Improving Security Visualization with Exposure Map Filtering

David Barrera, Mansour Alsaleh, Paul Van Oorschot
Carleton Computer Security Laboratory (CCSL)
Carleton University, Ottawa, Canada

Summary

Flow based analysis of network traffic is commonly used to analyze and understand security-related events. Graphical analysis helps analysts detect patterns or behaviors that would not be obvious in a text-based environment. The growing volume of network data generated and captured makes it increasingly difficult to detect stealthy network attacks. We propose a network flow filtering mechanism that leverages the exposure maps technique which in turn selectively reduces the traffic for the visualization process according to the network services being offered. This allows focus to be limited to selected subsets of the network traffic, for example what might be categorized (correctly or otherwise) as the potentially malicious portion. In particular, we use this technique to filter out traffic other than that from sources that have learned something from the network in question. We evaluate the benefits of our filtered visualization technique on different visualizations of network flows. Our analysis shows a significant decrease in the volume of network traffic that is to be visualized.

Description

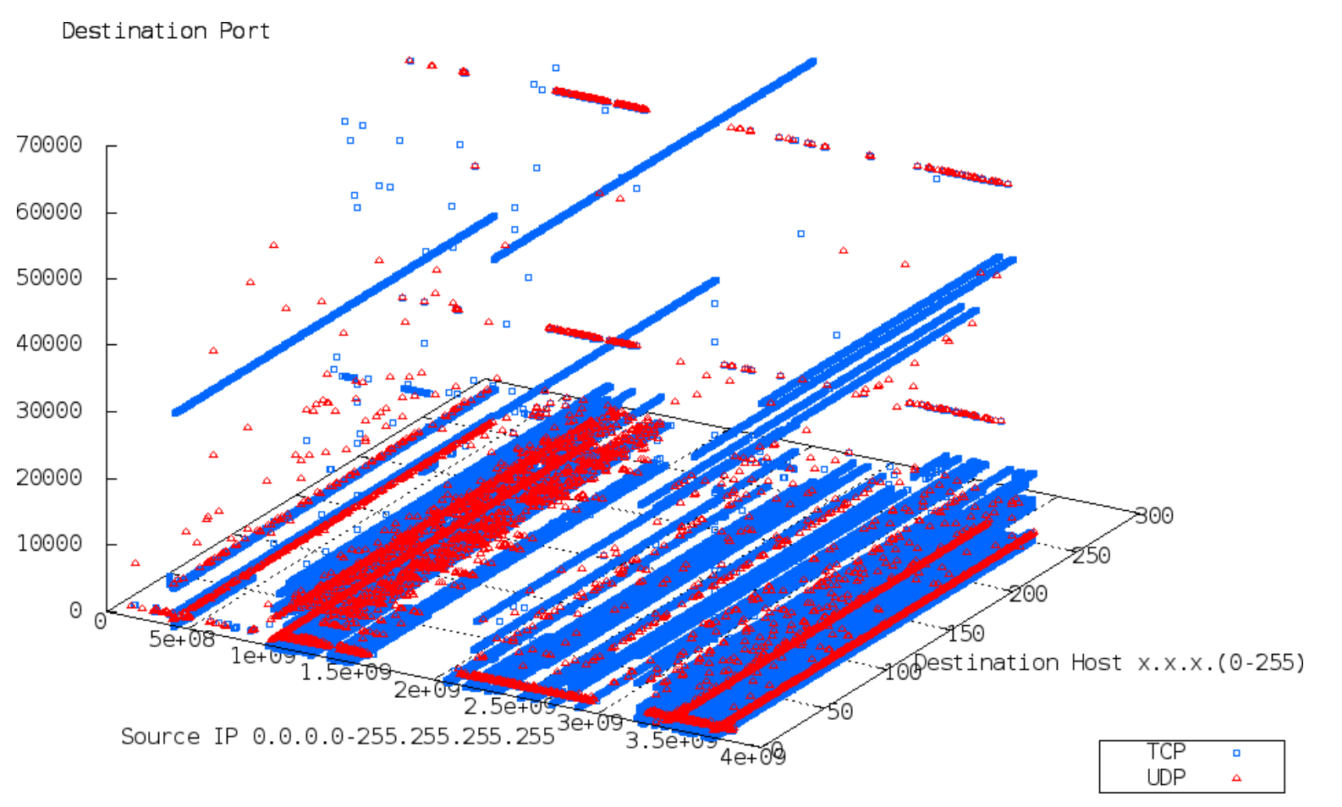
- Network security event monitoring is a time consuming and complicated process where security analysts are overwhelmed by massive amounts of audit log data that ideally would be analyzed for possible threats or malicious behavior.
- Visual representation of network data, as opposed to textual representation, can help in analyzing a vast amount of data in a shorter time (it takes humans much less time to recognize specific information or patterns in a picture than it would to detect the same information in text).
- The network exposure maps (NEM) is built over a training period during which outgoing TCP flows containing SYN-ACK flags are observed and recorded. Every host that was seen responding with SYN-ACK flags is added to the NEM.

Network Exposure Map (NEM)

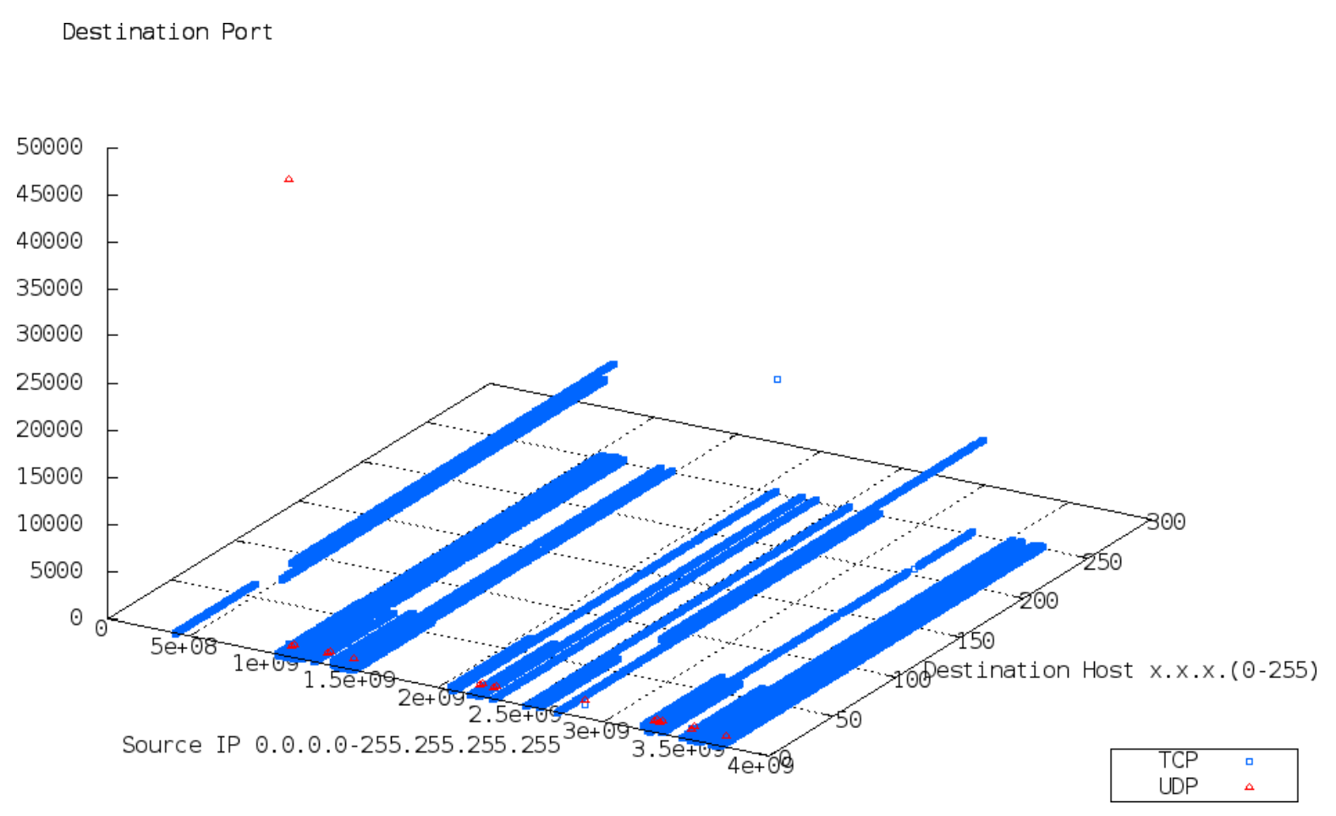
Source IP	Port	Protocol
x.x.x.11	25 (SMTP), 631 (IPP), 993 (imaps)	TCP
x.x.x.11	53 (DNS)	UDP
x.x.x.13	22 (SSH), 80 (http), 443 (https)	TCP
x.x.x.13	53 (DNS)	UDP
x.x.x.58	22 (SSH)	TCP

- The proposed visualization technique leverages network exposure maps to help filter raw network data, in order to focus visualization efforts on data given a preliminary classification of unknown or malicious traffic:
  - Network traffic from sources going to only closed ports or nonexistent internal hosts is assumed to be harmless and thus does not need to be visualized.
  - Network traffic from sources that go only to services offered at open ports in the network internal hosts is filtered out by using exposure maps due to its presumed legitimate nature.
  - This reduces the volume of the network traffic that is presented as requiring investigation as possibly malicious traffic.
  - Consequently, applying simple visualization techniques on only the filtered network traffic yields a much sparser dataset with less (presumably irrelevant) noise and data. It also helps network analysts better correlate malicious events and discover coordinated external hosts.

Visualization Process

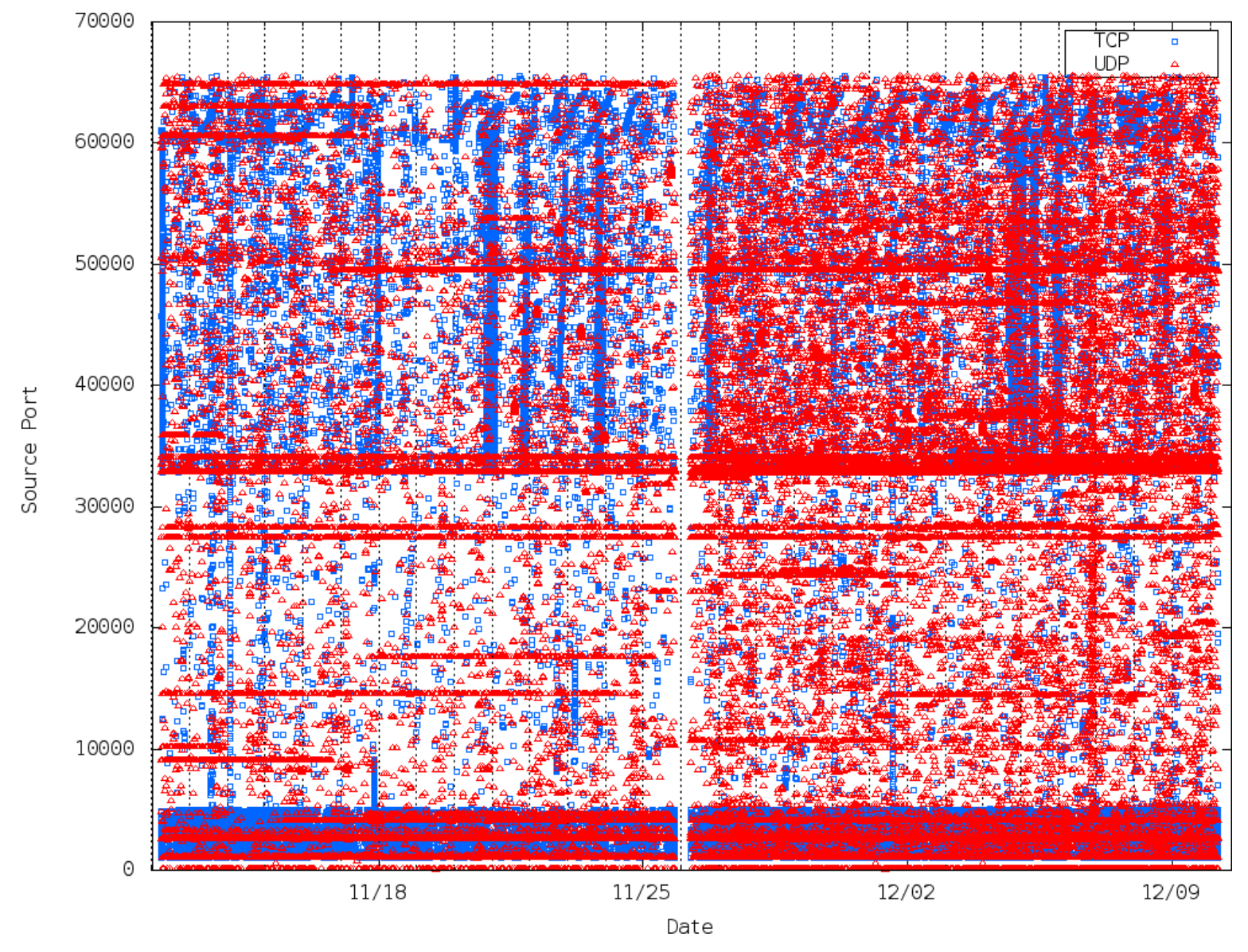


1.A Before Filtering

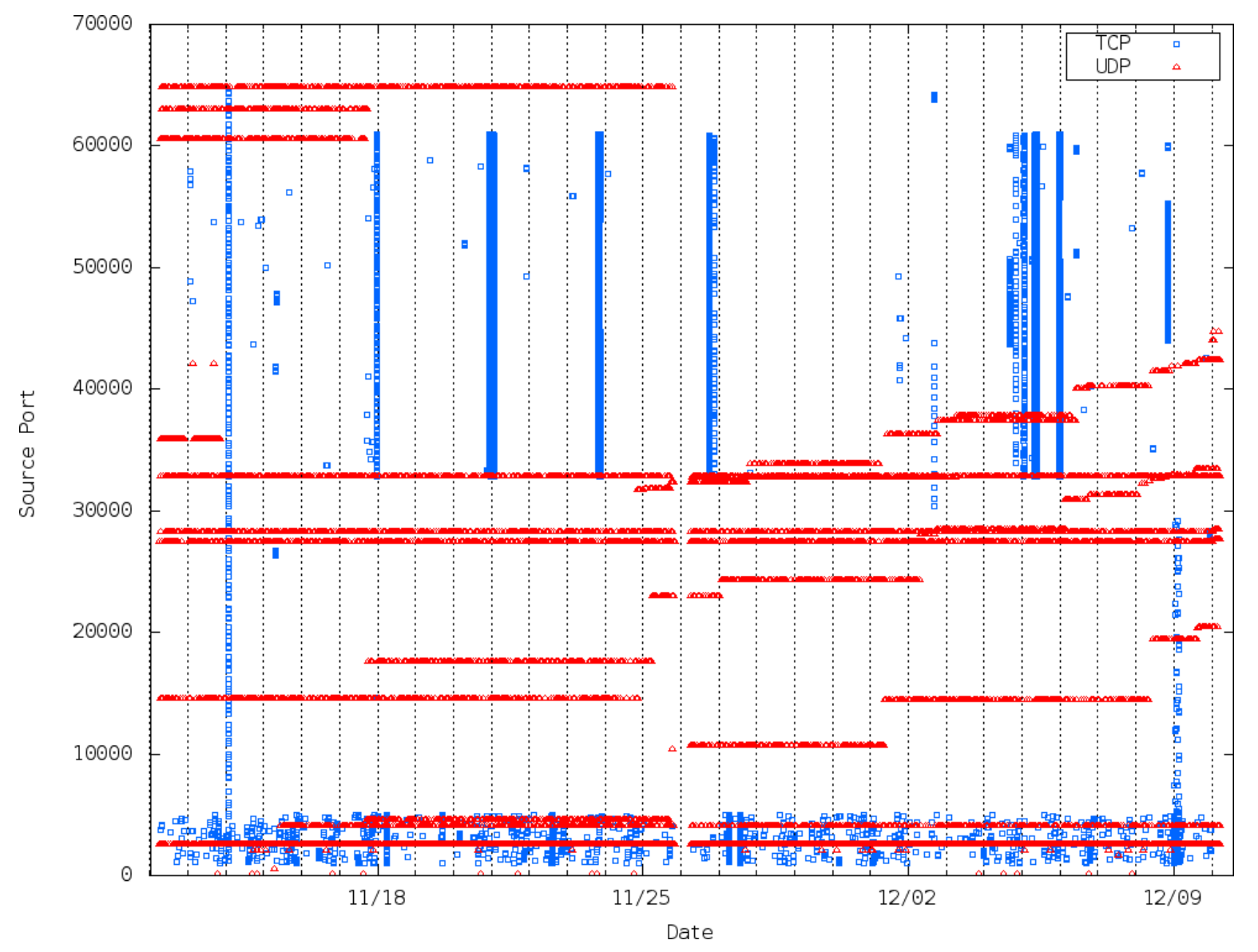


1.B After Filtering

The figure above presents the full source IP address (plotted as an integer from 0 to approximately 4.2 billion), the target destination host and the destination port. The graph shows a high number of source IPs probing a single port on the entire class C destination network and dense areas around low-order ports. It might be difficult for a security analyst to know which horizontal scan to select (probing a single port on all destination IP addresses as noted by bottom-left to top-right diagonal lines) for analysis in Figure 1.A, but most (if not all) horizontal scans in Figure 1.B likely reveal some type of malicious activity. The data that was removed from Figure 1.A was classified as non-threatening by the exposure map filtering.

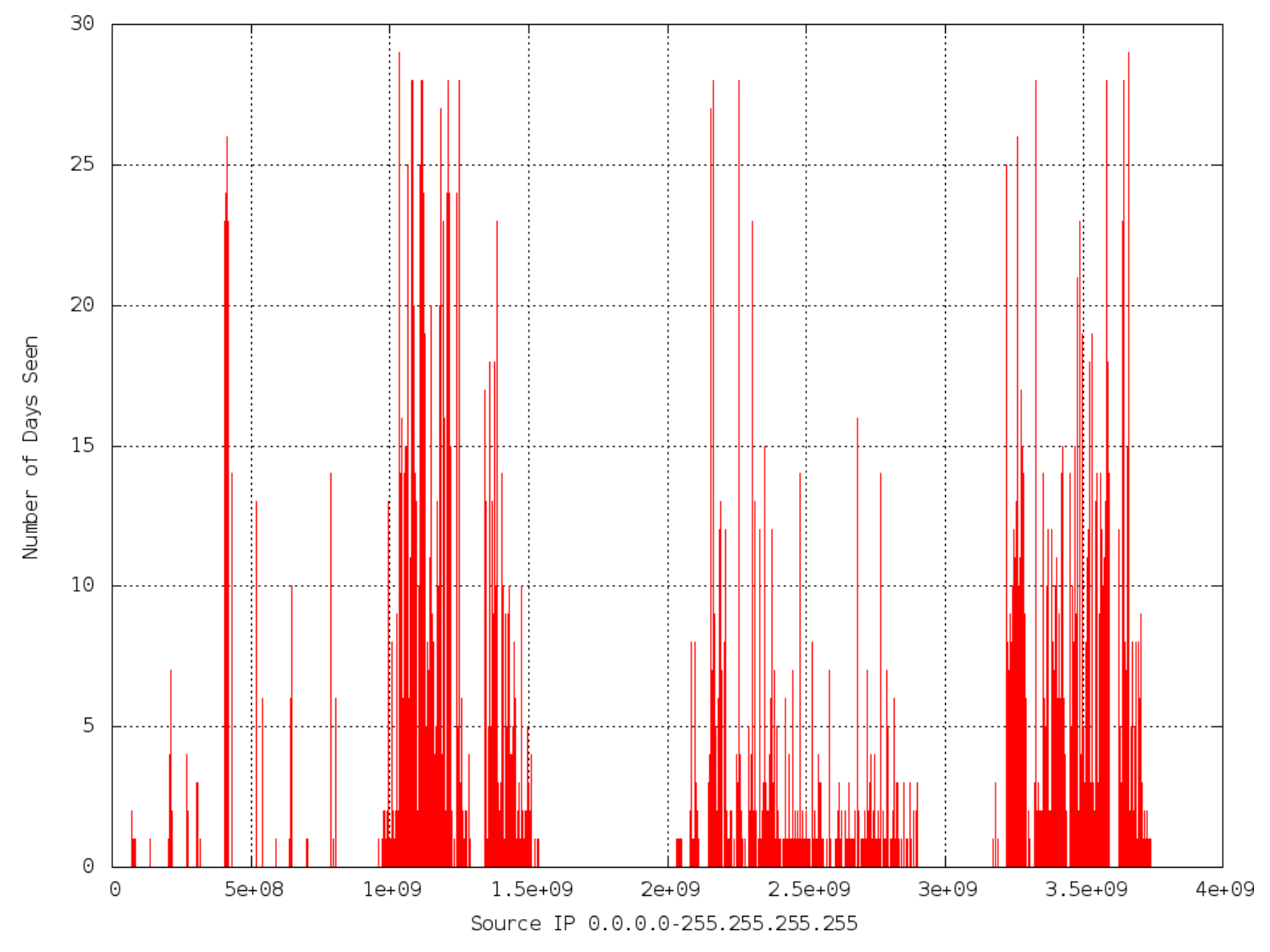


2.A Before Filtering

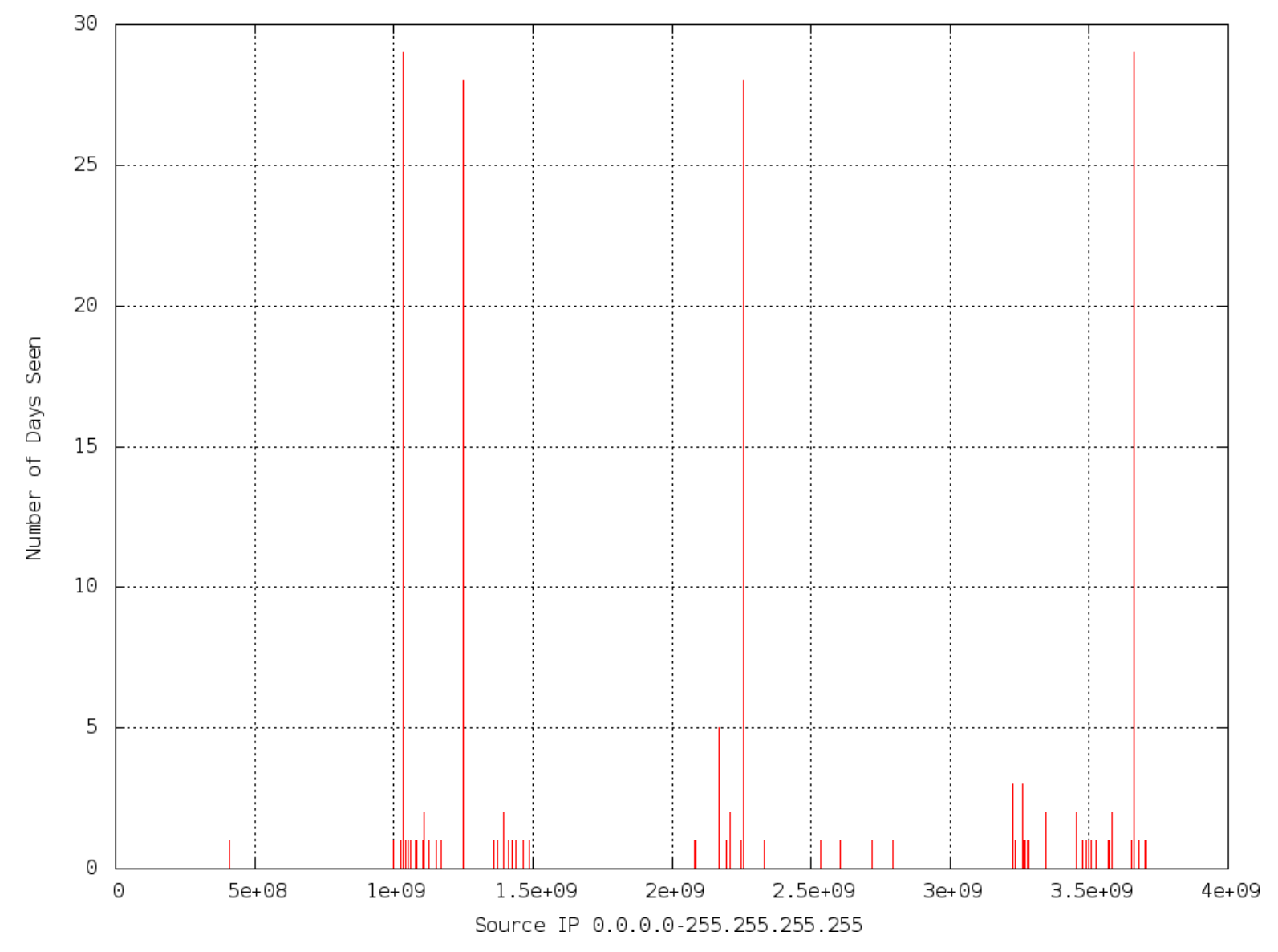


2.B After Filtering

This figure plots the source port used for incoming flows over the entire capture period. Source ports might give insight as to what operating system is being used, or what type of scan is being performed. Source ports are usually allocated randomly which explains the high large amount of clutter in Figure 2.A. On the other hand, Figure 2.B shows a clear view of the source ports being used (with some patterns). For example, we can notice the 4 evenly spaced vertical lines in the center of the plot. When we query the database, we find 4 distinct IP addresses in separate class A networks that attempt to bruteforce SSH logins. Upon further investigation through database queries, each one of the 4 sources is found scanning the full destination class C for any hosts who offer the SSH service, then focus the attack on the only 2 hosts who respond.



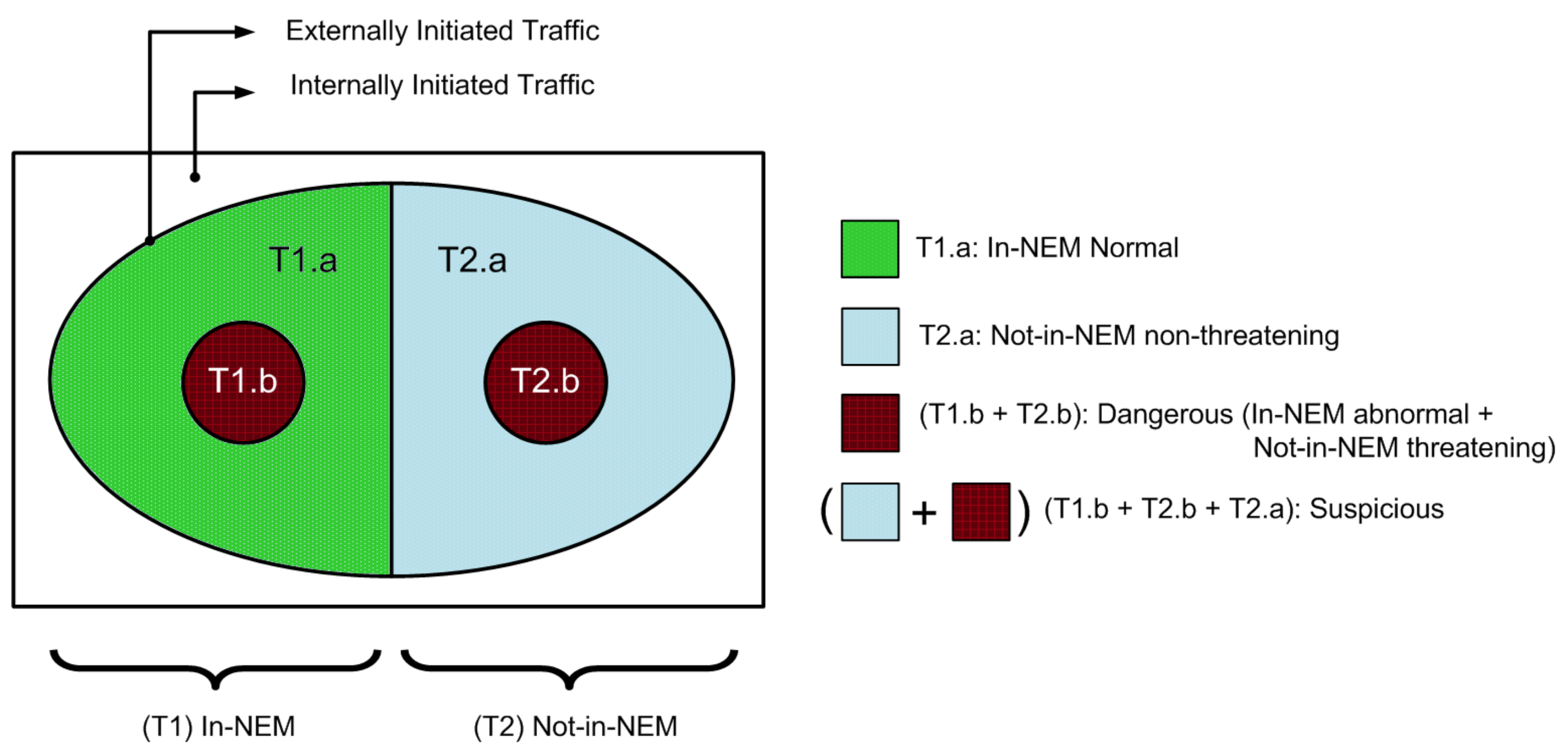
3.A Before Filtering



3.B After Filtering

The figure above gives the number of days a specific source IP address probed the network. IP addresses that probe the network repeatedly might be considered for further actions such as blocking or logging. In Figure 3.A, a large number of source IP addresses attempt connections over 15 times in the 4 week capture. This filtered view (Figure 3.B) leaves the analyst with far less information to analyze by removing legitimate traffic as well as probes from source IPs going to only closed ports. In this view we see either sources that return almost every day (4 peaks), or sources that come back less than five days.

Partitioning with Exposure Maps



🟩 T1.a: In-NEM Normal
⬜ T2.a: Not-in-NEM non-threatening
🟥 (T1.b + T2.b): Dangerous (In-NEM abnormal + Not-in-NEM threatening)
(⬜ + 🟥) (T1.b + T2.b + T2.a): Suspicious

Network flows are partitioned into logical tables T1 and T2 using the NEM:

Table T1: In-NEM. This table contains flows destined to a host/port combination offering an authorized service (i.e., to an authorized open port in the local network). This table is also logically partitioned into two sub-tables.

T1.a: In-NEM normal. This table contains flows that are considered ordinary, since their source IP addresses have only attempted connections to authorized services offered by the network in question (i.e., destined to an authorized open port).

T1.b: In-NEM abnormal. This table contains flows initiated by source IP addresses that also have flows in T2. We label these flows 'suspicious' because normally, a host does not attempt connections to closed ports while also accessing legitimate services.

Table T2: Not-in-NEM. This table contains flows destined to a host/port combination for which no authorized service is offered (i.e., closed port). This table is also logically partitioned into two sub-tables.

T2.a: Not-in-NEM non-threatening. This table contains flows in T2 and whose source IP addresses have no flows in T1. Exposure map filtering assumes these connection attempts are not a significant threat to the target network since sources, all of whose probes have been to closed ports, have not learned what is considered significant information from the target network (i.e., have not learned what services are offered).

T2.b: Not-in-NEM threatening. This table contains flows in T2 and whose source IP addresses also have flows in T1. Thus, the source IP address of these flows have queried both legitimate offered services and closed ports.

Table T3: Suspicious. This table includes all flows in T2 (T2.a and T2.b) plus T1.b. We call this 'suspicious traffic' because these source IP addresses have probed at least one closed port in the network.

Table T4: Dangerous. This table includes all flows in T1.b plus T2.b. This represents traffic from IP sources that probed at least one closed port and also attempted to connect to an open port. According to exposure maps, these are more likely to represent malicious flows since these IP sources, might attempt to send exploits to the discovered open ports.

Future Work

- Study more visualizations for network traffic security-related analysis and compare the output before and after applying our filtering mechanism, to test and demonstrate the effectiveness in segregating malicious traffic from harmless traffic.
- Use datasets for larger networks and evaluate the scalability of our mechanism.
- Filter network flows on the fly and generate visualizations of suspicious network flows in real-time.
- Explore if the proposed filtering technique improves existing visualization tools.
- Develop an interactive visualization tool using the proposed filtering mechanism.