Analysis of Internet-wide Probing using Darknets

Alberto Dainotti, Alistair King, Kimberly Claffy
CAIDA, University of California, San Diego
La Jolla, CA, USA
{alberto,alistair,kc}@caida.org

ABSTRACT

Recent analysis of traffic reaching the UCSD Network Telescope (a /8 darknet) revealed a sophisticated botnet scanning event that covertly scanned the entire IPv4 space in about 12 days. We only serendipitously discovered this event while studying a completely unrelated behavior (censorship episode in Egypt in February 2011), but we carefully studied the scan, including validating and cross-correlating our observations with other large data set shared by others. We would like to extend these strategies to detect other large-scale malicious events. We suspect the fight against malware will benefit greatly (and perhaps require) collaborative sharing of diverse large-scale security-related data sets. We hope to discuss both the technical and the data-sharing policy aspects of this challenge at the workshop.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network Monitoring;
C.2.5 [**Local and Wide-Area Networks**]: Internet

General Terms

Measurement, Security

Keywords

Darknet, Network Telescope, Botnet, Scan, Stealth, Probing

1. MOTIVATION

Recent analysis of traffic reaching the UCSD Network Telescope (a /8 darknet) revealed a sophisticated botnet scanning event targeting SIP servers (UDP port 5060), which we named "*sipscan*" [1]. The sipscan lasted from January 31 to February 12, 2011, and over these 12 days generated about 20 million probes from 3 million distinct source IP addresses; we later proved the Sality botnet was responsible for this scanning behavior [3]. Because the scan targeted a service running on UDP, each probing packet was carrying payload, which we used to extract a signature of the scan. Figure 1 shows (dashed blue line) the rate of UDP packets from the sipscan reaching our darknet.

We detected this malicious sipscan event serendipitously while studying the Egyptian Internet outage of January/February 2012 [2]. We noticed that after connectivity was restored, the amount of traffic from Egypt captured by our darknet was much larger than before the outage. The sipscan started during the Egyptian outage, making impossible for Egyptian hosts infected by the Sality botnet to communicate with the botmaster until Internet connectivity was restored. We later found that Egyptian IP addresses were heavily contributing to the sipscan, noticeably increasing the amount of unsolicited traffic from Egypt. Since we only accidentally noticed this activity while studying another phenomenon, we believe other stealth probing behavior may also be going unnoticed. The open question is: *how can we automatically detect such behavior?* Our experience studying this botnet highlights the potential power of collaborative data sharing to support automated discovery of behavior that is intentionally dissipated across the entire Internet address space to avoid detection.

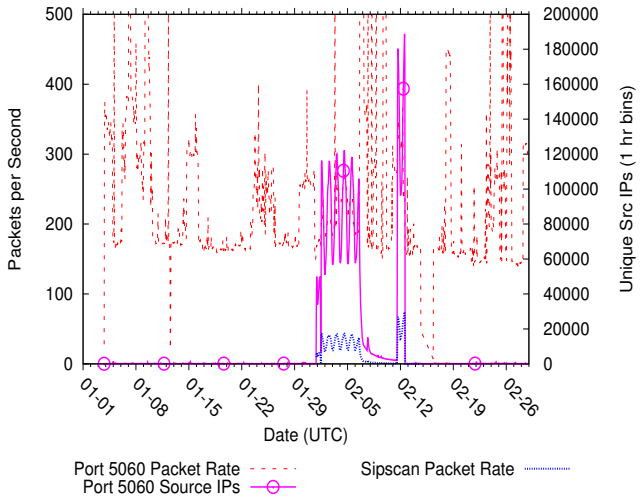

Figure 1: Finding the signal in the noise. The dashed red line shows the packets per second, in 1 hour bins, of UDP packets arriving on port 5060 observed by the UCSD Network Telescope. The dashed blue line is the rate of the subset of UDP packets to port 5060 that were identified to be a part of the sipscan. The solid pink line with circles is the number of distinct source IP addresses seen in UDP packets to port 5060 per hour.

2. ANALYSIS METHODS

In [1] we analyzed details of the event and how it was orchestrated by the botmaster to stealthily scan the entire Internet IPv4 address space using a combination of techniques not previously doc-

umented in research literature. Several aspects of the orchestration contributed to its stealth: (i) large bot turnover; (ii) little overlap in targets (iii) a sequence of target addresses that incremented in reverse-byte order (and took almost 12 days to cover each /24 IPv4 network). Our analysis required processing and visualizing high volumes of data, using multiple display methods in parallel to confirm or discover certain properties. We used a Hilbert-curve animation to understand and verify the covert reverse-byte incrementing of the target IP address scan. We projected the 16 million addresses sequentially probed in our /8 darknet onto a two-dimensional space using the Hilbert fractal curve to layout IP addresses [5], and generated an animation to illustrate the progression of the scan. We also geolocated the source IP addresses to simultaneously animate the botnet's scanning behavior on a worldmap.

Our analysis revealed insights into the properties of a large botnet, as well as an indicator of trends in botnet scanning behavior. We suspect the same type of analysis could reveal other patterns and trends of malicious stealth behaviors. Our ongoing work is on devising strategies for the detection and identification of similar events by focusing on separating traffic per port number and on observing the rate of distinct source IP addresses in fixed time bins. Figure 1 shows that, in terms of packet rate, traffic from the sipscan is hidden in the rest of the Internet background radiation (IBR) on port UDP 5060, which has large oscillations. However, the number of distinct source IP addresses hourly (a metric also used in [4]) starkly shows emergence of the sipscan behavior.

It is well-known that the nature of IBR is not uniformly distributed among different /24 networks. While the signal generated by the sipscan scales linearly, i.e., it is 2^{16} times smaller in a /24 subnet than in a /8, the background noise from which we must extract this signal does not necessarily scale in the same way. Thus, the resulting signal-to-noise ratio may not be amenable to change-point detection using a threshold-based approach. For example, in a /24 darknet, a rate of 0.9 distinct IPs per hour (that is, 256 IPs in 12 days) is difficult to detect when mixed with bursts of multiple packets from different sources, e.g. backscatter, misconfiguration, localized scanning. We are investigating the scaling properties of IBR across large darknets, as we try to automate detection of Internet-wide probing events on any transport-level port.

Even with a large darknet, a change-point detection approach may fail to detect phenomena targeted at popular ports (which SIP ports are not yet). For example, the number of distinct source IPs per hour observed at the UCSD Network Telescope is currently around 25,000 on port TCP 80 or 96,000 on port TCP 445 (publicized during the Conficker episode but a target of scanning activity for many years). These values are high enough to drown any incoming signal, rendering a threshold-based approach ineffective at detecting a sipscan-like probing event even on a /8. The difficulty with extracting relevant signal from such a diverse and high-volume source of noise motivates our current efforts to develop an approach that efficiently correlates information across multiple large darknets, such as analyzing the percentage of common source IP addresses observed at different vantage points.

3. CROSS-VALIDATION

To corroborate that the sipscan was targeting the entire IPv4 address space we looked for its presence in datasets provided by three other research groups/projects. First, aggregated data from the DShield project showed a large increase in the number of distinct source IP addresses on port 5060 during the 12-day interval of the sipscan activity, although their publicly shared data is too heavily aggregated to support finer-grained correlation. Second, the MAWI-WIDE project shares anonymized daily pcap 15-minute packet header traces from a trans-pacific link. Truncation of the payload in the traces prevented us from identifying sipscan packets using our payload signature, but we extracted an effective flow-level signature based on features of sipscan packets and flows uncommon to other (legitimate or malicious) SIP traffic. MAWI-WIDE's anonymization scheme preserves network class membership, which allowed us to prove that the sipscan was targeting several /8 networks. Third, Eduard Glatz and Xenofontas Dimitropoulos from ETH-Zurich University used our flow-level signature to identify the sipscan in netflow logs from their production network. Their traces have original source IP addresses, allowing us to collaboratively infer other properties of the scan (a work still in progress).

4. CONCLUSIONS

This study illustrated several lessons about Internet-wide probing: (i) attackers can make use of many resources (e.g. botnets made of millions of bots); (ii) they can be patient, easily trading task completion time for covertness; (iii) the level of sophistication in stealth behavior is growing (e.g., multiple source IPs, low overlap in targets, low re-use of the same source IPs, increments in reverse byte order in the progression of target IPs). Assuming unavailability of payload, these aspects challenge detection and identification of Internet-wide probing events. We are now working on devising mechanisms to automate their detection, which we could build on the analysis of *large* darknets and correlate information from multiple observation points in the address space.

As a data point, the UCSD Network Telescope collects approximately 3TB of data every month. Not only is processing the data computationally expensive, but even knowing what information to extract is a challenge. We are designing and developing an extensible tool, Corsaro, to efficiently analyze data collected by darknets and produce output that can facilitate research and collaboration. We are implementing our prototyped detection technique as a plugin of this platform.

Both our experience with the investigation of a specific case of a large-scale malicious event, and our considerations on devising automated techniques for the detection of this kind of events, motivate collaborative sharing of diverse large-scale security-related data sets. We hope to discuss both the technical and the data-sharing policy aspects of this challenge at the workshop.

Acknowledgements

UCSD network telescope operations and data collection, curation, analysis, and sharing is provided by NSF CRI CNS-1059439, DHS S&T NBCHC070133 and UCSD.

5. REFERENCES

[1] Alberto Dainotti, Alistair King, Kimberly Claffy, Ferdinando Papale, and Antonio Pescapé. Analysis of a "/0" Stealth Scan from a Botnet. In *ACM SIGCOMM Internet Measurement Conference*, 2012.

[2] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 1–18, New York, NY, USA, 2011. ACM.

[3] Nicolas Falliere. A distributed cracker for voip. `http://www.symantec.com/connect/blogs/distributed-cracker-voip`, February 15 2011.

[4] Zhichun Li, A. Goyal, Yan Chen, and Vern Paxson. Towards situational awareness of large-scale botnet probing events.

Information Forensics and Security, IEEE Transactions on, 6(1):175 –188, March 2011.

[5] Duane Wessels. IPv4 Census Hilbert Map, 2007. `http://www.caida.org/research/id-consumption/census-map/`.