THURSDAY, JANUARY

RESEARCH SPOTLIGHT: NEEDLES IN A HAYSTACK

This post was authored by Mariano Graziano.

Malware sandboxes are automated dynamic analysis systems that execute programs in a controlled environment. Within the large volumes of samples submitted daily to these services, some submissions appear to be different from others and show interesting characteristics. At USENIX Security 2015 I presented a paper in which we proposed a method to automatically discover malware developments from samples submitted to online dynamic analysis systems. The research was conducted by dissecting the Anubis sandbox dataset which consisted of over 30M samples collected in six years. The methodology we proposed was effective and we were able to detect many interesting cases in which the malware authors directly interacted with the sandbox during the development phase of the threats.

Another interesting result that came from the research concerns the samples attributed to Advanced Persistent Threat (APT) campaigns. Surprisingly, some of the malware samples used in these sophisticated attacks had been submitted to the Anubis sandbox months -- sometimes even years -- before the attack had been attributed to the proper APT campaign by a security vendor. To be perfectly clear, we are not saying that it took security vendors months or years to detect a threat. Most times, we are able to detect the threats in no more than a few hours. It is just that the malware samples were mislabeled and not properly associated with APT campaigns. In general, the same goes for non-APT malware campaigns. In this blog post, we tried to see if the same applied to the Cisco dataset. Specifically, we chose ten APT campaigns, -- some of which were already covered in the Usenix paper. We decided to inspect two different datasets: our incoming sample feeds / malware zoo, and the telemetry associated with our Advanced Malware Protection (AMP) solutions. Talos receives samples from over 100 external feeds ranging from anti-malware companies to research centers, while the AMP dataset contains telemetry from the Cisco AMP user-base.

The remaining part of this post is organized as follows. First, we show the APT campaigns we investigated. Second, we summarize the results of the analysis of the Talos dataset. Third, we show the results from the AMP dataset. Finally, we summarize our findings.

A P T   C A M P A I G N S

APT CAMPAIGN	MADE PUBLIC
Beebus	February 2013
Arid Viper	February 2015
Red October	January 2013
Equation	February 2015
Pacific RAT	July 2014
Regin	November 2014
Aurora	January 2010
Pitty Tiger	July 2014
Net Traveller	June 2013
BrutPOS	July 2014

The ten malware campaigns in the table above garnered significant media attention when discovered, with some of them clearly falling in the area of APT. They were found by different security companies between 2010 and 2015, having different levels of sophistication and different objectives. Moreover, these APT campaigns were not limited to western countries. They have affected organizations all over the world. Most of the time, connecting the dots and drawing relationships between samples and campaigns take months and many experts. This means the security company that releases a detailed report documenting the campaign is aware of it long before the information is made public. However, we believe the “public release” date is still a good metric, because it is the moment at which all the other security companies and the entire world are made aware of these threats.

Another important aspect during an APT investigation is attribution. While detection is done quickly, attribution for these campaigns is often an open and hard problem to solve. Most of the times the perpetrators remain unknown even after months of work by security researchers. However, sometimes researchers are able to connect the dots and attribute the attack to a threat actor. This was the case for some of the APT campaigns discussed so far. Some of these threats have been attributed to state-sponsored actors, others to cyber criminals or to espionage attacks. However, like in the USENIX publication, in this post we will make no speculation about attribution.

In the next paragraphs, we will present the results of searching for samples associated with these APT campaigns in our datasets.

T A L O S   D A T A S E T

APT CAMPAIGN	AVG DAYS BEFORE APT CAMPAIGN PUBLICALLY IDENTIFIED
Beebus	574
Arid Viper	178
Red October	68
Equation	1371
Pacific RAT	455
Regin	1018
Aurora	80
Pitty Tiger	602
Net Traveller	105
BrutPOS	68

This table shows the results of the analysis of our incoming sample feeds/malware zoo. For every campaign, we checked in our malware zoo to see when they had been initially submitted to us. Given that we know when information about these APT campaigns was made public, we can compute the number of days it took the security community to publicly tie the samples to an APT campaign, even though the samples had been marked malicious for other reasons. On average, these samples went for 458 days before being tied to an APT campaign. The table presents the average number of days for the entire campaign, and we go from a few months as in the case of "Aurora" to more than three years for "Equation". Notice that these figures come from our malware zoo which collects samples from external sources and in general are a good indicator given the amount of samples received per day. Notice that these numbers vary based on the dataset.

VIRUS TOTAL

The vast majority of the submissions come from big organizations such as Antivirus companies. Interestingly, a significant percentage is submitted by VirusTotal. For this reason we decided to check the submitters for possible links and intelligence information. As already documented by Dixon, information about the submitters of samples is not publicly available, but can partially be retrieved from their Intelligence service. For every sample, it is possible to know a hash (a hexadecimal unique identifier of the submitter), the country (from the geolocalization of the IP address of the submitter) and the method (the way the sample has been submitted, for instance via the web interface or the APIs). This opaque information complicates the analysis a little bit, but it is still possible to obtain interesting results.

SUBMITTER	CAMPAIGNS
6exxxxxx	AridViper Nettraveller RedOctober BrutPOS PittyTiger
14xxxxxx	AridViper Regin
22xxxxxx	AridViper Regin Nettraveller BrutPOS PittyTiger
20xxxxxx	AridViper Nettraveller PacificRAT BrutPOS PittyTiger
5exxxxxx	Equation Regin BrutPOS Auror
72xxxxxx	Equation Regin BrutPOS
4bxxxxxx	Regin
3bxxxxxx	Regin
cdxxxxxx	Beebus PittyTiger Nettraveller BrutPOS
b4xxxxxx	Aurora

The table above summarizes our findings from VirusTotal. The first column shows the hash of the submitter. This means that the submitter sent one or more samples of a given APT campaign to VirusTotal before its public release. One can only speculate on who these submitters are. They could very likely be the threat actors themselves, testing to see if their malware is detected by the AV companies. They could also be security researchers or vendors who are trying to get information from VirusTotal. It is noteworthy that in most of the cases the same submitters uploaded samples belonging to different APT campaigns.

CISCO AMP

We went through our logs to search for entries that contained hashes related to the ten APT campaigns we have been investigating. Interestingly, we got hits for eight different hashes belonging to three different campaigns that were discovered on Cisco AMP customer machines before the APT campaign was publicly identified.

APT CAMPAIGN (NUM OF SAMPLES)	DAYS BEFORE APT CAMPAIGN PUBLICALLY IDENTIFIED
Arid Viper (1 SAMPLE)	-50
Equation (1 SAMPLE)	+1
BrutPOS (6 SAMPLES)	-64

As illustrated in the table above, we identified eight malicious samples that were in the wild before being associated with APT campaigns. It is important to repeat that most of these samples were detected as malicious the moment they first appeared on our customers' machines.

Surprisingly, one sample of the Equation APT campaign (fanny worm) was found and blocked on a Cisco AMP customer's machine a day after the public release of the Kaspersky report.

HASH (SHA256)	DATE	DISPOSITION	APT
003315B0AEA2FCB9F77D29223DD8947D0E6792B3A0227E054BE8EB2A11F443D9	2015-02-17	MALICIOUS	EQUATION
003315B0AEA2FCB9F77D29223DD8947D0E6792B3A0227E054BE8EB2A11F443D9	2015-02-17	MALICIOUS	EQUATION
015FBC0B216D197136DF8692B354BF2FC7BD6EB243E73283D861A4DBBB81A751	2014-12-20	UNKNOWN	ARID VIPER
015FBC0B216D197136DF8692B354BF2FC7BD6EB243E73283D861A4DBBB81A751	2014-12-20	MALICIOUS	ARID VIPER
015FBC0B216D197136DF8692B354BF2FC7BD6EB243E73283D861A4DBBB81A751	2015-01-02	MALIICIOUS	ARID VIPER
015FBC0B216D197136DF8692B354BF2FC7BD6EB243E73283D861A4DBBB81A751	2015-01-16	MALICIOUS	ARID VIPER
015FBC0B216D197136DF8692B354BF2FC7BD6EB243E73283D861A4DBBB81A751	2015-02-12	MALICIOUS	ARID VIPER
14BFDA4A4ACA1276388702D0FB7629AF120FF34C1ACDEB7613815F2981C99832	2014-05-07	MALICIOUS	BRUTPOS
508909C8A00026C904F52099DD62BBF4062B4E8E40FC0601BD9E13570514B4F5	2014-05-06	MALICIOUS	BRUTPOS
7170A07BCB5B0467A75CBD17A1A1877AEC3C8EA43C45D3BED6AB5E6C95A62713	2014-05-06	MALICIOUS	BRUTPOS
9A10916AD0F43FA3376C2E54FD5CFDD06D684B3A19895ED4107FAF9F3313DCDA	2014-05-07	MALICIOUS	BRUTPOS
E28EABEB678AFB5E172F4127C5692E742809FD86DFA8478C1DC6F9C13B2A8E5F	2014-05-06	UNKNOWN	BRUTPOS
E28EABEB678AFB5E172F4127C5692E742809FD86DFA8478C1DC6F9C13B2A8E5F	2014-05-07	MALICIOUS	BRUTPOS

Based on our logs, Cisco AMP found the sample 015FBC0B216D197136DF8692B354BF2FC7BD6EB243E73283D861A4DBBB81A751 twice on 2015-12-20. It was "unknown" to AMP the first time, but detected as malicious the second time.
E28EABEB678AFB5E172F4127C5692E742809FD86DFA8478C1DC6F9C13B2A8E5F was "unknown" to AMP on 2014-05-06, but detected as malicious the next time it was seen on a customer's machine on 2014-05-07. In all the other cases the samples were already considered malicious.

C O N C L U S I O N

As the number of threats per day continues to increase, the number of malware samples security companies automatically analyze increases. Much of the analysis is comprised of dynamic analysis systems, such as sandboxes, to determine whether the sample is malicious or not. These samples are then stored for further analysis. Due to the large numbers of samples, in many organizations the vast majority of these samples remain categorized solely on the initial sandbox run. Even when these samples are shared among companies or via other services like VirusTotal some malware samples can go unnoticed for months because they are marked as malware but given some generic name, such as “Win.Trojan.Agent”. Then we are shocked when a security company discovers an APT campaign that has supposedly gone unnoticed for years.

The results of this post confirm the assumption of the Usenix paper, also based on a dataset of a big security company and similar results are expected throughout the security industry. Many times, malware is initially submitted to sandbox systems and marked as malicious based on the output of the sandbox. Then the authors use that information to tweak the sample to avoid detection in future sandbox runs through various evasion tactics. In other situations, the initial sample may not even be flagged as malicious due to evasion techniques being utilized. By performing statistical analysis and reducing the data through clustering, even samples that avoid initial sandbox detection can potentially be detected as malicious. There is clearly a need for more advanced analytical systems to identify campaigns and link the samples together.

Identifying today’s threats requires multiple layers of protection at various points across the network, along with constantly updated threat intelligence information. Cisco analyzes a massive amount of telemetry data and is able to flag malware as malicious based on multiple factors. By performing manual and programmatic analysis of sandbox data in conjunction with identifying behaviors which are associated with malicious activity, even unknown APT campaigns can be neutralized.

Comment as: Select profile...

Publish Preview

NEWER POST HOME OL PO

SUBSCRIBE TO: POST COMMENTS

SEARCH THE BLOG

Search

SUBSCRIBE TO OUR

Posts

Comments

Subscribe via Email

BLOG ARCHIVE

- Research Spotlight: Needles in a Haystack
- Microsoft Patch Tuesday - January 2016
- Rigging compromise - RIG Exploit Kit

- ► 2 (62) 1 5
- ► 2 (67) 1 4
- ► 2 (30) 1 3
- ► 2 (53) 1 2
- ► 2 (23) 1 1
- ► 2 (94) 1 0
- ► 2 (146) 0 9
- ► 2 (38) 0 8

RECOMMENDED BLOG

CISCO BLOG

Good People Goodness: A Toast to Kanouff, Clemson, and Raman!

SNORT BLOG

Snort Subscriber Rule Set Update for 05/11/2016

CLAMAV® BLOG

ClamAV 0.99.2 has been released!

CONNECT WITH US