International Association of Scientific Innovation and Research (IASIR)
(An Association Unifying the Sciences, Engineering, and Applied Research)

ISSN (Print): 2279-0063
ISSN (Online): 2279-0071

International Journal of Software and Web Sciences (IJSWS)
www.iasir.net

Network Traffic Auditing and Assessment in Cloud Environment

[1]Mansi Goyal, [2]Richa Chhabra
[1]Student, [2]Faculty
ITM University, Gurgaon, Haryana, India

Abstract: *Cloud computing permits users to remotely store their information into the cloud and provides on-demand applications and services from a shared pool of configurable resources. The protection of the outsourced information within the cloud is dependent on the protection of the cloud computer system and network. Though, there are varied efforts on securing data on the cloud computer system, assessment of information security on the network between cloud supplier and its user's remains a very difficult task. The audit of the cloud computing system and network can give insights on the protection and performance of virtual machines, the OS on multiple data centres, the intra-cloud network managed by cloud providers and also the wide-area network between the cloud user and cloud supplier. Thus, network traffic assessment for cloud auditing is of vital importance so users will resort to an external audit party to verify the info security on the network between cloud provider and its users. This paper confers the following important technologies needed to research network traffic in the cloud computing domain: IP geolocation of network appliances between cloud supplier and its users, observance the data security of the cloud network path, and on-line mining of gigantic cloud auditing logs produced by cloud network traffic.*

I. INTRODUCTION

Cloud computing is clearly one amongst today's most fascinating technologies attributed to its ascendable, flexible, and cost-effective access to computing resources. Massive and tiny enterprises are progressively leverage cloud computing to manage the computing and storage workloads. It's been utilized in the consolidation of IT resources, net based applications, smart phones, and a number of other innovative devices that heavily rely on the physiological condition and performance of network connections in clouds. As an example, cloud computing provides Web 2.0 technologies for audio and video applications that demand high performance/less latent connections. The cloud provider's network will be essential in VM migration as massive organizations move petabytes of internal data to public cloud suppliers and knowledge migrations between personal clouds for backup and disaster recovery. The performance of most applications utilizing the cloud computing domain crucially depends on the network infrastructure between cloud supplier and its users. For instance, a network, where the routers are at risk due to insecure or malevolent open ports, could lead to subscriber information being prone to threat. Also, if a router is troubled by packet losses and inefficient throughput, the various web applications on the cloud will show slow response time. This could cause discontented subscribers to choose alternate cloud suppliers. The first step in providing subscribers or supplier's ability to configure the network in present cloud computing environments is to spot the attributes of assured clouds and also the congruous security and availableness metrics. Historically, the network between cloud supplier and its users has largely centred on providing basic property and firewall support to client VMs at every virtual server. The network security and availableness metrics are needed to enhance access management, privacy, policy-based routing, audit compliance, protocol acceleration, path management, lack of redundancy and inadequate information measure estimation. The supply of the metrics makes it possible for cloud suppliers and its users to tack (attach) the network or assess the network victimisation Service Level Agreement (SLA) within the cloud computing environment. There is a necessity to develop cloud auditing techniques for calculating these metrics to improve safety, accessibility and performance in numerous multi-tenant cloud environments. Cloud auditing is a pressing research area space owing to rising challenges related to data exploitation of massive cloud audit logs. Bequest auditing isn't the answer and there are not any universal cloud auditing standards to evaluate the cloud's performance. Cloud auditing of huge logs needs analyzing knowledge volumes that habitually cross the peta-scale threshold.

In this paper, we present problems and approaches to develop security criterion for cloud auditing. The primary challenge is ascertaining the location of the user and examining the user's cloud service requests. Ascertaining the location of user is crucial for tracking malevolent cloud users. This is an extremely difficult task because of the capability of malevolent cloud users to change their locations and lack of correct information processing geolocation algorithms. An improved classifier primarily based on information processing geolocation rule to enhance the accuracy of locating cloud users has been introduced. The second problem is the incapacity to identify the cloud network's impact on the cloud safety and performance guarantees. There's a necessity to investigate the degree of security provided by the network to knowledge sharing applications installed in cloud environments that span body and network domains. A risk assessment methodology to measure the vulnerability of the cloud network has also been introduced. The third

problem is related to the necessity for progressive data processing techniques to analyze time period cloud audit logs. Traditional data processing techniques perform snapshot scrutiny of network traffic. Snapshot scrutiny is slow to retort to changes in network paths, addition of a brand new knowledge centre website, etc. An associate integrated supervised machine learning and control abstractive model is introduced to adaptively correlate the detection threshold of a time period Intrusion Detection System (IDS) in accordance with operational variations in cloud computing environment.

II. NETWORK TRAFFIC ASSESSMENT FOR CLOUD AUDITING

The cloud auditing logs generally contain data on network connections between cloud users and Virtual Machines instances on cloud supplier. Alas, the knowledge contained in these logs doesn't seem to be decent to perform cloud auditing. There is a requirement for added assessment to pinpoint location of cloud users, network path security, information measure estimation of network hops, accessibility of cloud ADPS and network routers, etc. as an example, queries that need geolocation data for cloud users or routers between cloud user and cloud supplier would rely upon the provision of information processing geolocation system. At identical time, queries which verify the safety and information measure of the network path between cloud user and cloud supplier would need the availability of router security and information measure metrics. This data isn't obtainable in cloud audit logs and requires further measurements and assessment to work out these metrics. Also, these metrics will often change due to operational changes within the network. Thus, there's a requirement for a web data processing model which is able to make sure that the IP geolocation and router security and information measure metrics are timely and correct.

A. IP Geolocation

The augmented concentration of business information and computing in cloud environments scales security risk similarly. Sophisticated cyber attacks are been installed to target the cloud computing infrastructure. For instance, attackers have abused many cloud computing platforms, like Amazon EC2 [1], [2] and Google's AppEngine [3], for botnet command-and-control (C&C). Defeating cyber attacks in cloud environments is difficult, thanks to the twin problems of magnitude and increasing sophistication of the attacks. One in all, the recent security considerations is attributed to the cloud user's information location. Cloud users source their information and computing workloads on VMs on a cloud provider's infrastructure. However, cloud auditing strategies require that some cloud users limit VM locations to sure locations, as specified by an SLA. Cloud users will use IP geolocation techniques to estimate the placement of their data.

IP geolocation provides the flexibility to estimate geographical location of hosts on the net by one IP address. This is a difficult method because of lack of an association between the IP address and associated geographic location. More recently, techniques primarily based on IP geolocation are deployed in cloud infrastructure services, like Amazon's EC2 service. Alas, the target of geolocation has incentive to deceive the geolocation system concerning its true location. Cloud suppliers or their staff might try to violate their SLAs to find client VMs to locations where the operational value of managing VMs is relatively less [4]. However, variations in laws governing problems like security, data discovery, compliance and audit need that some cloud users to limit VM locations to some jurisdictions or countries [4]. These location restrictions could also be a part of a SLA between the cloud user and supplier. Cloud suppliers might plan to break location restrictions in their SLAs to manoeuvre client VMs to cheaper locations. The IP geolocation algorithms got to maintain accuracy of estimating location of cloud knowledge in spite of tries by cloud supplier to manoeuvre VMs to locations not per the SLA. Thus, there's a demand for cloud users to have freelance verification of the placement of their VMs to meet audit needs. Notwithstanding the cloud supplier it is not malevolent; its staff may additionally try and relocate VMs to locations wherever they will be attacked by different malevolent VMs [5]. Thus, whereas cloud users may trust the cloud service supplier, they'll still be needed to own independent verification of the placement of their VMs to fulfil audit needs or to avoid legitimate liability.

Over the last decade, many IP geolocation approaches aimed at correct approximation of the location of network hosts have evolved. These approaches may be loosely classified into 2 groups on the basis of their technique to gather location data. One set of techniques leverages data from business databases to acquire data on the geographic location of IP addresses. These databases store organizational data appointed to IP domains and DNS names. These databases are liable to be coarse grained, generally returning the headquarters location of the organization that registered the IP address. This becomes a retardant when organizations dispense their IP addresses over a good geographic region, like massive ISPs or content suppliers. The databases may be simply fooled by proxies. In the other technique, active delay and topology measurements are used to determine the geographical location of the IP addresses.

An Improved Learning Classifier IP geo-location approach has been devised that extends an existing IP geo-location approach with add-on features and suitable landmark choice. Recently, it's been shown that precision of IP geolocation is improved by casting IP geolocation as a machine learning categorization challenge. This

approach makes it doable to include each network measurements (latency and hop count) and social characteristics (city population density) to find an IP address. The precision of the prevailing machine learning IP geolocation classifier is improved by increasing the list of options to include average delay, variance of delay, mode and median of delay and careful choice of landmarks. The addition of these options makes sure that the approach is less vulnerable to measure errors and different network anomalies affecting the space approximation. To signify the robustness and therefore the accuracy of the approach, the performance on PlanetLab nodes is assessed. The performance assessment shows that on ground truth knowledge sets, this approach provides location approximations with outstanding precision compared to techniques like CBG [6] and Machine Learning Based IP Geo-location [7].

With the dataset from 21,843 routers, the 70% routers are used as the training set and therefore the rest 30% as the testing set. Table I compares our initial results with learning based IP geolocation and CBG. As see within the table, the average error distance estimates created by our technique is lesser than learning-based IP geolocation and CBG. Also with same set of options as learning-based IP geolocation, the geolocation estimates created by our technique is healthier thanks to our landmarks choice policy.

Table 1: Comparison of Mean Error Distances (Miles) with Previous Methods

	Average delay	(Average delay, hop count)	(Average delay, hop count, population density)	(Average delay, hop count, population density, std)
Learning Based	278.96	261.89	253.34	-
CBG	322.49			
Enhanced Learning Classifier	270.35	216.80	206.55	176.33

B. Router IP assessment

In spite of the supply of various network applications on the cloud, key challenges persists when migrating line of- business network applications, together with lack of fine grained security, privacy, audit compliance, and poor liableness. The network applications are usually hosted on servers managed by cloud suppliers to supply service to cloud subscribers. The networking applications like internet services, instant electronic communication, monetary applications, and gaming, typically involve the exchange of sensitive information. The security of those applications depends on the provision of a reliable underlying network between the cloud supplier and cloud subscriber. The underlying network between the subscriber and supplier has mostly centred on providing basic property to client VMs, with basic firewall capabilities accessible at every server. Many key networks security capabilities for cover of information changed between the subscriber and supplier aren't accessible. This undependable network will compromise sensitive and high value info transmitted by the applications. Figure 1 shows the model of the system. The network traffic between the cloud subscriber and cloud suppliers flows through the network of core routers. The protection of the network traffic hinges on the protection level of the core routers in the path from the cloud subscriber to every of the cloud providers. Assessing the protection level of every individual router within the path can offer the cloud subscribers a mechanism to judge the protection of the network methods to each cloud supplier. The cloud subscriber can like better to conduct business with the cloud supplier, whose info traverses the foremost secure network route.

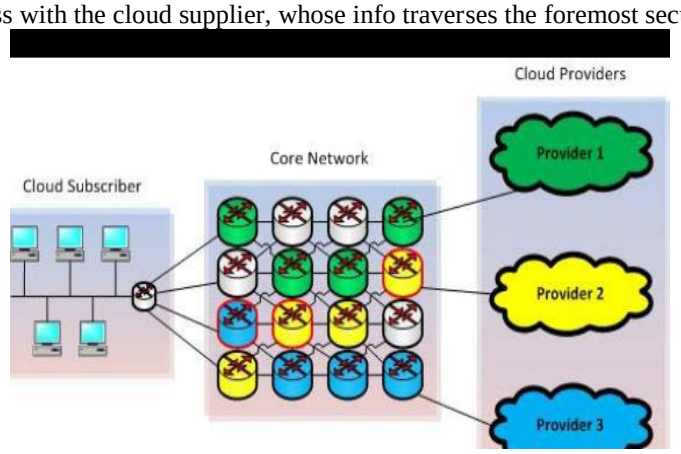

Figure 1: Inter Cloud Network

Cloud security problems have recently gained traction within the research community, wherever the main target has primarily been on protective servers on cloud providers (securing the low level operating systems or

VM implementations). Unsecured cloud servers are verified to be unfit with novel DoS attacks [8]. However, while such threats to cloud servers are widely understood, it is less appreciated that existing network infrastructure itself is prone to constant attack as well.

The safety level of the network between the cloud supplier and cloud user based on the data collected from routers within the core network has been assessed [9], [10]. This work is driven by the observation that the performance of the bulk of data-sharing applications deployed on cloud infrastructures can rely upon the secure underlying networks. The information collected from the routers helps in the assessment of software package and protocol vulnerabilities. The impact of the risks of routers on the performance of information-sharing applications in the cloud is also analysed. The router port vulnerability method demonstrates that insignificantly exploitable routers exist between the cloud supplier and user for adversarial attacks to be possible. The results additionally demonstrates that there enough unsafe open ports on routers which might be compromised on a large scale by technically unsophisticated threats. The data collection method is shown in Figure 2. Figure 3 shows the collation of Confidentiality risk of the cloud network for the two cloud suppliers. It is observed that the high risk and medium risk clusters are dominated by cloud B, on the other hand, the low risk cluster is dominated by cloud A. The identical trend is again observed when the Integrity risk as shown in Figure 4 is compared. The supply risk additionally follows a similar trend as shown in Figure 5.

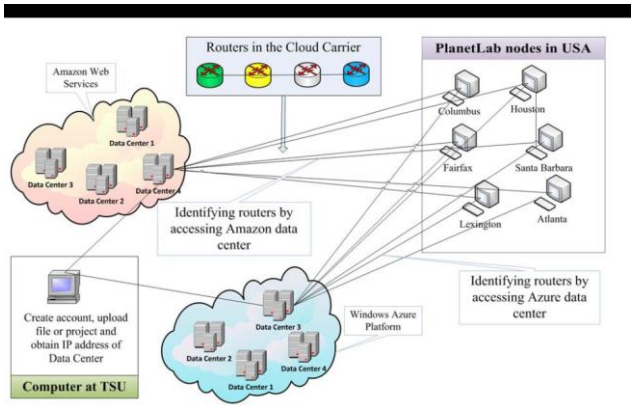

Figure 2: Router Data Collection

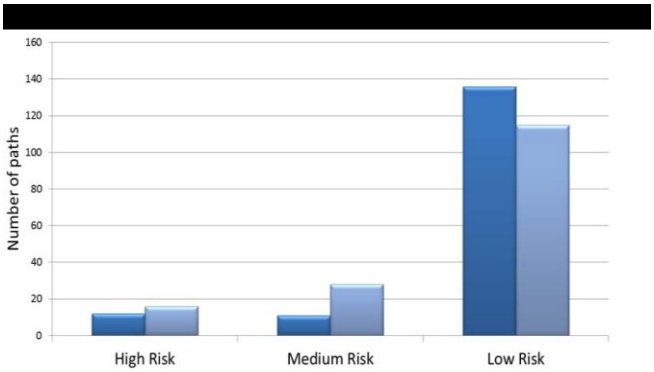

Figure 3 Comparison of Confidentiality Link

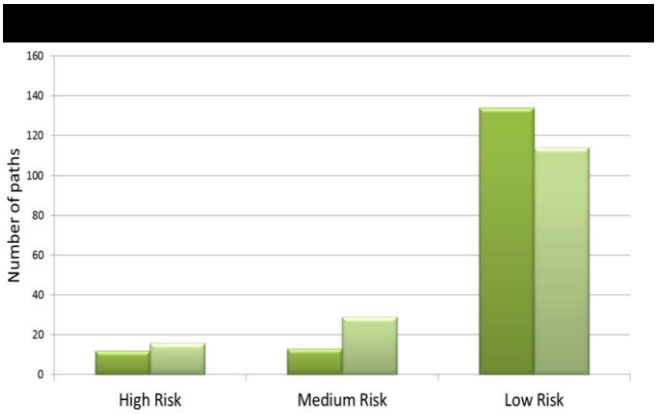

Figure 4: Comparison of Integrity Risk

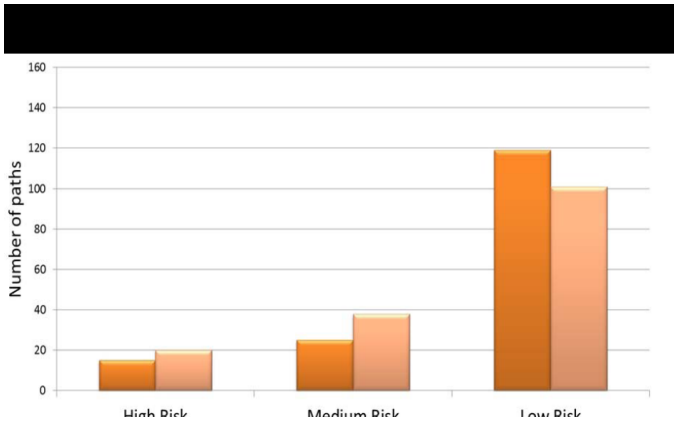

Figure 5: Comparison of Availability Risk

III. Online Data Mining:

Network traffic assessment in cloud computing atmosphere traditionally uses data processing techniques to master a model that precisely captures traffic patterns between cloud subscriber and supplier. However, these techniques rely upon the availability of the cloud system's traditional itinerary profile. However the statistical fingerprint of the traditional traffic pattern will change and shift over time due to various network level development. The changes in traditional traffic patterns over time lead to idea drift. Some changes will be temporal, cyclical and can be passing or they'll last for extended periods of time. Counting on variety of things, the speed at that the amendments in traffic patterns happens may also be variable, ranging from close to instant to the amendments occurring over the span of diverse months. These changes in traffic pattern square measure a reason for concern for network traffic assessment as they can result in a big increase in false positive rates. In turn, this result puts an important burden on the post-detection stages that examine the packets that have raised alarms therefore reducing the precision of network traffic assessment. In order to boost the precision of network traffic assessment, there is a desire for an automatic mechanism to sight valid traffic changes and avoid inappropriate circumstantial responses.

Receiver Operating Characteristics (ROC) curves are traditionally used as the factual method to judge the precision of network traffic analyzers. For real-time operation, the prime ROC operating point is selected to threshold count. But, in presence of concept-wandering network traffic streams ROC curves produced using fixed thresholds cannot provide the best possible precision for a network traffic analyzer. Also, a fixed classification threshold can limit the precision that the network traffic analyzer can achieve. Also, the threshold determination using ROC curves also bring out unwanted human interference in network traffic analyzer's functions.

In this section, the integrated supervised machine learning and control theory model is presented to adaptively tune the detection brink point of real time network traffic assessment in accordance with varied cloud user, supplier and network behaviour for drift in traditional traffic patterns. The model includes a classifier primarily based on a support vector machine that executes in a gateway router and performs as a part of the feedback management engine. Our model includes a relative entropy measurement scheme primarily based on Kullback-Leibler divergence for quantifying changes in traditional traffic. The feedback management engine relies on a PID controller which monitors changes in relative entropy metrics. Our model can be adjusted to varied range of network traffic assessment tasks on cloud computing systems. It is shown that, with some changes, our model can be accustomed to guide the network traffic assessed by providing data on when the traffic modification happens.

Traditionally, network traffic analysers are often updated in existence of a streaming traffic. There is a need for an all-encompassing technique that can accustom to the threshold in existence of concept-drifting network streams. This technique will be independent of the aberration detection algorithm and should simultaneously operate with great precision for any given algorithm. Such an all-encompassing concept-drift detection methodology to accustom the ROC threshold is largely unexplored.

A. Adaptive ROC thresh-holding in network traffic analyzer

In this section, the interspersed control-theoretic and machine learning model is presented which will be competent of accustoming the ROC threshold in network traffic evaluator in existence of concept-drifting network streams. The model consists of three interspersed components: online-SVM model, corresponding entropy evaluation scheme, and feedback control engine. On the basis of the statistical properties shown by the evolving data stream, the following section explains how the three components can precisely track the concept drift in real-time traffic evaluation. To put things in prospect, before proceeding with the explanation of the algorithm, we underscore our motive and high-level methodology.

B. Methodology

It is argued that a precise network traffic evaluator should fundamentally detect time-varying input traffic patterns and confirm its classifier model accordingly. If precise, such a robust thresh-holding mechanism will allow a network traffic analyzer to obtain higher average detection and false alarms precision. This leads us to the following explanation for the proposed robust thresh-holding technique. If we can precisely detect the statistical changes in the time evolving traffic under smaller perturbation conditions, the SVM model can be coached online using only the new information stream. Tracking the statistical changes requires a competence to imbibe the SVM model without retraining on the complete data stream. Once, a new SVM model is imbibed, a proportionate entropy measurement is needed to quantify the differences between previous and new SVM models. Figure 6 shows the proposed robust thresh-holding technique in an aberration based network traffic analyzer. The reader is referred to a recently published research paper for more details on the Adaptive Thresh-holding methodology [11]. In the next sections, we provide a comprehensive description of 3 important components of our technique.

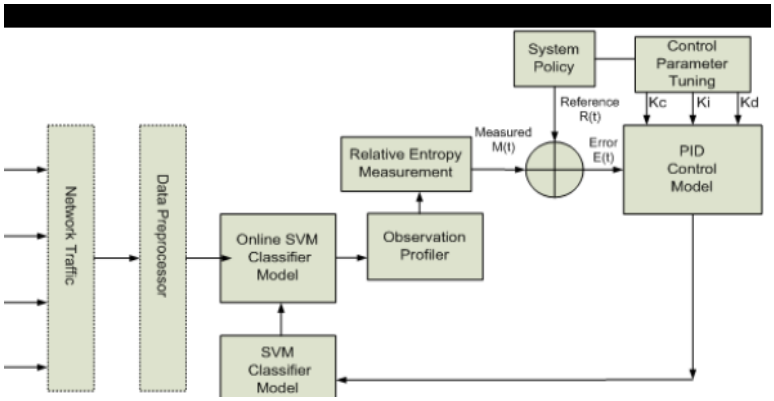

Figure 6: Adaptive Thresholding in Anomaly Based network traffic

1. Online SVM

SVMs are applied in batch mode in traditional machine learning systems. That means, an SVM is trained on a complete set of training data and it is tested on another set of training data. The learner distributes a new observation, gets feedback on the speculated result, updates its hypothesis appropriately and then waits for a new observation. Online learning allows an installed system to accustom itself in a variable environment. Re-training an SVM from starting on the complete set of previously viewed data for every new example is cost inefficient. But, using an old hypothesis as the initial point for re-training cut downs this cost to a great extent. We proposed a relatively smooth algorithm which is universally convergent from any initial point.

2. Relative Entropy Measurement

To detect the concept drifts in network traffic, relative entropy evaluation will be used. Analyzing changes to the normal profile are required for concept drift detection, where training set in itself is insufficient and the profile needs to be refurbished continuously. Using previous data unselectively helps when there is a time evolving concept drift, if previous and new concepts still have some congruity and the quantity of old data chosen randomly just happens to be correct. This needs an efficient access to data mining which helps in selecting a combination of previous and new data to make precise profiling and further grouping.

3. Feedback Control Engine

The feedback control engine will be devised to drive a conservative response primarily based on hysteresis to lower the frequency of erroneous positives caused because of concept drift, thereby avoiding unsuitable ad hoc responses. Increase in erroneous positives can reduce the speed of the system and badly impact the efficacy of the network traffic analyzer.

IV. CONCLUSION

This paper presents methodologies for network traffic assessment in cloud computing environment. The methodologies are IP geolocation, online data mining and Router IP assessment. The three methodologies are interrelated and required to analyse the security of outsourced information in the cloud. The information's' security in the cloud is reliant on a safe cloud computing system and network. The methodologies presented in this paper provide insights into impact of cloud network on the cloud data safety. IP geolocation provides a competency to locate the location of routers in the network path between cloud provider and user. Router IP assessment provides a system to perform software risk assessment on the routers. Finally, online data mining presents an access to assessment of cloud network traffic, on the other hand limiting erroneous positives from concept drifting streams.

References

[1] Amazon. Amazon virtual private cloud. 2012. http://aws. amazon.com/vpc/.

[2] Cloud Security Alliance, Top Threats to Cloud Computing.www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf, 2010.

[3] Google cloud platform used for botnet con- trol. http://www.infosecurity-us.com/view/5115/ google-cloud-platform-used-for-botnet-control, 2012.

[4] Gill, Yashar Ganjali, Bernard Wong, and David Lie. Dude, where's that IP? Circumventing measurement-based IP geolocation. In *Proceedings of the 19th USENIX Security symposium, 2010.*

[5] Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In *Proceedings of the 16th ACM conference on Computer and Communications Securit, NY, USA, 2009.*

[6] Bamba Gueye, Artur Ziviani, Mark Crovella, and Serge Fdida. Constraint-based geolocation of internet hosts. *IEEE/ACM Trans. Netw.*

[7] Brian Eriksson, Paul Barford, Joel Sommers, and Robert Nowak. A learning-based approach for IP geolocation. In *PAM*, 2010.

[8] Liu. A new form of dos attack in a cloud and its avoidance mechanism. 2010. ACM CCSW.

[9] Luna N. Xiong X. Shetty, S. Assessing network path vulnerabilities for secure cloud computing. In *Proceedings of the IEEE ICC Workshop on Clouds, Networks and Data Centers*. IEEE, *2012*.

[10] Shetty S. Reddy, S. and X. Xiong. Assessing Network path vulnerabilities for secure cloud Computing. In *CCGRID*. ACM/IEEE, 2013.

[11] S. Mukkavilli and S. Shetty. "mining concept drifting net- work traffic in cloud computing environments. In *CCGRID*. ACM/IEEE, 2012.