Flow Based Analysis of Advanced Persistent Threats

Detecting Targeted Attacks in Cloud Computing

Andrew Vance
Department of Cybersecurity and Information Assurance
University of Maryland University College
Washington, USA
andrew.vance@faculty.umuc.edu

Abstract— **Cloud computing provides industry, government, and academic users' convenient and cost-effective access to distributed services and shared data via the Internet. Due to its distribution of diverse users and aggregation of immense data, cloud computing has increasingly been the focus of targeted attacks. Meta-analysis of industry studies and retrospective research involving cloud service providers reveal that cloud computing is demonstrably vulnerable to a particular type of targeted attack, Advanced Persistent Threats (APTs). APTs have proven to be difficult to detect and defend against in cloud based infocommunication systems. The prevalent use of polymorphic malware and encrypted covert communication channels make it difficult for existing packet inspecting and signature based security technologies such as; firewalls, intrusion detection sensors, and anti-virus systems to detect APTs. In this paper, we examine the application of an alternative security approach which applies an algorithm derived from flow based monitoring to successfully detect APTs. Results indicate that statistical modeling of APT communications can successfully develop deterministic characteristics for detection is a more effective and efficient way to protect against APTs.**

Keywords— Advanced Persistent Threats, Cloud Computing, Cyber Security, Flow Based Analysis, Threat Detection.

I. INTRODUCTION

Meta-analysis of industry reports suggest that the distributed nature of cloud computing cause it to increasingly be the target of cyber-attacks; particularly Advanced Persistent Threats (APTs). APTs are a sophisticated cyber-attack that use multi stage techniques to target and compromise systems that often go undetected for months. These attacks are sometimes so advanced, that even organizations with cutting edge cyber defenses are vulnerable. Google, Adobe Systems, Juniper Networks, and Symantec were all victims of an APT attack called Operation Aurora. In this attack, valuable intellectual property was stolen and the attack went undetected for six months [1]. RSA was a victim of an un-attributable APT attack that compromised their proprietary authentication token mechanism putting nearly 40 million users at risk [1]. APTs are targeted attacks that characteristically rely on social engineering to deceive victims. APTs often use spear-phishing emails that contain an attachment or link with an exploit that compromised systems [2]. Traditional infocommunication system defenses are ineffective in detecting APTs. Innovative methods are needed to address APTs as leading cyber security companies' report that targeted attacks are on the rise and spreading [3].

This study is an extension of previous research involving a large regional Internet Access Provider providing cloud services that utilized deep packet inspection to investigate cyber-attacks [4]. That previous research revealed that retrospective network traffic analysis using Deep Packet Inspection (DPI) is an effective way to increase the detection occurrence of malicious activity. Ancillary observations during that research indicated that a specific type of malicious activity routinely exhibited unique network traffic behavior involving flow frequency and density. It was determined that this behavior was predominately associated with the multi stage communications of targeted attacks. It was considered that this could provide an effective approach in detecting APT activity in cloud computing.

In this paper, we analyze captured network traffic data and apply a statistical anomaly detection approach to analyze network based communications in order to detect malicious activity involving APTs. The paper is organized as follows. In Sections I and II, we introduce the problem, briefly describe the multi stage nature of APTs, and explain the current problem associated with effectively detecting APTs. In Sections III through VI, we survey APT activity within a cloud based ISP and detail the approach we used for detecting APTs. In Section VII we identify related work and propose future work. Finally, in Section VIII we conclude the paper by summarizing our observation, analysis, and recommendations towards a distinctive approach of an alternative security approach which applies an algorithm derived from flow based monitoring to successfully detect APTs.

II. PROBLEM

APTs establish persistent, covert links, to information technology infrastructures of targeted organizations in order to exfiltrate information which could undermine or impede the critical aspects of a mission, program, or organization [5]. APTs use multiple attack stages (e.g., social engineering, exploitation, and command and control communications (C2)) that routinely defeat security solutions [6].

A. Technique

APTs are covert and have ability to conceal themselves within enterprise network traffic. Detecting APTs presents a problem for current detection methods as these methods largely depend on known signatures of attacks. However,

APTs extensively exploit of unidentified security vulnerabilities during attacks thereby avoiding attack signature detection. Industry research showed that in 86% of the reported cases, symptoms of malicious activity associated with APTs, were logged by organizations security solutions. However, the security solutions neglected to activate security alarms [7]. A survey by the Ponemon Institute revealed that potentially up to 91% of APTs have evaded signature based security solutions such as Intrusion Detection Systems and Antivirus [8]. Prior detection methods utilized DPI which is acknowledged as a more effective way to inspect and detect malicious attacks on the network [9]. However, these solutions experience significant inefficiencies in performance and effectiveness in large data center environments due to processing delays, especially as dynamic data communication increases [10].

B. Challenges

Issues that create challenges for detection are encryption and cloud based communication networks. APTs drop malicious code that is encrypted. And after the targeted system executes the malware, the subsequently infected machines communicate with C2 systems over encrypted channels. These various layers of encryption make detection at the network level challenging. Security solution mechanisms such as signature string matching are not effective in this case and attackers are able to effectively conceal their activities from network-based detection. Additionally, cloud based communication architectures are intentionally distributed and intrusion detection sensors are not congregated and therefore unable to effectively correlate activity which allow APTs to evade detection.

III. SOLUTION

Flow based analysis detects targeted attacks by determining normal versus anomalous behavior. Typical network based APT detection involves discovering the C2 connections, data mining, and the exfiltration activity [11]. In flow based analysis, network traffic is aggregated so the amount of data to be analyzed is reduced. APTs typically "beacon" to C2 servers at given intervals which cause statistical anomalies in network flows. The basic structure of netflow records all them to be analyzed in near real time on even large networks. APT communications can detected by analyzing traffic based on the "volume of transferred data, timing, or packet size." The result is a high detection rate, low false positives, and in-depth incident reporting information designed to accelerate containment of an attack.

A. Framework

The proposed solution relies on a collection and detection framework that includes network gateways with flow collectors (FC) enabled to capture netflow packets (source and destination IP address, source and destination port, start time, end time, mac address, and byte count).

This framework (Figure 1) allows analysis to be performed without need for signatures or DPI. However, it can easily be correlated with other data sources to establish as baseline behavior; number of concurrent flows, packets per second, bits per second, telemetry, number of SYNs sent and received, rate of resets, duration of flow, and time of day.

Analyzing flows is more efficient because the method focuses on the behavior of the connection, not inspecting the payload. All network traffic generates netflow for analysis but the collectors must be enabled with Cisco NetFlow technology.

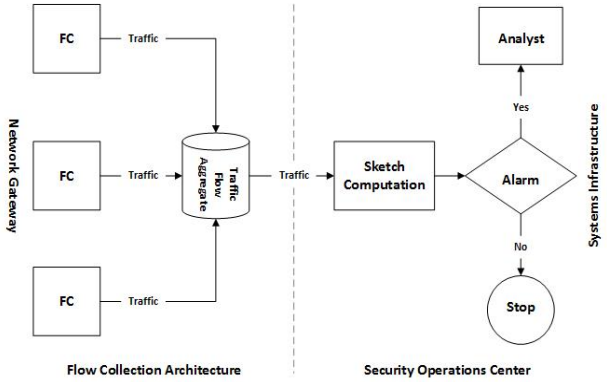

Fig. 1. Collection and Detection Framework

B. Algorithm

There are two major approaches to network anomaly detection: signature-based and non-signature-based. In signature -based approach, the anomaly is detected by looking for patterns that match signatures of known anomalies. In the non-signature-based approach, the anomaly is detected by using statistical techniques applied to network flow traffic. Unusual and/or significant changes in the traffic are classified network traffic anomalies. These anomalies can be changes in link traffic volume, distribution patterns of IP source and/or destination addresses and port numbers, etc. The sources of anomalies include both legitimate and illegitimate traffic. Legitimate traffic can include transient changes in provisioning demands or routing changes while illegitimate traffic can include unauthorized port scans, virus or worms.

Anomalies are comprised of traffic volume, time, series, and frequency at both the origin flow level and the link traffic level. Anomalies at the origin flow level, which are not directly measurable, causes anomalies at the link traffic level, which are measurable. This approach focuses on the link level and does not require any prior knowledge about the anomalies; therefore it is capable of discovering new anomalies.

The proposed solution measures non-signature based traffic and involves flow based measurements and applies a statistical approach for detection. It calculates and establishes standard statistical measurements for normal and anomalous network traffic, then applies sketch-based projections to aggregated traffic allows for more accurate detection capabilities.

1) Network Traffic Measurement

Network events can be represented as vectors in an n dimensional vector space. The dimension of this vector space is reduced to highlight detectable patterns. In the context of network anomaly detection, one or more quantifiable data fields in the IP header is key. Quantifiable data

can be the packet size, the total number of bytes or packets in a traffic flow. Due to the exponential increase in terms of the number of users and applications, it is not feasible to maintain statistics for each individual end-to-end traffic flow. This approach to detecting network anomalies, using aggregate traffic statistics, can more efficiently identify multi stage attacks such as APTs. Aggregation of end-to-end traffic flows at different levels, such as origin autonomous systems, ingress links, and applications, are more effective. For example, the origin-destination flow, is defined as all packets that enter the network at one origin gateway and exits at another destination gateway. The gateways collect traffic measurements on aggregated traffic flows in real time. Traffic measurement, denoted by *s*, can contain one or several traffic features,

$$s = \{c_1(\text{IP}), c_2(\text{Port}), c_3(\text{Size})\},$$

where $c_n(\text{x})$ denotes a function on IP addresses, TCP/UDP ports, packet size, or other traffic features. And where $c_n(\text{x})$ is computed within a time interval, such as count, entropy, or other quantifiable traffic features.

2) The Sketch-Based Measurement

Sketch-based measurement is a method of change detection. It works by deriving a model of normal traffic behavior based on past traffic history and searching for significant changes in observed behavior that deviates from the established baseline model. It is an extension of the Random Projection which projects an *n* dimensional vector into a random-selected one dimensional sketch with a set of random vectors. Subsequent to the projection, the distance between the original vectors is much less than the dimension *n* of the original vectors. The following method of change detection can be used to detect network anomaly.

$$z_{kj} = \frac{1}{\sqrt{l}} \sum_{i=t-n+1}^{t} r_{ik}(x_{ij} - \bar{x}_{tj}) = \frac{1}{\sqrt{l}} \sum_{i=t-n+1}^{t} r_{ik} y_{ij}$$

for i = t − n + 1, . . . , t, j = 1, . . . , m, and k = 1, . . . , l. It works by deriving a model of normal traffic behavior based on past traffic history and searching for significant changes in observed behavior that deviates from the model. Compared to traditional data structures, sketches are more efficient and provide great accuracy.

IV. ANALYSIS

Real-world traffic was analyzed from two Internet access points at two large cloud data centers and studied the results. The method was performed at both access points and evaluated for a period for 3 months. NetFlow records have been collected and archived from two border routers. The detection method was tested against actual data communications traffic which was collected from a major European Internet Access Provider that consisted of a large cloud data center with two geographically distributed Tier 2 network access points. Flow collections were connected to the each Tier 1 and Tier 2 network gateways using a passive network tap. This enabled the monitoring of traffic via the switch's mirroring/span port while still allowing the traffic to flow unimpeded between the data center and customers. A combination of meta-analysis and data analysis was performed. Comparative analysis was then conducted by performing vector distribution and standard deviation evaluations. Analysis of behavioral and statistical data reveal that characteristics could be derived and applied for more effective threat detection algorithm.

Results revealed that the method was effective in separating the considerable dimensional space of traffic measurements which calculated normal and anomalous variation, respectively. The success of the method was that flow aggregation did not significantly disrupt the variation in traffic properties that falls in the normal dispersion. This resulted in a vector B which was a successful detection of an anomaly (Figure 2). This vector was used to filter false alarms and increase confidence in actual anomalous detections. It lead to a low false alarm rate, since the sketches distinctly depicted the existence of anomalies. Plotting the joint distribution of two vectors illustrated normal (A) and anomalous (B) network traffic activity which established a baseline measurement for flows (Figure 2). However, this method did intermittently result in false alarms; the sketch by design is a random projection of the global traffic, and occasionally some sketches failed to signal a detection, even when a true anomaly exists.

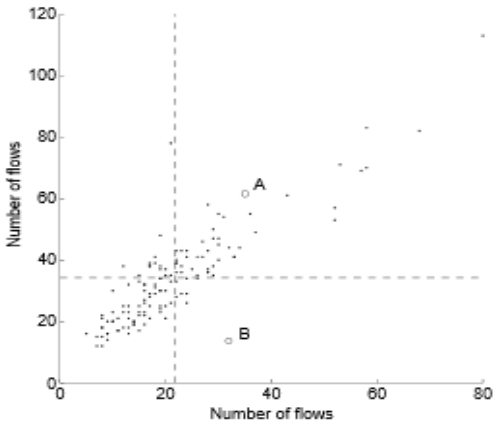

Fig. 2. A and B Vectors Distance Dimensions

False positive rates were significantly minimized, when traffic flows were calculated to illustrate standard deviation between vectors (Figure 3).

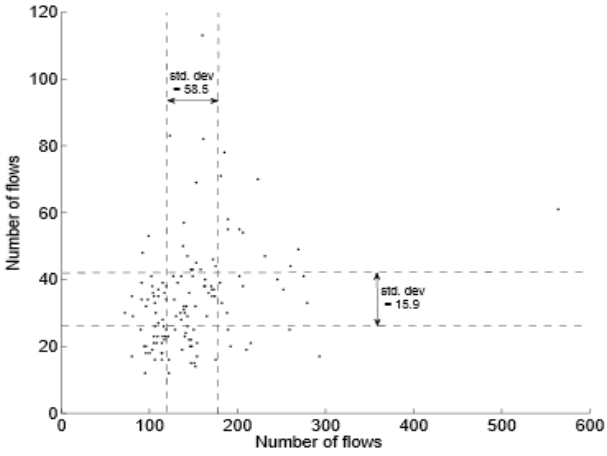

Fig. 3. Standard Deviation and Mean of Vector Clusters

Evaluation of the method determined that calculating for both the standard statistical measurement of anomalous

network traffic and sketch-based projections against aggregated provided a deterministic algorithm that increased APT detection. The collection and detection framework was more efficient at detecting the multi stage activities of APTs which allowed for more effective detection.

Detection performance was analyzed using comparisons of detected malicious activity correlated from the cloud provider's previous security incident reports (Figure 4) and anomalous traffic detected by statistical measurements. Data analyzed over the course of six months revealed that the number of documented investigations increased by 300 percent.

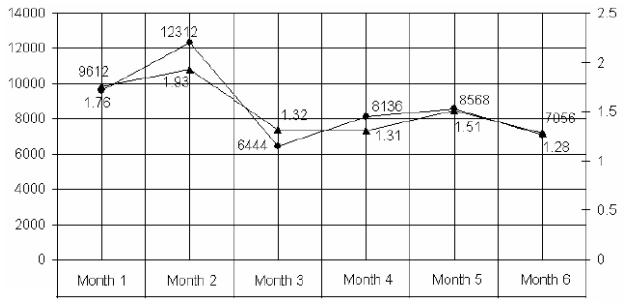

Fig. 4. Detections Correlation

The data was then normalized to isolate malicious beaconing, C2 communications, and data exfiltration. The number of documented APT related investigations was demonstrated to have exhibited an increase of 179% as compared to signature based anomaly detection alone. And an increase of 429% of flow based alerting (includes true and false positive results).

V. Related Work and Future Research

There is increasing research interest towards flow based analysis of target attacks within large scale data centers and cloud service providers [12]. A 2011 doctoral thesis, "Increasing Reliability in Network Traffic Anomaly Detection," by Romain Fontugne proposed statistical based anomaly detection has distinct advantages over signature based detection; such as commonly used in IDSs and AVs [13]. It is possible to detect APT activity by leveraging threat intelligence in network traffic analysis. Related research work conducted by Parth Bhatt, Edgar Yano, and Dr. Per Gustavsson proposed a detection framework that leverages an "Intelligence Module" that utilizes algorithms to correlate malicious events to more effectively detect APTs [14]. False positives are an inherent problem with all anomaly detection systems, as both normal and abnormal conditions can occasionally result in the same observable characteristics. Future work involving flow analysis should include correlated intelligence such as open source intelligence data from sources that collect destination addresses (e.g., domain, email, Internet Relay Chat [IRC]) information via dynamic blacklisting and whitelisting sources such as Symantec. The application of statistical and behavioral flow analysis will enable the development of communication fingerprinting rule use and reduce the rate of false positives

VI. Conclusion

In this research paper, we have demonstrated that flow based monitoring utilizing statistical anomaly detection approach is significantly more effective than signature based detection efficient than packet inspection based monitoring. Although some APT activities will inevitably continue to leverage zero day exploits. C2 IP addresses will continue to change, making it difficult to maintain a defense posture by blocking them alone, network patterns are less subject to change. A significant number of APT campaigns can be consistently detected with anomalous network indicators. A key recommendation for future research can be made based on the results of this research paper; analysis involving anomaly detection algorithms should be applied toward APT detection in conjunction with intelligence data.

References

[1] Fortinet, "Threats on the Horizon: The Rise of the Advanced Persistent Threat," 2013.

[2] FireEye Threat Intelligence, "Advanced Threat Report: 2013," http://www.fireeye.com, FireEye, 2014.

[3] McAfee Security, "Combating Advanced Persistent Threats," White Paper, http://www.mcafee.com/us/resources/white-papers/wp-combat-advanced- persist-threats.pdf , 2011.

[4] D. Smallwood and A. Vance, "Intrusion Analysis with Deep Packet Inspection: Increasing Efficiency of Packet Based Investigations," International Conference on Cloud and Service Computing, 2011.

[5] ISACA, "Advanced Persistent Threats Awareness Study Results," http://www.isaca.org/Knowledge-Center, 2014.

[6] U.S. Department of Commerce, National Institute of Standards and Technology, " Managing Information Security Risk," NIST Special Publication 800-39, March 2011.

[7] P. Giura and W. Wang, "A Context-Based Detection Framework for Advanced Persistent Threats," International Conference on Cyber Security, 2012.

[8] Ponemon Institute LLC, "The State of Advanced Persistent Threats," Ponemon Research Report, December 2013.

[9] A. Li, L. Gu, K. Xu, "Fast anomaly detection in large data centers," In Proc. of the 2010 IEEE Global Communications Conference (GLOBECOM'10), 2010.

[10] X. Yan and J. Zhang, "Early Detection of Cyber Security Threats using Structured Behavior Modeling," ACM Transactions on Information and System Security, Vol. V, No. N, Article A, 2013.

[11] J. Vries, H. Hoogstraaten, J. van den Berg, and S. Daskspan, "Systems for Detecting Advanced Persistent Threats; a Development Roadmap using Intelligent Data Analysis," International Conference on Cyber Security, 2012.

[12] J. Wells, "Combating Advanced Persistent Threats with Flow-based Security Monitoring," Lancope, 2011.

[13] R. Fontugne, "Increasing Reliability in Network Traffic Anomaly Detection," Thesis for Doctor of Philosophy, Graduate University of Advanced Studies (SOKENDAI), 2011.

[14] P. Bhatt, E. Yano, and P. Gustavsson, "Towards a Framework to Detect Multi-Stage Advanced Persistent Threat Attacks," International Symposium on Service Oriented System Engineering, 2014.