Insight-Driven Network Performance Management and Protection in the Cloud/IoT Era

An Operational Practice prepared for SCTE/ISBE by

Tony Kourlas
Director Product Marketing, ION
Nokia
600 March Road,
Kanata, Ontario
Canada K2K 2E6

Title **Page Number**

List of Figures

Title **Page Number**

Introduction

As streaming video, intelligent cloud and the Internet of Things (IoT) applications begin to dominate today's networks, they bring with them new challenges for service providers (SPs) to address. Subscriber demands for a perfect streaming and cloud experience is creating explosive growth in network bandwidth and complexity. A new generation of distributed denial of service (DDoS) attacks originating from cloud and IoT sources is bringing down critical parts of public network infrastructure. Any upsets in service increase customer dissatisfaction and churn.

At the root of the problem is lack of the visibility and control necessary to identify and resolve cloud/IoT network issues quickly and cost-effectively. Operators have petabytes of data at their fingertips. However, this data is collected in silos across multiple systems, requiring them to manually combine and correlate billions of data points amassed, and then organize them into a coherent report so they can try and sniff out issues. This process has been wrought with human error, and has not provided enough data to see exactly where problems lie. The primary tools of this process, deep packet inspection (DPI)-based appliances, were simply not designed to deal with cloud/IoT scale and complexity. In the end, problem resolution has become a costly guessing game that rarely resolves problems quickly enough to keep customers happy. That has led to ballooning capital and operational costs.

The industry has responded by evolving IP network analytics. A new generation of software-only solutions can ingest, combine, and correlate petabytes of siloed data from network, enterprise and cloud sources to provide a holistic view of the entire network and the applications that flow through it – in real time. SPs can, without hardware probes, track applications and services – not just at certain points in the network, but end-to-end, across 100 dimensions at the same time.

Network intelligence has also become actionable in real time. These solutions also provide SPs with the tools they need to quickly act on this data by creating baselines and triggers that alert on anomalies. Data from the network, data center and wide area network (WAN) can be used to trigger processes or real-time policies that increase customer quality of experience (QoE) and network security, while decreasing overhead and customer churn.

1. SP challenges and Needs in the Cloud/IoT Era

1.1. Ensuring Customer QoE

Cloud applications and services – including Netflix, Hulu, Twitch, YouTube and Facebook – make up more than 60 percent of network traffic today, and are expected to rise will rise to 80 percent by 2020, according to Nokia Bell Labs.[1] Yet in this environment, providers have very limited insight into which applications are running on their networks, and what impact this application traffic is having on performance and subscriber satisfaction. If their subscribers do not receive high-quality streaming, they will complain and eventually switch ISPs in search of a better experience. For them, "slow" is the new "down" when it comes to streaming speed and quality of over-the-top (OTT) content traffic. The *Nokia*

[1] *Bell Labs Consulting Inaugural Mobility Report* https://pages.nokia.com/1503.bell-labs-mobility-report.html

Acquisition and Retention 2016 Study found that internet quality is the most important driver for retention, and that those consumers most unsatisfied with that quality are more likely to churn.[2]

At the root of the quality problems is network congestion. To fix this, internet service providers (ISPs) have typically thrown more bandwidth and caches into the network, but they've done so blindly, making this an expensive and ineffective proposition.

Conventional methods for assessing network performance, including DPI, can be extremely expensive to deploy and scale, with limited visibility. DPI hardware cannot keep up with accelerating speeds and feeds, and is too expensive to deploy network-wide, which means it does not see much of the traffic flowing to and through a network. Because the classic DPI approach dissects every single packet in its path to see what's inside, it is completely blind to over 50 percent of network traffic. The end result is that SPs lack the data they need to pinpoint underperforming areas of the network. They know there is a problem because subscribers have complained, but they have no idea how much capacity is needed, or where to place it. Time and money is typically wasted on a trial and error process—deploying bandwidth and caches effectively in the blind—with the hope of solving the problem. The end result is that providers typically fail to resolve service issues in a timely manner, and they fight a losing battle to keep customers happy.

Getting the visibility necessary to ensure optimal QoE lies not in tens of thousands of expensive hardware probes, but in software-based solutions that can scale to the largest networks and provide real-time, multi-dimensional (cloud and network) data. Such a solution must monitor tens of thousands of popular cloud applications and services in real time, and run analytics that track how this traffic flows to and through networks to reach subscribers – without the need for expensive probes, taps and monitors. It must then combine this multi-dimensional visibility and analytics with the ability to create alarms and trigger policies that ensure superior performance management and customer QoE.

1.2. Securing Networks from DDoS Attacks

Driven by IoT security holes and 10G cloud server uplinks, DDoS attacks are growing in frequency and intensity. In these attacks, hackers set their sights on the source of the connectivity to disable as many end users as possible, as quickly as possible. They hijack thousands of unprotected IoT devices and cloud servers, and leverage the combined flows to spin up terabyte-level attacks that can disable entire data centers in moments

Verisign's *Q1 2017 DDoS Trends Report* identified a 23 percent decrease in the number of attacks in Q1 2017,[3] with average peak attack size increasing 26 percent compared to the previous quarter. Peak sizes were over 10 Gbp/s, while multi-vector attacks peaked at over 120 Gbp/s and around 90 million packets per second. DDoS attacks overall are expected to reach 100 million by 2019, up from 50 million DDoS attacks in 2016, according to Cybersecurity Ventures. [4] These attacks create a firestorm of outages, a deluge of angry customers flooding call centers and a serious impact on profitability.

---

[2] *Nokia Acquisition and Retention 2016 Study*
http://www.mediatelecom.com.mx/~mediacom/media/pdf/adquisition-retention-nokia-2016.pdf
[3] *Verisign Q1 2017 DDoS Trends Report*
http://forms.verisign.com/Q12017DDoSTrendsReport?utm_medium=Blog&utm_term=internal
[4] *DDoS Attack Report*, Cybersecurity Ventures http://bit.ly/2uA0b8A

As networks are threatened with the skyrocketing strength and intensity of such DDoS attacks, the immediate reaction for many operators is to disregard their previous investments and accumulate large collections of costly mitigation-specific hardware. The problem with that approach is that these appliances operate by sending all traffic through scrubbing centers, yet some vendors charge based on tonnage inspected—not traffic mitigated. Without proper detection, scrubbing will grow with the network while the expense of traffic monitoring skyrockets.

The same big-data visibility and context that now allows SPs to proactively ensure reliable network performance can also be utilized to surgically mitigate networks from DDoS attacks. A software-based multidimensional analytics approach to network security is the best defense, as it automatically combines many sources of real-time streaming datasets from the network to detect and mitigate DDoS attacks in seconds. This broad visibility into network and cloud behavior also serves to minimize false positives and false negatives. It has the cloud intelligence required to recognize when surge in internet traffic is an attack, and when it is just a normal behavior from cloud sources such as Facebook or Amazon. Because SPs understand how cloud applications and services flow to and through networks, they quickly can identify the presence or potential for DDoS attacks, along with the cloud and IoT sources that are responsible for them. Software alerting drives action – routers can be called upon to drop traffic at the edge, or traffic can be sent to mitigation devices for more stateful analysis. (Figure 1.)

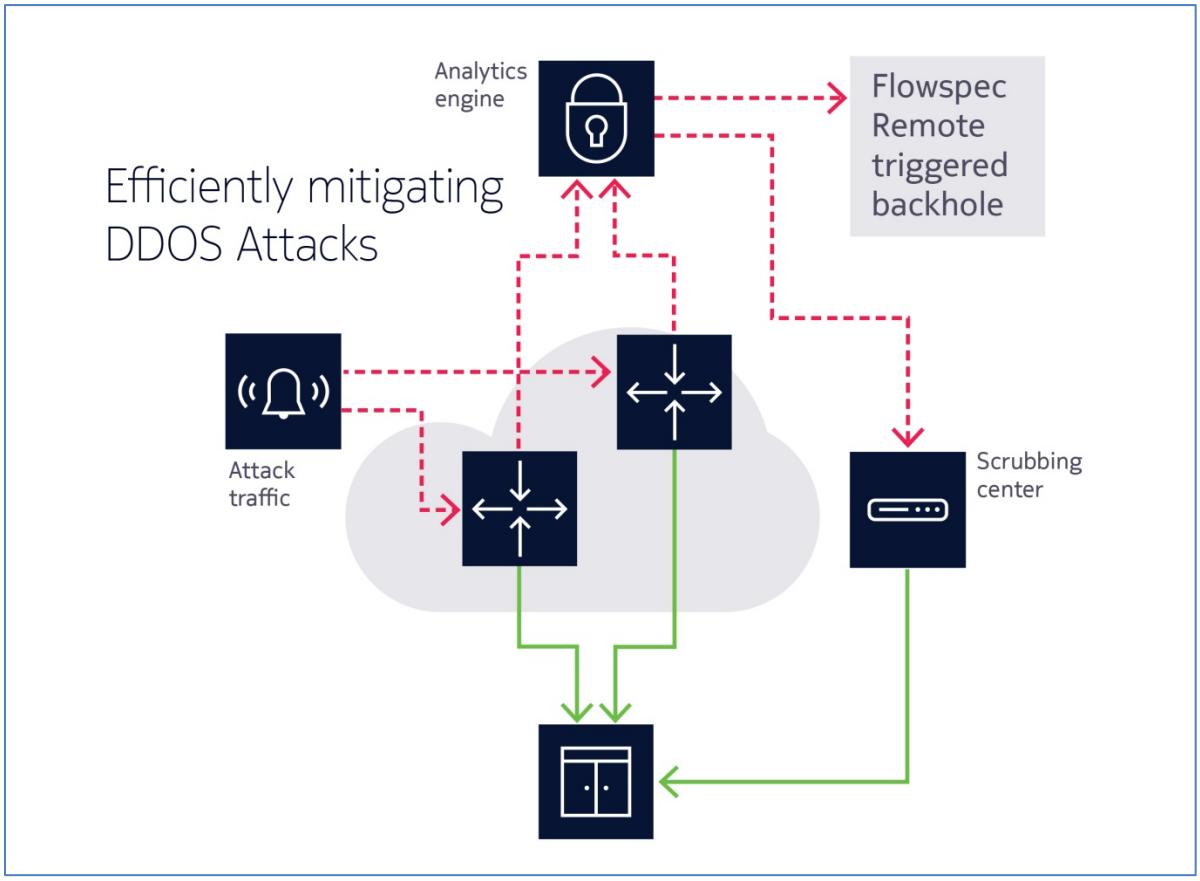

Figure 1 – Real-time insights allow ISPs to trigger on specific events and selectively mitigate DDoS attack sources.

1.3. Harnessing Big-Data Analytics to Drive SDN/NFV

Solving the network and service visibility problem requires coupling big data analytics with the dynamic control capabilities of software-defined networking / network function virtualization (SDN/NFV) orchestration platforms. Together, these elements become the cognitive "brain" that is capable of making real-time, automated corrections to critical networks so they can quickly adapt to changes in application demand, flow and traffic patterns, as well as automate actions that ensure ongoing service health and customer satisfaction. This all allows SPs to drive greater network efficiency, help assure quality and enhance security – all without manual intervention, and in real time.

Central to this process is automated monitoring, flagging and mitigation – the capability to set alerts on performance drops by any number of multidimensional variables, including customer segment, time of day when average bit rate (ABR) drops for a group of customers, and other key performance indicators. This requires rich analytics to drive the dynamic network configuration of software-defined SDN/NFV controllers.

2. The New Dimensions of Network Intelligence

Operations teams today typically have a lot of toolsets that primarily are focused on low-level sensing pieces, for example, "Is this interface up or down? How many tetrabytes are on the network currently? How is the CPU utilization?" and other variables. These are all useful metrics, but they don't divulge any business-level data such as how Netflix is streaming in Peoria, or whether Chicago is up or down, or whether you are seeing something different on this peer from what you saw yesterday. In isolation, these low-level metrics don't answer anything about customers' services, as they aren't mapped to today's business questions.

Effectiveness in today's business and technical landscape requires customizable network intelligence that offers offer several dimensions of capability:

- Network Intelligence: Visibility into tens of thousands of applications, without probes, in order to fully enable network optimization and informed discussions with content/network partners.
- Service Intelligence: Service assurance for cloud (OTT) applications ("How is Netflix or YouTube doing?"), while improving customer QoE and reducing churn, thereby lowering troubleshooting and support costs.
- Subscriber Intelligence: Visibility into granular usage dimensions such as daily traffic by subscriber (tonnage, category, site), which subscribers go over their plan tier and by how much, and use patterns – providing essential insight for how to give better service, bill more accurately, retain customers and predict cord cutters.
- Security Intelligence: Deeper, cloud-aware analytics, better attack detection, and more precise attack mitigation, minimizing the need for scrubbers, and eliminating the need for detection hardware.
- Operational Intelligence: Actively leveraging the same visibility and insight provided by all of these tactics to create custom alerts on business-level events.

All of these multidimensional analytics drive SDN/NFV operations, enabling dynamic network optimization and creating new service opportunities that will allow ISPs to better align network performance with customer expectations.

This new paradigm allows providers to bring topological and logical attributes to that lower-level data. Script-based application programming interfaces (APIs) allow them to answer questions about customers and their service, or above that, business-level objects, such as the performance of the OTT service as just a portion of the traffic – for example, Hulu on a certain node, or the YouTube peer. Simply being able to answer questions like that, and alert on abnormalities, is fundamentally a game-changing way of looking at the data. (Figure 2.)

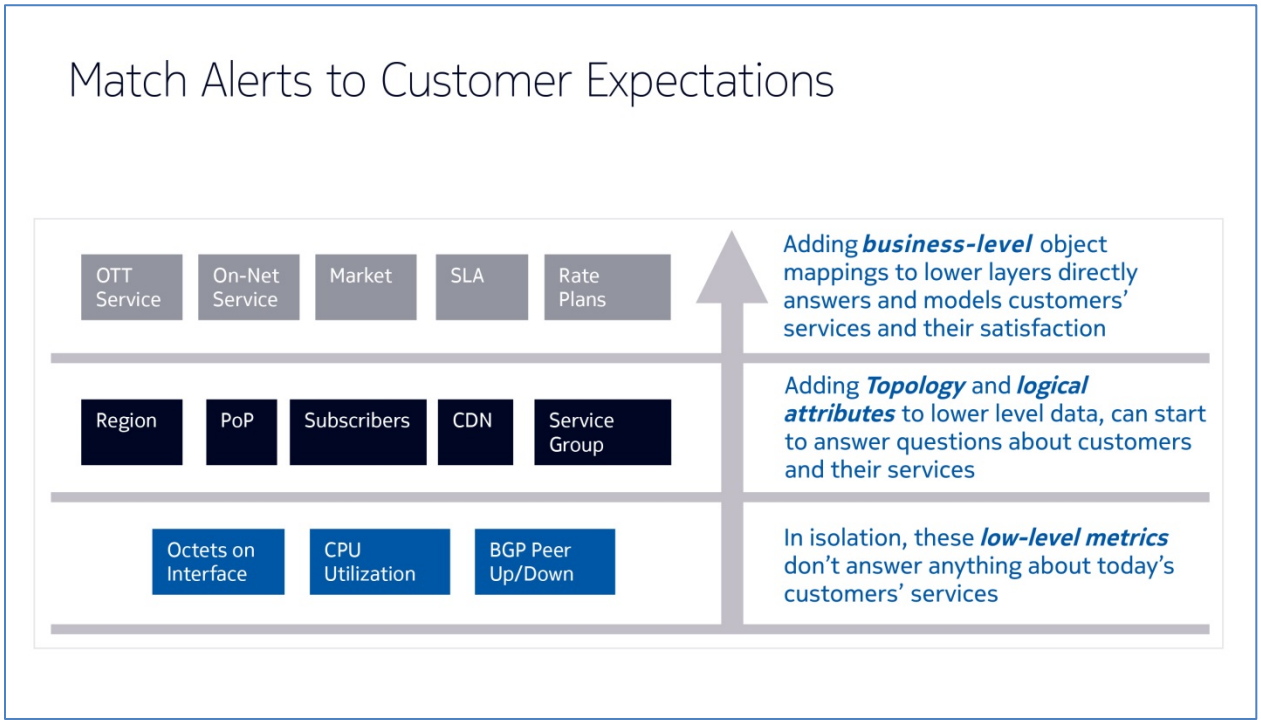

Figure 2 – Networks can be dynamically configured to better match alerts to customer expectations.

By employing an automated, dynamically configured network using a highly scalable software solution to leverage big data allows ISPs to build real-time alerts based on any number of those variables so that they can be proactive rather than reactive. Ultimately, they will also be able to create automated responses as well. The end result is a 360-degree solution that better understands the network to see what changes occur, and then repositions or reconfigures the network to make it more optimal for users.

2.1. A Modular Approach

One effective way to achieve this level of network intelligence is to combine several modules that work seamlessly together. An effective architecture includes a big-data engine/software platform; a massive map of the global service supply chain that adds visibility to all applications built onto the core platform; an analytics application that provides end-to-end network visibility and context-aware content

engineering; an analytics application that monitors customer QoE in real time; and a security module to enable real-time DDoS detection and mitigation.

A petabyte-scale big-data analytics engine can provide visibility into vast numbers of cloud applications and services, along with billions of IP addresses, tracking how traffic runs to and through networks to reach subscribers, in real time, and without the need for expensive probes, taps and monitors. Utilizing an ultra- high-capacity processor, it can analyze data against dynamic baselines based on network traffic, then provide alarms and, using existing router infrastructure to mitigate a bulk of traffic via flowspec or remote-triggered black hole, then take necessary actions when normal traffic flow and usage is disrupted.

This level of real-time, analytics-driven network and service automation can provide ISPs with greater network and application insight, control and DDoS protection.

2.2. Improving Network Performance

This approach takes advantage of information that is already available in internet infrastructure. By studying cloud applications and services, providers can unravel their supply chains to see what the related IP addresses are, where they're located and how they interact. When an IP flow reaches their network, those providers won't need DPI to tell them what application or service it is, how it landed on their peering router or how it traverses their network.

Armed with this data, operators can overlay this information onto their own topology to understand what traffic is on the network, where it is and its impact, even if it is encrypted—a level of visibility that's imperative for accurate performance management and quick identification of configuration issues. This approach also enables network issues to be resolved before customers complain about poorly streamed content.

2.3. Leveraging Distribution

This new level of dynamic network management employs multiple processors on a single multi-core unit – more on a single server than used to be available across an entire country. They operate as horizontally scalable clusters, providing resiliency that extends beyond what happens on any single server. With horizontal scaling, the cluster can adapt the resources and the compute as ISPs scale their services. This distributed file system provides the advantageous properties of sharing and replicating the data, fault tolerance and redundancy. Under this architecture, data can be *normalized,* meaning that domain name server (DNS), performance and topology data is converted into vectors, which then can be distributed across the streaming database, allowing all of the different parts of the cluster to operate on that data in parallel, both on the ingest and on the query. With these many dimensions at their fingertips, operators have the context needed quickly dive in deep and decipher service issues in an extremely flexible fashion. (Figure 3.)

Ultimately, this approach accomplishes three things:

1. Flexible pane-of-glass views across multiple data sources, enable extremely intelligent and deep analytics for an entire infrastructure.
2. Real-time correlation and an intelligent view into the data supports dynamic configuration.
3. Real -time alerts based on triggers and dynamic baselining allow operators to act on network drops or DDoS attacks in seconds.

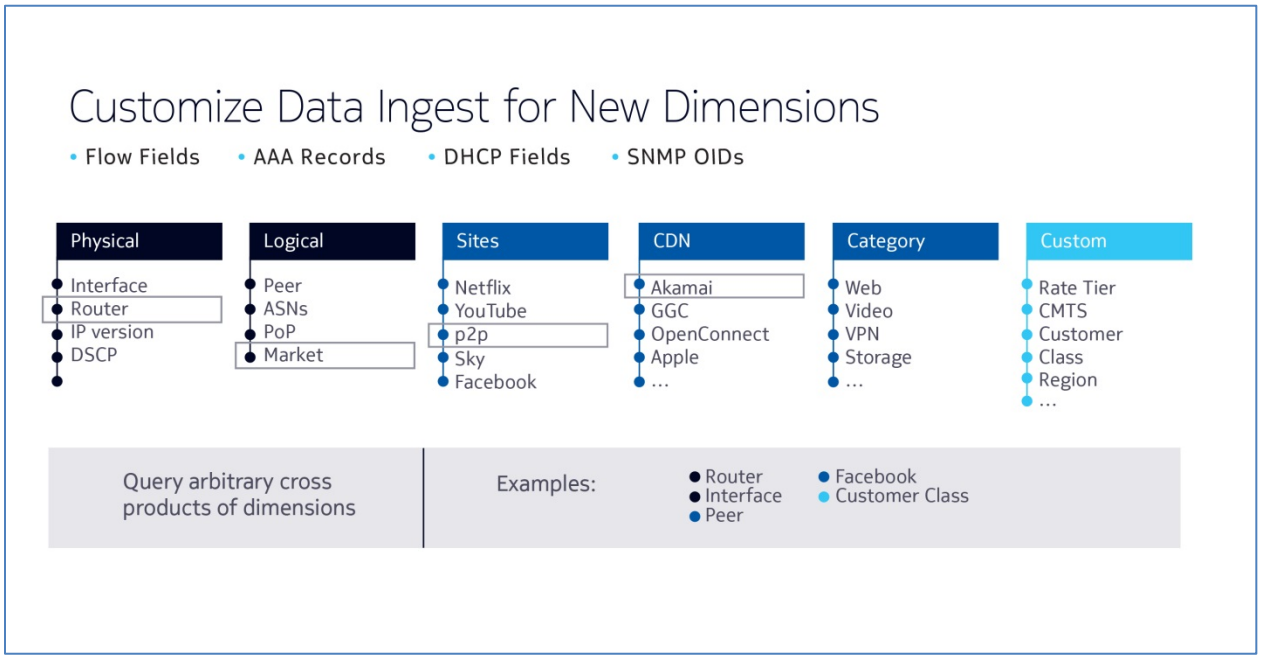

Figure 3 – Providers can use APIs to customize data queries across multiple dimensions.

2.4. Becoming Proactive Rather than Reactive

Typically, customer support teams actively monitor the network, as they have for many years, and when they determine that it is "green," they return to other daily tasks. In the current era of cloud and IoT, however, it is not always enough to simply see that routers are behaving correctly – there could be higher-level issues afoot that are not able to be tracked with old monitoring tools. One example is a network operator that went through the complete checklist of tasks needed to ensure the network was in order, yet turned to Twitter to learn that an entire market had lost their access to a popular OTT service.

This caused an uncomfortable scramble to solve an issue far after it had negatively affected a multitude of customers. This is a scattershot, reactive approach that does no favors for the customers or the business. Using the new model of actionable intelligence, operators receive alerts through external systems such as Webhook, email, syslog, and simple network management protocol (SNMP) traps. They also can utilize flexible alert settings using baselines, trending and thresholds, creating a proactive rather than reactive service environment.

As an example, consider the situation where there is a set of data on a router, providing a baseline for operation and performance. Traditionally, the provider has had to pull that back every 5 to 10 minutes. Now the industry is moving into telemetry streaming, where the provider will set some kind of persistent query that can put into the network, setting alerts for anomalies. One typical set of parameters would essentially translate to "Look for all the video data between Houston and Washington, D.C. and flag any kind of anomaly in that particular part of the network." The ISP will have an instant alert for that. When troubleshooting the root cause, operators can set alerts for Webhook, email, syslog, SNMP trap and other variables – whatever works in the system from a forensics perspective. The idea is that the

operator can build up different kind of queries to answer different questions about what has changed in the network, or what will trigger an alert and some kind of operations response. (Figure 4.)

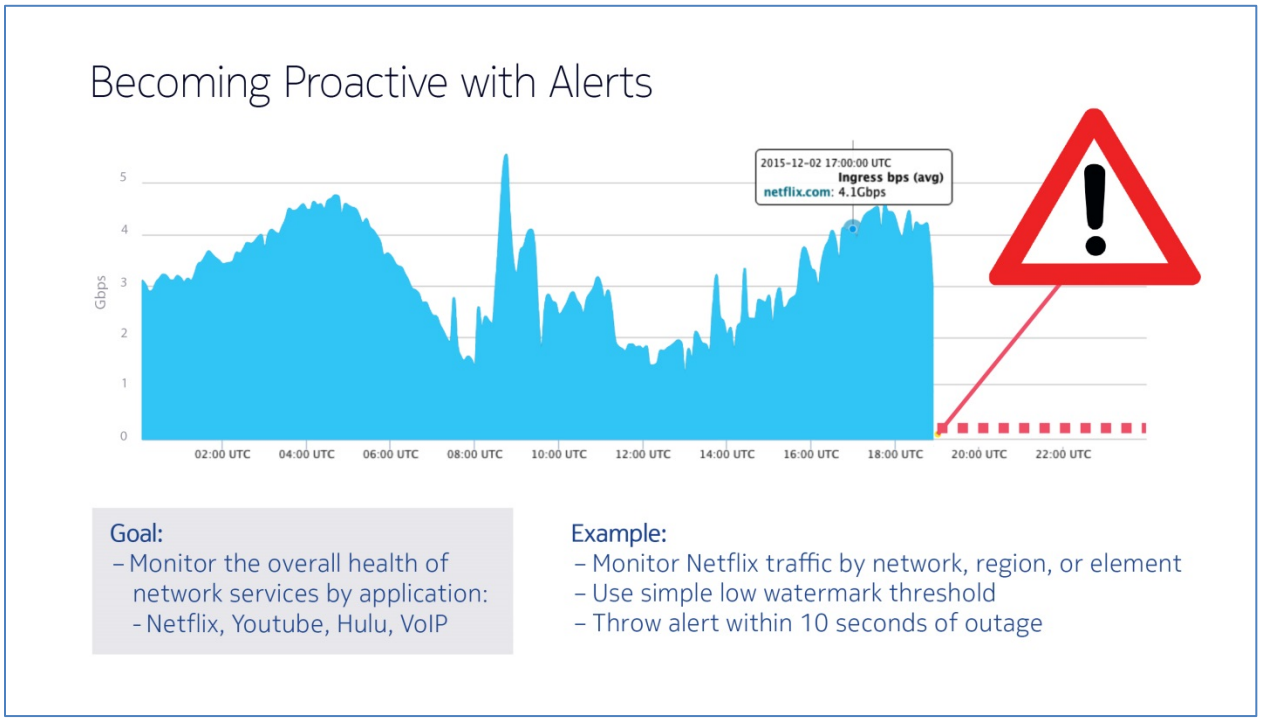

Figure 4 – Alerts based on a vast number of variables can monitor network health and trigger focused, proactive resolutions.

3. Use Cases

Many SPs are already benefiting from enhanced real-time network analytics to increase subscriber satisfaction, reduce churn and secure their networks (Figure 5.) A few examples follow.

3.1. Enhance Video Performance Drops by Market

As video content consumes increasing amounts of available bandwidth, ISPs must build out their networks to keep subscribers happy. However, they have a severe lack of visibility that prevents them from solving or proactively avoiding problems. ISPs need a way to understand how applications perform across all parts of the network so they can quickly identify business level events such as “Netflix is down in Chicago” for quick remediation.

Unfortunately, IP network analytics were not built with web-scale services in mind. ISPs have historically collected both application and network data. But they have stored this data in silos, and have not had sufficient cross-correlation to identify the specific OTT streaming issues caused by congestion on a particular link. This lack of insight has made troubleshooting a very costly and inefficient process that has done little to improve overall service quality.

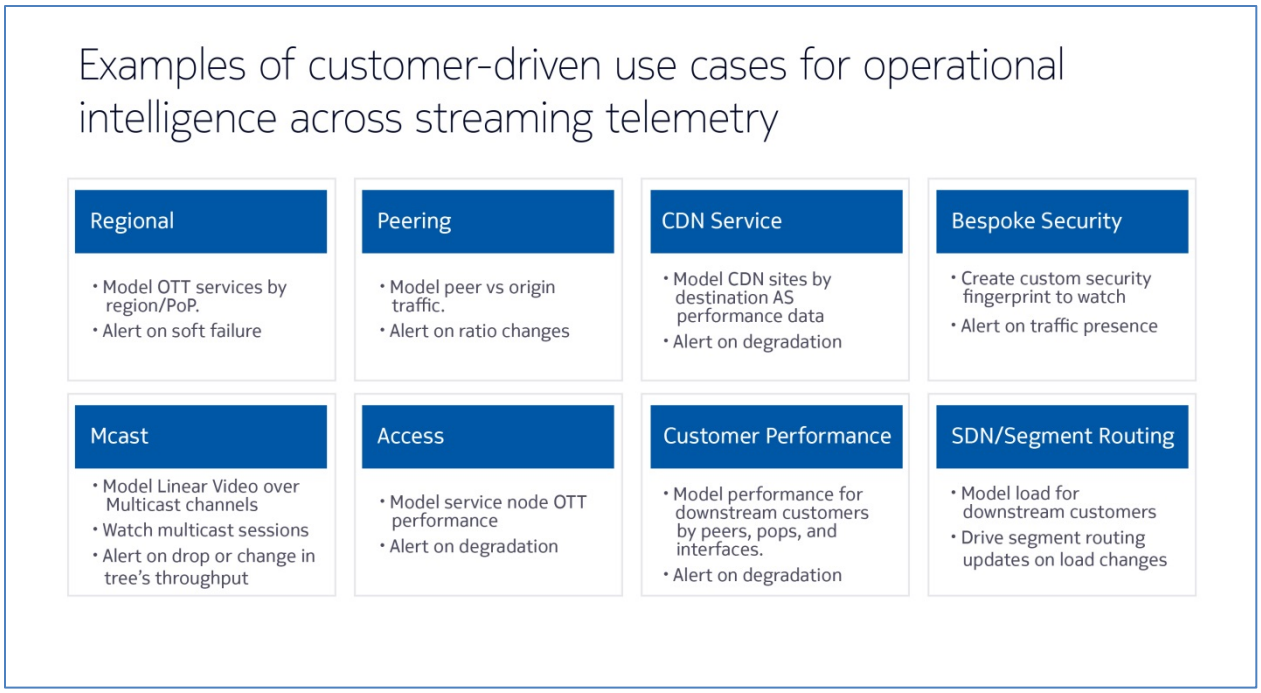

Figure 5 – Leading ISP and OTT providers are already using multidimensional intelligence for understanding and dynamically managing their networks.

One top ISP attempted to solve the visibility problem with DPI appliances. However, it was far too expensive across its entire network, and large swaths remained unmonitored. Since DPI is blinded by encryption, it provided no visibility into well over half of all video traffic flowing through the network. The ISP recognized that it needed a new software-based solution that could provide full multidimensional visibility and be far more cost-effective than customized hardware.

This ISP addressed the challenge with a massively scalable software-based solution, which identified applications and rapidly correlated those findings with all network data to immediately identify problems – all without looking at a single packet. This multidimensional analytics approach enabled the ISP to instantly monitor high-level events and visualize the impact of video traffic across any part of its network, alerting on any drops in any streaming degradation, and allowing speedy resolution of issues before customers had the chance to complain.

Using this solution, the ISP can now categorize every single flow to gauge and dramatically improve network performance by surgically adding the exact amount of bandwidth needed in just the right places. This ensures that all subscribers are receiving their desired content with the best possible quality. (Figure 6.)

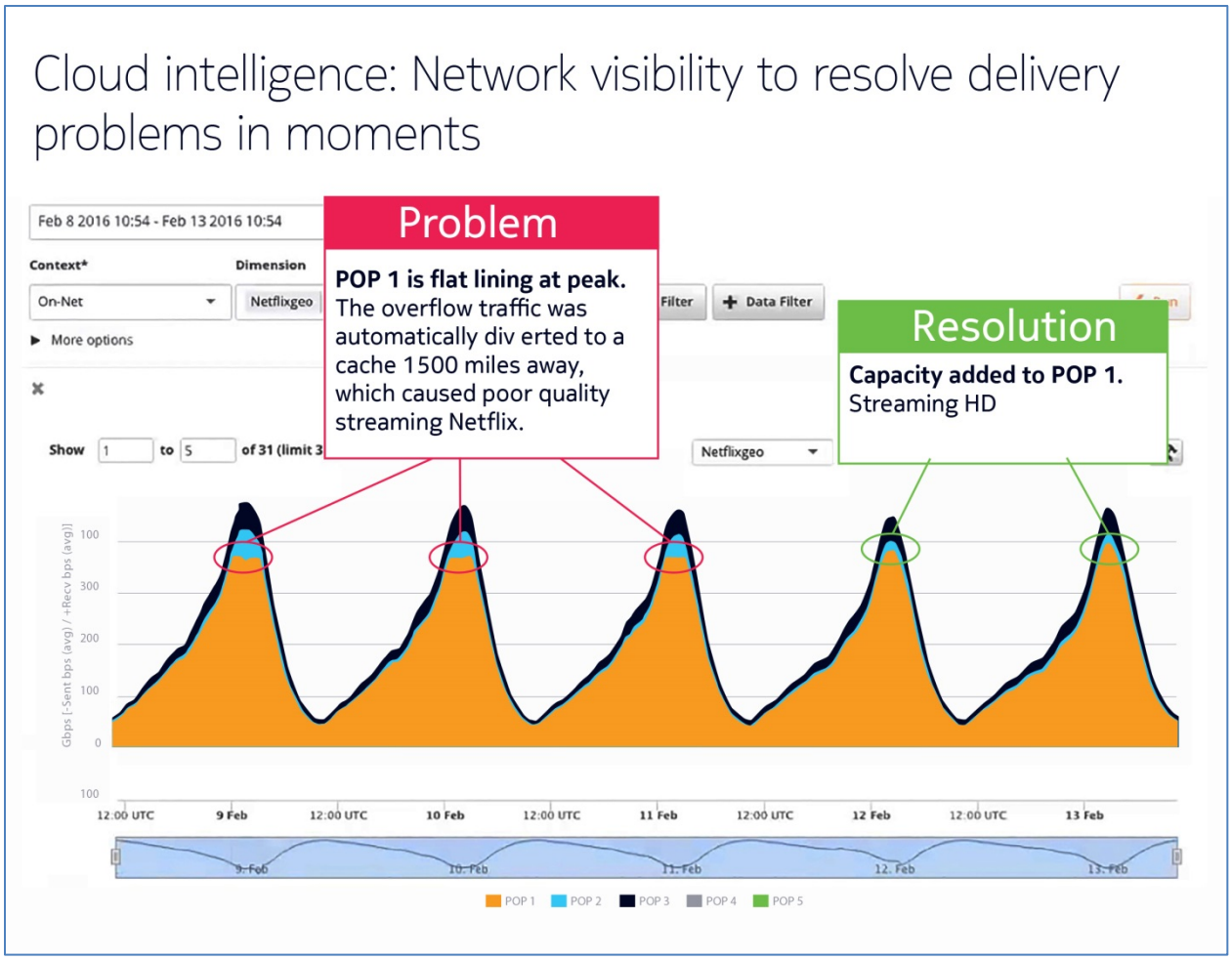

Figure 6 – SDN-based operational intelligence improves visibility to instantly identify and solve problems.

3.2. Take Immediate Action on Peering Misconfigurations

Historically, ISPs peered directly with one another to exchange traffic, and their contracts were based on tonnage exchanged. As the internet has become more complex, so have peering agreements. Ultimately, everyone involved is working together to deliver content to the end user when and where they want it, at the best quality. If peering relationships do not follow their contractual agreements, it could easily lead to clogged interfaces and poor streaming quality for the end user. Because of this, contracts detail what traffic can traverse which interface and when. It is imperative for an ISP to have a comprehensive view of its network that understands what traffic is flowing where if it is to actively monitor these relationships.

One ISP deployed a solution that alerted on configuration changes, and has found it to be highly beneficial. In one instance, Dropbox was just emerging, and at first used Amazon Simple Cloud Storage Service (S3) to distribute its content. However, as it grew, Dropbox decided build its own content delivery network (CDN), and began distributing traffic from both. This would look like suspicious brand-new traffic to a network that could not map and understand the change in IP flows. However, in this case an

operator could immediately see that it was the same traffic under a different name, and marked it as a valid traffic shift.

The same approach that identified all traffic and where it was traversing the network immediately alerted network operators to a change in peering traffic. There was a misconfiguration that caused peering traffic to enter the ISPs network on the incorrect interface, leading to a sharp drop in traffic entering at one port while flooding another. The real-time alert allowed for immediate intervention.

This same capability also can monitor the overall health of peers, alerting ISPs when performance thresholds are not met. (Figure 7.)

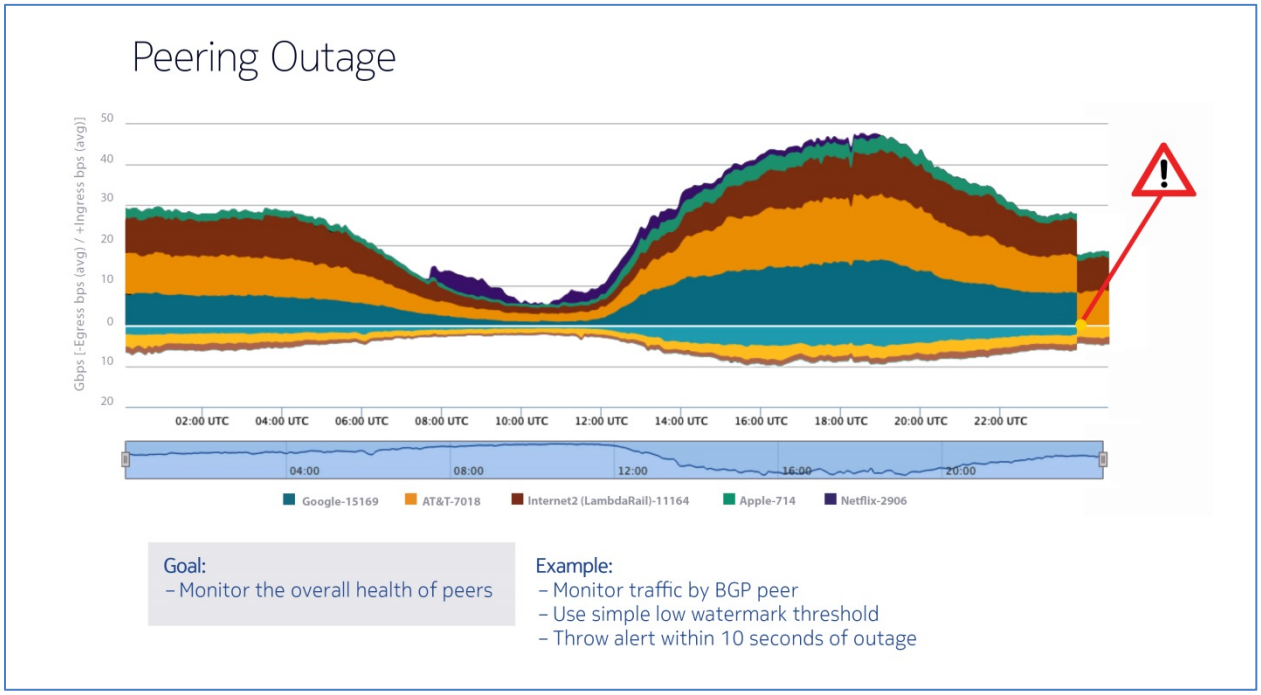

Figure 7 – ISPs can moitor peer health by setting a simple low watermark threshhold.

3.3. Leverage Insight-Driven DDoS Mitigation

DDoS detection appliances are struggling to keep up with the increasing frequency, size and complexity of attacks. Only a software solution that leverages the same visibility to alert on performance drops can provide the necessary insight to protect even the largest networks while lowering OPEX.

By combining multiple sources of data to contextualize and identify the exact origin of all traffic traversing a network, software can most accurately flag and categorize that which is associated with a DDoS attack. Then, it can whitelist traffic that has been known to create false positives (such as Facebook traffic from a geo-location at a given time). This, in turn, drastically reduces the amount of traffic that must be monitored, and the amount that would be mistakenly backhauled to scrubbing centers. This further reduces the need for costly, specialized devices.

The benefit of accurate detection is easy to see in an example such as this. A digital video recorder (DVR) might sporadically send one or two DNS requests, and a system with this aptitude understands the context in which a DVR must operate. Therefore, if a DVR suddenly begins flooding a DNS server with requests, the software will create a real-time alert for security operators to investigate. The operator then has the choice of which type of mitigation should be used. It can be dropped at the edge with router mitigation, or a portion of the traffic can be diverted to a partner scrubbing appliance to be cleaned and reinjected into the network.

This new and innovative concept is able to immediately identify attackers based on intuitively derived traffic fingerprints so the attack traffic can be cut off at the edge of the network – before it affects a single target.

4. Conclusions

The internet is constantly evolving. We have gone from T1 to the era of cloud computing. We are seeing explosive growth, including that of IoT and new security challenges. This is the most interesting time there has ever been in terms of how quickly the nature of traffic is changing, and how dynamic network management has become.

We live in the world where disk, memory, CPU and computation are essentially infinite, and where market priorities have radically changed. Traffic management now is not just looking at the peer, but at dynamic data flows coming from across the world. A problem may be generated though a vast number of internal or external sources, whether from OpenConnect, traffic flowing through metro-scale datacenters, or even with a DDoS attack. The bottom line: data traffic now is far more dynamic and requires new sets of technologies to understand it.

All of this means that ISPs need visibility and operational intelligence throughout the network – not just at a single point, but end-to-end, with the capability to radically scale capacity in the most cost-efficient manner.

Networks now are competing on the management capabilities of those running them, the cost of operations and the ability to drive superior customer QoE. Effective management and protection is not just about managing data traffic in one or two dimensions, but about how ISPs build visibility into the edge data centers, the level of inner-connection in the edge, and their focus on customers.

With a new approach that considers every piece of the network puzzle, service providers can more accurately, cost-effectively and efficiently manage their networks. That, in turn, will save them a substantial amount of money by doing away with unnecessary buildouts, protecting critical assets and reducing customer churn.

Abbreviations

ABR	average bit rate
AP	access point
API	application programming interface
BGP	border gateway protocol
CDN	content delivery network
DDoS	distributed denial of service
DNS	domain name server
DPI	deep packet inspection
DVR	digital video recorder
HD	high definition
IoT	Internet of Things
ISP	internet service provider
NFV	network functions virtualization
OI	operational intelligence
OTT	over the top
POP	point of presence
QoE	quality of experience
RTBH	remotely triggered black hole
SDN	software defined networking
SNMP	simple network management protocol
S3	Amazon Simple Cloud Storage Service
SP	service provider
VSP	virtualized services platform
WAN	wide area network

Bibliography & References

Bell Labs Consulting's Inaugural Mobility Report https://pages.nokia.com/1503.bell-labs-mobility-report.html

DDoS Attack Report, Cybersecurity Ventures http://bit.ly/2uA0b8A

Nokia Acquisition and Retention 2016 Study
http://www.mediatelecom.com.mx/~mediacom/media/pdf/adquisition-retention-nokia-2016.pdf

Nokia FP4 Routing Chipset, New Routers and New Operations Methods; Appledore Research Group
https://resources.ext.nokia.com/asset/201327

[1]*Verisign Q1 2017 DDoS Trends Report*
http://forms.verisign.com/Q12017DDoSTrendsReport?utm_medium=Blog&utm_term=internal

