L-025

Estimation of attackers' intentions based on observation of scanning activities

Ngo Kim Cuong
National Defense Academy
239-8686, Kanagawa Prefecture,
1-10-20, Hashirimizu, Yokosuka
em53039@nda.ac.jp

Yasuhiro Nakamura
National Defense Academy
239-8686, Kanagawa Prefecture,
1-10-20, Hashirimizu, Yokosuka
yas@nda.ac.jp

1 Introduction

Intention estimation is the ability to judge and predict the goal of attackers according to attack behavior and network environment. Knowing an attackers purposes can support the network security administrators make the right decision to protect the network resources. Additionally, intent analysis plays an important role in the calculation of the essential threat value and it has become a hot research topics in network security area recently. In this paper, a taxonomy of attack intentions in inbound network traffic is introduced, then the attack intention models based on network scanning activities is proposed and used to estimate attackers' intentions.

2 Related Work

There are many research papers proposed some different methods to detect network attack intention. Jordan Kiprop Koskei[1], used known information about an attackers behavior to create Hidden Markov Models, then decoded the alerts from an IDS (Snort) to discover the intruders high level intentions for the given alerts and predict the future intention. Qiu Hui and Wang Kun[2], proposed a dynamic real-time network attack intention recognition algorithm. By correlating real-time security alerts and vulnerabilities, they found the spread route and stage of attacks based on graph theory and probability theory, then identified the attack intention and predicted the possible transition of attacks. Xinzhou Qin and Wenke Lee[3], presented an approach to identify attack plans and predict upcoming attacks. They developed a graph-based technique to correlate isolated attack scenarios derived from low-level alert correlation based on their relationship in attack plans.

All of these three research papers above have one thing in common: only focus on detecting the attack activity, but not observing the scanning which could early warn of the next step in intrusion. Sometimes, the proposed method still have false discoveries.

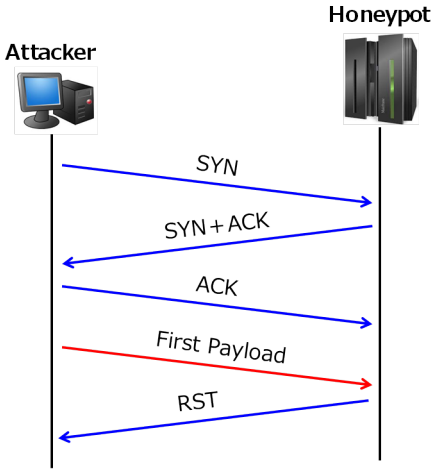

Figure 1: Observation of Scanning Activities

3 Proposed Method

3.1 Network Observation

To collect the network probes traffic data, a reactive observation system was setup to captures all packet that targeted to unused IP addresses as show in Figure 1. This system responds SYN+ACK packet to SYN request connection via TCP, then it will get the first payload which is sent from the attacker.

3.2 Features for Observation

In order to observe the network scanning activity in the inbound network traffic datasets, it needs to extract certain features that characterize the current trends in malicious traffic. This research targets to analyze the first payload from attacker, which includes the network scanning activities, malwares infection activities, and so on.

According to the results of preliminary investigation of scanning packets, the payload arrival ac-

tivities are investigated, then the following several facts are revealed:

- This reactive observation system can get only the first payload for each request connection.
- For each destination IP address can be received many payloads for any time.
- The same source IP address usually continuously sends the same payloads.
- Sometime, the same payloads are received from many different source IP addresses.

From these above conditions, 6 features are selected for each source IP address as shown below:

1. **dstip**: Number of destination IP address
2. **sport**: Number of source port
3. **dport**: Number of destination port
4. **hash** : Number of kind of Payload's Fuzzy hash value
5. **span** : Average time for each arrival time of Payload, [s]
6. **sdev** : Standard deviation of the **span**, [s]

Each features has three or four ranges based on its value, Combination of all these features would definite the scanning activity class of each source IP, then total number of classes are: 4x3x3x4x4x3= 1728 classes. To each source IP address, each packet will be extracted 6 features then be classified into one of scanning class based on prepared decision tree. This process repeats every 10 minutes to detect the change of scanning behavior for each source IP address, thus the scanning activities can be observed.

3.3 Experimental

To collect the real network traffic dataset, this research deployed the reactive observation system on a gateway of the network that allowed all incoming connections to the unused IP address range, but severely limited outgoing connections to minimize damage by the attackers.

Observed Data/Time	**2015/09/01 00:00 2015/09/07 23:59**
ML algorithm	Decision Trees C4.5
Total Number of dst. IP	1,501

Table 1: Experiments Environment

Number of Classified Class	**Number of Scanner's IP**
1	24,995
2	2,338
3	618
4	260
5	164
6	151
7	88
8	26
9	11
11	4
12	6
13	5
14	4
15	1
Total Number of Scanner: 28,707	

Table 2: Summary the Scanning Activity

4 Results

The analyzed data as show in the Table 2 has provided evidence that about 87% of scanners IP address only send the same payload, same interval and target to same port during all the time. It can refer that source IP addresses used only one scanning tool or malware activity.

These results also show that some scanners IP address keep changing the scanning activity to search for the potential system vulnerabilities to spreading malware. As it would seems the attacker changes the attack intention, start a new attack scheme.

5 Conclusion

This research investigated and analyzed scanning activities of each source IP address using decision tree algorithm. It found out how scanning activities are happened. Network administrator should beware of the packets that were sent from a few source addresses which keep changing scanning activities, that must be concerned for detection or prevention new attacks. Furthermore, continuous packet analysis helps estimating the attackers' intentions. My next research is to continue modeling the other attackers intention based on scanning activities classes in order to estimate other attack intentions.