A WHITE PAPER

>>> Packet vs flow-based anomaly detection

Why packet-based anomaly detection is superior to flow-based approaches

ESPHION
NETWORK DISASTER PROTECTION

>>> Contents

>>> Introduction

Operators of mission critical networks employ a variety of strategies to ensure uptime and availability. For network security, firewalls and intrusion prevention systems (IPSs) may be utilized, along with performance measurement tools and network infrastructure health monitoring systems.

However, to protect networks against threats such as DDoS attacks and worm outbreaks, intelligent, real-time solutions are needed.

Anomalies generate vast amounts of bogus traffic, which can overwhelm the network and any attached hosts. In addition, the traffic that is generated by anomalies may not have a signature, which is required by a typical IPS. It may also arrive on otherwise completely legitimate ports, passing the security checks of a firewall.

In the past, the only way to detect those anomalies was to watch the volume of data or packets on the network using various network monitoring tools. However, legitimate network traffic may be bursty or highly variable, rendering such naive approaches ineffective. Specially crafted attacks may be missed (false negative), and normal network activity may be erroneously classified as an attack (false positive). Constantly changing network utilization and configuration makes this approach inflexible and maintenance intensive.

As a result, a new category of network security systems has appeared, specifically geared to solve this problem. These systems utilize what is commonly known as behavioral anomaly detection. Rather than just looking at volumes of packets, these systems intelligently take into account the behavior of the network and the hosts that are attached to that network. Changes in network behavior are used to detect DDoS attacks, worm outbreaks and otherwise misbehaving hosts or network elements with dramatically improved accuracy.

More and more operators of mission critical networks recognize that an additional layer of security is needed above the traditional signature based systems, such as IPSs and firewalls. As a result it has become best-practice to deploy an intelligent behavioral anomaly detection solution in such networks, alongside the already existing security infrastructure.

In the field of behavioral network anomaly detection, two approaches have emerged. One is flow-based, usually relying on existing network elements, such as routers, to make so-called flow information available for analysis. The second approach is packet-based and does not rely on the capability or availability of other network elements. It observes network traffic for the detection of anomalies. This paper describes the main differences between the two approaches and outlines why a packet-based approach is faster, more accurate and leads to more useful, actionable information, which is more relevant in stopping an attack, securing the network and ensuring business continuity.

>>> 1. Flow-based anomaly detection

Flow-based anomaly detection centers around the concept of a network flow. A flow record is a summarized indicator that a certain network flow took place and that two hosts have communicated with each other at some point in the past. A flow record typically contains the IP network addresses of the two hosts, the network ports, the protocol, the amount of data that was sent as part of this connection, the time when the flow occurred as well as a few miscellaneous flags.

There are network devices, for example many Cisco routers, which are capable of observing network traffic and generating such flow records, which Cisco calls Netflow. The flow records are written into newly created network packets and sent off to a recipient for analysis.

Some behavioral anomaly solutions act as flow receivers and base their entire anomaly detection and analysis on the data contained in those records. Of course, network devices need to be reconfigured to generate those flows and send them off to the anomaly detection solution. Often, the amount of traffic that is generated by all those exported flow-records can itself place a significant strain on the network infrastructure.

Flow records are well suited to represent the interactions between hosts in a network. By analysing exported flow records and looking for unusual amounts, directions, groupings and characteristics of flows, an anomaly detection solution can infer the presence of worms or DDoS attacks in a network.

It is in the nature of flow records, however, that only a summary information is presented for analysis. In effect flow records present meta information about network traffic. The actual network packets, which were used in any communication, are not accessible for further examination.

>>> 2. Packet-based anomaly detection

Packet-based anomaly detection, unlike flow-based solutions, does not rely on third-party components to generate meta or summary information of the network traffic. Instead, all analysis is entirely based on actually observed raw packets, as they traverse the network links.

Observation of network traffic can occur in several ways. One is to configure a spanning port. A router or switch then makes a copy of every packet that is sent/received on one or more of its interface ports, and sends this copy out on the span port. A packet-based anomaly detection solution can then monitor the output of those ports.

Another method is the use of network taps. Those are passive devices, which allow the fully transparent observation of packets on a network link. For fiber links, a so called fiber-splitter is used. The advantage of network taps is that they can be used even when no network device is available to provide traffic via spanning ports.

Once the packet-based anomaly detection solution is set up, statistics about the observed packets are accumulated and analyzed by a variety of methods. For example, Esphion's netDeFlect uses sophisticated neural networks to detect the presence of any anomalous traffic. In addition, the content within those packets can be kept and used for advanced anomaly detection.

>>> 3. Packet-based vs. flow-based

When comparing these two methods of anomaly detection, the architecture of the network plays an important role. Some networks lend themselves more to one approach than to the other. However, most enterprise networks, ISPs and even larger service providers are typically able to deploy a packet-based solution in their network. A packet-based approach to anomaly detection has a variety of significant advantages, which are discussed in greater detail below.

>>> 3.1 Access to the raw packet data

Probably one of the most important aspects of packet-based anomaly detection is the access to the raw packet data that it allows. In contrast, flow-based solutions only see summary records, produced by a third party (routers, for example), and therefore never have access to this information, which often is vital for analysis and mitigation of an anomaly.

For example, advanced packet-based anomaly detection solutions, such as Esphion's netDeFlect, are capable of extracting signatures even for zero-day anomalies. This is done through a sophisticated statistical analysis of the observed packet content. As a result, fine-grained filter instructions are automatically produced within seconds of the onset of an anomaly. These filtering rules can be used by routers, but also by firewalls or IPSs to surgically remove the anomalous traffic. Some of these devices are capable of filtering traffic based on specific values in the packet headers or the packet content. Only a packet-based anomaly detection solution can provide filters that fully take advantage of the investment made in those inline devices.

Take the example of a large-scale SYN flood denial of service attack. Typically a huge amount of TCP-SYN packets (connection establishment requests) are generated by a number of compromised zombie machines. The source addresses are randomly generated. A flow-based solution only sees that there is a large number of flows, which are established from many clients to the specific server and port that is under attack. But no useful information beyond that is forthcoming from a flow-based solution. Therefore, the network operator has the choice to either rate-shape or block all traffic to that server, with disastrous impact on even the legitimate traffic.

A packet-based anomaly solution, however, can extract the signature of the offending packets. Often, large-scale attack tools initialize packet headers or content with certain, non-random data. For example, the TCP window size or sequence number, which is advertised in a TCP-SYN packet, could be fixed. A packet-based solution, which has access to the raw packet data, can detect this and provide a signature that very specifically only blocks the offending traffic, and leaves legitimate traffic untouched. >>

As another advantage, the ability to access the raw packet data also permits forensics examination of the actual attack traffic. For example, throughout an anomaly, Esphion's netDeFlect periodically records samples of the anomalous traffic, which are then used to extract updated real-time signatures. At the same time, these samples are kept for further analysis at a later time. Manual or automated payload analysis may reveal the exact application protocol, or other aspects of the anomaly, which help to reconstruct events or further secure the applications and the network.

>>> 3.2 Low-latency anomaly detection

Routers and switches usually export a flow after there has been a certain time of inactivity, typically some 15 seconds. Thus, a flow-based solution can at the earliest only begin to detect an anomaly at least 15 seconds after its onset. After that, the detection algorithms can start to do their work, which usually adds some more time before actually coming to the conclusion that there is an anomaly.

In contrast, a packet-based solution works in real-time. There is no 15 second lag before statistical data about the network traffic is available. The detection algorithms continuously work on this real-time data. As a result, a packet-based solution can detect network anomalies faster than flow-based solutions.

It might appear as if 15 seconds would not make a lot of difference. However, that is not so. Consider the case of an enterprise network. Coming back from a business trip, one of the employees plugs the laptop back in. Unfortunately, while on travels, this laptop was infected by an aggressively scanning worm, which now starts to look for new victims in the company's network.

Depending on the exact scanning algorithm of the worm, an infection of another machine in the network may now happen within seconds. Therefore, every second counts for the successful containment of the outbreak, and only the fastest detection solution will do. A packet-based solution can detect the infected machine and apply the necessary blocking instructions to switches and firewalls in less than 15 seconds. By that time, a flow-based solution will not even have seen the first indication of an anomaly.

>>> 3.3 No reliance on third-party network elements

A flow-based anomaly detection solution relies on third-party network elements, such as routers and switches, to produce the flow-records that are its only insight into the current network traffic. However, many attacks and anomalies are either designed to specifically affect those network elements, or are taking them down almost as a side effect. >>

But if the router fails, then who provides the flow-records? As a result, a flow-based anomaly solution essentially goes blind at the most inopportune time: Right in the middle of an attack. All insight is lost, along with any capability to track the attack and adapt to a changing attack profile. The same holds true if the router becomes too overloaded to produce more flow-records. Some or most of the traffic may continue to flow, albeit heavily congested by the attack, but no more flow-records are produced, leaving a flow-based solution incapacitated.

In general, it is a bad idea to rely on the network infrastructure to detect and observe attacks that are designed to take out that very infrastructure.

A packet-based anomaly detection solution on the other hand does not suffer from this flaw. It remains operationally independent of the state of the routers or switches. As long as traffic is flowing on the network links, it is seen and analyzed. This is possible, since a packet-based solution can operate via simple network taps, which are entirely passive devices, and which provide a raw copy of all network traffic to the anomaly detection solution without any additional computational requirement.

>>> 3.4 No overhead on routers, switches and network

Generating flow-records places a significant burden on the network infrastructure. Many routers' CPUs are heavily loaded when flow-generation is enabled, which can interfere with more important tasks of the router. During a denial of service attack, or even a worm outbreak, every single packet potentially represents a new flow. Flow-generation by the routers and switches acts in effect as an amplifier of the attack. The more attack packets are generated, the more the router CPU is loaded and the more the network is taxed by the presence of an increasing number of flow-records.

A packet-based solution does not exhibit any of these problems. Since packets are directly observed 'off the network', a packet-based solution does not cause any additional overhead on the routers, switches or the network itself.

Even in distributed deployments, as they are supported by Esphion's netDeFlect for example, the amount of data that is communicated between sensors and centralized database always stays at the same, low level. The detection and all necessary communication is entirely independent of the attack volume.

As a result, a packet-based solution operates without impacting any aspect of normal network operations. Specifically, it does not generate a significant load on the infrastructure elements at the worst possible time, during an attack, which unfortunately is the case with flow-based solutions.

>>> 3.5 High accuracy, even under heavy load

One solution that is often suggested to lessen the severe impact on network resources that is caused by the generation of flow-records, is the so-called sampled Netflow, or flow-sampling. The idea is not to consider every packet for the generation of flows, but only every n^{th} packet. For example, every 100^{th} packet.

Obviously, the number of generated flows is dramatically reduced, along with CPU load and network utilization. However, this comes at the price of lost accuracy. Consider that 99 out of a 100 packets will be completely ignored. As a result, most smaller flows will be seen only as a single packet, no matter if the flow does indeed only contain one packet, or not. Any information about the average flow length, for example, becomes entirely unreliable.

Large flows will obviously be over represented, since they have a higher chance of having at least one of their packets sampled, which represents a distorted picture of the actual state of the network.

As an additional downside, the detection time for an anomaly, such as a worm outbreak, is further increased, since more packets are needed to pass through the router to detect the anomaly.

Contrast this with packet-based solutions, which do not need to resort to sampling techniques. Their per-packet operations are quick and efficient, and therefore can run at full speed, and full accuracy, at all times.

>>> Conclusions

In comparing flow-based anomaly detection solutions with packet-based solutions, packet-based solutions are superior for a range of reasons. They give access to the actual packet data, which is vital for the generation of fine-grained filtering instructions, which enable the surgical removal of only the offending traffic. Packet-based solutions do not suffer from inaccuracies and dependency on the network infrastructure to the same extend as flow-based solutions. In addition, packet-based solutions do not cause the overhead on the network infrastructure, which so often is an issue with any flow-based deployment.

Packet-based solutions are preferable in almost all deployment scenarios, and should be the first choice for reliability, accuracy, efficiency and minimum impact on the network infrastructure.

> About Esphion

Esphion protects enterprises, service providers and Governments from network disaster. Esphion's breakthrough technology uses neural, behavior-based, real-time analysis to detect known and unknown threats within seconds. It then generates fine-grained signatures, allowing you to stop even the most aggressive attacks and eliminate internal network threats and insider misuse. Launched in 2002 by industry veterans, Esphion is backed by leading venture capital firms. Its customers include Fortune 100 enterprises and major service providers and financial institutions. For more information visit **www.esphion.com**

www.esphion.com
sales@esphion.com

Esphion Corporate Headquarters
20 William Pickering Drive, Albany
Auckland
New Zealand
Ph: +64 9 414 2060
Fx: +64 9 415 0228

Esphion Australia
Level 6. 90 Mount Street
North Sydney
NSW 2165
Australia
Ph: +61 2 9955 3611
Fx: +61 2 9959 5760