White Paper

Intrusion Prevention System Performance Metrics

The Importance of Accurate Performance Metrics

Network or system design success hinges on multiple factors, including the expected performance of the elements involved. Without proper throughput, alignment chokepoints can arise, impacting network traffic availability. Cables, interface cards, and other simple elements have fairly predictable and accurate performance guidelines. More sophisticated components such as switches and routers can exhibit a greater range of variance. Security elements have an equally if not greater range of performance results.

As more organizations move to build security into their networks, the importance of predictable performance grows. Networks must be engineered to assure that network and application traffic will continue through traffic surges and variations - and levels typical just a few years ago are far below today's needs. The mix of traffic types has changed as well, with more connection-intensive compound applications than ever before. Given the variability of traffic mixes, the growth in performance, and a cumulative gain in diverse applications plying those elements, it is increasingly critical to provide an accurate picture of how security devices will behave in these dynamic environments.

Overview of Performance Metrics

Given the need for accurate performance, it's vital to determine which metrics are important for the network and application environments where network security elements will reside.

Types of Performance Metrics

Network and application success is dependent on several performance attributes. The type and location of the network deployment will influence performance. The types of applications present, which are sometimes aligned with the network location, have equally important influence on the performance metrics most likely to result in a successful deployment.

The most frequently mentioned - and most commonly abused - performance metric is throughput. Throughput is measured in terms of traffic volume passing through a point in the network on a per-second basis. This coarse-grained measure is independent of traffic content.

Latency is another performance metric describing the amount of time it takes for the traffic to pass through the device. In effect, it is a measure of the amount of time the security device must take to perform its tasks.

Connection metrics come in either a "maximum count", or in the form of velocity representing connections per second. Connection metrics have significant impact on the types of applications being supported, as well as the types of devices participating.

There are also measures that address high availability, reliability, and recovery; from a network uptime perspective, these can be critical. Packet drop periods during failover or key transitions can make a critical difference in some cases. Mean time between failure (MTBF) values can be viable measures to help predict uptime.

Throughput

As mentioned earlier, few metrics in the IT industry are as consistently abused as system throughput. In response to this problem, computer system users, database users, and network users have attempted to create standards to provide common points of reference. For computing platforms, MIPS/MOPS/GIPs were introduced. For databases, Transaction Processing Council (TPC) standards are prevalent. Network RFC 2544 has also been suggested. The challenge with these reasonable attempts for commonality is that they rarely represent the deployment environment in which the system will operate.

Despite the aforementioned attempts at commonality, security vendors remain highly diverse in their definition of performance. The following sections will discuss what approaches are used in the industry.

Pure Network Throughput

Some vendors report performance numbers without any inspection activity. This measurement has potential value for network planning purposes, should the device experience a significant failure. Vendors generally avoid any specificity in describing this value, with the general assumption that it is essentially a wire equivalent through the device.

Even vendors that report pure network values can use multiple forms to describe their throughput. For example, one intrusion prevention system (IPS) security vendor uses a single User Datagram Protocol (UDP) and a single packet size of 1512 bytes. Another pure-play IPS vendor simply runs traffic in “wire mode” until packets are lost.

Single Traffic Standard

IPS vendors can be descriptive or opaque in describing their performance values. However, most will not describe the basis for their performance claims. Descriptive performance standards are available, but few IPS vendors are willing to expose their methods.

Of the major IPS vendors, few describe their performance values. The majority simply report a value without any explanation of their methodology. Further, these vendors typically make no attempt to qualify the likely deployment scenario for their performance benchmarks.

Third-party testing houses may or may not conduct multiple performance tests, but their results point to a single throughput value. For the most part, those traffic mixes, traffic change velocity, packet sizes, and protocol mixes are not described; thus, potential users are left wondering.

Not all vendors limit themselves to hidden performance benchmarks. For example, Cisco has evolved its performance metrics over the past year. Until recently, the performance measure was an Internet-edge scenario based on HTTP with varied return packet sizes - referred to as Transactional and Media-Rich. Since that time, Cisco’s IPSs now publish either an average or a range of numbers representing multiple performance tests based on third-party testing tools.

Deployment-Centric

Almost all IPS vendors keep their performance metrics vague and undefined. Only a handful will attempt to describe their methodology. And only one will incorporate multiple deployment scenarios.

Cisco recognizes the challenge that hidden performance metrics have imposed on customer deployments. This is why we have turned to publicly available third-party testing tools and deployment-specific tests.

Cisco IPS Sensors have been tested using five tests representative of these common deployment scenarios:

- Educational institution Internet edge
- Enterprise applications
- Enterprise data center
- Internet service provider (ISP) feeds
- Small and medium-sized businesses (SMBs) or remote office flows

Each test is published by BreakingPoint, a leading testing tool vendor. Specific tests were run through the BreakingPoint tools themselves.

Results of the five tests were used to generate either a range or an average, which serves as the basis for the performance ratings presented on Cisco IPS data sheets and other documentation. The intent is to give customers a better understanding of how their unique deployment is likely to be addressed by the IPS in question.

While some vendors treat customer deployments as though they are liabilities that are “outside their control”, Cisco sees this as a part of our partnership with our customers. Each individual result is available to your Cisco technical representative for proper sizing needs.

Latency

Delays in traffic can directly cause critical applications to fail as timeout conditions can be triggered. In some cases, time-to-live values may trigger traffic to be resent, potentially exacerbating challenging traffic problems. For all of these reasons, latency is an important consideration for an inline network security device.

Latency Measurement Conditions

Like throughput, most vendors hide latency measures. Some will make no reference to them at all or will simply deliver a value without any reference as to how it was determined. One vendor simply references 1518 byte size packets. Others say nothing.

As a gross statement, unless the device is deployed at a slow Internet edge, anything approaching a full millisecond is undesirable. However, even those values may be meaningless if the conditions in which they are determined are not similar to the environment in which the device will be deployed.

Cisco uses the same deployment-focused public tests and then generates an average. Again, these values are known to Cisco’s deployment specialists for reference when designing a network security solution.

Connections

The number of connections between hosts the system can track, and the rate at which those connections are established, can be critical for certain application types - and even for the type of devices involved. While sessions are generally the same as TCP connections, they have different meanings for different application types and uses. This is why most network equipment focuses on connections instead.

IPS vendors that understand network traffic and its relationship with connections will publish and explain these values. Unfortunately, many pure-play security vendors rarely do this.

Maximum Connections

When an IPS is placed in the path of a large number of clients reaching out to an application-heavy destination, connection counts - particularly, maximum connection counts - become highly important. While some vendors will purposely limit the overall connection count to improve their top-line throughput, this is a significant limiting factor for many deployments.

Connections per Second

A single-value maximum number holds some value, but is incomplete when building a network. Throughput and connections rarely occur in a slow, steady climb. For these reasons, the number of connections per second is important. If “bursty” traffic connections occur naturally in the deployment environment, the maximum value may not really matter, as the velocity of the growing connection rates will cause problems far before the theoretical maximum is ever reached.

Conclusion

Network security must simultaneously ensure that the traffic traversing the network is secured and performance and availability are assured. Any network or computing system design requires a clear understanding of how the components perform relative to one another.

Cisco has introduced a real-world, deployment-focused performance standard. This unique approach gives network and security architects an opportunity to preserve both their security and availability standards.

Customers should ensure the metrics they use reflect the environments where these systems are to be deployed. It is critical to understand how each security element describes its performance and if those descriptions are representative of the success criteria for the organization’s deployment.

References

iMix Example

iMix is an example of traffic associated with an Internet connection.

Packet Size	# Packets	Bytes	Distribution
64	67	4288	57%
570	1	570	7%
594	2	1188	16%
1518	1	1518	20%

RFC 2544

RFC 2544 is frequently referenced in networking equipment and can be found here: http://www.ietf.org/rfc/rfc2544.

The methodology specifically calls for a variety of packet sizes and protocols, but can be manipulated to a combination optimal for the device under test.

Both steady-state and bursty traffic pace are recommended, but burst measurements are not required.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-704003-00 05/12