SCTE CABLE-TEC EXPO 22
SEPTEMBER 19-22 • PHILADELPHIA, PA

2022 Fall Technical Forum
SCTE • CABLELABS • NCTA

Network Fingerprinting and Classification in Practice

An Operational Practice prepared for SCTE by

John Mansor
Vice President, Development Operations
OpenVault, LLC.
111 Town Square Place Suite 1180
Jersey City, NJ 07310
+1 (201) 677-8480
jmansor@openvault.com

Zach Simpson
Vice President, Engineering
OpenVault, LLC.
111 Town Square Place Suite 1180
Jersey City, NJ 07310
+1 (201) 677-8480
zsimpson@openvault.com

Table of Contents

List of Figures

List of Tables

1. Introduction

Network fingerprinting is an emerging classification and filtering process that utilizes standard flow protocols to extract and enrich traffic records for analytics purposes. The process utilizes both public and private enrichment resources to create a modular, templated framework for use by automated machine learning (ML) and artificial intelligence (AI) systems. The goal, to create a predictive, proactive and forecast ready system for network traffic analysis, including the evolving diversity of traffic.

Through flexible templating, network fingerprinting enables a system to rapidly identify destination bottlenecks, detect anomalies within traffic flows and even recommend package adjustments. This approach has no deep packet inspection requirement and leverages flow record and packet metadata to store and enrich existing flow sources. The separation of enrichment from machine and AI techniques supports the use of homegrown solutions such as forecasting or monitoring while also allowing the use of additional open-source models for quick deployment and rapid time to value. This flexibility is designed to enable use cases across a variety of network, threat assessment and quality of service spaces and includes models to address proactive network management, self-healing actions (platform to network connections), anomaly detection, traffic monitoring, customer churn and capacity management.

The proceeding sections introduce the basic elements needed to achieve network fingerprinting and classification processes and demonstrate possible outcomes when leveraging those resources in traffic flow environments. The processes outlined focus on enrichment and augmentation, leveraging standard traffic flow protocols at a software layer without the need for specific network inspection hardware.

2. Enrichment Templating

Enrichment templating is essential for fingerprinting exercises and begins with defining the resources, business and technical requirements necessary to augment traffic flow data from a specific system or process. Defining enrichment sources in a modular and easily parsed format provides flexibility for automation, continuous updates and analytics reuse. Aligning traffic protocol fields to enrichment sources is key and ensures accurate data discovery and analysis.

2.1. Augmenting Traffic Protocol Fields

Before beginning the enrichment templating process, it is necessary to understand the traffic flow protocol and associated fields of the system being analyzed. While there are many common fields and properties between the various traffic flow exporters, identifying and understanding both the nuances and similarities of traffic protocols such as: NetFlow, Internet Protocol Flow Information Export (IPFIX), sampled flow (sFlow) etc. will greatly expedite the templating and alignment processes. Fields available to augment and enrich may vary slightly by traffic protocol so it is important to understand availability and alignment to any specific analysis use cases.

With the traffic protocol identified, assessment of the available fields and their key properties can begin. Developing valuable enrichment templates begins with understanding the consistency of fields and data within the traffic export a system will be receiving. This consistency will be critical to the application of enrichment via systematic lookups, joining of data via common fields or similar methods used by the traffic analysis system. Examples of common fields include private Internet Protocol (IP) addresses, public IP addresses, network type, network protocols, network ports and even traffic direction.

Flow Record Field

Enrichment Source Fields derived from Client IP and Destination IP Addresses

client.ip	client.publicip	client.mac	client.package	client.monthusage	client.networkelement	org.name	application
10.0.0.1	192.168.0.1	00:00:00:00	GIGABIT PLUS - 1G	1.10 TB	CMTS-A01	MICROSOFT-CORP-MSN-AS-BLOCK	MS-Teams-Skype
10.0.0.2	192.168.0.1	00:00:00:00	GIGABIT PLUS - 1G	389.12 GB	CMTS-C01	MICROSOFT-CORP-MSN-AS-BLOCK	MS-Teams-Skype
10.0.0.3	192.168.0.1	00:00:00:00	750 MB PREMIER	753.24 GB	CMTS-G01	FASTLY	Disney Plus
10.0.0.4	192.168.0.1	00:00:00:00	500 MB Starter	1.34 TB	CMTS-A02	FASTLY	Disney Plus
10.0.0.5	192.168.0.1	00:00:00:00	GIGABIT PLUS - 1G	2.37 TB	CMTS-B01	FASTLY	Disney Plus

Figure 1 - Example flow record and enrichment source fields

Defining enrichment sources depends on analysis, modeling or automation requirements (see sections 4.2 and 5.1 for examples). Sources are normally qualified as either public or private. Public being sources or definitions broadly available or published openly by companies and individuals. While private sources normally represent internal network systems, purchased products or resources not publicly accessible.

The table below provides example definitions of several common enrichment sources, their classification type (public or private) and a brief description of their purpose in analysis systems.

Table 1 – Example enrichment source definitions

Enrichment Source	Type	Purpose
Autonomous Systems	Public	Align servers and destinations to known systems
Geographical IP Details	Public	Geographical data for public IP addresses
Application Identification	Private	IP address and protocol lookup for application enrichment
Cache Systems	Private	Internal network resource list for video or third-party service caching

When reviewing external enrichment sources, it is important to reference key fields and traffic flow protocol support to clearly decipher dependencies and alignment to protocol fields. Understanding how selected sources will impact and ultimately enrich the existing traffic flow data is crucial and plays a large role in defining not just data discovery but enrichment timing and performance. Identifying not just the fields but relevant aggregations required for analysis, can assist in preparation of the system and necessary classifications.

2.2. Leveraging Additional Network Resources

Enhancing a library of enrichment sources, begins by understanding the additional network resources capable of being leveraged for broader traffic intelligence. In some environments, traffic flow data may be the sole source of collection and in that case alternative public or private sources may be used for enrichment. However, in many network environments additional resources may be available or accessible through a variety of methods including Simple Network Management Protocol (SNMP) polling, Optical Line Termination (OLT) hardware or even Cable Modem Termination Systems (CMTS) via Internet Protocol Detail Records (IPDR). Additionally, Domain Name Systems (DNS) can provide insight into both local network destinations as well as public Internet resources, providing additional depth when analyzing traffic flow destinations.

Depending on policies and local regulations, another flexible and valuable resource for enrichment is billing system data. Even when anonymized or utilized only for subscriber package analysis, it can provide powerful insights into network health and ultimately customer experience within a network.

Billing systems may also include device and network resource details specific to subscribers, often simplifying the alignment process with traffic flow data.

Successfully leveraging the various formats and outputs of network resources normally requires a series of crafted templates. Depending on the volume of data desired for analysis and presentation, templates can be either individually constructed or merged to a single network enrichment source. The latter normally only being preferred if continuous updating of resources is not anticipated.

2.3. Aligning Template Sources

Template alignment begins with identifying fields capable of being joined together through common keys or via the fields present within traffic records. The exact method of enrichment may be dictated by the system or application being utilized for traffic flow inspection and analysis. While many traffic analysis systems support enrichment from multiple sources, some require a single formatted file. The alignment processes described assume a multi-templated approach but with the option to join to a single file location if required.

When initially developing a solution for enrichment alignment, one commonality is IP address information. While originating IP addresses can exist in traffic records in either public or private form (depending on network settings, environment), it is normally one of the simplest and most consistent origin identifiers. When combined with subscriber enrichment sources it can provide the basis for numerous enrichment field additions. Before selecting a single origin field such as IP address, it is important to understand network behavior that can contribute to field consistency, including client/server protocol behavior and Dynamic Host Configuration Protocol (DHCP) settings within the network. Since not all enrichment sources may be leveraging IP address as a primary method of joining data, various combinations of ports, protocols and network service information can also be used. When facing more complex or non-IP address enrichment scenarios, defining unique combinations of enrichment source elements that are aligned to flow record data can assist in rapid lookups at time of application.

Aligning application specific enrichment sources within templates is sometimes a broader challenge, as the variation and combination of server or destination ports and IP addresses can require continuous enrichment from both public and private locations. This continuous enrichment is critical for the effectiveness and accuracy of application specific details (see section 3.3 for additional context). However, introducing the results of more complicated enrichment sources to traffic flow data remains a clearer task when adhering to described IP address or unique string combination solutions.

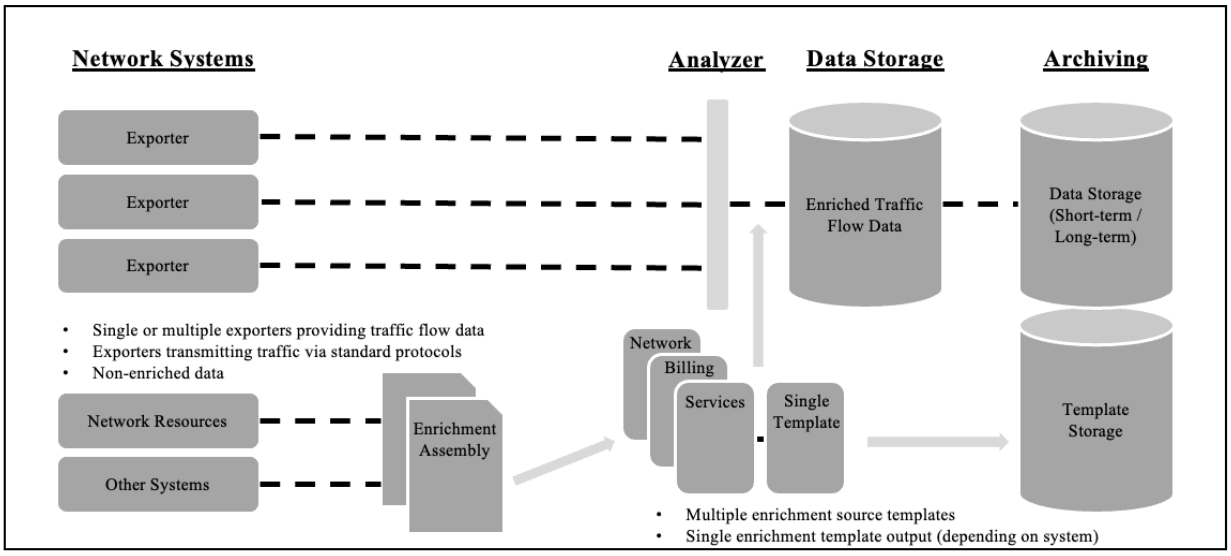


Figure 2 - Example template source alignments

2.4. Defining Enrichment Timing

There a variety of traffic analysis systems and solutions available in the market and defining when to perform enrichment of traffic flow data may depend on the solution deployed. Timing selection may also depend on the use cases and type of analysis being constructed. While real-time performance is desired or required for many scenarios, such as threat analysis or capacity monitoring, it is not always necessary for point-in-time analytics or broader experience measurements. Depending on the flexibility of the solution this can lead to other alternative time considerations for analysis that may be focused on past outcomes, point-in-time comparisons or historical grading. The templating process is an ideal time to assess when enrichment should occur and whether it is necessary for enrichment updates to be applied only to newly collected (ingested) data or whether all historical datasets should be updated.

Many enrichment sources that include IP address, service or application focused details are most valuable remaining real-time. Simply meaning that at the moment enrichment sources are updated, all traffic record data from that point forward is enriched with the source details, until the source is updated again. Network fingerprinting and classification is often the most valuable when there is a clear history of events. Whether it is IP addresses dynamically changing and being associated to a new subscriber, a content delivery network (CDN) change occurring for a particular site or application and even location based internal resources such as cache systems becoming visible in the network.

This does not mean that there are no scenarios where historical collection updates should not be planned. Adjustments due to billing, subscriber, network resources or other field classification changes are often common to ensure that analysis done for historical or trending purposes, remains accurate. When assessing historical data adjustment, it is important to understand the timing and impact on the system, volume of data that is required to be queried and ultimately updated. For scenarios where an enrichment source may only be updated weekly or monthly, planning for historical reconciliation or adjustment on a set schedule will likely be the most practical option. For further timing considerations and tradeoffs when applying enrichment to traffic flow records, see section 3.1 for additional context.

2.5. Retaining Templated Sources

Enrichment sources can be a valuable commodity over time. Depending on the source, having historical references of enriched values can aide other analysis efforts, metrics or grading. Determining an optimal process for storage, access and use within additional data lake or analysis systems can depend on the enrichment source types. Leveraging long-term or cold storage options may make sense for enrichment snapshots that are continuously refreshed while more accessible storage may assist sources that will be more commonly reused. When deciding on retention strategies for enrichment templates it is often valuable to define how often each enrichment source is updated by underlying code or processes. Sources that are refreshed on a daily or more often basis may be more desirable for short-term storage options where the data can remain accessible if required.

Example retention timeframes for templated sources, including possible storage classes and conditions can be referenced in the table below.

Table 2 – Template retention example

Archive Description	Archive Timeframe	Storage Conditions	Storage Class	Details
Autonomous Systems	On Update	Past two versions	Standard / Short-term	Two versions of the complete source stored
Application Identification	Daily	7 days	Standard / Short-term	Saved daily and stored for 7 days
Billing System	On Update	Past two versions	Standard / Short-term	Two versions of the complete source stored

When designing an effective templating process, it is important to plan for the unexpected. In a real-time system, what happens if an enrichment source becomes populated with incorrect information? Having a plan to not only rollback to a previous enrichment source but also knowing the optimal method for updating historical data to a very specific point-in-time is invaluable during a production event. While the actual restoration and historical update process remains specific to the application or system deployed for traffic analysis, ensuring that a standard set of processes accompanies enrichment retention plans will ensure enrichment accuracy or adjustment is always a possibility.

3. Applying Enrichment

The art of enrichment application begins with determining the frequency for updating traffic records and ends with the aggregation and storage of the combined flow data (original record plus enrichment, if desired). Understanding the impact that sampling rates can have on specific use cases, enrichment actions and ultimately flow record storage will help ensure required modeling, analysis or presentation function as desired. The ability to successfully enrich traffic flow records accurately and at scale depends on the consistency and quality of enrichment sources and their field alignment. This makes previous planning and templating efforts key to a performant system, capable of supporting detailed network fingerprinting and classification.

3.1. Approaches to Traffic Protocol Enrichment

As previously noted, real-time traffic enrichment is not always a reality. Sometimes the volume of data or complexity of an enrichment source requires a point-in-time or scheduled update. When the time comes to apply enrichment sources to traffic flow data, each of these three: real-time, point-in-time and scheduled,

time-based scenarios is important to consider and apply within enrichment and flow record data collection.

Real-time is normally representative of enrichment that occurs as flow records are ingested and processed by the traffic analysis system. Relevant enrichment sources are leveraged to populate supplemental traffic record fields and augment the existing traffic flow data. In a real-time enrichment scenario, historical data is not updated. Depending on collection and fps rates, the actual moment an enrichment source is updated, collected data is supplemented. Real-time enrichment allows for supplemental traffic fields to be used for immediate visualization, presentation and modeling as records are ingested by the system.

Point-in-time enrichment occurs after flow records are ingested and processed by the traffic analysis system. Relevant enrichment sources are leveraged to populate supplemental traffic record fields and augment the existing traffic flow data. In a point-in-time enrichment scenario, data updates are made to historical traffic flow records only. Supplemental fields are added or updated after collection and only a dependency on system ability, performance is needed to make the updates. Point-in-time enrichment allows for supplemental traffic fields to be used for visualization, presentation and modeling between two time periods and after records are ingested by the system.

Scheduled enrichment is similar to point-in-time and occurs only after flow records are ingested and processed by the traffic analysis system. Relevant enrichment sources are leveraged to populate supplemental traffic record fields and augment the existing traffic flow data. In a scheduled enrichment scenario, data updates are made to historical traffic flow records only but are done so on a consistent frequency in order to align or aggregate key fields. Supplemental fields are added or updated after collection and only a dependency on system ability, performance is needed to make the updates. Scheduled enrichment allows for supplemental traffic fields to be used for visualization, presentation and modeling between two time periods and after records are ingested by the system.

Understanding tradeoffs that need to be made for the performant function of the traffic analysis system will save time and may help simplify the approaches selected for specific enrichment sources. When combined with sampling and continuous refinement processes, these approaches assist in the overall accuracy and classification capabilities of the system.

3.2. Traffic Flow Sampling and Enrichment Impact

Sampling rates of exported traffic flows can have an impact on the type of enrichment being performed. The level of sampling often pertains to specific use cases or automated actions being taken. When assessing traffic flow collection and enrichment, it is important to understand the volume of devices, flows per second (fps) and even internal or external traversal of the network element being collected. Each of these factors represent considerations and influence use cases for network fingerprinting and classification exercises. For example, in threat analysis and security spaces, no to extremely low sampling rates are often preferred to ensure that malicious traffic is not missed and can be cataloged and actioned effectively. While higher sampling rates can provide adequate intelligence for prime time and capacity planning use cases. While there is a wide variation in collection and enrichment use cases for traffic analysis systems, there are a few common scenarios that can help guide enrichment planning.

Determining an optimal sampling rate is often dependent on business, network and retention requirements. The network device itself may also dictate maximum sampling levels based on protocol or vendor defined parameters. The scenarios and use case recommendations provided in the table below may not apply to all customer environments. Threat analysis with low or no sampling scenarios should be reviewed on a case-by-case basis to ensure optimal experience and data retention.

Table 3 – Traffic flow sampling scenario use case examples

Use Case	Sampling
Capture and aggregate detailed traffic patterns across device base. Monitoring of key internal cache resources for capacity management, alerting and team response.	1:512
Detailed traffic flows for client/server, protocol and application analysis. Determine traffic locations and key server destinations for capacity planning and design.	1:2048
Develop overall network insights particularly during peak/primetime hours without retaining larger quantities of traffic data. Analyze and aggregate with network usage and performance metrics.	1:10000

Sampling rate is an incredibly important consideration when designing and setting up the actual platform where traffic flow record and enrichment will take place. It cannot only determine the effectiveness of the overall solution but also influence enrichment value, modeling and analysis.

3.3. Continous Refinement and Update

Many enrichment sources require continuous care and feeding to ensure accurate results. As part of the templating process, enrichment sources should be clearly planned and defined. Once these sources are put to work enriching traffic flow records, keeping them up to date is critical to accurate fingerprinting of network resources. For certain enrichment templates, even a short lapse in updates can create data inaccuracy which may then require an adjustment to historical data. The complexity of continuous refinement within a traffic analysis system depends largely on the deployment and the context of the enrichment source. Enrichment sources focused on specific services classification, applications, CDNs and IP address ranges normally represent the highest risk for inaccuracy given change volatility. When finalizing enrichment application activities: documenting, alerting and implementing self-healing capabilities are often seen as best practices for these process critical sources.

Examples of refinement timing and associated enrichment sources are briefly described in the table below.

Table 4 – Enrichment source automated refinement example

Enrichment Source	Type	Timing	Details
Autonomous Systems	Public	Real-time	Align servers and destinations to known systems
Application Identification	Private	Real-time	IP address and protocol lookup for application enrichment
Billing System	Private	Weekly / Historical	Billing system reference and broadband package details

Leveraging built in functions of the traffic analysis system deployed or integrating surrounding scripts or functions to perform continuous enrichment on sources with the most risk, is always a recommended practice. Introducing more complex modeling, analysis and AI to enriched traffic flow data requires a high degree of refinement and accuracy.

3.4. Aggregation and Storage

A clear aggregation and storage strategy is necessary to ensure data remains accessible and retained for critical analysis and reuse. When considering aggregation and storage options it is often best to refer to enrichment template planning processes to guide decisions on what fields may require aggregation. Storing enriched traffic flow field data is not only valuable for historical purposes but often leads to an accelerated path for training AI models. Having aggregated data sets readily available to test or pilot various service techniques, can reduce time to value for analysis and modeling.

While simply storing all data, all the time, forever seems like the most ideal method for long-term historical reference, it is not always feasible for large traffic volumes. This means that for many traffic flow record scenarios, an approach that uses time-based aggregation can be considered a valuable alternative. Selecting the most ideal aggregate will depend on the use cases and traffic analysis system being utilized (support or performance of aggregation). There are many options for field aggregation in a traffic flow system, a time-based approach helps to ensure data can still be introduced to more advanced analysis models quickly. There any number of aggregate storage options based on roll-ups by minutes, hours, days etc. Choosing the right timeframe may also depend on audit scenarios, archive requirements and simply cost.

The table below introduces several timeframes and possible storage class options based on the type of aggregate data being stored.

Table 5 – Aggregate storage options and timing example

Archive Description	Data Timeframe	Storage Timeframe	Storage Class	Details
Enriched flow records	7 days	7 days	Standard / Short-term	Raw traffic and enriched records flow records available for 1 week
Raw flow records	7 days	7 days	Standard / Short-term	Saved for re-processing or comparison purposes
Enriched flow records aggregated	1 day	120 days	Cold / Long-term	Aggregated to a 24-hour period and stored
Enriched flow records aggregated	7 days	365 days	Cold / Long-term	Aggregated to a 168-hour period and stored

4. Introducing AI Service Models

Following an introduction to some of the techniques and processes needed to enrich and augment traffic flow records, it is time to leverage the available data for more advanced modeling and use cases. Introducing AI techniques to enriched traffic flow data enables network fingerprinting and associated classifications to take shape. The advent of numerous open data models as well as commercial solutions to assist with visualization and presentation can make detailed levels of classification a reality for almost any network. The success of each model and use case depends on the volume of data collected, availability of enriched fields and overall dependencies for the analysis being performed. Diligent and consistent enrichment processes help to ensure that the data is viable and more suitable for training and modeling processes. The process of data discovery and AI model selection normally begins with identifying key fields or aggregated points of interest within the enriched traffic flow data. These key data points will be used for model classification and training development activities. Assessing and preparing

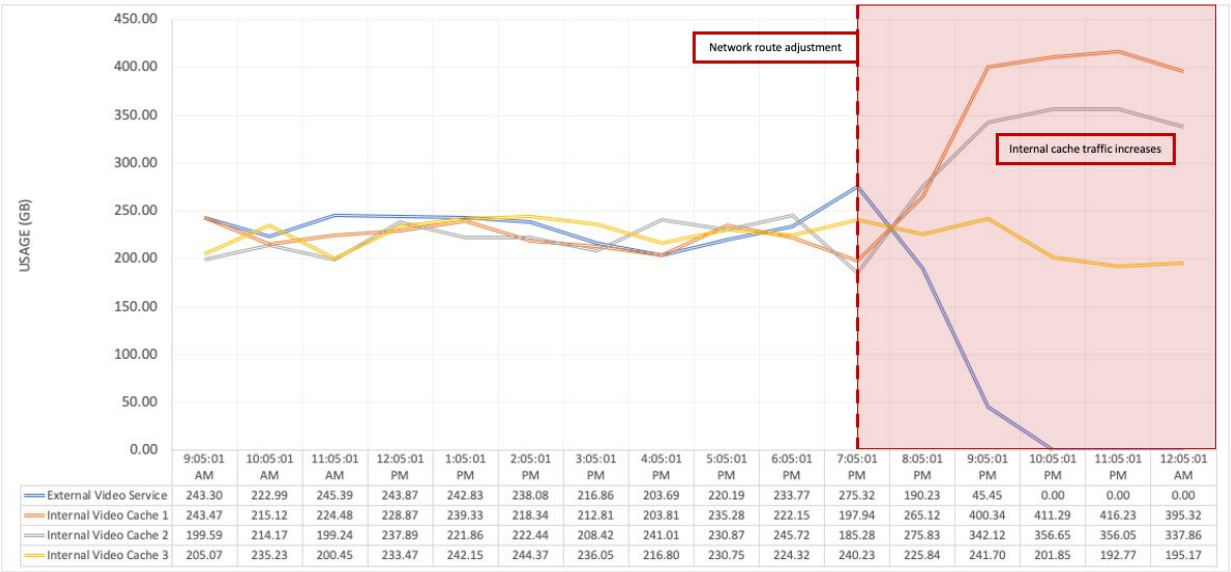
enriched traffic flow data for modeling may require exporting or saving data sets for analysis and testing prior to production implementation.

4.1. Fingerprinting Use Cases

While the use cases for fingerprinting the network is ever widening and can even be system specific, there are now a variety of paths supported by readily available market and open solutions. Building and maintaining accurate fingerprints for the network normally means having a system and templating process that is adaptable and capable of growing with the network analysis and classification needs of the environment. This section provides several use cases and attempts to provide insight into specific enrichment templates, fields and techniques utilized to achieve the results. These use cases require various forms of network or traffic enrichment to be successful and are provided as examples. Detailed analysis and response as described throughout the scenarios, may depend on a variety of business and system configurations.

Scenario 1: Internal resource optimization via network system enrichment

In this scenario, the goal is to understand how external video service traffic is traversing the network over a specified period. During this primetime window, a series of internal video cache servers is being observed along with the primary external video service server destination. The video cache servers should be inclusive of the relevant markets being observed and therefore should carry most of the traffic from client systems. However, during the observed window there is still a large quantity of traffic, preferring the external destination vs. internal caches. Leveraging real-time traffic flow enrichment, the capacity and network systems adjust network configurations, re-shaping the traffic and ensuring cache usage. This scenario leverages several enrichment templates: autonomous systems, application identification and internal network resources (video cache systems).



	9:05:01 AM	10:05:01 AM	11:05:01 AM	12:05:01 PM	1:05:01 PM	2:05:01 PM	3:05:01 PM	4:05:01 PM	5:05:01 PM	6:05:01 PM	7:05:01 PM	8:05:01 PM	9:05:01 PM	10:05:01 PM	11:05:01 PM	12:05:01 AM
External Video Service	243.30	222.99	245.39	243.87	242.83	238.08	216.86	203.69	220.19	233.77	275.32	190.23	45.45	0.00	0.00	0.00
Internal Video Cache 1	243.47	215.12	224.48	228.87	239.33	218.34	212.81	203.81	235.28	222.15	197.94	265.12	400.34	411.29	416.23	395.32
Internal Video Cache 2	199.59	214.17	199.24	237.89	221.86	222.44	208.42	241.01	230.87	245.72	185.28	275.83	342.12	356.65	356.05	337.86
Internal Video Cache 3	205.07	235.23	200.45	233.47	242.15	244.37	236.05	216.80	230.75	224.32	240.23	225.84	241.70	201.85	192.77	195.17

Figure 3 - Internal resource optimization via network system enrichment

Scenario 2: Capacity planning, analyzing past trends and known events to predict future growth

In this scenario, the goal is to understand how streaming video service traffic is traversing the network and how that traffic may take shape based on historical references. During this primetime window, streaming services traffic is being monitored and a forecast view overlayed on the typical chart to indicate expected traffic patterns beyond the current time. To accomplish this both application specific enrichment along with processed, historical traffic data is combined to provide a forecast of usage for the upcoming hours. Leveraging a multi-week baseline of the exact time, day of the week and allowing for any specific events or conditions, the data is then constructed and appended to applicable charting metrics. The accessibility of processed, historical traffic data enables the system to make continual refinements and updates increasing the accuracy of forecasted usage over time. This scenario leverages several enrichment templates: autonomous systems, application identification and processed, historical traffic data.

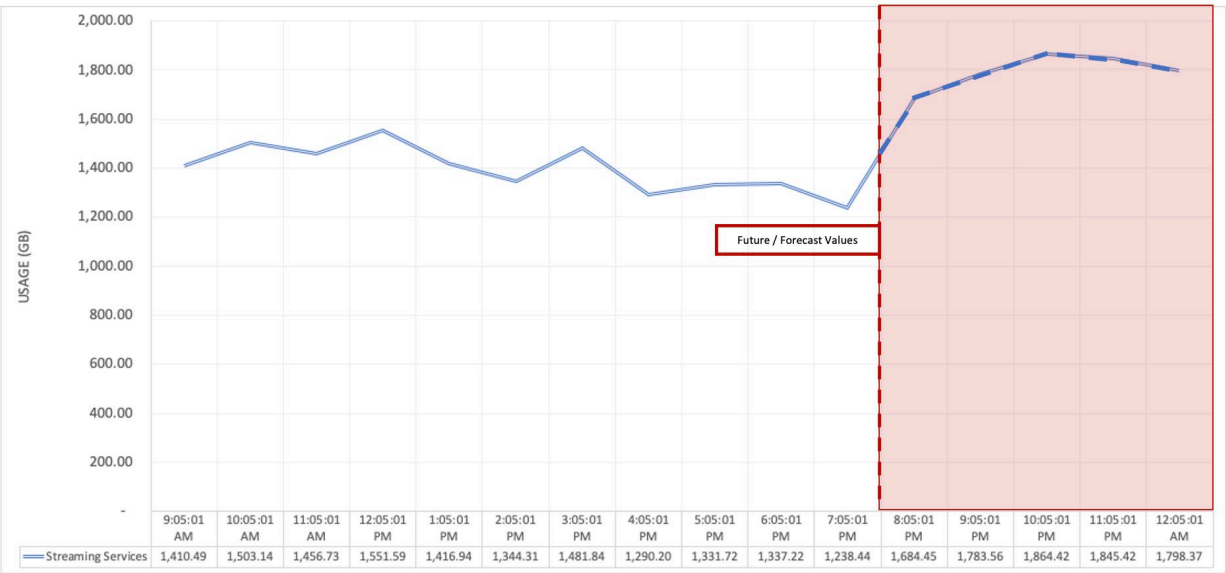


Figure 4 - Capacity planning, analyzing past trends and known events

Scenario 3: Anomaly detection across selected classifications

In this scenario, the goal is to identify traffic anomalies early and enable rapid response through either manual or automated intervention. The visualization represents a series of connections across sets of classified traffic. This includes traffic that is categorized as gaming, CDNs, streaming services etc. The anomaly detection process is examining the enriched traffic classifications for deviation at either the upper or lower bounds. In a real-time system, as soon as enriched traffic flow data is made available to the detection process, anomalies are identified and can be actioned. In this case, the classification is a particular CDN that has been categorized by the system and is no longer receiving connections. This scenario leverages several enrichment templates: autonomous systems and application identification.

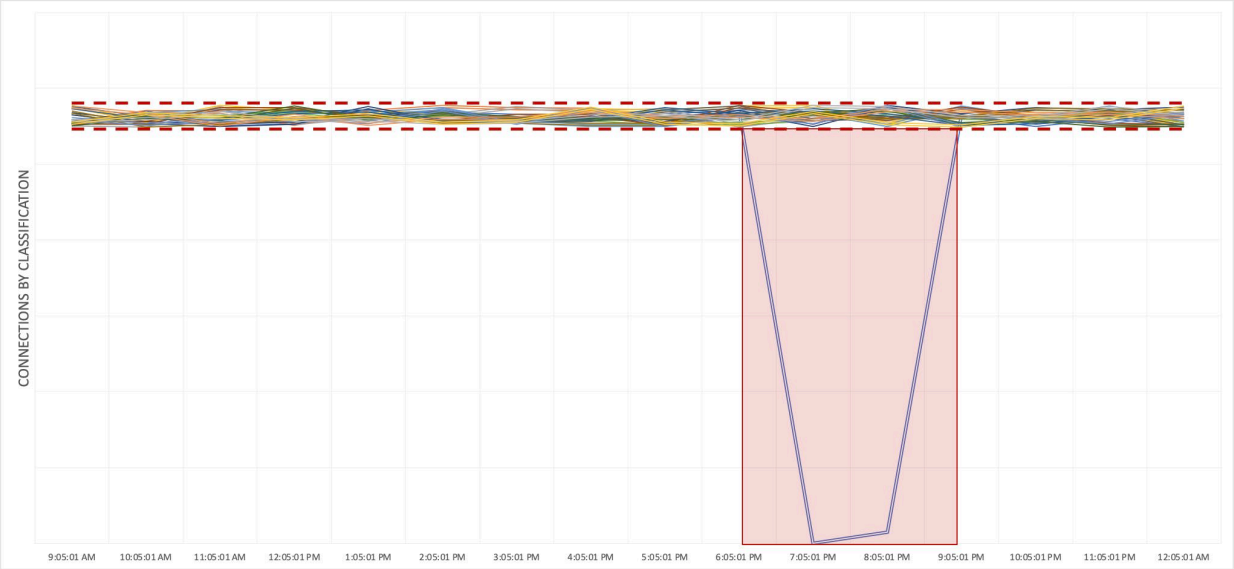
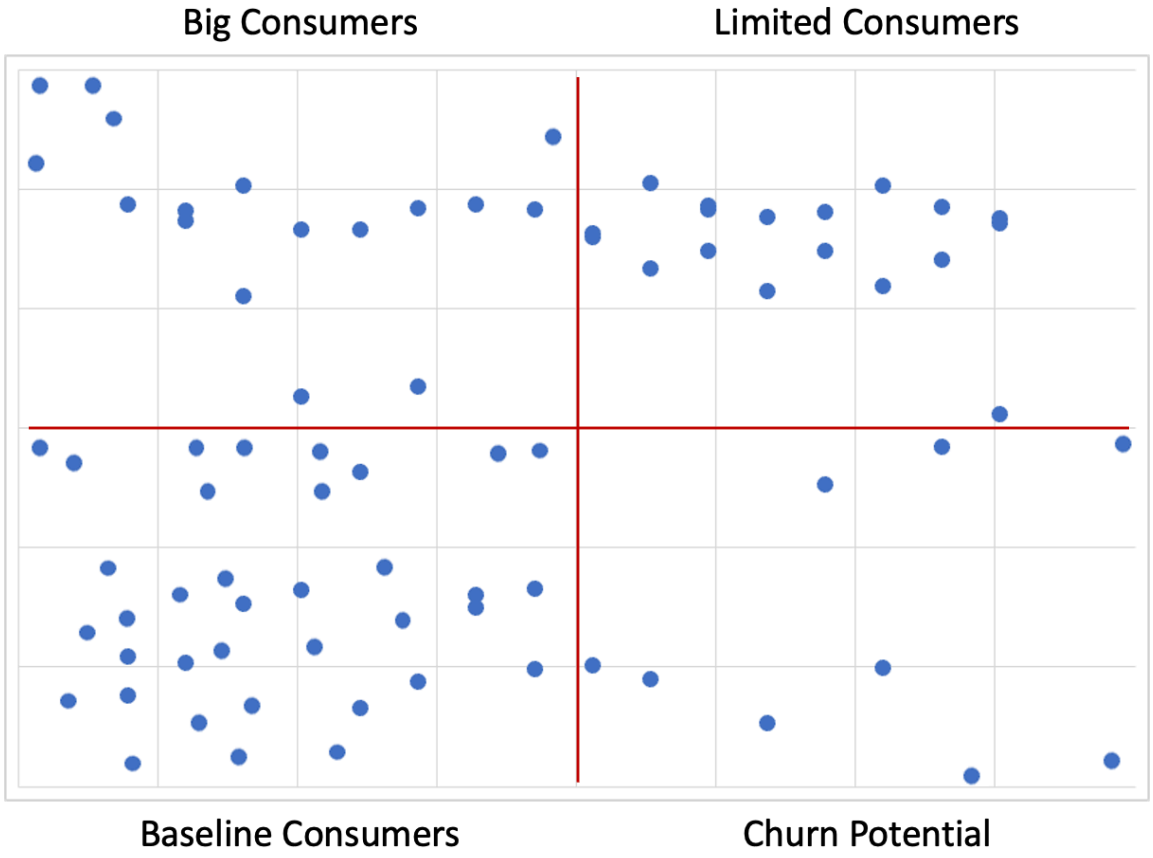


Figure 5 - Anomaly detection across selected classifications

Scenario 4: Identifying the potential for customer churn

In this scenario, the goal is to understand how a combination of traffic data, enrichment sources and network utilization information can help determine the potential for customer churn. This scenario provides a unique visualization, aggregating the various enriched data sources to plot client level data based on numerous activity conditions. These conditions include activity during peak and non-peak times, speed test usage and even references to supplemental customer support system resources. This provides a unique visualization of potential customers with activity that may represent the potential for churn. Drilling into the individual points of the data then provide inspection at the client level including all relevant enrichment detail available. This scenario leverages several enrichment templates: autonomous systems, application identification, internal network resources and customer support system data.



- **Big Consumers:** Power-users, routinely outside of assigned package
- **Baseline Consumers:** Consistent and expected activity based on assigned package
- **Limited Consumers:** Consistently below baseline activity of assigned package
- **Churn Potential:** Measurable decrease in activity, increased speed test usage and/or increased support interaction

Figure 6 - Identifying the potential for customer churn

5. Proactive Monitoring and Network Management

Proactive response remains one of the most powerful elements of network fingerprinting and classification. Leveraging enriched traffic flow data, modeling and analysis methods to enable proactive actions in many forms, should be the goal of any traffic analysis system. Whether those actions are designed to be fully automated, with intervention or in the form of customer experience metrics, there are any number of alerting and automation scenarios that can be developed.

5.1. Monitoring, Automation and Grading

Deploying monitors and automation within the network environment requires a functional traffic system with enrichment and/or deployed analysis use cases described in section 4.2. With an optimized traffic analysis system and with clear understanding of any sampling limits that may exist, a wide range of monitoring and automation use cases can be developed. Many of these use cases are made ever more powerful when combined with real-time monitoring and alerting functionality. This can include

automated dispatch of notifications, reports and information via team communication platforms, internal systems and devices. The ability to proactively tune and adjust devices based on analyzed data, either through internal processes or automated services, can assist in reducing manual intervention and drive self-healing abilities in the network. Example use cases for monitoring and automation range from capacity alerts to detailed threat assessments.

The table below provides several examples and includes common internal system dependencies.

Table 6 – Monitoring and automation use case examples

Use Case	Example Enrichment Dependency
Internal network capacity utilization (caches)	Internal network resources
Outage condition detection based on traffic and usage	Usage data, billing, IPDR
Outage detection for popular services / systems	Autonomous systems
Primetime or peak hour monitors	Applications
Congestion and network bottleneck visibility	Usage data, billing, IPDR
Top client / server monitors	Autonomous systems, Applications
Application and service optimization	Autonomous systems, Applications
Threat conditions and malicious traffic detection	Threat and security systems

In addition to monitoring and automation use cases, the availability of a broad range of enrichment templates can enable a variety of customer experience measurements. With network usage, subscriber or billing data availability it becomes possible to develop detailed experience and network grading components. Designing comprehensive experience monitoring and grading requires a deep understanding of not just the network system but relevant dependencies determining experience. This can include in-home devices, node, CMTS or optical layers, interoperability dependencies and even Over the Top (OTT) service performance. Establishing a grading process first requires assembling a list of available template fields that can be used either in real-time or as part of a point-in-time analysis. In its most simplistic form, scoring can be defined as a range of values equating to either a positive or negative network experience. Deciding on the fields to be leveraged may require an understanding of geographic, market, network or capacity constraints.

While the availability of enrichment sources and overall system capability may vary, the table below provides several example fields from previously discussed enrichment sources to attempt to aggregate a network experience score at a client level. Providing an overall score at the client level enables additional roll-up or aggregate opportunities and can provide additional granularity should it be required. In this example, a simple score of either a 0 or 1 is used per grading category or column based on pre-defined rules. This grading could be expanded to more sophisticated scoring and inclusive of many other factors and columns. These pre-defined rules can include evaluation of peak utilization of network elements and resources impacting a client over any number of days (30 days in the table below).

Table 7 – Deriving network resource scores from traffic intelligence

Client	Anomaly Impact	Anomaly Score	Peak Network Utilization	Network Score	Service Interactions	Service Score	Total
10.0.0.1	2.06 hours	0	93%	0	0	1	1
10.0.0.2	0 hours	1	71%	1	0	1	3
10.0.0.3	0 hours	1	68%	1	1	0	2

6. Conclusion

In summary, this practice has focused on enabling a structured process utilizing optimized enrichment templates and flow data analysis at the software layer. Network fingerprinting is an invaluable process for managing and enriching standard traffic flow protocol data. It is an inexpensive solution to stand up and does not require expensive deep packet inspection (DPI) appliances or additional hardware within the network. Most of the data provided by network routers are available from standard flow exporters and while the output can vary slightly by vendor, it is possible to rationalize those fields to get a full view of network traffic. While enrichment processes may need to run at varying frequency depending on source and desired outcomes, the combination of traffic flow data with additional network elements can help provide end-to-end network transparency. The inclusion of fiber and data over cable service interface specification (DOCSIS) termination systems can provide additional subscriber details that further augment network management potential. These classification and enrichment techniques help with fundamental understanding of traffic patterns, subscribers and can assist in improving quality of experience (QoE), by right-sizing internal vs external network utilization. The availability of collected data enables network, capacity and analyst personnel to better leverage traffic flow data through common AI and ML models. These models enable the visibility of anomalies and potential forecasting of customer churn. Proactive network management can be the culmination of data collection, enrichment, AI modeling and associated alerting. It enables continuous anticipation of networking issues and provides an additional mechanism for improving overall customer experience. Network fingerprinting as an emerging technique allows for multiple system operators (MSO) to anticipate issues, evolve network activities and ensure customers can continue to expand their broadband consumption.

Abbreviations

AI	artificial intelligence
ASN	autonomous system number
CDN	content delivery network
CMTS	cable modem termination system
DHCP	dynamic host configuration protocol
DOCSIS	data over cable service interface specification
DPI	deep packet inspection
fps	flows per second
IP	Internet protocol
IPDR	Internet protocol detail record
IPFIX	Internet protocol flow information export
ML	machine learning
MSO	multiple system operator
OLT	optical line termination
OTT	over the top
QoE	quality of experience
RDK	reference design kit
sFlow	sampled flow
SNMP	simple network management protocol

Bibliography & References

[Netflow] https://www.cisco.com/c/en/us/products/ios-nx-os-software/ios-netflow/index.html

[sFlow] https://sflow.org/sflow_version_5.txt

[IPFIX] https://en.wikipedia.org/wiki/IP_Flow_Information_Export

[RDK] https://wiki.rdkcentral.com/display/RDK/RDK+Documentation

[PHYv3.1] DOCSIS 3.1, Physical Layer Specification, CM-SP-PHYv3.1-I16-190121, January 21, 2019, Cable Television Laboratories, Inc

[PNM-3.1] Primer for PNM Best Practices in HFC Networks (DOCSIS 3.1), CM-GL-PNM-3.1-V03-220118, January 18, 2022, Cable Television Laboratories, Inc

[Anomaly Detection] https://aws.amazon.com/blogs/machine-learning/anomaly-detection-with-amazon-lookout-for-metrics/

[Customer Churn] https://aws.amazon.com/blogs/machine-learning/predicting-customer-churn-with-no-code-machine-learning-using-amazon-sagemaker-canvas/

