Changing the game: The art of deception against sophisticated attackers

Nikos Virvilis, Oscar Serrano

Abstract:
The number and complexity of cyber-attacks has been increasing steadily in the last years. Adversaries are targeting the communication and information systems (CIS) of government, military and industrial organizations, as well as critical infrastructures and are willing to spend large amounts of money, time and expertise for reaching their goals. In addition, we have recently witnessed sophisticated insider attacks that have resulted in the exfiltration of highly classified information to the public. Traditional security solutions have failed repeatedly to mitigate these threats in time. In order to defend against such sophisticated adversaries we need to redesign our defenses, aiming more on detective instead of preventive technologies. In this paper, we are focusing on Advanced Persistent Threats (APT) and malicious insiders, highlighting the common characteristics of these two groups. In addition, we are proposing the use of multiple deception techniques, which can be used to protect both the external and internal resources of an organization and significantly increase the possibility of early detection of such sophisticated attackers.

Keywords: Advanced Persistent Threat, Deception, Insiders, Honeypot, Honeynet, Honeytokens

1. INTRODUCTION

Since 2006, there have been a large number of advanced, well-orchestrated attacks against industry, military and state infrastructures. The main goal of most of these attacks is the exfiltration of large amounts of data, for example: in 2006, China was accused of downloading 10 to 20 Terabytes of data from the US NIPRNet Military Network [1] and in 2008, a USB drive left in the parking of a Department of a US Defense facility in middle East was used to infect a laptop computer connected to the United States Central Command, resulting in the exfiltration of sensitive information [2]. In 2010 “Operation Aurora” targeted more than 20 organizations including Google, Adobe and Symantec and US defense contractors [3]. Furthermore, we have witnessed attacks which are focused on causing physical destruction [4]. While it is believed that these attacks were perpetrated by different threat actors, they share certain common aspects and some of them have been categorized as Advanced Persistent Threats. The term “advanced persistent threat” (APT), a term coined by the U.S. Air Force in 2006[1], has been loosely used to highlight the commonality between these attacks. The definition of APT by the NIST [5] is the following: “... An adversary that possesses sophisticated levels of expertise and significant resources which allow it to create opportunities to achieve its objectives by using multiple attack vectors (e.g., cyber, physical, and deception). These objectives typically include establishing/extending footholds within the information technology infrastructure of the targeted organizations for purposes of exfiltrating information, undermining or impeding critical aspects of a mission, program, or organization; or positioning itself to carry out these objectives in the future. The advanced persistent threat: (i) pursues its objectives repeatedly over an extended period of time; (ii) adapts to defenders’ efforts to resist it; and (iii) is determined to maintain the level of interaction needed to execute its objectives.”

In addition, organizations face the always present threat of malicious insiders, a clear example of which is the recent case of Edward Snowden, who downloaded 50,000 to 200,000 classified documents belonging to the US National Security Agency [6]. This incident arose short before the conviction to 35 years in prison to Bradley Manning, after the largest data leak in US history [7].

The ability of current security solutions to address such attackers has been questioned openly [8][9], with the authors stating that prevention techniques (e.g. NIPS, antivirus products) - and especially the ones focused on signatures - will never be able to address sophisticated attacks.

---

1 Initially used as a generic term to describe intrusions without disclosing the classified threat name [32].
As the shortcomings of signature-based detection are well accepted, the research community has focused on the use of anomaly-based detection systems. However, the effectiveness of these systems has also been challenged: Sommer describes at [10] how anomaly detection is flawed on its basic assumptions: Research relies on the belief that anomaly detection is suitable to find novel attacks, however machine learning techniques are best suited to find similar events to something previously seen. APT on the other hand, tend to use unique attack vectors and custom built tools tuned for the particular target, making detection very challenging.

Our contribution in this paper is: a. the comparison between APT and malicious insiders, highlighting the common characteristics between these two attacker groups and suggesting that malicious insiders should be considered as a subcategory of APT and b. proposal of multiple deception techniques, such as the use of social network avatars, fake (honey token) DNS records and HTML comments - which to the best of our knowledge have not been proposed before - that can significantly increase the likelihood of early detection in every phase of the attack’s lifecycle.

The remainder of the paper is structured as follows: In section 2 we present the related work. Section 3 focuses on the similarities between APTs and insiders, as we believe that both can be treated equally for the purpose of detecting sophisticated attacks. In section 4, we propose a number of deception techniques for protecting both the Internet facing and the internal assets of an organization, following with our conclusion and further work in Section 5.

2. RELATED WORK

Decoys are a popular strategy long used in warfare, that played an important role during the Second World War [11] and the cold war [12]. Decoys are also an integral part of Electronic warfare strategies [13], however they are rarely used on the Cyber domain. The first general reference of Cyber decoys is attributed to Cliff Stoll and is detailed in his novel the Cuckoos Egg [14]. Years later, Spitzer describes mechanisms for the detection of insider attacks using honeypots [15] and honeytokens, which share similar characteristics with honeyfiles, as described by Yuill [16][17].

Similarly, honeypots [18][19] have been proposed for attack detection [20][21], including the detection and analysis of botnets/worms, while honeynets [22] have been proposed as an effective means for the classification of network traffic and the detection of malicious users on Wifi networks [23].

Honeyfiles including beacon signaling are discussed by Bowen [24], who propose an architecture for monitoring multiple system events, including user interactions with a set of previously marked honeyfiles. Similar work was pursued by Whitmam [25], introducing canary files which have similar characteristics with honeyfiles. Most of the published work, concentrates on the creation and distribution of perfect believable honeyfiles [26], which contain certain properties that make them indistinguishable from real files to malicious users and at the same time are enticing enough to attract attention. Finally, researchers have also proposed embedding code to legitimate documents, which gets automatically executed when the file(s) are opened and connects back to a monitoring server [27], as a means of detecting unauthorized access.

To the best of our knowledge, there has been no research on the use of deception techniques for the detection of Advanced Persistent Threats (APT).

3. ADVANCED PERSISTENT THREAT AND INSIDERS

The definition of a malicious insider based on Silowash et al. [28] is the following:

A current or former employee, contractor, or business partner who meets the following criteria:
- has or had authorized access to organization’s network, system, or data
- has intentionally exceeded or intentionally used that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization’s information or information systems

The motives of insiders vary, from financial or revenge to ethical and political [29].
APTs and malicious insiders share unique characteristics which significantly differentiate them from traditional (e.g. opportunistic) attackers:

- Their attacks require detailed planning [30] and are spread over a wide period of time, in an effort to evade detection. Insiders can have a potential advantage over APTs while planning for their attack, as they may be aware of the existing security controls. This is very likely if they hold a privileged role (e.g. a malicious administrator is expected to have a clear view of the deployed security mechanism and potentially has the access rights to control them, compared to a more limited user [30]). Nevertheless, experience has shown that APT have also managed to reach their goals, evading detection, without prior knowledge of the infrastructure [3].

- Both groups are willing to spend significant time and explore all possible attack paths for reaching their goals, including social engineering and deception [30]. APTs groups tend to have teams of highly skilled individuals with access to significant resources (financial, technical, intelligence). Equally, malicious insiders although they work mostly alone, as we witnessed in the case of Manning and Snowden, have significant technical skills.

- Both are interested in keeping access to the penetrated infrastructure and continue the exfiltration of data for as long as possible.

The main difference is that malicious insiders have by definition authorized access to the infrastructure and potentially even to the servers storing the sensitive information (e.g. fileservers, database servers), while APTs need to gain unauthorized access by any means in their disposal.

APT and insider threats are currently considered as two different threat groups. However, as we have already witnessed examples of APT groups blackmailing or bribing an insider to perform a malicious action on behalf of them [31], we strongly believe that malicious insiders should be regarded as a subset of Advanced Persistent Threats.

Robust models have been proposed in the past for the detection of insider threats [32], however they assume that the malicious insider(s) will perform the whole attack lifecycle on their own (e.g. information gathering, exploitation, exfiltration). Yet, in Stuxnet’s case [31], a malicious insider has only been used to deliver the payload, while the rest of the exploitation was performed in an automated way. Such modus operandi, which combines APT with the insider elements, possesses a significant challenge for insider threat detection models.

Taking into consideration the significant resources available to APT groups [33], we should expect similar attacks in the future and thus, we strongly believe that further research is necessary to augment the detection capabilities of such models against combined insider – APT attacks.

4. DECEPTION TECHNIQUES

The APT attack life cycle [34] consists of multiple stages: Attack preparation and initial compromise, establishing foothold, escalation of privileges, internal reconnaissance, exploitation of systems and exfiltration of data. For the sake of simplicity, in this paper we will group these stages into two generic phases: a. attack preparation - information gathering and b. exploitation and data exfiltration. For each one of these, we are proposing deception techniques, which can significantly increase the likelihood of attack detection.

A. Phase 1: Attack preparation - Information gathering

The initial step of an APT attack is the preparation phase, where perpetrators gather as much information as possible for their target. Identification of Operating System, third party software and publicly accessible services (e.g. Web servers, mail servers) of the organization is crucial for planning a successful attack. Additionally, any information related to the security solutions in use (IDS/IPS, endpoint protection, data leakage prevention) is significant for the attackers, as it allows them to test their tools and techniques beforehand.

An additional step of the preparation phase, is the collection of information about the employees, their position in the organization, their skills and their connections with other employees. With such information in hand,
APTs can create highly targeted spear-phishing campaigns. For example, assuming that attackers have identified a person working in the human resources department as well as her supervisor, they can send a spoofed email from the email address of the supervisor to the user, asking her to review an attached file (e.g. a CV). The CV can be a malicious Word or PDF file, which when opened, will execute the attacker’s payload. The fact that the email originates from a person known to the victim significantly increases the likelihood of accepting it as low risk/legitimate.

In order to address this first phase of the attack lifecycle, we are proposing the following deception techniques:

   a. **DNS honey tokens**

As attackers will try to identify internet facing systems/services belonging to the organization, defenders can deploy honeypots which are spread over the unused public IP range of the organization. Based on the fact that these systems won’t be publicly listed (e.g. returned as part of a search query, linked from the organization’s web site), any connection attempt could be due to: a. a user error (mistyping an IP address), b. an automated attack e.g. a worm scanning randomly the IP address space to find vulnerable hosts to compromise or c. an attacker, trying to identify all publicly accessible systems and services of the organization.

This approach is likely to cause significant noise due to the vast number of automated attacks on the internet [36]. In addition, it can be challenging to differentiate between an automated, non-targeted attack and targeted one. Our proposal is a simpler to implement technique, based on fake DNS records (honey tokens), that will significantly limit the number of false positives:

Attackers are very likely to bruteforce or attempt a zone transfer [37] on the organization’s DNS server(s), trying to identify interesting resources (e.g. subdomains, servers), as part of their information gathering process. By creating a handful of fake DNS records on the authoritative DNS server(s) of the organization and configuring them to alert when these specific records are requested, the defenders are able to get an early warning on DNS-related information gathering attempts against their infrastructure.

   b. **Web servers**

The public web servers of the organization is another fruitful source of information for the attackers thus, we propose three types of honey tokens which can help detect malicious web site visitors:

- Adding fake entries in robots.txt
- Use of invisible links
- Inclusion of honey-HTML comments

Robots.txt [38], is a simple text file located on the root folder of the web server, which legitimate bots (e.g. Google bot) are parsing to identify which folders on the webserver they should not access and indexed. However, it is one of the first places that attackers (and automated web vulnerability scanning tools) will look for potentially sensitive directories. By including non-existing directories such as “/admin” or “/login” in this file and monitoring for access requests to these locations, administrators can be alerted for mal-intended visitors.

Similarly, we propose the inclusion of invisible links (e.g. white links on white font) at random parts of the web site(s), pointing to non-existing (but interesting from the attacker’s perspective) resources. Although these links will be invisible to the legitimate visitors, they will be detected by crawling tools that attackers are likely to use. Any request for these fake urls, should raise an alert.

A final deception mechanism, particularly useful for web sites that support authentication, is the inclusion of fake accounts in html comments. Legitimate users have no need to review the source code of a web page however attackers frequently do, trying to identify vulnerabilities. The inclusion of a comment like the following in the html source of login page, it very likely to tempt the attacker to use it:

```
<!--test account: admin, pass: password123. Please remove at the end of development!-->
```
Once more, any attempt to login with these credentials as a clear indication of malicious activity.

c. Social network avatars

As already stated, social networks are an invaluable source of information for attackers. In order to identify malicious actions, we propose the creation of Avatars (fake personas) on the major social networks. It is important that these avatars look as realistic as possible, having connections with people from both inside and outside the organization and be given positions that are likely to be of interest for the attackers (e.g. HR department, financial department, developer etc). In addition, such avatars should have real - but very closely monitored - accounts in the organization (e.g. active directory account), as well as valid email addresses. Any interaction with the avatars should be regularly monitored (friend requests, private messages, attachments etc).

Although we have to assume that a legitimate person interested in applying for a position in the organization, may contact the Human Resources Avatar (false positive), there are a number of actions which should be considered as suspicious, such as any communication between the employees of the organization and the avatar (could be an indication that the employee’s account has been compromised), or any log in attempts using the avatar’s account(s).

B. Phase 2: exploitation and data exfiltration

The second step of the APT lifecycle is the exploitation of the target. The attackers, after gaining access to the internal network (e.g. 0-day vulnerability, social engineering, spear-phishing attack), will start their information gathering process trying to identify: a. systems that they can compromise to be used as alternative access points to the network (in case the initial one(s) are detected and quarantined) and b. systems which may contain the information they are after or they can help them access that information.

In order to address this phase of the attack we propose the following deception techniques:

Network layer defenses

In a medium to large organization where hundreds or even thousands systems are active, identifying where the targeted information is located, is not a trivial task. Attackers will need to explore the network, hop between networks and exploit multiple systems. Use of darknets/honeynets can be invaluable in detecting such actions, as eventually attackers will access them, raising an immediate alert:

Darknets

A darknet is portion of routed, unallocated IP space, where no workstations/servers or other network devices are located [39]. Access to such regions of the network can be due to a legitimate mistake (e.g. a user mistyping an IP address) however multiple connection attempts should be considered suspicious. Monitoring such segments for connection attempts, can be an effective and easy to deploy mechanism, however it is not guaranteed that attackers will actually access these parts of the network.

Honeynets

Honeynets [22] are used for monitoring a larger and/or more diverse networks in which one honeypot may not be sufficient. Defenders can use honeynets to create multiple fake systems in the same IP ranges as legitimate systems/servers. An attacker who gains access to the specific network segment, is very likely to access these systems among with the real ones. Any interaction with those systems should be very closely monitored as it is a very likely indication of an active attack.

Application layer defenses

The same techniques used for detecting malicious actions on external web servers can be used for protecting internal ones. Furthermore, as the majority of organizations make use of database and files servers on their internal networks, we propose the following deception techniques for the detection of malicious actions against those servers:
**Database Servers**

Use of honey tokens (bait-records) in the databases, can be used to highlight malicious activity. For example, a number of fake patient records (with fake patient names) can be introduced in a hospital’s patient database. Any access to these records should be considered as highly suspicious. Nevertheless, database auditing has to be enabled for logging the queries, which comes with a non-negligible performance hit.

**Honeyfiles**

As described in related work, a number of strategies for creating decoy (honey files) have been proposed, focusing either on the generation of perfectly believable decoys or the modification of legitimate files to include some alerting functionality. Although generation of perfectly believable decoys has been challenged, use of legitimate files has non negligible false positives [27].

We propose a combination of file system auditing and the generation of honey files with potentially interesting content for the attackers (e.g. passwords.docx, new_investements.pdf, etc). These files should be spread across the fileservers of the organization and/or even workstations, however the later will increase the false positive alerts [27].

A number of detection techniques can be implemented, including:

- File system auditing [40], which will log any access attempt to these files
- Inclusion of code which when executed, will report back to a monitoring server. This can be achieved by using JavaScript for pdf files and remote images, that are rendered when the document is opened for Microsoft office files [41] and,
- Inclusion of bait information, such as fake credentials that attackers may try to use.

**Honey accounts**

Creating bait accounts (such as accounts for the avatars), is also an additional way of detecting attackers, as any interaction (e.g. login attempts) with these accounts is a clear indication of an active attack. This could be combined with the aforementioned example of placing bait files on fileservers, where a file with fake credentials (usernames and passwords) could be created. An attacker who has gained access to the file, is very likely to try to use these accounts to gain further access to the network and as a result she will immediately raise an alert.

**Evaluation**

Preventive techniques will eventually fail against sophisticated attackers [9], thus it is critical to switch our focus on detective measures. Use of deception techniques like the proposed ones, will significantly increase the possibility of detecting attacks early in the attack life cycle, allowing the defenders to mitigate the threat before the attackers reach their goals.

Although the effectiveness of such measures against insiders can be challenged, based on the fact that insiders are likely to be aware of their use and may try to evade them, we believe that combining a wide number of deception techniques will make evasion very challenging, assuming they are not the ones that have deployed them.

While the aforementioned deception techniques, will rarely present false positives, we recommend to integrate them into a misuse detection system [35] incorporating a plethora of additional data sources, such as HR databases (e.g. user data, leave data), access rights matrices, netflow data, etc., as this would further increase the reliability of the system and reduce the number of false positives.

**5. CONCLUSION AND FUTURE WORK**

In this paper we highlighted the fact that insider threats and APT share multiple common characteristics and should be considered as a single threat. Furthermore, we have highlighted the inability of current security solutions to effectively address sophisticated attackers and we proposed the use of deception as a potential
solution to this multidimensional problem. Several deception techniques have been presented, which can be used to increase the possibility of early detection at any stage of the attack life cycle. Furthermore, such techniques can be combined with traditional collection and correlation systems to further increase visibility.

Finally, our future work focuses on the improvement of the existing insider threat detection models with the introduction of deception techniques, in order to address the very challenging combination of APT and insider attacks.

REFERENCES


A. Mandiant, *Exposing One of China’s Cyber Espionage Units (Feb. 2013)*. .


