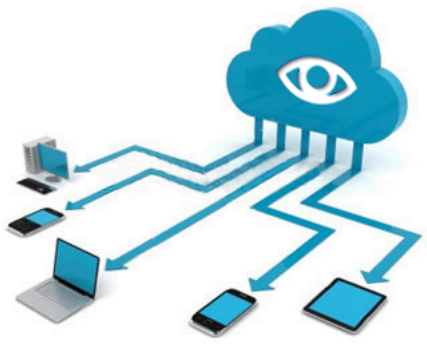
kentik

Network Traffic Intelligence at Tomorrow's Scale

AS NETWORK UTILIZATION SKYROCKETS, BIG DATA ARCHITECTURE IS ESSENTIAL FOR EFFECTIVE VISIBILITY.

Twenty-five years into the World Wide Web, networks have become indispensable. Daily life, both business and personal, now depends on the ability of providers and enterprises to manage networks effectively. But while rapid innovation has driven explosive growth in network utilization, advances in network management have not kept pace. The tools available to understand what's happening and what to do about it were built on assumptions and architectures that predate the current state of networking by a decade or more. That leaves existing infrastructure visibility systems without the scope and the capacity to operate efficiently at Web-scale or provide timely answers to critical questions. This paper introduces a solution to that problem.



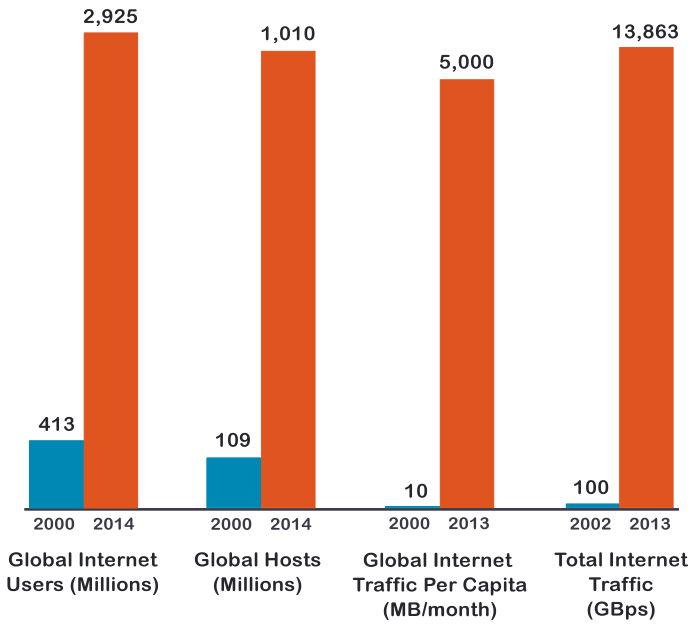
The founders of Kentik have spent decades in network management at service providers and at companies such as Akamai, Netflix, YouTube, AboveNet, and CloudFlare. They know first-hand the challenges of operating today's networks with yesterday's tools. Convinced that it's time for a fresh look at the visibility problem, they've proceeded from two core assumptions:

- The rapid increase in network utilization will continue, making scalability crucial.
- Flow is data, and lots of flow is big data.

Understanding network visibility as a big data problem, Kentik has responded by architecting a big data solution. It's a purpose-built platform, optimized for network data (NetFlow, SNMP, BGP), that is unified, open, and comprehensive. How does this new approach address the speed, scale, and usability requirements of network operators, now and into the future? We'll answer that by exploring the following topics:

- ***The growing challenge***: How skyrocketing network utilization impacts operators and enterprises.
- ***The visibility mandate***: Why effective visibility is required to meet the challenges of growth.
- ***A scalable foundation***: Scalability as the key enabler of effective visibility.
- ***State of the tools***: The inherent limitations of existing network visibility architectures.
- ***Rethinking the solution***: Kentik's big data approach to network visibility.

Fig 1: Global growth in network utilization.



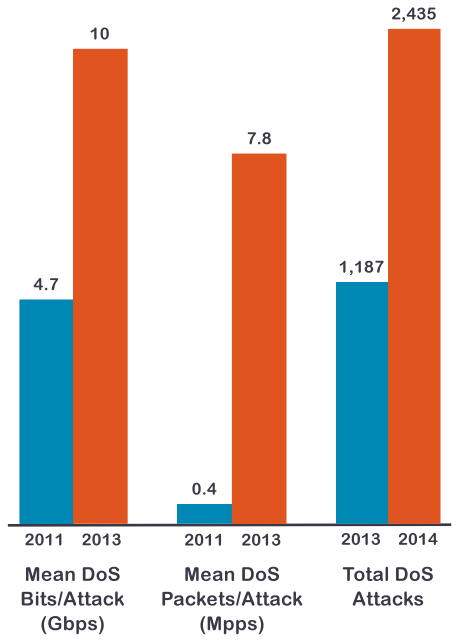
THE GROWING CHALLENGE

The cycle of innovation pushing utilization continues unabated.

Nearly two decades ago, rich media on the Web began driving up per-user bandwidth and attracting masses of new users worldwide. Network traffic and complexity have been growing robustly ever since. Accelerating demand drove the build-out of broadband infrastructure, creating a faster, more powerful Internet. And that laid the foundation for the online services, such as streaming, shopping, and banking, that are integral to today's network-centric lifestyle.

This cycle — innovation pushing bandwidth and enabling more innovation — has been paralleled in intranets and other private networks, and continues unabated to this day. The result is ever-increasing network utilization. As shown in Fig. 1, the number of global Internet users and global hosts have multiplied seven- and nine-fold respectively

Fig 2: Growth in DoS attacks.



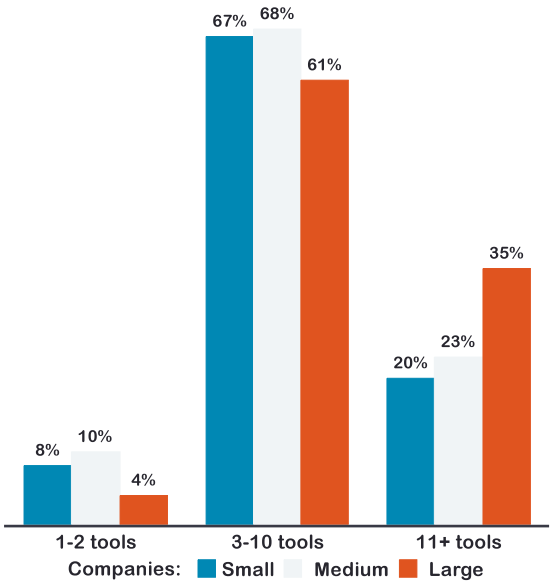
since 2000, while bandwidth-per-user reportedly rose from just 10 MB/month to 5 GB/month. Total Internet traffic, meanwhile, rose from 100 GBps in 2002 to 13,863 GBps in 2013.

As networks have grown, they've become a much bigger target for threats to both security and availability. Network operators across the spectrum — retailers, financial institutions, health insurers, media companies, etc. — live in the sights of skilled practitioners with malicious intent. As illustrated by Fig. 2, there's every reason to expect the continued proliferation of these attacks.

Threats to security and availability add to the challenges already facing network operators as they cope with rapid growth:

- More hosts, and more bandwidth demand per host, puts networks under continual pressure to scale.
- Increased use of streaming and online transactions brings increased sensitivity to performance issues. In some sectors (ad auctions, retail, etc.) slight delays can translate directly into loss of revenue.
- Traffic management hasn't been significantly upgraded since the introduction of Multi-Protocol Label Switching (MPLS) in the 1990s, and Software Defined Networking (SDN) has yet to be rolled out to the mainstream infrastructure.
- There's no single-pane solution providing fast, comprehensive insight into network status and traffic. As shown in Fig. 3, about two-thirds of companies — small, medium, and large — use between three and 10 network monitoring and troubleshooting tools. Most of the rest (21 percent of small and medium companies, 36 percent of large) use 11 or more tools, with some using more than 25.

Fig 3: Monitoring tools by company size.

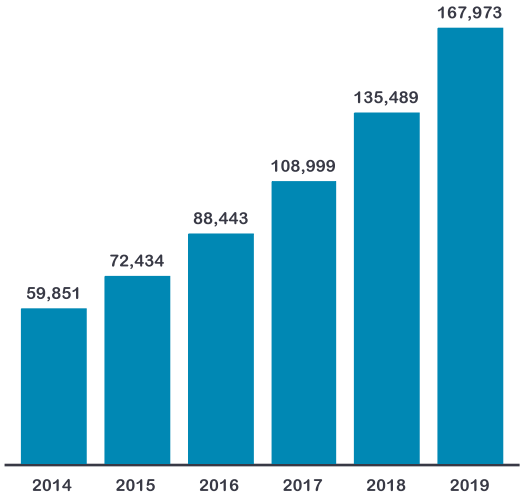


Growth typically flattens as markets and technologies mature, but current trends in the network space argue against that happening anytime soon:

- The range of connected activities on existing devices (mobile, tablet, desktop, etc.) continues to expand.
- The extension of new capabilities to mobile lets more people use more data in more places at more times.
- The up-and-coming Internet of Things (IoT) promises to extend connectivity to broad new classes of everyday objects.

The net effect is that there's no respite on the horizon for network operators. The Cisco VNI Global IP Traffic Forecast projects an increase in global IP traffic from 59 Exabytes/month in 2013 to 167 EB/month by 2019 (Fig. 4). As networks become more integral to our business and personal lives, it's increasingly critical to ensure performance, availability, and security. All of which rests on the shoulders of network architects, network engineering, network operations, and infrastructure security engineering. What they need are tools that can keep pace with the job.

Fig 4: Projected growth in global IP utilization.



THE VISIBILITY MANDATE

As network challenges intensify in coming years, visibility that's broad, deep, and fast will become ever more vital to network management. Visibility has direct impact across key areas of network operations, providing answers to the most critical questions:

- *Traffic Analytics:* Which users, applications, and services are driving network utilization?
- *DDoS detection:* Are attacks affecting our services? What's their signature? Are they happening right now?
- *Traffic engineering:* Are our links overloaded? Is re-engineering needed? Are more links required?
- *Peering analytics:* What changes can we make in our peering to increase performance and decrease costs?

The quality of visibility makes the difference between seeing and knowing.

Every network operator has visibility to some degree, but the quality of visibility makes the difference between seeing and knowing. It's one thing to be aware that there is, or has been, a problem. It's quite another to understand — in real time and/or forensically — where that problem is (internal, external, CDN, servers, stack, etc.), what's causing it, and how to fix it. In an ideal world, the same innovation that now makes possible near-instant execution of online transactions would enable near-instant insight into network performance issues, or expose precursors to brewing attacks. But so far that hasn't happened.

The problem isn't a lack of raw data but rather the abundance of it. The basic information needed to glean insights from traffic — NetFlow, sFlow, IPFIX, etc. — has long been available for harvest from routers and hosts. But even at modest sample rates, flow data adds up very quickly. The key for network visibility and analytics is to be able to collect and store that data in fine detail. That's already a big challenge, and as growth accelerates in streaming, mobile, and IoT, it's going to get bigger. That makes scalability a core necessity for any architecture deployed to enable visibility.

A SCALABLE FOUNDATION

Scalable architecture is a key enabler of effective flow-based visibility.

Scalability is what makes a flow-based data system capable of meeting key requirements:

- Elastic capacity: Keep full detail (e.g. unsummarized flow) for a long enough time-frame to analyze patterns and find causes. When capacity doesn't scale, you're left with less detail, shorter look-back, and compromised analytics.
- Fast ingest: Capture in full detail from thousands of devices simultaneously. The faster data can be ingested, the sooner that data is available for querying.
- Fast queries: The faster queries are handled, the sooner flow-derived insights can inform operational decisions (human or automated).
- *Multitenancy:* Serving unlimited clients without bottlenecks maximizes system utility and value.

Scalability must be built in from the outset. A system with inherent architectural obstacles to scaling is a system that will struggle to ingest data from multiple sources while simultaneously processing queries from multiple users. That said, scalability is only one of several essential traits of an effective network visibility solution. Additional requirements include:

- *Unified view:* A single platform to collect and store different types of data (flow, SNMP, BGP, etc.) lets you see and query all relevant information in a single environment. That's much more efficient than constantly checking multiple places to see what is or was going on. And it allows you to correlate between data types to reveal patterns you might not otherwise find.
- *Open access:* Industry standard interfaces facilitate integration with systems that perform complementary functions (e.g. business analytics or DDoS mitigation). They also maximize the value of collected data by allowing it to be analyzed for multiple purposes.
- *Cost-effective:* The net cost of deploying a solution must make it affordable to cover all traffic, which is the only way to fully understand and protect the network. A system that's cost-effective removes the financial incentive to accept blind spots, and it also increases operational margins.

How do existing tools measure up? Not well.

STATE OF THE TOOLS

Armed with the requirements of effective network visibility, let's look at how existing systems measure up. The short answer? Not well. Current offerings fall mostly into two broad categories: open source software and appliance-based enterprise systems. Neither is architected to enable a comprehensive solution at scale.

Open source software has contributed immensely to innovation in high-tech. But it's not typically oriented toward providing comprehensive solutions that are tested, supported, and ready to manage critical operations at scale with maximum uptime. The implementation of network visibility through open source alone runs up against significant obstacles in several areas:

- *Scaling:* Open source programs typically run on just one computer. They don't cluster, so they can't keep up with the massive flow volume generated by a network of any size.
- *Resources:* It takes significant time and/or money to adapt a generic open source program to handle the spectrum of visibility use cases. With no vendor support, tools groups spend much of their time developing a platform rather than enabling needed functionality.
- *Fragmentation:* Patching together a collection of individual small-scale tools that each target different types of data (flow, SNMP, BGP, etc.) does not yield a unified network view for monitoring or an integrated database capable of cross-type correlation.
- *User interface:* The UIs of open source tools are typically only partly realized, and their command-line interfaces aren't standardized. That makes them slower to use, delaying resolution of time-critical issues.

Tools without testing and support aren't ready for critical operations.

Enterprise appliance systems are architected with inherent limitations.

At the opposite end of the spectrum are appliance-based enterprise systems, which have their own set of inherent limitations:

- *Capacity:* Appliance architectures scale data capture and analysis capacity only in finite increments. When one appliance is full you have to add another appliance, at which point you don't have load balancing or full data sharing across your entire datastore.
- *Granularity:* Most systems can't index or search the full detail of every flow. Instead they render summaries (graphs, reports) and then discard the original data. So you have to choose in advance which aspects of flow to examine in full detail, and hope that you can accurately foresee every question you might later need to answer.
- *Integration:* Existing appliance-based offerings are closed, single-purpose systems that can't easily integrate with complementary solutions for functions such as business analytics and DDoS mitigation.
- *Deployment:* Appliance-based systems involve a cumbersome provisioning process. The lag between ordering and using is generally weeks or months. If a system is based on packet inspection it can only be installed in a maintenance window.
- *Cost:* Measured by the devices they can handle or the data they can process and store, appliance-based offerings are prohibitively expensive. As a result, they are typically relegated to a subset of possible viewpoints, which compromises visibility.

RETHINKING THE SOLUTION

Kentik has rearchitected visibility for today's networks.

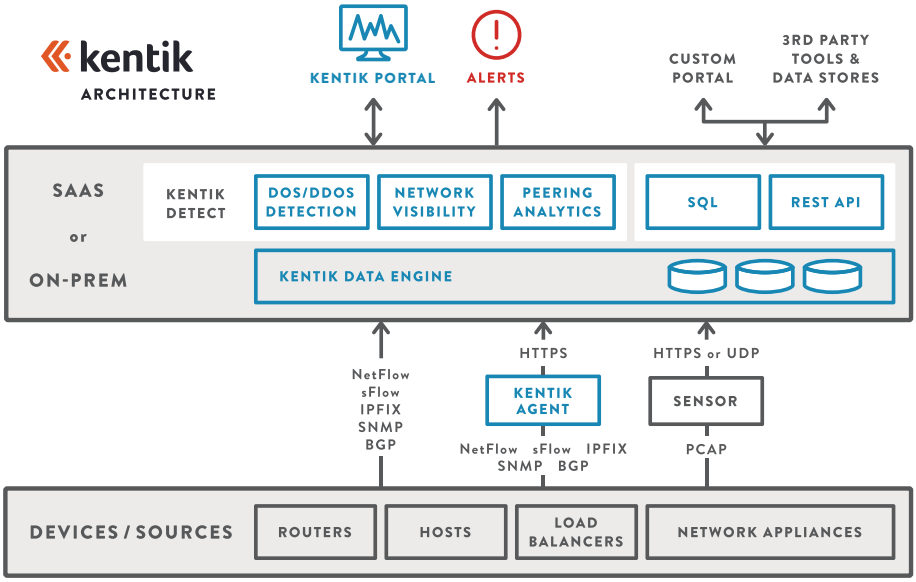
As we've seen, the inherent limitations of both OSS and appliance-based systems result in direct impediments to effective network visibility. It's possible to design around architectural constraints, but not without making trade-offs. What Kentik has done is the opposite: design an architecture that is expressly optimized for the requirements of the job at hand. For network visibility that means building a big data platform that enables users to:

- Detect, understand, and respond to current conditions (attacks, routing bottlenecks, congestion, etc.).
- Identify historical patterns and respond with defense and/or optimization.
- Improve operational efficiency and reduce costs by minimizing slow, labor-intensive tasks.
- Maintain compliance with contractual QoS guarantees and end-user performance expectations.
- Optimize revenue by applying business analytics in areas such as service tiers and pricing.

In the absence of an existing big data solution that can handle flow at Web-scale, Kentik's founders purpose built a new data platform. Harnessing technology that wasn't even on the horizon a decade ago, they've created a scalable, open, unified system called Kentik Detect, which is made up of a custom big data backend and an intuitive UI (portal).

The heart of Kentik Detect is the Kentik Data Engine, a clustered high-availability (HA) datastore. The datastore is engineered to provide:

Fig 5: Kentik Detect system architecture.



- Continual real-time ingest without database inserts, enabling sub-minute data availability for monitoring and queries.
- Unconstrained capacity via cloud-modeled architecture.
- Multitenancy to allow multi-user availability, with a secure, exclusive environment for every user.
- Low-latency querying using subqueries, caching, and rate-limited response to return near-instant answers.

- Open data access via SQL queries using the UI, REST APIs, or any PostgreSQL client.
- Data-source agnosticity, allowing querying across data types and seamless expansion of platform capabilities.

The Kentik Detect portal, meanwhile, provides a single-pane environment for multiple views of collected data:

- Configurable real-time dashboards display monitoring of key metrics.
- Query-based visualization tools graph flow at full granularity (unsummarized), allowing accurate drill-down forensics over multi-month time frames.
- User-configurable, query-based alerting monitors network status.
- Fast onboarding eliminates both the barriers to trial and the wait for a fully functional system.

In sum, Kentik Detect is the solution that finally brings network visibility into the modern age. Drawing on the capabilities that have made modern networks so powerful, compelling, and popular, it provides a common platform that can unify all network data and make it available for all use cases.

Kentik Detect is the solution that finally brings visibility into the modern age.

CONCLUSION

As networks grow and traffic skyrockets, effective network management is essential to the smooth functioning of business and personal life across much of the globe. Network operators need fast, clear visibility to detect attacks and performance issues, and they need flexible access to detailed long-term data for forensics and analytics. There's currently a disconnect between these requirements and the capabilities of available tools. Architected with inherent limitations, existing systems can't keep pace with the steep growth of network utilization. What's needed instead is a scalable, open, multi-tenant solution, built on the realization that network visibility is fundamentally a big data challenge. Kentik Detect is that solution.

To learn more about Kentik Detect and how it can address your network visibility needs, please contact sales@kentik.com.

SOURCES:
Fig 1a. ITU and United Nations, from Internet Live Stats, http://www.internetlivestats.com/internet-users/#trend
Fig 1b. Domain Survey, Internet Systems Consortium (ISC), https://www.isc.org/services/survey/
Fig 1c, d. Cisco VNI Global IP Traffic Forecast, 2013–2018,
http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html
Fig 2: Verizon 2014 Data Breach Investigations Report (DBiR),
http://www.verizonenterprise.com/DBIR/2014/reports/rp_Verizon-DBIR-2014_en_xg.pdf
Fig 3: Enterprise Management Associates, 2014
Fig 4: Cisco VNI Global IP Traffic Forecast, 2013–2018

KWP1-150721