SANDVINE

Advanced Traffic Classification

Providing a reliable and future-proof foundation for Active Network Intelligence

To power real-world use cases, a traffic classification capability must:

- Provide broad coverage of at least 95% of internet traffic – including regional services and niche applications
- Provide billing-grade accuracy with zero false positives for key applications
- Provide insights that go deeper than surface-level identification

Sandvine's advanced traffic classification:

- Recognizes almost 5,000 protocols, services, and applications
- Is maintained and extended through weekly updates – plus urgent updates when key applications change
- Is future-proof against encryption

INTRODUCTION

Accurately identifying internet traffic is the foundation of Active Network Intelligence. Without this ability, operators are unable to confidently take action to optimize their networks, offer innovative service plans, comply with regulatory requirements, defend against fraud, or even study usage trends.

Functional requirements to enable Active Network Intelligence

To serve as an effective foundation for real-world use cases, a network intelligence platform must:

1. Deliver sufficient performance to identify traffic in real time and at the throughput of the network
2. Provide broad coverage of protocols, applications, and services – including regional and niche offerings – so that operators can act on practically all traffic flowing through their networks
3. Be trustworthy, with as close to zero false positives as possible, both for effective traffic management and for accurate charging
4. Provide very granular classification that extends into individual service and application providers, distinguishes between types of traffic within a single application, and can extract meaningful metadata

Building and maintaining an effective traffic classification capability

The composition of internet traffic varies considerably from region to region and is ever-changing. In recent years, the impact of the arrival of mashup applications, the growing adoption of protocols that use flow multiplexing, and increasing rates of encryption – due in combination to TLS 1.3, DNS encryption, Google QUIC and IETF QUIC, VPN services, and masquerading applications – has been felt by operators.

Many legacy systems are unable to cope with these development – particularly encryption – and operators who rely on these solutions have watched as reports show that previously recognized traffic shifts into the catch-all of "unknown" or "unrecognized" buckets.

But vendors must go beyond just developing the technology to recognize traffic: operational processes must be in place to maintain the ever-growing library of signatures, to continually test recognition as new consumer devices appear, and to deliver updates to systems already deployed in the field.

Sandvine has an unmatched traffic classification track record resulting from:

- Operational experience powering business intelligence, traffic optimization, and billing/charging solutions for the world's largest and most innovative communication service providers
- Two decades of ongoing research and development
- Proactive monitoring of global traffic trends

Regular signature updates every Monday – plus emergency updates when key applications change – ensure operators have the trusted visibility they need to monitor and manage their networks with confidence.

It is extremely difficult to distinguish between different services within a multiplexed flow – but there is value in being able to do so

CHALLENGES IN THE REAL WORLD

Accurately identifying internet traffic and extracting important information is rarely simple. In recent years, a combination of specific developments and broad trends has made the task even more difficult – and in doing so has rendered legacy traffic identification solutions obsolete.

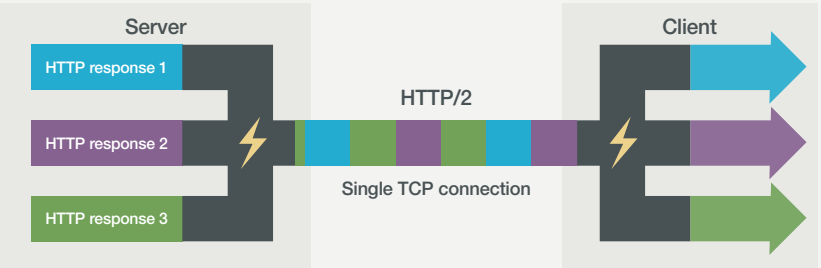
Flow Multiplexing

Flow multiplexing was introduced in RFC 7540 – Hypertext Transfer Protocol Version 2 (HTTP/2) – and is increasingly employed by major applications.

It aims to increase transport efficiency by reducing the overhead associated with opening multiple connections by instead delivering multiple services over a single flow. However, it has the consequence of making it very challenging to distinguish between different types of traffic within the multiplexed flow. Many popular applications use flow multiplexing, for example WhatsApp text messages, contacts syncing and other control messages may be delivered over the same flow. Facebook web browsing (scrolling), video and Facebook live may similarly be delivered over the same traffic flow.

Figure 1

In flow multiplexing, multiple services are delivered over a single flow; for instance, a single HTTP/2 connection can contain multiple concurrently open streams, with either endpoint interleaving frames from multiple streams



Accurately identifying mashup applications requires the ability to correlate multiple seemingly independent flows to different services

Mashup Applications

A mashup application uses content or services from more than one source to deliver a single user experience. Perhaps the most famous mashup application is Pokémon GO, which stormed to prominence in 2016 and employed several different APIs and services (including Google Maps) to deliver an augmented reality gaming experience.

On the wire, mashup applications manifest as multiple connections to separate services, and being able to associate together these seemingly independent services is an essential element of recognizing such applications accurately.

The implication of encryption for network operators is clear: applying ANI requires the ability to recognize encrypted applications, services, and traffic types

Encryption

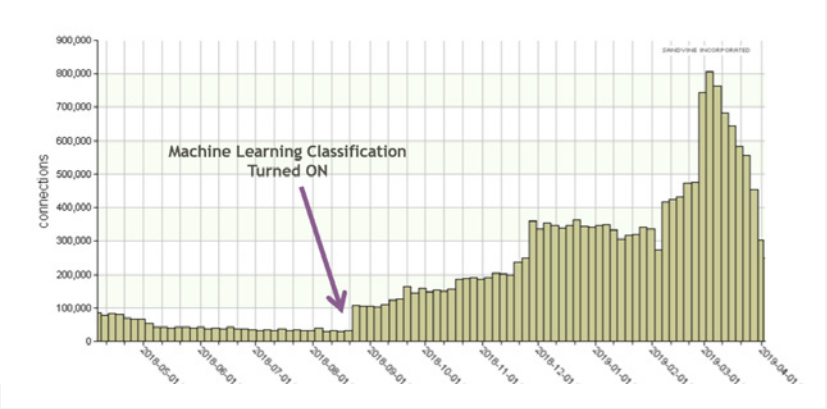
While encryption itself is not new, it has become more prevalent in recent years as a result of several developments and trends:

- TLS 1.3 introduced new standards for how encryption is implemented; significantly, certificates are no longer visible
- Encrypted DNS is gaining traction in many forms, including DNS over HTTPS, DNS over TLS, and DNS over QUIC
- Encrypted SNI (ESNI) encrypts the Server Name Indication (SNI) field within a certificate
- Google QUIC and IETF QUIC (with header protection) have had a significant impact on how over-the-top services deliver content and are particularly relevant on mobile networks
- VPN services – whether simply for privacy or to bypass content restrictions – and masquerading applications (e.g., Psiphon, Lantern) seek to deliberately obscure the nature of internet traffic and have gained widespread adoption (and notoriety) in recent years

Although the journey towards an increasingly encrypted internet has been dotted with fits and starts, the direction is clear, as is the implication for network operators: applying Active Network Intelligence requires the ability to recognize encrypted applications, services, and traffic types.

Figure 2

Sandvine's experience is that network operators always underestimate – by a significant multiple – the amount of VPN traffic on their networks; this chart shows both the immediate impact of introducing Sandvine's VPN Classification feature and the rapid growth in popularity of VPNs



Even the most basic traffic identification systems can apply regular expressions across multiple packets

IDENTIFICATION AND CLASSIFICATION TECHNIQUES

The building blocks of traffic classification include the ability to apply regular expressions within and across multiple packets, to measure and track packet sizes, and to examine headers and to look within packet payloads.

Advanced traffic classification, however, extends much deeper – combining multiple techniques that both leverage and go beyond these building blocks.

Stateful recognition is a cornerstone of many traffic classification solutions; however, many legacy systems lack the memory resources to apply complex state machines at meaningful scale

State Machines

Many protocols are stateful; that is, as information is transferred, the parties on either end of the exchange (e.g., a client and server, or two peers) each have a shared understanding of the 'state' of the communication. TCP is a familiar example, although it is relatively simple compared to many other protocols.

Sandvine's traffic classification technology employs state machines to track, in real time, the state of a protocol. Knowing the state is a fundamental enabler of distinguishing between similar protocols and to accurately identify different traffic types within a protocol, because the state determines how and where to apply the building blocks outlined above.

Parsers and analyzers read information from flows to power insightful business intelligence and to enable more accurate identification

Parsers and Analyzers

Parsers and analyzers take state machines a step further: beyond understanding the state of each protocol, parsers extract key pieces of information which – in addition to having value in and of themselves – enable analyzers to 'link' separate flows.

For instance, a parser can extract fields including service name, user agent, referrer, URL, SSL fields, QUIC fields, and others – all of which have value from a business intelligence standpoint.

But a parser can also extract information from a control stream that tells an analyzer where to look for a corresponding data stream. Without this vital insight, it would be very difficult to associated the data stream with the appropriate protocol. Analyzers, beyond providing enhanced stateful capabilities, are also crucial for measuring flow and session characteristics.

Years of research and development equip Sandvine to use behavioral correlation to reliably link together related flows from different protocols and services – a capability which is critical for identifying mashup applications

Behavioral Correlation

While parsers and analyzers 'read' data fields to link together flows belonging to a single protocol, behavioral correlation allows Sandvine's traffic classification to link together associated flows spanning multiple protocols.

As a simple example, consider a DNS request for YouTube, followed by an SSL exchange to known YouTube servers. If these are the only two flows observed during a particular period, then it is known with very high confidence that they are related. If there are many flows happening concurrently, as is more likely the case, then correctly associating related flows becomes significantly more challenging – and Sandvine's years of research and development become critical.

Machine Learning

Sandvine uses supervised machine learning models which have been pre-trained in-house and validated for accuracy. These proprietary models and techniques are built upon 150 different flow parameters.

Employing these techniques, Sandvine is able not only to broadly classify encrypted traffic into categories (e.g., Web Browsing, Video Streaming, VoIP, etc.) but also to accurately classify unique applications within categories (e.g., Facebook vs. Instagram, WhatsApp vs. Lime, Netflix vs. YouTube) – even when the traffic is encrypted and ESNI is in use.

Plus, machine learning techniques can also be applied to distinguish between authentic traffic and traffic masquerading as something else. For example, when an evasion tool like Psiphon is employed, then flows may represent themselves as a certain protocol or service (e.g., Facebook); however, the deception becomes clear to a sufficiently advanced machine learning algorithm because the true underlying behavior of an authentic service is extraordinarily difficult to emulate.

Machine learning is the pinnacle of traffic classification technology; Sandvine's machine learning algorithms leverage up to 150 flow parameters, none of which are affected by encryption – ensuring a dependable, future-proof investment

Crucially, these techniques do not rely on reading or extracting fields from flows; as such, Sandvine's machine learning is impervious to encryption. Because many services are already encrypted and VPN use is widespread, these techniques are already important today, but they will be essential – and will perhaps even account for the majority of traffic recognition – in the future.

LEVERAGING THE GLOBAL INTERNET PHENOMENA PROGRAM

Sandvine's Global Internet Phenomena program is an industry-renowned examination of internet traffic and consumer usage trends.

By analyzing aggregated usage data from customer networks and probe sites around the world, Sandvine is able to project internet traffic trends and changing usage characteristics – even months and years in advance. This foresight informs research and development activities to ensure Sandvine's solutions are prepared for changes as they become mainstream, rather than playing catch-up when it's already too late.

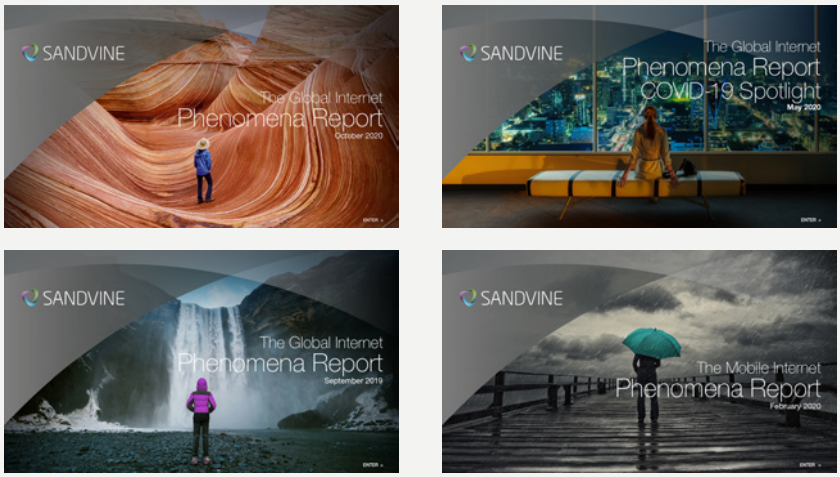
For instance, even by 2010 it was clear that while online video would be dominated by a few major global players, there would nevertheless be a long tail consisting of smaller content providers and regional services. With this forthcoming reality in mind, Sandvine invested heavily in research to build video analyzers that could automatically identify new services.

Similarly, the rise of encryption – while gaining mainstream attention only recently – was forecast many years ago. Again, Sandvine used this time to prepare and the result is a platform that can recognize the nature of encrypted traffic through behavioral analysis alone.

These two trends converged with the eruption of IPTV and streaming fraud. Again, Sandvine's Global Internet Phenomena research provided early warning; this warning was heeded and Sandvine quickly emerged as the leader in helping to identify fraudulent services.

Figure 4

Sandvine's renowned Global Internet Phenomena reports analyze and explain internet traffic trends, providing insights that extend beyond an operator's own network borders and direct experiences



ABOUT SANDVINE

Sandvine helps organizations run world-class networks with Active Network Intelligence, leveraging machine learning analytics and closed-loop automation to identify and adapt to network behavior in real-time. With Sandvine, organizations have the power of a highly automated platform from a single vendor that delivers a deep understanding of their network data to drive faster, better decisions. For more information, visit **sandvine.com** or follow Sandvine on Twitter at **@Sandvine.**

SANDVINE

EUROPE
Svärdfiskgatan 4
432 40 Varberg,
Halland
Sweden
T. +46 340.48 38 00

CANADA
408 Albert Street,
Waterloo,
Ontario N2L 3V3,
Canada
T. +1 519.880.2600

ASIA
RMZ Ecoworld,
Building-1, Ground Floor,
East Wing Devarabeesanahalli,
Bellandur, Outer Ring Road,
Bangalore 560103, India
T. +91 80677.43333