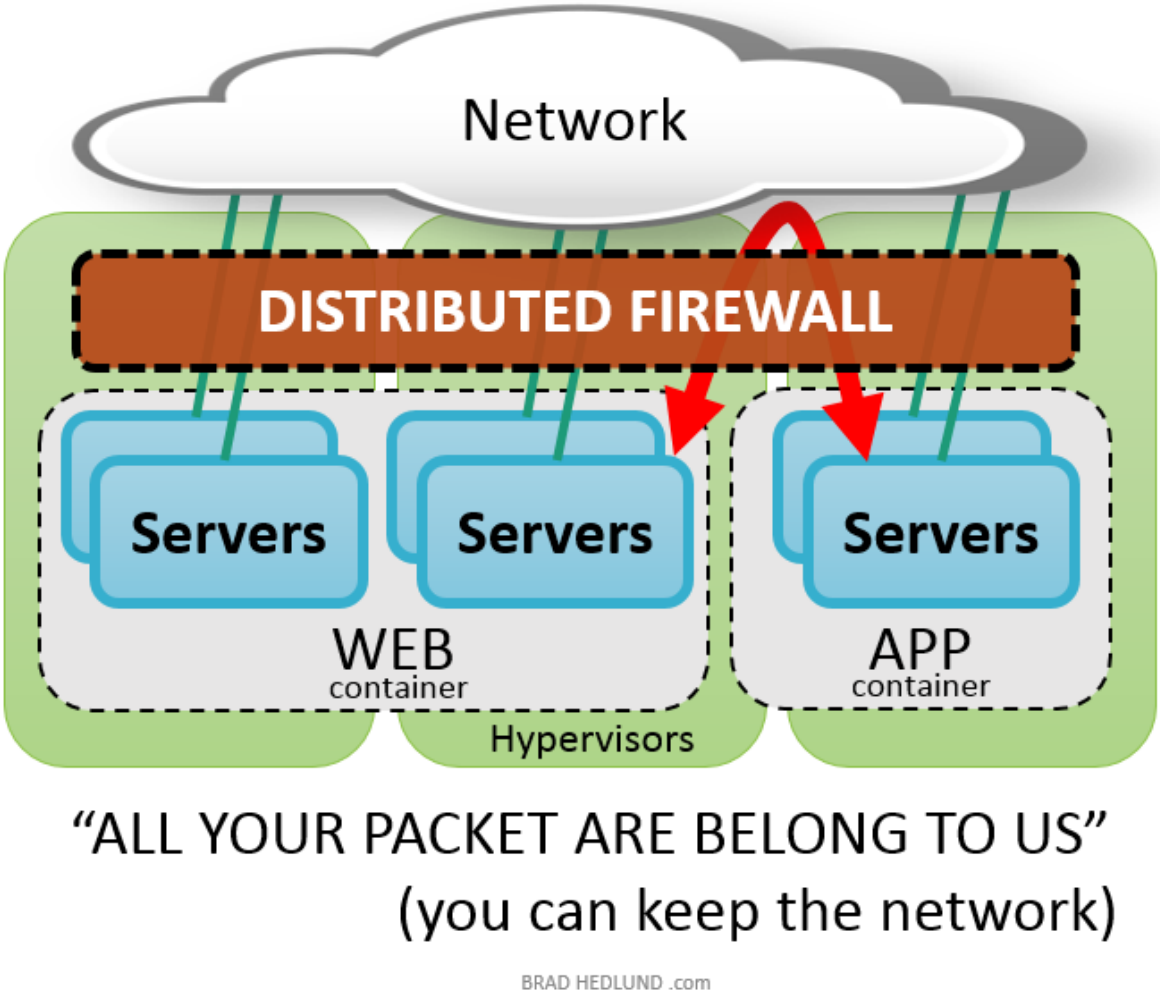

Figure 5: Diagram showing a Distributed Firewall acting as a service during a hop[14]

4. Summary

Cloud Computing is a rapidly growing industry because of its undeniable efficiency and practicality. With more companies and individuals entrusting their data to Cloud environments, it is imperative we continue to understand security threats, and continue searching and learning about counteracting them with all the tools available, old and new.

5. References

[1] Popek, G., Goldberg, R. (1974) “Formal Requirements for Virtualizable Third Generation Architectures”, Communications of the ACM. Volume 17, number 7, pages 412-421.

[2] Sugerman, J., Ganesh, V., Beng-Hong L. (2001). Virtualizing I/O Devices on VMware Workstation’s Hosted Virtual Machine Monitor. Proceedings of the 2001 USENIX Annual Technical Conference.

[3] Kelem, N., Feiertag, R. (1991) “A Separation Model for Virtual Machine Monitors”, Research in Security and Privacy. IEEE Computer Society Symposium, pages 78- 86.

[4] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.9231&rep=rep1&type=pdf

[5] E. Naone, Technology overview, conjuring clouds, MIT Technology Review, July–August, 2009.

[6] http://ac.els-cdn.com/S0167739X10002554/1-s2.0-S0167739X10002554-main.pdf?_tid=bb17206c-a302-11e5-8b39-00000aacb35e&acdnat=1450166963_a7a52440d5649030830d8912c3803e78

[7] http://download.microsoft.com/download/c/2/d/c2d2d5fa-768a-49ad-8957-1a434c6c8126/the_microsoft_modern_data_warehouse_white_paper.pdf

[8] http://www.tomsitpro.com/articles/big_data-business_intelligence-business_analytics-hadoop-virtualization,2-557-5.html

[9] https://dzone.com/refcardz/data-warehousing

[10] http://www.forbes.com/sites/louiscolumbus/2015/01/24/roundup-of-cloud-computing-forecasts-and-market-estimates-2015/

[11] http://searchsecurity.techtarget.com/answer/Host-based-IDS-vs-network-based-IDS-Which-is-better

[12] http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/VMDC/2-0/design_guide/vmdcDesignGuideCompactPoD20/design.html#wp1100968

[13] https://www.vmware.com/files/pdf/network_segmentation.pdf

[14] http://bradhedlund.com/2013/07/07/what-is-a-distributed-firewall/

[15] http://www.cloudcomputing-news.net/news/2014/nov/21/top-cloud-computing-threats-and-vulnerabilities-enterprise-environment/ Android platform.

6. List of Acronyms

EDW	Enterprise Data Warehouse
VM	Virtual Machine
VMM	Virtual Machine Monitor
BI	Business Intelligence
IaaS	Internet as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
ETL	Enrichment Transformation Load
MPP	Massive Parallel Processing
IDS	Intruder Detection Systems
DOS	Denial of Service

Last Modified: December 15, 2015