Future Generation Computer Systems 28 (2012) 833–851

Contents lists available at SciVerse ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing

Md. Tanzim Khorshed, A.B.M. Shawkat Ali *, Saleh A. Wasimi

School of Information and Communication Technology, CQUniversity QLD 4702, Australia

ARTICLE INFO

Article history:
Received 15 August 2011
Received in revised form
11 January 2012
Accepted 18 January 2012
Available online 27 January 2012

Keywords:
Security
Threats
Machine learning
Trust
Cloud computing

ABSTRACT

The long-term potential benefits through reduction of cost of services and improvement of business outcomes make Cloud Computing an attractive proposition these days. To make it more marketable in the wider IT user community one needs to address a variety of information security risks. In this paper, we present an extensive review on cloud computing with the main focus on gaps and security concerns. We identify the top security threats and their existing solutions. We also investigate the challenges/obstacles in implementing threat remediation. To address these issues, we propose a proactive threat detection model by adopting three main goals: (i) detect an attack when it happens, (ii) alert related parties (system admin, data owner) about the attack type and take combating action, and (iii) generate information on the type of attack by analyzing the pattern (even if the cloud provider attempts subreption). To emphasize the importance of monitoring cyber attacks we provide a brief overview of existing literature on cloud computing security. Then we generate some real cyber attacks that can be detected from performance data in a hypervisor and its guest operating systems. We employ modern machine learning techniques as the core of our model and accumulate a large database by considering the top threats. A variety of model performance measurement tools are applied to verify the model attack prediction capability. We observed that the Support Vector Machine technique from statistical machine learning theory is able to identify the top attacks with an accuracy of 97.13%. We have detected the activities using performance data (CPU, disk, network and memory performance) from the hypervisor and its guest operating systems, which can be generated by any cloud customer using built-in or third party software. Thus, one does not have to depend on cloud providers' security logs and data. We believe our line of thoughts comprising a series of experiments will give researchers, cloud providers and their customers a useful guide to proactively protect themselves from known or even unknown security issues that follow the same patterns.

1. Introduction

Cloud computing can be viewed as the transformation into reality of a long held dream called "Computing as Utility", it emerged into the market with a huge potential to fulfill this dream. It promises on-demand services for a customer's software, platform and infrastructure needs. In its fold, companies do not even need to plan for their IT growth in advance with this new "pay as you go" system. Already, there has been upbeat assessment about its great potential for utility, scalability and instant access features; but on the flip side, some are also apprehensive of security gaps involving for instance, trust, threats and risks.

While cloud computing has received mixed reviews from its customers, some experts describe it as the reinvention of distributed main frame model [1]. It could be the most significant shift in IT infrastructure area in recent times as it appears promising but still a great deal of work is warranted in the domain of security to minimize the gaps. At the time of writing this paper, we discovered a propensity in many small or midsized organizations to adopt cloud computing mainly to reduce upfront investment costs, minimize maintenance work in IT infrastructure and to enhance on-demand capabilities. However, there is a risk of depredation for not doing an assessment on security and privacy. Before we explore the security and privacy issues in cloud computing, it is worthwhile to revisit the definition of cloud computing.

In our quest for the definition of cloud computing, we perused books and articles [2–7] and came up with our own definition that is easy to comprehend and yet broad in its scope, which can be visualized in graphical form as described in Fig. 1. Put in words:

Cloud computing is a system, where the resources of a data center is shared using virtualization technology, which also provide elastic, on demand and instant services to its customers and charges customer usage as utility bill.

* Corresponding author. Fax: +61 7 4930 9729.
E-mail addresses: t.khorshed@cqu.edu.au (Md.T. Khorshed), s.ali@cqu.edu.au (A.B.M.S. Ali), s.wasimi@cqu.edu.au (S.A. Wasimi).

0167-739X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2012.01.006

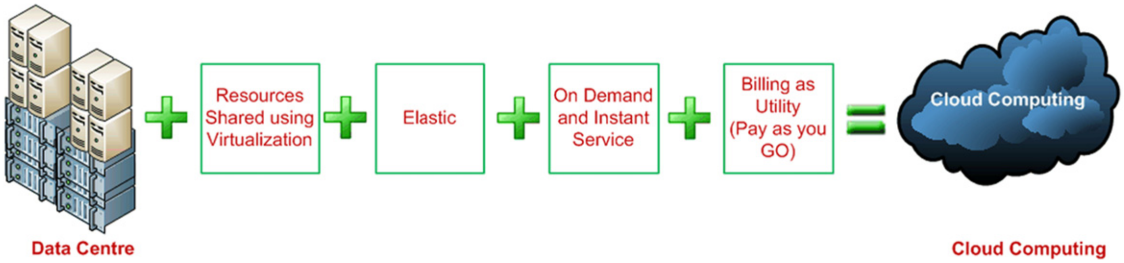

Resources Shared using Virtualization
Elastic
On Demand and Instant Service
Billing as Utility (Pay as you GO)
Cloud Computing
Data Centre
Cloud Computing

Fig. 1. Schematic definition of cloud computing.

Virtualization, elasticity, on-demand, instant service and pay as you go are the main characteristics that convert a data center into cloud computing. In a typical depiction, the word 'data center' may be restrictive because it could be any IT resource that can be shared using virtualization technology. But if we walk through any of today's cloud provider's office we will witness a large data center full of computer systems in the racks which are used to share resources. So, we may as well include the word "data center" to make our definition more relevant to the real world. We have noticed that some existing data center providers are already rebranding themselves as cloud providers taking advantage of their existing infrastructure as they do not wish to miss out on the "next big thing" in IT industry.

In some definitions, we found that experts have added the phrase "using internet technology" [8,9] as a must for cloud computing. But our interpretation does not make that feature imperative because on-premise single organization's private cloud would not need internet to access cloud services. Thus, we exclude internet from our definition. Furthermore, in cloud computing, virtualization is used to create multi-tenant architecture, but we did not use the word 'multi-tenant' in our definition to keep it simple as the encompassing word 'virtualization' is already there.

As with any change in IT infrastructure where there are accompanying novel risks and opportunities, cloud computing is no exception. Shared, on-demand nature of cloud computing expose it to some unique risks that have not been experienced before. In this paper, a survey of cloud computing with the main focus on gaps and their proposed solutions are presented. The presentation of the paper is in two discourses. The first discourse is on the survey for an easy but comprehensive definition of cloud computing and understanding its main aspects and gaps. The second discourse is on thoughts for some novel approaches to identify cyber attack types using modern machine learning techniques including rule-based learning and statistical learning theory. We believe our thoughts encapsulated through a series of experiments will give researchers, cloud providers and their customers the initiative to proactively protect themselves from known or even unknown security risks.

2. Review of cloud computing standards

Cloud computing standards are currently the topic of research of several groups and organizations. 'Cloud Standard Coordination' was formed in July 2009, their main "goal is to create a landscape of cloud standards work, including common terminology" [10]. In that vein they created a Wiki page where different cloud oriented Standard Developing Organizations (SDOs) can update their part of research [11]. We have visited each of these SDOs websites and attempt here to capture the essence of their areas of research. The intent is to give aspiring researchers a lead as to where to start in the sky of clouds.

Cloud Security Alliance (CSA) is a non-profit organization promoting the use of best practices, common level of understanding, awareness and guidelines for cloud related security threats [12–14]. The goal of CloudAudit working group, which has been working under the guidance of CSA from October 2010, is to provide a common interface and namespace for cloud providers to automate the audit, assertion, assessment, and assurance of their service environments so that their authorized clients can access the services using a similar secured interface [15].

Distributed Management Taskforce's (DMTF's) cloud efforts are focused on standardizing management protocols for interactions and development of cloud environments. To reach this goal, they have formed two working groups—Cloud Management Work Group (CMWG) and Cloud Audit Data Federation (CADF) work group [16]. To address convergence issues between cloud computing and telecommunications the European Telecommunications Standards Institute (ETSI) established the cloud project. Their particular interest is on the Infrastructure as a Service (IaaS) delivery model [17]. The National Institute of Standards and Technology (NIST) is a United States government agency; their long term goal is to provide specific guidance to the industry and government, they aim to shorten adoption cycle and identify gaps in cloud standards [18].

Open Grid Forum's (OGF) Open Cloud Computing Interface (OCCI) working group was originally formed to create remote management API for Infrastructure as a Service (IaaS), but their current release of open computing interface is even suitable for other service delivery models such as Platform as a Service (PaaS) and Software as a Service (SaaS) [19]. Open Cloud Consortium (OCC) works with development of standards for interoperability, benchmarks and open source reference implementations. They have several working groups working at the moment, such as The Open Science Data Cloud (OSDC) working group, The Open Cloud Testbed working group and Intercloud Testbed working group [20].

The Storage Networking Industry Association (SNIA) has created the Cloud Storage Technical work group with the aim of developing SNIA architecture related to cloud storage technology [21]. In May 2010 "the open group" merged "SOA and Security" and "Security in cloud" projects to form "Security for Clouds and SOA". Their main objective is to develop best practices and to describe and understand security and cloud security architecture [22]. The Open Cloud Manifesto group is working on a set of principles for the cloud community "in the belief that cloud computing should be as open as all other IT technologies". In their document, they pointed out choice, flexibility, skills and speed, and agility as goals for open cloud with six principles [23].

Before we discuss the gaps and unique security concerns of cloud computing, it is imperative that we portray the main aspects of cloud computing as many of those are actually generated because of its unique features.

3. Main aspects of cloud computing

'Cloud Computing' can be viewed as the evolution of 'Grid Computing'. Foster et al. [24] argue that "Cloud Computing is not a completely new concept; it has intricate connection to the thirteen-year established Grid computing paradigm, and other

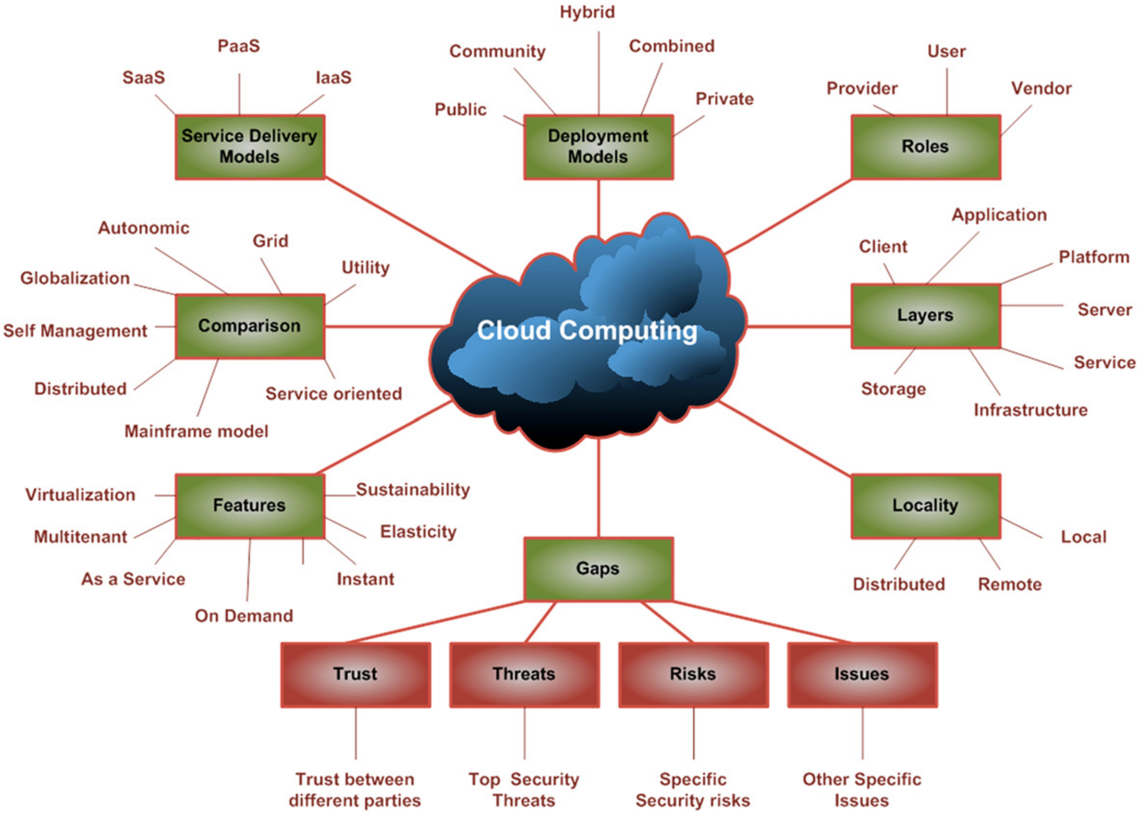


Fig. 2. Understanding cloud computing.

relevant technologies such as utility computing, cluster computing, and distributed systems in general''. The *European Grid Initiative (EGI) Design Study* represents an effort to establish a sustainable grid infrastructure in Europe. The latest version, EGEE-III project and other previous grid projects, nowadays is one of the largest existing distributed computing infrastructures for e-Science in the world supporting a variety of different international research collaborations [25]. Throughout the South–Eastern Europe, in the grid arena, the South–East European GRid e-Infrastructure Development (SEE-GRID) series of projects have established a strong human network in the area of scientific computing [26]. The latest project on grid, EELA-2 e-Infrastructure is comprised of a service grid and an opportunistic grid that federates computing resources from scientific institutions in both Europe and Latin America [27]. While development in grid computing can be considered as phenomenal, the development in cloud computing, which goes beyond the resource management aspect of grid computing, is just gathering pace. Essentially, in cloud computing an application accesses a service that allows on-demand resource provisioning and everything else that it entails.

Cloud computing inherits all the security issues from existing systems, for instance grid computing, plus the security issues that has been created due to its unique architecture and features. To understand these unique features, we first need to put into context the main aspects that form a cloud system. Jeffery et al. [28] have drawn a picture which makes it easy to understand the cloud system and its main aspects. There are other researchers [29,14,3,5,24,30,31,7,32] who also tried to organize and capture different aspects of cloud computing. After doing an extensive review of all these works we have constructed a new framework to visualize every detail of a cloud system as shown in Fig. 2.

In Fig. 2 we have categorized a cloud system into eight main aspects. These are features, comparison, service delivery models, deployment models, roles, layers, locality and, gaps which is a new addition. Each of these main aspects has at least three sub aspects. We linked all sub aspects with the relevant main aspects. The new contribution of this paper is the introduction of gaps as one of the main aspects because we believe it is too important to ignore. Furthermore, our assertion is that trust issues, security threats, security risk and some other specific cloud computing related issues are the main gaps of cloud computing.

4. Cloud computing gaps

Despite the huge potential that cloud computing has, so far, it has not been adopted by the consumers with the enthusiasm and pace that it deserves. This can be attributed to the gaps. The National Institute of Standards and Technology [18] contends that security, interoperability, and portability are the major barriers to a broader cloud adoption. A group of researchers from the University of California at Berkeley [29] identified 10 obstacles to cloud computing. These are: availability of service, data lock-in, data confidentiality and auditability, data transfer bottlenecks, performance unpredictability, scalable storage, bugs in large distributed systems, scaling quickly, reputation fate sharing, and software licensing. Ness [33] associates three major barriers to cloud computing, first, cloud depends on new approaches to security, second, cloud can break static networks, and third, network automation is critical. By contrast, Leavitt [34] describes six challenges which are: control; performance, latency and reliability; security and privacy; related bandwidth costs; vendor lock-in and standards; and transparency.

There may be so many ways to define gaps, compounded further by the fact that many parties are involved other than cloud providers and customers. However, more importantly, it is the perception of the customers which dictates whether they or their organizations are willing to join cloud computing that matters. What their organizations' expectations are and what services they

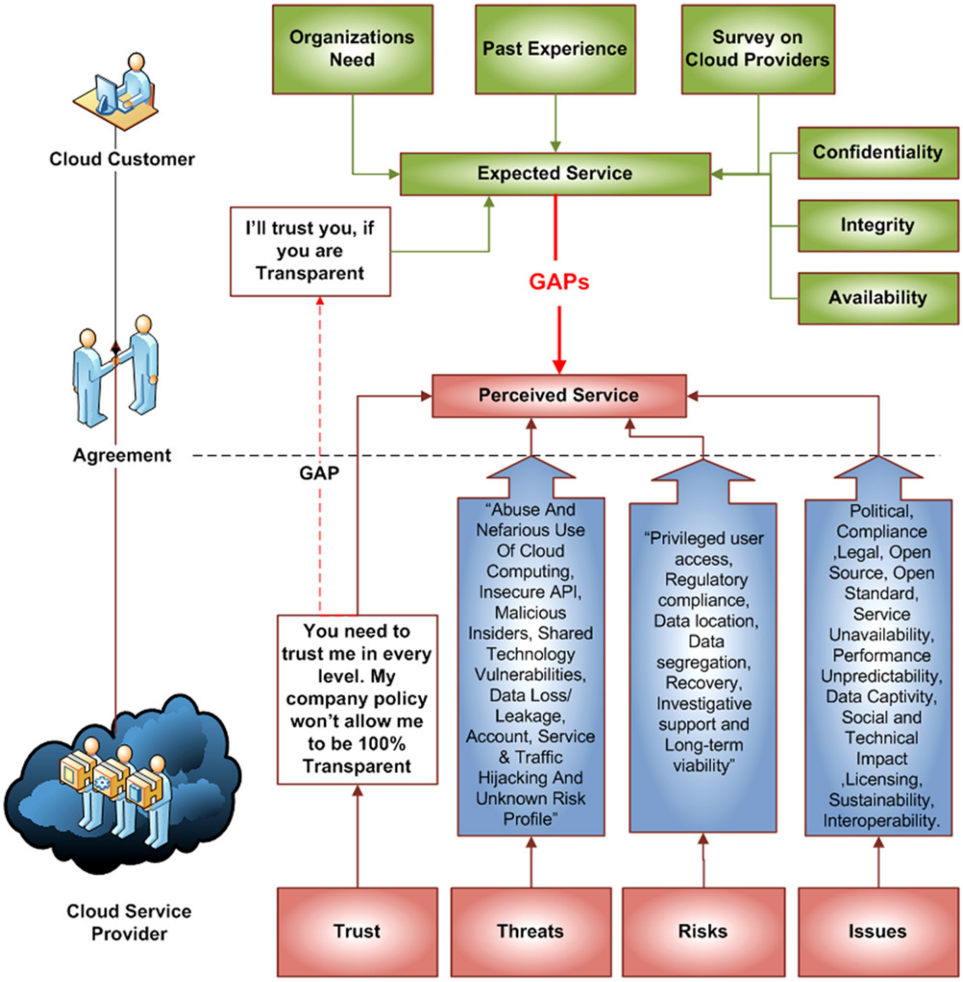

Organizations Need
Past Experience
Survey on Cloud Providers
Cloud Customer
Expected Service
Confidentiality
Integrity
Availability
I'll trust you, if you are Transparent
GAPs
Perceived Service
Agreement
GAP
"Abuse And Nefarious Use Of Cloud Computing, Insecure API, Malicious Insiders, Shared Technology Vulnerabilities, Data Loss/ Leakage, Account, Service & Traffic Hijacking And Unknown Risk Profile"
"Privileged user access, Regulatory compliance, Data location, Data segregation, Recovery, Investigative support and Long-term viability"
Political, Compliance ,Legal, Open Source, Open Standard, Service Unavailability, Performance Unpredictability, Data Captivity, Social and Technical Impact ,Licensing, Sustainability, Interoperability.
You need to trust me in every level. My company policy won't allow me to be 100% Transparent
Cloud Service Provider
Trust
Threats
Risks
Issues

Fig. 3. Understanding cloud computing gaps.

are going to receive from a particular provider are likely to be the key deciding factors in choosing a cloud provider. From a rather extensive review [29,34,33,18] we can define the cloud computing gaps succinctly as follows:

The factors that are slowing down migration to cloud computing from existing systems are cloud computing gaps.

In Fig. 3 we have drawn a diagram showing the gaps between cloud customers' expectations and deliverable services based on our understanding [12,13,29,35,14,5,30,31,34,33,18].

Cloud customers may form their expectations based on their past experiences and organizations' needs. They are likely to conduct some sort of survey before choosing a cloud service provider similar to what people do before choosing an Internet Service Provider (ISP). Customers are expected also to do security checks that are centered on three security concepts: confidentiality, integrity and availability. On the other hand, cloud service providers may promise a lot to lure a customer to sign a deal, but some gaps may manifest later as insurmountable barriers to keep their promises. As we can well appreciate, there is a gap between a customer's expectations and deliverable services. Many potential cloud customers are well aware of this, and consequentially, still sitting on the sidelines. They will not venture into cloud computing unless they get a clear indication that all gaps are within acceptable limits.

Notwithstanding the fact that cloud computing is still nascent in the information technology landscape and so many of the future threats may still be unknown, we can always analyze the gamut of lessons learned in the past and incorporate that in the new architecture. Mikkilineni and Sarathy [36] draw a parallel between the state of the data centers today and the evolution of the Intelligent Network (IN) infrastructure in telecommunications. They believe that the next generation cloud evolution would be a fundamental transformation. Chonka et al. [37] note that as security experts their experiences premonish the same mistakes that occurred during the development of Internet being repeated with cloud computing. They pointed out that functionality and performance are receiving unduly higher priority than security. Unfortunately, customers are less aware of the risks.

Tim Watson, Head of the computer forensics and security group at De Montfort University notes, "although one provider may offer a wonderfully secure service and another may not, if the latter charges half the price, the majority of organizations will opt for it as they have no real way of telling the difference" [38]. While referring to the seriousness about the security threats, George Wrenn, Security Solutions Director at Unisys suggested, "Customers must evaluate cloud infrastructure vendors on more than price and top feature sets before deciding to move critical systems and applications" [39].

4.1. Review of cloud computing security

Choubey et al. [40] have done a short but very specific review of cloud computing security and identified the key advantages, disadvantages and tradeoffs between cost and security. Subashini and Kavitha [9] have done a recent survey on the security risks that have been created by the sheer nature of different service delivery

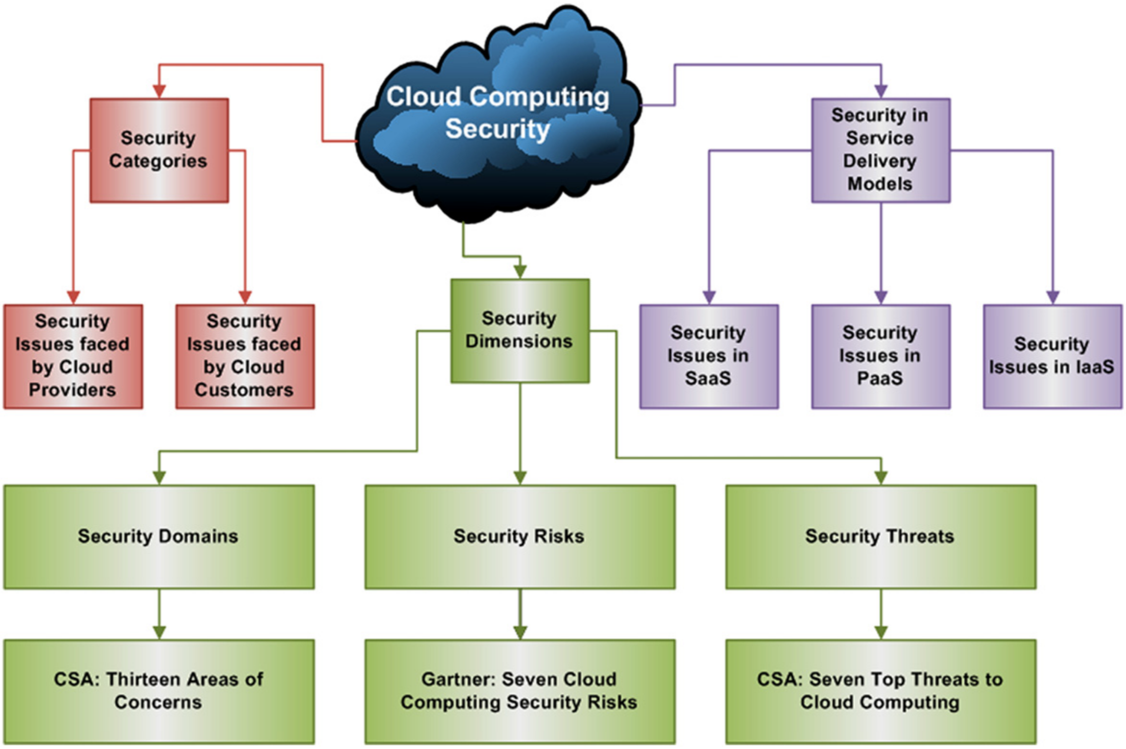

Cloud Computing Security
Security Categories
Security Issues faced by Cloud Providers
Security Issues faced by Cloud Customers
Security Dimensions
Security in Service Delivery Models
Security Issues in SaaS
Security Issues in PaaS
Security Issues in IaaS
Security Domains
Security Risks
Security Threats
CSA: Thirteen Areas of Concerns
Gartner: Seven Cloud Computing Security Risks
CSA: Seven Top Threats to Cloud Computing

Fig. 4. Graphical view of cloud computing security.

models. Rad et al. [41] have done a survey of cloud platforms that mainly focused on foundation, storage system, infrastructure service and integration. Chang et al. [42] have done a review on cloud business models where they classify these business models into eight types. They also discuss how the cloud cube model fits into each of these eight models. Srinivasamurthy and Liu [43] have done another survey on secure cloud architecture advantages and different security threats with some existing ways to minimize these threats. Rimal et al. [32] have attempted to classify the cloud architecture followed by the addition of their own survey findings on existing cloud services using their classifications.

From our observation of these and other analyses [12,13,35,14,5,42,40,44,41,32,43,9], we tried to integrate all the information and visualize cloud computing security in a snapshot which is presented in Fig. 4. We organized cloud computing security into three sections: security categories, security in service delivery models and security dimensions.

The existing surveys in published literature are mainly on categories and service delivery models. Though there are some very good research works on security dimensions, they are rather limited in scope and incomplete. Essentially, this has motivated us to write a review on cloud security dimensions. In this effort, we found two research works as most cited in literature, which are Cloud Security Alliance's research on top threats [13] and security farm Gartner's research [35,44]. Coincidently, both of these organizations picked the top seven threats and risks respectively.

4.2. Major security concerns in cloud computing

Some organizations and security farms have released their research findings on major security concerns to assist companies interested in joining cloud computing for them to make wise decisions being fully cognizant of the associated risks. This urges new customers to ask tough questions and consider getting a security assessment from a neutral third party before committing to a cloud vendor. In June 2008 the security farm Gartner published a report entitled "Assessing the Security Risks of Cloud Computing" [44]. In it, they identify seven specific security issues that customers should raise with vendors before selecting a cloud vendor. The specific issues are "privileged user access, regulatory compliance, data location, data segregation, recovery, investigative support and long-term viability" [35,44].

In November 2009, the European Network and Information Security Agency (ENISA) published another research document entitled "Cloud Computing: Benefits, Risks and Recommendations for Information Security". The document lists eight important cloud-specific risks which are: loss of governance, lock-in, isolation failure, compliance risks, management interface compromise, data protection, insecure or incomplete data deletion and malicious insiders. They have also discussed about risk management and provided recommendations [5].

Cloud Security Alliance (CSA) released version 2.1 of their document "Security Guidance for Critical Areas of Focus in Cloud Computing" in December 2009, where they identified thirteen areas of concerns in three major sections. These are "Section 1: Cloud architecture, Domain 1: Cloud computing architectural framework; Section 2: Governing in the cloud, Domain 2: Governance and enterprise risk management, Domain 3: Legal and electronic discovery, Domain 4: Compliance and audit, Domain 5: Information lifecycle management, Domain 6: Portability and interoperability; Section 3: Operating in the cloud, Domain 7: Traditional security, business continuity, and disaster recovery, Domain 8: Data center operations, Domain 9: Incident response, notification, and remediation, Domain 10: Application security, Domain 11: Encryption and key management, Domain 12: Identity and access management, Domain 13: Virtualization" [14].

CSA have published their research findings on the top threats to cloud computing in March 2010. The purpose of the research was to assist cloud providers as well as their potential customers in identifying the major risks and to help them decide whether or not to join in cloud infrastructure, and also, how to proactively protect them from these risks. The top seven threats they mentioned are

"Abuse and nefarious use of cloud computing; Insecure application programming interfaces; Malicious insiders; Shared technology vulnerabilities; Data loss/leakage; Account, service and traffic hijacking; and Unknown risk profile" [13]. However, it has also drawn criticism from some experts. Lacey [45] wrote in his blog about CSA research that it contains information on many general IT security problem areas but little on specific threats to cloud computing. He also stated that 'Unknown risk profile' (the absence of a risk assessment), 'Malicious insiders', 'Data loss or leakage' and 'Abuse and nefarious use of cloud computing' are too general to any network security issues.

It is our contention that the four research works mentioned above [13,14,5,44] on cloud computing security could be regarded as pioneering work to guide aspiring future researchers in this area. This however is not intended to downplay the importance of other security issues. For instance, Choo [46], Senior Research Analyst at the Australian Institute of Criminology, has pointed out that attacks targeting shared-tenancy environment, Virtual Machine (VM)-based malware, Botnet hosting, launch pad for brute force and other attacks, data availability, and rogue clouds are key risks. He also identifies espionage and regulation and governance as other potential risks. Moreover, some experts compare cloud computing with the old mainframe model and fear that data could be held captive by the providers [1]. Additionally, some professionals have voiced their concerns on the current privacy laws strangling the process of addressing some of the cloud computing specific risks [47].

As we have found CSA's research on top threats [13] the most recent among the notable research works in cloud computing security area, and not many reviews are available on it, we have decided to expand on the top seven threats in the following section.

5. Survey on top threats to cloud computing

Securing computer networks and data centers has never been an easy task. Shared on-demand nature of cloud computing makes it an even more challenging job. Selecting an appropriate security procedure requires correct judgment of the threat environment [13]. In this section our main intention is to address the gaps in implementing threat remediation.

5.1. Abuse and nefarious use of cloud computing

CSA mention as some Infrastructure-as-a-Service (IaaS) providers do not maintain enough control, hackers, spammers and other types of people engaged in criminal activities can take advantage of the opportunities such as free limited trials. CSA propose strict registration and identity check procedures, enhanced monitoring for possible credit card frauds, comprehensive introspection of network traffic, and monitoring of public blacklists [13].

In a round table meeting, Microsoft representative John Howie once complained, because of the privacy laws they are not allowed to look at what customers are doing. So, if a malicious individual or organization is performing something nefarious (malware, phishing attacks etc.) using their service, they cannot immediately know, and have to rely on other mechanisms such as notifications and abuse reports [48]. Monfared [49] echoed John's sentiments in his research by saying, "cloud customers may abuse services which they are paying for, hosting a phishing website is an example of it". He pointed out "Communication between different stakeholders play a vital role in mitigating the threat, interest of stakeholders are not necessarily in the same direction. Therefore conflict may happen".

Findings of a few researchers [13,48,49] corroborate the fact that even though CSA propose enhanced monitoring, comprehensive introspection of network traffic, and other actions, present privacy laws are restricting cloud providers to become the first to know if some abuse and nefarious activities are in progress in their cloud.

5.2. Insecure application programming interfaces

As cloud providers provide some kind of software interfaces to a customer to manage and interact with their services, a relatively weak or too much user friendly interfaces may expose different kinds of security issues. Suggested solutions to address the problem are to analyze the security model of API, strong authentication and access control with encrypted transmission, and understanding of the dependency chain [13].

Wrenn [39], Security Solutions Director at Unisys, pointed to the fact that the security control mechanism (authentication and authorization mechanism) may not be adequate to counter bypass attacks and API hacks. This may lead to unauthorized access to even privileged user functions. Attackers can steal session cookies for access to customer systems and data. He also identified two barriers to securing API, which are the inability to audit events associated with API use and incomplete log data to enable reconstruction of management activity. The worst case scenario could be the complete loss of control over the customer cloud infrastructure [39].

Andrei and Jain [50] praised the API use in cloud computing for its centralized model. They said cloud computing helps software developers in creating multiple evaluation environments for their applications, software monitoring can be done by monitoring API calls for server requests. If there is a centralized architecture for data storing, all efforts can be focused in one direction resulting in better monitoring. While discussing vulnerabilities in web application layer in their paper, Grobauer et al. [51] said that API functions share many vulnerabilities with web application layer. As most cloud services are likely to be web services, to access web URL, customers need to use web applications through web browsers which share more vulnerabilities than other avenues of sharing vulnerabilities.

From the works of [50,13,51,39] we can summarize that there are some advantages of monitoring API in cloud computing based centralized system, but web application based API mostly share more vulnerabilities. Though CSA suggested a few remediation measures, there are still some gaps. These gaps are the inability to audit events associated with API use and incomplete log data to enable reconstruction of management activities.

5.3. Malicious insiders

It is usual for a provider to hide its own company policy on recruiting employees and what level of access it provides to them, but with higher level of access an employee can gain access to confidential data and services. CSA suggest enforcing strict supply chain management, specifying human resource requirements as part of Service Level Agreement (SLA), transparency in overall information security and management practices' compliance, reporting, and determining security breach notification processes [13]. Wrenn [39] mentioned if someone gets a job with a cloud service provider with ill intent, it may be much easier for him/her to engage in malicious activities than what people can anticipate. Things can get even worse if this intent conjoins with the cloud provider's inability in monitoring its employees, especially the ones with privileged access. If a cloud provider lacks a breach notification policy and practice, a time may come when a customer may not even be informed of a serious security incident.

Spring [52] suggested ten best practices for cloud providers to handle malicious insiders. These are: separation of privileges, least privilege, access control systems, alarm systems, administrator logging, two-factor authentication, codes of conduct, confidentiality agreements, background checks, and visitor access. Rocha and Correia [53] discussed three solutions, which are: not to allow any physical access to the servers, zero tolerance policy, and logging all accesses to the server. However, they demonstrated in their paper that an incident can happen remotely (without physical access) and the last two solutions come into effect after the incident occurs, which is too late. They also discussed how recent research mechanisms fail to protect the confidentiality of previous attacks on users' data.

From the research works discussed so far [13,53,52,39] we can infer that most suggestions are mainly on monitoring employee activities and formulation of cloud providers' policy such as zero tolerance. Customers can only make sure that they sign up an agreement that contains all of the proposed solutions including transparency of the cloud provider. But some of these solutions will only come into effect after a serious security breach occurs. Unfortunately, in the foreseeable future, it is likely to continue to be a natural tendency of a cloud provider to hide its company policy regarding hiring of employees and put in place insufficient measures to monitor them because of economic reasons.

5.4. *Shared technology vulnerabilities*

Shared on-demand nature of cloud computing needs virtualization, and this virtualization technology uses hypervisors to create virtual machines and operating systems. But flaws in a hypervisor sometimes allow someone to gain inappropriate access and control to the platform that impacts other customers as well. CSA suggest implementing security best practices for installation and configuration, monitoring for unauthorized changes, promote strong authentication, SLA for patching and vulnerability remediation, and vulnerability scanning and configuration audits [13].

Wrenn [39] pointed out that cloud computing was designed for infrastructure sharing in a cost effective model, which inherently lacks basic protection and customer compartmentalization. This class of vulnerability is evident at all levels of the infrastructure stack. Shielding customers' network traffic, data and applications is very difficult because of the hardware limitation. Attackers can hijack privileged user accounts, run other customers VM, and intercept network communication. Yildiz et al. [54] cited the example of mainframes where secure separation is possible but the cost is always unacceptable to the SaaS providers. This is the reason for the introduction of lower cost equipment with emerging virtualization capabilities that can offer business competitors separate virtual machines on the same physical hardware. They also think coexistence of manufacturing and retail sector clients is a problem, as the former's quiet time does not match with retail demand, resulting in issues on applying security patches to shared equipment.

Chow et al. [55] think many adoption problems of cloud computing are essentially old problems in new settings. They also claim that virtual machine attacks and web service vulnerabilities existed long before cloud computing became fashionable. Grobauer et al. [51] expect future virtualization to develop into virtualized servers from computational resources. They are concerned about VMs image handling. As a common practice, cloud providers create a template image of Operating System (OS) and clone it to multiple machines. This is a vulnerable VM template image that may spread over many systems. An attacker can rent one of these VMs and can analyze all the important configurations including administrative rights. Another important issue they raised is that an image can even be taken from an untrustworthy source, which may provide back-door access to an attacker.

From the works of the aforementioned researchers [13,55,51,39,54] we can summarize that there is a hardware limitation of compartmentalization. However, there is an expectation for future virtualization technology—virtual servers to be developed from computational resources.

5.5. *Data loss/leakage*

Cloud customers need to make sure that costs saving methods never compromise their valuable data as there are multifarious ways to compromise data. The ways specially increase in a cloud environment because of the number of interactions between risks and challenges. An example can be deletion or alteration of records without backup and another example can be, not able to restore large context after a disaster. Loss of the encoding key can be very painful too. Some of these may be unique to a cloud system as well as too complex to restore because of its architecture [13]. Proposed solutions by CSA include implementation of strong API access control, encryption and protection of integrity of data in transit, analyses of data protection at both design and run time, implementation of strong key generation, storage and management, and destruction practices—contractually demand providers wiping persistent media before it is released into the pool. Contractually here specifies provider backup and retention strategies [13].

Wrenn [39] sees protection of cloud based data from unauthorized access as top priority in cloud security. He identified two types of risks that cloud providers must adequately address in their cloud implementations. These are data theft and data loss. Reasons behind data loss can be corrupted storage, failure of drives, accidental deletion of partition, providers' lack of adequate backup capability, untested procedures, poor policy, and inadequate data retention practices. Dahbur et al. [56] raised the issue that if users and cloud employees are not educated enough on processes and procedures, they can make intentional or unintentional mistakes that can cause data loss inflicting a devastating impact on a business. Wang et al. [57] mentioned two different sources of data loss and leakage. First, a cloud provider can be self-interested, untrustworthy and possibly malicious and store data in a lower tier of storage than agreed, the provider can also hide a data loss incident due to management errors. Second, there could be someone with capability to alter (modify or delete) cloud data in different time intervals and still remain undetected by a cloud provider for a while.

In making an inference from the already cited research [13,56,57,39] we need to be aware about [45] criticism that this type of threat is too general to any network security issues. Surely, data protection is the top priority in network (not just in cloud) security, but it reaches a much higher level of challenge in cloud computing due to the number of interactions between risks and challenges. While untested procedures, poor policy and inadequate data retention practices are too general, they may be critical in cloud computing because of policy issues, complex infrastructure and customers demand. A cloud employee's lack of knowledge or understanding on cloud related processes and procedures can prove to be very costly. There is also a trust issue with the cloud providers, who may become too commercial and store customers' data in a lower tier of storage than agreed.

5.6. *Account, service and traffic hijacking*

These kinds of attack are usually perpetrated with stolen credentials. There are different attack methods for stealing someone's credentials such as phishing, fraud, Denial of Services (DoS), finding vulnerabilities, and account hijacking. In a cloud, if an attacker can gain access to someone's credentials, he or she can eavesdrop on a customer's activities, transactions, and alter data. Remediation proposed by CSA include prohibition on sharing of account credentials between users and services, leveraging strong two-factor authentication techniques where possible, employing proactive monitoring to detect unauthorized activity, and understanding cloud provider security policies and SLAs [13].

Table 1
Gaps in threat remediation.

Threats	Challenges in implementing threat remediation or gaps
Abuse and nefarious use of cloud computing	• Privacy laws are restricting cloud providers from instant monitoring • Interest of different stakeholders are not necessarily in unison
Insecure application programming interfaces	• The inability to audit events associated with API use • Incomplete log data to enable reconstruction of management activity
Malicious insiders	• Providers are naturally inclined to hide their own company policies for recruiting employees • Solutions come into effect after the incident occurs, which is too late • Cloud providers' inability to monitor its employees
Shared technology vulnerabilities	• Shared elements were never designed for strong compartmentalization • Business competitors using separate virtual machines on the same physical hardware • Coexistence of manufacturing sector and retail sector
Data loss/leakage	• Trust issue with the cloud providers that they may become too commercial and store in low security area than agreed • Untested procedures, poor policy and inadequate data retention practices • Lack of knowledge
Account, service and traffic hijacking	• Rapid development of cloud computing also opens some new loopholes • Present practice of digital identity management is not good enough for hybrid clouds
Unknown risk profile	• Cloud providers' unwillingness to provide log and audit data; and security practices • Lack of transparency

Srinivasamurthy and Liu [43] found four attack types that match these kinds of threat. Those are: man-in-the-middle attacks, phishing, spam campaigns, and DoS attack. Wrenn [39] raised the concern that a company's cloud infrastructure can be targeted as a staging ground for these kinds of attacks, and all these can happen under the identity of the company. He proposed three defensive actions to deal with these types of attack, which are: providing strong cryptographic authentication of systems and users in the cloud, user and system level strong defense against account hijacking, and only authorized systems belonging to a company's interest can access and manage cloud resources for a given customer. Shin and Kobara [58] raised another issue, if the cloud provider also provide Single Sign-On or ID management services, then this type of attack can cause more significant damage. While referring to ID management, Yan et al. [59] emphasized that quick development of cloud computing brought some security problems. Currently, the majority of cloud computing systems use digital identity for their users to access cloud services, this could be a disadvantage for a hybrid cloud. In their research they proposed the use of federated identity management together with hierarchical identity-based cryptography.

From [13,58,43,39,59] we can see that on one hand, there is nothing new, all these types of attacks mentioned above has been encountered before, but on the other hand, rapid development of cloud computing brought in some new problems, such as, the prevalent ways of identity management are not adequate for hybrid clouds.

5.7. Unknown risk profile

One of the major benefits of cloud computing is the reduction of hardware and software needs, which lead to financial savings for a cloud customer as well as relieve them of some complexity to focus more on their actual business. However, the transition into cloud computing may not ensure the efficacy of the security procedures that the company used to maintain by itself, and can entail unknown risks. CSA suggested having disclosure of applicable logs and data; full or partial disclosure of cloud infrastructure; and monitoring and alerting on necessary information [13]. CSA actually raised in this context their concern about customers' questions that are not clearly answered by cloud providers. In another report they mentioned "Unless cloud providers can readily disclose their security controls and the extent to which they are implemented to the consumer, and the consumer knows which controls are needed to maintain the security of their information, there is tremendous potential for misguided decisions and detrimental outcomes" [14].

Wrenn [39] delved into well known and less known features of cloud computing. While well known features are deciding factors for choosing a cloud provider, there are always some less known features such as the details on auditing, logging, security policies, vulnerability and incident response. Without these details, it is essentially an incomplete picture of cloud providers' security practices. He further pointed out that for the unknown risk profile, traditional risk managements are ineffective because it happens without customer awareness. He proposed transparency into the cloud provider's infrastructure management practices and audit data as a solution, an aspect highlighted in recent press reports and technology publications, many cloud providers do not disclose those details to their customers.

We therefore see that an unknown risk profile is somewhat creation of the cloud providers' unwillingness in providing details about security logs, audit report, security practices etc. [13,14,39]. Without these details customers cannot grasp the full extent of the security procedure and may be exposed to unknown risks. There is no easy solution because of the aversion of cloud providers to be transparent in this matter.

Our observations and reviews on the top threats to cloud computing are surmised in Table 1. We have enlisted the challenges, which is the first step in contemplating solution strategies.

6. Attack detection for cloud computing using machine learning techniques

Given the inherent deficiencies of cloud computing such as, remediation only comes into effect after a successful attack happens and cloud providers are unwilling to provide security related data to its customers, we propose a "Proactive Attack Detection" model with three goals. Firstly, it will be able to

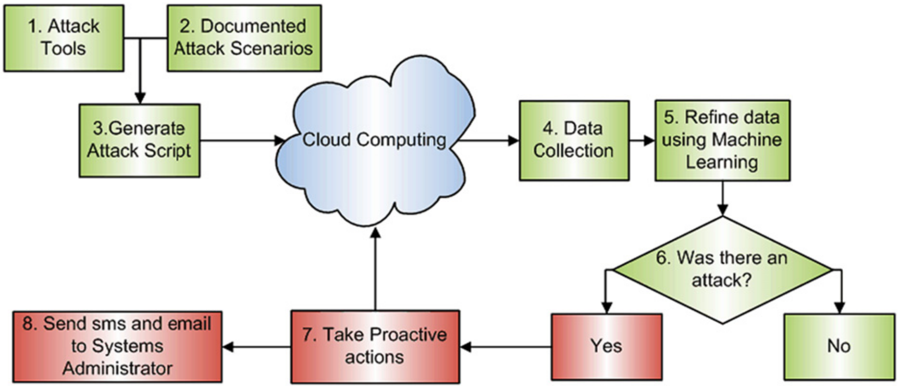

1. Attack Tools
2. Documented Attack Scenarios
3.Generate Attack Script
Cloud Computing
4. Data Collection
5. Refine data using Machine Learning
6. Was there an attack?
Yes
No
7. Take Proactive actions
8. Send sms and email to Systems Administrator

Fig. 5. Attack detection and proactive resolution in single cloud environment.

detect an attack when it starts or at least during the time of its perpetuation. Secondly, it can alert system/security administrators and data owner about the attack type with possible action needed. Thirdly, if cloud providers try to hide attack information from customers, this model will be able to tell customers on the kind of attack that happened by looking at the pattern of attack. Our experience on machine learning techniques suggest that modern machine learning techniques including rule based learning and statistical learning theory are capable of achieving these goals.

6.1. Background

Despite our awareness on threats and our efforts to tackle them, cyber attacks are not vanquished, and we believe this is due mainly to the gaps. In their research Rimal et al. [32] presented eight examples of outages in different cloud services with date and duration. Dahbur et al. [56] presented three other scenarios of cloud computing outage and data loss with the number of customers affected. It is not clear whether these outages were caused by attacks, but nevertheless, outages and data losses are surely basic security concerns and can be put into CSA's data loss/leakage threat category.

Researchers at the University of California, San Diego and the Massachusetts Institute of Technology, Cambridge [60] showed in experiments with Amazon Elastic Compute Cloud [61] that it is possible to map the internal cloud infrastructure and find out the location of a particular virtual machine. They also showed how such findings can be used to mount cross-virtual machine side-channel attacks to collect information from a target virtual machine residing on the same physical machine. In a recent research, Rocha and Correia [53] showed how malicious insiders can steal confidential data. They demonstrated a set of attacks with attack videos, showing how easily an insider can obtain passwords, cryptographic keys and files etc. Chonka et al. [37] recreate some of the recent real world attack scenarios and demonstrate how HTTP-DoS and XML-DoS attack can take place in cloud computing.

We discovered some commonalities among these [37,60,53] attack models in that all of them used attack tools and followed organized attack procedures. We have attempted to design our experimental setup in the same pattern.

6.2. Experiment design

In our experiment, the first step has been to collect attack tools such as Hping, Socket Programming, Httping, Unix shell scripts, side channel attack tools etc. The next step has been on generating attack scripts from the information described in documented attack scenarios in different internet security related websites and blogs such as Dancho Danchev's blog [62] or Jeremiah Grossman's blog [63] and their research works [64–67]. We may not know if some of these attacks happened in cloud computing because of lack of transparency from the cloud providers but it would surely help us from our novelty detection graph. One of the benefits of generating attack scripts is less human effort and these can be programmed to run according to the actual attack timing and duration over multiple virtual machines simultaneously. We have designed our experiment as given in Fig. 5 for a single cloud.

The next step was data collection, the type of data would determine which data collection tools are to be used. The most common type of data collection in an attack scenario could be the number of packets sent and received, processing time, round trip time, CPU usages etc. Machine learning techniques can then be used to investigate if there was an attack. If there is a known type of attack, machine learning can take proactive action to address the issue, and at the same time, notify systems/security administrators. If an unknown type of attack happens, machine learning will still be able to detect it as an attack from the data variations from usual usage, and can notify the designated person with the closest type attack known to its database. It would make the security administrator's job easier to fight against unknown types of attacks.

For data communication between multiple clouds, also known as InterCloud communication, our proposed experimental design is given in Fig. 6. In this scenario, an attacker may attack data sent from one cloud to another. Here machine learning needs to undertake proactive action on both clouds. To achieve this there must be some kind of trust relationship between both the cloud providers. The proactive action on both clouds is called for because if one is infected, it would become an attacker's target for his/her next mission.

To create a virtual cloud environment, we have chosen a HP ProLiant DL380 G4 Server as shown in Fig. 7, with following features: dual Intel Pentium IV Xeon 3.2 GHz Processors, 6 GB RAM, 2x 72.8 GB Hot Plug SCSI Hard Drives, Integrated Smart Array 6i Plus RAID Controller, Dual network interface cards. The main reason for choosing server hardware is for not making hardware limitation a bottleneck, which may provide incorrect data. We also chose VMWare ESXi 3.5 [68] Hypervisor as Virtual Machine Manager (VMM) and Windows 7 as guest Operating System (OS).

We designed this process on the belief that customers need to know all attacks striding on their VM and the physical machine they are co-residing with others. If their business competitors get co-residence on the same physical hardware, or their machine is being cloned without prior notice, there is always a threat. Our main goal for this experiment is to enlighten cloud customers with some basic ideas about how they will be able to detect different attack types with the limited resources and access they have. Fig. 8 shows a screenshot of taking guest VM snapshot, and in Fig. 9 we put Hypervisor performance plots at the time of taking this snapshot.

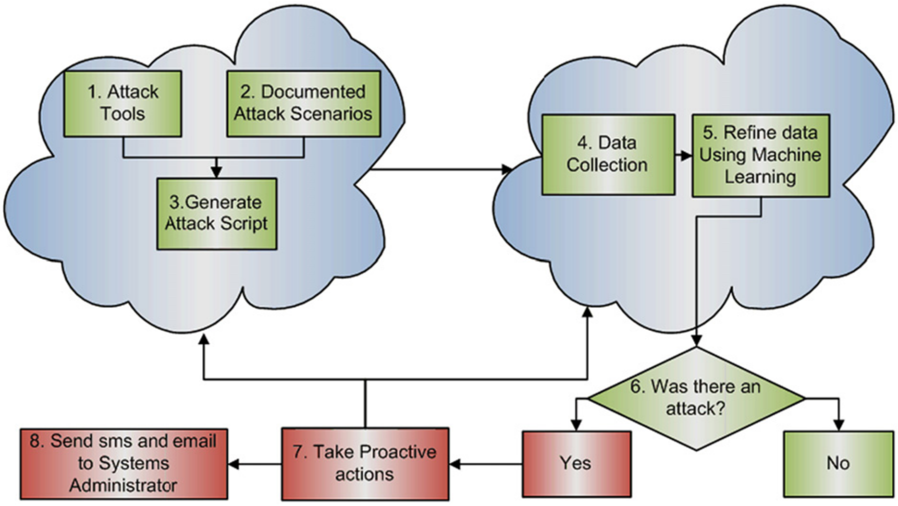

1. Attack Tools
2. Documented Attack Scenarios
3.Generate Attack Script
4. Data Collection
5. Refine data Using Machine Learning
6. Was there an attack?
Yes
No
7. Take Proactive actions
8. Send sms and email to Systems Administrator

Fig. 6. Attack detection and proactive resolution for InterCloud.

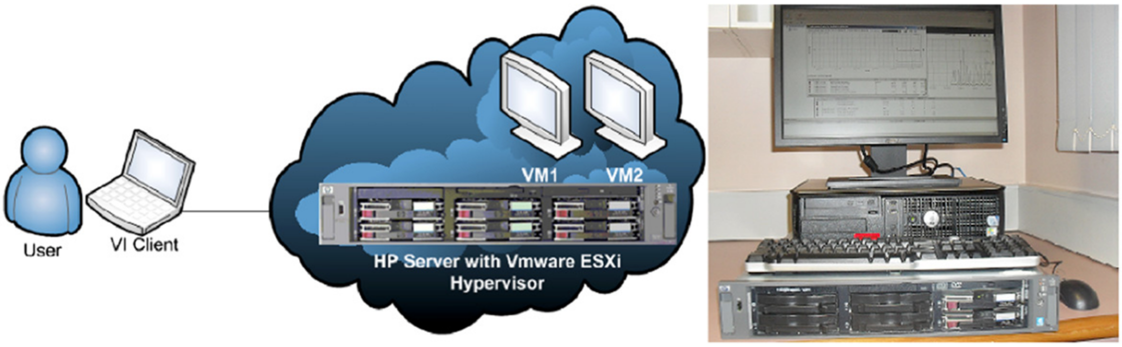
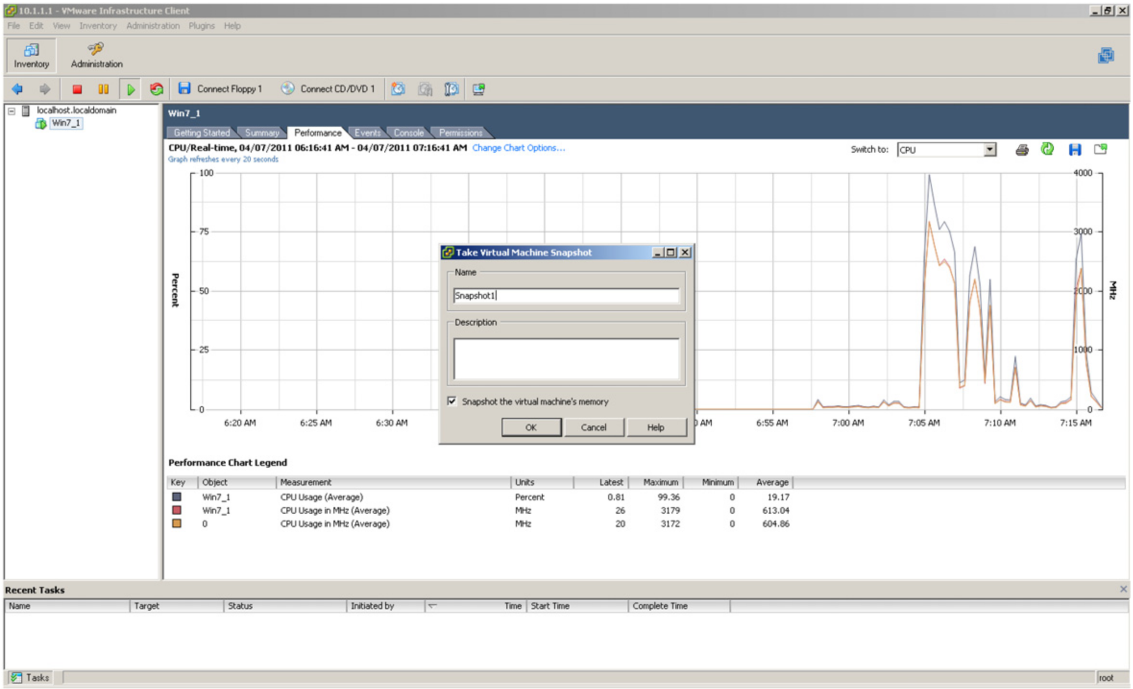


Fig. 7. Logical and physical diagram of our experiment design.



CPU/Real-time, 04/07/2011 06:16:41 AM - 04/07/2011 07:16:41 AM Change Chart Options...
Graph refreshes every 20 seconds

Name
Snapshot1

Description

☑ Snapshot the virtual machine's memory

Performance Chart Legend

Key	Object	Measurement	Units	Latest	Maximum	Minimum	Average
■	Win7_1	CPU Usage (Average)	Percent	0.81	99.36	0	19.17
🟥	Win7_1	CPU Usage in MHz (Average)	MHz	26	3179	0	613.04
🟧	0	CPU Usage in MHz (Average)	MHz	20	3172	0	604.86

Recent Tasks

Name	Target	Status	Initiated by	Time	Start Time	Complete Time

Fig. 8. Taking snapshot of guest VM.

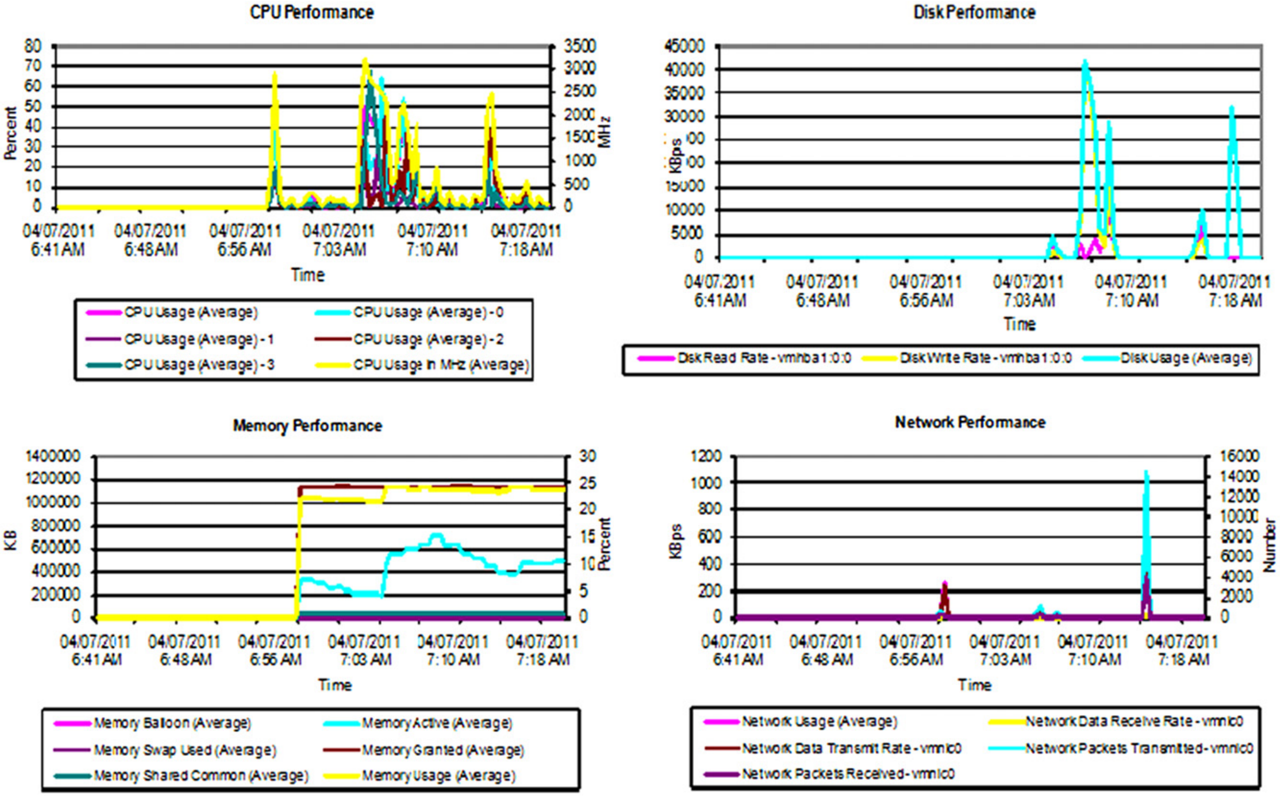


Fig. 9. Performance chart of Hypervisor at the time of taking guest VM snapshot.

6.3. *Data preparation*

To identify the nature of the attack in a cloud environment, we generate an attack dataset for the experimental demonstration by simply gathering performance data of CPU, memory, disk and network usage from hypervisor and guest OS, and choose an appropriate technique for activity classification as shown in Fig. 7 (Logical and physical diagram of our experiment design). The aim is to detect activity pattern and, alert on the type of cyber attack that happened by looking at the change of parameters in the computer and network systems. To start with, we considered five types of most discussed attacks in cloud computing [13,37,69,60,53]. These are as follows:

6.3.1. *Denial of service (DoS) attack*

According to the United States Computer Emergency Readiness Team (US-CERT) DoS attack is a type of attack where an attacker attempts to prevent legitimate users from accessing network or computer resources. Distributed Denial of Services (DDoS) means, the attacker is using multiple computers to launch the denial-of-service attack [70]. CSA raised their concern about this type of attack in their first of seven top threats [13]. Chonka et al. [37] demonstrated how the two types of DoS attacks can take place in cloud computing. However, there are several other types of DoS attacks and attack tools which are worth testing in an experimental cloud environment. McDowell [70] listed few symptoms of DoS and DDoS attacks such as unusually slow network performance, unavailability of a particular website, inability to access any website, and dramatic increase in the amount of spam.

6.3.2. *Cross VM side channel (CVMSC) attack*

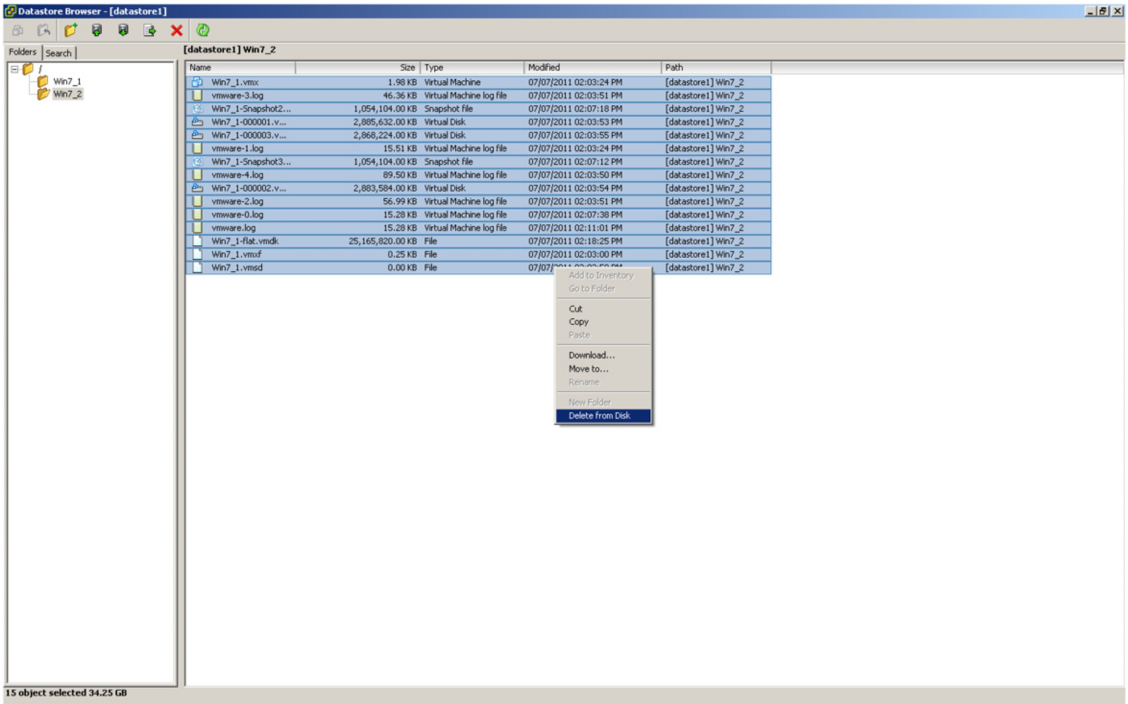
Ristenpart et al. [60] showed how to run this kind of attack in Amazon EC2 to collect information from a target VM where an attacker can reside on a different VM on the same physical hardware. One of the authors of their paper, Professor Stefan Savage from the University of California, mentioned in an interview, "A virtual machine is not proof against all of the kinds of side-channel attacks that we've been hearing about for years" [71]. Some security experts commented about this experiment that, though these attacks developed by the researchers are minor, the techniques could lead to more significant concerns in cloud computing [71].

6.3.3. *Malicious insiders (MI) attack*

This is one of the most widely discussed and most difficult to detect attack types in any network, where an attacker is an insider and therefore bestowed with trust and access. We have discussed about this type of threat in cloud computing in Section 5.3 and mentioned that a cloud provider's lack of transparency makes this threat detection even more complex. Rocha and Correia [53] recently demonstrated execution of this type of attacks using XEN hypervisor. However, we found in our survey that the UNIX commands and procedures they used in their experiments to obtain passwords, cryptographic keys and other confidential data are not generally used by an insider with administrative privileges once a physical machine with customers VMs is in production environment. So, we can definitely monitor those kinds of attempts in order to detect insiders attack.

6.3.4. *Attacks targeting shared memory (ATSM)*

In this type of attack, an attacker takes the advantage of shared memory (physical and cache memory) of a physical/virtual machine. This is an initial level of attack in cloud computing and can lead up to several other types of attacks. For example, while performing CVMSC attack on Amazon EC2, Ristenpart et al. [60] measured cache activity of other users [71]. Rocha and Correia [53] also used information from memory dump while doing MI attack.



[datastore1] Win7_2

Name	Size	Type	Modified	Path
Win7_1.vmx	1.98 KB	Virtual Machine	07/07/2011 02:03:24 PM	[datastore1] Win7_2
vmware-3.log	46.36 KB	Virtual Machine log file	07/07/2011 02:03:51 PM	[datastore1] Win7_2
Win7_1-Snapshot2...	1,054,104.00 KB	Snapshot file	07/07/2011 02:07:18 PM	[datastore1] Win7_2
Win7_1-000001.v...	2,885,632.00 KB	Virtual Disk	07/07/2011 02:03:53 PM	[datastore1] Win7_2
Win7_1-000003.v...	2,868,224.00 KB	Virtual Disk	07/07/2011 02:03:55 PM	[datastore1] Win7_2
vmware-1.log	15.51 KB	Virtual Machine log file	07/07/2011 02:03:24 PM	[datastore1] Win7_2
Win7_1-Snapshot3...	1,054,104.00 KB	Snapshot file	07/07/2011 02:07:12 PM	[datastore1] Win7_2
vmware-4.log	89.50 KB	Virtual Machine log file	07/07/2011 02:03:50 PM	[datastore1] Win7_2
Win7_1-000002.v...	2,883,584.00 KB	Virtual Disk	07/07/2011 02:03:54 PM	[datastore1] Win7_2
vmware-2.log	56.99 KB	Virtual Machine log file	07/07/2011 02:03:51 PM	[datastore1] Win7_2
vmware-0.log	15.28 KB	Virtual Machine log file	07/07/2011 02:07:38 PM	[datastore1] Win7_2
vmware.log	15.28 KB	Virtual Machine log file	07/07/2011 02:11:01 PM	[datastore1] Win7_2
Win7_1-flat.vmdk	25,165,820.00 KB	File	07/07/2011 02:18:25 PM	[datastore1] Win7_2
Win7_1.vmxf	0.25 KB	File	07/07/2011 02:03:00 PM	[datastore1] Win7_2
Win7_1.vmsd	0.00 KB	File	07/07/	[datastore1] Win7_2

15 object selected 34.25 GB

Fig. 10. A screenshot of deleting a VM.

6.3.5. *Phishing attack (PA)*

According to [72] "Phishing is an attempt by an individual or group to solicit personal information from unsuspecting users by employing social engineering techniques". This kind of attack is mainly done by sending links of a website in emails or instant messengers. Such a link looks the same as the original website of a bank or a credit card verification site for example. Resorting to this deception, an attacker can obtain passwords, credit card information etc. [72]. In cloud computing, phishing attacks can be classified into two threat categories. First, as an abusive behavior where an attacker can use the cloud service to host a phishing attack site, and we discussed in Section 5.1 how present privacy laws are preventing cloud providers to detect that instantly. Perpetrators can take advantage of the free trial and instant access. Second, phishing attacks can also be performed to hijack account and services of cloud computing [13].

Sometimes there could be a combination of attacks. Also, for each attack type, different set of parameters of the computer/network system may change, which requires collection of data on what parameters are changing compared to usual/average usage.

Based on the attack symptoms described above about five widely discussed attack types in cloud computing, we select eight attributes to construct the dataset. These are: number of packets transmitted, number of packets received, number of packets lost, disk read rate, disk write rate, memory usage, CPU usage and number of failed administrative log on attempt. In practice, all the data points are considered as real values. The total number of instances in our dataset is 5000. The highest numbers of attacks is 1762 which belongs to class 3 MI attack and the lowest number of attacks is 217 which belong to class 5 PA.

In the set of figures, Figs. 10–14, we present some other screenshots and performance charts of our experiment. Fig. 10 shows a screenshot of deleting a VM; Fig. 11 shows a screenshot of removing a VM from inventory after deletion; Figs. 12 and 13 show screenshots from two steps of cloning a VM and Fig. 14 shows a performance chart of the Hypervisor at the time of installing new VM.

The novelty of our planned series of experiments is to identify an insider's activities and other cyber attacks using performance data. The reason for using performance data rather than traditional logs and security related data is that the performance data can be collected by the customers themselves without any help from cloud providers. To the best of our knowledge no one has tried to detect activities or cyber attacks using these data.

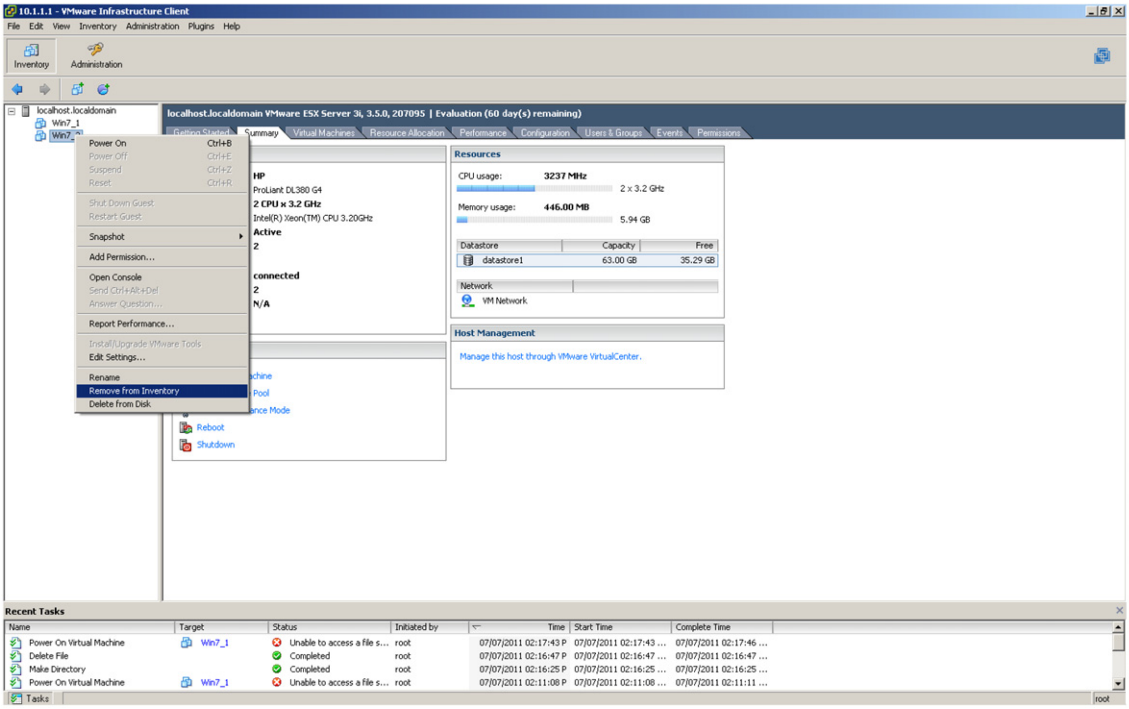
Figs. 15–17 show performance charts of some of the well known cyber attacks. Fig. 15 shows CPU and Disk Performance plots during Ping flood and RDoS attacks, Fig. 16 shows a Network Performance chart of the hypervisor at the time of a TCP SYN Flood Attack, Fig. 17 shows CPU and Network performance plots of the victim at the time of HTTP-DoS Attack.

7. Attack classification

Classification of any attack based on predefined classes of attacks can be solved successfully using machine learning techniques. These techniques are widely available from the data mining community. From the available list we have chosen Naive Bayes [73], Multilayer Perceptron [74], Support Vector Machine [75], Decision Tree [76], and PART [77] to classify into attack type. Naive Bayes is a probability based technique, Multilayer Perceptron and Support Vector Machine (SVM) are function estimation based techniques, Decision Tree and PART are basically rules based data mining techniques.

All these techniques have been implemented in WEKA, which is a Java based popular data mining tool. Weka uses C4.5 James and Barbara [78] algorithm for decision tree implementation. Initially, we carried out some experimental tests to identify the best suited technique for attack classification. The details of performances are provided in Table 2. Performance indicators considered are classification accuracy, number of unclassified instances, and the Area Under Receiver Operating Characteristic (AUROC).

The classification accuracy gives the percentage of attacks which are classified correctly by the data mining techniques. The number of unclassified instances measures the technique's limitations, which means failures in classifying some attacks.



10.1.1.1 - VMware Infrastructure Client

localhost.localdomain
Win7_1
Win7_2

localhost.localdomain VMware ESX Server 3i, 3.5.0, 207095 | Evaluation (60 day(s) remaining)

HP
ProLiant DL380 G4
2 CPU x 3.2 GHz
Intel(R) Xeon(TM) CPU 3.20GHz
Active
2

connected
2
N/A

Resources

CPU usage: 3237 MHz
2 x 3.2 GHz
Memory usage: 446.00 MB
5.94 GB

Datastore	Capacity	Free
datastore1	63.00 GB	35.29 GB

Network
VM Network

Host Management

Manage this host through VMware VirtualCenter.

Recent Tasks

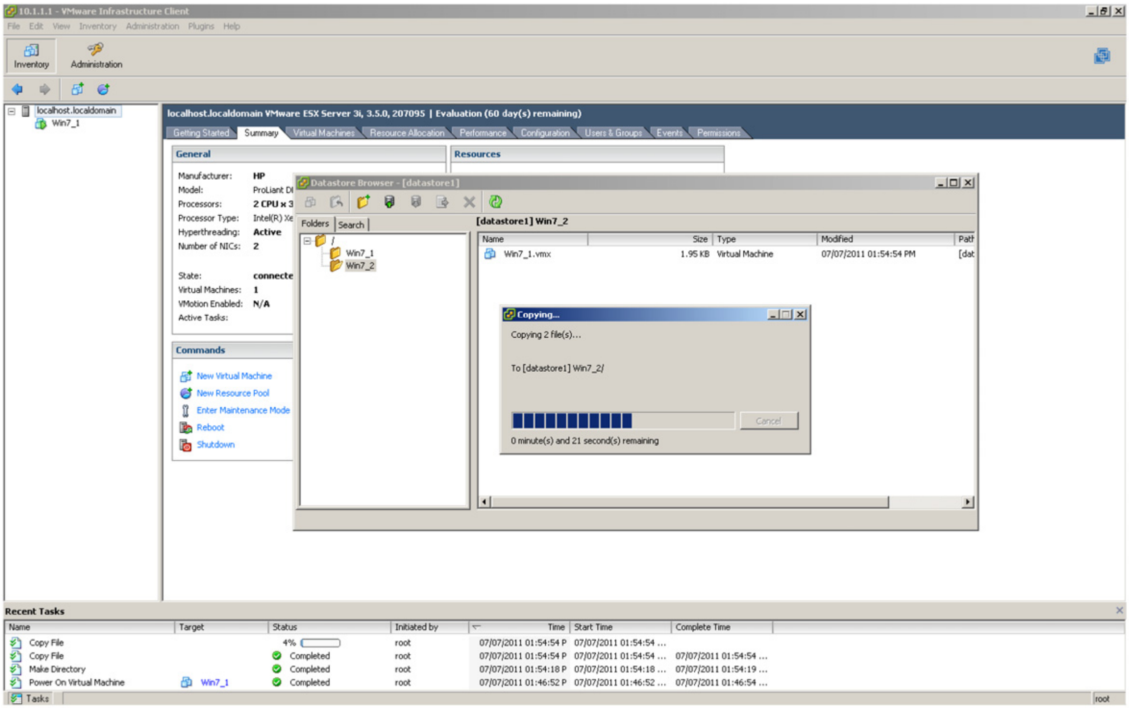
Name	Target	Status	Initiated by	Time	Start Time	Complete Time
Power On Virtual Machine	Win7_1	Unable to access a file s...	root	07/07/2011 02:17:43 P	07/07/2011 02:17:43 ...	07/07/2011 02:17:46 ...
Delete File		Completed	root	07/07/2011 02:16:47 P	07/07/2011 02:16:47 ...	07/07/2011 02:16:47 ...
Make Directory		Completed	root	07/07/2011 02:16:25 P	07/07/2011 02:16:25 ...	07/07/2011 02:16:25 ...
Power On Virtual Machine	Win7_1	Unable to access a file s...	root	07/07/2011 02:11:08 P	07/07/2011 02:11:08 ...	07/07/2011 02:11:11 ...

root

Fig. 11. A screenshot of removing a VM from inventory after deletion.

Table 2
Classification performances of attack data.

	Naive Bayes	Multilayer perceptron	Support vector machine	Decision tree	Part
No. of unclassified instances	0	0	0	0	0
Area under ROC	0.97	1	1	0.90	0.89
Model building time in seconds	0.02	10.90	0.65	0.26	2.76
Model testing time in seconds	0.02	0.01	0.01	0.00	0.02



10.1.1.1 - VMware Infrastructure Client

localhost.localdomain
Win7_1

localhost.localdomain VMware ESX Server 3i, 3.5.0, 207095 | Evaluation (60 day(s) remaining)

General

Manufacturer: HP
Model: ProLiant Dl
Processors: 2 CPU x 3
Processor Type: Intel(R) Xe
Hyperthreading: Active
Number of NICs: 2

State: connecte
Virtual Machines: 1
VMotion Enabled: N/A
Active Tasks:

Resources

Commands

New Virtual Machine
New Resource Pool
Enter Maintenance Mode
Reboot
Shutdown

Datastore Browser - [datastore1]

/
Win7_1
Win7_2

[datastore1] Win7_2

Name	Size	Type	Modified	Path
Win7_1.vmx	1.95 KB	Virtual Machine	07/07/2011 01:54:54 PM	[dat

Copying...

Copying 2 file(s)...

To [datastore1] Win7_2/

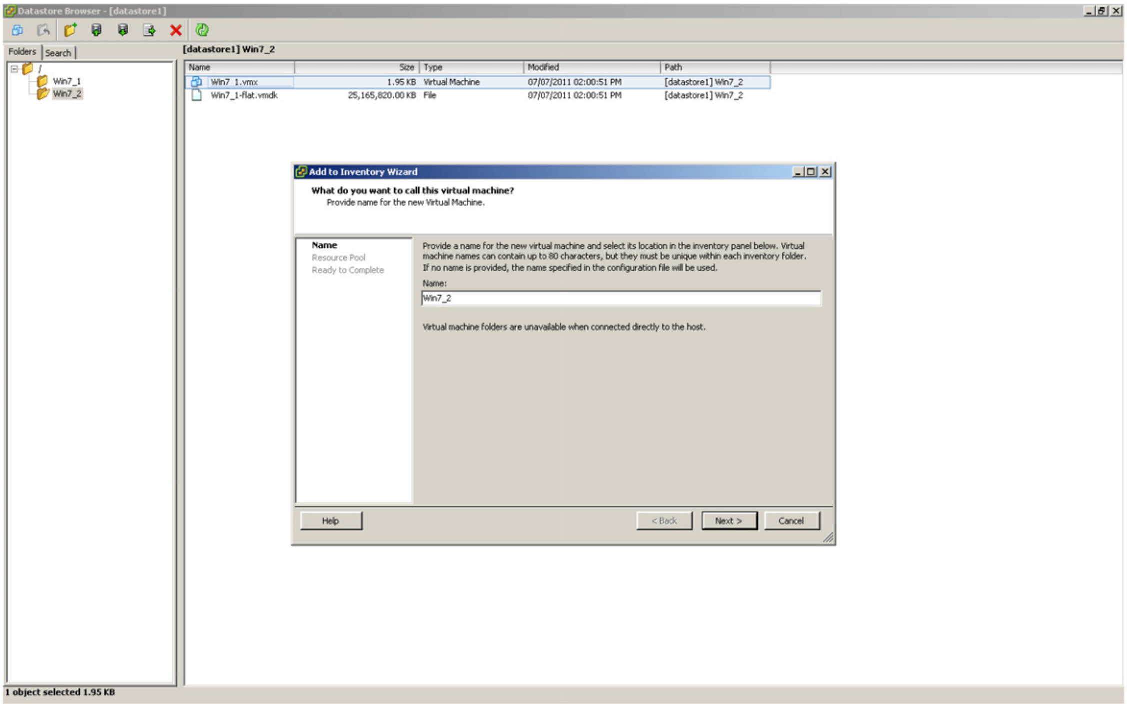
0 minute(s) and 21 second(s) remaining

Recent Tasks

Name	Target	Status	Initiated by	Time	Start Time	Complete Time
Copy File		4%	root	07/07/2011 01:54:54 P	07/07/2011 01:54:54 ...	
Copy File		Completed	root	07/07/2011 01:54:54 P	07/07/2011 01:54:54 ...	07/07/2011 01:54:54 ...
Make Directory		Completed	root	07/07/2011 01:54:18 P	07/07/2011 01:54:18 ...	07/07/2011 01:54:19 ...
Power On Virtual Machine	Win7_1	Completed	root	07/07/2011 01:46:52 P	07/07/2011 01:46:52 ...	07/07/2011 01:46:54 ...

root

Fig. 12. A screenshot of cloning a VM.



[datastore1] Win7_2

Name	Size	Type	Modified	Path
Win7_1.vmx	1.95 KB	Virtual Machine	07/07/2011 02:00:51 PM	[datastore1] Win7_2
Win7_1-flat.vmdk	25,165,820.00 KB	File	07/07/2011 02:00:51 PM	[datastore1] Win7_2

What do you want to call this virtual machine?
Provide name for the new Virtual Machine.

Name
Resource Pool
Ready to Complete

Provide a name for the new virtual machine and select its location in the inventory panel below. Virtual machine names can contain up to 80 characters, but they must be unique within each inventory folder. If no name is provided, the name specified in the configuration file will be used.

Name: Win7_2

Virtual machine folders are unavailable when connected directly to the host.

1 object selected 1.95 KB

Fig. 13. A screenshot of the last step of cloning.

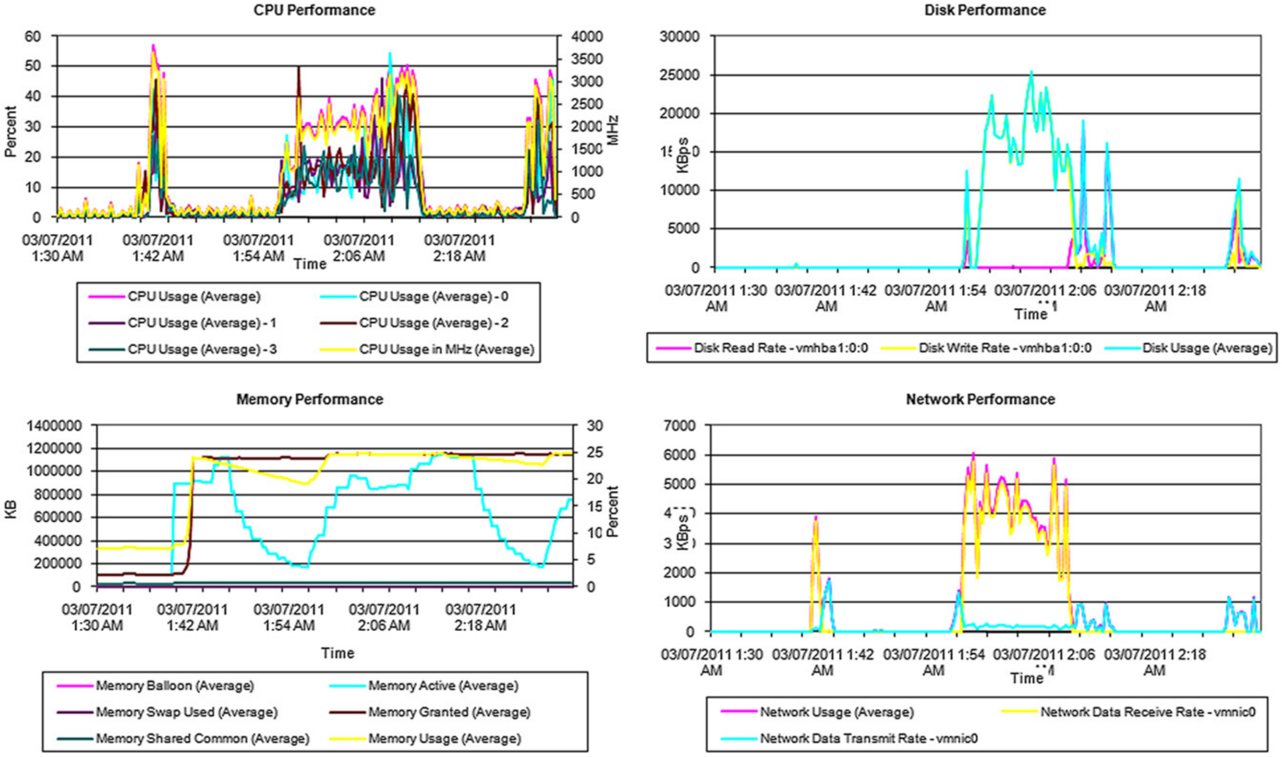


Fig. 14. Performance chart of the hypervisor at the time of installing new VM.

AUROC is a two dimensional popular method of a classification technique's performance measure. In its simplest form, it is a parametric plot of the true attack versus the false attack rate, as a decision threshold is varied across the full range of a continuous classification quantity. It is often taken as a scalar measure [79]. An AUROC of 0.5 reflects random classification, while AUROC $= 1$ implies perfect classification. We also need to consider the computational efficiency of the algorithms and how well they learn since we are dealing with comparatively large datasets. Therefore, we measured the model building and testing times, which are listed in Table 2 along with the number of unclassified instances and AUROC. The attack classification accuracies in percent are shown in Fig. 18.

Based on the classification accuracy, number of unclassified instances and AUROC we found multilayer perceptron and support vector machine are better choices for attack classification in

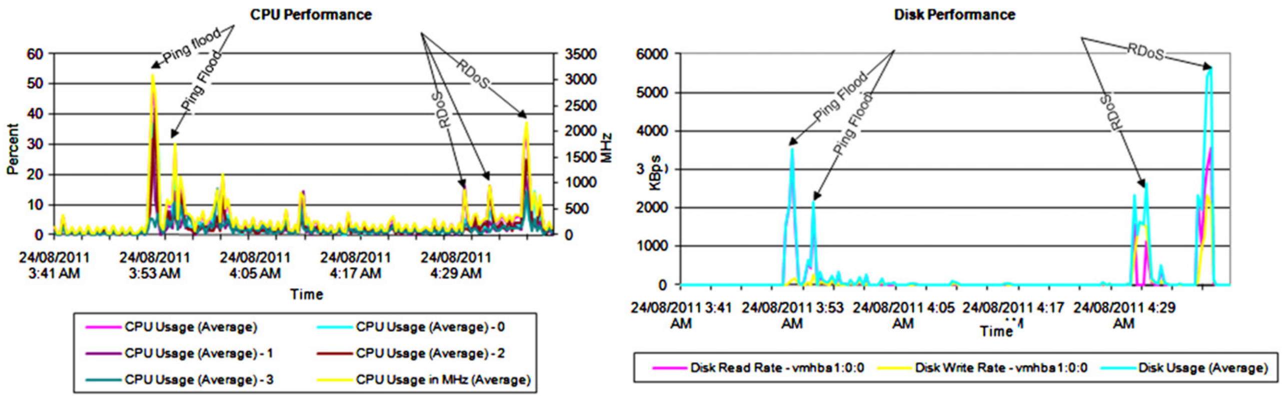
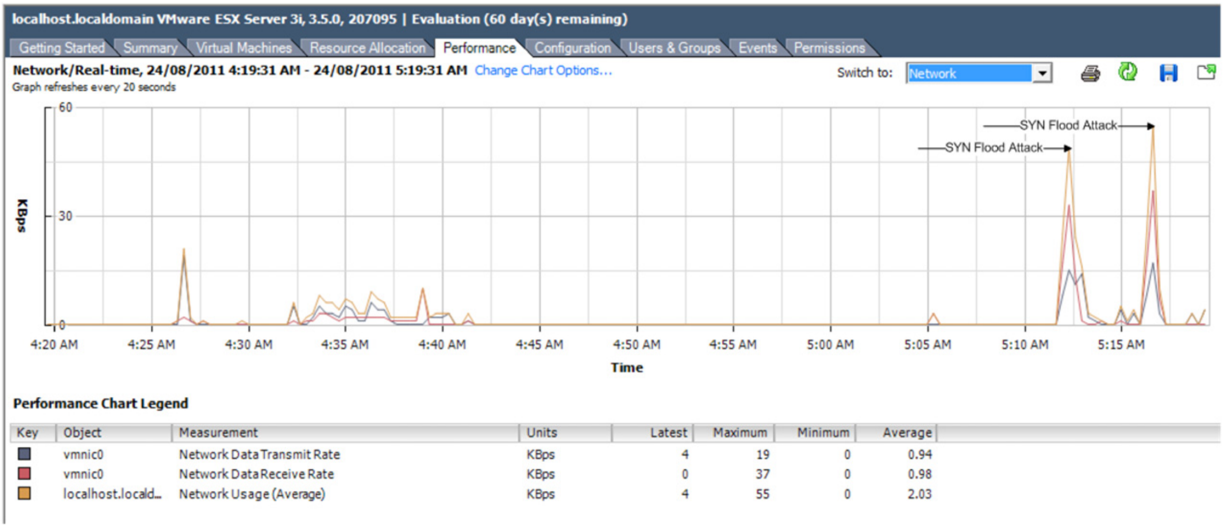


Fig. 15. CPU and Disk Performance plot during Ping flood and RDoS attacks.


localhost.localdomain VMware ESX Server 3i, 3.5.0, 207095 | Evaluation (60 day(s) remaining)

Network/Real-time, 24/08/2011 4:19:31 AM - 24/08/2011 5:19:31 AM

Graph refreshes every 20 seconds

Switch to: Network

Performance Chart Legend

Key	Object	Measurement	Units	Latest	Maximum	Minimum	Average
■	vmnic0	Network Data Transmit Rate	KBps	4	19	0	0.94
■	vmnic0	Network Data Receive Rate	KBps	0	37	0	0.98
■	localhost.locald...	Network Usage (Average)	KBps	4	55	0	2.03

Fig. 16. Network performance chart of the hypervisor at the time of TCP SYN Flood Attack.

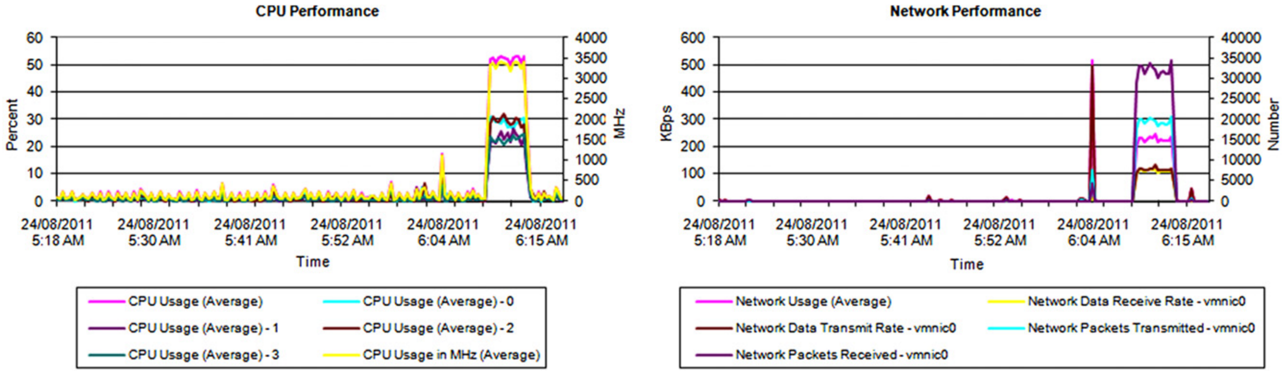


Fig. 17. CPU and Network performance plot of the victim at the time of HTTP-DoS Attack.

the cloud computing area. Further comparison among these two techniques demonstrated that SVM is the best choice for attack classification. However, in terms of computational complexity we found both techniques have spent more time to build their models compared with other techniques. Multilayer perceptron is an extremely slow technique for our task. But the computational complexity of SVM is relatively close to other techniques. Therefore, we choose SVM as the final selection for our task.

The classification accuracy, number of unclassified instances and AUROC basically captured the average performances of the techniques for our problem. But we made further attempts to take a closer look at the performances on attack classification. In that vein, we employed Confusion Matrix [80] analysis to study the details of the techniques' performance measures. This matrix offers a detailed picture on the actual and predicted classification task done by any classification technique classwise. The confusion matrix based performances are provided in Tables 3–7.

We found naive Bayes classified 56.80% of DoS, 73.00% of CVMSC, 73.60% of MI, 67.80% of ATSM and 35.50% of PA attacks successfully. CVMSC and MI were comfortable tasks for naive Bayes. However, PA attacks appear to be a difficult task for this technique. Moreover, DoS attack performance is not high enough.

By contrast, multilayer perceptron has classified 98.40% of DoS, 99.30% of CVMSC, 99.20% of MI, 97.60% of ATSM and 0% of PA

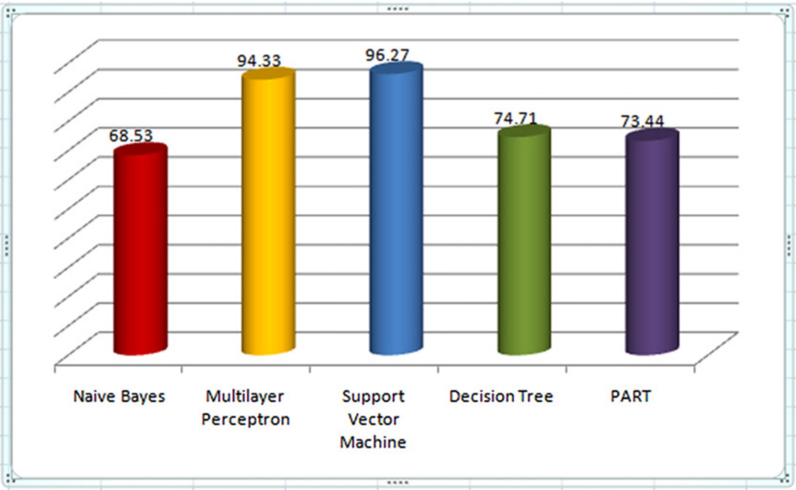


Fig. 18. Attack classification accuracy.

Table 3
Confusion matrix based performance for naive Bayes algorithm.

	DoS	CVMSC	MI	ATSM	PA	Classified
DoS	104	75	4	0	0	DoS
CVMSC	6	325	114	0	0	CVMSC
MI	0	81	377	54	0	MI
ATSM	0	3	90	198	1	ATSM
PA	0	0	4	40	24	PA

Table 4
Confusion matrix based performance of multilayer perceptron.

	DoS	CVMSC	MI	ATSM	PA	Classified
DoS	180	3	0	0	0	DoS
CVMSC	1	442	2	0	0	CVMSC
MI	0	2	508	2	0	MI
ATSM	0	0	7	285	0	ATSM
PA	0	0	0	68	0	PA

Table 5
Confusion matrix based performance of support vector machine.

	DoS	CVMSC	MI	ATSM	PA	Classified
DoS	171	12	0	0	0	DoS
CVMSC	0	435	10	0	0	CVMSC
MI	0	5	506	1	0	MI
ATSM	0	0	13	277	2	ATSM
PA	0	0	0	13	55	PA

Table 6
Confusion matrix based performance of decision tree C4.5.

	DoS	CVMSC	MI	ATSM	PA	Classified
DoS	137	46	0	0	0	DoS
CVMSC	32	339	73	1	0	CVMSC
MI	0	66	400	46	0	MI
ATSM	0	0	52	221	19	ATSM
PA	0	0	0	21	47	PA

Table 7
Confusion matrix based performance of PART.

	DoS	CVMSC	MI	ATSM	PA	Classified
DoS	130	52	1	0	0	DoS
CVMSC	38	333	72	2	0	CVMSC
MI	1	75	382	54	0	MI
ATSM	0	0	64	218	10	ATSM
PA	0	0	0	21	47	PA

attacks successfully. DoS, CVMSC, MI, ATSM were very comfortable tasks for multilayer perceptron. But, it failed miserably to classify any attack of PA. All PA attacks were classified as ATSM attacks.

Support vector machine appeared to have high level of performance across all classes. It classified 93.40% of DoS, 97.80% of CVMSC, 98.80% of MI, 94.90% of ATSM and 80.9% of PA attacks successfully. SVM classification performance for all categories – DoS, CVMSC, MI, and ATSM – is very similar in level and acceptable. Only the classification rate of ATSM was relatively lower than others.

Decision tree classified 74.90% of DoS, 76.20% of CVMSC, 78.10% of MI, 75.70% of ATSM, and 69.10% of PA attacks correctly. DoS, CVMSC, MI, ATSM classification performances were all at a similar level for the decision tree. PA attack performance is slightly lower for this technique, but still it is better than naive Bayes or multilayer perceptron.

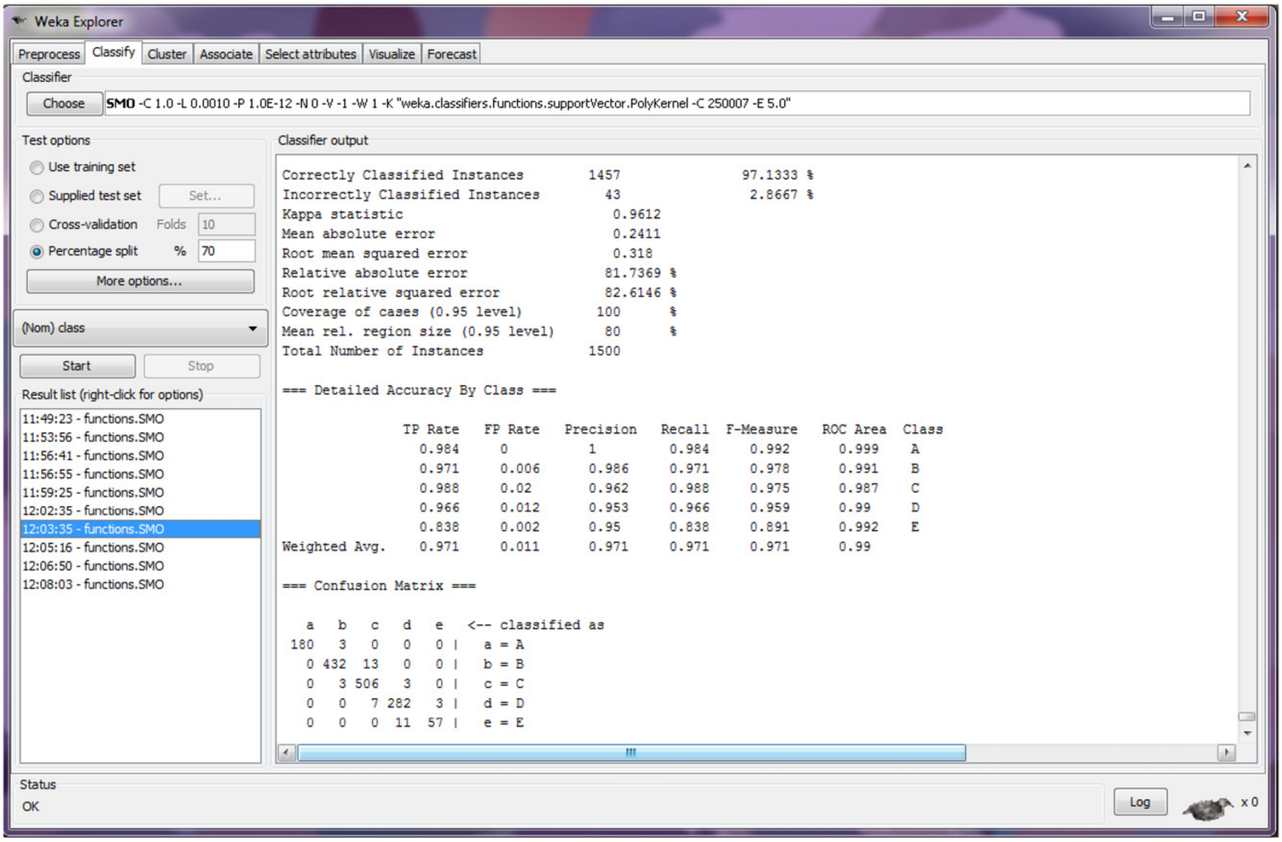
PART classified 71.00% of DoS, 74.80% of CVMSC, 74.60% of MI, 74.70% of ATSM, and 69.10% of PA attacks successfully. DoS, CVMSC, MI, ATSM, and PA classification performances were all at similar level for PART, which we have seen earlier with the decision tree C4.5. DoS, CVMSC, MI, ATSM attacks classification performances of PART are slightly lower than the decision tree but the PA attack classification performance is similar to the decision tree. These results of the confusion matrices corroborate the fact that SVM is an efficient classifier to classify the attack types in the cloud environment.

SVM is a statistical learning theory based data mining technique, which was first introduced for data classification only. With time it has been expanded with full functionality in the areas of regression analysis and clustering tasks. Even though it is considered that multilayer perceptron (artificial neural network) is easier to use than this, experimentally it has been found that SVM is more efficient in many cases than neural networks in terms of accuracy and even computational complexity [81,82].

Because we have finally adopted SVM, we provide here a brief description of SVM. Let us consider a training sample: $D_l = \{x_i, y_i\}_{i=l}^{l}$, x_i is the ith input vector, $x_i \in R^n$, $y_i \in [+1, -1]$, l is the total number of input vectors and n is the dimension of the input space. Suppose the relation between x and y is $y = \mathrm{sgn}(f(x) + \varepsilon)$, where $\mathrm{sgn}(x) = l$, if $x \geq 0$ and $\mathrm{sgn}(x) = -1$, if $x < 0$, the task uncovering function f is called classification.

SVM basically minimizes a tradeoff between empirical error and complexity of hypothesis space in the training phase. Formally, this is done by solving the following minimization problem:

$$\min_f \|f\|_K^2 + C\sum_{i=1}^{l}|1 - y_i f(\mathbf{x}_i)|_+ \qquad (1)$$



Classifier

SMO -C 1.0 -L 0.0010 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -C 250007 -E 5.0"

Test options

○ Use training set

○ Supplied test set

○ Cross-validation Folds 10

● Percentage split % 70

(Nom) class

Result list (right-click for options)

```
11:49:23 - functions.SMO
11:53:56 - functions.SMO
11:56:41 - functions.SMO
11:56:55 - functions.SMO
11:59:25 - functions.SMO
12:02:35 - functions.SMO
12:03:35 - functions.SMO
12:05:16 - functions.SMO
12:06:50 - functions.SMO
12:08:03 - functions.SMO
```

Classifier output

```
Correctly Classified Instances        1457               97.1333 %
Incorrectly Classified Instances        43                2.8667 %
Kappa statistic                          0.9612
Mean absolute error                      0.2411
Root mean squared error                  0.318
Relative absolute error                 81.7369 %
Root relative squared error             82.6146 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      80      %
Total Number of Instances             1500

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.984     0          1         0.984     0.992      0.999    A
                 0.971     0.006      0.986     0.971     0.978      0.991    B
                 0.988     0.02       0.962     0.988     0.975      0.987    C
                 0.966     0.012      0.953     0.966     0.959      0.99     D
                 0.838     0.002      0.95      0.838     0.891      0.992    E
Weighted Avg.    0.971     0.011      0.971     0.971     0.971      0.99

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 180   3   0   0   0 |   a = A
   0 432  13   0   0 |   b = B
   0   3 506   3   0 |   c = C
   0   0   7 282   3 |   d = D
   0   0   0  11  57 |   e = E
```

Status

OK

Fig. 19. A screen shot of SVM with polynomial kernel performance.

where C is a so called "regularization parameter" that controls the tradeoff between empirical error and complexity of the hypothesis space used.

We present the quadratic programming formulation for SVM classification, and we refer the interested reader to [83].

Eq. (1) can be rewritten as follows:

SV classification:

$$\min_{f,\xi_i} \; \|f\|_K^2 + C\sum_{i=1}^{l}\xi_i \tag{2}$$

subject to:$y_i f(\mathbf{x}_i) \geq 1 - \xi_i, \quad$ for all i

$\xi_i \geq 0.$

Variables ξ_i are called slack variables and they measure the error made at point $(\mathbf{x}_i, y_i)$.

A sequential optimization method was proposed by Platt initially to solve the above problem [75].

SVM classification, dual formulation:

$$\min_{\alpha_i} \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

subject to:$0 \leq \alpha_i \leq C, \quad$ for all i

$$\sum_{i=1}^{l}\alpha_i y_i = 0$$

while $K(\mathbf{x}_i, \mathbf{x}_j)$ is called the kernel function. This is the most important ingredient of SVM, which basically transforms the data from a nonlinear space into a linear or near linear space. As a result the learning space becomes more tractable. Some common kernels are shown in Table 8. In our study we have experimented with polynomial and radial basis function (rbf) kernels.

Table 8
Common kernel functions for SVM.

Kernels	Functions
Linear	$x \cdot y$
Polynomial	$[(x^* x_i) + 1]^d$
RBF	$\exp\left\{-\gamma \, \lvert x - x_i \rvert^2\right\}$

In Table 8 we can observe that d and γ are the polynomial and rbf kernel parameters, respectively. Both kernels' performances are highly reliant on the tuning of these parameters.

In WEKA the default kernel for SVM is the 1st degree polynomial. We tested the polynomial kernel (in a normalized form) and rbf kernel for our attack classification problem. The performances are reported in Table 9. We found polynomial kernel is the best choice for our problem.

For the final selection of SVM with polynomial kernel for the attack classification, the value of d was varied from 1 to 5. This is the normal practice in SVM applications to keep the polynomial degree range within these limits [81]. The performance of the polynomial kernel with different degrees are summarized in Table 10.

A screenshot of WEKA during the calibration process is given in Fig. 19. The series of experiments in the screenshot demonstrated that the performance of the polynomial kernel deteriorates with increasing d values. Therefore, we deduced that SVM with polynomial degree 2 performs the best for attack classification in a cloud.

8. Conclusions

Despite the great potential that cloud computing holds, we are witnessing a lack of enthusiasm among the consumers; a phenomenon we believe is largely attributable to data security

Table 9
SVM kernel performances.

% accuracy | time in seconds	Types of SVM Kernels		
	Polynomial	Normalized polynomial	rbf
	96.27 | 1.45	84.46 | 55.95	66.6 | 96.67

Table 10
Best polynomial degree performance.

Degree	Polynomial kernel				
	1	2	3	4	5
% accuracy | time in seconds	96.27 | 1.45	97.13 | 10.03	97.06 | 11.01	96.87 | 10.65	96.8 | 10.03

concerns. Organizations desire assurances that their valuable data are not compromised. Because of many benefits that cloud computing can offer, it is of critical importance that the gaps in security measures be identified and addressed. Unfortunately, cloud services do pose as an attractive target to any cyber criminal because it is a one-stop shop to perpetrate all kinds of criminal activities since these sites contain many user and organizational data. To address the problem, lessons learned from the past on internet are always beneficial.

This research focused on an extensive search on gaps, identify prevalent types of attacks, and seek solutions for the cloud environment. We identified five common types of attacks, which are Denial of service attack, Cross virtual machine side-channel attack, Malicious insiders attack, Attacks targeting shared memory, and Phishing attack. These are the top threats for the real world cloud implementation. To develop a procedure for the automatic identification of these attacks we generate a database from our experience by including number of packets sent, number of packets received, number of packets lost, number of open ports, difference in VM file size, network usage, CPU usage, and number of failed administrative log-on attempts. We set up an actual cloud environment and performed cyber attacks on it to simulate the real world attack scenarios. With the data generated, machine learning techniques were employed for detecting top and known attack types as well as some unknown attacks that follow the same pattern.

We have presented the performance of SVM technique using different kernels on our attack dataset and compared with other conventional machine learning techniques. Through the process, we not only established that SVM is the best choice but also found that polynomial and rbf kernels are most suitable for the purpose. We evaluated polynomial kernel for different values of degree and discovered that second degree is the most appropriate.

However, our experimental outcomes are by no means conclusive because of the limitations on the depth and volume of trials. As a future task when more data become available we intend to focus on optimizing the naive Bayes, multilayer perceptron, decision tree C4.5, and PART techniques by adopting their parameters for our attack classification problem. We also hope to be able to collect real world cloud environment data and test how many attack traffic we can identify within a short period of time. It may so happen that we would discover different methods are best suited for different platforms. Indeed, it would be an exciting experience to be able to travel through the real world experimental environment.

References

[1] B. Schneier, M. Ranum, 2009, Face-off: assessing cloud computing risks retrieved 9 MAY 2011, from http://searchcloudsecurity.techtarget.com/video/Face-off-Assessing-cloud-computing-risks.

[2] M.M. Boroujerdi, S. Nazem, Cloud computing: changing cogitation about computing, World Academy of Science, Engineering and Technology (2009) 58.

[3] R. Buyya, C.S. Yeo, S. Venugopal, Market-oriented cloud computing: vision, hype, and reality for delivering it services as computing utilities, 2008.

[4] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems 25 (6) (2009) 599–616.

[5] D. Catteddu, G. Hogben, Benefits, risks and recommendations for information security, European Network and Information Security Agency (ENISA) (2009).

[6] D.S. Linthicum, Cloud computing and SOA convergence in your enterprise: a step-by-step guide: Addison-Wesley professional, 2009.

[7] P. Mell, T. Grance, The NIST definition of cloud computing, National Institute of Standards and Technology 53 (6) (2009).

[8] J. Heiser, What you need to know about cloud computing security and compliance, Gartner, Research, ID, 2009.

[9] S. Subashini, V. Kavitha, A survey on security issues in service delivery models of cloud computing, Journal of Network and Computer Applications (2010).

[10] S. Covert, Press release retrieved 20 May 2011, 2009 from http://cloud-standards.org/wiki/index.php?title=Press_Release.

[11] M. Rutkowski, A. Sill, M. Edwards, L. Vreck, C. harding, P. Lipton, et al., 2011, Cloud standards wiki retrieved 20 May 2011, from http://cloud-standards.org/wiki/index.php?title=Main_Page.

[12] J. Archer, A. Boehm, Security guidance for critical areas of focus in cloud computing, Cloud Security Alliance (2009).

[13] J. Archer, A. Boehme, D. Cullinane, P. Kurtz, N. Puhlmann, J. Reavis, 2010, Top threats to cloud computing, version 1.0. cloud security alliance retrieved 7 May 2011, from http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf.

[14] G. Brunette, R. Mogull, Security guidance for critical areas of focus in cloud computing V2. 1. 2009 CSA (cloud security alliance), USA, Disponible en: http://www.cloudsecurityalliance.org/guidance/csaguide, v2.1, 1.

[15] CloudAudit, 2010, CloudAudit (codename: A6) retrieved 8 May 2011, from http://www.cloudaudit.org/index.html.

[16] DMTF, 2011, Cloud management retrieved 21 May 2011, from http://www.dmtf.org/standards/cloud.

[17] ETSI, 2011, Grid and cloud computing retrieved 21 May 2011, from http://www.etsi.org/WebSite/Technologies/GRID_CLOUD.aspx.

[18] NIST, 2011, NIST cloud computing program retrieved 21 May 2011, from http://www.nist.gov/itl/cloud/.

[19] OCCI, 2011, Open cloud computing interface retrieved 21 May 2011, from http://occi-wg.org/.

[20] OCCI, 2011, Open cloud consortium retrieved 21 May 2011, from http://opencloudconsortium.org/working-groups/.

[21] SNIA 2011, Cloud storage initiative retrieved 21 May 2011, from http://www.snia.org/forums/csi.

[22] O. Arasatnam, S. Boardman, 2010, Security for the cloud and SOA retrieved 8 May 2011, from http://www.opengroup.org/soa/projects/security.htm.

[23] O.C. Manifesto, Open cloud manifesto, 2009, Available online: www.opencloudmanifesto.org/Open, 20.

[24] I. Foster, Y. Zhao, I. Raicu, S. Lu, Cloud computing and grid computing 360° compared, 2008.

[25] T. Ferrari, L. Gaido, Resources and services of the EGEE production infrastructure, Journal of Grid Computing (2011) 1–15.

[26] A. Balaž, O. Prnjat, D. Vudragović, V. Slavnić, I. Liabotis, E. Atanassov, et al., Development of grid e-Infrastructure in South–Eastern Europe, Journal of Grid Computing (2011) 1–20.

[27] F. Brasileiro, M. Gaudencio, R. Silva, A. Duarte, D. Carvalho, D. Scardaci, et al., Using a simple prioritisation mechanism to effectively interoperate service and opportunistic grids in the EELA-2 e-Infrastructure, Journal of Grid Computing 9 (2) (2011) 241–257.

[28] K. Jeffery, H. Schubert, B. Neidecker-Lutz, The future of cloud computing opportunities for European cloud computing beyond 2010, Expert Group Report, public version, 1 2010.

[29] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, et al., Above the clouds: a berkeley view of cloud computing, EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28, 2009.

[30] M.T. Khorshed, A.B.M.S. Ali, S.A. Wasimi, Monitoring insiders activities in cloud computing using rule based learning, Paper presented at the Proceedings of IEEE TrustCom-11, Nov. 16–18, Changsha, China, 2011.

[31] M.T. Khorshed, A.B.M.S. Ali, S.A. Wasimi, Trust issues that create threats for cyber attacks in cloud computing, Paper presented at the Proceedings of IEEE ICPADS, December 7–9, Tainan, Taiwan, 2011.

[32] B.P. Rimal, E. Choi, I. Lumb, A taxonomy and survey of cloud computing systems, 2009.

[33] G. Ness, 2009, 3 Major barriers to cloud computing retrieved 22 May 2011, from http://www.infra20.com/post.cfm/3-major-barriers-to-cloud-computing.

[34] N. Leavitt, Is cloud computing really ready for prime time? Growth 27 (2009) 5.

[35] J. Brodkin, Gartner: seven cloud-computing security risks, Infoworld (2008) 1–3.

[36] R. Mikkilineni, V. Sarathy, Cloud computing and the lessons from the past, 2009.

[37] A. Chonka, Y. Xiang, W. Zhou, A. Bonti, Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks, Journal of Network and Computer Applications (2010).

[38] C. Everett, Cloud computing-a question of trust, Computer Fraud & Security (6) (2009) 5–7.

[39] G. Wrenn, 2010, Unisys secure cloud addressing the top threats of cloud computing retrieved 25 May 2011, from http://www.unisys.com/unisys/common/download.jsp?d_id=1120000970002010125\&backurl=/unisys/ri/wp/detail.jsp\&id=1120000970002010125.

[40] R. Choubey, R. Dubey, J. Bhattacharjee, A survey on cloud computing security, challenges and threats, International Journal on Computer Science and Engineering (2011) 3.

[41] M. Pastaki Rad, A. Sajedi Badashian, G. Meydanipour, M. Ashurzad Delcheh, M. Alipour, H. Afzali, A survey of cloud platforms and their future, Computational Science and Its Applications—ICCSA 2009 (2009) 788–796.

[42] V. Chang, G. Wills, D. De Roure, A review of cloud business models and sustainability, 2010.

[43] S. Srinivasamurthy, D. Liu, Survey on cloud computing security, 2010.

[44] J. Heiser, M. Nicolett, Assessing the security risks of cloud computing, Gartner Report, 2009.

[45] D. Lacey, 2010, Top threats to cloud computing? Retrieved 21 May 2011, from http://www.computerweekly.com/blogs/david_lacey/2010/03/top_threats_to_cloud_computing.html.

[46] K. Choo, Cloud computing: challenges and future directions, Trends and Issues in Crime and Criminal Justice (2010).

[47] D. Svantesson, R. Clarke, Privacy and consumer risks in cloud computing, Computer Law & Security Review 26 (4) (2010) 391–397.

[48] E. Grosse, J. Howie, J. Ransome, J. Reavis, S. Schmidt, Cloud computing roundtable, Security & Privacy, IEEE 8 (6) (2010) 17–23.

[49] A.T. Monfared, Monitoring intrusions and security breaches in highly distributed cloud environments, 2010.

[50] T. Andrei, R. Jain, Cloud computing challenges and related security issues, A survey paper, 2009, DOI: http://www.cse.wustl.edu/~jain/cse571-09/ftp/cloud.pdf.

[51] B. Grobauer, T. Walloschek, E. Stöcker, Understanding cloud-computing vulnerabilities, IEEE Security and Privacy (2010).

[52] J. Spring, Monitoring cloud computing by layer, part 1, Security & Privacy, IEEE 9 (2) (2011) 66–68.

[53] F. Rocha, M. Correia, Lucy in the sky without diamonds: stealing confidential data in the cloud, 2011.

[54] M. Yildiz, J. Abawajy, T. Ercan, A. Bernoth, A layered security approach for cloud computing infrastructure, 2009.

[55] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, et al., Controlling data in the cloud: outsourcing computation without outsourcing control, 2009.

[56] K. Dahbur, B. Mohammad, A.B. Tarakji, A survey of risks, threats and vulnerabilities in cloud computing, 2011.

[57] C. Wang, Q. Wang, K. Ren, W. Lou, Ensuring data storage security in cloud computing, 2009.

[58] S.H. Shin, K. Kobara, Towards secure cloud storage, Demo for CloudCom2010, 2010.

[59] L. Yan, C. Rong, G. Zhao, Strengthen cloud computing security with federal identity management using hierarchical identity-based cryptography, Cloud Computing (2009) 167–177.

[60] T. Ristenpart, E. Tromer, H. Shacham, S. Savage, Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds, 2009.

[61] Amazon, 2011, Amazon elastic compute cloud (Amazon EC2) retrieved 27 June 2011, from http://aws.amazon.com/ec2/.

[62] D. Danchev, 2011, Dancho Danchev's blog—mind streams of information security knowledge retrieved 31 May 2011, from http://ddanchev.blogspot.com/.

[63] J. Grossman, 2011, Jeremiah Grossman retrieved 19 June 2011, from http://jeremiahgrossman.blogspot.com/.

[64] D. Danchev, Coordinated Russia vs Georgia cyber attack in progress, Retrieved October, 25, 2008.

[65] D. Danchev, The DDoS attack against CNN. com, 2008.

[66] J. Grossman, Cross-site scripting worms and viruses, Whitehat Security (2006).

[67] J. Grossman, T. Niedzialkowski, Hacking intranet websites from the outside, Talk at Black Hat USA, 2006.

[68] VMware, VMware vSphere hypervisor retrieved 16 July 2011, from https://www.vmware.com/tryvmware/?p=esxi&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a.

[69] F. Lombardi, R. Di Pietro, Secure virtualization for cloud computing, Journal of Network and Computer Applications (2010).

[70] M. McDowell, 2009, Understanding denial-of-service attacks retrieved 21 June, 2011, from http://www.us-cert.gov/cas/tips/ST04-015.html.

[71] R. McMillan, 2009, Researchers find a new way to attack the cloud retrieved 21 June 2011, from http://www.computerworld.com/s/article/9137507/Researchers_find_a_new_way_to_attack_the_cloud.

[72] US-CERT, 2011, What is phishing? retrieved 22 June 2011, from http://www.us-cert.gov/nav/report_phishing.html.

[73] G.H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, Paper presented at the Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995.

[74] R. Lopez, E. Onate, A variational formulation for the multilayer perceptron, Artificial Neural Networks—ICANN 2006 (2006) 159–168.

[75] J.C. Platt, Fast training of support vector machines using sequential minimal optimization, 1999.

[76] J.R. Quinlan, C4. 5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993.

[77] E. Frank, I.H. Witten, Generating accurate rule sets without global optimization, Paper presented at the Fifteenth International Conference on Machine Learning, 1998.

[78] A. James, J. Barbara, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1982) 29–36.

[79] J.A. Hanley, B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology 148 (1983) 839–843.

[80] R. Kohavi, F. Provost, Glossary of terms, Machine Learning 30 (1998) 271–274.

[81] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines: and Other Kernel-Based Learning Methods, Cambridge university press, 2006.

[82] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, 1973, Wiley, New York, 1973.

[83] V.N. Vapnik, Statistical Learning Theory, Wiley-Interscience, 1998.

Md. Tanzim Khorshed is a Ph.D. candidate at the School of Information and Communication Technology, CQUniversity, Australia. He holds a Master of Computing (Networking) degree from the University of Western Sydney, Australia. He was working with a top level systems administrator group in an enterprise cloud environment before starting his Ph.D. candidature. His research interest includes Cloud computing, Network Security and Virtualization.

A.B.M. Shawkat Ali is currently working as a senior lecturer with the School of Information and Communication Technology, CQUniversity, Australia. He holds a Ph.D. in Information Technology from Monash University, Australia on Statistical Learning Theory: Support Vector Machine. He is an author and editor of two Data Mining books published by Thomson and IGI-Global and has published over 85 book chapters, journals and conferences papers in the area of Data Mining, Bioinformatics, Telecommunications and Sensor Networking. He has served as the PC Chair DMAI 2008, 2009 and also PC member for many international conferences such as IEEE and ACS. He is a regular reviewer of the various IEEE Transactions and Elsevier journals. He is the Editor-in-Chief of the International Journal of Emerging Technologies in Sciences and Engineering. He is also a Senior Member of IEEE.

Saleh A. Wasimi is an Associate Professor and Head of Program, Mathematics and Statistics at CQUniversity, Rockhampton, Australia. He received his Bachelor and Master degrees from Bangladesh University of Engineering and Technology, Dhaka and his Ph.D. from The University of Iowa, USA. He has published 7 books and over 50 refereed articles. He received national and international awards for his research work. Besides being a lifelong academic, he has done consultancy work for government, autonomous and private organizations.