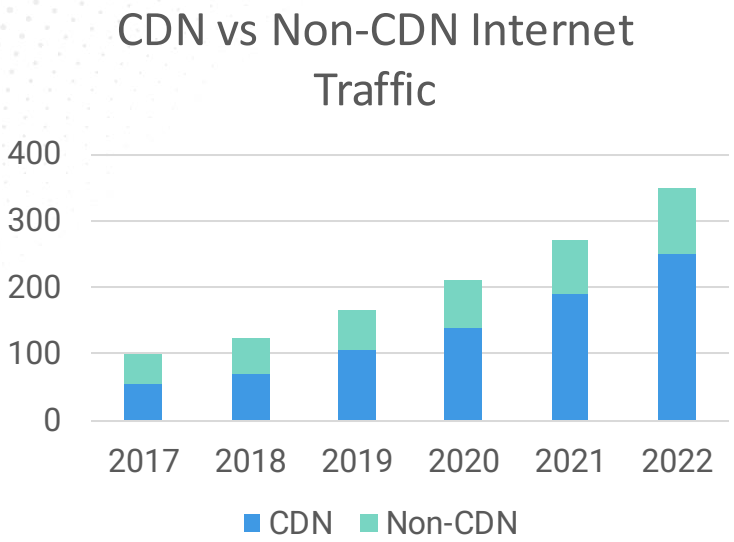
Enrich NetFlow by Metadata

Jacob Chiang
CTO, Genie Networks

Application And Platform Visibility



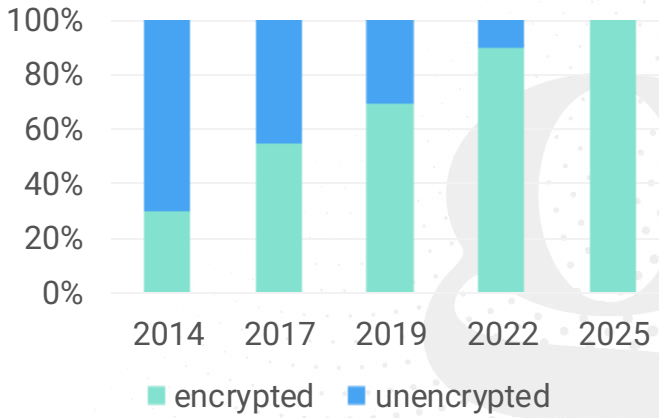
- CDN Planning
  - Which CDNs to be embedded or peered through which IXP? What applications can they deliver?
- CDN Monitoring
  - Do the embedded or peered CDNs deliver through agreed servers and link to agreed targets? How much bandwidth is delivered?
- Application Monitoring
  - How are critical applications delivered? Are there any anomalies?

Challenges of DPI

- ISPs have used DPI technology for more than a decade: lawful intercept, policy enforcement, quality of service, tiered services, copyright enforcement, statistics, etc.

- Almost all agree that at least 80% of the worldwide internet traffic is encrypted, and Google reports 95% encryption over all its services.

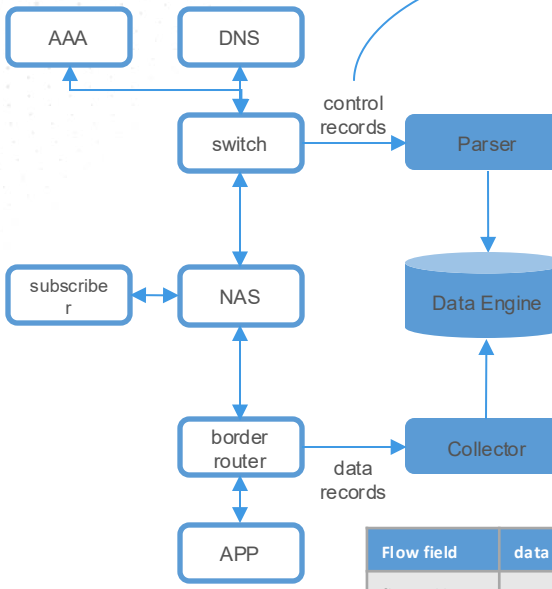
- More and more passive DPI solutions rely on DNS and other technologies coupled with AI/ML to infer applications.



Using NetFlow As Data Source

- If we can only infer the application of encrypted traffic, NetFlow is a more cost-effective data source.
- NetFlow is standardized and supported by all major router vendors. NetFlow can be collected from border, core, and access routers.
- NetFlow can sample packets to reduce data volume and thus reducing costs.
- Many ISPs are already collecting NetFlow enriched by BGP for route optimization, DDoS detection, and other applications.
- But NetFlow only collects L3/4 header fields. The requirement is to enrich it with metadata learned from the control plane and/or other sources.

Enrich NetFlow By Metadata



DNS name	type	data
download.skype.com	CNAME	download.skype.com.edgekey.net
download.skype.com.edgekey.net	CNAME	e4707.dspg.akamaiedge.net
e4707.dspg.akamaiedge.net	A	184.51.240.147

GEO/IRR/OTT

...	
prefix	city
184.51.240.0/24	Hong Kong

enriched flow records

AAA attribute	value
Framed-IP-Address	10.20.89.215
3GPP-IMSI	284011234567890
...	...

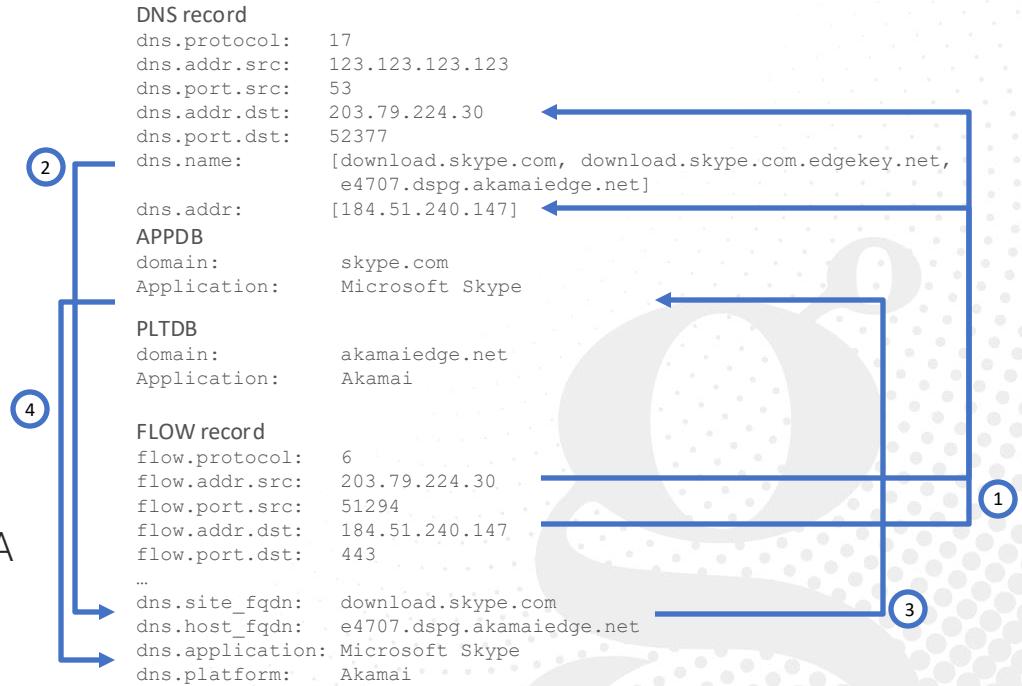
NAT attribute	value
IP-Port-Int-IPv4-Addr	10.20.89.215
IP-Port-Ext-IPv4-Addr	203.79.224.30
IP-Port-Range-Start	51200
IP-Port-Range-End	53247

Flow field	data
flow.addr.src	203.79.224.30
flow.addr.dst	173.222.180.229

field	203.79.224.30	173.222.180.229
site_fqdn	-	download.skype.com
application	-	Microsoft Skype
host_fqdn	-	e4707.dspg.akamaiedge.net
platform	-	Akamai
inside_addr	10.20.89.215	-
imsi	284011234567890	-
city	...	Hong Kong

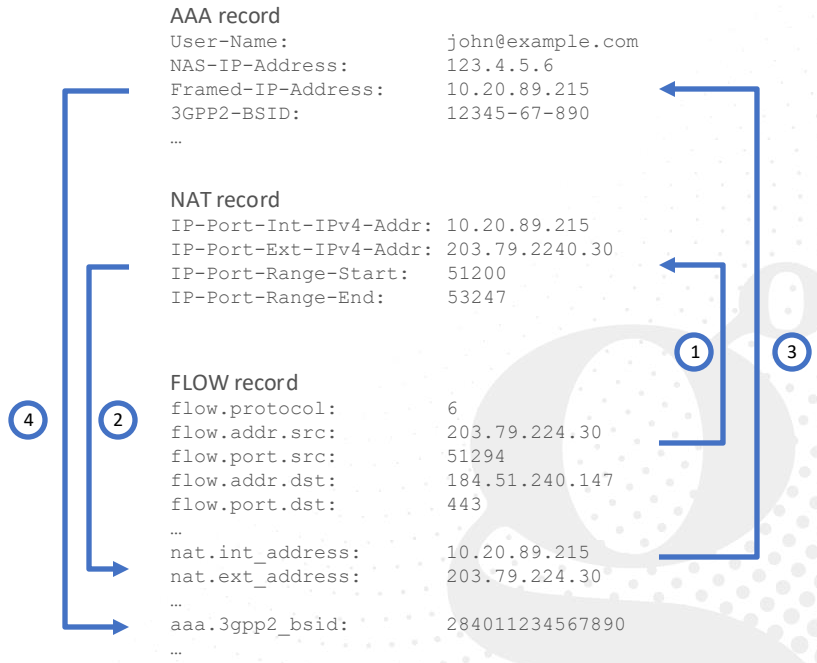
Enrich NetFlow By DNS/NAT

- Application – the first in CNAME chain
  - Site FQDN – download.skype.com
  - Application – Microsoft Skype
- Platform – the last in CNAME chain
  - Host FQDN – e4707.dspg.akamaiedge.net
  - Platform – Akamai
- Reverse DNS Table – A/AAAA
  - For dedicated servers
- Reverse DNS Table – CLIENTIP + A/AAAA
  - For shared servers



Enrich NetFlow By AAA/NAT

- NetFlow records enriched by AAA/NAT provides finer granularity of network traffic than CDR.
- Can be used to detect issues like congestion and provide insight to causes.
- Can be useful for access network planning.



Enrich NetFlow By Data Feed

- NetFlow can be enriched by 3rd party data feed for more context information. Popular data feeds are :
- GEODB – To enrich NetFlow by GEO info such as country or city. Useful for peering or transit analysis.
- IRRDB – To enrich NetFlow by registrant information. Useful for forensic applications.

prefix	country	subdivisions	radius
122.175.164.0/32	India	[Madhya Pradesh]	20.0
12.144.146.128/25	United States	[Missouri]	100.0
84.86.49.128/25	The Netherlands	[South Holland]	100.0
2a09:bac1:76e1:17a2::/63	United States	[Alabama]	1000.0
74.103.194.0/23	United States	[Rhode Island]	5.0
12.196.19.88/29	United States	[Michigan]	20.0
2a02:a450:1200::/40	The Netherlands	[South Holland]	5.0
2a09:bac1:76e0:4b28::/61	United States	[Texas]	1000.0
170.169.126.0/23	Mexico	[Querétaro]	200.0
92.76.184.0/24	Germany	[Lower Saxony]	200.0
89.114.118.0/24	Portugal	[Porto]	20.0
78.99.193.0/24	Slovakia	[Bratislava Region]	20.0
50.45.10.0/23	United States	[Illinois]	100.0
2a02:26f7:bc08:50c6::/64	Canada	[New Brunswick]	100.0
2804:14c:ce87::/48	Brazil	[Sao Paulo]	10.0
2804:14c:ce88::/45	Brazil	[Sao Paulo]	10.0
45.63.78.182/32	United States	[Illinois]	20.0
2a02:26f7:f6f3:a62b::/64	United States	[Rhode Island]	100.0
24.111.33.0/25	United States	[Minnesota]	50.0
201.145.144.0/23	Mexico	[México]	20.0

Example: GEODB Data Feed

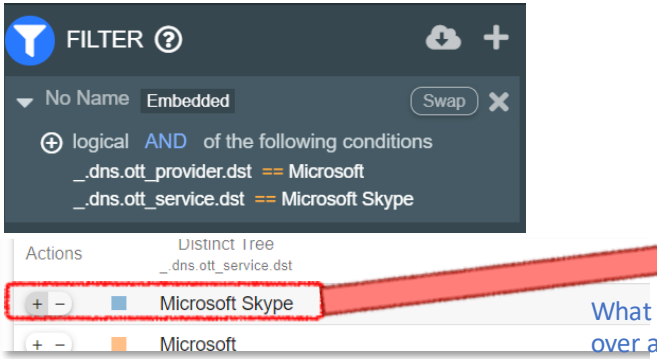
Capacity Planning | Hotspot Analysis

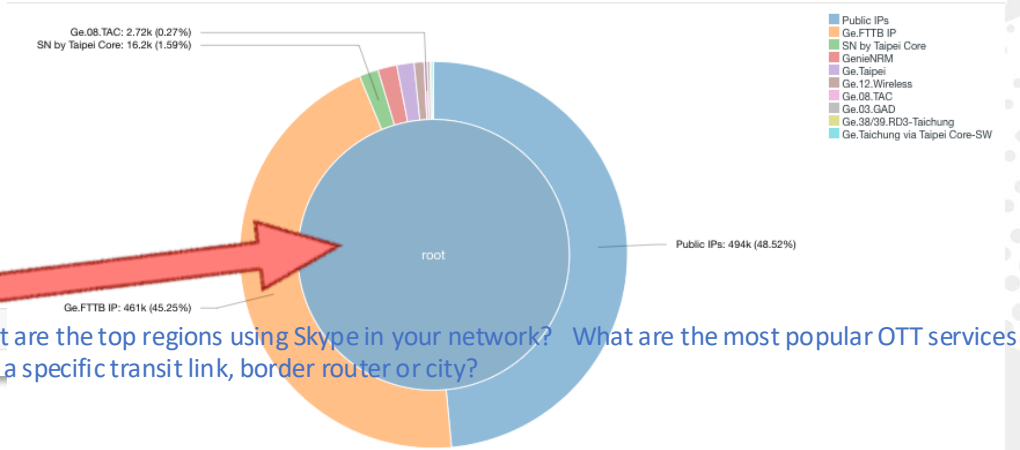
Correlate service precisely (not just via ASN) w/ various network resources/attributes

- What are the network service (e.g., OTT) preferences of users in different regions for resource deployment planning
- Need end-to-end traffic analysis that can correlate service, CDN network path, and subscriber identity

Benefits

- Help make subsequent adjustments to optimize network performance and user experience, such as locating popular services optimally to reduce traffic delivered poorly across the backbone and reduce high-cost outbound traffic





What are the top regions using Skype in your network? What are the most popular OTT services over a specific transit link, border router or city?

Capacity Planning | Traffic Analysis for Optimization

Flexible traffic aggregation of various attributes with intuitive visual presentation

- How much traffic is transiting my network at where (which links) (e.g., traffic going to a prefix, an ASN, an AS Path, a BGP Community, or a Next-hop on different links/interfaces?_
- What the traffic distribution would look like if we move a certain share (e.g., a prefix-worth, an ASN-worth, a community-worth) of traffic from one link to another?

Benefits

- Help deterministically move certain share of traffic around for optimizing the resource utilization

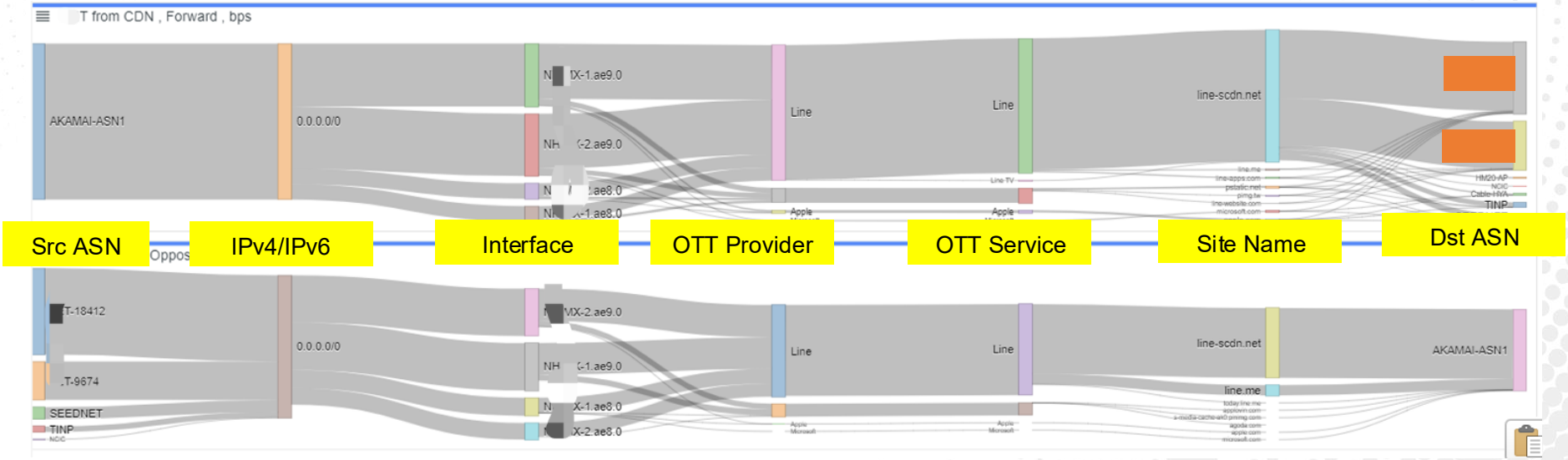
Capacity Planning | Hot Service per CDN Analysis

Break-down analysis of hot services vs. CDN providers

- How different hot services (e.g., OTT) are delivered through CDNs' on-net caching?

Benefits

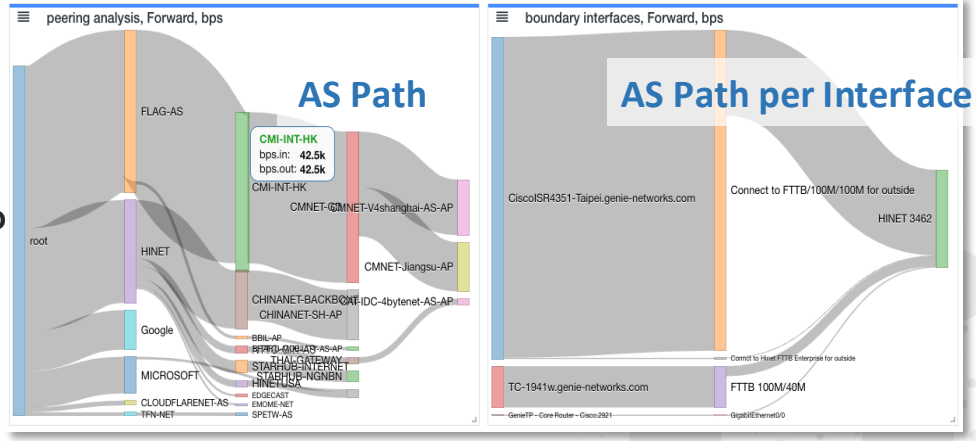
- Help plan the on-net caching resources to prevent over-loading busy cache servers and to deliver popular service traffic directly to avoid costs of distant content fetching or delivering



Peering Analysis

Full delivery paths the traffic leaving your network to its destination

- Understand the path by AS path, device/link, geography, service provider names, etc.
- Understand the destination by IP address, ASN, host/site name, service, etc.

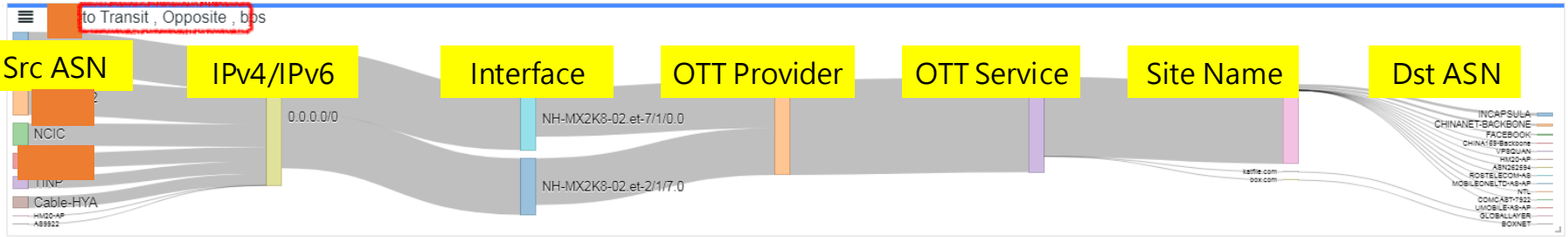


Benefits

- Help find new peering candidates to reduce transit costs
- Help find transit customer prospects to increase revenues
- Help identify unreasonable customer/partner behaviors:
  - An IDC host keeps accessing services acting like an ISP/eyeball-network An on-net CDN sends its traffic through the home network's peering links with its peers

Peering Analysis

Services traffic with full delivery path through Transit networks



Src ASN | IPv4/IPv6 | Interface | OTT Provider | OTT Service | Site Name | Dst ASN

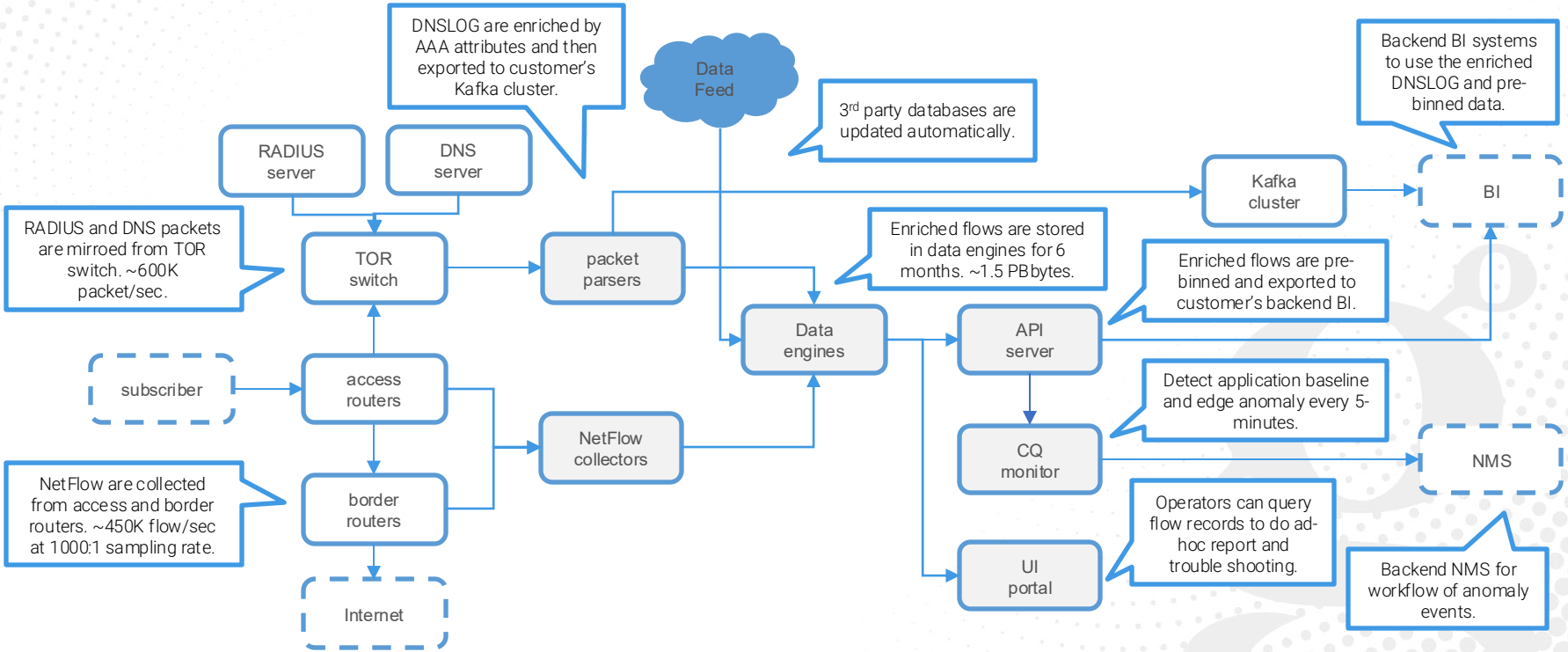
From Transit , Forward , bps

Actions	Distinct Tree Node Key	Distinct Tree Node Field	bps forward.left	bps forward.right
+ −	QUANTILNETWORKS	flow.as.src.name()	0.000	4.32G
+ −	COGENT-174	flow.as.src.name()	0.000	3.61G
+ −	AKAMAI-ASN1	flow.as.src.name()	0.000	2.70G
+ −	CMNET-GuangDong-AP	flow.as.src.name()	0.000	2.49G
+ −	Taobao	flow.as.src.name()	0.000	1.64G

to Transit , Opposite , bps

Actions	Distinct Tree Node Key	Distinct Tree Node Field	bps opposite.left	bps opposite.right
+ −		flow.as.dst.name()	0.000	1.64G
+ −		flow.as.dst.name()	0.000	1.41G
+ −	NCIC	flow.as.dst.name()	0.000	907M
+ −		flow.as.dst.name()	0.000	586M
+ −	TINP	flow.as.dst.name()	0.000	490M

Case Study – 10M Subscribers ISPX



Conclusion

- Most Internet traffic is encrypted, and passive DPI can no longer peek at the data plane. Therefore, NetFlow is a more cost-effective data source.
- By enriching NetFlow with DNS, we can still maintain enough visibility for critical applications such as CDN planning, CDN monitoring, and application monitoring.
- By enriching NetFlow with AAA, we can have finer granularity compared with CDR, which is useful for access planning and troubleshooting.
- By enriching NetFlow with other external data, we can have traffic visibility on other dimensions, which is useful for operational or business intelligence.

FAQ

- Q: Is it mandatory to mirror DNS and AAA packets?
  A: DNS and AAA log are also feasible if they are exported in real-time.

- Q: CDN edge server serves many applications. How to resolve the ambiguity?
  A: Matching by CLIENTIP+A/AAAA reduces ambiguity significantly, but not to zero.

- Q: Can NetFlow monitor application performance?
  A: NetFlow can measure elephant flow throughput, but not for RTT.

- Q: Is it legal to sniff users' DNS and AAA packets?
  A: It is legal to sniff your own DNS and AAA servers.

- Q: Flow records with DNS and AAA info is sensitive. How to protect user privacy?
  A: Hash user identity to make them anonymous, or store tags instead of identity.

THANK YOU!

www.genie-networks.com