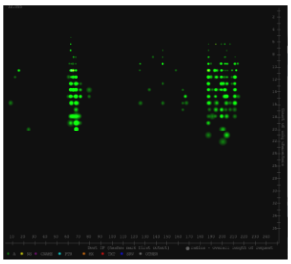
Armatum Networks
Management consultants for information security officers

dns part ii: visualization

Posted on February 13, 2009 by Jeffrey J. Guy



Typical view of DNS visualization

The commonality between science and art is in trying to see profoundly – to develop strategies of seeing and showing." — Edward Tufte

As I wrote in a study of DNS, current tools don't detect DNS tunnels. After opening the post ranting about static signatures, I recommended alerting on any hostname request longer than 52 characters or with more than 27 characters. *(ahem)* It's fair to call my recommendation a signature.

Static signatures are great at identifying an already-known attack. A signature allows you to search immense data for specific details, but the precision provides little awareness of your network traffic in general. By the measure of the tools they've provided, few in industry appreciate the value of context for decision making.

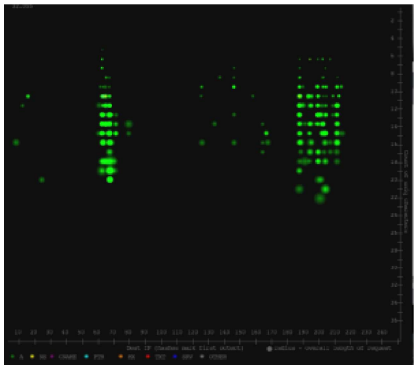
After completing the DNS traffic analysis, I was still unsatisified. I studied the statistical outliers, but it was too hard to understand the nuances, too hard to put those outliers in context with the normal traffic. My eyes were crossing studying lines of text, endlessly cut-ing, grep-ing, sort-ing and uniq-ing to manipulate the text. meh.

Visualization

Using processing, I graphed 4 characteristics of each request and displayed them in a 2d x-y grid in real-time.

- x-axis: destination IP
- y-axis: character count
- radius: hostname length
- colour: request type

Areas with multiple requests increase in intensity and become white-hot as new types appear, so rate indirectly becomes a 5th characteristic.

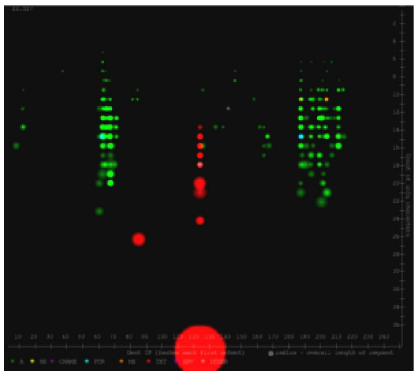


Typical view of DNS visualization

I started analysis focused on DNS tunnel detection, so the demonstration is not complete until we see DNS tunnel traffic. Using dns2tcp, I setup the DNS client/server, ssh'ed into my own machine over DNS and captured the result.

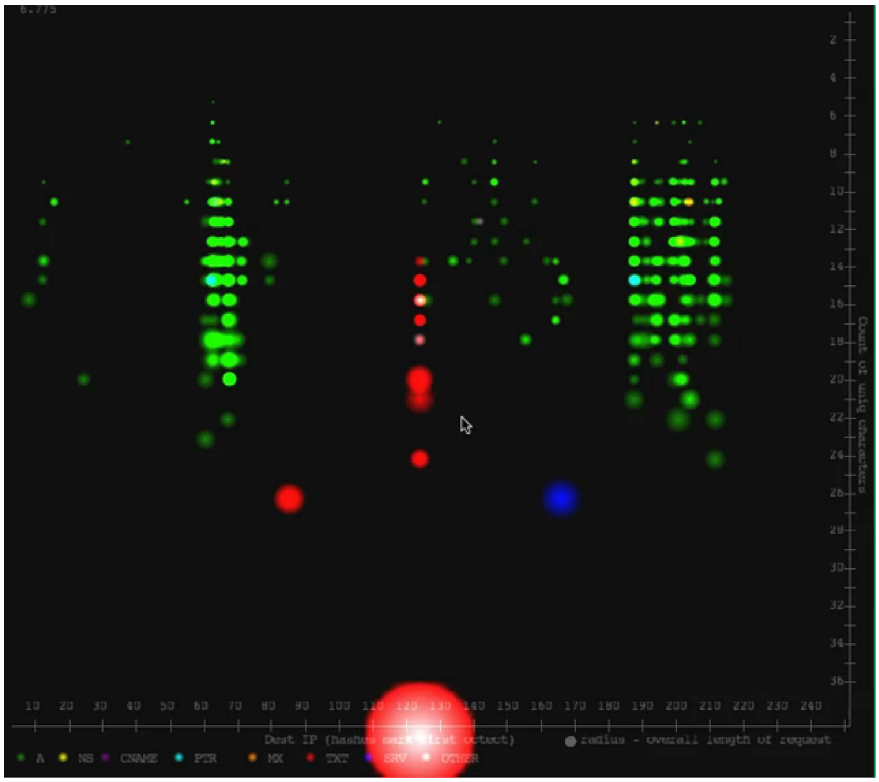
```
guy:~ guy$ ssh -p 2200 guy@127.0.0.1
Password:
Last login: Sat Feb  7 12:53:01 2009 from localhost
guy:~ guy$ ls -laRtp
...^C
```

For a realistic view, I put the DNS tunnel traffic in context with the sample data from the applet above. This is real DNS tunnel traffic in context with DNS traffic from a real network perimeter.



DNS Visualization with DNS tunnel traffic included

Pictures are a poor representation of the live animation. Playback the video demo. (The base32'ed TXT record request you'll see was in the real capture.) The data is a sanitized sample of 1800 requests from the same dataset discussed before. DNS tunnel traffic has been artificially injected.



Thoughts...

I studied the same 97,000 requests for several weeks before graphing the results. Even after weeks of study, I learned even more about the dataset in just a few minutes.

My graph layout went through several iterations. Some good things about where I ended up:

- *DNS request anomalies:* Abnormal traffic stands out naturally, without any threshold or static values. The DNS tunnel traffic is startlingly abnormal. The base32'd TXT record requests to smupdate.net and the TXT record requests to mac.com are less so, but still stand out as unusual. Neither are suspicious, but both are likely indications of a security policy violation.
- *SRV records:* The SRV records you see are not for the organization. Those were answered by the internal DNS server and did not get forwarded. Every SRV record collected is a laptop from another company's domain. Attaching a non-company laptop is aganist the company security policy, but it's obviously happening. In the course of the hour-long capture, there were SRV record requests for 4 other domains.
- *Rate consolidation:* If you refer to the hostname distribution analysis in a study of DNS, you will see the huge spike of 5,000ish requests for myspace records on Limelight Network's CDN. Where they overwhelmed the distribution chart, they have been nicely consolidated in the graph around (65, 20) into a single bright greeen area. If you hover over the point, you'll see the hostnames under your cursor shoot off the screen. I stripped out the sorbs and PTR lookups; they were equally nominal when put in context.
- *Destination context:* The heavily-populated strips coincide with servers located in North America, consistent with the organization's clients. While we have significant data occlusion in these areas, any activity in the the lesser-tracked corners of the Internet naturally stands out more.

The data representation is not perfect:

- *Rate:* The rate representation does not have much fidelity. Two is distinguishable from one and ten is distinguishable from two, but two hundred is not distinguishable from ten. In the case of the myspace/llwnd anomaly, this is perfect. But even though we've significantly lowered the noise floor, an attacker could still hide in the noise at the cost of exfil bandwidth. This is mitigated by "flashing" new requests, but that technique only applies to real-time analysis. ***UPDATE:*** *More recent research has identified another characteristic that would be of use: count of hostnames per domain. Any DNS tunnel will have one unique hostname per request (as long as there is a data payload), while most domains will have just a few. A good update to this research would be to include this metric, it would also reduce the of the Rate gap.*
- *Window of analysis*: In it's current form, an analyst must watch the console at all times. If the request data was maintained in persistent storage instead of temporarily stored in memory, someone could view arbitrary time windows. The same visualization scales to represent thousands of requests without much visual clutter.
- *Presumption of capture location*: by giving the destination IP such predominance, I've limited the location your collector can sit to outside your external DNS server.

If I were responsible for managing a network perimeter day-to-day, I would want something like this to monitor my DNS traffic. Unfortunately, industry has shipped very few tools to get this kind of insight. ***Admins***: ask industry for better traffic monitoring tools. Be smart enough to use them and give good feedback. ***Vendors***: ignore the media-generated fear of day from the latest malware author to grab headlines and study the adversary. Network security is not a boring support function, but a tactical engagement. The ramifications of a tactical mindset are immense, and we need the industrial base to adopt it.

I appreciate criticisms, suggestions or screenshots of your own network traffic!