BEFORE THE
DEPARTMENT OF COMMERCE
NATIONAL TELECOMMUNICATIONS & INFORMATION ADMINISTRATION

In the Matter of
Promoting Stakeholder Action Against Botnets and Other Automated Threats

Docket No. 170602536–7536–01

Comments authored by Mina J Hanna, Synopsys Inc.
mhanna@synopsys.com

July 20th, 2017

I. INTRODUCTION

We[1] greatly appreciate the opportunity to comment on the Department of Commerce, National Telecommunications and Information Administration ("NTIA") Request for Comment ("RFC") on the matter of "Promoting Stakeholder Action against Botnets and Other Automated Threats". In part, the RFC seeks to identify actions that can be taken to address automated and distributed threats to the digital ecosystem as part of the activity directed by the President in Executive Order 13800, ''Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure.''[2] Through this (RFC), NTIA seeks broad input from all interested stakeholders—including private industry, academia, civil society, and other security experts—on ways to improve industry's ability to reduce threats perpetuated by automated distributed attacks, such as botnets, and what role, if any, the U.S. Government should play in this area."

As indicated by the Presidential Executive Order, NTIA should expeditiously suggest risk management measures and mitigation strategies, to guard the Nation's critical IT infrastructure and secured national data from unauthorized access, intended to cause mischief, by maliciously disclosing or modifying such data and/or perturbing related services. Such perturbations and illegal access to restricted data could potentially elicit local, regional or national catastrophes on public safety, public health, national and economic security. The Presidential Executive Order urges Federal agencies to seamlessly integrate the Cybersecurity Framework[3] and key cybersecurity risk management standards and guidelines defined by the National Institute of Standards and Technology ("NIST"), to develop, implement, and continuously improve agency-wide cybersecurity risk management processes that inform strategic, operational, and other enterprise risk decisions.

Accordingly, we composed the following comment in response to the NTIA RFC. We herein propose a strategy to subvert possible automated botnet attacks on critical national infrastructure. This strategy is an emerging long-term approach, which may prove to be promising for future research and development ("R&D"). Hence, we briefly discuss the role of the Federal government in promoting innovation in the cybersecurity field to secure the continuous development of future applications such as the Internet of Things ("IoT"), that will greatly impact the United States technological leadership and economic growth, and which may be also vulnerable to malicious attacks. In addition, we put forth policy suggestions for NTIA in leading interagency collaboration, promoting Federal investments in future R&D into this approach and maintaining a comprehensive understanding of cybersecurity risk.

---

[1] This response to the NTIA Request for Comments represents the sole views of the author(s) and not the views, opinions or positions of the organizations of which the author(s) is a member, including Synopsys Inc.

[2] Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure, Exec. Order 13800, 82 FR 22391 (May 11, 2017).

[3] The Cybersecurity Framework: Implementation Guidance for Federal Agencies, NIST Interagency Report 8170.

II. BACKGROUND

Cybersecurity risk management comprises the full range of activities undertaken to protect IT and data from unauthorized access and other cyber threats, to maintain awareness of cyber threats, to detect anomalies and incidents adversely affecting IT and data, and to mitigate the impact of, respond to, and recover from incidents. Information sharing facilitates and supports these activities. the ever-evolving cyber threats, vulnerabilities, and opportunities that our nation faces.[4]

Cyber intelligence tools have markedly evolved over the past two decades. Big data systems and practices have been implemented within the cyber tools realm to detect anomalies on secured IT networks whereas machine learning algorithms have been deployed to predict future trends in the cybersecurity field. Despite the exceptional advancement in cyber tools and IT security, current risk management and mitigation technologies are largely reactive. They lack the ability to proactively differentiate between legitimate and malicious traffic from the source similar to in network address translation ("NAT") cases, which is commonly known as a Proxy problem. In often cases, this leads to the decisions undertaken by threat analysts to either block, pass, or redirect the suspect traffic to impact a large number of network users, not only the attacker.

The assets that may be targeted with such attacks render the matter of great importance to the United States national interest and constitute a grave threat to our national security. IoTs and mobile networks are the most at risk, and may be compromised with Botnets and Distributed Denial of Service Attacks ("DDOS"), which can also have catastrophic repercussions on the American public safety and public health. Today, automakers are increasingly outfitting their modern motor vehicles with sophisticated connected technologies that gather, process, store and transmit vast amounts of information for augmenting passenger entertainment and enhancing safety and vehicle performance. Many of these technologies that include Vehicle-to-Vehicle communication pose challenges in the cybersecurity, privacy and security domain. Enabling interconnected automated vehicles and enabling IoT applications in health, education, transportation and business render the need for fending attacks that might threaten the national infrastructure all the more necessary.

The last year, a number of government reports were released that acknowledge the growth and advancements in machine-learning, neural networks and artificial intelligence. The Defense Science Board ("DSB") Summer Study on Autonomy, and reports from the Joint Staff and Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (USD(AT&L)), all emphasize the need to promote the application of ML and AI in DoD defense systems. Similarly, the Obama Administration released the "National Artificial Intelligence Research and Development Strategic Plan" and "Preparing for The Future of Artificial Intelligence", highlighting the importance of promoting the application of these technologies in many other non-defense industries.

---

[4] Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure, Exec. Order 13800, 82 FR 22391 (May 11, 2017).

We therefore propose to combine machine learning and cybersecurity. We propose a machine learning enabled honeypot scheme, designed to identify potential attacks as a cyber defense technology. Machine Learning unsupervised learning algorithms, such as logistic regression, Naïve Bayes classifiers, Support Vector Machines and Neural Networks can be devised to eliminate suspicious traffic and route it to a honeypot. Continuous training of the machine learning module using cyber threat intelligence provided by Akamai and Arbor, can train the machine learning module on distributed network traffic anomalies, and accelerate the learning curve of the routing machine. Equipping the honeypot with a machine learning module has the potential to make the honeypot more proactive in identifying and eliminating suspicious traffic aimed at compromising the compute resource or network to launch automated attacks on other nodes within the same or different networks. This scheme can better predict suspicious traffic by appropriately self-creating user profiles and ensuring false positives are kept at a minimum.

Instead of blocking traffic on perimeter which an attacker can evade by changing the source IP, this machine learning scheme will trap the attacker inside the honeypot. This design can potentially be capable to differentiate the incoming and outgoing traffic based on its behavioral characteristics. If a computer is infected by a rootkit or a bot, this scheme can differentiate between traffic initiated by the original user and traffic initiated by the bot from the same IP intended to infect other computing resources. This would resolve NAT problems as well.

III. TECHNICAL PROPOSAL

The core of the machine learning enabled honeypot scheme we are proposing in this comment, is a Cyber Intelligent Router which consists of a network switching hub, a honey net, external cyber intelligence feed and a machine learning framework. At a high level, this router should be able to differentiate between legitimate and potentially malicious traffic from the same source, and send the legitimate traffic to the production network while forwarding the malicious traffic to the honey net.

The proposed solution is composed of the following:

1) **Router running a machine-learning algorithm**
2) **Honey net**
3) **Threat intelligence**

The router is an Open Systems Interconnection ("OSI") model layer three device, typically placed in front of the network perimeter. This router runs a routing policy based on source Internet Protocol ("IP") address, and not on the destination IP address as the normal routers run. It is also running a Machine Learning module that can identify a malicious session from a stream of legitimate sessions, based on a set of traffic evaluation criteria. Based on the output of the session traffic, a decision to either pass the traffic to production or redirect it to a honey net will be taken.

High interaction honeynets employ a defense strategy that is passive in nature. They attempt to exhaust the attackers and offload the production servers from malicious traffic, by hosting a variety of services that imitate the production environment and making them available to the attackers. However, the honeynet we are proposing has several advantageous characteristics:

a) Shadow copy of production: The honey net must be an identical copy of the running servers including but not limited to listening ports, running code, and database types. The goal behind having an identical honey net is to not allow the attacker to differentiate between the production services and the honey net.

b) Scrambled data: To increase the level of camouflage of the honey, we add feed and/or store real data on the honey net but in a scrambled format. For instance, fake social security numbers mapped to incorrect names and addresses. The records integrity is modified in a way that it appears to provide a real value to the attacker but in effect, these records are of useless.

c) Separation from production: An air gap should exist between the honey net and production servers. There shouldn't be any common infrastructure between the honey net and production environment or a network path from both entities. The Only connecting point is the routing machine.

Companies like Arbor and Akamai claim they have visibility into big swaths of internet traffic which uniquely positions them to run traffic analysis and assign reputation levels for IP addresses over the internet. The reputation evaluation process takes in consideration the type of traffic initiated and received by a particular IP address, location and other evaluation criteria.

The routing machine is to be placed as module of reverse proxy. Reverse proxy is a network device employed to publish applications to the internet. The reverse proxy terminates the connections and performs inspection to the payload then re-write the packet if necessary and send it to the backend application. The routing machine will be operating in two distinct modes. The learning mode, where the routing machine learns about the nature of the traffic and performs a baseline study for the nature of traffic hitting the network perimeter, and the decision-making mode.

Learning mode

The routing machine needs to learn about the type of traffic hitting the network perimeter in terms of resources, services, speed of request, payloads, and size. In addition, it accepts feeds from threat intelligence operators. Feeding all these parameters to the intelligent routing machine so it could be able to identify the malicious from legitimate sessions and access requests. The goal of the learning mode is to train the machine and fine tune the thresholds so decrease the possibility of receiving false positives or passing true positives.

Decision making mode

Once the routing machine matures enough and is ready to make decisions, the operation mode is switched from learning to decision making mode. This mode allows the routing machine not only to identify the malicious session but also redirect the traffic to the honey net.

Use case

Group of internet users are placed behind IP address translating device, typically a firewall or proxy where each utilizes the same egress IP address. One of the users attempted to attack the web site of the popular shopping company ABC. Company ABC has placed intelligent routing machine on its network perimeters. The routing machine will identify considerable traffic initiating from the same IP address. The routing machine will be able to identify the traffic of the attacker from entire legitimate traffic and forward it to honey net.

The attacker started to probe the production, however when his traffic has been redirected to the honey net, he didn’t experience any difference in response because the honey net is responding in the same manner as production environment.

This solution was able to identify the malicious traffic not on the IP level but on the session level. Once the packet has been terminated at the reverse proxy, the routing machine decides if this traffic will be forwarded to the production or to the honey net. Therefore, increasing the burden on the attacker to detect if his traffic is going to the real targets or not. In

the majority of current scenarios, when detecting they have been blocked, attackers can change their IP address. The security administrator cannot block IP addresses as they may block entire subnets with legitimate users to use their services due to the fact that one malicious user is hidden inside the network.

Additionally, this scheme will make attackers inadvertently work for us for free. Since the honey net represents an identical copy of the production, except that it does not store correct integral records. Assuming one of the attackers was able to compromise any of the honey net resources, it is likely that the production shares the same vulnerability. In essence, this should trigger the vulnerability management process to investigate, respond and remediate the production environment if necessary.

Currently available solutions make decisions on the network layer information, and this is insufficient to defend network perimeters against cyber-attacks. This proposal suggests making the decisions on the session layer and utilizing machine learning techniques to identify the malicious traffic, therefore enhancing the security outcome and increasing the robustness of critical IT infrastructure.

IV. POLICY RECOMMENDATIONS

INTERAGENCY COLLABORATION

NTIA should promote interagency collaboration, with the Secretary of Homeland Security, in coordination with the Secretary of Defense, the Attorney General, the Director of National Intelligence, the Director of the Federal Bureau of Investigation, the heads of appropriate sector-specific agencies, as defined in Presidential Policy Directive 21 of February 12, 2013 (Critical Infrastructure Security and Resilience), including the Federal Communication Commission and the Federal Trade Commission. This collaboration should aim at sharing knowledge, defining techniques, standards, best methods of employing long-term approaches and innovative technologies, such as Machine Learning and Artificial Intelligence, as outlined in this comment, to strengthen cybersecurity and risk management defenses for critical national infrastructure.

MODERNIZATION OF FEDERAL IT

NTIA should study the technical feasibility, and cost effectiveness, including timelines, Federal budget limitations, and milestones, of consolidating any number of Federal agencies IT backend infrastructure to more modern architectures or transitioning a subset of IT support services, such as email, cloud storage and cybersecurity defense to these architectures.

NTIA's recommendations in this respect, should all be consistenc with section 227 of the Homeland Security Act (6 U.S.C. 148) and compliance with policies and practices issued in accordance with section 3553 of title 44, United States Code.

PUBLIC-PRIVATE PARTNERSHIPS

NTIA should take a leadership position and promote an effective partnership with the private sector, especially companies focused on developing Machine Learning algorithms aimed at strengthening cyber defense tools.

FEDERAL GOVERNMENT FUNDING

The Federal Government should aim to double its investment in the National Science Foundation ("NSF")'s Cybersecurity initiative and partnership to secure the Internet of Things from $74.5 M to $150M, in addition to supporting other R&D programs within DARPA and the DoD.

V. CONCLUSIONS

We hope that this information, as outlined in this comment, has been of assistance in furthering NTIA's effort in identifying a strategy and actions to subvert threats by automated attacks and Botnets in order to eliminate the risk, such attacks impose on the national security, privacy and safety of the public, critical data and national IT infrastructure. We wish to see NTIA collaborate with other federal agencies (the FTC, FCC, DoD, DHS) and the private sector to accelerate R&D and innovation leadership in the cybersecurity field and enable wider adoption of machine learning and artificial intelligence in cyber defense.