A Comprehensive Model for Elucidating Advanced Persistent Threats (APT)

Rudra Prasad Baksi, and Shambhu J. Upadhyaya
Dept. of Computer Science and Engineering, University at Buffalo, SUNY, Buffalo, NY, USA
Email: {rudrapra, shambhu}@buffalo.edu

Abstract - Advanced Persistent Threats or APTs are not only a matter of concern for the government organizations but also to the industries as well. Recent studies show that many companies have suffered financial damage at the face of an attack by an APT and sustained damage to their reputation and brand value. Detecting the attack and building a defense system against APT is difficult due to the fact that these attacks are very stealthy and targeted. This paper presents a holistic approach via a parameterized model to identify an APT and then to review and assess the vulnerabilities, attacker resources and probable targets. This model would help defenders to build a strong defense mechanism for already seen attacks as well as for future attacks. Several case studies are conducted to support the efficacy of our model.

Keywords: Advanced Persistent Threat (APT), Computer Security, Cyber-security, Petya, Stuxnet, WannaCry

1 Introduction

Advanced Persistent Threats (APT) are putting systems at great risk. They stealthily infiltrate the system, perform reconnaissance to gather information, gain access to critical infrastructure and mount an attack that poses grave danger to the functioning of the entire system. APTs often carry out targeted attacks to achieve their goal. The APT campaigns are typically carried out against government agencies, military systems and commercial entities. The attackers can be individuals, or nation state actors. Malware-as-a-service (MaaS) has made such attacks prevalent by facilitating the attackers with a framework and infrastructure, which would help them to easily mount an attack [1][2].

FireEye, a publicly listed cyber-security company detected 4,192 attacks involving groups, which can be confidently classified as APTs [17]. They also detected 17,995 unique infections caused by malware, which belong to different APT groups. In 2011, RSA Security was attacked by an APT group, which led to financial damage amounting up to $66.3 Million [15][18]. One of the most expensive harms caused by an attack from an APT to an organization is the damage to its reputation. A study by Ponemon Institute states that on average the cost of recovery of the reputation of the company, after an attack by an APT has taken place, often amounts to $9.4 Million [16]. The cost of damage is expected to go higher than the aforementioned numbers, due to the fact that the attacks from the APTs are becoming deeper and adaptive in nature. To the best of our knowledge, there is no formal model to classify a threat as an APT or not; security experts often apply some ad hoc rules after analyzing the impact of the attack. Without a proper threat or attack model, the defense mechanism also becomes unstructured and ineffective. In this paper we propose a model to identify a threat, as and when the attack is unfolding, and determine whether it is an APT or not. With such a full attack comprehension, a more meaningful response can be generated to deal with the attack via the assessment of system vulnerabilities, attacker resources and probable targets. A formal analytical approach is needed for calculating the risk associated with the attack as well as to give the defender enough information and time to design an effective defense mechanism.

This paper takes a parameterized approach to model APTs. Unlike the traditional models, the model proposed in the paper takes into account the fact that the sophisticated attacks may come with a contingency plan. It also puts forward an algorithm to classify a threat as an APT or not. Finally, a risk model is developed to estimate the risk associated with the threat, which would not only analyze the ongoing attack but also prepare the defenders for future attacks exploiting similar vulnerabilities and/or attacking similar targets. To summarize, the main contributions of the paper are a parameterized approach to model APTs and a detailed risk model. Section 3 enlists the characteristics and parameters associated with them to analyze an APT. Section 4 gives an algorithm to determine whether a threat is an APT or not. Section 5 describes the risk associated with each characteristic and finally calculates the total risk, which can be expressed in tangible terms. Section 6 discusses the significance of the model in the face of real world and most recent sophisticated attacks. Finally in Section 7 we conclude the paper and discuss the future work.

2 Preliminaries and Related Work

An APT has a multi-stage life cycle. The authors of [3] came up with the “Cyber Kill Chain” model, which is patented by Lockheed Martin. The life cycle of an APT proposed by Lockheed Martin has 7 stages and that by LogRythm has 5 stages [4]. Rashid et al. [5] came up with a 3-stage life cycle for an APT. These aforementioned models analyze the impact of the attack and then use some rules to classify the threats associated with the attacks as an APT or not. These models do not provide any guidance to recognize and classify sophisticated attacks in a generic sense. Thus, they are not quite useful for analyzing an ongoing attack that may seem to be an APT. Chen et al. [24] suggested Big Data
Analytics and Security Information and Event Management (SIEM) as a potential technique to detect the presence of an APT. But it requires one to know the characteristic features of an APT. APTs are constantly evolving in their characteristics, and are able to mount sophisticated attacks.

National Institute of Standards and Technology (NIST) recently issued a draft document on cyber resiliency [25] and put forward a policy for mitigation of an APT. They have identified ten tactics of staging an APT. This is the first work on parameterizing an APT from a defense framework point of view. They took one tactic at a time and discussed corresponding damage mitigation. Basically, for every tactic there are multiple ways to tackle them. Often, the mitigating process for one tactic is similar to that of another tactic. So, in the event of an attack by an APT, which manifests multiple tactics, similar mitigating processes could be run for different tactics. This will make the defense mechanism slow and the quality of service (QoS) of the processes deteriorate. The NIST model has acknowledged the fact that newer attacks from APTs are adaptive in nature, but they have not taken into consideration probable contingency plan on the part of the attacker.

In this paper, we propose a model that is based on attack parameterization, and one that can help develop a holistic defense mechanism should there be an APT for an organization. This model will account for and has dedicated parameters for probable contingency plans by the attacker, among others. This paper also puts forward a risk model in order to estimate the risk associated with each parameter. The holistic approach based on parameterizing an attack and risk estimates renders scalability to the model, which can be used not only for an attack which is unfolding, but also for future attacks which might exploit similar vulnerabilities and/or have similar targets.

3 A Comprehensive Model: APT Characteristics

Identification of an APT while the attack is unfolding is not an easy task due to the dynamically changing landscape of sophisticated attacks. In order to address this, we first define a set of characteristics of an attack. If the attack displays certain well-defined characteristics, we can say that the threat is an APT within the bounds of our definition and an action can be taken to mitigate the after-effects of the attack. The proposed parameterized model is a customization of the stealth attack model of [14] to fit the new breed of attacks, viz. APT.

Most published cyber-attacks in the literature are single-shot, staged by running certain attack scripts and are instantaneous. The typical defense mechanism is reactive in nature. On the other hand, the APTs are generally multi-shot and spread over multiple stages [14]. The defense against an APT is different from that of the non-APT attacks. It is possible to proactively handle an APT if it is recognized at an early stage. So an early detection and classification of the threat is important from the defender’s point of view. For the purpose of classification, we are going to consider five main characteristics shown in Table 1. For each characteristic we define certain parameters, which would help in identifying the threat. We then use each parameter to calculate the risk.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sy.</th>
<th>Param.</th>
<th>Parameter Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconnaissance</td>
<td>P1</td>
<td>P1.1</td>
<td>External Reconnaissance</td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>P1.2</td>
<td>Internal Reconnaissance</td>
</tr>
<tr>
<td>Exploit</td>
<td>P2</td>
<td>P2.1</td>
<td>Hardware Vulnerabilities</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>P2.2</td>
<td>Software Vulnerabilities</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>P2.3</td>
<td>HW Timing Constraints</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>P2.4</td>
<td>Target Acquisition</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>P2.5</td>
<td>Affected Resources</td>
</tr>
<tr>
<td>Authorized Access</td>
<td>P3</td>
<td>P3.1</td>
<td>Admin Rights</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>P3.2</td>
<td>Backdoor Implants</td>
</tr>
<tr>
<td>Honeypot Interaction</td>
<td>P4</td>
<td>P4.1</td>
<td>Low Interaction Honeypots</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>P4.2</td>
<td>High Interaction Honeypots</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>P4.3</td>
<td>Activity Logging</td>
</tr>
<tr>
<td>Contingency Plan</td>
<td>P5</td>
<td>P5.1</td>
<td>Abort Mission</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>P5.2</td>
<td>Different resources but same mode of attack</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>P5.3</td>
<td>Different modes of attack but same resources</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>P5.4</td>
<td>Search for different Attack Scenarios</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>P5.5</td>
<td>Decoy Attack</td>
</tr>
</tbody>
</table>

Table 1: Characteristics and Parameters of an Attack

3.1 Reconnaissance

Reconnaissance is a characteristic possessed by almost all kinds of attacks. Most of the attacks perform some external
reconnaissance on the system and many do internal reconnaissance once inside the system. Therefore, we define two parameters for this characteristic, namely, P1.1 and P1.2 to recognize both of these activities.

3.2 Exploit

The Exploits characteristic encompasses all the exploitable vulnerabilities and the parameters defined here are used to calculate the risk associated with the vulnerabilities. In this characteristic, five parameters are defined. Each parameter enlists the vulnerabilities associated, which can be exploited by the attacker to mount an attack. Parameter P2.1 enlists all the exploitable hardware vulnerabilities. Parameter P2.2 encompasses all the exploitable software vulnerabilities. Parameter P2.3 gives the time constraints regarding exploitation of the vulnerabilities. It denotes the time required for exploiting the hardware as well as software vulnerabilities. This parameter is particularly very important from the attacker’s point of view. It gives them an idea to run a feasibility test. It is also important from the defender’s perspective as well. This parameter helps a defender to understand whether a system is unconditionally secured, computationally secured or not secured. Parameter P2.4 enlists all the target components, which had either been acquired or will be acquired. This parameter gives an idea to the defender about the components, which are to be secured on a priority basis at the event of an attack. The last parameter of this characteristic is P2.5, which lists the affected resources.

3.3 Authorized Access

The above two characteristics are generic and common to most threat families. Authorized Access is a major characteristic of a quiet invader [6] and/or an APT. This characteristic signifies the capability of the threat to gain authorized access to the system and maintains the foothold. The parameters defined in these characteristics are P3.1 and P3.2. P3.1 is different from the other parameters in the sense that it is a Boolean parameter. P3.1 is True if the attacker is able to gain administrative privileges, else it is False. P3.2 enlists all the backdoor implant tools used to install backdoors. This parameter is very important from defender’s perspective as it gives an idea about the nature of attack being carried out. This parameter also opens up the avenue for the defender to snoop into the communication between the malware installed by the attacker and the Command and Control (C&C) centers. If this communication between the malware and the C&C centers can be breached, then the attack can not only be thwarted but also the family as well as the nature of the attack can be known easily.

3.4 Honeypots

For the purpose of defense many systems use deception as a weapon. Use of honeypots and honeypot farms is such an application [14]. Parameters P4.1 and P4.2 enlist the threat and honeypot pairs and log their interaction. Parameter P4.3 is another Boolean parameter, which is True if the honey-pots are logging the activities of the malware and is False otherwise.

3.5 Contingency Plan

A distinguishing feature of our APT model is the recognition of the presence of a contingency plan in an attack campaign [12]. A defender might come up with a defense mechanism while under attack, but the attacker realizing that a defense is building up might change the mode of attack. The renewed attack might be totally different in nature throwing the defender off guard and making the system more vulnerable than it was before. Therefore, to understand this characteristic, we define 5 sub-parameters. The parameter P5.1 is a Boolean parameter. P5.1 is True when the contingency plan of threatening attacker is to abort mission when discovered, and False otherwise. This may be viewed as a default plan on the part of the attacker. Parameter P5.2 enlists all the resources, which can be attacked using the same mode of attack. Parameter P5.3 enlists all the different modes of attacks, which can attack the same resources, which are currently under attack. Parameter P5.4 enlists all the attacks and resources, which can be attacked except for the current mode of attack and the resources, which are currently under attack. Parameter P5.5 addresses a certain plausible scenario. Though this parameter is enlisted under Contingency Plan it can become the main plan as well depending upon attacker skills and the dynamics of the application environment. It enlists all the probable decoy targets, which are currently under attack in order to take the focus away from the main target. It primarily enlists all the sets in which one or more targets are the main targets and other targets, which can be probable decoys for the main target. It also states which sets are currently active, which means that the decoys are currently under attack and the main target is at risk.

4 Identifying an APT

Determination of the type of attack will require a review of system vulnerabilities, assessment of the attacker skills and capabilities. A risk assessment needs to be performed for the purposes of mitigation. In this section, the model described in Section III is used to develop an algorithm for the identification of the type of threat.

4.1 Step 1: Reconnaissance

The attack identification process begins with the detection of suspicious activities, which are indicative of certain interactions with the system. These could be passive or active processes and can be detected by network or host based intrusion detection systems (IDS). If a process is found to be a legitimate process intended for interaction with the system, it is classified as a legitimate process, else it is classified as a reconnaissance process. Depending upon the signature of the IDS alert, the activity can be categorized as an internal or external reconnaissance. The external and/or internal reconnaissance processes are to be listed under the parameters P1.1 and P1.2, respectively.

4.2 Step 2: Vulnerabilities

Once reconnaissance is detected and recorded under P1.1 and P1.2 (see Table 1), the next step is to list the
vulnerabilities, which can be exploited by the attacker to mount an attack. A list of hardware vulnerabilities that can be exploited is compiled under P2.1 and a list of software vulnerabilities is compiled under P2.2 along with the time to exploit information in P2.3.1 and P2.3.2, respectively. Given the reconnaissance processes, vulnerabilities and timing constraints to exploit each of the vulnerabilities, we can compile the list for probable targets in P2.4. The defense mechanism can scan the system to carry out an audit to know the affected resources and can list them in P2.5.

4.3 Step 3: Authorized Access

The reconnaissance and exploitation of the vulnerabilities are part of any sophisticated attack. But the goal of our model is to find out whether a threat is an APT or not. In order to determine with confidence whether the attack is an APT or not, we look at the next characteristic, viz. Authorized Access. Parameter P3.1 checks whether or not the attacker has acquired administrative access. If it is found that the attacker has gained administrative rights, then the system scans for backdoors installed and/or backdoor implant tools in use. Discovery of the backdoor and/or backdoor implant tools, and figuring out if the malware has gained authorized access or not, implies that the malware might mount a multi-stage targeted attack. This step can occasionally result in false-positives in identifying an APT but rarely false negatives. At this point, the system can start preparing itself to circumvent an APT.

4.4 Step 4: Activity Logging

Systems with high value targets often maintain honeypots and honeypot farms to divert the attacks. Such systems are also targets for sophisticated attacks such as an APT. The honeypots help in activity logging of the processes, which are malicious in nature. There are two types of honeypots, low interaction honeypots and high interaction honeypots. Parameter P4.3 checks if the malicious processes are interacting with the honeypots or not; their activities are then logged for analyzing their behavior and to know the family of attacks they belong to. This way attacks from the same family can be thwarted in the future. The model can be a part of any APT Detection System, and therein Parameter P4.3 will be important because if it is TRUE, then it will trigger the analysis of the honeypot logs. Commercial implementation of honeypots logs the activities of interacting traffic but in this model, the analysis of the logs begins only when it is triggered by Parameter P4.3 so as to preserve the quality of service (QoS). Parameter P4.1 compiles the list of processes interacting with the low interaction honeypots and parameter P4.2 compiles the list of processes interacting with the high interaction honeypots. Both the parameters compile the activities of the processes as well.

4.5 Step 5: Contingency Plan

The contingency characteristic is a true indication of an APT according to our model. With more sophisticated and targeted attacks, the attackers are expected to be equipped with a contingency plan, should they be discovered by the system. At the face of an attack, which comes with a contingency plan, the system may be rendered defenseless. This is because, traditionally defenders prepare for the primary attacks from an APT and ignore the fact that there can be contingency plan in another form or a decoy attack. Parameter P5.1 provisions for checking for a contingency plan of the attacker. The model checks for the contingency plan, if the plan is to “Abort Mission”, which means retreat on the part of the attacker at the event of discovery by the defender, it is listed under parameter P5.1. The system continues to look for other possible attacks, which can potentially be the contingency plan of the attackers. Parameter P5.2 scans and lists all the resources that can be affected by the same mode of attack. Parameter P5.3 scans and lists all the attacks, which can possibly affect the resources that are currently under attack. Parameter P5.4 compiles and lists all the high value resources, which can possibly come under attack, that are listed in P5.2 and P5.3. If there exists some resources that can be affected by the current mode of attack, then parameter P5.5 checks for it and scans for the decoy attack. Below is the pseudo code of this step.

```plaintext
P5.1 = FALSE;
Check for Contingency;
if (Contingency == Abort Mission) {P5.1 = TRUE;}
if (P5.1 != TRUE)
{
    P5.2 ← list of resources affected by the same mode of attack;
P5.3 ← list of attacks affecting the same resources;
P5.4 ← list of different attack modes, which can affect different resources other than the currently affected resources with higher risk value;
if (P5.2 != NULL && P5.3 != NULL)
{
    Check for possible decoy attacks;
    if (decoy attack detected)
    {
        P5.5 ← list possible decoy attacks;
    }
}
System_under_Attack = FALSE;
if (P1.1 != NULL || P1.2 != NULL) {
    if (P2.1 != NULL || P2.2 != NULL) {
        System_under_Attack = TRUE;
    }
}
Is_APT = FALSE;
if (System_under_Attack == TRUE) {
    if (P3.1 == TRUE || P5.1 == FALSE) {
        Is_APT = TRUE;
    }
    else if (P3.1 == TRUE || P5.5 == NULL) {
        Is_APT = TRUE;
    }
}
```

5 Risk Assessment

We use a simple method for risk assessment by defining risk for each parameter of the threat model developed in Section 3. The risk associated with each characteristic is the sum of the risks calculated for each parameter. The total risk for a system is the sum of all the risks for each characteristic in this simple assessment method. It shows that making a system more secure incurs more cost for the defender. The purpose of this risk assessment is to get a fair idea about the gap between the vulnerability and the security of a system in terms of the cost incurred. The cost discussed here is from the defender’s perspective and can be realized in tangible terms by using an appropriate unit of measurement.

5.1 Risk: Reconnaissance

Risk associated with reconnaissance is:

\[ R_{\text{recon}} = R_{p1.1} + R_{p1.2} \]

Where \( R_{p1.1} \) and \( R_{p1.2} \) are the risks associated with external and internal reconnaissance missions. The attackers carry out reconnaissance missions. If these missions are thwarted then most of the attacks can be prevented in their early stages. The associated risk with a parameter here is the amount of resources that needs to be spent to thwart the reconnaissance missions from taking place.

5.2 Risk: Exploit

Risk associated with exploitation of the vulnerabilities is:

\[ R_{\text{exploit}} = R_{p2.1} + R_{p2.2} + R_{p2.3.1} + R_{p2.3.2} + R_{p2.4} + R_{p2.5} \]

Vulnerabilities are the weaknesses in the system, which are exploited by the attackers to mount the attack. \( R_{p2.1} \) is the total cost to replace the vulnerable hardware, which are found to be either affected or infected. \( R_{p2.2} \) is the total amount of resources needed to reproduce or replace the software once the software vulnerabilities are exploited. \( R_{p2.3.1} \) and \( R_{p2.3.2} \) relate to the man-hour spent and the cost incurred per hour to completely exploit the hardware and the software vulnerabilities. This gives a fair idea of the resources to be invested in order to prevent the vulnerabilities being exploited. \( R_{p2.4} \) is the risk associated with the acquisition of the probable targets. \( R_{p2.5} \) is the risk associated with the affected resources.

5.3 Risk: Authorized Access

Risk associated with the malware gaining authorized access is:

\[ R_{\text{access}} = R_{p3.1} + R_{p3.2} + R_{\text{backdoor}} \]

\( R_{p3.2} \) is the risk associated with the number of backdoors being installed in the system by the malware to establish communication with the Command and Control (C&C) center. It is the cost incurred by the defender to find and prevent the backdoors from being installed in the system. \( R_{\text{backdoor}} \) is the risk associated with the acquisition of the backdoor implant tool. This is the cost incurred by the defender to acquire, and learn about the backdoor implant tool used to install the backdoors. If the defender is ready to incur this cost, then it might give the defender an opportunity to monitor the communication between the malware and the C&C centers.

5.4 Risk: Honeypots

Risk associated with interaction between the malware and the honeypots is:

\[ R_{\text{honeypots}} = R_{p4.1} + R_{p4.2} + R_{p4.3} \]

\( R_{p4.1} \) is the cost associated with the process of concealing the identity of the low interaction honeypots. \( R_{p4.2} \) is the cost associated with the process of camouflaging the identity of the high interaction honeypots. \( R_{p4.3} \) is the cost associated with the maintenance of honeypots which are interacting with the malware.

5.5 Risk: Contingency Plan

Risk associated with the attacker having a contingency plan is:

\[ R_{\text{contingency}} = R_{p5.1} + R_{p5.2} + R_{p5.3} + R_{p5.4} + R_{p5.5} \]

If the attacker is spooked during the attack process, then the attacker can often resort to a contingency plan. The first and the foremost contingency plan can be “Abort Mission” and the risk associated with it is denoted by \( R_{p5.1} \). It reflects the cost incurred by the defender if the attack mission is aborted before the defender gets to identify the family and the type of attack carried out. But this risk can have negative value if the defender acquires the knowledge about the identity and the family of attack being mounted. This comes from the fact that once the defender acquires the knowledge about the family and the type of attack, then for any future attack by the same family of attackers, the defender could prepare itself faster than before. But if the contingency plan is something other than aborting the mission, then the remaining parameters take care of the different forms of the contingency plan. \( R_{p5.2} \) is the risk associated with different probable resources being affected by the same mode of attack. \( R_{p5.3} \) is the risk associated with different probable attacks that can again affect the same resources as the current attack. \( R_{p5.4} \) is the risk associated with different resources, which could be of higher value than the current one. \( R_{p5.5} \) is the associated risk if the attacker triggers a decoy attack.

5.6 Risk: Total

The total risk to the system is the sum total of all the risk calculated from each characteristic of the model. This gives an estimate of the total risk to the system or the organization in some tangible terms. Total risk from an APT is:

\[ R_{\text{total}} = R_{\text{recon}} + R_{\text{exploit}} + R_{\text{access}} + R_{\text{honeypots}} + R_{\text{contingency}} \]

6 Analysis of the Model’s Utility

The APT identification algorithm in Section IV not only puts forward a detection mechanism, but also paves way for the defender to come up with a defense model by providing
the fundamental details of the threat. Unlike traditional models, this model takes into consideration the attacker skills and the fact that the attackers could come with a contingency plan. For building a defense mechanism against an APT, one needs to know all the features of the APT. Traditionally a threat is classified as an APT once the attack has occurred and it matched certain observations, but does not account for attack follow-ups or for the possibility of a contingency plan. But the presented algorithm is an online version that works as the attack is unfolding. It can be used to develop preemptive measures to tackle any ongoing attack by giving the defender ample time and information.

The presented security model is a holistic one in the sense that it consists of a threat identification algorithm and the risk model. The model itself defines five major characteristics in order to determine whether it is an APT or not. It begins with the detection of the reconnaissance processes of the malware. The algorithm detects and lists the reconnaissance processes both external as well as the internal reconnaissance processes. The risk model calculates the risk associated with the reconnaissance processes. Following that, the algorithm looks for the hardware and software vulnerabilities, and the time to exploit the vulnerabilities. The risk model calculates the risk associated with the aforementioned vulnerabilities and also provides the defender with an estimate on the time it requires to gain authorized access to the system by exploiting the vulnerabilities. Given the type of vulnerabilities being exploited, the model gives guidance on the probable targets and affected resources. The algorithm then checks if the attacker has gained authorized access in the system and installed backdoors to communicate with the Command and Control (C&C) centers. At this point, if the malware portrays all the characteristics mentioned, then the threat is probably an APT. The defender can start preparing for a defense against an APT. If the system has honeypots then the activities of the malware can be logged by the honeypots and the information can be conveyed to the defender for building the defense mechanism. The final characteristic that the model investigates is the presence of a contingency plan. Given the characteristics portrayed by the threat until this point, the algorithm proceeds to list the probable resources that can come under attack from the same and/or different modes of attack. It also evaluates the probable targets to figure out if the current attack is a decoy attack or a parallel attack. The risk model calculates the financial losses that can be incurred from the contingency characteristics including a decoy attack. This helps the defender to prepare not only for the current attack but also for some different follow-up attacks or a decoy attack. It also helps the defender to be alert against attack from the same family of the attackers, or for any future attacks, which are similar in nature.

We now look at some real world, sophisticated attacks to determine the efficacy of our APT Model. The following attack campaigns are advanced enough to mount sophisticated attacks and persistent enough in their pursuit of the end goal, and according to our algorithm, are Advanced Persistent Threats (APTs)

6.1 WannaCry

WannaCry ransomware [8] looks for exploitable vulnerabilities to infiltrate a system (in reality, it exploited the TCP port 445 or Server Message Block (SMBv1) vulnerability of Windows operating system to carry out the attack [10]). Once inside the system, the malware uses a backdoor implant tool (e.g., DoublePulsar) to install a backdoor in order to communicate with its Command and Control Center [11] [19] [20]. The malware then gathers information about the environment where it is being executed [6]. The malware then makes a query to establish a connection to the domain \texttt{hxxp://www[.]iuqerfsodp9ifjaposdfjhgosuriuforiaewrwegwea[.]com}. This is done to verify that the environment in which it is being executed is a sand-box or a real system. If the domain does not exist, the connection fails and it carries out the attack. The goal of this malware is to encrypt the system with a public key, demand a ransom, and when the payment is received, release the private key to decrypt the data.

Our algorithm recognizes WannaCry as an APT. The WannaCry worm looked for DoublePulsar backdoor implant tool and Eternal Blue exploits of the Server Message Block (SMB) for infiltration [8]. If a flag is raised for a process querying for the vulnerabilities one can list the existing vulnerabilities, and start eliminating the vulnerabilities on the non-infected systems. At this point it is clear the system is under attack. So all possible resources and probable targets should be under the scanner. Risk value for all the parameters can be calculated. The attacks started on May 12, 2017 and by May 13, 2017 the kill-switch was accidentally discovered [21] which was the domain the malware queried to, was registered deceiving the malware into believing that it is being run in a controlled environment. On May 19, 2017, the attackers using Mirai botnet launched a DDoS attack on the server hosting the domain. On May 22, 2017, Marcus Hutchins of MalwareTech protected the site by moving it to a cached version of the site [9]. The presence of a contingency plan on the part of the attackers with targeted attacks exploiting a certain type of vulnerability, the WannaCry campaign can be confidently designated as an APT. It is an advanced threat as it is using zero-day vulnerabilities and persistent enough to have a contingency plan to execute the attacks. If our model had been applied to study the WannaCry campaign, then the attacks caused by NotPetya could have been thwarted, which we would discuss in the next sub-section.

6.2 Petya and NotPetya

Petya was a ransomware attack that exploited the vulnerabilities in the master boot record and infected mainly the Windows based systems. The attacks were discovered in 2016. The targets mainly were the file system tables of the hard drive [22] [23]. If the model had been used to analyze the attacks then the probable targets, vulnerabilities and the resources exploited could have been recorded and stored. In May 2017 WannaCry ransomware mounted attacks globally affecting considerable number of systems. WannaCry exploited the EternalBlue vulnerability of the SMBv1 of Windows [8]. If the model had been used at the time of the
attack, then again the list of vulnerabilities and targets could have been created and stored. In June 2017, NotPetya mounted attacks using the same vulnerabilities used by WannaCry in May 2017 and attacking similar targets as in the case of the attacks mounted by Petya in 2016. If the model, presented in this paper, had been used to analyze the previous attacks by Petya and WannaCry, then the defender could have had ample information regarding the vulnerabilities and the targets. Therefore, attacks by NotPetya could have been easily mitigated.

7 Conclusions

The paper provides a holistic approach to identify an APT, record the vulnerabilities, and list the resources and the probable targets, so as to design a defense mechanism. The paper also gives an algorithm to carry out the aforementioned actions, as and when the attack is unfolding. This gives the defender valuable time and information about the attack. Finally, the model also provides a risk assessment method, which gives a financial estimate of the potential losses that an organization can incur at the face of an attack. The model is useful to analyze attacks mounted by APTs and also by the attacks from non-APT sources. The model provides scalability of attack analysis, as it takes care of the resources under attack and figures out the probable targets, and also takes into consideration the fact that the attackers may have a contingency plan or a decoy attack if the ongoing attack is discovered by the defender. The model also imparts flexibility in terms of analysis by considering both APTs and non-APTs. There are several future directions for this research. The risk analysis presented is very simplified. A more sophisticated risk assessment is possible by considering the various detection methods and probability of success of the attack penetration at the various stages. We also have plans to investigate several other real world attacks to further generalize/refine our threat model.

8 Acknowledgment

This research is supported in part by National Science Foundation Grant No. DGE-1241709.

9 References