Using the Cloud: The Cost of Encryption in IaaS

Kyle Cronin
Kyle.Cronin@dsu.edu

Wayne Pauli
Wayne.Pauli@dsu.edu

Michael Ham
mjham@pluto.dsu.edu

College of Business and Information Systems
Dakota State University
Madison, SD 57042, USA

Abstract

In recent years, the panacea that is cloud based computing has become the prescribed solution for many applications and services. Benefits such as greater availability, ease of maintenance, and reduced costs over physically owning and maintaining such systems has proven to be the remedy that many consumers have chosen to use. Despite the benefits to consumers using IaaS, there are downsides; chief among them being the security of data stored in cloud computing environments. This research examines the actual cost of encryption in IaaS, and in order to do so, a commercial IaaS vendor was chosen for the study. By utilizing a commercial IaaS vendor the study was completed in a real-world environment and the results then become more applicable and realistic. The time invested in these tests, compounded with the usage of guest-based encryption, could ultimately lead to a significantly more secure cloud-computing environment in a cost-effective manner.

Keywords: infrastructure as a service, IaaS, guest based encryption, cloud computing

I. INTRODUCTION

In recent years, Infrastructure as a Service (IaaS) has seen huge growth (Krigsman, 2012). With this growth both business and consumers alike realize the benefits of greater availability, ease of maintenance, and overall reduced costs of operations. Despite the many benefits to consumers using IaaS, the industry exhibits several downsides; chief among them being the security of data stored in cloud computing environments (O'Neill, 2011).

With various platforms for cloud computing available, we must define the specific type of cloud environments by which our study focuses. Based upon the NIST approved definitions, Infrastructure as a Service is defined as a cloud environment where a provider provides computing hardware and the appropriate connections for the consumer (Mell & Grance, 2009). Often times the expense for this service depends on the hardware allotted and the operating system provided. The IaaS provider often delivers these resources in the form of a virtual machine to the consumer for the use of storage and processing of data.

In a previous study, the problem of data security was clearly defined (Cronin, Pauli, & Ham, 2012). The individuals owning/operating/administering IaaS systems

have complete access to the data stored by the customers. In situations where commercial entities are using publically available cloud computing solutions, the observed risk is very high-- prohibitive in some situations. Related risks range from simple misconfiguration, outside attackers, and rouge insiders. Though separate situations, all of these factors exemplify attacks against the underlying infrastructure of the cloud environment in which the consumer has no control. These risks typically are applied to business entities. While the observations can also apply to consumers, the purpose of the study is to determine the impact on business entities.

A previous study defines three levels of protection in IaaS environments. These levels are: no encryption, host-based encryption, and guest-based encryption (Cronin et al., 2012). No encryption is the essence of zero protection. In the event of a compromise, no protections are in place to prevent the data from being readable by the attacking party. The overarching idea serves as the baseline for performance and cost metrics. A baseline of pre-encryption performance can be established for prospective customers by implementing no encryption whatsoever.

Host-based encryption provides minimal protection for the end user or consumer (Cronin et al., 2012). In the host-based encryption scenario, the owner of the IaaS solution provides encryption that is under his or her control. While host-based protection is better than no encryption, the control over encryption is still outside IaaS consumers' control. Such situations frequently exist in the software as a service (SaaS) setting.

Finally, guest-based encryption provides encryption within the guest operating system of a virtual machine (Cronin et al., 2012). This allows the consumer of the IaaS offering to control aspects of the encryption of his or her data. With this 'trust no-one' approach, the Iaas consumer pertaining to his or her data since the IaaS provider has no access to the keys required to decrypt data. In the event that the IaaS provider would attempt to read the data, they would only have access to the encrypted data.

To reiterate O'Neill's work (2011), a large concern is the security of the data that is stored in cloud computing environment. One of the foci of this applied research is to suggest and document that proper security of data stored in the cloud computing environments does come at a cost. Properly securing IaaS does erode the reduced cost of operation associated with the service. The actual cost of security could vary from provider to provider, but we must recognize that increased cost is an issue when applying appropriate security.

2. USER CONTROLLED ENCRYTPION IN SaaS ENVIRONMENTS

In the industry, Software as a Service (Saas) vendors are providing solutions for consumer-grade cloud storage. Popular SaaS products such as SpiderOak, SugarSync, and Dropbox provide such services, with varying implementations of protection (Ion et al., 2011). These different levels of encryption provide differing levels of data security.

An overarching focus of encryption methods used in SaaS providers is ownership of the encryption keys. SpiderOak takes advantage of a "Trust No One" or TNO approach ("Reasons Behind SpiderOak," n.d.). In such instance, only the user, not the SaaS vendor has access encryption keys to secure data. In contrast, SugarSync and Dropbox, use a differnet approach where both the user and SaaS vendor have the ability to decrypt stored data by maintaining copies of the encryption keys .

By studying encryption models that define which entities may access encrypted data in SaaS, we derived a similar framework for our IaaS study. This study is based on a TNO encryption implementation as a high level of security is desired. Industry-accepted and currently used encryption schemes in SaaS can be effectively applied in IaaS services.

3. IMPROVING PRIOR RESEARCH

As a continuation of research, the previously published study must be analyzed for faults within the process. Additionally, new consideration may be made for newly created technologies or processes. As a part of this analysis, the empirical results of the study conducted by Cronin, Pauli and Ham (2012) could be enhanced by these three observations:

- The study made assumptions based upon a small dataset
- The environment was isolated from any other users or participants

- The study was conducted in a highly artificial environment

In the initial study, the observations were highly supportive of the results. However, since only four observations were made, the data may be missing various statistical anomalies that would have risen in multiple observations.

The environment in which the study took place leveraged hardware that would be realistic in an IaaS environment. The workload on the hardware would be anomalous when compared to real world environments. In a real world setting, IaaS environment support several users, if not hundreds or thousands of users consuming resources within the same hardware set. These users would have significant (but not deliberate) natural influence on the observed results.

Finally, the environment in which the observations were made was highly artificial. The environment did not account for influence by outside users, and was 100% under control of the observers. In a more realistic IaaS environment, the users of the system have very little choice in the hardware or resources presented ("Amazon EC2 FAQs," n.d.). In essence, these users are given the opportunity to select the amount of memory provided to their virtual machine, the amount of disk space, a particular processing allotment, and which operating system will be used. Otherwise, the consumers of IaaS are unaware of the underlying infrastructure to the system.

While these observations are not critical to the research, they allow for grounds for improving the data collection process. These ideas contribute to a more thorough approach to be made in furthering the research conducted.

4. TESTING IaaS IN THE WILD

In order to calculate the cost of data operations in an IaaS environment, one must solicit a commercial IaaS vendor. For the purpose of this study, Amazon's Elastic Compute Cloud was selected. Leveraging a commercial IaaS provider allowed for the study to be completed in a production, real-world environment, making the numerical results more applicable to realistic scenarios as opposed to the study being limited to its overall conclusions.

The Elastic Compute Cloud (EC2) IaaS from Amazon is a service well suited to both evaluating IaaS with guest-based encryption and representing other IaaS providers. Amazon's EC2 allows consumers to select various levels of memory, storage, and processor allotment ("Amazon EC2 FAQs," n.d.). These options give enough flexibility to create a proper environment while maintaining a real-world approach where resources are actively in use and shared by other users around the world.

For the purpose of the study, a standard "small" and "medium" configuration was used. The small configuration consisted of 1.7 GB of memory, 1 virtual core with 1 EC2 Compute Unit (one EC2 compute unit is equivalent to the "CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor") ("Amazon EC2 Pricing," n.d.). These particular resources were judged to be comparable to a basic server configuration that would be common in a small to medium sized business. The cost for this instance was $0.115 per hour of the instance's operation. The medium sized instance had 3.75 GB of memory, 2 EC2 compute units with 'moderate' IO performance with an overall cost of $0.23 per hour (it should be noted that these costs are both in the US East region, prices vary by region). This was selected as a cost versus performance comparison due to the increased CPU capacity.

In addition to the base instance configuration, two 20GB volumes were added for each virtual machine using Amazon's Elastic Block Store ("Elastic Block Store," n.d.). These volumes were used for the simulated data storage. One volume was be encrypted while the other volume was left in an unencrypted state. This configuration allowed for the testing of guest-based encryption as well as providing a control group (performance of unencrypted data) for each iteration of the test.

Microsoft's BitLocker drive encryption software was used to implement guest-based encryption. Each virtual machine was configured with multiple storage volumes. The protected data in the test scenario resided on the additional storage volumes (not the operating system volume). This configuration emulated environments that frequently store data on disk volumes separate from the operating system.

This configuration provided the base test framework. A template instance was configured that was preconfigured to log in and run each individual test. Once the data I/O tests were

ran, the results of each test iteration were stored in a centralized location. After the results were saved, the instance was configured to automatically shutdown and terminate. In the end, each instance was active for less than ten minutes. The actual I/O test began after the operating system was fully booted for a three-minute period. Afterwards, the testing utility activated and completed a five minute simulation.

Careful creation of data reading and writing patterns was essential for the test to be successful and was accomplished by using Microsoft's SQLIO Disk Subsystem Benchmark Tool. SQLIO is designed for the purpose of simulating online transaction processing (OLTP) traffic on a particular system's hard disks or storage array (Ruthruff, 2007). SQLIO allows careful configuration of very specific variables, including testing read or write patterns, varying the number of threads active, the duration of the test, the specific size of the IO requests, and the type of IO requests (sequential or random). This is essential because varying types of file IO have different traffic patterns. SQLIO is able to simulate all of the traffic patterns necessary.

For the tests conducted, SQLIO was configured to test over a duration of five minutes per test. A five-minute duration allowed enough time for natural influences from other customers of Amazon's IaaS offering. The system was configured to operate with a database 40 megabytes in size. While 40 megabytes is not an excessively large database, many customers of a hosted IaaS solution are generally small or medium sized businesses, so the size is appropriate. Database certainly vary in size, but it was determined that this was a sufficient size to test.

As for reading and writing data, SQLIO was configured to read random data patterns with a block size of 8 kilobytes bytes (Wilson, Ruthruff, & Kejser, April 2010). Typically DBMS servers frequently perform read operations that are generally 8 KB in size. As for write sizes, 64 kilobyte blocks were used. As with read sizes, DBMS, or Microsoft's SQL Server specifically, tends to write sequential blocks of 64 KB chunks of data. While all DBMS installation and configurations may vary, these were determined to be the most appropriate sizes for use in the testing environment.

With the test bed implemented, some discussion of what guest-based encryption accomplishes is required. The notion of protecting data requires that a control maintained by the consumer (and not the IaaS provider) be implemented that prevents access to data. This task is accomplished by using encryption, such as that implemented in SaaS environments. By encrypting data, the consumer can prevent other unauthorized individuals from accessing the data.

With the benefits of guest-based encryption, two primary drawbacks must be considered: support and performance. In the case of support, it should be noted that Microsoft does not specifically support using its BitLocker technology within a virtual environment ("Planning for Hyper-V Security," 2009). One potential reason for Microsoft's lack of support revolves around the second primary drawback: performance. As with any encrypted system, the higher the levels of encryption, the greater impact to performance will be observed.

With the study conducted, the performance of guest-based encryption was compared with the control group. The control group ran the exact same data IO tests without any encryption in place. Using the testing utility, several key metrics are monitored:

The end results produced from SQLIO give us two primary metrics: MB/s and IO/s. When comparing storage mechanisms, these two values are essential for comparison. Megabytes per second (MB/s) define how much data can be transferred per second (Lowe, 2010). While similar, MB/s is different from input/output operations per second (called IOPS or IO/s). IOPS define the number of operations that were attempted per second. Generally, these two values are mutually inclusive- high MB/s rates can correlate with higher IOPS, though this is not always the case. For the purpose of this study, our primary focus is on the specific MB/s patterns established. For our results, a higher throughput in MB/s will mean a better performing system.

5. RESULTS: THE ACTUAL COST OF ENCRYPTION

Upon the collection of all the results, they were averaged together for easy comparison. Table 1 displays the aggregated data for each testing platform.

Table 1 Aggregated Performance Data

Instance Size	Small Instance	Medium Instance
Encrypted Read	3519 MB/s	6892 MB/s
Unencrypted Read	4812 MB/s	11639 MB/s
% Change Read	**-27%**	**-40%**
Encrypted Write	523 MB/s	628 MB/s
Unencrypted Write	590 MB/s	644 MB/s
% Change Write	**-11 %**	**-2%**

As seen in the results pertaining to the small instance, guest based encryption caused a performance drop in the input/output rates of 27% in the data read rates with the small instance. Similarly, an 11% decrease is observed with the medium instance. This drop is caused specifically by the overhead required when encrypting data to be written to the virtual disk since it was the only change in the test scenarios. As with any encrypted system, a negative performance impact is expected.

The purpose of completing the test with two different sized instances is to observe the offset. Ideally, if an organization requires the performance of a small instance but needs encryption a significant impact will be observed. Viewing these results, it can be determined that a medium sized instance would be sufficient to fill the performance gap.

With these results in place, we have determined that a medium sized instance is an effective capacity increase when using guest-based encryption. A medium sized increase would allow the data IO rates to essentially be unaffected by guest-based encryption in a situation where data protection is important.

The overall cost of guest-based encryption in this situation can be determined by taking the cost of a small instance compared with the cost of a medium sized instance, since the medium instance establishes the baseline performance. Thus, the overall annual costs can be computed as:

365 days $\times$ 24 hours $\times$ \$0.115=\$1007.40

$365\ days \times 24\ hours \times \$0.23 = \$2014.80$

$\$2014 - 1007.40 = \$1007.40\ additional\ cost$

With these basic calculations, in a situation where guest based encryption is required the additional cost of a small instance would be \$1,007.40 annually.

6. CONCLUSION

To summarize the data collected, it can be determined that the cost of guest-based encryption is relatively low when implemented with Amazon's EC2. While a 50% increase initially seems excessive, the cost of operating a server for 24 hours a day and for 7 days a week given the resources granted is relatively cost effective for small to medium sized businesses. To enunciate this fact, it can be observed that no upfront hardware cost is required to launch such instances.

In the overall scope of this research, we have determined that in the situation where a small Amazon EC2 instance is required, only a small upgrade is needed to prevent performance loss with guest-based encryption. More significantly, we have created a testing procedure for evaluating these systems. While Amazon EC2 was the focus company of this study, we have clearly demonstrated that an IaaS provider can be evaluated for performance with guest-based encryption with very little cost and minimal time involved. It should be noted that each business entity should evaluate the system to determine if any additional factors introduced by differing business software will influence the results. This scenario specifically emulates a SQL Server environment's storage capabilities. The time invested in these tests, compounded with the usage of guest-based encryption, could ultimately lead to a significantly more secure cloud-computing environment. With this in mind, we look towards industry to provide IaaS solutions with the security features in place to easily provide the protection required for small business database systems.

7. REFERENCES

Amazon EC2 FAQs. (n.d.) Retrieved 27 June 2012, 2012, from http://aws.amazon.com/ec2/faqs

Amazon EC2 Pricing. (n.d.) Retrieved 25 June 2012, 2012, from http://aws.amazon.com/ec2/pricing/

Cronin, K., Pauli, W., & Ham, M. (2012). Using the Cloud: Keeping Enterprise Data Private

Journal of Information Systems Applied Research, 5(3), 24-31.

Elastic Block Store. (n.d.) Retrieved 25 June 2012, 2012, from http://aws.amazon.com/ebs/

Ion, I., Sachdeva, N., Kumaraguru, P., Srdjan, #268, & apkun. (2011). *Home is safer than the cloud!: privacy concerns for consumer cloud storage*. Paper presented at the Proceedings of the Seventh Symposium on Usable Privacy and Security, Pittsburgh, Pennsylvania.

Krigsman, M. (2012). Research data on cloud growth and focus Retrieved 9 July 2012, 2012, from http://www.zdnet.com/research-data-on-cloud-growth-and-focus-7000000488/

Lowe, S. (2010). Calculate IOPS in a storage array [What drives storage performance? Is it the iSCSI/Fiber Channel choice? The answer might surprise you. Scott Lowe provides insight into IOPS.]. Retrieved from http://www.techrepublic.com/blog/datacenter/calculate-iops-in-a-storage-array/2182

Mell, P., & Grance, T. (2009). *The NIST Definition of Cloud Computing*. (Special Publication 800-145). Retrieved from http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

O'Neill, M. (2011). SaaS, PaaS, and IaaS: A security checklist for cloud models. *Cloud Security*, from http://www.csoonline.com/article/660065/saas-paas-and-iaas-a-security-checklist-for-cloud-models?page=1

Planning for Hyper-V Security. (2009), from http://technet.microsoft.com/en-us/library/dd283088(v=ws.10).aspx

Reasons Behind SpiderOak. (n.d.), from https://spideroak.com/whyspideroak

Ruthruff, M. (2007). SQL Server Best Practices Article. Retrieved from http://technet.microsoft.com/en-us/library/cc966412.aspx - EDAA

Wilson, E., Ruthruff, M., & Kejser, T. (April 2010). Analyzing Characterizing and IO Size Considerations. Retrieved from http://technet.microsoft.com/library/cc850692(office.12).asp