Advanced Intrusion Prevention for Geographically Dispersed Higher Education Cloud Networks

C. DeCusatis, Marist College, Poughkeepsie, NY casimer.decusatis@marist.edu

P. Liengtiraphan, Marist College, Poughkeepsie, NY Piradon.Liengtiraphan1@marist.edu

A. Sager, BlackRidge Technologies, Reno, NV tsager@blackridge.us

Abstract.

We present the design and implementation of a novel cybersecurity architecture for a Linux community public cloud supporting education and research. The approach combines first packet authentication and transport layer access control gateways to block fingerprinting of key network resources. Experimental results are presented for two interconnected data centers in New York.. We show that this approach can block denial of service attacks and network scanners, and provide geolocation attribution based on a syslog classifier.

Keywords: Authentication, Identity Management, Attribution

1 Introduction

Higher education institutions in the U.S. are expected to spend about $10.8 billion on information technology (IT) in 2016 (up from $6.6 billion last year), primarily driven by investments in enterprise networks [1]. Globally, the higher education market is expected to spend over $38.2 Billion on IT in 2016 alone [2]. According to EduCause, a nonprofit organization of IT leaders from higher education [3], the leading issue driving upgrades for these organizations is information security. Security concerns among higher education institutions appear to be well justified; the environment in which higher education institutions operate, and the data which they store, has made them prime targets for cyberattacks. Recent survey data indicates that 35% of all security breaches take place in higher education [3]. Among those institutions suffering a breach, over 46% verified advanced persistent threat (APT) activity taking place in their environment [4]. Higher education institutions collect and retain valuable data such as student, alumni, and faculty personally identifiable information (PII) including medical records; research data which may be subject to export control regulations; financial and accounting data including student tuition, loans, and institution accounting records; and critical infrastructure or intellectual property information including analytic systems used for grading and research. This type of information is subject to various local,

national, and international security and privacy compliance regulations, including the NIST 800 series of security guidelines [5]. In some ways, higher education can be considered a large enterprise; despite this, higher education is not currently classified as a "mission critical" application by the U.S. federal government [5]. In fact, many large enterprises employ security policies based on the principle "exclude everything, allow specific", while the nature of higher education is just the opposite, and often implements policies such as "allow everything, exclude specific" in an attempt to promote shared academic research and education. This can make it particularly challenging to develop effective security policies for higher education institutions.

A recent example involves the Linux One Community Cloud, a collaboration between industry and academia to provide free access to an open source Linux development environment [6]. In August 2015, IBM announced a series of enterprise-class servers which run only the Linux operating system. The Linux One platforms currently support SUSE, Red Hat, and Ubuntu Linux distributions, along with a variety of supporting tools such as Apache Spark, MongoDB, and Chef. In order to promote development and research on this platform, the Linux One Community Cloud makes it possible for anyone to request a free instance of the Linux One servers and toolsets. This environment is hosted at the New York State Center for Cloud Computing and Analytics (CCAC) at Marist College (a private, 6,000 student institution in upstate New York), and is managed from a IBM development location in Poughkeepsie, NY. However, this open innovation initiative also means that the cloud hosting Linux One is subject to continuous cyberattacks from bad actors who attempt to exploit the open access privileges in this environment. There is a need for an intrusion prevention and authentication solution which limits access to the cloud development code to only authorized users, while at the same time preventing malicious reconnaissance attempts to fingerprint the cloud infrastructure or launch denial of service (DoS) attacks.

In this paper, we present results of a cybersecurity testbed deployed in production for the Linux One community cloud. Our research addresses the unique cybersecurity requirements of this environment, including improved authentication as well as identity and access management within a cloud data center. The key points of novelty for this work include the use of network-based identities in a hybrid public/private cloud; specifically, we demonstrate a combination of BlackRidge Technology first packet authentication and transport layer access control (TAC) technologies. We experimentally demonstrate user identity management in the Linux One community cloud, including the novel ability to prevent unauthorized fingerprinting of key network resources. Further, we have developed original software to parse the logs from these appliances and related honeypots, performing geolocation and botnet classification. This work is intended to address the leading concerns expressed in recent surveys of chief information security officers in academia, and enable replication of our security solution at other colleges and universities. We deploy BlackRidge Technology TAC virtual appliances throughout the network which manage user identity based on the first packet used in transport connection requests. This solution including software developed specifically for this project which performs geolocation and attribution for all unauthorized access attempts, and enables collection of analytic data on attempted attacks which can be processed into actionable threat intelligence. Experimental results are presented,

demonstrating that our approach detected and blocked 1,161 unauthorized access attempts in the first twelve hours of production deployment. Over a period of ten days, our approach successfully blocked over 18,000 attacks, which we have attributed to locations in China, Korea, Brazil, Vietnam, and elsewhere. We also demonstrate the ability to identify insider threats by running our authentication technology inside the college firewall (an essential enabling feature for a NIST zero trust network [7]). We present data demonstrating that this approach successfully prevents IP Spoofing and Denial of Service attacks, and identifies network scanners such as Nessus if they are operating on the cloud network. This functionality was not possible using conventional network security approaches.

The paper is organized as follows. Section I provides an introduction and motivation for this work, and an overview of our novel contributions. Section II describes TAC and first packet authentication technologies in more detail. Section III provides experimental results obtained from the Linux One higher education cloud deployment over a 30 day period. Section IV includes a summary and conclusions.

2 BlackRidge Technology Transport Access Control (TAC) Architecture

Our approach is based on a novel combination of two technologies, namely transport access control and first packet authentication. In our proposed explicit trust model, each network session is independently authenticated at the transport layer before any access to the network or protected servers is granted. Unauthorized traffic is simply rejected from the network, and there is no feedback to a potential attacker attempting to fingerprint the system. Explicit trust is established by generating a network identity token during session setup. The network token is a 32 bit, cryptographically secure, single use object which expires after four seconds. Tokens are associated with identities from existing Identity Access Management (IAM) systems and credentials, such as Microsoft Active Directory or the IAM system used by Amazon Web Services [8]. Explicit trust is established by authenticating these identity tokens on the first packet of a TCP connection request, before the conventional 3-way TCP handshake is completed and before sessions with cloud or network resources are established.

Tokens are generated for each unique entity requesting access to a network resource; these entities are generally a user or device. An in-line virtual security gateway is then implemented between the equipment being protected and the rest of the network. The approach is illustrated in Figures 1 and 2, which show a conventional security architecture before addition of the TAC gateways and our new approach following addition of the TAC gateways. In Figure 1, a conventional security architecture would simply place a commercially available intrusion prevention system (IPS) such as the Juniper 3600 platform between the untrusted Internet and resources connected to an education network (for example, the three Linux servers). However, conventional IPS systems cannot block network reconnaissance and scanning attempts, or perform first packet authentication when a user requests a secure session. To improve on the conventional approach, Figure 2 shows the placement of two BlackRidge Technologies

TAC gateways within a higher education cloud network architecture. A TAC gateway appliance is connected in the path between this user and the remaining network, and a second gateway is positioned before the protected resources. The first gateway inserts an identity token in the first packet of the TCP connection request. The second gateway enforces the network access policy by extracting the token, resolving the token to an identity, and determining the identity's authorizations. Trusted users (attempting to access the education network) have identity tokens inserted by Gateway A; untrusted users receive no such authentication tokens. The TAC gateways are configured to protect sensitive resources, such as the cluster of Linux servers. When the second gateway receives a connection request, it extracts and authenticates the inserted identity token and then applies a security policy (such as forward, redirect, or discard) to the connection request based on the received identity. This gateway acts as a policy enforcement point transparent to the rest of the system architecture and backwards compatible with existing network technologies. Trusted users will be authenticated by Gateway B, allowing them full access to the Linux server cluster. Untrusted users are not recognized by Gateway B, and their first packet requesting a new session is dropped, along with all responses at or below the transport layer. In this manner, the untrusted user is unable to determine that the Linux server cluster exists, and cannot begin to mount an attack. The attempted access is logged in an external syslog server, which allocates enough memory to avoid wrapping and over-writing log entries. Existing security information and event management (SIEM) tools can still be used to analyze the logs or generate alerts of suspicious activity. We note that continuous logging of all access attempts is consistent with the approach of a zero trust network (i.e. not allowing any access attempts to go unmonitored). Conventional denial of service (DoS) and port scanner attacks from an untrusted user are similarly blocked, effectively cloaking the presence of the Linux server cluster in this example. Note that the conventional IPS platform is no longer required, but may remain in place since it is transparent to the TAC gateway authentication process. We may also add features such as honeypots which accept redirect requests from a failed access attempt at the TAC gateway (for example, SSH honeypots may be configured in this manner). This enables the collection of attack data which may subsequently be used to craft actionable threat intelligence, such as attack signatures. Both the identity insertion gateway and identity authentication gateway appliances can be implemented as virtual network functions (VNFs) hosted on a virtual server, router, or similar platform.

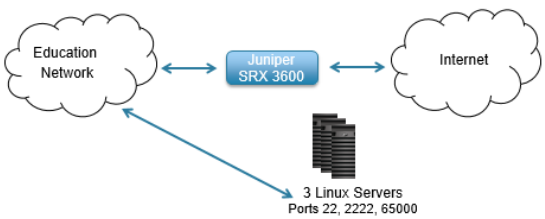

Fig. 1. Conventional network IPS

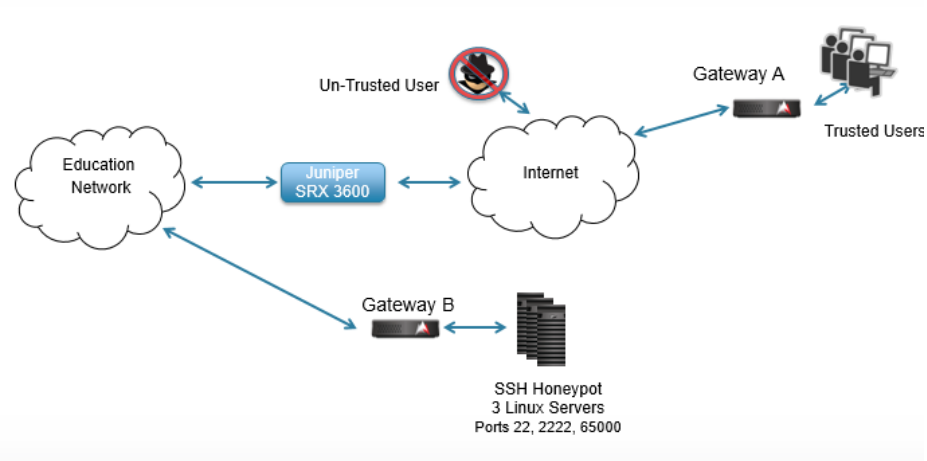

Fig. 2. Deployment of TAC gateways in the education network

This approach has several advantages, including separation of security policy from the network design (i.e. network addresses and topologies) [7]. This approach works for any network topology or addressing scheme, including IPv4, IPv6, and networks which use the Network Address Translation (NAT) protocol and is compatible with dynamic addressing often used with mobile devices. This approach extracts, authenticates, and applies policy to the connection requests, not only protecting against unauthorized external reconnaissance of the network devices but also stopping any malware within the protected devices from calling home (exfiltration). Security policies can be easily applied at the earliest possible time to conceal network attached devices from unauthorized awareness. By preventing unauthorized scanning and reconnaissance, TAC disrupts the attacker's kill chain, blocks both known and unknown attack vectors, and stops lateral attack spreading within a data center. This approach is low latency and high bandwidth since packet content is not inspected. Since the network tokens are embedded in the TCP session request, they do not consume otherwise useful data bandwidth. The combination of transport access control and a segmented, multi-tenant network implements a layered defense against cybersecurity threats, and contributes to non-repudiation of archival data. These techniques are also well suited to protecting public and hybrid cloud resources, or valuable, high performance cloud resources such as enterprise-class mainframe computers and higher education data centers. Further, this approach can be applied to software defined networks (SDN), protecting the centralized SDN network controller from unauthorized access, and enabling only authorized SDN controllers to manage and configure the underlying network. Further, our implementation of TAC uses an innovative identity token cache to provide high scalability and low, deterministic latency. The token cache is tolerant of packet loss and enables TAC deployments in low bandwidth and high packet loss environments.

3 Experimental Results

The Linux One geographically distributed community cloud (Phase One production environment) created for these experiments is shown in Figure 3. This cloud interconnects two physical data centers, namely the Linux One cloud data center hosted at Marist College near Poughkeepsie, NY; the IBM data center hosted in their Poughkeepsie, NY facility. The Marist College and IBM Poughkeepsie data centers are located approximately 8.5 km apart in upstate New York.

Users connect to the Linux One Community Cloud via a secure Internet portal to an Apache web server at the Marist College data center. Content management servers in this data center host instances of OpenStack (Liberty and Juno releases), Maria database server, IBM Java Development Kit (JDK), and IBM BlueMix DevOps Build Engine. These applications are hosted on virtual machines (VMs) partitioned in an IBM z Systems 113 enterprise server. It is necessary to securely authenticate the long distance connection between the Marist College data center and IBM Poughkeepsie data center, (which houses a processing server and content fulfillment engine), To authenticate traffic between these two data centers, BlackRidge appliances implementing TAC and first packet authentication were implemented between these locations as shown in the figure. A physical appliance was installed at the edge of the IBM Poughkeepsie data center network, and a virtual appliance hosted in an IBM z13 enterprise server Zvm virtual partition was installed at the corresponding edge of the Marist College data center network.

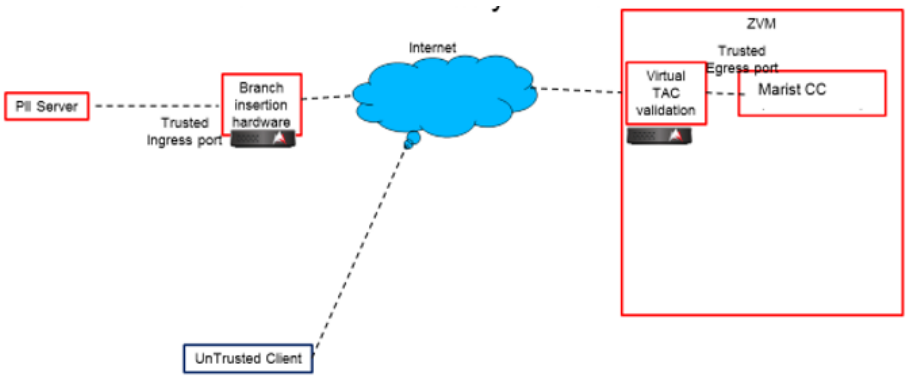

Fig. 3. Linux One Community Cloud Architecture

To determine the effectiveness of the TAC appliances at cloaking attached systems, we performed nmap scans of both the Marist College and IBM Poughkeepsie data center networks before and after implementing the TAC appliances. Representative scans from the Marist College and IBM Poughkeepsie data centers before implementing TAC are shown in figures 4 and 5, respectively. From these scans, an attacker can clearly

see the open port 22 on the Marist network, running OpenSSH 6.6.1, and a traceroute showing network hops within the IBM network, among other reconnaissance data that would be useful in planning an attack on these systems.

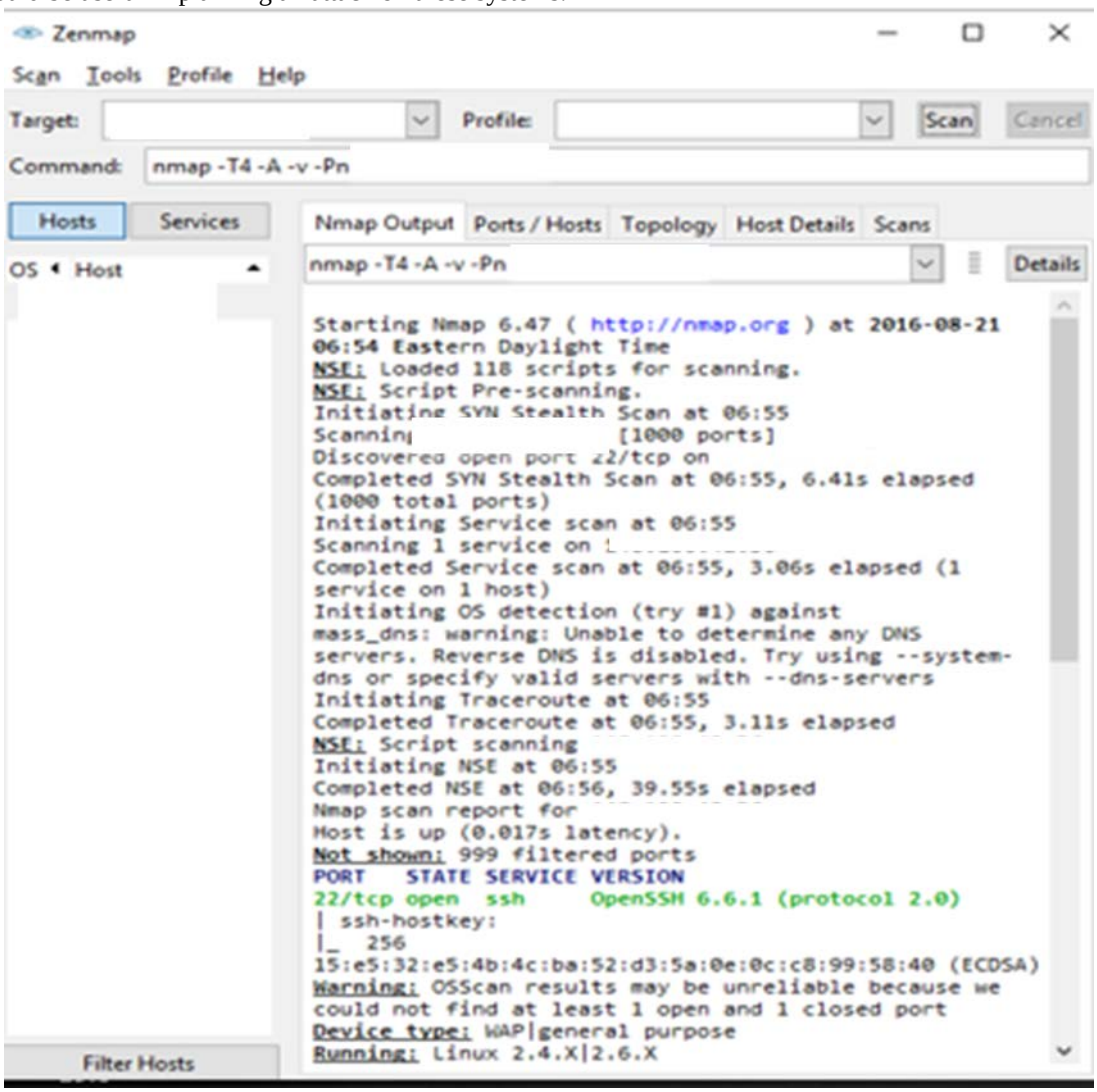

Fig. 4. Marist network scan prior to implementing TAC

A representative scan after implementing TAC on this network is shown in figure 6 (results are equivalent for both the IBM and Marist network segments). Note that we can no longer detect any open ports, including the exposure previously reported on port

22. All attempts to scan these hosts were successfully blocked by first packet authentication, and all responses from the host due to these scans were successfully blocked by TAC. The scan is now unable to determine the host operating systems, port or IP addresses, or services running on the host. These results show that we can effectively block fingerprinting of all devices located behind the TAC gateway.

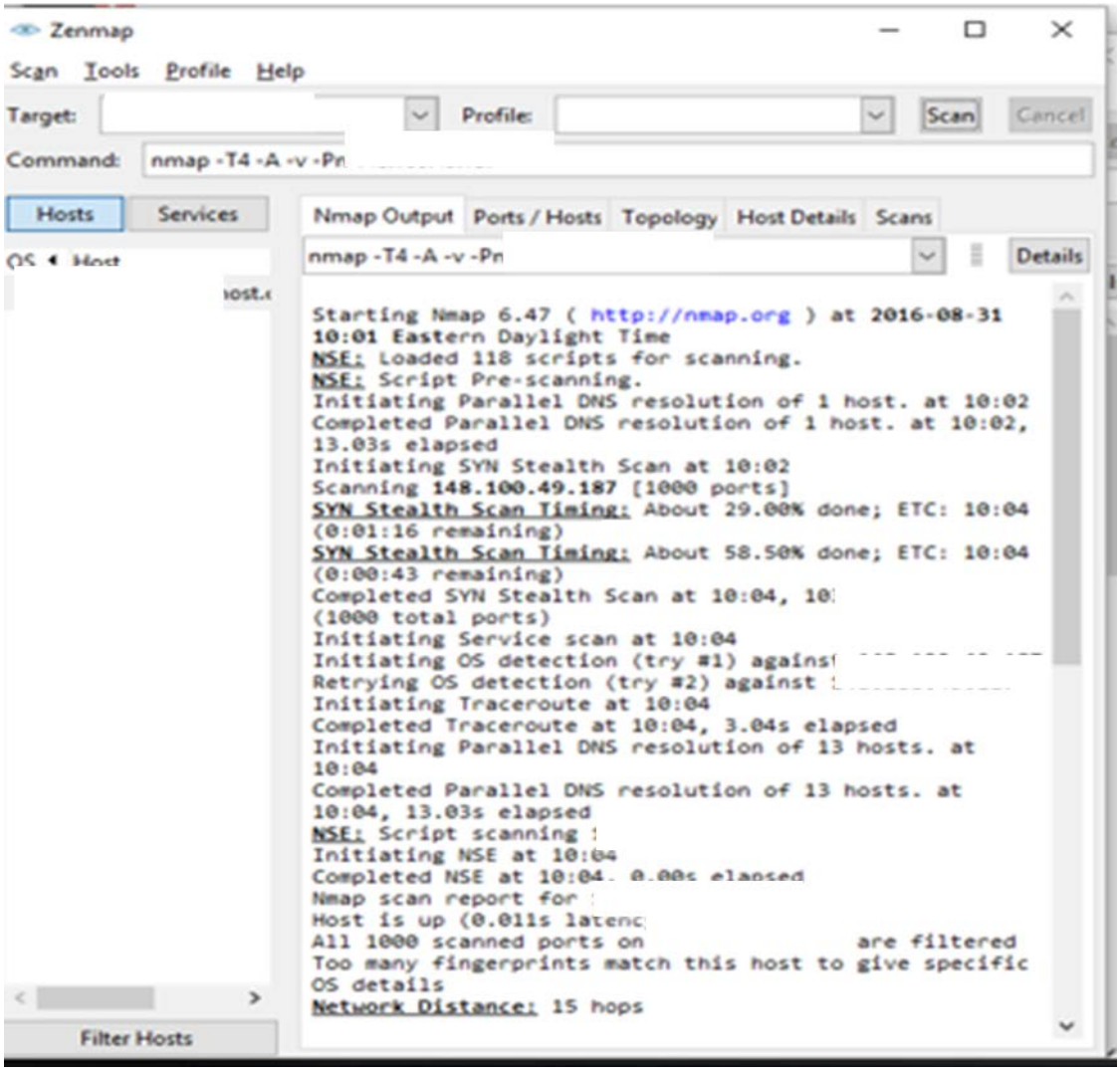

Fig. 5. Marist/IBM network scan after implementing TAC gateways

In order to better understand the attack vectors being used against this higher education cloud, we created a script in Python 2.7 to parse the syslog from a TAC appliance. This script uses the Python regular expression operator ReGex to retrieve data from the syslog including source and destination IP address and port numbers. This data is subsequently processed through a geolocation module which we created for a related project [7] to generate a report of the ISP, ASN, hostname, latitude, longitude, country, state/province, and city of each attacker in JSON format. The TAC appliance was programmed to automatically blacklist any IP address which attempted more than 100 accesses to the network within 30 seconds. The log parser which we have created also classifies blacklisted IP addresses as potential DoS attacks or port scanners. We also collect data on the number of attacks generated from unique IP addresses. All of this data is used to create a profile of the attacker, which can be correlated with known botnets or hacker groups.

For example, during the first 12 hours of monitoring the Linux One cloud after installing the TAC appliance, there were numerous unauthorized attempts to access the system. At this point the TAC system was placed into enforce mode, and successfully blocked all subsequent unauthorized access attempts. The TAC appliance remained in enforce mode for the next 10 days; a list of the top attacking IP addresses, and the top 10 attacking countries, is shown in figures 6 and 7, respectively.

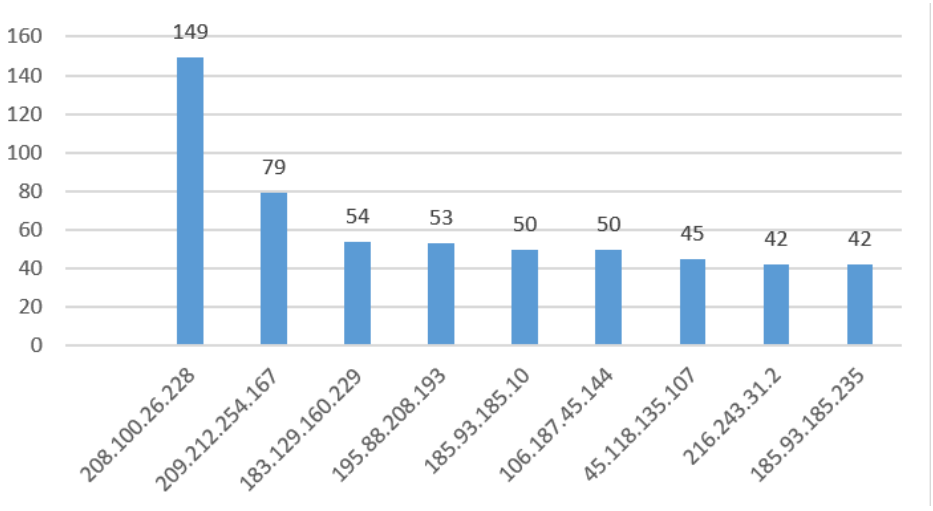

Fig. 6. Number of attacks attempted from the top attacking source IP addresses

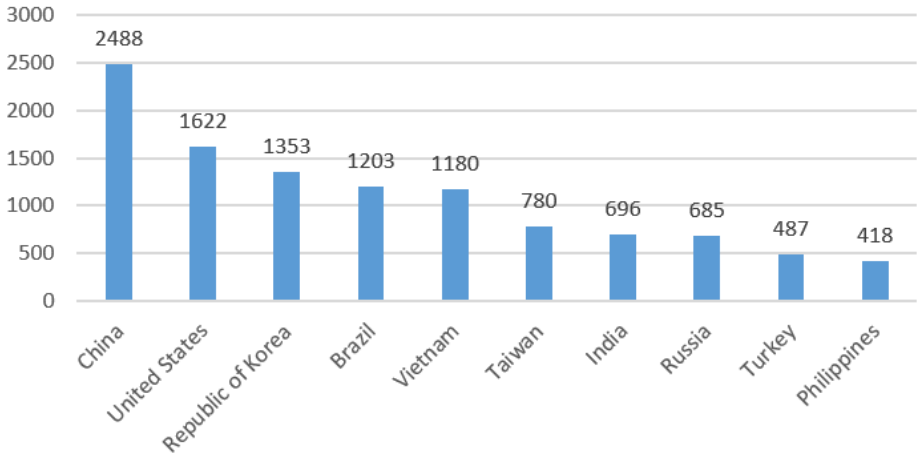

Fig. 7. Number of attacks attempted by each of the top attacking nations

For example, analysis of the TAC appliance logs revealed a DoS attack against port 23 (originating from the Shangdong provide in China). We configured the TAC appliance to block unauthorized access attempts after 10 seconds of continuous attempts from a given site, and to keep these sites blacklisted for one hour. Using this technique, we successfully blacklisted the DoS attacker while continuing to collect log information on the attack. In this manner, we have demonstrated that the TAC appliance provides improved protection by identifying and blocking attacks which were previously undetected on the education network.

Further, we assessed the performance logs of the IBM z Systems enterprise server in the Marist College data center before and after these attacks. Prior to implementing the TAC appliance, the server attempted to block unauthorized attacks using network appliances (such as intrusion prevention systems). This approach was replaced with a single TAC gateway, protecting all VM's on the server at the point of entry. We further demonstrated that the TAC appliance was able to block IP spoofing on the network. By comparing nmap scans of the network before and after implementing the TAC appliance, we can show that attempts to perform IP spoofing are effectively blocked by the TAC appliance. A scan of the network using the Spoofer tool (part of BGP-38 recommended by the National Science Foundation [7]) confirmed that both IPv4 and IPv6 packets attempting to spoof the network were blocked (including private and routable addresses). In a related test of egress filtering depth, the BGP-38 tracefilter test found the network unable to spoof valid, non-adjacent source addresses through even the first IP hop.

Additional statistical data on attacks against this system was obtained using LongTail, an open source botnet classifier which we developed for a related project at Marist College [7]. This classifier was used to identify SSH brute force botnet attacks against the Linux One educational network, and to evaluate the effectiveness of blocking these attacks using a conventional intrusion prevention system and the TAC appliance. For

this test, we first monitored the total number of attacks against an SSH honeypot deployed in the Marist College network ingress from the IBM Poughkeepsie site; statistical analysis of these attacks is shown in Table 1. We then evaluated a commercially available intrusion prevention system, the Juniper SRX 3600, under the same conditions; results are shown in Table 2. We can see that the IPS helped reduce the number of attacks, but did not eliminate them completely. Finally, we deployed the TAC appliance under the same conditions; results are shown in Table 3. In this case, the combination of first packet authentication and transport access blocking was able to successfully block all brute force SSH attacks against the network, and demonstrated a significant improvement over the commercial IPS system alone.

Table 1. Total number of attacks against the Marist education network

Time Frame	Number of Days	Total SSH attempts	Average Per Day	Standard Deviation	Median	Max	Min
Past Day	1	4394	N/A	N/A	N/A	N/A	N/A
This Month	18	183649	10202.72	5447.31	7022.5	20352	5294
Last Month	30	165593	5519.77	7196.19	1194	24666	0

Table 2. Attacks against the Marist education network mitigated by conventional IPS

Time Frame	Number of Days	Total SSH attempts	Average Per Day	Standard Deviation	Median	Max	Min
Past Day	1	30	N/A	N/A	N/A	N/A	N/A
This Month	18	897	49.83	39.75	35.5	124	0
Last Month	30	369	12.30	12.22	10	43	0

Table 3. Attacks against the Marist education network mitigated by TAC gateways

Time Frame	Number of Days	Total SSH attempts	Average Per Day	Standard Deviation	Median	Max	Min
Past Day	1	0	N/A	N/A	N/A	N/A	N/A
This Month	18	0	0	0	0	0	0

Last Month	30	0	0	0	0	0	0

We have also demonstrated that a TAC gateway placed just inside the Marist College firewall is useful in nonrepudiation of insider threats. When a bad actor inside the Marist firewall is detected, efforts to trace the source of the attack traditionally stop at the Marist NAT gateway. It can be a difficult, time consuming process to trace the IP address which originated such an attack. However, a TAC gateway placed behind the Marist firewall (on the Marist side of the NAT) can be used to authenticate the attacker's source IP address much more quickly and efficiently. This new functionality should be helpful not only in discouraging insider threats, but also in helping the college comply with requests and subpoenas from law enforcement agencies investigating such attacks.

4 Conclusions

Recognizing the importance of cybersecurity for higher education, we have developed a novel approach to intrusion prevention and authentication for multi-site, multi-tenant educational cloud computing environments. In particular, we have designed, tested, and implemented this approach for a Linux community public cloud supporting education and research, spanning two locations in New York. The approach combines BlackRidge Technology first packet authentication and transport layer access control gateways to block fingerprinting of key network resources. We have shown experimentally that this approach can block denial of service attacks and network scanners, and provide geolocation attribution based on a syslog classifier. Further, this design offers lower server utilization compared with conventional alternatives. We have also demonstrated that a TAC gateway placed just inside the higher education institution's network firewall is useful in nonrepudiation of insider threats.

5 Acknowledgments

The authors gratefully acknowledge support of the National Science Foundation grant Cloud Computing – Data, Networking, Innovation (CC-DNI), area 4, 15-535, also known as "SecureCloud".

6 REFERENCES

1. S. McCarthy, "Pivot Table: U.S. Education IT Spending Guide, version 1, 2013-2018"; IDC publication GI255747, April 2015, http://www.idc.com/getdoc.jsp?containerId=GI255747

2. J. Lowendahl, T. Thayer, and G. Morgan, “Top ten business trends impacting higher education”, Gartner Group white paper, January 2016, https://www.gartner.com/doc/3186325/top--business-trends-impacting
3. J. Grama, “Data breaches in higher education”, Educause Center for Analysis and Research, May 2014 https://library.educause.edu/resources/2014/5/just-in-time-research-data-breaches-in-higher-education
4. Fireye white paper, “Cyber threats to the education industry”, March 2016, https://www.fireeye.com/content/dam/fireeye-www/current-threats/pdfs/ib-education.pdf
5. G. Stoneburner, A. Goguen, and A. Feringa, “Risk management guide for IT systems”, NIST special publication 800-30, September 2012, http://csrc.nist.gov/publications/PubsSPs.html#800-30
6. A. Guilen and P. Rutten, Driving Digital Transformation through Infrastructure Built for Open Source: How IBM LinuxONE Addresses Agile Infrastructure Needs of Next Generation Applications, IDC white paper, December 2016 https://public.dhe.ibm.com/common/ssi/ecm/lu/en/lul12345usen/LUL12345USEN.PDF (last accessed October 22, 2016)
7. C. DeCusatis, P. Liengtiraphan, A. Sager, M. Pinelli, “Implementing zero trust cloud networks with transport access control and first packet authentication”, Proc. IEEE International Conference on Smart Cloud, New York, NY (November 18-21, 2016)
8. Amazon Web Services Identity and Access Management, April 2016 https://aws.amazon.com/iam/ (last accessed May 20, 2016)
9. BlackRidge white paper, “Dynamic network segmentation”, August 2012 http://www.blackridge.us/images/site/page-content/BlackRidge_Dynamic_Network_Segmentation.pdf