White Paper: ScienceDMZ and IPS Bypass

Hyojoon Kim
Princeton University

Abstract

ScienceDMZ is becoming the *de facto* infrastructure design pattern for enabling faster data transfer for science flows. However, there are important questions and concerns that left unanswered. For example, it is unclear how to provide tighter security measures to the flows that are once tagged as "safe" flows. Realizing that SDN, the underlying technology often used in ScienceDMZ operations, is underutilized compared to its capabilities, we focus on research directions that aim to better utilize the flexibility and programmability of SDN.

1 Research Direction

Automatic detection of safe traffic. Current ScienceDMZ solutions make a strong assumption that the network will be aware of all science flows beforehand, mostly by asking them to submit a request to be prioritized (and/or bypass security appliances). There are several major problems here: (1) adds an additional step for researchers of submitting a formal request, (2) requires a researcher to submit a **correct** request who will likely have less or no background knowledge about the network and traffic, (3) trust is given to the flow indefinitely long without automatic detection of end-of-transfer.

Princeton University is working on a project to automatically identify safe traffic as well as large science flows by collecting and analyzing data from various types of data sources, including NetFlow, packet dumps, IPS/IDS event logs, and logs from data transfer tools like Globus. Through this project, we aim to build a system that will correctly identify unreported safe science flows and automatically reconfigure the network to prioritize and redirect such flows to bypass security appliances. The project also aims to steer or shunt traffic that does not have to go through security appliances to lower the load on campus IPS device. (*e.g.*, inspection of first few packets identifies the flow as malicious, thus it is needless to inspect further).

Consistent security guarantees. ScienceDMZ redirects science flows so that they bypass various security appliances. However, the idea of allowing inspection-free traffic makes network operators nervous, deterring the deployment of such infrastructure. We aim to build a system that allows traffic to bypass firewalls without losing much security robustness. We will investigate a design of deploying a Bro [1] cluster that will also consistently analyze "trusted" traffic and revoke its trusted state when necessary to guard against a case where end-hosts of such flow gets unexpectedly compromised and become malicious. Closely related work includes Indiana University's SciPass project [3].

Optimizing data transfer schedules. Intelligent scheduling is unnecessary if few scientific data transfers occur at a time. However, it is likely that multiple scientific data transfers will overlap, each with different latency, size, frequency, and time requirements. Naive scheduling of such data transfers (*e.g.*, batching or round robin) will likely result in an unoptimized, ineffective utilization of the network bandwidth, possibly missing many desired requirements of such data transfer requests. The suboptimal usage of network will exacerbate as the number of simultaneous transfers increases.

Princeton University plans to tackle this problem by building a mathematical model using data transfer requests and network status information, and formulating an optimization problem that will maximize network bandwidth utilization, maximize the chance of satisfying all transfer requirements, and minimize overall transfer latency.

2 Infrastructure Requirements

Princeton University is in the process of building a ScienceDMZ testbed. The testbed diagram can be found at [2].

References

[1] The bro network security monitor. https://www.bro.org/. (Cited on page 1.)

[2] Sciencedmz testbed. https://goo.gl/Fgxz5a. (Cited on page 1.)

[3] Scipass. https://globalnoc.iu.edu/sdn/scipass.html. (Cited on page 1.)