ACG RESEARCH

Powering Intelligent Network Services with Real-Time Visibility, Analytics and Automation

EXECUTIVE SUMMARY

The rapid migration of content, application and services to the cloud is driving network transformation. Consumers have embraced over-the-top video streaming, online gaming and a deluge of mobile applications. Enterprise IT managers are leveraging software-as-a-service while migrating mission-critical applications to hybrid and multi-cloud environments. At the same time, a proliferation of Internet of things (IoT) devices is driving exponential growth in the number of end points and network connections, most over wireless networks, where mobile operators are planning new 5G services for low-latency, high-bandwidth use cases leveraging computing resources closer to the network edge.

Cloud-native applications and services are highly dynamic and distributed across hyperscale, software-driven, virtualized application and network infrastructure. Continuously shifting traffic flows and bandwidth demands are placing a premium on the ability of network service providers and webscalers to respond rapidly to changing network conditions, driving an urgent need for intelligent network automation to overcome serious challenges related to service assurance, end user quality-of-experience and security.

This paper examines four key building blocks for powering insight-driven network services that combine to provide the foundation for a wide range of use cases leveraging intelligent network automation. The paper concludes by describing three practical scenarios for streamlining network operations and traffic engineering.

KEY FINDINGS

Four key building blocks power intelligent network services:

- A more capable data plane and highly instrumented network fabric
- A simplified and programmable control plane
- Unified, multi-domain, multi-layer network topology visibility
- Big data analytics to derive real-time intelligence from a wealth of data sources

Insight-driven network automation enables service providers and webscalers to overcome serious operational challenges created by cloud-native services and IoT

SOFTWARE, CLOUDS AND IOT ARE TRANSFORMING NETWORKS

In a widely quoted Wall Street Journal essay, Marc Andreessen wrote in 2011 that “software is eating the world.” Today, network service providers are confronted with the full effects of this phenomenon. Open source software running on merchant silicon-based hardware now powers the largest global data centers at webscale giants such as Google, Facebook, Amazon and Apple. Consumer and business users are benefiting from cloud-based delivery of a vast array of content, applications and services. In less than a decade, the market landscape has shifted dramatically with the rise of a new generation of cloud-native businesses, such as Uber and Spotify, that reach customers exclusively via the Internet. With streaming video traffic dominating the Internet, Netflix and other content providers have built out vast content distribution networks (CDNs) that overlay the Internet backbone.

Cloud computing is transforming enterprise IT with software-as-a-service (SaaS), a widely accepted model, but businesses have also started to migrate critical applications and data to public and hybrid clouds. Cloud-centric enterprise IT, software-defined networking (SDN) and network functions virtualization (NFV) are enabling the adoption of SD-WAN services for flexible network overlays that connect users to applications in the cloud. The proliferation of connected devices for sensing and controlling the physical world is driving exponential growth in Internet of things (IoT), and mobile operators are planning to roll out new 5G services to support a wide range of IoT applications.

These changes are driving network transformation from the edge to the cloud, impacting both traditional network service providers and the webscalers that rely on many different networks to deliver services to users from the cloud. Figure 1 shows the different types of networks involved in connecting end points to the cloud. Service providers manage access and IP transport networks and webscalers manage connections on the cloud side that typically traverse some combination of content distribution networks, peering and transit networks. End-to-end connections span a complex multi-domain, multi-layer topology with paths traversing both the IP and optical layers. In addition, these paths and the network topology itself are constantly changing in response to network conditions and application behavior.

Both service providers and webscalers face the predicament of operating in this environment while relying on switches, routers and network management tools designed in a bygone era of residential Internet access, private enterprise WANs, static optical transport and an Internet that served mainly email, chat, e-commerce and web browsing applications. Service providers and webscalers have an urgent need for software-driven, intelligent network infrastructure that allows them to meet the operational demands of networks supporting cloud-centric applications and services.

SOFTWARE-DRIVEN APPLICATIONS AND SERVICES CREATING NEW CHALLENGES

Cloud-scale infrastructure is software-driven across multiple layers, with open source software frameworks facilitating the DevOps deployment model for applications composed of microservices, while SDN and virtualization techniques are pervasive across networks both inside and outside of the cloud. Software-driven infrastructure increases network agility, flexibility and scalability but also drives increasing complexity, resulting in a new set of challenges for network operators.

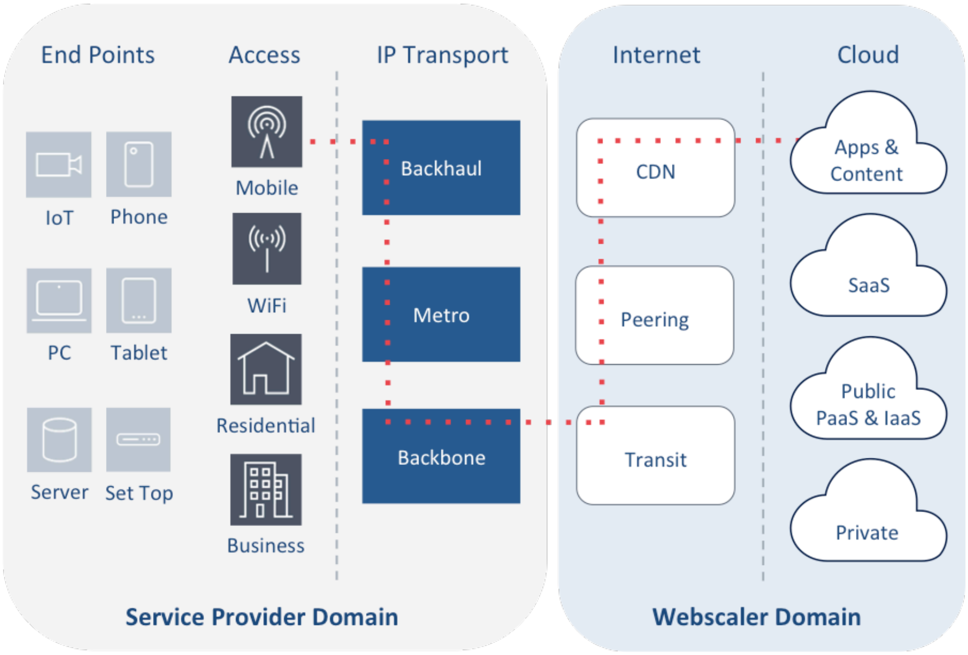

Figure 1. Flows Traverse Complex Paths from End Points to the Cloud

In public and hybrid cloud environments, microservices applications are highly dynamic and distributed, with traffic flowing between many different services that may span multiple data centers. As businesses adopt SaaS and migrate applications to cloud computing platforms, enterprise IT traffic will continue to shift to the Internet, driving traffic to many different data centers in the cloud. The net result is increased variability and unpredictability in terms of traffic patterns and network load as application usage ebbs and flows across the globe every day.

Networks support an increasing variety of end points accessing the cloud, including mobile devices, laptops, home gateways, enterprise CPE and a vast array of IoT devices and sensors. IoT will increase the number of end points on the network by several orders of magnitude and will drive more traffic into the cloud. Streaming video accounts for about 70% of network load and typically traverses a complex network path from content provider to consumer. Real-time voice and video communications are moving to cloud-based delivery and network operators are challenged to deliver these services without impacting user quality of experience (QoE). SD-WANs carrying enterprise traffic over the Internet can react rapidly to changing network conditions and business needs by bringing up new connections on demand, resulting in more dynamic, unpredictable traffic patterns.

In a cloud-centric world, security is obviously a major challenge for all concerned: consumers, businesses, service providers and webscalers. Services and applications are being delivered from publicly accessible platforms spanning multiple networks and data centers, presenting a vastly larger attack surface. In addition, the full stack of supporting infrastructure is becoming increasingly software-driven, leading to additional points of vulnerability in the attack surface. At the same time, the rapid growth in the number of IoT devices has unleashed the potential for hackers to create huge botnets that can be harnessed for malicious and crippling DDoS attacks.

Given these challenges, there is an overarching need for service providers and webscalers to gain real-time visibility into network behavior and extract the actionable intelligence needed to react immediately to performance anomalies, changing traffic patterns and security threats. This involves taking action leveraging insight-driven network automation mechanisms or at minimum, simplifying operator workflows guided by the insights extracted. More specifically, operators are typically preoccupied with these challenges:

- Root cause analysis to rapidly detect, isolate and recover from errors and faults
- Performance monitoring for service assurance and end-user QoE
- Traffic engineering and network optimization for efficiency and performance
- Security threat detection and mitigation, including DDoS attacks

Service providers and webscalers are dealing with these challenges while operating in an environment where they lack end-to-end visibility into the complete service delivery path from cloud to end point, which may span multiple clouds, data centers, CDNs and networks. In addition, because of the widespread adoption of end-to-end encryption, they are also unable to access encrypted packet payload data that could be used to gain visibility into this path.

BUILDING BLOCKS FOR THE INTELLIGENT AUTOMATED NETWORK

Four key investments will help service providers overcome these operational challenges and power the delivery of intelligent, automated networking services: a more capable data plane and highly instrumented networking fabric, a simplified and fully programmable control plane, unified network topology based on multi-domain, multi-layer visibility and real-time Big Data analytics. Throughout all four areas, access to information is paramount, beginning with instrumenting the network fabric at fundamental and granular hardware and software levels. Streaming telemetry is used to deliver network state, flow metadata and other visibility data. Operators gain visibility into applications and service utilization without dedicated Deep Packet Inspection (DPI) platforms. The result is a network that eliminates the need for expensive hardware probes while increasing visibility and network intelligence, even with end-to-end encryption for applications and flows.

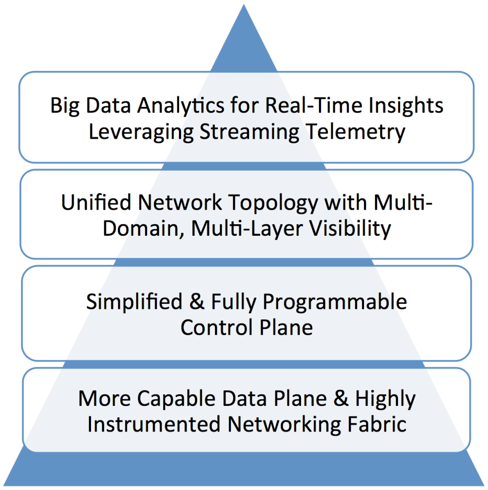
Big Data Analytics for Real-Time Insights Leveraging Streaming Telemetry
Unified Network Topology with Multi-Domain, Multi-Layer Visibility
Simplified & Fully Programmable Control Plane
More Capable Data Plane & Highly Instrumented Networking Fabric

Figure 2. Building Blocks of the Intelligent Automated Network

A More Capable Data Plane and Highly Instrumented Networking Fabric

At the heart of IP/MPLS routing and switching products for service provider networks is the data plane. The ability to read, manipulate and act on flows and packets is at the center of vendors' developments. However, with so much recent focus going into the separation of data plane and control plane as part of the journey toward SDN, it is easy to lose sight of the need to evolve and enhance the data plane. Multi-dimensional data plane evolution is critical and includes increased bandwidth, increased logical scalability, real-time performance monitoring and real-time flow replication.

Overall IP traffic is expected to grow at 25% compound annual growth rate (CAGR) through 2022.[1] Peak data rates are growing even faster, at nearly 40% CAGR.[2] Optical networking equipment is delivering 100G, 200G and now 400G single-wavelength connectivity with advanced coherent dense wave division multiplexing solutions. The Optical Interworking Forum (OIF) is working on making a sub-120 km, 400G single-wavelength solution economical with its collaborative 400G-ZR specification activity. The OIF also announced support for a new generation of link aggregation with the adoption of the Flex Ethernet Implementation Agreement 1.0 in March 2016. Flex Ethernet enables bonding of multiple subrate paths to create the appearance of a single, larger interface. In addition, the IEEE 200GbE and 400GbE Task Force continues its work on the 802.3bs draft specification that includes support for 400Gb/s Ethernet interfaces. With evolution to 100G, 200G and 400G interfaces and continued bandwidth growth, there is a clear requirement for multi-terabit packet processing engines in next-generation networks. There is also a need for programmability. To maximize optical network performance and adapt to changing traffic patterns and networking conditions, the underlying optical infrastructure must provide programmatic and granular tunability to support modifications to transmission settings such as modulation schemes, baud rates and forward error correction algorithms.

Beyond raw bandwidth, the data plane also needs increased logical scalability. The diversity of connected devices (for example, IoT) and types of traffic requires an increase in the number and sophistication of traffic flow classifiers and access control lists (ACLs). The ability to establish classifiers for each flow through the packet engine and the sophistication to match on any portion or depth of packet header or payload provides maximum flexibility, enabling the data plane to be programmed to manipulate and expedite the right traffic while blocking malicious or aberrant traffic from entering the network right at the ingress point.

To measure total QoE and diagnose potential issues, operators must be able to look at performance throughout the network path from source to destination. Real-time, per-flow statistics and metadata collection from IPFIX, NetFlow, sFlow and other telemetry enable performance monitoring and service assurance software to measure, analyze and diagnose per-flow performance.

Real-time flow replication at scale is also a requirement. With the overall size of interfaces moving toward 400G and beyond, simply turning on port mirroring, which might have been done in the past, is not appropriate and can create its own set of network scalability issues. With IoT and hybrid cloud service delivery, network security perimeters are dissolving. Operators need the ability to isolate and replicate traffic on any port or flow in the network at any time for additional analysis and scrubbing.

[1] Nokia Bell Labs Traffic Forecast 2017–2022.

[2] Nokia Bell Labs Traffic Forecast 2017–2022.

With real-time per-flow replication, the network can continue forwarding traffic for a flow while in parallel sending identical traffic through a secondary analysis engine. If the secondary analysis positively identifies a virus or DDoS attack, additional action can then be taken to install granular ACL rules to isolate and block the original flow to prevent network impact.

Although merchant silicon network processors have evolved significantly since 2000, in the near-term, merchant silicon will not achieve the combined multi-terabit throughput and logical scalability performance as custom ASICs. ACG Research expects industry-leading IP/MPLS routing and switching vendors to provide custom ASIC developments. Nokia's recently announced 2.4 Tb/s FP4 silicon chipset with support for millions of flexible classifiers and ACLs is one example of a solution that brings raw scalability, fine-grain visibility and control and improved security mechanisms to next-generation data planes.

Simplified and Fully Programmable Control Plane

One of the original design tenants of SDN was the clear separation of data plane and control plane. Although SDN is being realized today differently than early advocates may have expected and with different protocols, two basic concepts remain at the heart of the SDN transformation: utilizing open application programming interfaces (APIs) to abstract the network and unified data modeling of network elements such as routers and optical gear to create topology and hierarchical control.

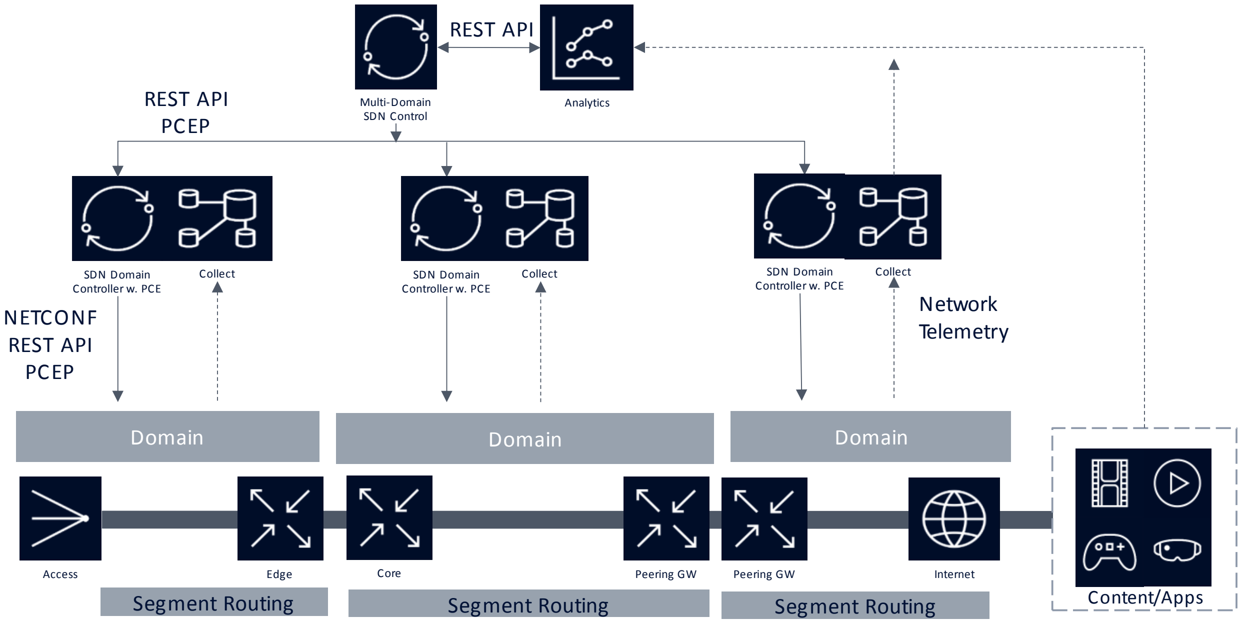

Figure 3. Hierarchical SDN Control Framework

As service providers migrate IP/MPLS and optical networks toward SDN, two powerful approaches have emerged: utilization of segment routing (SR) to reduce control plane protocol complexity and utilization of external path computation for increased scale and enhanced best path calculations. With segment routing, LDP and RSVP-TE can be eliminated from the network. Instead of utilizing RSVP-TE to distribute labels for traffic engineered paths, segment routing assigns SR I.D.s to nodes and links that are then advertised into the domain by the interior gateway protocol that is running. The headend node at entry to the network can then express a path through the network by utilizing a destination SID or a stack of

SIDs (segment list) to express a specific path from source to destination. Only the headend node is required to maintain flow state information. Intermediate nodes utilize the outermost label to forward the packet to the appropriate link or node.

By relocating path computation out of the headend node and centralizing it for a domain or multiple domains, operators can enable the path computation engine (PCE) to scale its compute capacity independently from the network elements while also providing a broader network view in determining the optimal traffic engineered or policy constrained path through the network. Although a PCE can be stateful or stateless, a stateful approach is most powerful because it maintains a synchronized view of existing label switched paths, thus enabling optimal initial path computation and future reoptimization to maximize network efficiency and performance. Industry-leading SDN controller solutions with PCE integrate real-time network link utilization, performance data and flow policy/intent to construct the best path through the network and perform dynamic network optimization to ensure service health and network efficiency. With protocol extensions as outlined by the Internet Engineering Task Force (IETF)[3] to enable the communication of SR segment identifiers (SIDs or labels) between the PCE and any client network elements (PCC or path computation clients) the path computation element protocol (PCEP) has gained recent industry support. PCEP also supports communication between two PCEs to support multi-domain or hierarchical deployments.

The combination of segment routing with external PCE is not limited to the network. Data centers will also participate in this solution mix. One can envision best-of-breed approaches bringing together segment routing and hierarchical multi-domain, multi-layer SDN control with path computation to deliver end-to-end control across the network and the data center from user to application.

Unified Network Topology with Multi-Domain, Multi-Layer Visibility

With traffic flows between end points and cloud-based applications and services traversing many different paths across multiple network domains and layers, both service providers and webscalers need to maintain real-time visibility into network state and topology.

Streaming telemetry is displacing SNMP as the preferred method for collecting statistics and state data from network infrastructure, overcoming the inherent scalability and performance limitations of polling. Formal methods for network configuration, such as the OpenConfig initiative, use YANG modeling for network configuration, the NETCONF protocol for distributing configuration data to network elements, and gRPC protocol buffers for streaming telemetry. These methods are integral to intent-based networking solutions that rely on telemetry-driven insights for network automation.

Service providers and webscalers need real-time visibility into network topology at multiple layers. Monitoring tools for multi-layer network topology ingest BGP and interior routing protocol data for a comprehensive view of the IP layer topology and available paths, which can be correlated with the topology of underlying transport networks, as well as the topology of virtual overlays. A cross-domain, multi-layer view of network topology is critical for root cause analysis, traffic engineering and network optimization applications.

[3] draft-ietf-segment-routing-09, IETF.

A recent advancement in topology visibility is the new BGP monitoring protocol (BMP), which provides a method for collecting BGP routes and statistics from BGP peers that previously had to be derived by screen scraping the CLI of a router connected to these peers. A key benefit of BMP is full visibility into all possible BGP paths as opposed to only the best paths advertised in peer router BGP updates, which is valuable for traffic engineering and peering analytics applications.

Big Data Analytics for Real-Time Service Orchestration and Operational Intelligence

Given the scale, complexity and highly dynamic nature of cloud-based services and applications, service providers and webscalers need to leverage Big Data analytics for real-time service orchestration and operational intelligence, ingesting, processing and storing a huge volume of data from a variety of sources at high velocity. This typically involves accumulating petabytes of data over weeks and months that can be analyzed rapidly for actionable insights. Figure 4 shows the wide array of data sources feeding a typical Big Data cluster supporting automated service orchestration, ad-hoc multi-dimensional analytics and operational applications such as service assurance and traffic engineering.

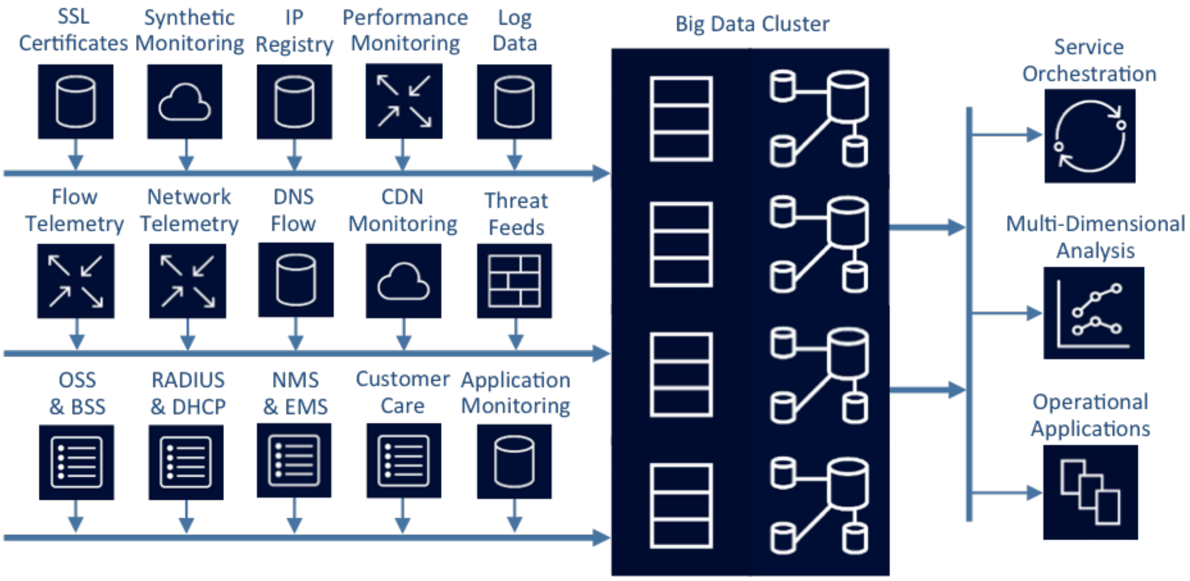

Figure 4. Data Sources for Big Data Analytics

Flow telemetry and network telemetry enable service providers and webscalers to monitor traffic flowing over various paths across multiple networks, which is critical for use cases such as detecting and mitigating against DDoS attacks and for peering analytics to select the most cost-effective peering and transit connections based on current and historic traffic trends. Big Data analytics on aggregated time series telemetry data provides network operators with real-time insight into network behavior and conditions, typically within seconds. Visibility into how much traffic is traversing paths from IP source to destination requires correlating BGP path visibility derived via route data ingestion with IP flow metadata collected from network elements using protocols such as NetFlow or sFlow. The latest generation of edge routers and high-end switches can export sampled flow metadata without significantly impacting forwarding performance, eliminating the need for external probes.

Network and flow telemetry data is time series in nature and well suited to real-time streaming analytics techniques and storage of retained data in a column-oriented database. Data from multiple sources is typically aggregated into a Big Data cluster front-ended by a high performance publish/subscribe pipeline that feeds data to analytics consumers within the cluster. For example, streaming analytics can be performed to detect behavioral anomalies, track performance trends and incorporate machine learning algorithms for historical and predictive analytics.

Column-oriented databases are an ideal way to collect and retain a high volume of many different types of time series data, with records of each type stored sequentially in separate columns in the database. This allows for efficient queries and rapid response when correlating data across multiple dimensions over specific intervals. Modern, well-designed column-oriented databases can respond to complex multi-dimensional queries in seconds.

Use cases such as traffic engineering, forensic analysis and regulatory compliance require access to time series data going back weeks and possibly months. Therefore, it is common for Big Data clusters to incorporate sufficient disk-based storage to retain large volumes of data for long periods. In addition, frequently accessed data for streaming analytics and real-time, multi-dimensional analysis is often stored in solid-state drives inside the cluster to boost performance.

Network and flow telemetry must be supplemented with and correlated against many other data sources, such as application and network performance monitoring data, syslog records, subscriber-specific OSS/BSS data and customer care data. These other sources provide the necessary context and business intent for different operational use cases and applications, providing service providers and webscalers with a more complete picture of real-time network conditions and behavior. For example, service providers can use large datasets of anonymized network utilization and subscribers' demographic data for service and capacity planning. A Big Data column-oriented database can readily incorporate these for rapid multi-dimensional analytics spanning a wide range of data types.

Big Data analytics frameworks are built on an extensive foundation of open source software with well-defined APIs between layers and components. Therefore, Big Data analytics solutions should be implemented using an API first approach in which all data ingestion and analytics functions are exposed via well-defined REST APIs. Figure 4 shows APIs being use to extract data from a cluster for service orchestration, multi-dimensional analytics tools and various operational applications. These APIs play a key role in insight-driven network automation, triggering the orchestration layer to take remedial action in response to failures, anomalies or changing network conditions.

Big Data is also critical for providing the context needed for service providers and webscalers to gain complete end-to-end visibility into specific applications, services and subscriber end points. Telemetry-based techniques for traffic flow visibility operate at the IP layer, but due to the complex paths that traffic flows traverse from source to destination, it is not straightforward to correlate flows with specific applications, services and subscriber end points. In the past, operators would use DPI probes to inspect packet payloads and identify flows, but this does not work for encrypted flows, and so operators need to use other methods for mapping server IP addresses to applications and services in the cloud and end-point IP addresses to subscribers, as shown in Figure 5.

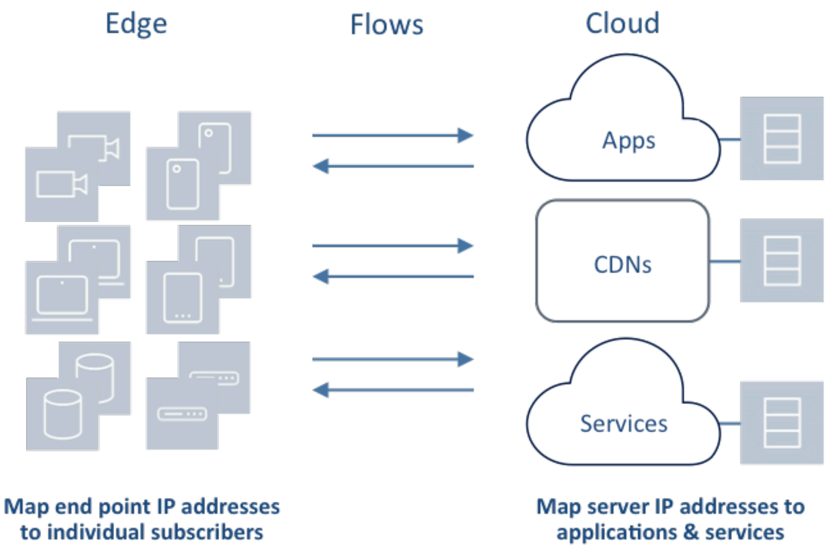
Edge
Flows
Cloud
Apps
CDNs
Services
Map end point IP addresses to individual subscribers
Map server IP addresses to applications & services

Figure 5. Visibility into Applications, Services and Subscriber End Points

The methods are nontrivial, but these mappings can be discovered using standard protocols by mining data from the variety of sources shown on the left side of Figure 4, particularly DNS servers for mapping IP addresses to domain names and also many other repositories containing information that can be used to identify servers and end points by owner, location, type and purpose.

Nokia Deepfield's Cloud Genome is an innovative Big Data solution to this complex Internet visibility problem that employs patented techniques involving software agents that continuously scan and probe Internet servers and end points while monitoring the behavior of traffic flows across the vast expanse of many networks. The solution also discovers the dynamic relationships and dependencies between cloud applications and services, CDNs, and peering and transit networks, essentially mapping the many diverse Internet supply chains that traffic flows traverse between thousands of servers in the cloud and millions of end points. Armed with this information, which is updated constantly, service providers and webscalers can immediately correlate any given traffic flow with a specific cloud application or service while identifying the subscriber end point. This capability is not only valuable for operational use cases, such service assurance and traffic engineering, but is especially critical for DDoS detection and mitigation.

INSIGHT-DRIVEN ACTIONABLE INTELLIGENCE AND NETWORK AUTOMATION

With investments in the four key building blocks in place, one can envision an endless number of real-time and non-real-time use-case scenarios where insight-driven intelligence and automation can be used to deliver a better end-user experience and protect the network.

SaaS/PaaS/IaaS Workload Migration and the Hybrid-Cloud

Enterprise workloads are shifting from traditional data centers to cloud computing platforms and from private to public clouds. By 2020, 80% of traffic will originate from cloud applications and services.[4] Enterprises are not just moving to a single public cloud but are also negotiating arrangements with multiple data center providers to obtain the best price and performance for their specific application

[4] Nokia Bell Labs Traffic Forecast 2017-2020.

needs. In addition, public cloud data center providers are expanding their footprint by opening data centers in new geographies and by partnering with Interexchange Providers (IXP), such as Equinix, EdgeConneX or Digital Realty, to deploy compute/storage pods inside IXP partner locations. All of this adds up to the fluid-like movement of virtualized workloads among private and public data centers, creating a hybrid, multi-cloud environment.

Best-of-breed networking solutions will utilize analytics to detect the shift in workload locations and reprogram the network accordingly. If workloads move to a new location, the network can be programmed to increase bandwidth to that location or take an alternative route for lower latency. The same approach can be applied to application and network resiliency in the event of a broad geographic outage. If workloads need to shift from New York to rural Iowa, the analytics engine can detect the geographic change, inform the SDN controller, which can update segment routed paths and install new classifiers and ACLs for the appropriate flows. End-to-end visibility and intelligence combined with a programmable control plane and a scalable instrumented data plane means that the network can be just as programmable and fluid as the virtual workloads it serves.

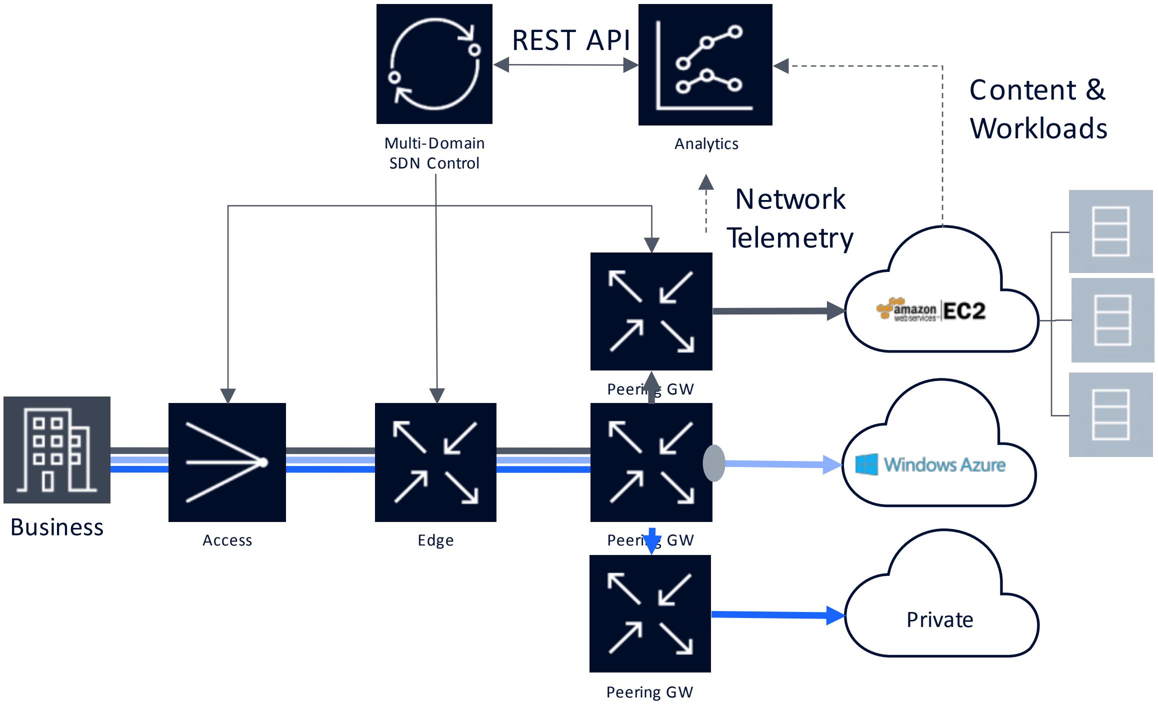

Figure 6. Cloud Workload Migration with Network Intelligence and Control

Security: DDoS Attack Detection and Mitigation

The widespread deployment of cloud-based applications and services has resulted in a dramatic expansion of the attack surface exposing service providers, webscalers and their customers to a myriad of potential security threats. At the same time, the proliferation of consumer IoT devices for home automation and security has compounded the problem by presenting hackers with a vast pool of inadequately secured end points that they can infiltrate and exploit for launching botnet-based DDoS attacks.

The escalating frequency and scale of attacks has created an arms race in DDoS defense, with vendors racing to build faster and more powerful hardware appliances for detecting attacks and scrubbing traffic flows to mitigate them. Not only is this hardware-based approach increasingly costly, but existing detection techniques that monitor for abnormal spikes in traffic flowing to a specific IP address are not foolproof and too often result in both false positives and missed negatives, leading either to legitimate traffic being blocked or worse, allowing malicious traffic to continue to flow.

Multi-dimensional Big Data analytics offers service providers and webscalers a more accurate, cost-effective and scalable software-based approach to detecting DDoS attacks in real time. Big Data streaming analytics based on flow and network telemetry can rapidly detect anomalous traffic flows directed at one or more servers. Operators can then perform real-time, multi-dimensional analytics using data collected from a range of sources, including threat feeds (see Figure 4) to identify the specific application or service being impacted and whether the traffic is from legitimate users as opposed to hijacked servers or a botnet. DDoS attack detection can also involve correlating historical flow data and known attack profiles stored in the Big Data cluster with observed traffic flows to validate a suspected attack.

Hardware-based mitigation of DDoS attacks requires backhauling all suspected flows to a scrubbing center where offending traffic is filtered and legitimate flows are allowed to pass through, but this approach is becoming increasingly costly given the scale of recent volumetric attacks. As a result, service providers and webscalers are turning to software-driven mitigation techniques that also leverage programmable data plane mechanisms in the underlying network infrastructure. These new methods rely on the Big Data visibility techniques previously described for mapping IP addresses to subscriber end points to determine if malicious or legitimate hosts are generating a spike in traffic.

One software-based approach takes advantage of the BGP Flowspec protocol in routers to configure fine-grained rules for filtering traffic flows based on specific BGP routing criteria, including source and destination prefix, effectively implementing ACL like filtering rules at the routing layer. Another approach leverages highly scalable data plane programmability emerging in the latest generation of Internet routers to configure ACLs containing up to one million IP addresses compiled from the list of IP addresses positively identified as malicious, enabling the router to filter these flows at line rate at the ingress point of the operator’s network. Highly granular data plane programmability would also enable routers to inspect packets for known patterns correlated with DDoS attacks and then apply filters based on detected patterns to automatically drop packets associated with malicious flows.

Service providers and webscalers can automate the process of DDoS attack protection by combining multi-dimensional Big Data analytics for attack detection with software-driven mitigation techniques that leverage data plane programmability in the routing infrastructure. Flow and network telemetry drive streaming analytics for detecting attacks and compiling a list of IP addresses for malicious hosts. An SDN controller can then automatically distribute this list to routers using the NETCONF protocol so that routers are configured to filter packet flows originating from these addresses. This is a prime example of streaming telemetry driving real-time intelligence for insight-driven network automation that exploits advanced programmability in the data plane.

Beyond Optimization: Proactive Anticipation

Intelligence and insight do not always need to happen in real time. Two examples are traffic engineering and network optimization. The traditional approach to optimizing the network has involved the periodic ritual (for example, quarterly) where the network engineering team pours over numerous spreadsheets and databases looking for opportunities to increase the utilization of underutilized links and network elements while also moving traffic off links that are exceeding a predefined utilization threshold (for example, 40% due to failure reroute considerations). Just as the need to periodically reallocate the holdings of a financial portfolio, networks also need to reallocate bandwidth over time as circuits are deleted or added and traffic growth occurs in a nonuniform manner.

With telemetry, instrumentation and analytics, the task of rebalancing and optimizing the network can be drastically simplified and automated. Optimization can also occur with greater frequency. Although the result is a reduction in the time spent analyzing traffic and more efficient network utilization, the biggest benefit lies in the ability to anticipate future events and demands. There is an opportunity for the network to move from being reactive to proactive. By correlating data from sources beyond traditional routing and networking equipment, it is possible for the analytics engine to identify patterns or behaviors that would go undetected or escape the purview of traffic engineers. The ability to correlate telemetry data from the network with applications such as Netflix or Amazon Prime Video or SaaS applications such as Microsoft 365 can provide insights that would never be found with traditional traffic engineering approaches. One example could be correlation of Microsoft 365 updates to a temporary increase in enterprise bandwidth utilization. The analytics engine could detect this correlation, and enterprise customer flows could be automatically and proactively adjusted to provide an improved Office 365 experience. The same approach could occur through a self-care portal if someone first identified the correlation. However, such identification would still require human intervention to increase the bandwidth for a time and reduce it later. With an analytical and programmatic approach, the entire process can be automated from identification and correlation through to network reconfiguration.

A more powerful extension of this example is the combined use of network telemetry with anonymized subscriber demographics and application usage data. By correlating network and non-network information, a more comprehensive picture can be formed among subscribers, applications and the network. The analytics engine is better able to identify patterns or behaviors that can lead to proactive adjustments to existing services or the launch of completely new services. Just looking at averages or network trend charts cannot provide this type of insight. Imagine being able to truly segment geographic and anonymized subscriber demographic information so that instead of delivering services for the average customer one delivered services tailored to classes of customers in specific geographies of the network. User QoE experience would improve as would net-promoter scores and service provider profitability.

CONCLUSION

The migration of application and services to the cloud and widespread adoption of IoT are driving network transformation. Cloud-centric networking is characterized by massive-scale, software-driven infrastructure and constantly shifting traffic flows, bandwidth demands and network topologies. As a result, service providers and webscalers face serious operational challenges related to service assurance, end user QoE and security. There is an urgent need for intelligent network automation solutions that enable operators to respond rapidly to changing network conditions with a minimum of operator intervention.

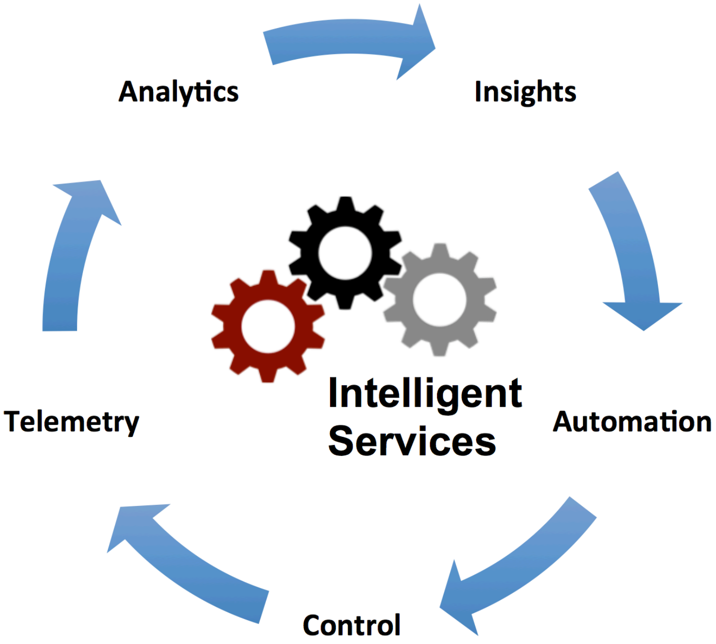

Figure 7. Insight-Driven Network Automation

Insight-driven network automation is based on four key building blocks: a simplified and fully programmable control plane; a more capable, highly instrumented data plane; unified visibility into multi-domain, multi-layer network topology; and real-time Big Data analytics. Figure 7 illustrates the closed-loop feedback model that can be implemented by combining these building blocks, enabling operators to deliver intelligent network services.

Analyst Biography: Stephen Collins is Principal Analyst at ACG Research, leading the firm's practice in network visibility and analytics. He has more than three decades of networking and telecommunications industry experience across many segments of both the enterprise and service provider markets. Stephen has worked in business and technical organizations for many leading hardware and software infrastructure vendors, serving in executive and managerial roles, including: general manager, VP of marketing, VP of product marketing, VP of business development, product line manager and software engineering manager. He has extensive experience bringing new products to market with technology-driven startups and emerging growth companies as a company founder, member of the senior management team, independent consultant and advisor to early-stage investors.

Stephen is a frequent speaker at industry conferences and has authored numerous articles for trade publications. He holds an M.S. in Computer, Information and Control engineering from the University of Michigan and a B.S. in Computer Systems Engineering, Summa Cum Laude, from the University of Massachusetts, Amherst. He currently serves as an advisor to the ECE department at UMass Dartmouth and also mentors students in technology innovation and entrepreneurship at Brown University.

Analyst Biography: Tim Doiron is principal analyst for ACG's Intelligent Networking practice which includes Packet Optical solutions, Data Center Interconnect, Transport/Multi-Layer SDN, Mobile Anyhaul and enterprise services virtualization with NFV. Doiron was most recently vice president, product management for Coriant's Data line of business. Doiron has more than 25 years of telecommunications industry experience. In addition to previous general manager, product planning and product management posts at Tellabs, Doiron has held a variety of leadership positions in software engineering, product management, business development and marketing at ARRIS, Cadant, Ericsson, and Southwestern Bell Wireless. Doiron holds a Master of Business Administration from Webster University, a Master of Science in electrical engineering from Virginia Polytechnic Institute and State University and a Bachelor of Science degree in electrical engineering from Southern Illinois University. Doiron also holds eight U.S. patents and is a member of IEEE, OSA and the Electrical and Computer Engineering Industrial Advisory Board at Southern Illinois University.

Authorship: This paper was authored by ACG Research, which is solely responsible for its contents.

Sponsorship: Nokia, November 2017.

ACG Research is an analyst and consulting company that focuses in the networking and telecom space. We offer comprehensive, high-quality, end-to-end business consulting and syndicated research services.