Data-Driven Network Operations

France-IX 2016

Avi Freedman

kentik

Summary

kentik

- Why Data-Driven Network Operations?
- The food: data types and sources
- Requirements and tool types
- Data Fusion
- Business-driven use cases: Review
- Use case: Traffic exploration and anomaly detection
- Use case: Planning and peering
- Use case: Network performance analytics
- Use case: Security and breach detection
- Use case: Revenue enhancement
- Summary

Why
Data Driven
Network Operations?

What is Data-Driven Network Operations?

kentik

Using your data to drive your ops + business!

Most content companies and enterprises are data and analytics driven.

Devops is as well (APM, metrics-at-scale).

But the network world has some catchup to do.

We’re now at the point where CPU, systems and methods exist for NetOps.

But most of the work remains to be done.

The Food:
Data Types and Sources

What kind of data?

kentik

- *flow
- BGP
- SNMP and streaming telemetry (i.e. queue depth)
- Event logs
- User tags (“Application A”, “Peer”, “My Network”,Rack#1”)
- Server Metrics (RRD, tsdb, Grafana, …)
- Active measurement data (probes)
- LLDP and IGP-based topology data
- SDN overlay/underlay mappings
- DNS queries
- Threat intelligence

Enhanced NetFlow: Performance and Semantics

kentik

Imagine if we could get performance data from the network:

- Queue Depth
- Retransmits per flow
- TCP latency
- Application Latency

You can (somewhat):

- Host software (Nprobe)
- Sensors / Taps
- Webserver logs (Nginx)
- Cisco AVC supported routers

Requirements
and
Tool Types

Key Network Operator Requirements

kentik

Key requirements for modern Data-Driven Network Operations:

- No data aggregation or pre-filtering.
- Correlation (fusing) between data types.
- Full resolution searchable and stored for months.
- FAST. Less than 10s for results. Cannot wait minutes while spelunking.
- Network-savvy Uis and APIs (understanding routing and prefixes).
- Detect anomalies. Should not have to watch graphs manually.

Anomaly Detection

kentik

You shouldn't have to stare at dashboards or watch logs to detect badness.

Should be able to look at dimensions such as interface+path+type, or cidr/prefix, detect unusual patterns and perform an action. Continually.

Be able to look at things beyond pps/bps such as retransmits, latency, log.s

Be able to detect shifts such as did an asn or IP on a particular interface suddenly move from top-x #200 to #2 and that is unusual for this time of day.

This is doable today, though requires network-savvy data backends and is not yet "magic".

Available Tool Types for Traffic-Centric Analytics

kentik

- **Current Open Source:** pmacct, ntop, SiLK
- **Older Open Source:** cflowd, MrTG, AS-PATH, swatch
- **Commercial software:** Arbor, SevOne, Solarwinds, ManageEngine
- **Do It Yourself Big Data:** Kafka + {ELK, Hadoop, Druid, Cassandra, Spark}
- **SaaS Big Data:** Kentik, Appneta +(Splunk [for security])
- **On-Prem Big Data**: Cisco Tetration, Kentik, Deepfield…

- We're making progress as most many "big data" tools support high volumes of data, a large set support high speed queries, and a (not identical) large set support non-aggregated storage and high volumes of data.
- Most don't yet support data fusion and "network savviness" about prefixes and paths yet, so DIY efforts must implement.

Data Fusion

Fusing Data

kentik

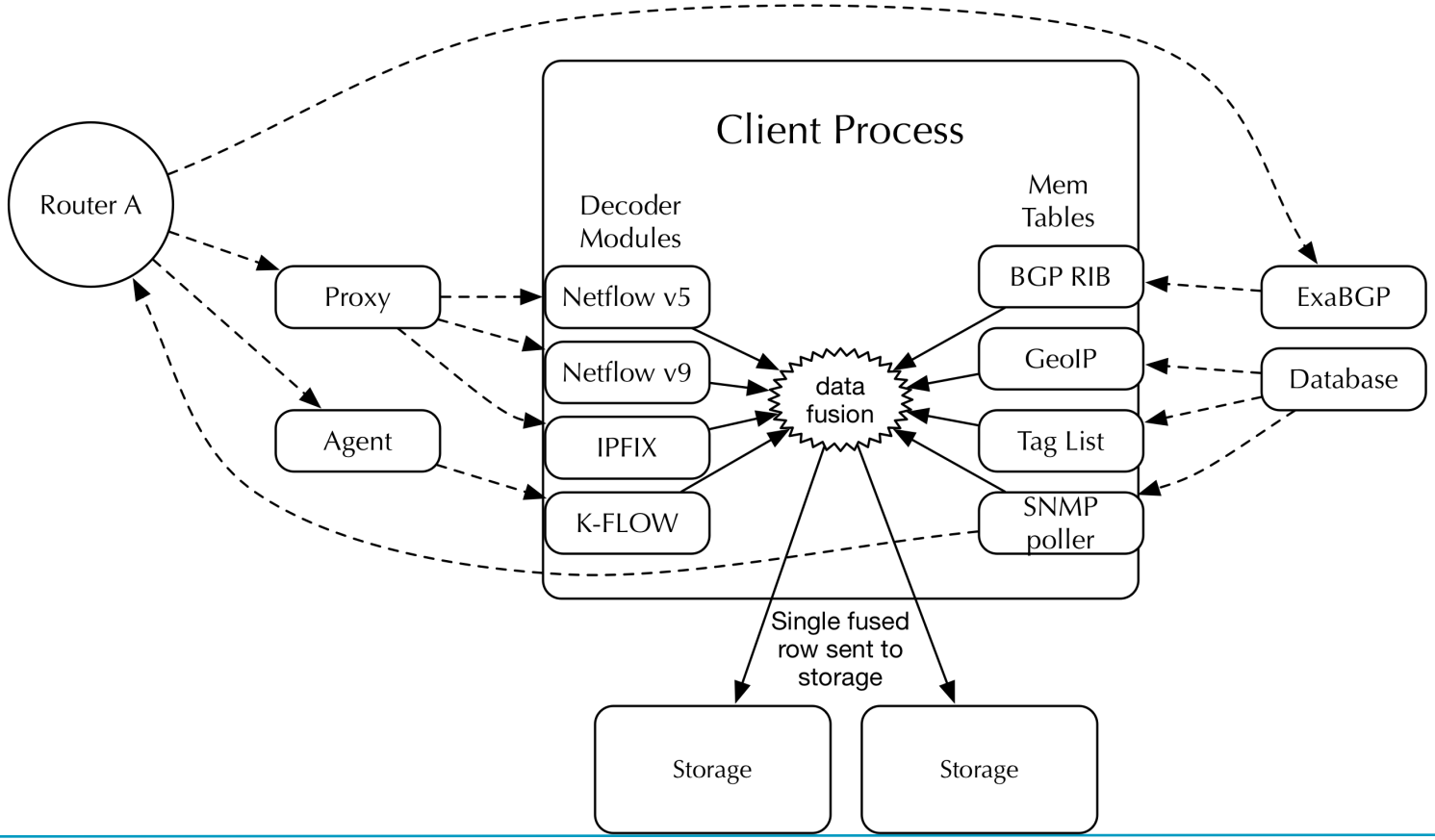
Flow or BGP or SNMP or DNS or logs alone are not enough. Starting with NetFlow/sFlow/IPFIX alone, which is (IP src/dst, ports, bytes, packets, and sometimes VLAN, MPLS, …).

This becomes much richer when combined with:

- BGP attributes
- Geography
- Tags (rack, department, customer…)
- Config changes and software versions
- Threat intelligence and known-bad IPs

- Fusing should be near real-time, performed at ingest and data specific

Distributed system data ingest



Pulling it together

kentik

Legacy non-distributed network tools:

- Don’t scale to handle large amounts of data (space/IO/CPU limited).
- Typically store only aggregates or pre-filtered data.
- Very limited fusing, only 1 or 2 data types per tool.
- Limited dimensions and filtering depth, often slow.
- Often require large amounts of configuration + tuning.

Big-Data approach:

- Large scale, billions of records per hour, no aggregation, complex fusing.
- Distributed micro-service architecture. Can scale very “wide”. ++Hardware.
- If done right, fast. Real-time ingest and < 5s queries.
- Big challenge is adding fusion, network savviness, and of course hardware.

Business-Driven Use Cases

Data-Driven Ops + Business Use Cases

kentik

- Network Planning
- Peering Analytics and Abuse
- Congestion detection
- Is it the network?
- Where on the network?
- Proactive alerting
- Distributed DDoS Detection
- What Changed Post Deploy?
- Security and Breach Detection
- Cost Analytics
- Revenue Identification (New + Risk)
- Enabling Internal Groups

Use Case:
Traffic Exloration
and
Anomaly Detection

Use case: Traffic debugging and inspection

kentik

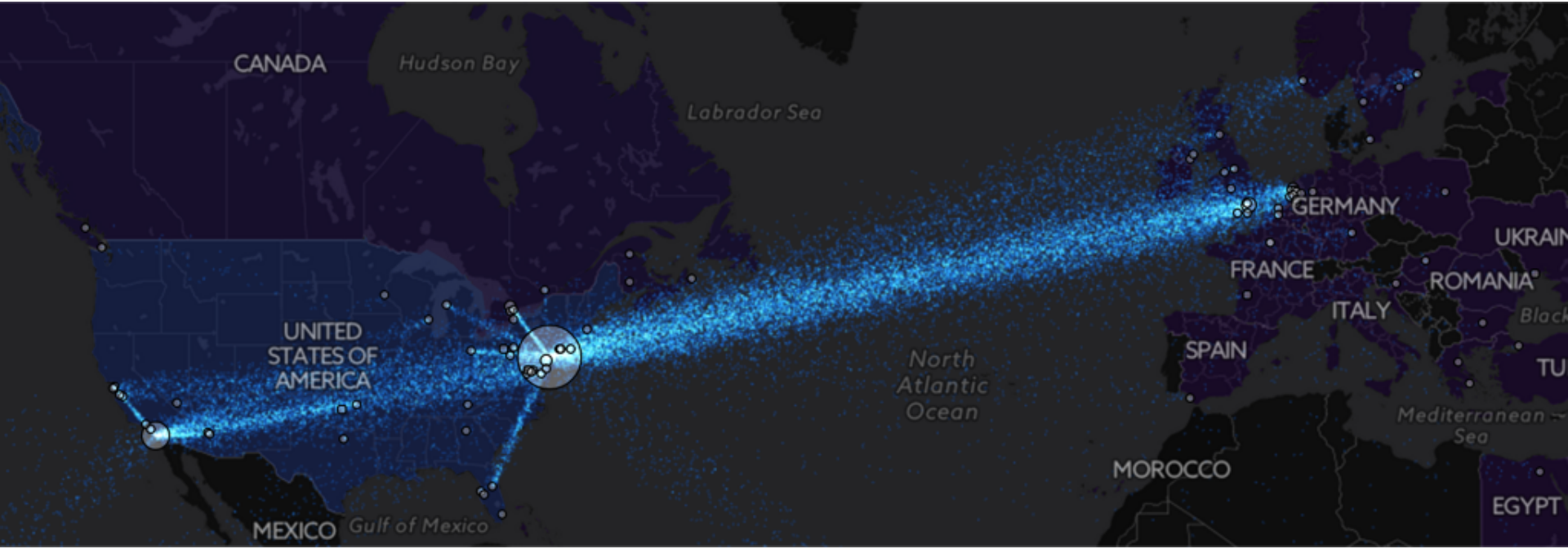
Why did the interface just double its traffic, now saturated?

Customer renewal re-pricing. Where is the traffic leaving my network?

Is a peer sending me traffic they shouldn't be?

Did a content provider shift their traffic path to me?

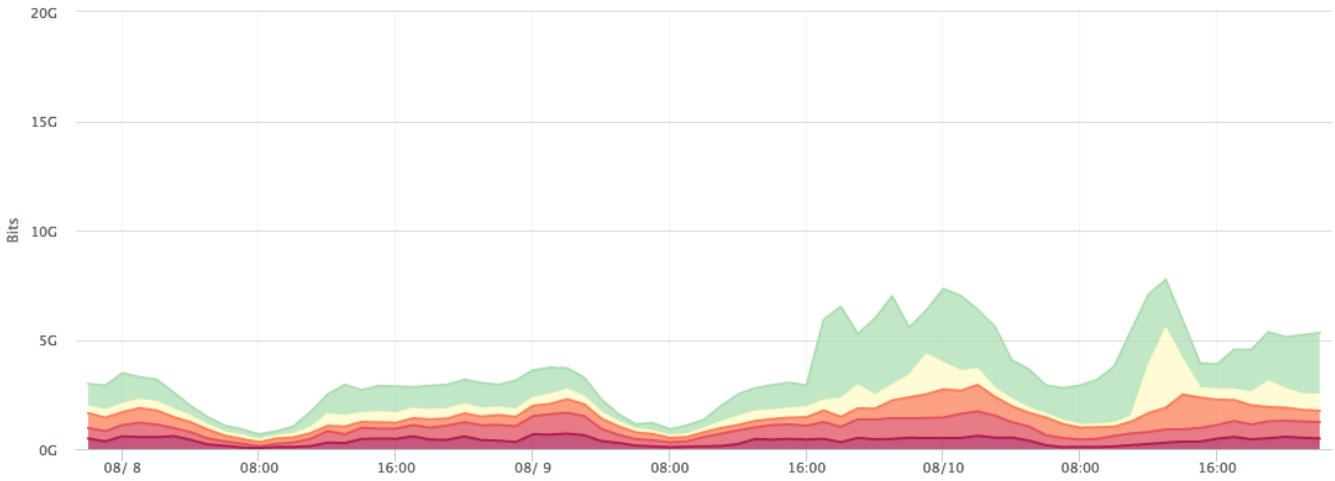
Traffic by geo, filtered by one source interface



Traffic debugging and inspection

kentik

Top Source AS Number by Max Bits



Left +Y Axis

key	Avg Mb/sec	95th Percentile	Max Mb/sec	Last Datapoint
Total	*10,930.56*	*16,106.00*	*18,662.48*	*14,177.33*
MICROSOFT-CORP-MSN-AS-BLOCK - Microsoft Corporation,US (8068)	1,544.89 (14.1%)	3,603.76	4,211.69	2,843.10
HIGHWINDS3 - Highwinds Network Group, Inc.,US (20446)	*2,030.78 (18.6%)*	*3,243.14*	*3,778.61*	*1,616.54*
MICROSOFT-CORP-MSN-AS-BLOCK - Microsoft Corporation,US (8075)	519.44 (4.8%)	1,381.69	3,611.92	723.76

Use case: Anomaly detection

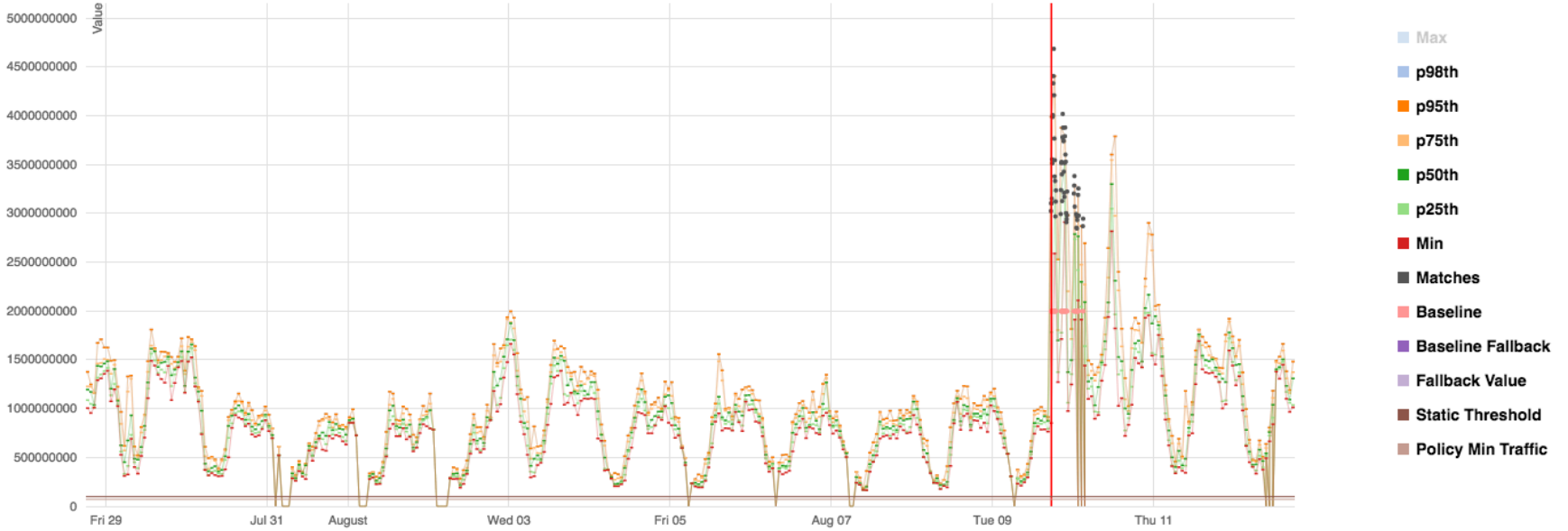
kentik

Traffic from individual top-20 ASN over transit unusually high. Operator notified at red line.

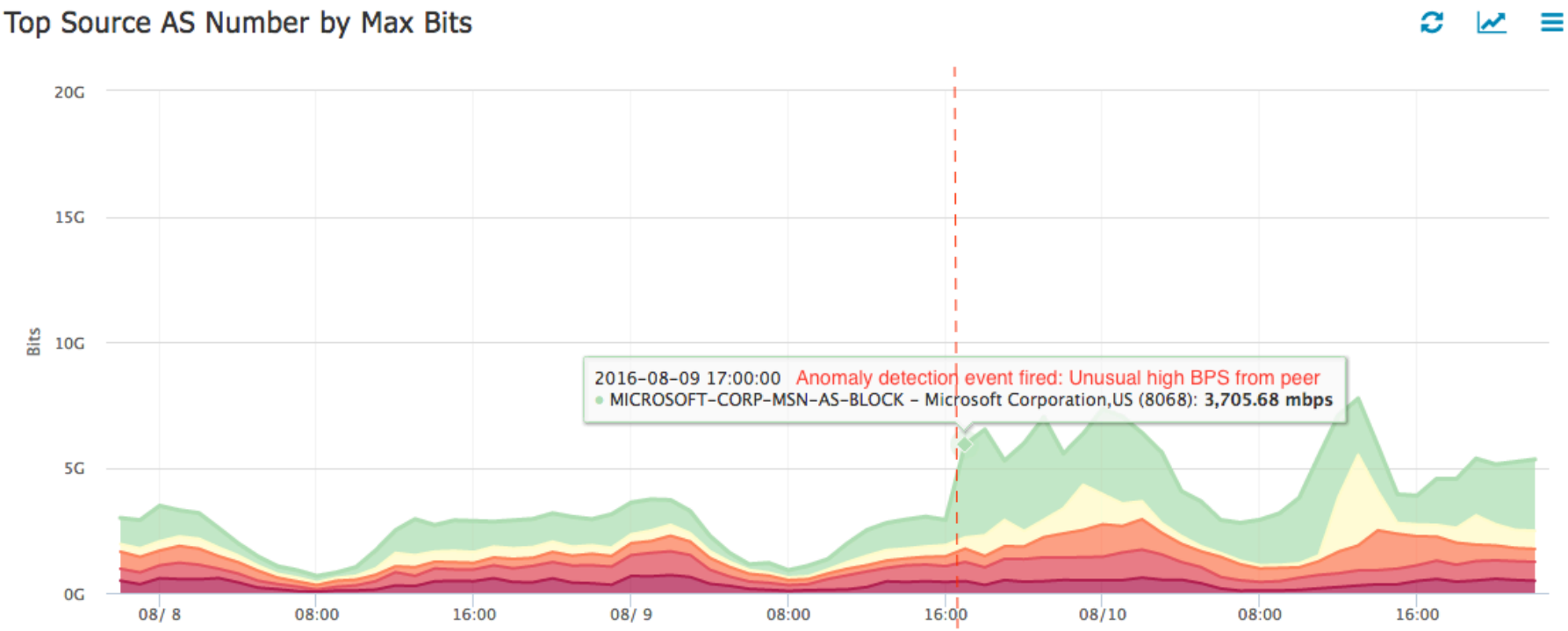
Unusual High BPS traffic from ASN per site (210)

	AS_src: MICROSOFT-CORP-MSN-A... i_device_site_name: Chicago, IL	major Matches: 5	757.22 Mbps	488.28 Mbps CALCULA...
Unusual High BPS tra... CLEAR -> ALARM	AS_src: MICROSOFT-CORP-MSN-A... i_device_site_name: Ashburn, VA	major Matches: 17	2,960.19 Mbps	1,905.01 Mbps CALCULA...

15 Days Graph

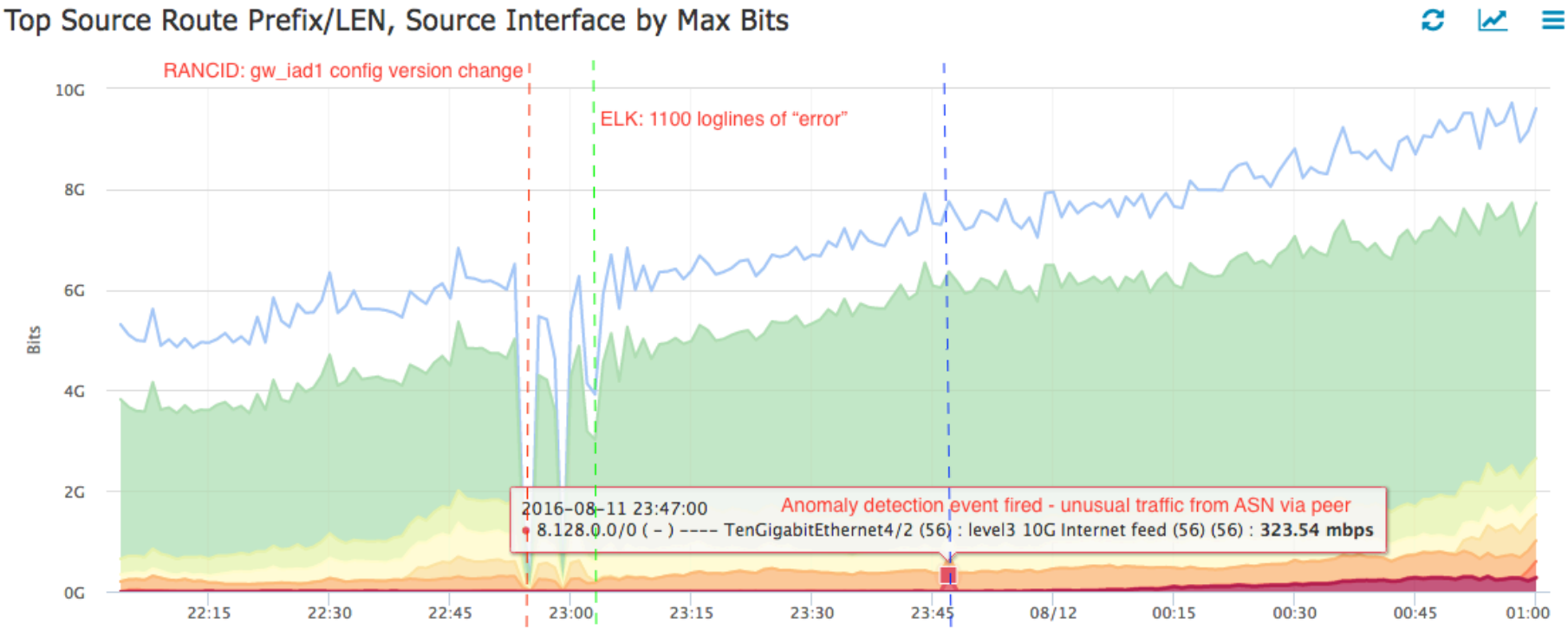


Traffic, anomaly detection and annotation



Traffic annotated with multiple events

kentik



Anomaly detection: DDoS detection and characteristics

UDP_BADPORTS **CLEAR -> ALARM**	IP_dst: **70.44.101.7**	major Matches: 9	**2,445.91 Mbps** 8,370 unique_src_ip	381.47 Mbps DEFAULT...
UDP_HIGHBPS **CLEAR -> ALARM**	IP_dst: **70.44.101.7**	major2 Matches: 9	**3,914.95 Mbps** 10,340 unique_src_ip	95.37 Mbps DEFAULT...

Use Case:
Planning and Peering

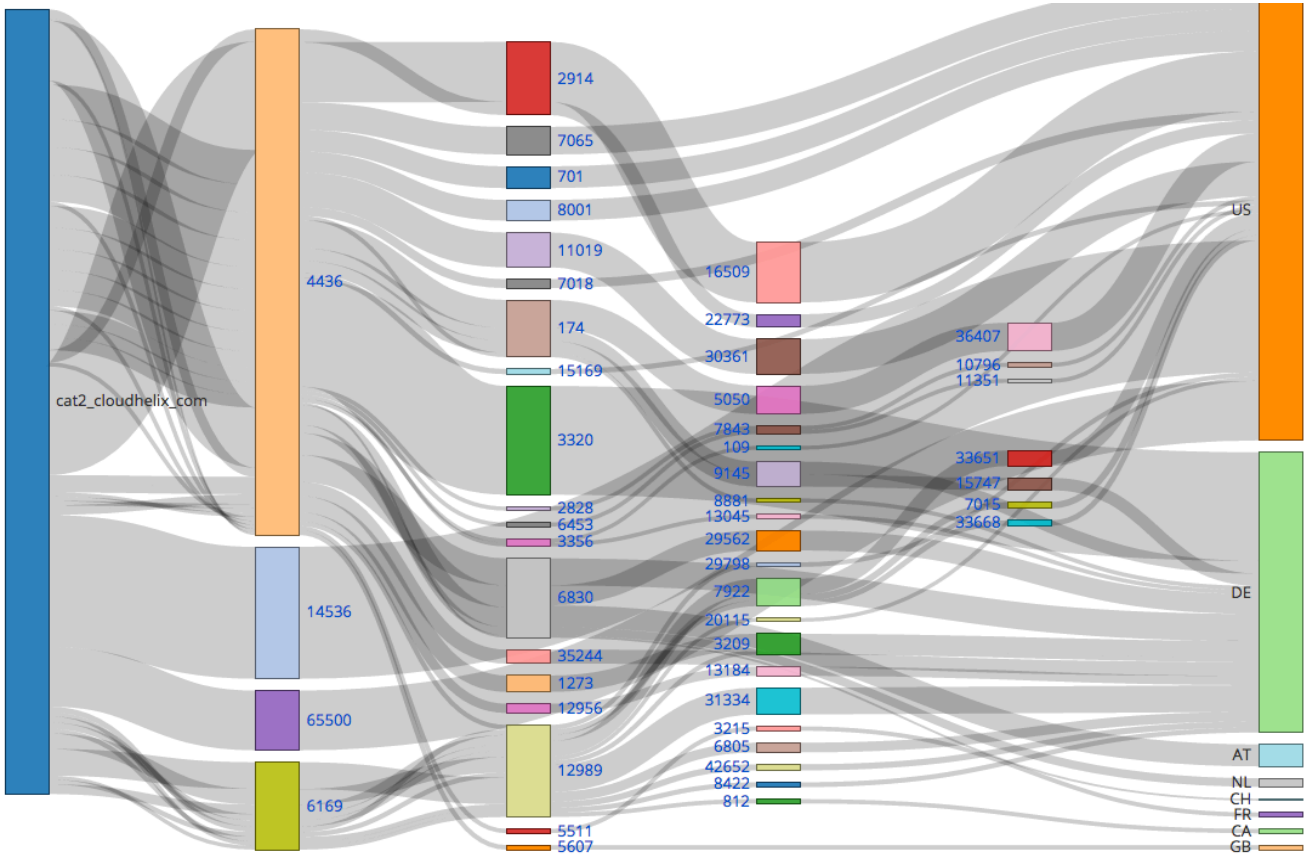
Use case: Network Planning

kentik

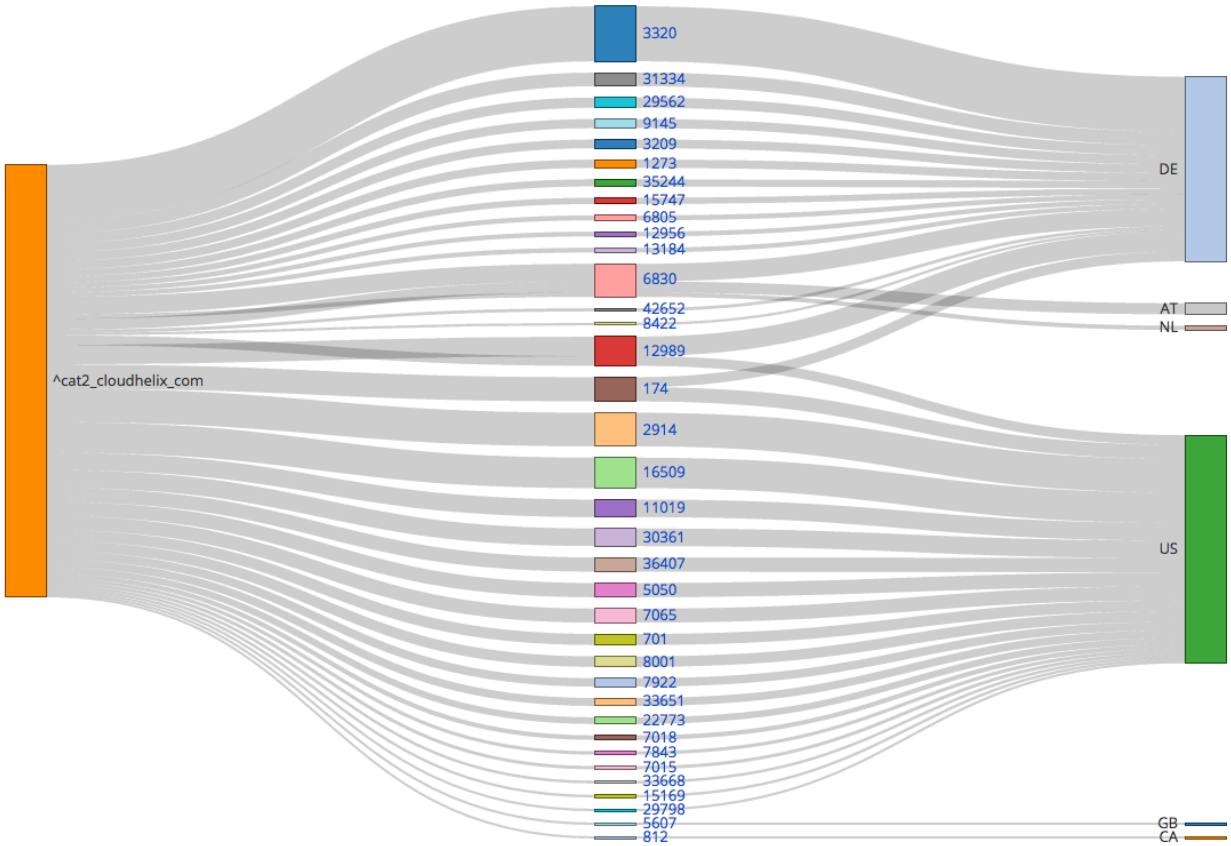
Flow-based traffic + BGP can be used to help show:

- Path, neighbor, transit, origin, and country of traffic.
- Strategic peering and transit changes that can improve performance and reduce costs.
- Potential new peers and locations to peer.
- Evaluate the potential of new peering exchanges or facilities.
- Transit relationships that are of high or little value.
- Understand ROI before extending backbone links or capacity.

Network Planning, traffic by BGP HOP



Network Planning, collapsed path, exclude 1st



Use Case:
Network Performance Analytics

Use case: Network Performance Analytics

kentik

Flow-based traffic + BGP + network performance data can show:

- Whether issues are in the application or network layer
- If network, where?
- And in a way exposable to internal development + app operations
- And to pinpoint where…
- Performance issues by peer, or even remote AS hops behind peers
- Or prefix
- Or DC
- Or not (in the network)

Retransmits enhanced flow example

kentik

Apply Changes

Query — Dataseries: Auto

Group By Dimensions [Clear All] — CIDR

Destination : IP/CIDR — 32

Metric — PPS Threshold

% Retransmits (Beta) — 10

Visualization depth: Normal

Display and Sort By

98th Percentile

Time — Range: Custom — Zone: UTC

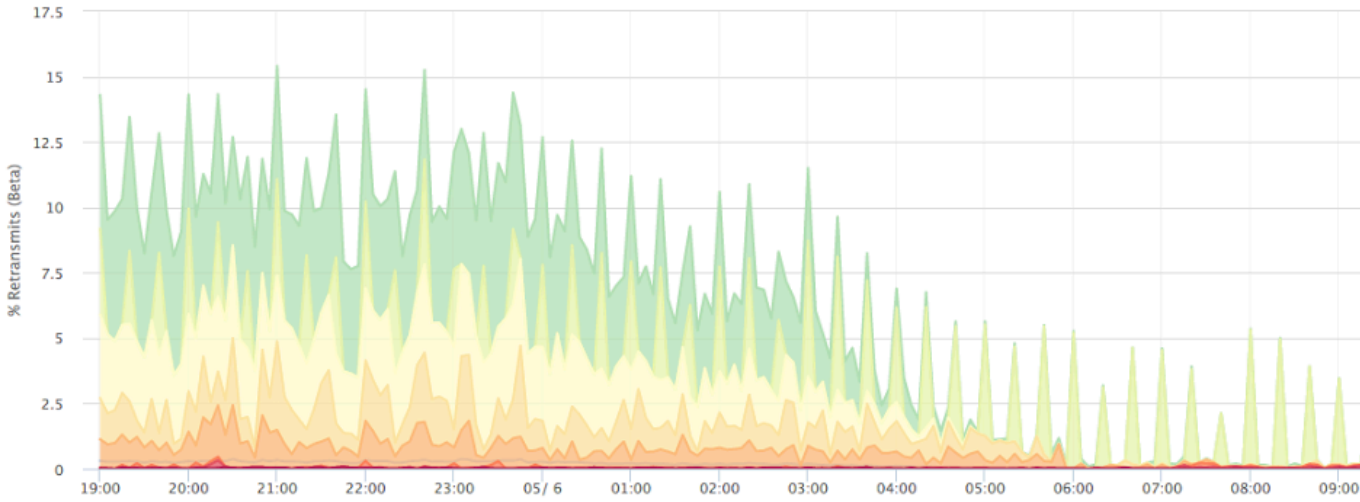
From 2016-05-05 19:00

To 2016-05-06 14:28

Filters

Devices (1)

Top Dest IP/CIDR by 98th Percentile % Retransmits (Beta)



	% Retransmits				Retransmits/s			
ipv4_dst_addr	Avg	p98th	Max	Last Datapoint	Avg/sec	p98th	Max/sec	Last Datapoint
Total	0.139	0.341	0.376	0.030	8.490	24.910	27.410	2.137
Historical Total: 7 days back	0.002	0.006	0.008	0.002	0.015	0.110	0.150	0.043
10.1.1.88/32 (s17.iad1)	2.105	5.212	6.249	0.166	3.072 (36.2%)	9.053	10.647	0.243
10.1.1.15/32 (www1.iad1)	3.824	5.059	5.512	5.037	0.188 (2.2%)	0.992	1.083	0.993

Use Case:
Security and
Breach Detection

Use case: Network Security Analytics

kentik

Flow-based traffic +/- threat intelligence can show:

- Potentially compromised servers, desktops, or IoT devices.
- Anomaly detection can often uncover these even without threat intel.
- Ideally, this can be used to help downstream or internal customers.
- And feed DDoS response if there are local sources/sinks.
- To be a good citizen, BCP38 violation detection can run across the network, even with simple heuristics like “how many /8s as sources”?

Use Case:
Revenue Enhancement

Use case: Revenue Enhancement

kentik

Sometimes not discussed in polite company, but great traffic-based analytics can help with the top line as well as margin:

- Offering high-margin customers lower rates to attract more traffic.
- Identifying large 2nd and 3rd hop AS sinks or sources behind peers, to convert to customer relationships.

Summary

Summary

kentik

Networks can produce large amounts of data that will make your life easier.

Big Data platforms are able to consume this data.

Specific tools for Network Operators are beginning to appear (free and paid).

Paid tools are more specific to network use (UI, easy setup, etc).
Free tools have the “power” but require cobbling together pieces.
Much work to be done re fusing data such as logs, changes, alerts, DNS.

SaaS providers will provide community views and enable data-sharing.

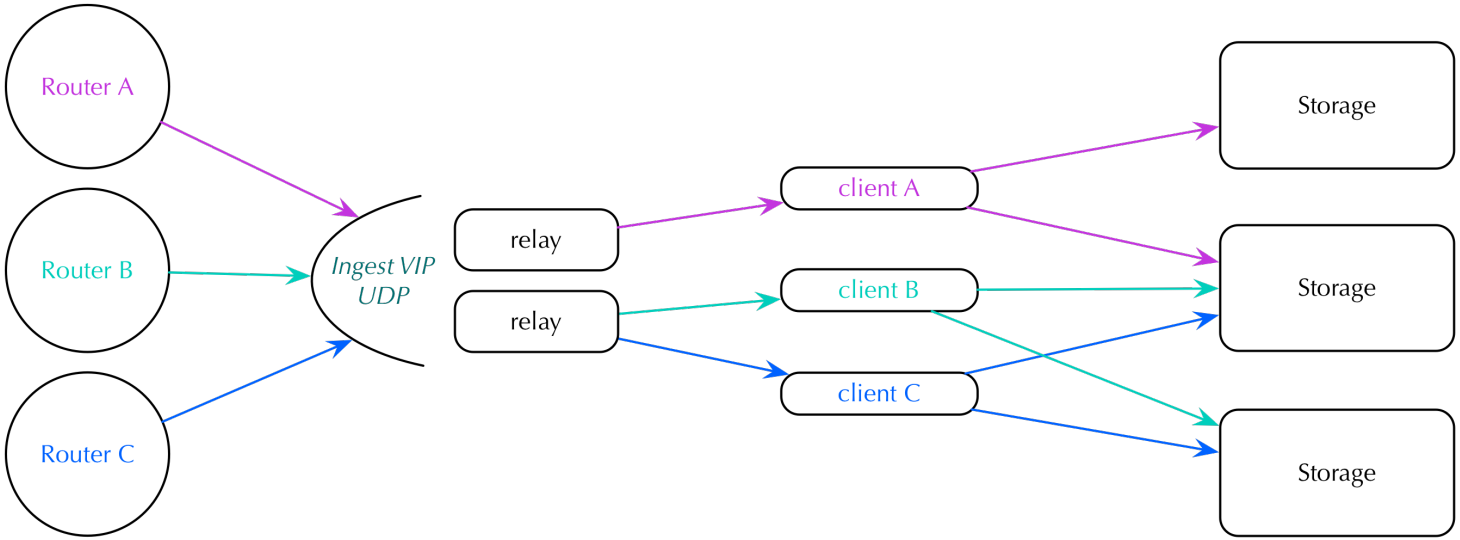
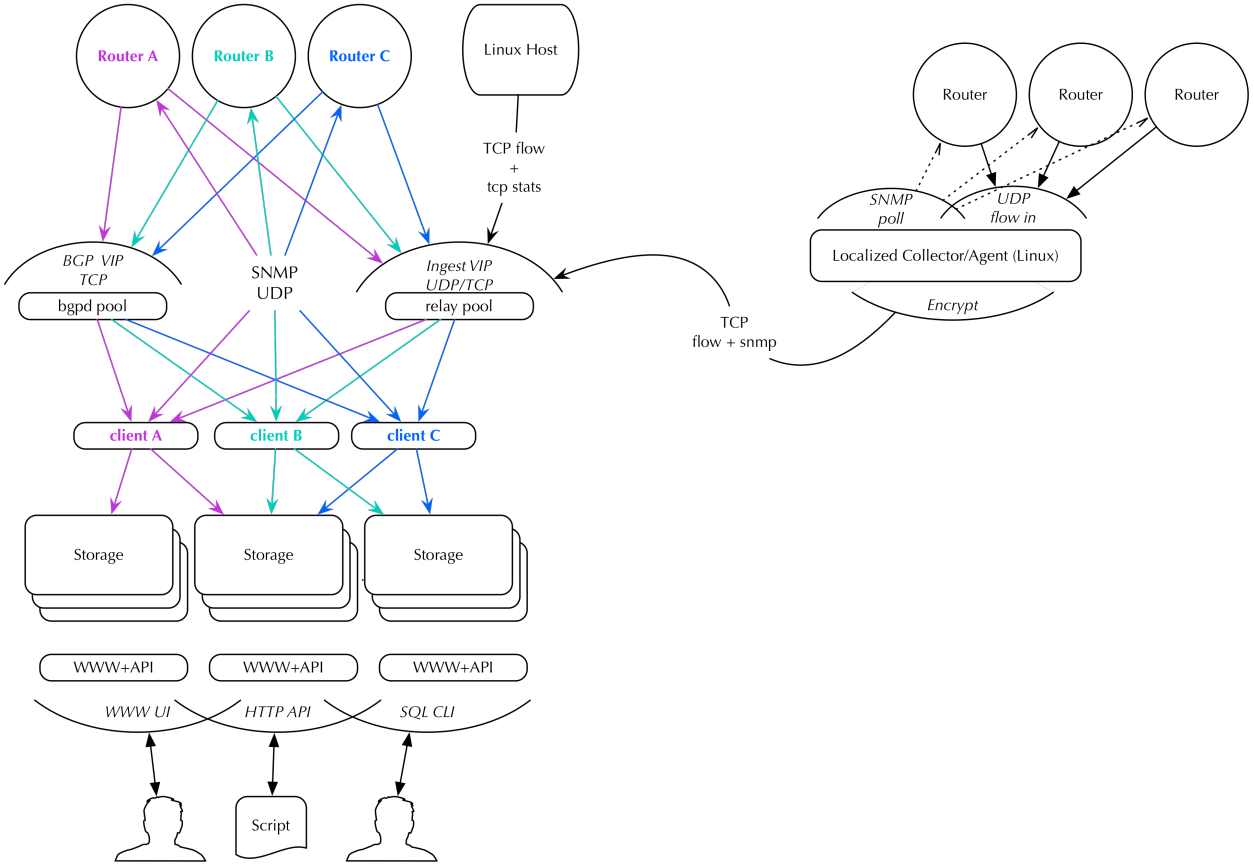
Multi-source correlated at-scale data is heavily in use for non-network systems.
It’s time vendors and network folks catch up and end the NetOps suffering.

Questions?

Dan Ellis
dan@kentik.com

Appendix: Architecture

Distributed system data ingest



Distributed system fusing