Applying Internal Traffic Models to Improve Identification of High Fidelity Cyber Security Events

Abstract

Effective Security Operations throughout both DoD and industry are requiring and consuming unprecedented amounts of telemetry and intelligence to both protect assets from attack and also identify anomalous events. While the cyber intelligence community is vast, comprising government and commercial feeds, the notion that internal traffic modeling may complement existing intelligence feeds and further lend credibility toward identifying high fidelity security events represents a potential cyber security force multiplier. Existing security controls, telemetry aggregators, and analytics engines are pushing the boundaries of security operations capabilities. Industry experts, including Bruce Schneier[1], have recently noted that there remains a necessity to continue the evolution of security tools in order to aid operators in the detection of anomalous traffic. One promising approach to respond to this challenge is to utilize internal traffic modeling as a complimentary layer of identification for anomalous security events that must be elevated for operator / security analyst attention. If such traffic models were capable of corroborating suspicious events and reducing the criticality of more common events, they could create an increasingly sophisticated security operations environment that further reduces the haystack to a more manageable volume.

Security Operations Is Not Easy

In the current threat environment, Cyber Security is extremely challenging. Simply put, security practitioners play from behind on a daily basis. To say that protecting assets and detecting threats is a struggle isn't something that will receive much rebuke. What was once an environment of fences, moats, and other perimeter defenses has matured, by necessity, into more holistic approaches to protecting, detecting, and responding to/from attacks. It is commonplace to layer security controls throughout the architecture within an enterprise not only to protect, but also to detect. Long gone are the days of simple detection where an Intrusion Detection System would alert and/or block. Detection today ventures into the Big Data realm of gathering large amounts of security intelligence in addition to the enormous amounts of telemetry numerous devices and applications throughout the environment are generating.

This flood of intelligence and telemetry data often originates from numerous sources. Many security control vendors provide inherent intelligence feeds via subscription with their products. We even see this sort of capability at the consumer level. Take for instance a consumer antivirus software product.

[1] Bruce Schneier. The Future of Incident Response; Schneier on Security, Nov. 10, 2014

We're all aware that these products regularly update the DAT files which provide, at the very least, signatures of known malware. This capability extends to the enterprise quite easily, but introduces the challenge of monitoring, or consuming, log and event data from each of the endpoints within the environment. Further consider that layer 7 firewalls, Web and Email Proxy servers, IDS/IPS, and numerous other vendor products all provide regular updates with regard to security intelligence. Another example is the Web Proxy. While updates and feeds to such devices may not provide a signature for malware, what they do provide is a recent (typically within hours or sometimes minutes) evaluation of the threat landscape on the Internet. These updates allow the Web Proxy, if configured to do so, to block attempts from clients to reach known bot nets, phishing sites, and other malicious web sites. Again, extend the volume of Internet traffic that one endpoint can generate and extend that to a typical large enterprise and you have a small sample of telemetry that is fed into Big Data Analytics. Other telemetry sources may be more straightforward such as server event log data, data from a Syslog server, firewall logs, or other security controls. Lastly, security intelligence can come from the community in a manner that allows organizations to make decisions about how they consume this brand of intelligence. Information about IP addresses, domains, and other IT-centric nomenclatures could lead a practitioner to introduce DNS blackholing, Null routes, or other security measures to prevent threats from entering the environment or more specifically to prevent valuable information from leaving their environment.

This overwhelming glut of cyber security telemetry and intelligence has inundated our enterprise security architectures over time and is now presenting a challenge of its own – how do I find the critical threat needle in the large cyber security haystack? In enterprise environments with real-time operations requirements, the issue is even more acute.

On the operations side of security, there are tools that are communicating via common languages, sharing information whether in the form of raw or meta data. Analytics platforms are popping up like daisies as the good guys continue to enhance the microscope to find the needle. Most recently, an open source "OpenSOC" repository has been populated with software to help organizations establish capability to consume and analyze high volumes of data and telemetry. These analytics platforms then begin to whittle down the masses by removing duplicate entries (even if from separate sources). A subsequent pass through the resultant set aims to target well known signatures. These automated actions provide identity of high fidelity security events. Nonetheless, the resultant set of high fidelity events must often then receive attention from an operator or security analyst to further validate the criticality of the event. This is where additional contextual information may lend further credibility to the analytics engine and further reduce the set of high fidelity security events in order to optimize the human interaction within the scope of security operations and, more specifically, incident analysis.

Within the Navy alone, the Navy Cyber Defense Operations Command receives and attempts to adjudicate 10s of millions of cyber "alerts" from its cyber instrumentation infrastructure every day. Sophisticated analytics appliances and teams of analysts struggle to separate threat "signals" from routine cyber "noise". In the end, a small number of events are identified for human analysis – but, larger patterns and events that represent potentially anomalous (and dangerous) activity within the many smaller scope environments that make up the larger whole are often overlooked.

Has Big Data Analytics reached its potential? With certainty, the industry indicates it has room to mature. Where does the industry go next to improve intelligence, telemetry, communication among partners, and/or analytics to mature beyond the current state? With respect to automation and aiding the reduction of events to high fidelity sets, does internal traffic modeling present sufficient value so as to pursue a solution? Finally, what are the constraints around internal traffic modeling in both

enterprise and smaller scale tactical environments and how does introducing such personal intelligence help security operations teams break free from such challenges and present a unique value proposition?

Internal Traffic Modeling As An Intelligence Feed

There are already solutions that identify anomalous traffic and create security events based on non-standard submissions. However, this proposed effort does not focus on identifying anomalies by inspecting packet structure or payload. Simply put, it is conceivable to assume that internal traffic patterns form consistency over time and could lend credibility to existing security events to further flag them as high fidelity or to reduce a high fidelity event to a lower rating.

From a high level, there exist numerous security controls that consider internal traffic flows in one way or another. However, there may be a gap with consideration for an entire traffic profile versus single characteristic consideration such as payload or packet structure. To that end, an opportunity may exist to introduce internal traffic profiling as an intelligence feed. This could be achieved by utilizing four-tuples or five-tuples[2], applying analytics and data normalization, and possibly even introducing the concept of access control matrix referencing. If one were able to establish predictive (and permissive) traffic patterns, it could then extend to standard exchange formats and contribute as a unique intelligence feed. While the majority of tools build intelligence around the entire community, this potential solution remains fruitful only for the inherent environment. The value of this solution is optimized for the smaller scale organization. This approach is ideal for sites with unique and/or very deterministic operational functions (e.g. submarine Broadcast Command Authorities (BCAs)) or afloat platforms with limited or no ability to provide cyber data in real time to a centralized analytics warehouse.

It should be noted that similar modeling exists within the framework of IDS/IPS devices and other security and/or networking tools. However, the primary distinction with the proposed solution is found in the capability to extend a traffic model beyond short-term input, possibly consume dynamic access control policy as a reference, and provide an additional source of intelligence for Big Data Analytics engines. While the event logs from firewalls, routers and switches (ACLs), IDS/IPS, and other security devices may provide a collective snapshot of the environment, an internal traffic model would be able to further validate or invalidate suspicious patterns extending the identification of anomalous communications beyond the IP packet structure and/or payload. Further, this approach compliments the use of host-based security controls or alternative tools that evaluate behavior and detect anomaly. Visibility into the environment is key to understanding communications between assets. Employing both host-based and network-based measures provides a sort of “Visibility-in-Depth” approach. The goal with all collective efforts with data and telemetry and the subsequent automation (analysis) is to make the human interaction with security events both efficient and effective by reducing the resultant set of high fidelity security events that ultimately fall into the lap of an analyst.

High Fidelity Anomaly Identification

Security telemetry, specific to one’s own environment, targeted to aid qualification on high fidelity security events has significant potential value. If automation were able to further reduce the working set of events that required human review, it could reduce the workload and help focus on those

[2] This is a reference to terminology commonly used to define the elements of a connection (source IP, target IP, source port, target port, protocol).

events that are both anomalous with regard to community intelligence, and further anomalous to an internal profile. Take into account the improved assurance this solution could provide by applying local intelligence. No longer would an organization solely rely on the analyst, or overworked watch stander, to apply localized intelligence, but this solution would automate, to some degree, this capability. Zero-day attacks often rely on a 'soft-middle' and leverage known protocols and applications. Traffic profiles might identify an anomalous traffic flow, regardless of the packet structure or payload, and add an additional layer of detection to the infrastructure. In addition to the reduction in workflow, the internal traffic profiling could lend itself beneficial in other areas of the Information Technology Operations, though that and other benefits are not the focus of this effort.

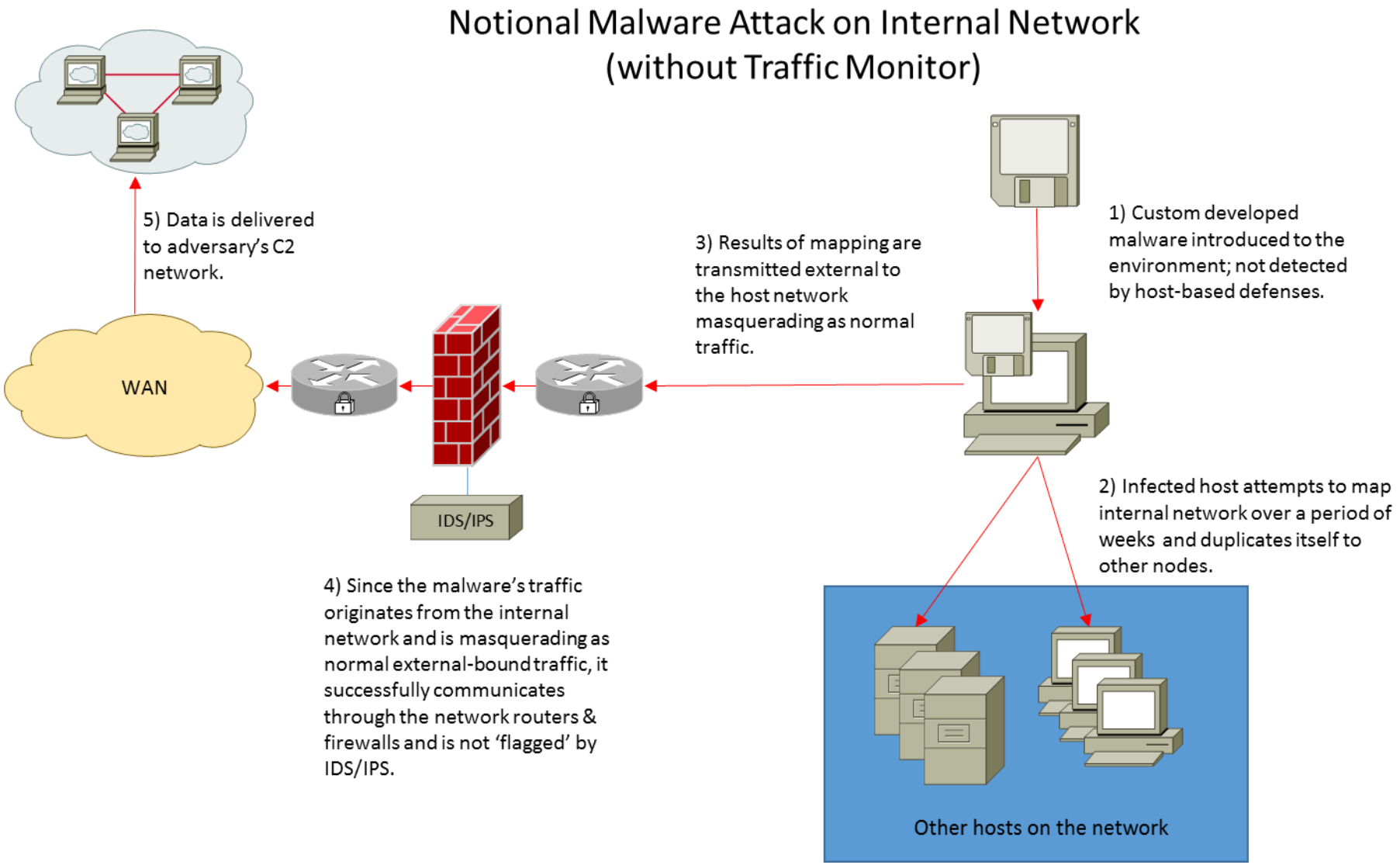
Summary & Subsequent Work

There is most certainly a significant amount of work to determine the efficacy of using internal traffic models to complement existing intelligence feeds within the security operations domain. While perimeter-style security controls remain a necessary tool in preventing attacks on protected assets, the perimeter continues to dissolve. Attacks that do penetrate protective measures often encounter a 'soft middle' of security controls within the infrastructure. Such controls are typically passive detection or accounting (logging) capabilities. These security controls feed telemetry and/or raw intelligence to an aggregator. Thus begins the task of identifying which events require valuable attention from an analyst. Automation entails the majority effort toward identifying high fidelity security events. While increasingly more intelligence may ultimately reach the point of diminishing returns, it seems unlikely we have arrived at such a point. To that end, it may very well be the case that internal traffic modeling is a likely next step in the evolution of security intelligence, and therefore, worth an investment of time to research existing solutions, determine the viability of such customized intelligence feeds, and design a solution that will enhance the identification of high fidelity security events to increase the efficiency and effectiveness with which security analysts evaluate a never-ending work stream.

The diagrams that follow are intended to provide a sample use case that shows network responses with and without an internal traffic modeler. The scenario depicted is as follows:

> *An insider introduces custom developed, targeted malware to the internal network through removable media. The malware is designed to perform reconnaissance on the internal network to discover the architecture, possible vulnerabilities, etc. and report its findings to an adversary C2 node via a standard HTTPS connection (thus masking itself as normal traffic).*

Notional Malware Attack on Internal Network (without Traffic Monitor)



Notional Malware Attack on Internal Network (with Traffic Monitor)