A framework for data security in the cloud using Collaborative Intrusion Detection scheme

Upasana Nagar
School of Electrical and Data Engineering,
Faculty of Engineering and IT,
University of Technology Sydney
Upasana.T.Nagar@student.uts.com.au

Xiangjian He
School of Electrical and Data Engineering,
Faculty of Engineering and IT,
University of Technology Sydney
Xiangjian.He@uts.edu.au

Priyadarsi Nanda
School of Electrical and Data Engineering,
Faculty of Engineering and IT,
University of Technology Sydney
Priyadarsi.Nanda@uts.edu.au

Zhiyuan(Thomas) Tan
School of Computing,
Edinburgh Napier University
z.tan@napier.ac.uk

ABSTRACT
Cloud computing offers an on demand, elastic, global network access to a shared pool of resources that can be configured on user demand. It offers a unique pay-as-you-go feature which is based on measured usage and can be compared to other utility services like electricity and water in everyday life. The advantages of cloud computing are lucrative for well-established organizations looking to reduce infrastructure cost overheads. It is equally appealing for start-up organizations as they need not invest in infrastructure and take advantage of the cloud. Thus, cloud computing promises huge cost savings and minimal management efforts. However, the users are not quite confident in entrusting their data to the cloud due to security threats and risks perceived in the cloud domain. Issues involving privacy requirements for the cloud and best practices in the cloud are suggested in this paper. Although the cloud provider ensures security in the cloud yet the flow of data, storage location, data computing process and security breaches are not transparent to the cloud customer. This distrust and lack of control on data is a major hindrance for potential cloud customers in adopting the cloud models for their businesses. Hence there is a need to research this security gap. Further cloud systems are also susceptible to the existing network attacks such as the Distributed Denial of Service (DDoS) attacks. Intrusion Detection Systems (IDSs) are widely used to detect malicious activities and are classified as Host based or Network based. However existing solutions with IDSs involving DDoS and other non-detectable events may not be suitable in applying to the cloud due to distributed data storage and a major shift in Internet access mechanisms offered by cloud providers. Hence there is a strong need to analyze an appropriate IDS to counter DDoS attacks in the cloud. In this paper we propose a novel framework for data security in the cloud using Collaborative Intrusion Detection (CIDS) scheme. 1

KEYWORDS
Cloud Security, Collaborative Intrusion Detection, HIDS, NIDS, Alert Correlation

ACM Reference format:
https://doi.org/10.475/123_4

1 INTRODUCTION
Cloud computing represents one of the most significant shifts in information technology and is of great interest for academic researchers and IT business community. The National Institute of Standards and Technology (NIST) provides the definition of cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [16]. The cloud computing delivery model offers computing as a utility, on demand and pay as you go service which is lure for small and medium IT organizations to get on board. Cloud computing can drastically reduce the infrastructure overheads, reduce capital cost and focus on business competencies for organizations. As much as the popularity of cloud is increasing, it is observed that potential customers are still sitting on the fence. They seem to be apprehensive and nervous in adopting cloud computing for their organization mainly due to the security gaps in cloud. Hence the importance of cloud computing security cannot be ignored.

Cloud computing security deals with securing various aspects of security comprising of data transparency, data authenticity, data authorization etc. It essentially covers the existing security challenges

1This is an abstract footnote
that are extended to the cloud as well as security risks that are introduced due to the unique nature of cloud architecture and deployment methods. Several researchers [2][7][8][12][13][15][19][22][25][27] have discussed the security threats and risks within the cloud and proposed approaches to make cloud computing a secure and trustworthy domain. However, there is much more to be achieved, in order to ensure that the fears of potential cloud costumers are addressed particularly in the face of new applications and the ways data are accessed.

The threats and risks in cloud computing are discussed further in the following sections. The main cause of concern in the cloud is the lack of user control over their data once it propagates to the cloud. Users are unsure of the level of accessibility of the cloud provider in their data. For an organization / cloud customer, this is the biggest obstacle in the cloud domain as the cloud customer does not want the cloud provider to access their personal and sensitive data and manipulate it for monetary gains or malicious acts. The security measures currently in place even do not leave much control in the hands of the data and thus makes them feel vulnerable and unsafe in the cloud. The data owners have rather very little choice but to trust the cloud providers and the security they offer in the cloud. The data owners have little or no visibility of the security and processes once their data is uploaded to the cloud and hence feel dependent upon the cloud provider. This report presents and investigates future research that aims to propose a scheme by which users can have better access control over their data in the cloud, minimizing the cloud provider’s access to user data and yet effectively reaping the benefits of cloud computing.

Intrusion detection can be defined as the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions, which attempts to protect confidentiality, integrity, availability, or even bypassing the security mechanisms of a computer or network[1]. IDS can be based on software or hardware that is strategically placed at appropriate detection points in a system or networks and automatically detect possible intrusions attempt and stop them from any future attacks. There are several factors based on which the IDS can be classified. Depending upon the target that needs to be monitored and protected, IDS can be classified as Host-based Intrusion Detection Systems (HIDS), Network-based Intrusion Detection System (NIDS) and Distributed Intrusion Detection System (DIDS). IDS can be further classified based on the deployment methods such as; software based IDS, hardware based IDS or VM based IDS[23].

Any IT system is threatened by either outsider attacks or insider attacks. Researchers [5][27] believe that the concept of cloud computing is derived from the existing grid computing and distributed computing system. However they also note that implementing existing IDS may not be effective in the cloud environment due to the unique characteristics of the cloud such as visualization, cloud data storage and computing over various geographical locations. Thus we understand the need to have an IDS for the cloud environment, which will address the gaps and challenges presented by the cloud architecture and its implementation.

2 RELATED WORKS

2.1 Cloud architecture and deployment models

It’s argued that cloud computing is not completely a new concept; it has intricate connection to the thirteen-year established Grid computing paradigm, and other relevant technologies such as utility computing, cluster computing and distributed systems in general[5]. In the light of this argument it may well be said that cloud computing essentially inherits all the security issues from the existing systems. But novel security issues are introduced due to its architecture and features which will be discussed further in the paper. Researchers point out that cloud computing being an advance model, it targets improving features of the existing security model. However it also threatens the security of existing technologies when deployed in the cloud environment[25].

2.1.1 Cloud service delivery models. The cloud architecture can be divided into three categories (1) Infrastructure as a Service (IaaS) (2) Platform as a Service (PaaS) (3) Software as a Service (SaaS)[24]. These three cloud delivery models are depicted in Figure 1.

Infrastructure as a service (IaaS) is the bottom most layer of cloud architecture and provides computer infrastructure like servers, data center space or network equipment as a service. Amazon EC2 is an example of IaaS. The customer does not control the infrastructure in the cloud but merely uses the rented machines. The ability to provide physical and virtual resources provides the consumers benefits of saving cost and time to set up the infrastructure[25].

Platform as a service (PaaS) is the middle layer of the cloud service delivery model. It offers customers the flexibility to use the cloud infrastructure and deploy self-created or purchased applications using programming languages and use the libraries, services and tools supported by the provider. Customers can only control their own applications and the associated environment but not the underlying cloud infrastructure[25]. Google App Engine is a good example of PaaS which provides a development environment using Python, Java and Go Programming languages[24].

Lastly Software as a service (SaaS) is the top layer of cloud architecture and lets the customer use the software applications provided by the cloud vendor via a client interface such as a web browser or program interface. The customer does not control the cloud infrastructure or the application configurations. Salesforce.com is an example of SaaS[25].

2.1.2 Cloud deployed models. Further as seen in Figure 1. cloud services can be deployed as (1) Private cloud (2) Public cloud (3) Community cloud (4) Hybrid cloud. Private cloud is available for employees of single specific organization and can be located in the organization or at an offsite location and may or may not be managed by a third party. Public cloud is available for access by the general public and can be owned, managed and used by government, academic or business organization. It is not located on the consumer site but at the cloud provider premises. Community cloud is deployed to be used by groups or communities having a common mission or activities. It can be located on site or off site. Hybrid cloud is built with combination of two or more either private, public and/or community cloud, which have some common technology to bind the two distinct cloud infrastructure to provide data and application portability[16].
2.1.3 Characteristics of cloud computing. Cloud computing is gaining popularity due to its several characteristics like on-demand pay-as-you-go service, resource pooling, rapid elasticity, broad network access, fast deployment, virtualization and multi tenancy. Despite all these positives, the consumer is not yet convinced to adopt cloud computing with vigour. This reluctance can be attributed to the underlying security gaps in cloud computing. [12] mention in their survey paper that consumer perception is the main factor that inclines the potential consumer to opt for cloud computing or vice versa. Customers conduct survey of the cloud provider and base their decision on the organizational need, past experience and the facilities offered by the cloud provider. This forms a basis of expectation for the customer. Cloud providers however are unable to provide a service as promised to the customer due to the security gaps. Also there is lack of transparency and trust between the customer and the cloud provider in terms of how and where customer data is stored and which employees on the provider side have got access to it.

2.2 Cloud computing security gaps

Cloud Security Alliance (CSA) has prepared and published a document on the top threats to cloud computing[8]. This research lists the top security threats as

- Abuse and nefarious use of cloud computing
- Insecure application programming interfaces
- Malicious insiders
- Shared technology vulnerabilities
- Data loss/leakage
- Account, service and traffic hijacking
- Unknown risk profile

If the cloud providers are not strict with security governance then malicious users and criminal minds can exploit and abuse the loopholes in cloud computing many ways. The cloud provider does not have authority to monitor the user activity based on the privacy laws[12]. Users with ill intent can take advantage of this weakness and use free trials offered by the cloud provider to launch attacks in the cloud. The attack cannot be detected immediately in real time and the provider is notified only after the attack is commenced via the security log or notification, which might be too late. Another threat to cloud computing is posed by insecure application programming interfaces provided by the cloud provider. In an attempt to provide an extremely user-friendly cloud interface, the cloud vendor might compromise the security[12]. The authors mention that in such cases the damage can be to the extent of losing complete control of the cloud infrastructure.

Even though vendors do not reveal what access levels are provided to their employees, someone with greater access level can easily engage in malicious activities and gain access to customer data and tampers with it. The authors mention that if the cloud provider does not have a breach notification policy in place, chances are that the customer may not be notified about the security breach. Hence, the responsibility to some extent lies with the customer to thoroughly read the service level agreement before getting on board[12].

Shared technology vulnerabilities are introduced in cloud due to its virtual nature. The same survey paper mentions hypervisors
being used for virtualization technology to create virtual machines and operating systems. It is highly possible for the attacker to take control of a privileged user account and further run virtual machines (VM) that belong to other customers. Another security lapse can happen when the attacker rents out a VM on which the template image of operating system (OS) has been copied. This allows him to access the entire configuration and rights and can even let the attacker use an image from an untrustworthy source.

Data loss and leakage is of prime concern for the cloud customers. If the customer does not exercise caution then the low cost cloud solution may result in data loss due to damaged/ corrupt storage, alteration or deletion of records without backup, failure of drives and accidental deletion of partition. Lack of procedure and process knowledge of the cloud employees can result in unintentional accidental loss of data which can cause great damage to the customer. Data theft is equally an important concern for the cloud customer. It is highly possible for the cloud provider to have ill intent and store data or alter it for their self-interest.

Account, service and traffic hijacking attacks are inflicted when the user credentials are stolen and misused. Attackers can use several methods for stealing like phishing, fraud, Denial of Service (DoS) and account hijacking[12].

The inability or sometimes reluctance from the cloud provider to disclose the cloud infrastructure security procedures and releasing the security logs and data to the customer, leads to the risk of unknown profile. The customer does not get the real picture of the security measures provided by the cloud provider and hence exposes himself to unknown risks.

A survey paper[25] has described the security issues based on the service models of cloud computing. According to the authors the key security elements that need to be considered for the SaaS application development and deployment are data security, network security, data locality, data integrity, data segregation, data access, authentication and authorization, data confidentiality, web application security, identity management and sign on process. The consumer gets some application control in PaaS service model and is responsible for the application security however the vendor is liable to ensure security for the layers below the application layer and should ensure that the customer data is protected and not exchanged across applications. In IaaS service model, data is stored within the provider hardware. Indeed the developer at consumer side has much control on the security of the hardware, but the owner of data faces a big challenge in controlling the data in a virtual environment. Cloud architecture is not restricted by geographical location of its VMs and this factor increases the severity of the risk faced by the data owner. It is highly possible for the data owner to lose data due to a security lapse in cloud and lack
of transparency from the cloud provider. Further to this survey, Figure 2. lists the various attacks that are based on the cloud service delivery model[9].

Another survey paper[11] has listed attacks based on the cloud component under threat. According to the authors, the attacks in cloud can be network based attacks, VM based attacks, Storage based attacks and application based attacks. Port scanning, botnets and spoofing attacks can be classified as the network based attacks in cloud. VM based attacks in cloud systems can be of various types as Cross VM side channel attacks, VM creation attacks, VM migration and rollback attacks and VM scheduler based attacks. Data scavenging and data duplication are included under the storage based attacks in cloud. Application based attacks compromise of mainly three types, namely, Malware injection and steganography attacks, shared architectures and lastly web services and protocol based attacks. The paper further lists the implications of these attacks as violation of data protection, malicious manipulation of data, denial-of-service and theft-of-service.

2.3 Intrusion detection schemes in cloud environment

As discussed earlier in the literature review, cloud computing is seen as an extension of the existing grid and distributed computing system; hence it also inherits the security issues of the existing systems. However the cloud characteristics, framework and the potential threats and attacks described above, differentiate it from the existing system. It is easy to infer that implementing existing standalone IDS systems directly within the cloud environment might have its shortfalls. Various traditional attacks that endanger cloud security are IP spoofing, Address Resolution Protocol (ARP) spoofing, Routing Information Protocol (RIP) attack, DNS poisoning, flooding, Denial of Service (DoS) and Distributed Denial of Service (DDoS), user to root attack, port scanning, attacks on virtual machine (VM) or hypervisor and back door channel attacks [18]. Firewall can be used as a defense mechanism in cloud; however it provides detection and protection for outsider attacks and is not very efficient when there is an insider attack. Thus building and implementing IDS for the cloud infrastructure is of great priority. Signature based detection and Anomaly based detection are two of the main techniques used in traditional IDS and can also be applied in the cloud at certain level.

2.3.1 Signature based detection. Signature based detection is also referred as misuse based detection and involves the comparison and matching of the incoming network traffic with an existing set of rules (signatures) written by domain experts. If there is a signature mismatch between the incoming packet and the domain signature database, then it is classified as an attack. Hence signature based detection has high detection accuracy and low false positives. However the drawback of this system is its inability to detect any new attack or attacks whose signature is not known and thus require continual update to the signature database for latest attacks. SNORT and Bro are examples of traditional signature based IDS commonly used[10]. A number of researchers have done some useful work on using signature based IDS in the cloud [2][13][15][23]. It is noted that signature based IDS can be positioned at the front end in the cloud (at VMs), enabling the detection of outsider attacks.

It can also be positioned at the back end of the cloud for detecting internal or external intrusion.

2.3.2 Anomaly detection. Anomaly detection based IDS learn the normal behaviour of a user profile, and based on this they compare the incoming user profile to detect if the behaviour is legitimate or not. If there is any anomaly or deviation from the normal behaviour profile, alarm is generated. The advantage of anomaly detection based IDS is its capability to detect new attacks and variations of the existing attacks. The main drawback however is increase in false positives which might be caused by detection of a new traffic which is not malicious. A number of authors have proposed several intrusion detection systems that are based on anomaly detection technique[3][6][7][27]. Anomaly detection can be applied in the cloud environment at various layers of the cloud architecture, which makes it a challenge to monitor intrusions over multiple layers of the cloud[18].

One of the recent survey papers[17] has taken into consideration the novel architecture of cloud and has classified the IDS deployment approaches in cloud into the below categories

- In-Guest agent based approach
- IN-Virtual Machine Monitor (VMM) agent based approach
- Network Monitor based approach
- Shared technology vulnerabilities
- Collaborative agent based approach
- Distributed approach

These approaches give a better understanding of the ways in which IDS can be implemented in the cloud. It is noteworthy that unlike traditional system security which uses HIDS and NIDS approaches, the above classification caters to the architecture or the cloud and its unique features of virtualization, scalability, elasticity and multi tenancy. The In-Guest agent based approach can draw parallel with the traditional HIDS as it is configured inside the Virtual machine (VM)[17]. Similar to HIDS, this method monitors the incoming and outgoing packet on the host VM and performs the checks and analysis for detection of any intrusion. So it is under the complete control of the cloud machine user. However, modern day attacks are more distributed in nature and hence the In-guest agent based approach may not be best suited to detect the complex distributed attacks that may occur in the cloud. In-VMM agent based approach requires nested virtualization of the IDS in or below the hypervisor. Network Monitor based approach is similar to the NIDS as it requires the IDS to be positioned at the network entry points in cloud[17]. Distributed approach involves the IDS to be deployed on the host VMs; however the IDS policies and alerts are created and analysed by the cloud administrator. The Collaborative agent based approach in Cloud is of particular interest as it deploys the IDS at various locations in the cloud. The concept is based on collaboration of the various IDSs located at multiple cloud network points and exchanging the alert details amongst themselves to generate a complete attack map based on the alerts from various IDS. The IDS can be deployed at the VM, network switches and the hypervisor and hence the collaborative agent based approach is most suitable to detect the distributed and co-ordinated attacks in the cloud environment[17].
3 PROPOSED COLLABORATIVE AGENT BASED IDS FRAMEWORK FOR CLOUD SECURITY

Our proposed CIDS framework for cloud is illustrated in Figure 3. This framework is based on the idea proposed in [26] and involves two types of nodes (a) Cooperative nodes and (b) Central coordinator. As shown in Figure 3, a dedicated NIDS is deployed on a virtual switch or VM at the cloud entry point. This monitors the incoming network packets. We propose to use SNORT tool for this function[21]. The host VMs are monitored by individual HIDS. We propose to use open source tool HIDS in our framework. The alerts from the NIDS and HIDS are collected by the central coordinator. The alerts are in the Intrusion Detection Message Exchange Format (IDMEF)[4]. The central coordinator is the heart of the framework and is responsible for monitoring, managing, analysing and correlating the alerts and hence creating a comprehensive information of attacks within the public cloud. The central coordinator reports the comprehensive attack information with the user who can then initiate required preventive measure to contain the attack. However, the drawback of this scheme can be single point of failure at the main central coordinator. However as proposed in [26], we can incorporate a backup central coordinator. This may make the CIDS framework somewhat bulky, but we can try and reduce the weight of our framework in future. CIDS can be implemented as a centralized, hierarchical or decentralized topology[26]. We propose to deploy the centralized topology.

As pointed out earlier, public cloud has multiple entry points and multiple active instances at a given point of time. Hence any potential intruder has several pathways to launch an attack in the cloud. These intrusions can be coordinated and spread out in parts over the various VMs in the cloud; hence they can go undetected by the conventional standalone HIDS or NIDS[26]. The collaborative IDS scheme for cloud proposed above plugs in this loophole of the conventional IDS. It does so by collecting alerts from all the different IDSs located strategically across the cloud network and correlating this alert information to provide detection of valid attacks[26]. The main advantages provided by our CIDS framework include the high intrusion detection efficiency with efficient use of the IDS resources in the cloud, low false alarms, low computational costs. The framework employs fusion of alerts collected across the IDS, thus deriving a more comprehensive knowledge of the attacks in the cloud[14]. The cooperative intrusion detection schemes proposed by researchers in [20] [13], deploy NIDS in the cloud and aim to detect attacks in the cloud. Our proposed scheme on the contrary provides the benefits harnessed by deploying both HIDS and NIDS in the cloud at strategic points. The alert correlation analysis will be rich with attack information and will enable the cloud user to take quick and timely preventive action to counter any large scale coordinated attack in the cloud.

4 CONCLUSION

In this paper we have proposed a collaborative IDS framework for public cloud environment. We are currently developing the laboratory test environment for the public cloud and setting up the framework. We also aim to simulate attacks in this test environment and evaluate our framework for large cloud based coordinated DDoS attacks. Our focus in future will be more invested on the fusion, correlation and analysis of alerts in the central coordinator and hence improve the detection rate of our CIDS framework for cloud.
REFERENCES


[26] Zhuyuan Tan, Upasana T Nagar, Xiangjian He, Priyadarshi Nanda, Ren Ping Liu, Song Wang, and Jiankun Hu. 2014. Enhancing big data security with collaborative intrusion detection. IEEE cloud computing 1, 3 (2014), 27–33.