State of the Hack: Maltego Use Cases

Over the past several years, data has quickly gone from being sparsely collected to hoarded in mounds, locked away in datacenters and spread across the world. Is there really a need to store so much information? That question has yet to be answered, but with advances in modern technology, it is relatively cheap to store every byte just in case it is needed later. Like other industries that collect data, the information security and threat intelligence industries are no exception. The issue with large stores of data comes in making sense of it all. How does one sift through terabytes of domains, IP addresses, malware strings, assembly code and netflows to find the one indicator driving the next attack? In this week's State of the Hack article, iDefense will cover a process best described as visual data analytics, which is a way to visualize data during analysis to find what data is most critical.

Problem Highlight

Security analysts often observe attacks as they are happening or shortly after they have occurred. It is up to analysts to identify the “who” and “why” of an attack based on the indicators collected. These indicators not only range in type but also in confidence level, as analysts collect and store some indicators based on tangential relationships. To provide the best possible intelligence, data collected from each malicious file, attack, public database and threat feed is stored in a single database that makes connections between the data.

At present, databases holding indicator information hold upward of billions of records. When analysts quickly assemble all of the information about an attack and its related infrastructure, it becomes difficult to find the most meaningful data when querying data collections this large. Assume for a moment that a query is fast and an analyst is able to obtain the results he or she requested. Even in some cases, results come back with thousands of related indicators that may or may not be related to the overarching incident.

Beyond collecting all the important indicators about a particular attack, analysts need to structure their results and analysis so that users can easily follow that analyst’s conclusions. This process of creating a visual of data is both an art and an engineering feat. An analyst must choose a medium and style that will most effectively display that data between the plotting of important data and its relevant connections, and the space in which it is represented. For example, representing several domains and passive IP address associations in a three-dimensional space is not going to work for a report meant to be sent out via e-mail or printed offline. Dealing with such limitations further increases when attempting to show highly connected data or data that is important but several layers outside the starting indicator.

Solution

One of the largest benefits of looking at visualized data is that it provides viewers with the ability to tap directly into the mind of an analyst. Technological tools are excellent, but nothing compares to the human mind and its ability to identify patterns within milliseconds of seeing a well-developed visual aid. It is for this reason that iDefense engineers have chosen to use Maltego, a visualization framework capable of taking custom data feeds and turning them into well-connected graphs, as a visual platform for doing analysis. Maltego allows a user to specify transforms, which are bits of custom code that run locally or on a server, that can take the supplied input, frame it into a query and then return results to the user in the form of visual connections inside a graph.

Using a visualization platform when doing analysis provides three advantages over the traditional method of taking notes. First, a visual platform capable of processing independent data feeds is quickly able to assemble, deduplicate and show associations among data. Second, no matter what the task, if the results of the analysis are worthwhile, an analyst will need to convert them into some type graph for others to visualize the data. By using a visualization platform during the analysis process, an analyst saves time, as the platform will build a visual while the analyst is conducting the research rather than after having conducted the research. Finally, because an analyst can see the data as a visualization platform is plotting it, he or she can quickly use his or her experience to filter or dismiss faulty results—something that could take hours to determine without visuals.

Case Study: itsec.eicp.net Infrastructure and Malware Connections

On Feb. 21, 2013, iDefense identified a password-protected PDF document named "Mandiant_APT2_Report.pdf" that was attempting to pass as a second iteration of Mandiant’s APT1 report released just days earlier. Upon entering this document's associated password, the document would install known cyber espionage malware on the victim computer observed back in November 2012.

The malicious payload used the domain “itsec.eicp.net” for command and control (C&C). eicp.net is a dynamic domain name system (DNS) domain owned by the Shanghai Best Oray Information S&T Company (上海贝锐信息科技有限公司). A quick look at the fully qualified domain name (FQDN) in iDefense’s passive DNS data resources shows that this domain has resolved to more than 800 different IP addresses over the course of its existence. The majority of these IP addresses resolve to five net blocks that Exhibit 4-1 outlines.

Net Block	Owner
1.202.0.0 - 1.203.255.255	China, Beijing - Chinanet Beijing Province Network
111.192.0.0 - 111.207.255.255	China, Beijing - China Unicom Beijing Province Network
114.240.0.0 - 114.255.255.255	China, Beijing - China Unicom Beijing Province Network
115.168.0.0 - 115.171.255.255	China, Beijing - Chinanet Cdma Network
123.112.0.0 - 123.127.255.255	China, Beijing - China Unicom Beijing Province Network

Exhibit 4-1: Top Net Blocks for itsec.eicp.net

Walking through this passive DNS data in an attempt to identify connections to other known advanced persistent threat (APT)[1] intrusion sets and targeted attacks, along with related malware samples, would take an analyst a lot of time conducting data discovery and an initial analysis. Visualizing this data and automating the collection process using Maltego allows an analyst to quickly identify those IP addresses that are most relevant to the research and that expedite follow-on analysis of those addresses. Plotting this IP address in Maltego and running a transform to grab all known IP addresses in passive DNS that are associated with the domain results in the image in Exhibit 4-2. While in and of itself this graph is no more helpful than a static list of passive DNS data, the ability to then run a transform querying for all know domains associated with these IP addresses is more helpful, making this automation and visualization a drastic improvement to an analyst's workflow.

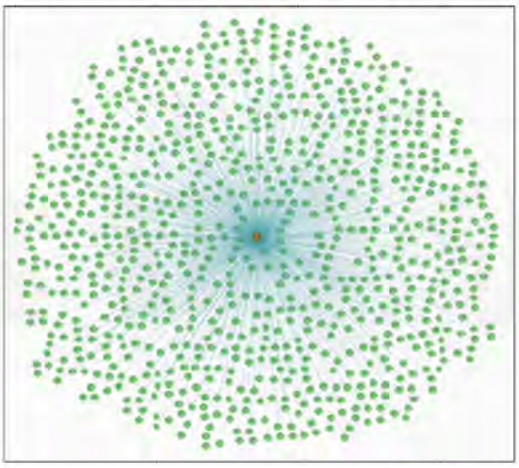

Exhibit 4-2: Itsec.eicp.net Passive IP Connections

As one can see in Exhibit 4-3, iDefense analysts were able to gather approximately 300 FQDNs associated with this infrastructure in minutes and plot it to a graph. The results of the data collection are still somewhat mixed, as they contain dynamic DNS domains, actor-registered domains and some possible crimeware or knockoff-goods domains; this is not surprising since this analysis covers more than 800 IP addresses on major Chinese net blocks.

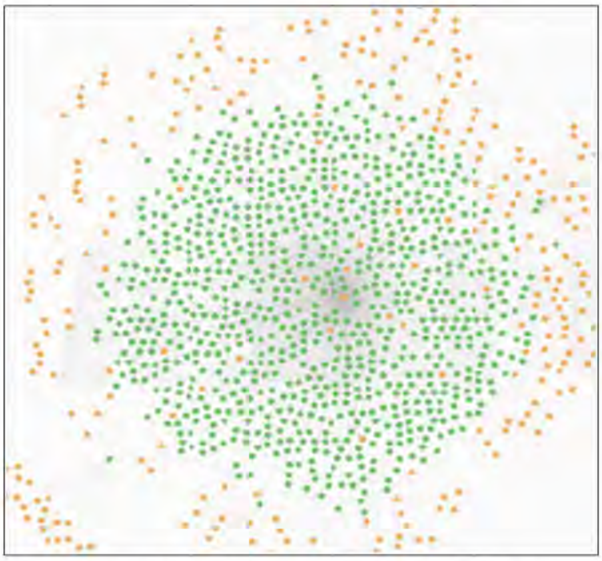

Exhibit 4-3: Tier-2 Domain Connections to itsec.eicp.net

As the graph in Exhibit 4-3 demonstrates, this large data set is still difficult to manage and, in its current view, does not exactly provide an analyst with a good starting point from which to conduct further research into this infrastructure. An analyst could start cherry-picking domains at random to analyze, but that method does not provide solid leads and could take a significant amount of time. Luckily for iDefense analysts, Maltego aids in the process of quickly narrowing those analysts' search. Changing the graph view from normal to "bubble" allows an analyst to view the chart highlighting weights and connections. While these connections may have previously been found using a more manual and less visual process, with Maltego, analysts are able to make connections much quicker. As one can see in Exhibit 4-4, using a weighted view of the connections between IP addresses (green) and domains (orange), an analyst can quickly focus his or her attention on the upper-right quadrant of the graph where the domain and IP address bubbles are larger, signifying that they are more interconnected than other connected IP addresses. That analyst can then work his or her way to the smaller "bubbles" as time permits.

[1] While many countries conduct cyber espionage, the expression "APT" is an umbrella term that the US first used as an unclassified means to discuss the strategic intent and share tactical threat indicators regarding the probable, ongoing intelligence-gathering campaign the People's Republic of China (PRC) is carrying out. The term APT and the various intrusion sets that this term houses are directly analogous to the use of computer network exploitation (CNE) as the cyber arm of their intelligence-collection operations. The primary goal of these cyber espionage operations is to obtain and then maintain access to target networks to exfiltrate intellectual property, personally identifiable information (PII), and financial or targeted strategic information (or both) from governments, corporations and individuals.

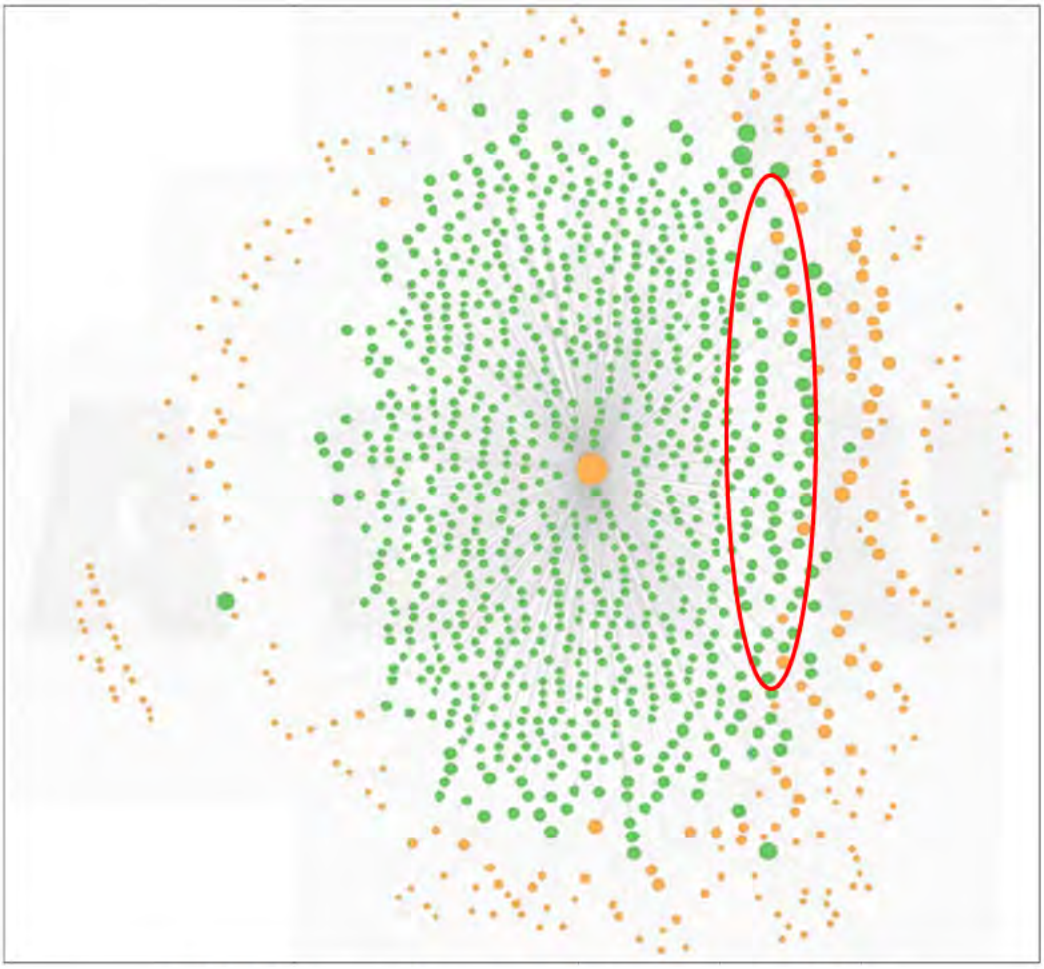

Exhibit 4-4: Cluster of Interesting IP Addresses Based on Weight

Upon looking at the weighted area circled in red in Exhibit 4-4, iDefense analysts identified the following three domains of interest:

- news[.]live-msn[.]net
- help[.]abobesuit[.]com
- zc[.]antivirusbar[.]org

Once analysts have identified the domains of interest, Maltego allows an analyst to transplant those domains and their connections to a new graph (see Exhibit 4-5) to further develop the research in a less-cluttered workspace.

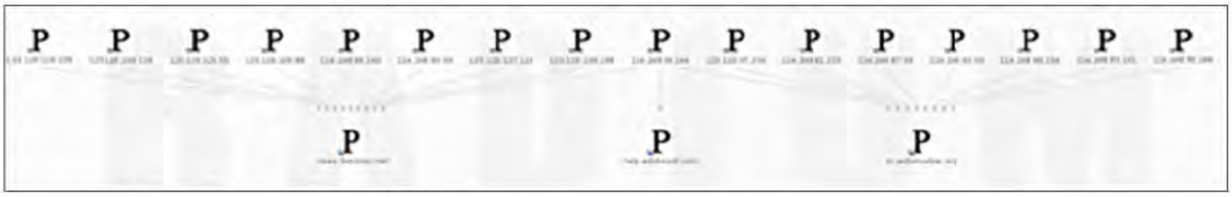

Exhibit 4-5: Malicious Domains Discovered from the revious Result Set

Next, as one can see in Exhibit 4-6, an analyst can quickly access WhoIs data on all the domains, allowing quick identification of any registrant e-mail addresses that iDefense may associate with known intrusion sets. In the case of these domains, iDefense has been tracking the registrant e-mail address dnsjacks@yahoo.com for some time now; however, analysts have not connected it with one single intrusion set. The other two e-mail addresses are new to iDefense, and the company's analysts will track them going forward and will use them in iDefense's proactive monitoring operations.

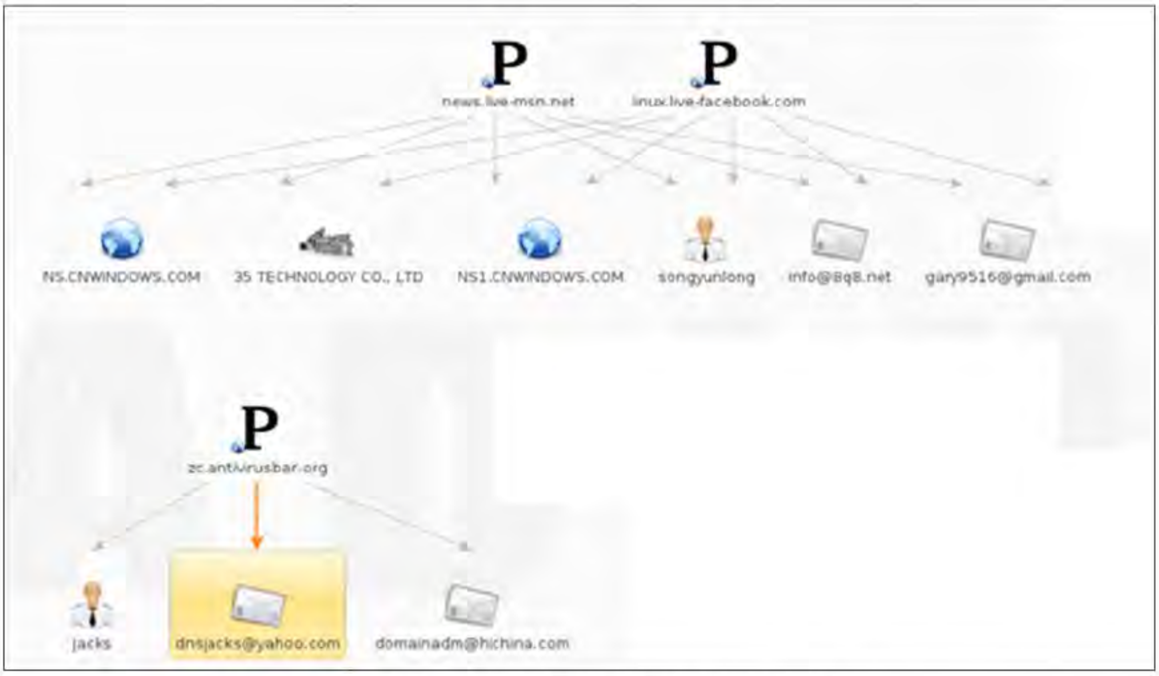

Exhibit 4-6: Registrant Data for Malicious Domains

Finally, an analyst can directly query both iDefense databases and public malware repositories to identify malware samples using these domains for C&C in an effort to allow those analysts to more quickly connect attacks and campaigns and to close any intelligence gaps around the identified domains that may exist. As Exhibit 4-7 shows, two malware samples were associated with "zc.antivirusbar.org," and initial analysis ties these samples to a report published by Dell Secure Works Mirage Campaign, which iDefense tracks under the Lurid intrusion set.

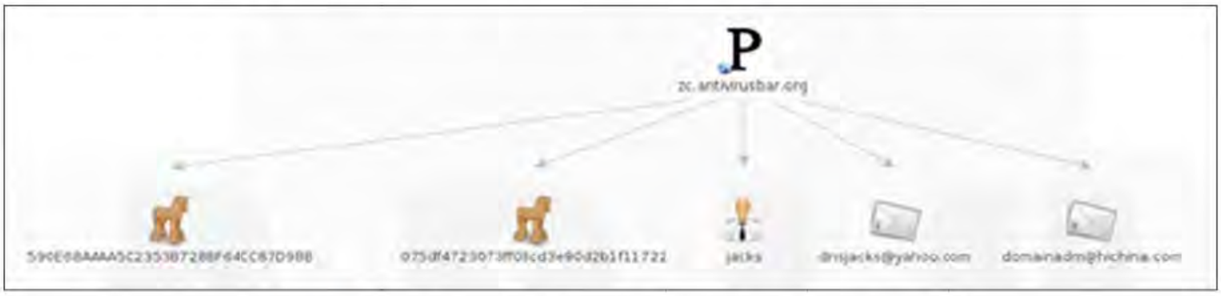
zc.antivirusbar.org
590E68AAAA5C2353B7288F64CC87D988
075df4723073ff08cd3e90d2b1f11722
jacks
dnsjacks@yahoo.com
domainadm@hichina.com

Exhibit 4-7: Malicious Samples using zc.antivirusbar.org for C&C

Besides being able to query multiple data sources and automate certain aspects of data collection, one of the other important pieces of this research is understanding the analyst's workflow and removing those steps that are easily accomplished by a computer. One such example of this workflow automation involves hosting-provider parking sites. When threat actors are not using their C&C domains, they often park them on hosting-provider IP addresses along with hundreds, if not thousands, of domains. In cases like the example in this report with large data sets that need analysis, an analyst cannot check all IP addresses before running passive DNS queries; however, at the same time, an analyst does not want a graph of his or her IP addresses for analysis to be cluttered (even more so that it currently is) with parked domains that may not have any bearing on that analyst's investigation. To prevent this from happening, iDefense has capped all queries to 100 domains.

As Exhibit 4-8 shows, the system alerts an analyst when a query has returned more than 100 results, allowing that analyst to carry out additional research on the entity before plotting it to the graph or to ignore the finding, such as in cases of local host IP addresses. If the analyst decides the query is worthy of additional research, he or she can run the search again, and Maltego will plot the entities to the graph. This saves significant time when dealing with passive DNS data, as most queries returning more than 100 additional entities do not often provide valuable research leads.

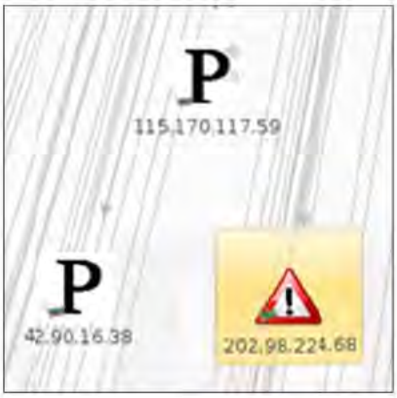

Exhibit 4-8: Example of a System Alert

Having found leads and interesting data, an analyst would normally need to create a visualization ɔf this information, share it with his or her team, take the high-confidence indicators gathered from that analysis and insert them into threat indicators; however, when an analysis is done within a visual engine, there is no need to spend time creating a new graph, as that visual will have already created such a graph. Additionally, when using a visual engine, an analyst can easily share the engine's graph or exported report with his or her team, thereby reducing the amount of time spent on communicating the results of that analyst's research. Lastly, with the use of a visual engine, an analyst can export in different formats, making it easy to distribute those results to the proper systems for proactive blocking.

Conclusion

Dealing with large databases of data can be difficult, but iDefense has identified a way that not only assists with this difficulty but that also provides analysts with additional valuable information. The case study that this article discusses would have normally taken an analyst hours, if not days, to identify the most useful information from the "itsec.eicp.net" infrastructure; instead, Maltego handled the data in a matter of minutes. Additionally, all information obtained from such research could easily be exported into a report or comma separated value (CSV) list that could then be shared among other teams or added to iDefense products like threat Indicators.

iDefense uses Maltego to achieve its visual data analytics, but other platforms also exist for doing so. The benefit to using Maltego over some of the other products on the market is the ability to query multiple custom data sources and plot the results on an interactive graph that could later be exported and saved. Using visual data analytics, it is possible to significantly reduce the amount of time spent looking for leads and instead spend that time researching more-important information. iDefense will continue to add additional data transforms and use the Maltego platform to enhance its intelligence capabilities.