Offensive Threat Modeling for Attackers

Turning Threat Modeling on its Head

Black Hat Europe 2012

Rafal Los

Shane MacDougall

PREAMBLE

Modern threat modeling is a defensive response to understanding a threat so as to prepare yourself, your network, and your assets from attackers. This paper describes how threat modeling can also be used as an offensive weapon. While traditional models look at the attacker, the asset and the system – offensive threat modeling looks back at the defender to understand his tactics and expose weaknesses that can be leveraged for successful exploitation. With our approach we will utilize a methodology we describe as "The Five P's: People, Points, Position, Posture, Pwn and Poll."

Some of the methodologies outlined in this paper are illegal and unethical. The authors do not endorse the undertaking of any criminal activity; the purpose of this paper is to identify how attackers (be they white hat or black hat) can use the offensive threat model to effectively launch and manage attacks.

OFFENSIVE THREAT MODELING: A HIGH LEVEL OVERVIEW

The main difference in offensive threat modeling is to classify the defenders as the asset, rather than just the target system. To achieve this, we break the threat model into five distinct tasks.

In fact there is a sixth "P", which is actually our first step, and is important, however, it usually is the same for most attacks, so we omit it from the "Five P's."

- *Purpose: Identify Objective (more of a formality)*
- People: Identify Assets
- Points: Decompose assets into various "points" of opportunity
- Posture: Identify the posture of the points (can they be compromised)
- Pwn: Compromise the asset
- Poll: Monitor and update the asset list

The Omitted P: "Purpose" - Identify The Objective

In many, if not most cases, our objective is not to just compromise a specific application or database, but rather the entire operation. We seek to infiltrate the enterprise from many disparate and diverse avenues in a pervasive and persistent manner, such that if one compromise is discovered that other compromised assets will not be affected, or ideally discovered since they are different in nature. If your goal is to target a specific application or asset, this is an important step, since it will assist you in identifying when the rest of your "P's" go "off the reservation" or begin to move away from the end goal.

The First P: "People" - Identify The Target Assets

Again, in offensive threat modeling we are not looking at products or systems as assets, but rather the defenders and the defense capabilities. First create a HPTL (High Payoff Target List) – those assets that give the biggest bang for the buck when compromised. Security personnel and Senior Executives usually populate this list. Then create a tangential list of targets, that while might not necessarily hold the keys to the kingdom, can allow access to the enterprise with access to some proprietary information or access, be it logical or physical. Sales personnel, support staff, and vendors most often appear in this list, although that is not an exhaustive list by any means. Finally, create a list of targets of opportunity, the "low hanging fruit" of the enterprise.

In this phase we also map out defensive capabilities, such as security infrastructure like IDS, firewalls, physical plant defenses (CCTV, proxcards, guards, etc).

The Second P: "Points" - Decompose Assets Into Points of Attack

We break down each of the assets into base components to identify what parts can be readily compromised. Family members, hobbies, conferences, behavioral analysis, psychological/sociological profiling, sentiment analysis and other areas of the human asset are targeted. For physical assets we decompose them in a similar manner to identify

The Third P: "Posture" - Identify The Posture of Assets

Now that we have broken each asset down into individual components, we assess the state or posture of each component: is it ready to be broken? It is important to identify schedules for postures of the assets. Are firewalls fixed at a specific period of time? Is there a release schedule? When are employee performance reviews performed? Do employees attend specific conferences on a regular basis (hello BlackHat!).

The Fourth P: "Pwn" - Compromise The Asset

Hax0r those assets! Utilizing the discovered points of attack, leverage the known weaknesses to compromise the various assets whose posture lends them to attack. These attacks can range

from logical attacks to social engineering to physical on-site attacks. The attacks can also be in the form of blackmail, bribery, or other incentivization.

The Fifth P: "Poll" - Monitor and Update The Asset List

The attacker must continuously monitor and update the asset list to identify if any of their states have changed, gauge their effectiveness, and perform a cost-benefit analysis on underperforming assets. Lost assets need to have damage assessments performed to ensure no attack leakage has occurred and to identify possible replacements.

OFFENSIVE THREAT MODELING: A DETAILED LOOK AT THE FIVE P'S

I - Modeling The "People"

Identifying the defender is accomplished through a variety of traditional and non-traditional methods of footprinting a company's network, organization, and physical plant. Scanning corporate websites, press releases, conference presentations, dialing through voicemail and phone directories, and social engineering are all commonly used methods to begin to map out a company's org chart, as well as identifying secondary attack vectors such as suppliers/vendors and customers.

Social media sites are an obvious resource for the attacker. Twitter, Facebook and LinkedIn are the mother lodes of sites which serve up an abundance of data. Less commonly searched sites such as flickr.com can yield tremendous amounts of actionable intelligence. Company events where ID cards or badges are visible not only give you employee names, employee numbers, and badge layouts (which can be replicated and used for on-site attacks), but also give the attacker reference able events that they can use in a social engineering attack. Address, phone, email and other background information can be easily cultivated from sites such as beenverified.com, spokeo.com, emailfinder.com, and many other sites. For a comprehensive list of sites and their specific functions, refer to Appendix A.

To identify hierarchy within an organization without identifiers such as titles or an organization chart is still possible. Recent research by Eric Gilbert at Georgia Institute of Technology has identified certain phrases used in email and electronic messaging have a very high correlation to workplace hierarchy [a]. For example phrases like "thought you would" are strong indicators the recipient outranks the sender, while "let's discuss" implies the opposite. The data set of 7,222 phrases has been released online [b] along with the associated weightings to assist the reader in "rolling their own" solutions to determine the hierarchical status of the sender of messages.

It is also important that the attacker identify both clients and vendors/suppliers, since they are often the path of least resistance to a successful penetration of an organization. Vendors will routinely use clients as case studies.

After all the potential assets have been identified, the attacker needs to prioritize them, and place them in the appropriate list. The High Payoff Target List (HPTL) is for those assets that give the biggest ROI when compromised. The Secondary/Tangential Target List contains those targets, that while might not necessarily yield a huge ROI, still have the potential, if compromised, to get the attacker access to the enterprise with access to some proprietary information or access, be it logical or physical. The remaining assets go on the Targets of Opportunity/Low Priority List, and are generally the "low hanging fruit" of the enterprise which will yield the attacker either some information or limited access, and are mainly used for either throwaway operations such as one-time phishing attacks or onsite access attacks.

Physical plant surveillance is also helpful in identifying the company's security posture. Is CCTV in place, are the cameras fixed or PTZ? Do employees wear badges, and if so what can you glean from them? Are the employee names on them? Do they display a barcode or photo? Job title? Do they use biometrics or mantraps at their entrances? Are there guards on duty? Are couriers received at the front desk or are they sent to a shipping area? Each of these pieces of information help build the profile of the target.

II - Modeling "Points of Attack" - Decomposing Assets Into Points of Attack

Effective threat modeling is predicated on breaking every potential asset into as many possible elements as possible, and then assessing each element for weaknesses that can be exploited. In the case of employees at a target company, this can include identifying family members or roommates who may share a home network with the employee, allowing indirect targeting of the asset through targeted spearphishing. These people can also be targeted via social engineering to elicit information that can be leveraged in an attack, even to the point of granting physical access to the home.

Identifying hobbies, routines, favorite hangouts, religious preferences, all can be very useful in identifying pretexts to be used in social engineering attacks, or to track an individual physically. Identifying behavioral issues such as substance abuse, gambling, or sexual activities such as the use of prostitutes or extramarital affairs can be extremely helpful to the attacker, since it opens up the possibility for blackmail or extortion.

Psychological profiling from social networking presence is an ever growing field, with sites such as tweetpsych.com, and automated user profiling has long been established. [e] Recent research from The Online Privacy Foundation also has shown some correlation to psychological profiles and social media presence, although they argue the correlations are not as strong as previously believed. [f] No matter the psychological profile one discerns from social media, proclivity to certain activities and political beliefs are often easy to ascertain, if not directly from the target user, then from their common online associates. For example, a user with a Guy Fawkes mask as their Twitter avatar, who follows command and control accounts from Anonymous, can probably be profiled as a follower of Anonymous, and is more than likely more susceptible to being successfully recruited into a "hactivist" operation against a bank than is a user who follows Andrew Breitbart and the Wall Street Journal.

III - “Posture” - Identifying The Posture of Assets

Identifying as many users most at risk of compromising is critical to an ongoing effective penetration of an organization. With the explosion in social networks over the past few years, this has become much easier for attackers.

Identifying the general employee contentment within an organization is a critical first step. Job boards such as Glassdoor.com, insidebuzz.com, and jobitorial.com are all excellent sources of gauging whether or not there is negative sentiment against the employer, and whether it is restricted to particular groups, or widespread throughout the organization. They are also excellent resources for attackers to gather company lingo such as facility and project nicknames, which can be invaluable when launching social engineering attacks.

Identifying users with open social media profiles who are liberal with their information leakage allows an attacker to rapidly create a list of “low hanging fruit” to target with basic social engineering attacks.

If an attacker is able to grab emails, chat logs, or message board information from a compromised resource, they can then determine users within an organization that are disgruntled and possibly open to being compromised by bribes or being social engineered into performing a revenge attack.

To automate analysis of messages for sentiment an attacker can utilize several tools. Lymbix provides its Tonecheck plugin for Outlook/Gmail/Lotus Notes (and other coming up) that does basic and some extended analysis. The results, however are less than stellar, with the crux of the software appearing to be highly dependent on extreme emotional words such as hate, despise, love, etc. Unfortunately it doesn’t always properly determine the context of these key words. For example, “I’d love to burn this company down” actually scores as a neutral sentiment. A slightly better tool is Muse [c][d] from Stanford’s Sudheendra Hangal and Monica Lam which allows analysis of some chat, mbox format inboxs, and mailing lists. Its accuracy also seems to be a bit better than the Lymbix products.

Ultimately, we have found that manually searching through mail archives using keywords yields a much more effective accuracy rate. A list of keywords that we use is found in Appendix B. While it obviously takes more time to parse results, it doesn’t appear that current sentiment analysis is good enough to find nuanced or outright disgruntled emails unless specific words are triggered, and context is often missed. That said, an attacker should always do a first run through the tools to identify high indicators of negative sentiment.

Some readers may point out that if we have access to emails from the client, chances are we have already gained access to systems within the company, and thus don’t really need to be doing this exercise. First, we may have gained access to an individual resource such as a laptop

which does not have access to the desired systems. In this case we can use the information gathered to further our attempts to escalate privilege. Secondly, the goal of an "APT" attack is not simply to gain root access, but rather to create many different vectors of infiltration. Sentiment analysis allows us to identify other targets within the enterprise that we can determine are good targets for exploiting their disenchantment with the organization.

Identifying when targets are in a broken or vulnerable state is essential to a successful attack. On a human level, we can use the conference experience as an example. If a target routinely attends a conference, an attacker can plan two methods of attack: local physical plant/social engineering which rely on the target not being around (i.e. dropping the unavailable target as the reference), or a physical attack on the target at the conference. Said physical attack can be gaining access to their physical machine, sexual honeytrap, social engineering, etc. If the target is speaking at a conference that is even better, since it is often possible to social engineer information about internal initiatives, evolving research, etc.

From a logical security point of view, identifying maintenance windows can be gleaned either from social engineering or monitoring website behavior. Even if the window of opportunity is just a matter of minutes, identifying the window can be make the difference between a successful attack and failure.

Similarly, delivery of new systems, construction projects at the corporate facilities, and mass hirings or layoffs also create windows of opportunity for an attacker to exploit. These all can be identified easily from websites, physical surveillance or social engineering.

IV- P is for "Pwn" - Compromising The Asset

Upon building the target lists, the attacker can then target specific individuals. Sending trojaned antivirus software branded as coming from the organization as an effort to help secure employee's home machines are particularly effective, as are mailing branded USB thumb drives as tokens of appreciation for the company. Of course the drives are preloaded with nasty malware.

Other options include attacking the employee's home network, especially if the user is utilizing wireless networking, and even physical break-ins to the employee's house in order to install malware, keyloggers, or just clone drives.

Realtime tracking is often enabled by the target personnel themselves, via social media such as FourSquare. If we know the IT team from our target is going to be out at a bar, the attacker can ingratiate themselves with the team and buy them several rounds of alcoholic beverages with their newfound "friends". Later in the night when the attacker launches an attack on the network, he can have a higher degree of confidence that immediate response on the part of the IT team may not be forthcoming. Additionally, identifying the location of the team while out drinking also gives the attackers the opportunity to get physical access to laptops and other devices while the targets are otherwise occupied.

Other avenues of attack involve identifying employees at risk of being extorted/blackmailed. This vector often involves luck, although with the proper resources and targeting can be very effective. By utilizing websites that specialize in illegal activities or "alternative lifestyles", an attacker can sometimes identify potential victims within an organization with relative ease. This is because these sites such as escort sites usually allow for "providers" to register with minimal hassle (as long as they pay for their advertising), while forcing their members to register with full, (quasi) verifiable information. This is to protect the providers, and as such the providers usually have access to the user database to confirm their customers aren't law enforcement or known abusers. Information gleaned over many of these sites in cities where a target has a presence can be used to identify persons at the organizations with relative ease. Obviously the larger the target company is, the higher the chance of finding a hit.

Honeytraps can also be deployed in a manner to attract employees already disgruntled with the organization. Creating a website that purports to be opposed to the organization due to some perceived grievance (even a manufactured grievance should work), will sometimes draw current or former employees willing to divulge information that attackers can use. Additionally, the site can be used in tandem with social media poisoning to attempt to change both public and employee perception of the target, incentivizing additional attacks both externally and from within. Reports of pending layoffs, financial misdeeds, environmental damage, and the like can all be effectively used to mold perception.

This ploy can even be extended, by in turn leveraging a reverse honeypot. For example, the "activists" can claim they have found a backdoor into the company or have compromised and mirrored an entire server. The organization in question will no doubt be monitoring this site, and when the reverse honeypot is announced, will invariably attempt to login to the site in order to validate the "breach." The login attempts will also invariably yield valid credentials to the site, which in turn can be used to in turn breach the actual site. This method would be effective for systems with 2-factor authentication too, as long as the supplied credentials are relayed immediately to the target systems.

Counterintelligence, misdirection, weaknesses of other attackers (if they exist and can be identified/created), and false flag attribution can also be utilized to increase the effectiveness of the attack, tie up defenders, and minimize detection. "Leaking" data and attributing it to specific individuals within the target company through "accidental incomplete cleansing" can sow doubt within an organization. Not only does this isolate the individual within the company and create havoc, it in turn can cause the employee to become disenchanted with the organization if they doubt his integrity. The employee is thus incentivized to get revenge.

V - "Poll" - Monitor and Update The Asset List

Continuously monitor and update the asset to identify if its state has changed. Has a compromised asset been "fixed"? If so can it be re-compromised? If an asset is lost, perform a damage assessment to ensure no attack leakage has occurred (i.e. corporate security bulletin sent out), or if it was fixed as part of a routine remediation/audit process. Identify any possible

replacement assets that will replace the compromised asset's function in the attack.

Perform an asset effectiveness assessment. Identify what assets are performing the best, identify and risks to the asset which could lead to the asset being lost. It is critical that the attack team develop a standard system to rank effectiveness in a quantifiable manner as applicable to the project.

Next, identify any non-performing or underperforming assets and decide if the asset has any chance of becoming more effective given a specific impetus or time period. If it can, a cost-benefit analysis should be done to determine if the required resources to turn the asset into a performing one is worth it. If not, then it is prudent to backburner the asset or potentially drop it altogether.

Finally, identify if postures have changed of previously non-compromised assets? What might have not been an available asset might become open to acquisition. This happens frequently with employees, especially when companies are "right sizing", during acquisitions, and after annual performance reviews. It is also important to continuously monitor for new assets that come online.

The larger the attack team's resources are, be it in manpower, cash, or other operational support assets, the more field assets they can manage. It is obviously in the team's best interest to keep as many assets functioning as possible, but only if they can be monitored and maintained effectively. As available resources change, the attacker must reassess the ability to either maintain or expand the asset base, or begin to cull assets.

Summary

The attack methodology described herein is a new way of looking at the defensive posture of an organization, and picking the most strategic place to attack with minimal risk. This happens regularly against your defenses in the real world, so exposing and educating on this topic is critical to defending yourself. While this paper is purely offensive in nature, it like many other tactical positions are essential in creating good defense.

Remember that real-life is not a spy movie, and attacks aren't nicely contained within the parameters with which your defenses are built to handle. This means that the tools and technologies you've likely employed today only protect you against those who don't have advanced attack strategies such as the ones presented here. This should serve as a wake-up call to anyone who is building and maintaining a security program who is genuinely interested in creating a culture of awareness, and ultimately a strongly secured posture for their organization. Both authors firmly believe that the only way to defend oneself properly is to understand and learn the offensive tactics and methodologies that the *"bad guys"* will use against you, since again, they often don't play by any convenient rules.

We encourage you to learn these tactics, understand the methodology and mindset. If you are an ethical attacker whose role it is to penetrate defenses professionally – this should help you

significantly get an edge over those you're attacking – even if their defensive posture is strong. While most organizations don't take this much effort to penetrate successfully, stronger industry awareness and education is making it more difficult to penetrate the well-defended organizations. If you are a defender, this should provide you with a glimpse into what you're *really* up against. This paper and presentation provides you with a peek into the darker side of the adversarial world of those who seek to do you harm, and penetrate your defenses. Use offensive threat modeling education to better arm yourselves.

--

References:

[a] "Phrases That Signal Workplace Hierarchy", 2012, Eric Gilbert, School of Interactive Computing & GVU Center, Georgia Institute of Technology, gilbert@cc.gatech.edu

[b] http://comp.social.gatech.edu/hier.phrases.txt

[c] "Sentiment Analysis on Personal Email Archives", 2011, Hangal and Monica Lam, Stanford Computer Science Department, ACM 978-1-4503-0268-5/11/05

[d] http://mobisocial.stanford.edu/muse

[e]" Abusing Social Networks for Automated User Profiling", 2010, Marco Balduzzi, Christian Platzer, Thorsten Holz, Engin Kirda, Davide Balzarotti, and Christopher Kruegel, Institute Eurecom, Sophia Antipolis/Secure Systems Lab, Technical University of Vienna/University of California, Santa Barbara

[f] "Determining personality traits & privacy concerns from Facebook activity", Chris Sumner, Alison Byers, Matthew Shearing, The Online Privacy Foundation, Black Hat Briefings 2011, Abu Dhabi

--

Other resources:

In many cases, data culled from social media, supply chain, satellite imagery, property management, conference attendance, personal browsing predispositions, sales literature, even political campaign donations can be aggregated and prioritized to increase the likelihood of success.

Targeting supplier chain – posing as client - -contacting sales guy - Bill's no longer with the company; it's kind of ugly to be honest. Please take him off the contact list, do not give him

http://www.icwsm.org/papers/3--Gosling-Gaddis-Vazire.pdf