

Network Security Analysis at Internet Speed

Problem:

Network analysts are overwhelmed with the amount of data available. Current analysis systems are not designed for the rapidly increasing data sets created by modern networks. A variety of rudimentary tools are available, but require manual analysis and patchwork, which is labor intensive. Most analysis tools are strictly text based and are missing the benefits of visual analysis. For example, there are tools that read netflow or BGP but don't have the ability to compare that data with attacks and virus outbreaks. Most visualization solutions are 3-D engines that accept limited data types from text based tools and fail to provide useful results.

A Perfect Solution:

A next generation system is needed to advance analysis to the level of providing actionable information. This new system should be able to adapt to the speed and growth of the Internet while also improving analytical capabilities. An application that automates analytical tasks and event correlation can assist analysts by reducing the amount of manual research required. This application should allow an analyst to look at the big picture and match events across disparate data sets, while streamlining the entire process. A next generation system should have both textual and visual components, because patterns aren't always obvious in text reports and visual reports aren't easily fed to other systems for automation. Visually a user can measure quantities, map similarities and identify hidden relationships. A user should be able to start at a high level and drill down or go straight to the results of any query. The application should be flexible and dynamic to allow for data analysis on new and evolving sources. As an example, overlaying BGP data, with traffic flows and attack data can reveal emerging threats and their location in cyberspace and possibly the real world. The following data sets are important for the system to accept: Border Gateway Protocol (BGP), netflow, attack data, malicious data, traceroutes, Internet Protocol (IP) address to country mapping, Distributed Denial of Service (DDoS) information, internal and external routing protocols, data from Regional Internet Registries (RIRs) and Internet Routing Registries (IRRs). Any data that is useful for establishing network awareness is a possible data source. Network cognizance goes beyond the enterprise view. Allowing corporations to see beyond their perimeters and be prepared for traffic that is incoming or outgoing. A system that strives to provide all of these functions begins to look like a network management system, network monitoring system and data warehouse combined.

Understanding the Need:

Current network security challenges can be summarized with the four V's: Volume, Velocity, Variety, and Veracity (precision). Networks of all sizes are presented with these challenges daily. The exploding growth of the Internet has made effective and timely human analysis impossible. Digital Network Intelligence (DNI) is a term the

intelligence community uses to describe any data related to computers and computer networks. To deal with huge amounts of DNI, tools need to sift through this data and provide summaries, correlation and pattern recognition quickly so that analysts are able to keep ahead of emerging threats. Tools that can help analysts do their jobs faster must be able to handle the four V's, provide useful features to automate routine tasks, and allow analysts to provide intelligence back into the system so it evolves with their needs. Automation is key to a system that can evolve with current and future types of analysis. An effective tool can replace a human when the overall speed and accuracy of analysis is improved. Automation doesn't remove the need for a human, rather it focuses those analysts on tasks where human intervention is necessary. Planning for the unexpected helps define how a system is designed so it doesn't become obsolete as soon as it's deployed. In order to stay relevant, an analysis system has to be flexible so that new data types and relationships can be integrated quickly without impacting analysts.

A system that addresses the four V's is positioned to automate all aspects of DNI analysis and in turn, free up manpower for analyzing the output of the system. Allowing analysts to work on new and unique methods of network analysis keeps them ahead of evolving threats. The ultimate goal of the system should be to improve the response to all types of threats, pinpoint troubleshooting areas and manage defensive asset placement. The only way to improve threat response with the huge amounts of data is to automate the ingesting, processing, merging, correlating and presentation of the data.

Unique Strengths in Design:

An effective analysis system handles the Volume of data by organizing, aggregating and handling duplicate data efficiently. Data correlation is important when dealing with large data sets with varying formats, such as those produced by netflow, BGP, and Intrusion Detection/Prevention Systems. When predefined events occur, the system will react faster and handle more data than a human. Presenting data relationships in a meaningful way improves analysis, by allowing analysts to see patterns in the data they might otherwise miss. Large data sets of duplicate data are helpful in determining the validity of a network event. Storing multiple copies of the same data is cumbersome. Aggregation of these data sets reduces storage requirements and speeds up queries. Responding quickly to emerging threats can be difficult with large data sets, simply because of how much data has to be processed. A good system design has to be balanced to allow fast loading while also providing fast access to data.

An effective analysis system handles the Velocity of data by consuming data rapidly, allowing for quick data queries and providing related change alerts in near real time. Networks connected to the Internet are producing and processing billions of packets of data daily. As the rate of traffic increases, systems doing analysis need to be able to process these large data sets quickly, while also delivering results rapidly to analysts. As more and more traffic migrates to the network, changes in topology becoming increasingly more important. These changes can be due to network attacks, service provider policies on traffic loads, network failures and physical infrastructure incidents. Providing alerts in real time is critical for a system that addresses threats as they happen

and allows analysts to stay on top of network problems and changes.

An effective analysis system handles the Variety of data by avoiding the trap of building a system that handles a limited range of data types. The ability to handle new data sets quickly and easily will set the system apart from previous tools that required time consuming code changes to incorporate new data formats. A flexible method for adding new data types will enable the system to handle the diverse data sets that are encompassed by DNI. A system that handles diverse data sets will help fill analysis gaps and produce a more comprehensive network picture by combining seemingly unrelated data. It is vital that the system have an API for both sending data to the system and retrieving data from the system. This allows the tool to become a gateway for data automation, which in turn will speed the analysis process. Combining data such as, BGP, netflow, packet capture, IDS/IPS data into one complete picture allows relationship connections to be made in seconds, as opposed to weeks. Designing the system for unknown data types frees analysts to concentrate on how to use the data to make new connections. A system that allows virtually any data type opens new avenues of analysis never before realized.

An effective analysis system handles the Veracity of data by automating as much analysis as possible, while allowing analysts to augment data sets with manual analysis results. The accuracy of the system is essential. Humans make mistakes and process automation reduces the possibility of human error. As the amount of data being processed, and the reliance on the system to produce reliable information increases, the system becomes the cornerstone of analysis. As more tasks are automated by the system, the need for analyst feedback also becomes more important. A method for analysts to input information back into the system makes the system a valuable tool as its use grows. Multiple source conflict resolution is important to maintaining system accuracy. When two or more data sources have different information about a network event, it's essential that the system apply rules to determine which source is correct or which is more reliable. With the amount of data that is being produced on networks connected to the Internet, analysis produced by such a tool must be accurate.

The Mission Piece:

An often overlooked part of a well built analysis system is a method of reporting the completeness of consumed data sets. A common phrase in the intelligence community is applicable here; *"Know what you don't know"*. The core idea is that knowing that there is missing information, is itself a helpful part of analysis. A system should have the ability to compare data sets and report their completeness and what data might be missing. For example, information about IP address allocation by region, unregistered autonomous system announcements, conflicts between regions where prefixes are registered and where they are being announced make up valuable information sets. Commonly an analysis tool is so focused on data loading and querying that it overlooks the importance of analyzing how much value the data sources add.

Built in Functionality:

A well-designed system will have many benefits:

In depth analytical capabilities with useful visual displays - Being able to quickly determine how vulnerable a network is to cyber attack and allow an analyst to take steps to mitigate or automatically respond.

- Ability to visualize network status
- Improved BGP route monitoring
- In depth peering information
- More detailed topology data
- More reliable geolocation data
- Increased accuracy of matching IP to router interface.
- Identification of critical network infrastructure
- Identification of connectivity richness among networks both public and private
- Improved router level maps
- Quick and efficient drill down to views of interest, such as:
 - Known bad actors
 - Inbound/outbound attack types (scanning, beaconing, worm sign)
 - DDoS
 - Routing anomalies
 - Identification of Botnets and their command and control
 - Virus and worm outbreaks
- Visualization of Botnets, attackers and other network threats
- Customizable reporting tools that can be tailored by the user

Summary:

A useful analysis system can evolve with the changing cyber landscape. It's essential that this system be designed with flexibility in mind and a modular architecture that allows it to scale quickly and easily. A system that frees up analysts and improves analysis techniques, while keeping up with the growth of the Internet is a valuable addition to the network analysis process. A system that is capable of doing all these things effectively is positioned to radically change analysis techniques. The ideal application enables analysts to discover new cyber security events and analysis methods and feed those techniques back into the system. Integrating global Internet data with data from an enterprise is essential to seeing how the enterprise network fits into the big picture. The merging of public and private data allows for a high level of network cognizance. Network cognizance leads to network situational awareness. Keeping systems up and running while servicing users and clients in a safe manor is the outcome of network situational awareness. Therefore network security analysis has a huge impact on intrusion prevention and system survivability and the insight described in this paper is absolutely necessary to protect enterprise networks and provide sustained services.