Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 45 (2024), No. 7, pp. 1843–1856
DOI : 10.47974/JIOS-1734

Statistical method for prediction of a targeted spear phishing attack

Vijayant Alok *
Amity Business School
Amity University
Noida 201303
Uttar Pradesh
India

J. S. Sodhi †
Amity University
Noida 201303
Uttar Pradesh
India

Abstract

Targeted Cyber Attacks normally are placed on the intersection of the cognitive as well as the technical domains of hybrid warfare. With a view to glean super normal information for strategic as well as tactical advantages, the cognitive factors activate the technical resources in the domain to achieve one or more objectives. The paper is intended to provide an insight and relationship between the mind-set of an attacker and the resources that he activates to fulfil his desires – why, when and from where.

Subject Classification: *62P30.*

Keywords: *Bots, Botnets, Cyber Warfare, Spear Phishing, Intelligence, Active and passive reconnaissance.*

1. Introduction

The domain of cyber has emerged as the fifth domain of warfare after the realms of Land, sea, air and space. It has proven to be asymmetrically aligned unlike the other four. The ingenuity, innovation and the technical knowhow are the key parameters of defining the lethality and the strike

* *E-mail:* **me@alokvijayant.in (Corresponding Author)**
† *E-mail:* **jssodhi@akcgroup.com**

© Taru Publications

range of the cyber weapons. The battle is more on the cognitive surface while the tools and techniques deployed in this battlefield revolves around a complete asymmetric deployment of technical resources.

Intelligence or fore-knowledge is therefore an important aspect of this domain of warfare. The reconnaissance both active as well as passive is one of the first start points of an attack thread. The process of passive reconnaissance has more or less standardized in terms of the processes that need to be carried out while carrying out such recon exercise, whereas the active reconnaissance and exploitations thereof are heavily dependent on human abilities and the weapons of the game. Passive reconnaissance entails a step wise activities to be performed on the target system and therefore the same can be coded as a tool. Some of the activities such as ping sweep operations, network scanning as well as port and services discovery are some of the activities that can be coded and made to be operated upon by bots and other automated systems. The controller has the responsibility to assign tasks to these bots and thereafter the bots are responsible for carrying out the repetitive tasks on their system using their own bandwidth and resources to obtain results for their respective tasks assigned. The results from the task are obtained by a remote controller and are filtered for further activities on the system to obtain super normal information and intelligence form those systems

As it is often said that this is the battle of the mind, fought on the surface of Ether, with resources hitherto unknown and uncharted. What it means is that one can utilize the resources that come his way to fight and win the battle. This lends to the concept of bots and botnets. Bots are systems that take command from their remote controllers to carry out certain defined tasks for their controllers. The tasks may range from scanning the desired network to creating, dispatching and attacking the targeted computers/systems as per the command of their controllers. Network of such Bots are called Botnets. The automated tasks are carried out by the bots and the controller just reaps the fruits gathered by the bots in their respective areas of operations. The distribution of cyber tasks in real time is therefore done through a command and control interface that the attacker controls. While the bots are performing the routine task, the attacker has enough time to pursue some other non-repetitive and non-automated tasks of building up more blocks.

In the last three decades of percolation of cyber into the warfare realms, it has been observed that most of the attacks that surface have some triggers points, be it political, economic, social or technological.

Estonian Attack of 2007[1] was characterized by a deployment of more than a million Bots distributed across the globe. The distributed denial of service caused heavy losses and crippling of the Estonian Cyber Information Infrastructure effecting the public utility services.Cyber Attack on Georgia was once again characterized by the physical movements of Russian tanks into Georgia thereby sparking waves of cyber attacks on the information infrastructure. Dancho Danchev[2], claimed that the Russian attacks on Georgian websites used "all the success factors for total outsourcing of the bandwidth capacity and legal responsibility to the average Internet user."[1]

During the common wealth games in 2010 in India, there was a massive campaign against the Indian Information Infrastructure almost six months in advance so as to create conditions conducive to the attackers during the games. "India warded off 8,000 cyber attacks during Commonwealth Games" quoted Shri Shiv Shanker Menon, the then National Security Advisor [2].Tokyo Summer Olympics [3] witnessed a surge of cyber attacks with an estimated 450 million attempts on the critical information infrastructure games preparations. What we observe here is very clear that most of the attempts were made with a definite intention and also such massive volumes of attack can not be an effect of a single day. There has to be preparations for the event and also such attacks need to be coordinated properly to ensure the desired effect.

Therefore, the objective of this paper is to understand the dynamics of the behavioural aspects of an attacker, be it a state or individual, the planning that undergoes to bring in the desired effects during the cyber operations and the economic activities that drive past the event.

Literature Review:

A number of cyber attacks that surfaced on the Indian Information Infrastructure were studied and were correlated with the events that were going on at that particular time in the country. The political and other diplomatic events had definite bearings on the attack and its pattern.

---

[1] On 27 April 2007, a series of cyberattacks targeted websites of Estonian organizations, including Estonian parliament, banks, ministries, newspapers and broadcasters, amid the country's disagreement with Russia about the relocation of the Bronze Soldier of Tallinn, an elaborate Soviet-era grave marker, as well as war graves in Tallinn [7]

[2] Dancho Danchev is a Bulgarian Cyber Security Researcher whose work often focuses on understanding the tactics, techniques, and procedures (TTPs) used by cybercriminals and threat actors, as well as the implications of cyber attacks on individuals, organizations, and nations.

Some of the major attack, their strategies, the command and control and other settings that were used in such attacks were reviewed during literature review. Some of the major attacks during this period were – (a) Operation Shady Rat (2010) – the attack targeted several Indian Government organizations compromising sensitive data and networks (b) Operation Red October (2012) – Indian Government entities and the diplomatic missions were targeted in this cyber espionage campaign, resulting in the theft of classified information (c) Operation Hangover (2013) – this cyber espionage campaign targeted various Indian Defence establishments, aerospace and telecommunication sectors causing significant data breaches (d) Shadows in the Cloud – wherein a global report identified several computers within the Indian Government and Tibetan setups been compromised as a part of large scale espionage operation. In addition to all these attacks there were several financial espionage based cyber attacks too, however the paper references a pattern in cyber espionage and the reconnaissance exercise that the attacker carries out before the actual attack.

More than 10000 Indian Government Emails were hacked by supposedly state sponsored hackers in 2012 and was adequately reported by "The Register" then where D4 strategies of Denial, Disruption, Degradation and Destruction was spelt out as a mode of conduct of these cyber-attacks.[4]Cyber Guards have also reported sequentially about the attacks that were observed and detected by different agencies at different times. However, some of the attacks quoted on this site are factually incorrect in their contents.

Most of the contents available on attacks on Indian Information Infrastructure draw our attention to the fact that such attacks are motivated and specifically targeted. The targets are very carefully chosen and the baits selected for the purpose are very specific. Spear phishing attacks therefore suggest that the payloads chosen are attractive enough for the target to bite it and get infected for exploitation by the attackers. In many of these attacks, when the payloads were analysed, it was observed that some of the attacks were very expensive in nature and deployed many of the zero days' exploit that are bleedingly expensive. Therefore, a pattern of state involvement or state sponsored mechanism is observed in such attacks.

The levels of sophistication, the expenses involved and the handling of exfiltrated data and information also speak a great deal about the modus operandi of the attacker during a cyber-espionage attempt of such scale.

Methodology

The research paper is divided into two parts – The first part makes reference to the thought process of a hacker while he plans a spear-phishing attack on its target and the second part deals with the working of a sniffer software created for identifying passive reconnaissance being done by external systems on specific network on which the software is planted. Spear Phishing attacks [5] are a form of cyber-attack on the targeted system wherein the attacker throws a bait to the target waiting for him to consume and then get exploited. Normally, the attacker tricks the target to click on certain links to download malicious codes thereby infecting the computer or the system with malwares [6].

Thought Process of a Hacker – The first activity while performing a targeted espionage operation begins with the generation of interest on the part of the attacker. The attacker tries to identify the results and the return on investment of his time and effort so as to align his strategy towards deployment of cyber resources during the cyber operation. The deployment of resources on the part of the attacker is completely commensurate with the returns that are likely to follow. The numbers of bots, the command and control system, the capacities in terms of the bandwidth and other storage requirements are decided by the attacker depending upon the results and the returns from the cyber operations.

The next step in the life cycle of the cyber operation is that of identification of the target system deep up to the levels of entities that are to become the carrier of the attacks. In order to wage a successful attack that are likely to accrue the desired and anticipated results, the first task for a hacker is to identify the digital contours of his target system and this would include identification of the network that the target is on, identifying the systems that are active on the network and then identifying the ports open on the available systems for further exploitation of the services running on those ports. Till this stage, it is a routine task for any hacker to perform steps in sequence. This is often referred to as the Cyber Kill Chain Framework. The depiction of the process adopted by the hacker as a reconnaissance exercise is given below as a diagram.

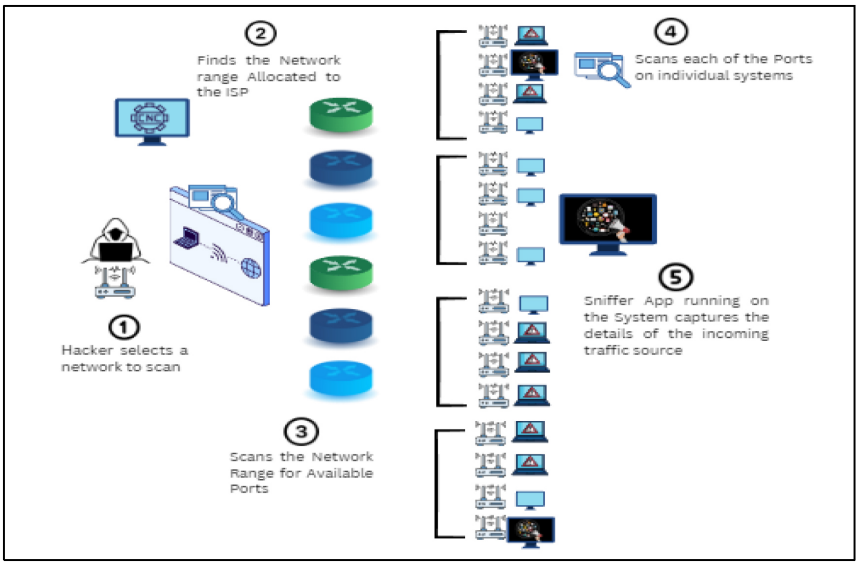


Figure 1

Anatomy of a Cyber Attack

Error! Not a valid bookmark self-reference.In the second phase the attacker identifies the systems that are active and lists out the open ports and the services running on the ports. Thereafter, he targets these systems through specifically targeted attack vector deploying his payloads for exploitation. He exploits the vulnerabilities identified on the system through port and services scan and then exploiting the system for specific purposes such as information extraction, disruptions, denials, degradation and destructive purposes (Figure 1).

The emphasis of the current paper is on understanding the hackers psychology and behavioural patterns while conduct of a cyber operation exercise. The initial steps are more manual in nature and therefore need hackers intervention. Hacker always tries to identify processes that can be programmed to be conducted without his intervention. Therefore, in the entire cyber kill chain framework, he finds that the second stage of reconnaissance can easily be programmed wherein botnets could be programmed to identify among targets, the ports and services running on the system that can be subsequently exploited by the attacker.

Normally, the attacker would codify the steps in the second stage and deploy botnets for carrying out these repetitive tasks of finding the targets. Therefore, while the attacker is trying to generate botnets for running

these repetitive tasks, he would perform a very noisy scans on the network wherein he would run port scanners on each of the available systems. In order to identify and create a network of bots, the attacker always tries to identify the easily available systems that are prone to being converted as a bot. While creating such botnets for subsequent spear phishing attacks, the network is very verbose and therefore generates tremendous traffic on the network. This is the stage at which we set up a trap framework that would capture all such scans and everytime the attacker runs a scan on the network, the trap system would capture his details such as IP Address and other meta data that a network sniffer captures. Every time, he runs a scan on the network, the trap system would capture his details such as IP Address and other meta data that a network sniffer captures. In the current scenario, the trap system with sensor capturing the incoming scans on the system was placed on three ISPs network to widen the network range that the attacker scans.

Technical Details of the Trap System placed on Networks – Basic Windows based system with a packet sniffer running on the system with its port 4000 opened. Port 4000 was chosen at random and any other port could have been chosen for the purpose. One reason for choosing port 4000 was because most of the port scanners that the attackers choose to deploy normally run the scanners on every available port where ports below 1000 are reserved for special services. The port ranges from 1- 65535 and any port in this range could have given similar results. The idea behind selection of the port was that it was not a standard port on which standardized services run on systems and therefore anyone who is scanning on port 4000 definitely had a malicious intent.

The screen shot of the application written for the purpose of trapping the system scanning port 4000 is given below. The language used for coding was C++ and the database behind the same was chosen to be Postgre SQL. The database has the GeoDB Lite imported into it so that the IP Addresses can be converted into the country and city information to be displayed on the panel. The graph below is a depiction of the study that we were envisaging to conduct on the data and would be described in the succeeding paragraphs.

The data was collected over a period of almost 18 months on MTNL and Reliance network based systems. Therefore, in the current study only network scans on these two ranges of IP were captured and the actual scans on the entire Indian network could be many times more than the results that we obtained.

Data Collection Process

Two computers were placed on the two different networks of MTNL[3] and Reliance[4]. The software was installed on the systems. Port forwarding was enabled on the ADSL router so as to let the external traffic reach the intended sniffer software. The IP Address of the incoming vector was identified and then written to the database. After a period of time, the data from the database was analysed and deductions made on the data.

Observation

The data collected during this process had the following format

IP Address	City	Country	Date of Scan	Number of Scans

The panel in the figure below has the following details –

(a) List of connected machines display the current port scanners IP. In the parenthesis is the number of scans made by the IP after the start of the sniffer software. The next is the country code to reflect the country to which the IP belongs to.

(b) The second block provides the details of the IP Address selected in the first block. Like in the current case that is displayed, when 59.103.75.26 [1] PK is clicked, the display shows the details of this IP Address along with its geo-coordinates.

(c) The Dashboard contains the comprehensive summary of total number of countries observed on the scanner, the number of local systems, total systems as well as the ones from China.

(d) The first frame on the second row reflects the current political and other activities of that period drawn from an automated google search and Bloomberg sources.

(e) The graph shows a real time plotting of the Chinese systems and Indian systems where the red lines denote the number of Chinese systems scanning and the green reflects the Indian systems scanning the sniffer module. The x-axis of the graph denotes the time line whereas y-axis denotes the number of systems active at that time frame.

[3] MTNL (Mahanagar Telephone Nigam Limited) is one of the Government owned Internet and Mobile Service Provider in India with its presence mainly in Delhi and Mumbai

[4] Reliance is a private Internet and Mobile Service Provider in India

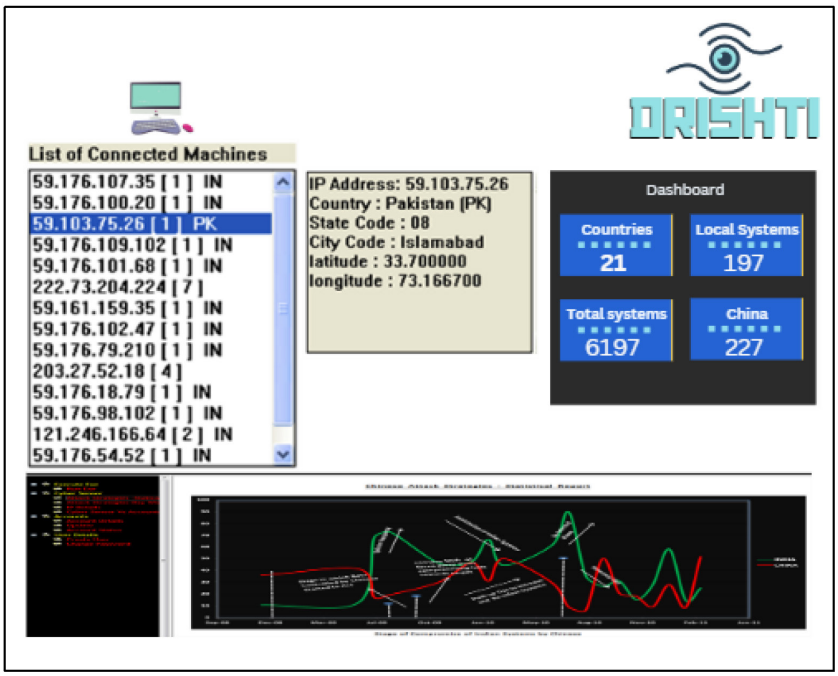


Figure 2

Custom Software DRISHTI

The IP Address from where the scan was originated was captured as the first field, thereafter the details pertaining to the city and the country was populated using the GeoDB Lite database. The system date of scan was also picked up. Number of scans field was also captured and every time on repeat of the IP Address on the screen, the field was incremented. This was set up to give an idea of the timed bots operating on behalf of the hacker (Figure 2).

During this process, the trap system detected multiple IP Addresses from across the globe which was to define that the information infrastructure were continuously on scan by different players spread across the globe. However, for the current study our focus was mainly on the behavioural pattern of a hacker where it was assumed that the passive scan of the network through automated systems are the pre-cursor to a major targeted spear phishing attack on information infrastructure. Therefore, we focussed our attention primarily on IP Addresses originating from China and India alone.

Over a period of time, when the number of IPs that were found touching the sniffer program were populated, the graph emerged as a cress cross between the number of Chinese IPs and the number of Indian IP Addresses. As and when the number of Indian IPS that surfaced on the sensor were high, the corresponding number of the Chinese IPs appearing on the sensor dipped down and vice versa. The result was peculiar and therefore a study was conducted on the data in terms of the cognitive factors governing the hackers mind set as well as the technical process of attack and penetration.

In order to study the behaviour and the graph pattern, some of the events that took place during that time in the cyber space were superimposed on the graph and it was found that the incidents of cyber-attacks and their types had a relationship. Logically, if the attacker was involved in a target based attack, he would not surface on the DRISHTI interface. Therefore, the corollary was that during a spear-phishing targeted attack, the number of times the details of the attacker would surface on DRISHTI would be lesser than at those times when he is actively targeting the network for creation of bots. Also, bots in a selected range are utilized for doing the repetitive tasks which was otherwise being done directly by the attacker. Therefore, it is very clear that while the botnets are appearing on the sniffer more number of times, the attacker's system does not touch the network for the repetitive tasks and therefore, the dip in the count of their appearance is justified.

When the cyber incidents were superimposed on the graph on the timescale, a distinct pattern emerged thereon. While the number of the attacker system dipped on the trap sniffer system DRISHTI, almost the same time witnessed cyber espionage attempts through spear-phishing on

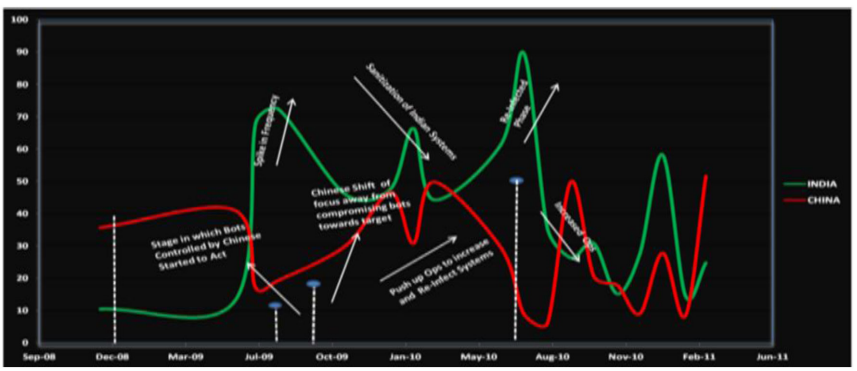


Figure 3

Statistical Chart of Observation

sensitive and identified entities. This pattern was observed almost for four cycles over 18 months of data collection and without fail, the pattern observed no anomalies.

Analysis of the Cyber Spear Phishing Attacks on the timeline of the Statistical Curve Observed (Figure 3) - It was an interesting observation on the time line of the statistical curve when we started observing the behaviour of bots and the actual cyber spear phishing attacks that were detected in India during those periods.

December 2008 – November 2009 – There was a high volume of Chinese IP that appeared on the sensor. This phase was that of a Indian Bot Creation Phase. When in July 2009 the numbers almost were touching the peak (signifying the increase activity of the Indian Bots), the number of Chinese instances on the sensor started dipping downward (reflecting a decreased activity in terms of blind reconnaissance of the network). At the same time when the Chinese IP had a reduced presence on the sensor, a massive targeted spear phishing attack at this time was being planned against the Indian entities, the reconnaissance of which probably would have already been done by the botnets conducting passive scans.

December 2009 – A spear phishing attack on Indian Prime Minister's Office with a payload "Save Ganga Movement" was detected with almost 38 entities on prominent positions being the target was observed and detected.

December 2009-February 2010 – Due to the detection of the attack, a massive correction exercise was undertaken by the Indian Government and many of the systems were scanned for appearance of any malware on the systems. This led to sanitization of many of the machines. Many of the systems that were a part of the Botnet were sanitized and seized to exist on the network as controlled system by the attacker. The reduced number compelled the attackers to recreate the threshold number of bots in their control, thus observing a surge in the chinese behaviour duly supported by the upward movement of the red line on the graph.

February 2010 – May 2010 – With the upward movement of the red line depicting Chinese IPs increase on the sensor, the green line (depicting Indian bots) started rising upward. On reaching a threshold value of Indian IPs, the red line began to dip downward suggesting that the Chinese IPs were now ready to focus on targeted entity exploitations.

July 2010-December 2010 – A major spear phishing attack was observed on the Indian Government Computer Systems and installations. Major organizations of the Ministry of Defence and Home were targeted during this spear phishing attack with the payloads name "Army Cyber

Security Policy". Later in the month of November 2010 another attack on Ministry of External Affairs, Ministry of Defence, Home and NSCS was observed as a part of the spear phishing attack with a payload named "National Security Doctrine".

Therefore, it appeared somewhat like a pattern wherein the sensor detected more of Chinese IP while the attackers were preparing Botnets for their reconnaissance exercise in order to identify potential targets, thereafter passing the job of identification of the target system by the Indian Bots prepared during this time. While Indian bots were carrying out routine scanning of their respective networks and identifying vulnerable systems, the Chinese attackers focussed their attention on sending out spear phishing attack vectors to the identified target. During this time, they appeared less number of times on the sensor deployed. Also, when the attacks were detected by the Indian agencies, the sanitization drive led to sanitization of many of the bots, thereby pushing the attacker to get involved again in scanning the networks and thus increased appearance on the sniffer system. (Figure 4)

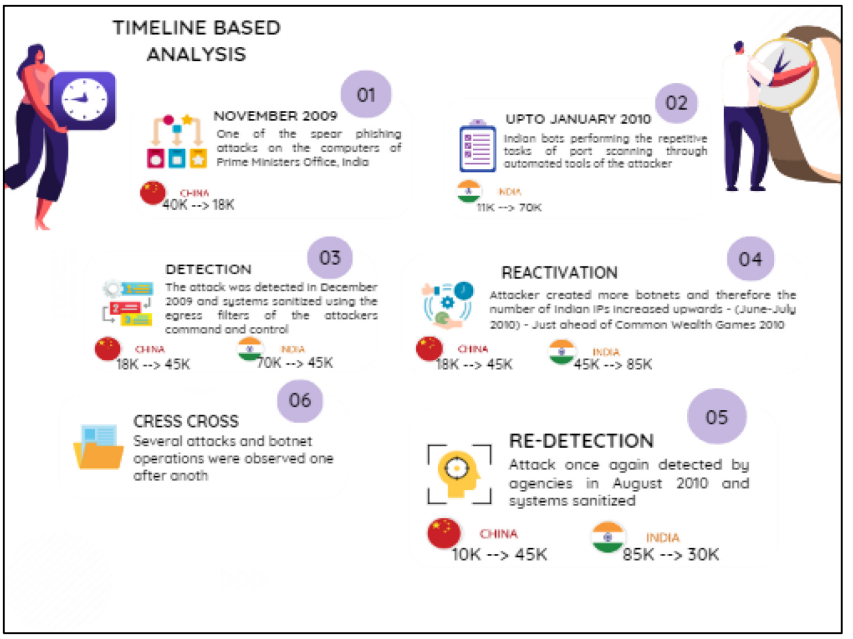


Figure 4

Timeline of Analysis

Results

The above experiment over a period of almost 18 months of data collection using one's own sniffer application on three different networks of ISPs (Internet Service Provider) established relationship between the hackers behaviour as against the resource utilization methodologies in the cyber domain. The deployment of Botnets for repetitive task and utilization of the free time and filtered results derived from the botnet operations for conduct of targeted attacks on entities of importance is now the norm of the day and this exhibit was clearly demonstrated through this experiment. The paper provides a visual and graphical relationship between the events such as cyber-attacks to a pre-task of automated bot-based network scanning.

Derivations and Conclusions

From the observation during the experiment, if such sniffers are run on the nationwide networks and graphs plotted accordingly, it is very likely that the statistics may provide an insight into the possibility of a spear-phishing or targeted attack on entities of importance.

Further, the research establishes a trend of a type of attack that the larger networks are going to face depending upon the numbers of internal bots and external entities appearing on the sensor. It would therefore be helpful in predicting an attack that can be prevented.

There is yet another utility of such systems wherein the egress filters may be set wherein traffic from and to the external systems could be observed which would further detail the quality and type of the incoming attack thread from the hacker. If observed carefully, it would also ultimately give out the details of the command and control of the hacker.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] D. Danchev, "Coordinated Russia vs Georgia cyber attack in progress," ZDNet, Aug. 11 (2008). [Online]. Available: https://www.zdnet.com/article/coordinated-russia-vs-georgia-cyber-attack-in-progress/. [Accessed: Oct. 23, 2024].

[2] C. W. Attacks, "India cyber attacks Commonwealth Games," India Today, May 16 (2012). [Online]. Available: https://www.indiatoday.in/india/north/story/india-cyber-attacks-commonwealth-games-nsa-shivshankar-102376-2012-05-15. [Accessed: Oct. 23, 2024].

[3] "Tokyo Olympics and Paralympics cyberattacks," The Japan Times, Oct. 5 (2021). [Online]. Available: https://www.japantimes.co.jp/news/2021/10/05/national/2020-games-cyberattacks/#:~:text=The%20Tokyo%20Olympics%20and%20Paralympics%20saw%20around%20450,the%20attempts%20were%20all%20blocked%2C%20organizers%20said%20Tuesday. [Accessed: Oct. 23, 2024].

[4] P. Muncaster, "Indian government email hacked," The Register, Dec. 21 (2012). [Online]. Available: https://www.theregister.com/2012/12/21/indian_government_email_hacked/. [Accessed: Oct. 23, 2024].

[5] Panda Security, "Panda Security," [Online]. Available: https://www.pandasecurity.com/. [Accessed: Oct. 23, 2024].

[6] Panda Security, "Panda Security," (2023). [Online]. Available: https://www.pandasecurity.com/. [Accessed: Oct. 23, 2024].

[7] Wikipedia, "2007 cyberattacks on Estonia," [Online]. Available: https://en.wikipedia.org/wiki/2007_cyberattacks_on_Estonia. [Accessed: Oct. 23, 2024].

Received April, 2024