Pipeline
Technology for Service Providers.
www.pipelinepub.com Volume 14, Issue 2

Turning Siloed Data Lakes into Actionable, Real-Time Network Analytics

By: Naim Falandino

Cloud-based OTT applications and the Internet of Things (IoT) are putting demands on the network that are fundamentally changing not only the way it is architected, but also the way it is analyzed. Just as network silos and overlays are converging into a seamless dynamic mesh, operators need to join the isolated data lakes that currently exist — using big data to create a holistic view of the network.

However, operators need to embrace that network analytics can no longer be reactive; they must be proactive and generate actionable insights that can accommodate and drive the dynamic nature of the new network.

Simpler Times

Not too long ago, Simple Network Management Protocol (SNMP) and telnet were the principal tools used by network operators to analyze the performance and health of their links. The main concern was being able to identify problems. SNMP was designed to work on sick links. It was lightweight and had a minimum of features so as not to overload what might already be a failing connection. If a link was fully functional, there was no need to analyze it in real-time. As a result, point monitoring — or polling every five minutes — was sufficient.

As network and data center architecture has evolved over the past 10 years, running a cluster with headroom has been replaced by running at maximum utilization. With the ability to scale up and orchestrate traffic at a moment's notice, hyper-scale cloud companies — such as Google, Amazon and Facebook — have passed the limits of SNMP and need to know instantly what is happening and when changes need to be made to react to load.

Instead of polling, they need asynchronous, event-driven updates that inform them of failures and their locations as they occur. And because their subscribers are intolerant of performance issues, they need to have a real-time view of the operational state of their entire network as well as the ability to precisely troubleshoot and remedy those issues.

Streaming Telemetry

Thankfully, the industry has recognized the limits of SNMP and is adopting streaming telemetry, with all router and switch vendors scheduled to support it by 2018. This will improve the performance over SNMP by up to two orders of magnitude and provide more granular and flexible data — a welcome change over the limited SNMP dataset, as streaming telemetry will provide cloud operators with more timely data.

State-of-the-art streaming telemetry will support thousands of network elements and services and collect data from hundreds of different data sources, as opposed to just existing SNMP Management Information Bases (MIBs). It will not only monitor points, but literally everything flowing to and through the network from end to end. This data will be streamed in real-time, rather than polled every five minutes. The consequent richness of this data will allow it to not just be reported, but also analyzed in detail to provide actionable data that allows operators to reach the ultimate goal: integration with software-defined networks alongside real-time analytics that provide the intelligence they need to make them truly dynamic.

All of these possibilities sound fine in theory, but what is an operator really supposed to accomplish with all of this data and why do they need to seriously consider that? In addition to the needs typical of hyper-scale cloud application providers, there are other needs to consider, including solving network outages, instilling proactive customer satisfaction, analyzing encrypted traffic and improving security from DDoS attacks.

The Show Must Go On

Of the 60 percent of internet traffic currently represented by streaming cloud apps and services, binge-watching video content accounts for the majority, and Nokia Bell Labs predicts that it will rise to 80 percent by 2020. Meanwhile, Internet Service Providers (ISPs) have very little visibility into these flows, which is not only a problem for companies such as Netflix and Amazon, whose customers are very intolerant of interruptions to their favorite shows and movies, but also for network operators, whose customers will churn if dissatisfied.

A common problem for ISPs deluged with customer complaints about poor streaming quality is the task of figuring out the cause of the diminishing quality. Sometimes throwing more bandwidth at the problem works, but this approach can be hit or miss, not to mention expensive. A better approach is proper analysis, which can be provided by holistic network analytics. Then, it becomes possible to isolate, for instance, a single router with a poorly configured cache, or a fail-over router that is too distant from the problem to be effective. In such a case, a simple cache reconfiguration and a bit of bandwidth in the right place can translate into happier customers and less churn, and all at a reduced cost to the ISP.

DPI and Encryption

The aforementioned 60 percent of internet traffic represented by streaming cloud apps and services is also increasingly encrypted — and there will come a point in the not-too-distant future when most internet traffic will be encrypted. Although that is good for security, it is not so good for DPI-based analysis.

This is a critical point because many operators have relied on DPI to overcome the shortages of SNMP. One of the key problems with DPI is that it is expensive to implement and most operators can only use it to spot check their networks. However, the root problem with using DPI for network analytics is encryption. Signature-based classification on a packet's payload can only work if it is in plain text. Otherwise, DPI is simply blind.

For example, one operator had 100 percent more Facebook traffic than it was catching with its DPI-based analytics because over half of it was encrypted. Only a holistic view of its network correlated with data and context gathered from the entire Internet was able to accurately track the encrypted Facebook traffic to and through their network, providing more accurate data for network planning. Across the entire network, its DPI-based analytics solution was leaving 70 percent of top-application traffic flows unclassified.

Using a more comprehensive data set makes it possible to map flows from source to destination. By correlating telemetry from internet endpoints, DNS requests and a host of other information sources, it becomes possible to create a historically rich analysis that identifies up to 90 percent of the traffic — whether it is encrypted or not. This not only avoids all of the privacy issues raised by DPI technology, but also represents the only effective way to understand what is actually flowing to and through the network. That gives operators almost surgical precision in their troubleshooting and results in more accurate planning and optimized delivery, even of encrypted cloud applications and services.

Security and IoT

The Mirai malware DDoS attack on the managed DNS provider, DYN, in October 2016, was not the first DDoS attack carried out using a combination of IoT devices and cloud servers, but it received widespread media attention due to its scale and impact. Internet services were disrupted for millions of users in many areas of North America for several hours.

If there were a silver lining to the Mirai attack, it would be the attention brought to the lack of security of many IoT devices and the problems they pose for network security overall. For example, IoT devices do not effectively provide reporting statistics like network equipment does. They are clients of the network, and the volume of these devices — and their unfortunate lack of security — make them a liability.

Big data is crucial in understanding how myriad devices utilize the network, providing a historical record of device interactions and requests so as to answer questions about whether particular devices are operating normally or not. For instance, in the Mirai case, even the manufacturer of the hijacked devices could not understand what was happening because it could only see that they were sending traffic, not what kind of traffic or where.

As it turned out, because of a back door left open by the manufacturer in the firmware, Mirai was able to bypass the users' password controls and gained access to the root level of the devices, which was subsequently used to carry out the DDoS attack on DYN. The nature of this event in particular is additionally complicated since the attackers targeted DYN with a flood of DNS requests, making it incredibly difficult to decipher between the good traffic and the bad.

Network analytics based on a big data view of the network would have understood immediately that the half million infected devices were acting out of character, sending repeated DNS requests to servers that they normally would have no business contacting.

Protecting against next-generation DDoS attacks requires this type of contextual awareness. There needs to be visibility into services, CDNs and sites — not just IP addresses. The increasing complexity and frequency of attacks requires an ability to reduce false positives and negatives with the highest accuracy and enable real-time, surgical mitigation, otherwise, the cost to operators will balloon and response times will be inadequate to stop the damage.

Big Data, Better Analytics

The mesh of technologies, players, applications, and protocols is increasing the complexity of network operations — and it is only getting more complex with time. However, the tools available to manage this complexity have been evolving as well. For example, advances in database technologies, such as streaming vector, column-store databases, make it possible to map out the entire internet, using only publicly available data of the same type that Google collects. Through this, the database can create a catalog of the structure of the internet and how services are delivered, leveraging it in ways that are tremendously useful and valuable to network operators.

As a result, this data — combined and correlated with network telemetry and enterprise data — becomes a crucial enabler in solving the biggest challenges facing networks today.

The next generation of network analytics, based on standardized developments such as streaming

telemetry and big data, are already proving critical for network operators. Looking to the future, they will be foundational for realizing the promise of software-defined networks, providing the real-time analytics for correlating quality of service, routing costs and traffic demands to automatically optimize paths through the network dynamically.