WHITEPAPER

Realistic Application Traffic Reproduction

A New Gigamon + Spirent Solution

Executive Summary

In today's networks, the need to reconstruct live traffic for assessment is critical for successful deployment of network configurations, topologies and new network services. Modern networks now contain multiple state-aware devices that require realistic upper layer application traffic to fully exercise the network elements.

Together, Gigamon and Spirent have created the industry's first solution for recreating actual application traffic in pre- production environments. The joint effort will characterize the live application flows in a production network using the Application Intelligence module within the Gigamon Visibility Fabric. This data is exported to Spirent's market-leading test tools, such as Spirent's CyberFlood and Spirent TestCenter, where it is used to statefully recreate the client and server behavior application flows. This recreation includes use of full stack TCP, coordinated sub-flows, messaging, application data payloads, and stateful emulation of real clients and servers.

By combining their core competencies, Gigamon and Spirent are allowing customers to characterize complex, interleaved stateful traffic and reproduce application-aware flows in a repeatable fashion. The result is that network operators—including service providers, enterprise IT and governments—can adopt, deploy, and operate new networking technologies faster, more efficiently, and with less risk.

Challenges and Limitations of Traditional Approaches

Every organization has its own set of applications that make up the majority of an organization's application traffic mix. It is a classic case of the 80/20 rule, where 20% of the applications create and drive 80% of the network traffic. This mix is not just a simple list of applications, but also includes the number of packets, flows and bandwidth generate by each individual application.

Networking appliances are generally assessed using generic application mixes provided by test generation tools. The results are often used for marketing purposes and show up in datasheet specifications. Most vendors generally place a disclaimer regarding these specs stating they were achieved in a controlled lab environment under ideal conditions.

These datasheet numbers provide a useful starting point for understanding performance, but they are typically not an accurate reflection of how the vendor's products will perform in a network operator's unique environment. Simply put, claimed performance is often not achieved in real-world conditions, which makes it hard for operators to evaluate and adopt new networking products and technologies. This situation is not ideal for the vendors or their customers, since the overall effect is increased cost and delayed time to revenue.

When these networking products are placed inside a production network with actual applications and their realistic usage levels, their performance may degrade from what is achieved in a lab setting. For example, in a production network where encrypted traffic is around 40 percent of the total traffic, a firewall may perform differently than the lab environment where it was just tested with 20 percent encrypted traffic and without any attack/bad traffic.

If a company knows its application traffic mix, then the networking appliances can be tested with their specific traffic mix so that there is no performance differential when put in the production environment. However, it is quite challenging to know the actual application traffic mix on your network. Most of the time a rough estimate based on the business-critical applications involved is all that is available. And beyond such an estimate, network operators have not traditionally had a way of knowing all of the applications that are actually present on the network, not to mention the exact flows, packets or bandwidth for each application. The usage of these applications also changes over time; for example, during peak times, depending on the nature of business, some applications will consume dramatically more bandwidth than during off-peak times.

Gigamon's Application Intelligence overcomes this challenge; it identifies application traffic in a company's network and provides the bandwidth percentages for each application along with granular details on flows and packets. And it is able to provide this detailed information with potentially thousands of applications contributing to the overall traffic mix. This eliminates guess work and provides a realistic view of the network performance posture at different times for various parts of an organization's network.

Gigamon customers who are also working with Spirent are now able to use their unique traffic mix information from Gigamon's Application Intelligence to create more powerful pre-production test scenarios. Specifically, test scenarios which use the company's own actual traffic mix. Devices under test can now be evaluated prior to deployment in a production like environment. This method removes the uncertainties and saves the company on capital and operating expenses.

In addition, precisely reproducing what is seen on the live wire has always been difficult. Live traffic is complex, interleaved and temporal in nature. In addition, modern networks are extending state awareness from a centralized "appliance" such as a firewall, to services widely dispersed, even down to access ethernet ports. The challenge using traditional traffic generators is that they can typically only successfully reproduce a very small portion of the input stimulus across the device under test (DUT). For example, fixed frame size and rate traffic generator test cases will only produce a very specific input to the DUT but will not reproduce the interconnected real-world nature of production traffic where frame size, instantaneous rate and address are dynamic. At best, this traffic pattern will lead the user to a false sense of "testing completeness". Even if this were possible, it would take a long time to characterize and reproduce in the testing, making the usefulness of the exercise nominal.

Another challenge using traditional modeling with an L3 traffic generator is that the network is starting to make routing and switching decisions at Layers 2-7, not just Layers 2-3. There might be a firewall where the network port will only allow approved stateful TCP traffic, or even approved applications over TCP. In addition, the network may expect coordinated TCP with application flows, bundled as an application with context and order between the connections. Because an application is a bundle of TCP and application flows, using a classic L3 traffic generator to generate unconnected, stateless flow may result in traffic being simply dropped on the first hop into the network.

Next, the user perceives quality at the "application layer". Although L3 flows do measure path QoS, it is more accurate to say they will measure "conditions of failure" as opposed to "conditions of success." For example, if an L3 QoS test measures 10% packet drop, it is reasonable to assume that application flows over this path will fail quality of experience (QoE). Conversely, low packet drops such as 0.1% do not predict the success of the pathway. Specifically, an L3 tester will generally neither tell the user the distribution of loss, nor will it replay the effects of loss such as TCP recovery. L3 traffic generators are a useful part of testing as they can measure failing pathways and allow for rapid fix and retest in a controlled state, but they never measure success.

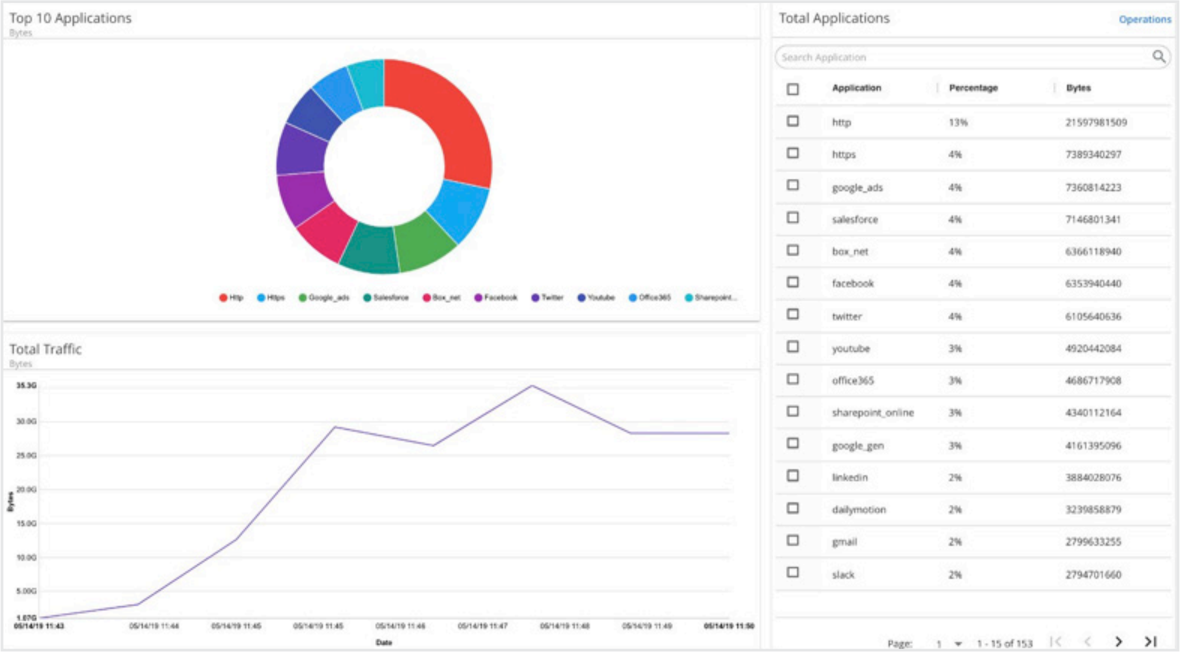
Yet another challenge is to correctly load the elements of the Device Under Test (DUT). The DUT is like a SQL database in that it has tables and relations, consumes compute and RAM, and performs self-tuning such as table pruning. A simple L3 stream will only consume a few rows in a few tables (ARP, FIB Table). In comparison, one Application workflow can contain many TCP connections containing application client requests and server responses. This will consume many more kinds of tables but fill them with more entries and stress more cross table relationships. This will consume more RAM and compute power. Further, the application workflow is a unit. Any sub-flow within the application workflow that is degraded will impact the overall application QoE. Contrast this with isolated L3 stateless flows, and the loading difference is clear.

The last challenge is how to accurately measure traffic scale. If we use metrics from the middle of the OSI stack such as bandwidth (L1/L2/L3) or open connections (TCP) we will certainly measure a narrow KPI. Upon inspection of results we will find that the measurement is necessary but not enough to properly measure traffic scale. For example, if one measured 40 Gbps of total bandwidth at L3, it is true that 40 Gbps of ethernet are forwarding across my DUT. But, it is likely that the QoE is so degraded (for example, taking 20 seconds for an application page to load) that in the production network one would have to keep rate-limiting traffic until the application became useful for the users. This kind of measurement tends to be overly optimistic of performance, and in many cases, real-world bandwidth with an acceptable user experience may be 70% of what was measured in the lab. We need a realistic way to accurately and predictably measure real-world conditions.

Other techniques such as PCAP playback fall short. As a simple dump of packets, there is no state. This means that whenever a network element induces a fault, the consequences (e.g. TCP retransmission) will not occur. This induces several errors. First, consumption of future performance in exchange for recovery is not processed. Second, recovery and fault signaling does not occur, keeping the TCP state table inaccurate. Lastly, any faults knock the PCAP "off alignment" and is no longer valid after a fault.

Gigamon - Spirent Solution Description

The combined solution of pairing the Gigamon Visibility Fabric with the Application Intelligence module and Spirent CyberFlood empowers the user for the first time to successfully and swiftly characterize stateful, interleaved application level traffic in the production network and then successfully reconstruct this traffic (i.e. in the right ratios) as a predictive, reproducible test pattern with Quality of Experience (QoE) measured results.



Total Applications

□	Application	Percentage	Bytes
□	http	13%	21597981509
□	https	4%	7389340297
□	google_ads	4%	7360814223
□	salesforce	4%	7146801341
□	box_net	4%	6366118940
□	facebook	4%	6353940440
□	twitter	4%	6105640636
□	youtube	3%	4920442084
□	office365	3%	4686717908
□	sharepoint_online	3%	4340112164
□	google_gen	3%	4161395096
□	linkedin	2%	3884028076
□	dailymotion	2%	3239858879
□	gmail	2%	2799633255
□	slack	2%	2794701660

Page: 1 ▾ 1 - 15 of 153

Figure 1: Gigamon's Application Intelligence identifies and classifies unique application traffic

The Gigamon Application Intelligence taps and aggregates live traffic in the production network and through deep packet inspection (among other techniques) identifies the application workflow based on content. This is true application visibility as opposed to other schemes that simply rely on port mapping as a crude method of identifying the applications present. By automatically identifying the application workflows and their relative percentages of all traffic, Gigamon classifies and reports the specific application distribution over the selected sample period with a much greater level of precision (see figure 1). In addition, Gigamon Application Intelligence can sample different points in the network at different times.

Spirent CyberFlood can recreate traffic at the application workflow level, allowing the user to mix traffic according to the measured ratios between apps in the live network. In addition, Spirent CyberFlood has an extensive TestCloud library of applications, allowing the user to closely model apps. Once a traffic mix is created, it can be played back exactly or scaled up according to bandwidth or client and server IP addresses while measuring crucial KPI metrics. Other Spirent tools including Spirent Test Center are also planned to take advantage of the rich Gigamon traffic mix information to enable even more customer testing scenarios. Figure 2 shows how Gigamon and Spirent test tools can be used in concert to allow network operators to design, deploy, and operate their networks faster, more efficiently, and with less risk.

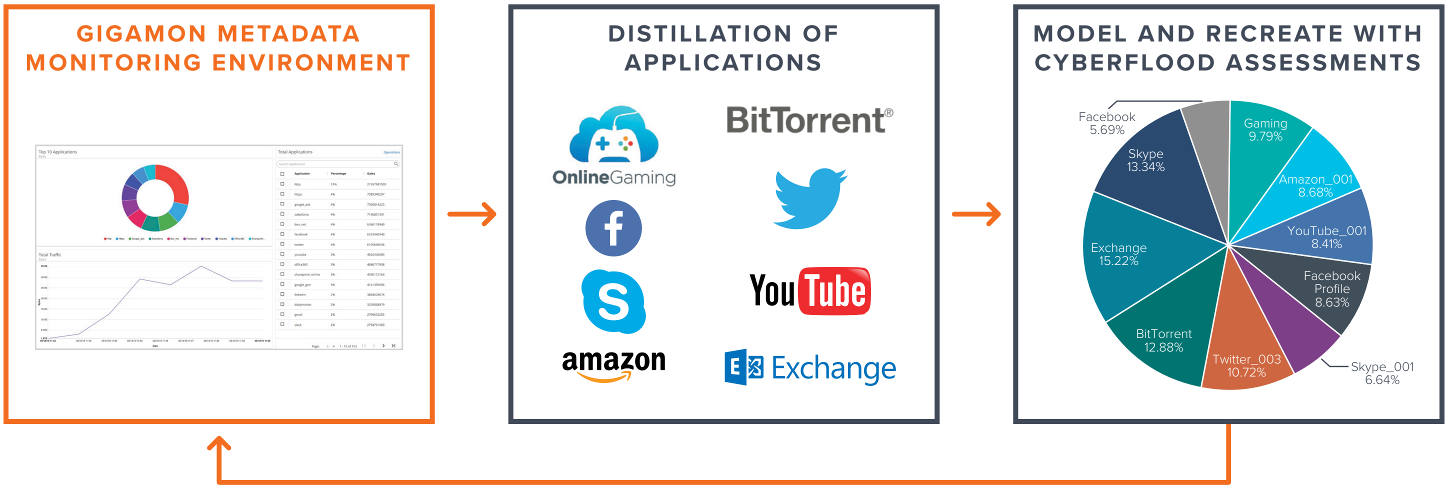


Figure 2: Gigamon Application Intelligence used in concert with Spirent's CyberFlood

Example Use Case

The use cases for playing back stateful application workflows are described by, but not limited to, the following example. They illustrate the benefits of precise stateful reconstruction of the live network.

In a support scenario, the support department can remotely schedule classification using the Gigamon Application Intelligence for specific network segments, subnets and time periods causing issues across the network. Once application identification is complete, the user can construct the traffic mix in Spirent CyberFlood and play back the remote network's traffic in the lab. The benefit of the solution is that the customer's traffic mix will be reproduced and replayed under controlled conditions, creating a higher quality fix in a shorter period.

In the case of RFP (Request for Proposal) testing, the IT department can classify their traffic specific to the target subnet, and even choose multiple scenarios representing nominal and high utilization time periods using the Gigamon Application Intelligence. Now the IT department will have truly comparable test traffic with meaningful QoE-oriented metrics for measuring vendor-to-vendor scale.

In live network debugging, the user can classify and reconstruct traffic allow for rapid debugging and prototyping of impact of device settings on application traffic. The benefit is that the user can rapidly reduce debugging time and assure that settings will be accurate, and impact will be measurable. If the user is considering migrating a physical DUT to an NFV/VNF virtual appliance, the user can harness the Gigamon Application Intelligence to classify peak traffic across the existing device and use the reconstructed traffic across the virtual equivalent to measure impact of underlay, DPDK/SR-IOV, CPU/RAM pinning, and the overall efficiency of the virtual machine or container.

In the case of a security device, the user can classify real traffic at high stress points during the day. Then, they can validate security appliance for new versions of code or new settings by using the reconstructed application mix and adding any combination of malware, DDoS, and attacks in line with application traffic to see how well the security appliance is mitigating attacks. The user can also see the impact of QoE on their specific applications. Lastly, this system can be used to assess WAN and SD-WAN circuits by capturing typical traffic and playing a real application mix across the test circuit measuring scale and SLA compliance. This is especially useful if the SD-WAN circuit is shaping application traffic or if the SD-WAN provider is offloading a service such as firewalling.

For compliance purposes, its always best to use lab traffic in test environments since the live corporate feeds contain sensitive consumer data such as financial transactions, medical records, and personal information. Once the live feed is characterized by Gigamon's Application Intelligence, and emulated using Spirent's CyberFlood, the organization can be rest assured that no sensitive data is being used for testing purposes.

Summary

The ability to identify live application traffic and statefully recreate and play back is substantially more realistic and accurate than using basic simple PCAP replay. By capturing the application state within real TCP, the consequences of faults can be reproduced and assessed, ensuring future performance of a production network. By correctly aligning accurate stateful recreation and a total QoE measurement, the user can uniquely measure true scale of a network element. This is accomplished by Gigamon Application Intelligence providing deep traffic identification and Spirent CyberFlood statefully reconstructing the scenario for reproducible testing.

About Spirent Communications

Spirent Communications (LSE: SPT) is a global leader with deep expertise and decades of experience in testing, assurance, analytics and security, serving developers, service providers, and enterprise networks.

We help bring clarity to increasingly complex technological and business challenges.

Spirent's customers have made a promise to their customers to deliver superior performance. Spirent assures that those promises are fulfilled.

For more information, visit: www.spirent.com

About Gigamon

Gigamon is the recognized leader in network visibility and control solutions, providing the application intelligence required to optimize the security and performance of your Digital Enterprise. With Gigamon solutions that deliver rich network data while ensuring complete visibility across physical, virtual and cloud networks, our customers are empowered to solve the complex business challenges of a digital transformation. Since 2004, our 800+ employees have earned 66 technology patents and cultivated a global customer base which now includes more than 80 percent of the Fortune 100 and 10 out of the top 10 government agencies. For the full story on how our network visibility and security solutions can help evolve your Digital Enterprise, visit our website, follow our blog, and connect with us on your favorite social channels Twitter, LinkedIn and Facebook.

Gigamon®
Worldwide Headquarters
3300 Olcott Street, Santa Clara, CA 95054 USA
+1 (408) 831-4000 | www.gigamon.com