RECONNAISSANCE

A Walkthrough of the "APT" Intelligence Gathering Process

October, 2015

Content and liability disclaimer

This Research Paper is for general information purposes only, and should not be used as a substitute for consultation with professional advisors. EMC has exercised reasonable care in the collecting, processing, and reporting of this information but has not independently verified, validated, or audited the data to verify the accuracy or completeness of the information. EMC shall not be responsible for any errors or omissions contained on this Research Paper, and reserves the right to make changes anytime without notice. Mention of non-EMC products or services is provided for informational purposes only and constitutes neither an endorsement nor a recommendation by EMC. All EMC and third-party information provided in this Research Paper is provided on an "as is" basis.

EMC DISCLAIMS ALL WARRANTIES, EXPRESSED OR IMPLIED, WITH REGARD TO ANY INFORMATION (INCLUDING ANY SOFTWARE, PRODUCTS, OR SERVICES) PROVIDED IN THIS RESEARCH PAPER, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, AND NON-INFRINGEMENT. Some jurisdictions do not allow the exclusion of implied warranties, so the above exclusion may not apply to you.
In no event shall EMC be liable for any damages whatsoever, and in particular EMC shall not be liable for direct, special, indirect, consequential, or incidental damages, or damages for lost profits, loss of revenue or loss of use, cost of replacement goods, loss or damage to data arising out of the use or inability to use any EMC website, any EMC product or service. This includes damages arising from use of or in reliance on the documents or information present on this Research Paper, even if EMC has been advised of the possibility of such damages

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Every meticulous APT attack starts with a comprehensive intelligence gathering that includes getting to know the target before proceeding to a more invasive act. In this research paper, we will discuss the reconnaissance process performed on a potential target from the perspective of the adversary. This demonstration will show how much information can be harvested from a hypothetical targeted entity, using techniques, tools, and procedures (TTP) which are available to literally anyone.

Some may suggest that certain threat actors do not necessarily use conventional means/sources since they can afford more elaborate means for collecting their intelligence ("intel"). While this may be true, such means are generally unnecessary due to the amount of intel that can be gathered by open source intelligence (OSINT).

First, let's go over the basic terms, what exactly is an APT?

According to Wikipedia, an APT means –
"set of stealthy and continuous computer hacking processes, often orchestrated by human(s) targeting a specific entity"[1]

History teaches that an "entity" could be anything or anyone from an individual to a small business, scaling up to a large corporation/organization or even a government agency.

This clarifies one edge of the puzzle. The next important step is to ascertain the identity of the attacker (or as it has become more readily known, the "threat actor" or "TA"). Past incidents and many media covered events teach that TAs may range from a single basement-dwelling hacker to state-funded agencies with a wide array of motivations.

So before we can dive into the actual intelligence work, we need to know what means are available to the individual TA.

KNOW YOUR INTEL

Intelligence work is basically the art of connecting the dots: the mastery of data collection, correlation and the ability to make educated assumptions on the missing parts. In order to do this properly, TAs need to know what resources are available, especially when talking about open source intelligence. Thus, knowing the intel is realizing what kind of resources are available and what type of information is expected to be found.

When performing intelligence gathering, there are two primary types of information – human and digital intel. Different tools, techniques, and procedures may be employed for each.

The tools a TA is going to use to perform human intelligence on the targeted entity are nothing more than a well-known search engine and a popular business-oriented social network. Crawling through these sites focused on a hypothetical TA's target, a TA can gather material on both the human and the company itself.

For the digital part of our reconnaissance, a hypothetical TA would use common search engines to seek leads on certain employees of the target and further explore these with "traditional" hacker tools.

TARGET IDENTITY

A TA will likely explore the target victim systematically, using both human and digital intelligence. The first step of the reconnaissance process is to assemble a sort of identity profile.

A quick search of the hypothetical company name reveals its official website, which includes the location of its headquarters site as well as the formal contact information. In this hypothetical, it is located in New York, USA.

Knowing this, the TA could determine the time zone and the estimate the working hours of the hypothetical target. In this case, it is New York time - UTC-05:00. Timing is everything when choosing the most suitable time for any operation. The type of attack would dictate the timing. If it were a spear phishing mail sent to an employee, the TA would like do this during working hours. If the TA were to undertake an intrusive act that may take several hours and/or may cause disruptions to the target victim's service/network, he would probably choose to do it off working hours or during the weekend.

[1] https://en.wikipedia.org/wiki/Advanced_persistent_threat

Continuing with the intelligence gathering, the TA would then examine company's page on a business-oriented social network. A short review indicates it was founded about 5 years ago and that it has fewer than 50 employees. Such information provides the TA with the maturity of the organization and its scale, as well as the industry category and the estimated value of the company.

Illustration1: The Headquarters location

This publicly available information would be enough for the TA to assemble the target basic identity, which would include:

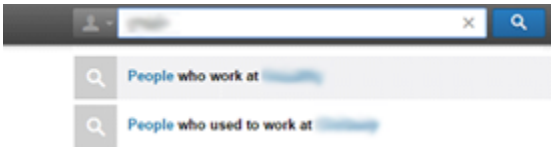
Name: Victim Corp
Location: New York, NY 10001, United States
Time Zone: UTC-05:00
Industry Category: Internet
Company size: <50 employees
Exist since: 2010
Estimated value: $5,000,000 (USD)

TARGET PERSONAL PROFILING

Why collect intelligence on the human factor of our target?

The answer for this question is rather simple, if even a bit cliché, because the human factor is the weakest link. The deluded tendency to think that APT attacks are all conducted using sophisticated technological means with yet-to-be-seen techniques and vulnerabilities (also known as 0-days[2]) often leaves companies blind to the actual non-sophisticated foundational elements of some attacks.

An analogy to better illustrate the spectrum of techniques employed in an APT: a spear phishing email is to a rocket what a 0-day exploit is to a warhead. If one gathers enough intelligence on an employee victim, he can surely form an attack scenario customized to gain trust and trick the user to click on a malicious link, open a malicious file, or simply download and run a malicious executable file. Ultimately, a simple click could result in a compromise to the employee's personal credentials and/or infecting his or her end-point device with malware.



In addition to gathering information about the target company, business-oriented social networks are also used by TAs to gather information about individual employees.

As we can see, regular search can reveal two types of individuals related to the target: both current and former employees, which may provide the TA with insight into the target's organizational structure. The TA can make further use of information about former employees later on when assembling its attack scenarios.

Examining the listings of employees, the TA can identify about fifty individuals from different functions of the company – Developers, Product management, Architects, IT, Marketing, HR, Executives and many more. The TA would collect all the names, professional descriptions, and identifiable interpersonal relations.

Now, since it is a rather small company with fifty or so employees, a few individuals would be of interest for a TA. They could be key-positioned employees, or just employees who are close enough or have access to the TA's objective. These individuals can range from the IT guy (usually considered as the holy grail - from the APT perspective), to an HR professional, to the Server Developer engineer.

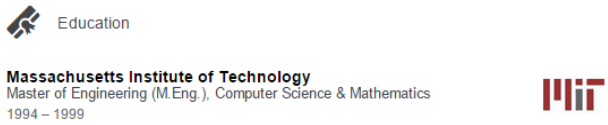
Considering how much information the TA might be able to gather, a likely target would be the Server Developer engineer. And for the sake of this example and our amusement, let's call him - Jesse Pinkman.

Illustration 2: Targeted company employees

[2] Zero-day (also known as zero-hour or 0-day) is a computer threat that exposes undisclosed or unpatched computer application vulnerabilities. Zero-day attacks can be considered a great threat because they take advantage of computer security holes for which no solution is currently available.

Before collecting intel on Jesse from new sources, let's make sure we squeeze any detail we can from his social network profiles. First, we'll concentrate on his personal details rather than his technical mastery; since the later may reveal more about the targeted company than Jesse himself.

So we can see that Jesse's employment history begins in the year 2000, preceded by his education at MIT (1994-1999) resulting in an MS in engineering, Computer Science & Mathematics - highly educated.



Assuming he was around the age of 22-23 when he started at MIT (in 1994), we can make an educated guess that his current age is somewhere around 43-44. We should also note the company he founded in 2006. This may warrant further investigation later.

The profile also lists "Liked pages". Nearly every modern social network offers a similar section…. where the user willingly reveals his areas of interest.

Today's "like" and "favorite" buttons can make profiling an individual quite easy: a list of topics that might attract our prey.

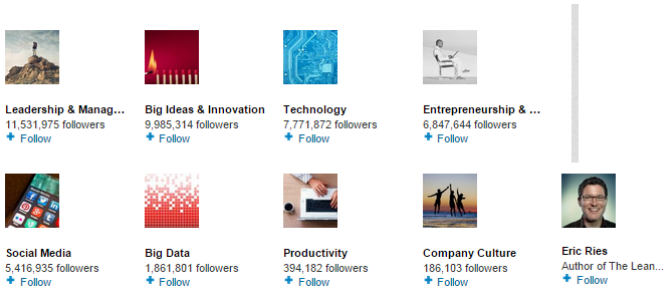


Illustration 2: Show me your "Likes" and I'll tell you who you are.

Now, let's return to a lead identified earlier, the Company Jesse founded in 2006. The business-oriented social network page of the company offers a link to a website. A quick glance at the site reveals there there's a conspiquous lack of content. The company may have never really existed. If it existed only to pad Jesse's resume, perhaps that information could be exploited later. There isn't much else to glean from the website, but we can check the *Whois* record of the domain.

Whois is a query and response protocol that is widely used for querying databases that store the registered users or assignees of an Internet resource, such as a domain name or an IP address block.

Using the whois command line tool on Jesse's website reveals some useful info: a full address, phone number and even a personal e-mail address for Jesse.

```
attacker@evil $ whois employeesite.com

Admin Name: Jess Pinkman
Admin Organization: [CENSORED]
Admin Street: 127 [CENSORED] rd.
Admin City: Newton
Admin State/Province: MA
Admin Postal Code: 024[CENSORED]
Admin Country: US
Admin Phone: +00[CENSORED]
Admin Phone Ext:
Admin Fax:
Admin Email: [CENSORED].mit.edu
…
```

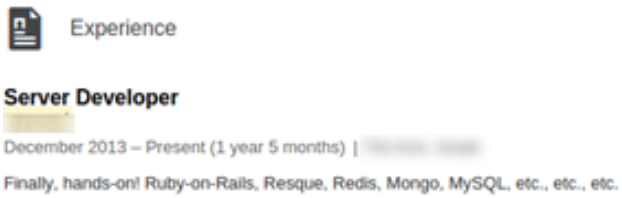
The residence of the target can be partially validated by using external street-view services, ensuring that the address is, in fact, real.

Given Jesse's background (computer science major, professional software developer), one can anticipate that he is active in various online platforms. Searching for information about Jesse on common search engines using his email address and full name reveals several social network profiles, personal pages, and records.

From his personal social network profile, we can determine his marital and family status. His privacy settings are configured such that anyone can read his posts. Information can be gathered about his daily routine and political views. We can even correlate between his friends and colleagues and gain insights into their relationships.

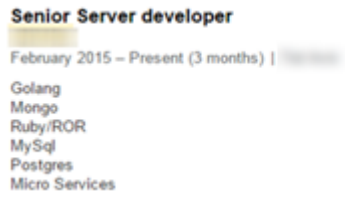
DIGITAL INTELLIGENCE

The digital intelligence part of reconnaissance is where all the digital resources involving the target are gathered and mapped. This is the search for every piece of information about the digital / technology side of the potential victim.



Let's start this process from where the human intel analysis started – Jesse's social network profile. Judging from his experience at the current company, the TA can form hypotheses about the technology employed by the targeted entity. Judging from the work Jesse claims to be doing, the TA can expect that the server applications are probably written in Ruby using Ruby-on-Rails framework and Resque framework for background apps & cron[3]. Further, they are likely using Redis as cache server and finally, their database architecture is a combination of MongoDB and MySQL.

In examining some of Jesse's colleagues, the TA can find a teammate and fellow Server Developer, who although more experienced, has been with the company for only three months. He lists the following:



This is another indicator that the target is using Ruby based technology and MongoDB + MySQL in combination. He is generous enough to hint the presence of PostgreSQL DB as well as the usage of GoLang (Google's programming language).

In viewing a few more developers/software architects profiles, the TA can get a general idea of the overall specialization of the target's development team.

[3] A cron is a time-based job scheduler in Unix-like computer operating systems.

Let's now move onto more proactive means. The next step is to map any resources in use by the targeted entity. Beginning with their website[4], I use a tool called dig to resolve the IP address of their website server.

```
attacker@evil $ dig victim.com

; <<>> DiG 9.9.5-3ubuntu0.2-Ubuntu <<>>

;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 43150
;; flags: qr rd ra; QUERY: 1, ANSWER: 6, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4000
;; QUESTION SECTION:
; victim.com.                    IN      A

;; ANSWER SECTION:
victim.com              299     IN      A       54.216.[CENSORED]
```

From here a TA would try to determine the scope of his target. Sometimes companies/organizations have entire IP blocks available at their disposal, while others use cloud services which provide their own IP's. In any case, the TA should be able to find some indication for that using whois tool.

```
attacker@evil $ whois 54.216.[CENSORED]

NetRange:       54.208.0.0 - 54.221.255.255
CIDR:           54.220.0.0/15, 54.216.0.0/14, 54.208.0.0/13
NetName:        [CENSORED]-2011L
NetHandle:      NET-54-208-0-0-1
Parent:         NET54 (NET-54-0-0-0-0)
NetType:        Direct Allocation
OriginAS:       AS16509
Organization:   [CENSORED]Technologies Inc. (AT-88-Z)
RegDate:        2013-02-19
Updated:        2013-02-19
Ref:            http://whois.arin.net/rest/net/NET-54-208-0-0-1
```

[4] For this example, the author has used his own website and run the cited tools in a secure lab environment.

The result of the whois shows us that the company website is hosted on a 3rd party hosting service. So there is no use for scanning the whole block. Instead, we are going to use another tool called dnsmap. This tool 'brute forces' subdomain names using a preset dictionary. By doing so, a TA might be able to expose more servers used by our target.

```
attacker@evil $ dnsmap victim.com

dnsmap 0.30 - DNS Network Mapper by pagvac (gnucitizen.org)

[+] searching (sub)domains for victim.com using built-in wordlist
[+] using maximum random delay of 10 millisecond(s) between requests

testing.victim.com
IP address #1: 54.247.[CENSORED]

cc.victim.com
IP address #1: 46.137.[CENSORED]

owa.victim.com
IP address #1: 54.231.[CENSORED]

mail.victim.com
IP address #1: 54.125.[CENSORED]

reports.victim.com
IP address #1: 54.170.[CENSORED]

www.victim.com
IP address #1: 54.216.[CENSORED]
```

And so it has! It is clear that 'owa.victim.com', their Outlook Web Application server, is exposed. This could be used as a good attack target in the future.

Next, we could scan all of the open services on these servers using nmap to fingerprint each one (platforms, versions, etc.).

```
attacker@evil $ nmap –sV –P0 testing.victim.com,cc.victim.com,owa.victim.com,mail.victim.com,
reports.victim.com,www.victim.com

Starting Nmap 6.40 ( http://nmap.org ) at 2015-05-28 15:40 IDT

Nmap scan report for beta.victim.com  (54.247. [CENSORED])
Host is up (0.23s latency).
rDNS record for 54.247.[CENSORED]: [CENSORED]-54-247-107-7.eu-west-1.compute.[CENSORED].com
Not shown: 847 closed ports, 150 filtered ports
PORT     STATE SERVICE      VERSION
80/tcp   open  http         nginx 1.2.9
443/tcp  open  ssl/http     nginx 1.2.9
1720/tcp open  H.323/Q.931?

Nmap scan report for dl.victim.com  (54.231. [CENSORED])
Host is up (0.00081s latency).
Not shown: 998 filtered ports
PORT     STATE SERVICE      VERSION
80/tcp   open  http         Amazon S3 httpd
1720/tcp open  H.323/Q.931?

Nmap scan report for mail.victim.com   (74.125. [CENSORED])
Host is up (0.00074s latency).
Not shown: 997 filtered ports
PORT     STATE SERVICE      VERSION
80/tcp   open  http         Google httpd 2.0 (GFE)
443/tcp  open  https?
1720/tcp open  H.323/Q.931?
Service Info: OS: Linux; CPE: cpe:/o:linux:linux_kernel

Nmap scan report for ntp.victim.com  (54.216. [CENSORED])
Host is up (0.00082s latency).
rDNS record for 54.216.[CENSORED]: [CENSORED].compute.[CENSORED].com
Not shown: 997 filtered ports
PORT     STATE SERVICE      VERSION
22/tcp   open  ssh          (protocol 2.0)
80/tcp   open  http         nginx
1720/tcp open  H.323/Q.931?
1 service unrecognized despite returning data. If you know the service/version, please submit the following fingerprint at
http://www.insecure.org/cgi-bin/servicefp-submit.cgi :
SF-Port22-TCP:V=6.40%I=7%D=5/28%Time=55674634%P=x86_64-pc-linux-gnu%r(NULL
SF:,29,"SSH-2\.0-OpenSSH_6\.6\.1p1\x20Ubuntu-2ubuntu2\r\n");

Nmap scan report for reports.victim.com  (54.170. [CENSORED])
Host is up (0.23s latency).
rDNS record for 54.170.[CENSORED] : [CENSORED]west-1.compute.[CENSORED].com
Not shown: 995 filtered ports
PORT     STATE  SERVICE      VERSION
22/tcp   open   ssh          OpenSSH 5.9p1 Debian 5ubuntu1.4 (Ubuntu Linux; protocol 2.0)
80/tcp   open   http         nginx 1.2.9
443/tcp  open   http         nginx 1.2.9
1720/tcp open   H.323/Q.931?
8080/tcp closed http-proxy
Service Info: OS: Linux; CPE: cpe:/o:linux:linux_kernel
```

A TA may try to try to fuzz web servers using a tool such as wfuzz to uncover vulnerable services or useful files and directories which may lead to a valuable/useful information.

Identifying employee e-mail addresses would be a good next step. We need to figure out the company's e-mail naming convention. Luckily we have fifty of the company's employee's names that we can use for testing purpose.

In some cases we can exploit a common technique for validating e-mail addresses. This technique exploits the fact that some mail servers send back a `bounce` email in case the recipient does not exist. A TA could search for an existing email address of one of

the employees using search engines and imitate the convention, or he could simply iterate through some of the most common conventions as shown below until the e-mail doesn't bounce.

jesse.pinkman@victim.com
jessepinkman@victim.com
jesse@victim.com
pinkman@victim.com
jpinkman@victim.com
j.pinkman@victim.com

Once an email address is attained, the task shifts to contacting the target directly and establishing trust until the TA is able to trick him into opening a document containing a malicious macro, or even simply running an executable file.

SUMMARY

Awareness of open source intelligence is increasingly relevant. As demonstrated, a tremendous amount of intelligence could be gathered on almost every company/organization and their employees. A potential threat actor could leverage this information to form a specific and persuasive attack scenario against a few carefully chosen individuals. It requires no advanced knowledge, and merely takes one weak link to lead to compromise.

MITIGATION

Unfortunately, the human intelligence attack surface is not one that can be remedied simply by buying more security technology.

Instead, mitigation should center on education programs that raise awareness of the following topics:

- Online Privacy and Information sharing
- Implications of personal privacy on the enterprise/organization
- The nature of APT attacks, and how this information can be exploited

The effectiveness of such programs can be measured by professional 'red' teams who attempt to penetrate the organization, focusing on human intelligence and social engineering. Information gathered during these evaluations can also be leveraged for training purposes.

RSA's Advanced Cyber Defense Practice trains and enables security practices to analyize and reduce exposure, including hardening against reconnaisance. Organizations must evolve their security programs to account for their advesary's tactics, which may include addresses these less-technically-sophisticated methods. Empowering operations teams to collect and analyze online information related to personel may ultimately serve to reduce risk by preventing cyber attacks, very early in the attackers' process.

AUTHOR

Rotem Kerner, RSA Research