sandvine®

Intelligent Broadband Networks

2016 Global Internet Phenomena

SPOTLIGHT: ENCRYPTED INTERNET TRAFFIC

Introduction

In 2015, Sandvine revealed some of the most detailed data about the growing trend of Internet traffic encryption. This paper aims to build upon that release and use real network data to shine a spotlight on just how much Internet traffic is currently encrypted as well as provide a high-level overview of some of the current and emerging techniques used to provide such encryption.

Sandvine believes that encrypting traffic to protect subscriber privacy is a good thing, and while there has been a lot of talk on how information on the Internet can be hidden or guarded, there is still a great deal of misunderstanding on the topic.

Two related concepts related to protecting the privacy of subscriber Internet traffic are:

- Encryption: encoding information such that it can only be read by an authorized party
- Obfuscation: hiding or disguising information to prevent detection

Either or both of these general techniques might be used by any particular application, and the lines sometimes blur. For instance, consider:

- Encryption to preserve content privacy: Some applications encrypt user data and content as a privacy measure, but don't attempt to evade detection and management. As a significant example, YouTube traffic is currently carried via HTTPS (or QUIC) which prevents third-parties from inspecting video title information and revealing detailed individual viewing habits. The encryption method can be proprietary or based on a standard. Additionally, encryption is frequently employed as part of a digital rights management (DRM) strategy, in an attempt to control access to and reproduction of information.[1]
- Encryption as a means of obfuscation: Some applications such as Tor apply encryption in an attempt to evade detection and the application of traffic management.

It is important for subscribers, operators, and politicians to understand that encryption does not mean something is undetectable or unidentifiable, it just means that the content is private. Because most encrypted traffic relies on accepted standards (e.g., IPSEC, TLS), it is generally easy to detect the application being used, although capabilities do vary by solution vendor.[2]

1. Encryption both helps and hinders, Digital Rights Management (DRM) depending upon who is applying the encryption. Encrypted peer-to-peer filesharing defeats DRM strategies that inspect data for identifiers that correspond to licensed content, and laws/regulations that require CSPs to filter unlicensed content are ignorant of this technical reality. However, when the encryption is part of the DRM strategy itself it prevents unauthorized access and copying.
2. For instance, the "server_name" field is visible in TLS, but exists at a variable offset. As a consequence, solutions with hardware fast-paths for TLS traffic will struggle, as they typically lack the flexibility to handle non-fixed offsets.

Common Encryption and Obfuscation Techniques

Due to the prevalence of encryption measures, subscribers and operators concerned about privacy should understand the differences between the various technologies commonly in place today in order to understand how their traffic is being encrypted or obfuscated.

- **SSL/TLS (Secure Sockets Layer and Transport Layer Security)**[3]: These are cryptographic protocols designed to provide secure communications, and are used extensively in applications where security is required (e.g., banking, exchanging private data, etc.). HTTP Secure (HTTPS) adds the security capabilities of SSL/TLS to HTTP communications. HTTPS is technically not a protocol by itself, as it is simply HTTP on top of SSL/TLS. Historically, getting and maintaining an SSL certificate was cost-prohibitive for all but the larger web properties, but the Electronic Frontier Foundation's (EFF) HTTPS Everywhere initiative[4] looks to change that and will lead to wider adoption and use of SSL.

- **Virtual Private Networks (VPNs)**: A VPN extends a private network across a public network, and includes security elements such as authentication and encryption (typically using SSL/TLS). VPNs are used extensively by enterprises to provide connectivity between sites and remote workers, but private VPN services are available specifically to provide encryption for Internet content and are being increasingly used by subscribers to access content not available in their region.

- **Data Compression Proxies**: These are proxy services that provide data compression to users (with the intent of reducing bandwidth usage), and have the same practical impact to traffic classification as encryption. For instance, Google has a data compression proxy for Chrome[5], which can use a variety of protocols depending on what's available.

- **Proxy Applications**: These are applications (eg. Opera Mini) that can be installed on (typically mobile) devices to provide users with privacy and more efficient data usage. Similarly, add-ons/plug-ins or configurations can instruct web browsers to use certain optimization protocols or techniques. For instance, Windows Phone has a Browser Optimization Service[6] that compresses data.

State of Encryption Adoption in 2016

Sandvine worked with several operators globally in January 2016 with the goal to measure the amount of encrypted traffic on fixed and mobile networks across the globe.

One common misinterpretation from previous Global Internet Phenomena Reports made by some readers was that an application listed as "SSL" encapsulated the entirety of encrypted traffic on the Internet. The reality is that, in Sandvine's reports the data presented are direct outputs of Sandvine's reporting products, and that the "SSL" category listing typically represents the very long tail (thousands of websites or applications, representing a fraction of Internet traffic each) of SSL traffic that Sandvine has consciously chosen not to separately classify (for example, your bank's encrypted traffic, secure payment systems, etc.) as individual applications.

At the same time, leading encrypted applications such as Facebook, YouTube, or Twitter, have used SSL for many years and have been reported accurately and separately under their own proper names because of Sandvine's decision to assign an application name to them in our reports. To arrive at an accurate total, the traffic related to the "SSL" category and these major applications must be added together.

3. An overview is available at: http://en.wikipedia.org/wiki/Transport_Layer_Security; the IETF RFC can be found here: http://tools.ietf.org/html/rfc5246
4. You can learn more about this initiative here: https://www.eff.org/https-everywhere
5. Learn more about this service here: https://developer.chrome.com/multidevice/data-compression
6. More information is available at https://dev.windowsphone.com/en-US/OEM/docs/Driver_Components/Browser_Optimization_Service

North America, Fixed

Figure 1 below shows a breakdown of our research from a North American fixed access network and how 37.5% of total traffic is now encrypted. This marks a moderate increase over the 29.1% traffic that was observed in the same network observed in April of last year.

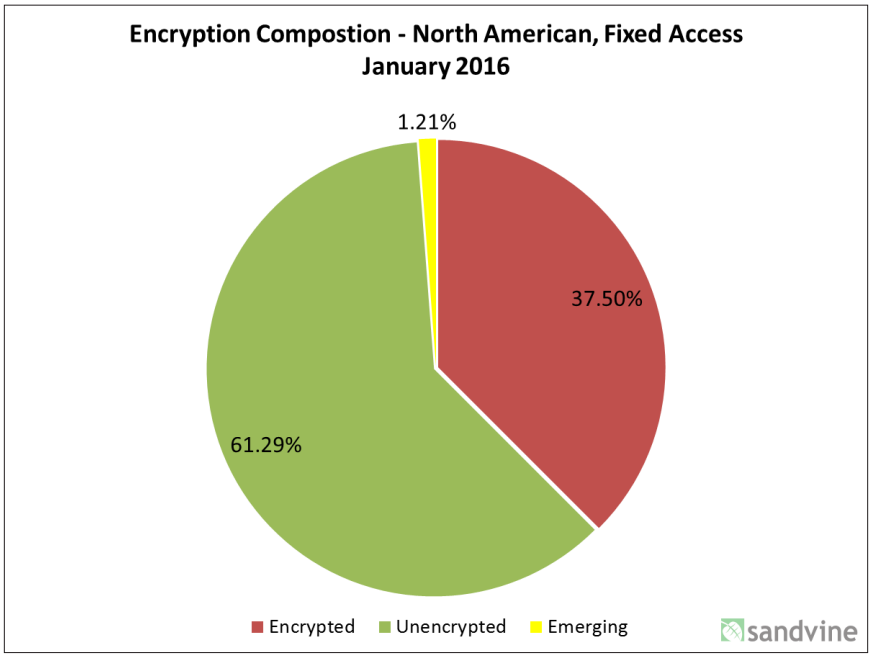


Figure 1 - Encryption Composition - North America, Fixed Access – January 2016

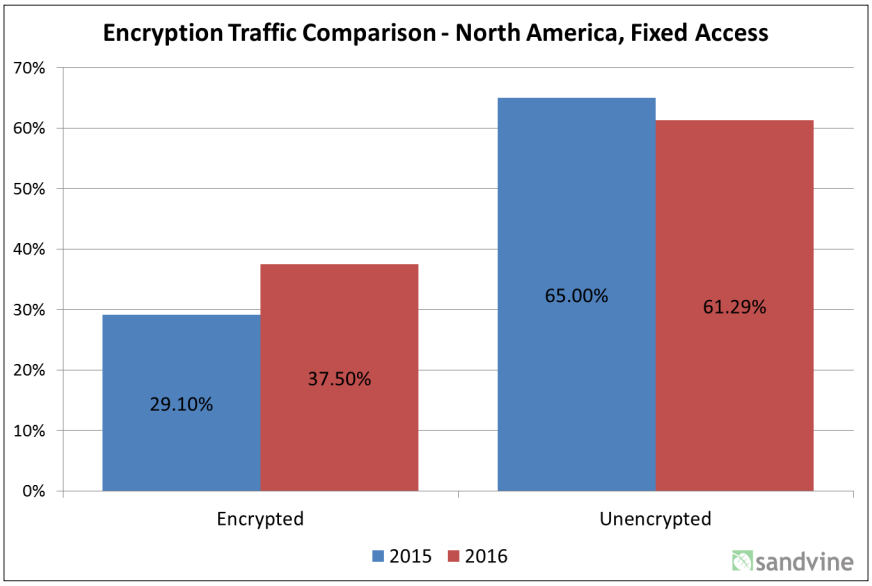


Figure 2 - Encryption Comparison - North America, Fixed Access

The biggest driver of encryption's growth has been Netflix's transition towards HTTPS delivery. In our most recent Global Internet Phenomena Report released in December, we observed that Netflix accounted for 37.1% of peak downstream traffic. In April 2015, Netflix's CEO revealed plans over the next year to move to using HTTPS with the aim to "protect member privacy, particularly when the network is insecure, such as public Wi-Fi, and it helps protect members from eavesdropping by their ISP or employer, who may want to record our members' viewing for other reasons." [7]

In Table 1 below, the top 10 applications by aggregate traffic during peak period from Sandvine's most recent Global Internet Phenomena report are listed. Of the top 10 applications, only four YouTube, BitTorrent, Facebook, and SSL – Other) are encrypted.

	Aggregate	
Rank	Application	Share
1	Netflix	34.70%
2	YouTube	16.88%
3	HTTP - OTHER	6.05%
4	BitTorrent	4.35%
5	Amazon Instant Video	2.94%
6	iTunes	2.62%
7	Facebook	2.51%
8	Hulu	2.48%
9	MPEG - OTHER	2.16%
10	SSL - OTHER	1.99%
		72.89%

sandvine

Table 1 - Top 10 Applications – Aggregate Traffic – Peak Period - North America, Fixed Access

For the purpose of this report, for Sandvine to consider an application as encrypted, over 80% of its traffic must be encrypted. Using these criteria, we consider YouTube to be encrypted and Netflix not yet encrypted, although they are in transition.

To date, approximately 9% of Netflix traffic is encrypted, the majority of which is from browser-based streaming. At the time of data collection, the majority of other client devices used by Netflix, had not yet transitioned to HTTPS, although Sandvine expects that to change in this year. Once completed, over two-thirds of North American fixed access traffic will be encrypted.

For comparison, YouTube now has 98% of its traffic encrypted compared to 83.6% last year, making it the largest provider of encrypted video on the Internet.

The 1.2% of "emerging traffic" referred to in the chart above (and elsewhere in this report) is traffic yet to be classified by Sandvine, so a determination of whether it is encrypted was not possible.

7. More details from Netflix here: http://www.theverge.com/2015/4/15/8422889/netflix-https-coming-within-one-year

North America, Mobile

Unlike fixed access networks, which are dominated by unencrypted Netflix traffic, mobile networks contain a far higher share of encrypted traffic.

Figure 3 below shows that the majority of traffic on North American mobile networks is now encrypted, with only one-third remaining unencrypted.

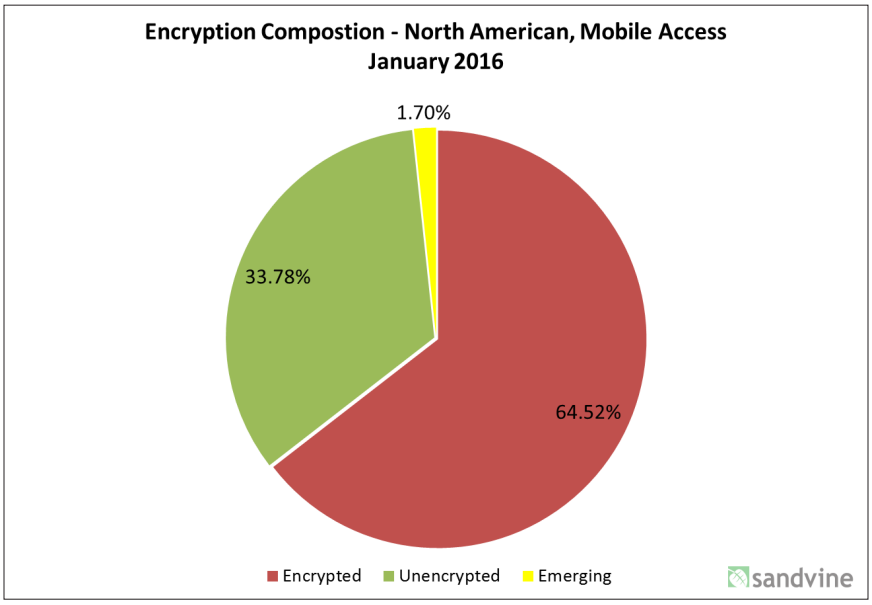


Figure 3 - Encryption Composition - North America, Mobile Access – January 2016

Table 2 below shows the list of the top ten aggregate traffic generating applications during peak period from our most recent Global Internet Phenomena report. The data shows that five of the top six applications (YouTube, Facebook, SSL – Other, Google Cloud and Snapchat) are encrypted.

	Aggregate	
Rank	Application	Share
1	YouTube	19.59%
2	Facebook	16.35%
3	HTTP - OTHER	10.69%
4	SSL - OTHER	7.81%
5	Google Cloud	4.33%
6	Snapchat	4.11%
7	MPEG - OTHER	4.09%
8	Pandora Radio	3.95%
9	Instagram	3.79%
10	Netflix	3.22%
		74.00%

sandvine

Table 2 - Top 10 Applications – Aggregate Traffic – Peak Period - North America, Mobile Access

As with fixed networks, where many Real-Time entertainment applications have yet to transition towards encryption, the mobile network also sees streaming applications like Pandora Radio as hold outs as well. It will be interesting to see which of these streaming applications will be the first to to encrypt their traffic now that the two largest sources of traffic (Netflix and YouTube) have committed to using HTTPS.

One standout application from the top 10 list is Instagram. Owned by Facebook (who has long been an evangelist of encrypted traffic) only a small portion of Instagram traffic is actually served over HTTPS. Based on our measurements, only 8.1% of Instagram traffic is encrypted, which means the bulk of their traffic, pictures and videos shared on the service are unencrypted. In July 2014, Instagram publicly stated that Instagram Direct, the private messaging portion of the service does use HTTPS, and that they are "doing the technical work that is necessary to add HTTPS protection across the remaining parts of the Instagram app, while still ensuring stability and performance".[8]

Global Data

In addition to an in-depth analysis of North American networks, Sandvine also obtained data from select operators in Latin America and Europe as well, with both datasets having similar encryption share on both fixed and mobile networks.

Fixed, Access

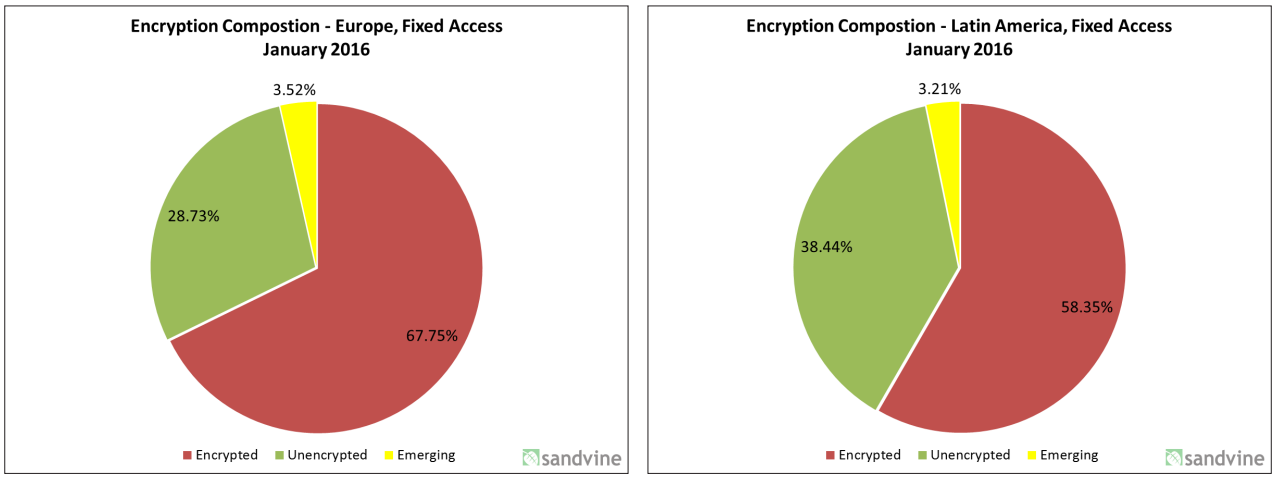


Figure 4 - Encryption Composition – Europe & Latin America, Fixed Access – January 2016

On fixed networks in both regions, approximately two-thirds of all traffic is encrypted. These figures differ significantly from the North American fixed access figures because at the time of data collection, Netflix, the largest source of traffic in North America, had comparatively low bandwidth share in Europe and Latin America. This lower Netflix share, combined with higher shares of encrypted traffic from YouTube and BitTorrent explain the vast difference between fixed access figures. While North American fixed access networks currently have the lowest share of encrypted traffic of any around the world, we expect North America to equal and even surpass other regions once Netflix completes their HTTPS transition.

Mobile, Access

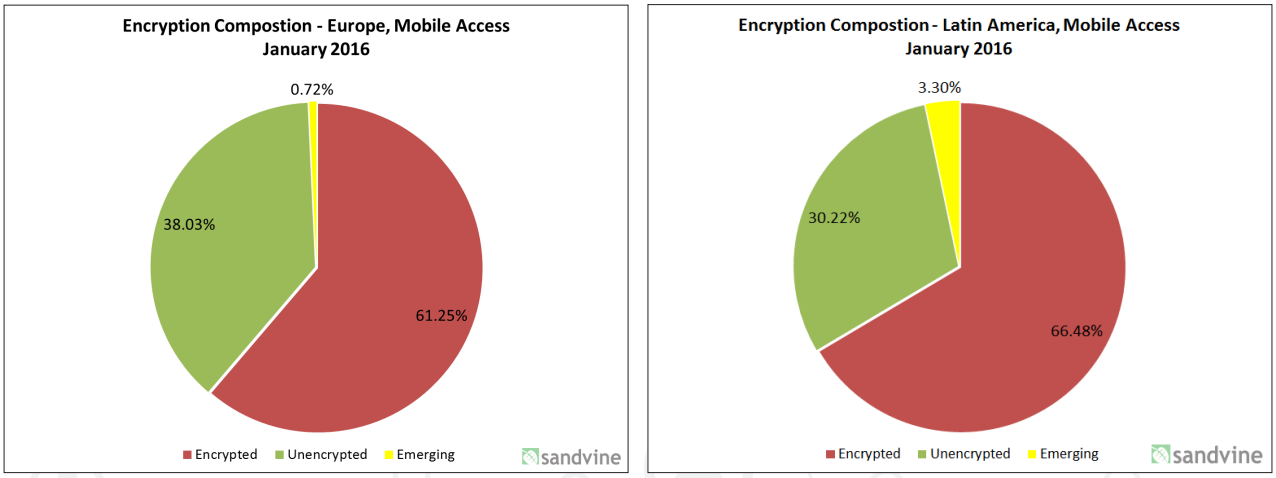


Figure 5 - Encryption Composition – Europe & Latin America, Mobile Access – January 2016

8. Instagram planning to complete HTTPS rollout 'soon,' after developer exposes iOS flaw - http://thenextweb.com/facebook/2014/07/29/instagram-planning-complete-https-rollout-soon-developer-exposes-ios-vulnerability/

Like on fixed networks, the data from mobile networks in Latin America and Europe are broadly inline with each other and slightly below North American levels with 60% of total traffic being encrypted.

As noted in the North American mobile section of this report, a big chunk of the unencrypted data comes from streaming audio and video applications that have yet to make the transition. The remaining traffic comes from generic HTTP browsing which Sandvine expects to gradually transition to SSL in the future.

Respecting Content Privacy

While some of Sandvine's competitors have attempted to delve into traffic content (i.e., going beyond simply identifying "this is Netflix HD" and saying "this is Netflix and is episode 4 of House of Cards"), Sandvine has never had such an interest. This is an important distinction, for at least two reasons:

1. Sandvine respects users' privacy: We seek to identify traffic, its attributes, and its measured characteristics, because those are needed to achieve operator use cases including business intelligence, the creation of innovative subscriber service creation tiers, traffic optimization during times of network congestion, and network security. Revealing the precise content of a traffic flow does not advance any of these use cases.

2. Content intelligence solutions are incredibly vulnerable to proprietary encryption[9]: Most content providers have an incentive to protect details of their service usage, and will actively take measures to prevent third-parties from extracting and revealing this information. For operators, this reality means that a content intelligence solution bought today can be rendered mostly useless tomorrow. In contrast, most content companies have no incentive to prevent mere identification of their traffic – they just don't want anyone investigating more deeply into the exact content itself.

Encryption's Impacts on Service Creation and Subscriber Experience

Even with the majority of Internet traffic now encrypted, it should not negatively impact a network operator's ability to offer innovative service plans for subscribers, providing it has deployed a network policy control solution capable of accurately identifying encrypted traffic.

Below are two examples of current in-market plans powered by Sandvine at Econet Wireless in Zimbabwe, and Smart Communications in the Philippines. In both cases, Sandvine helps enable subscribers to purchase access to encrypted applications such as Facebook, Twitter, or Gmail. Sandvine is able to provide accurate traffic classification so that the service works as expected for the subscriber, and the operator experiences no revenue leakage.

One area where encryption does provide challenges is in the area of video, especially now that the two largest sources of Internet video (YouTube and Netflix) are now committed to encrypting their traffic.

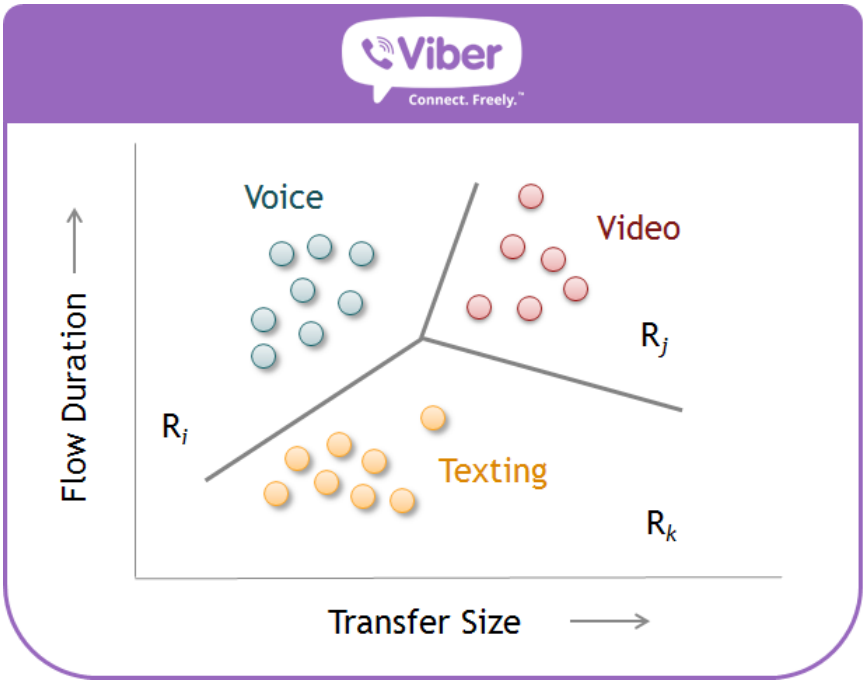
The challenges don't lie in the area of classification and billing. For example, in the Smart Philippines example above, Sandvine enables YouTube-based service plans to subscribers. However, as observed in a Infonetics report, "the practice of third-party content encryption has created challenges in managing third-party video content." [10]

9. This is also why no network-based DRM enforcement system has ever been successful.
10. Infonetics Service Provider Deep Packet Inspection Products Report (subscription required) - http://www.infonetics.com/pr/2014/1H14-Service-Provider-DPI-Market-Highlights.asp

The Infonetics report highlights how encrypted video makes video optimization solutions ineffective since they can't transcode encrypted video, and it also may present challenges to video caching solutions.

Additionally, as more applications begin to diversify the type of content they provide consumers, operators and network equipment vendors will be faced with challenges in providing the same level of traffic classification they had than before they were encrypted. As an example of the diversifying functionality, you can look towards messaging apps like WhatsApp and Viber who over the years have transitioned from offering one type of communication (text or voice), to being all-in-one voice, text, and video messaging services.

For operators looking to understand how these messaging apps are being used by subscribers, they will have to rely on new advanced traffic classification methods.



One such advanced traffic classification method would use machine learning in an attempt to classify traffic based on a set of known attributes of subjects. Using Viber as an example above, by examining both the transfer size and flow duration, it is possible to determine if a subscriber is using the service to send a text message, or participating in a voice or video call. While the subscriber's content would always be encrypted, protecting the subscriber's privacy, the information on what type of communication is being used could still be determined in order for in order for an operator to understand the adoption and use of these services by their subscribers.

Another advanced traffic classification that could be applied to encrypted traffic is heuristics. As an example, YouTube provides various levels of video quality that can either be selected by the end user manually or automatically determined by the client device. By measuring the bitrate of an encrypted YouTube video, it is possible to classify whether the video is SD, HD, or UHD.

In order to accurately classify and sub-classify traffic using a technique such as machine learning or heuristic, service providers will need to deploy solutions that offer both a high amount of both CPU and RAM to the handle the computing complexity needed to provide real-time classification.

Conclusions and Projections

Based on spot checks with our existing customers around the world, and data presented in this report, Sandvine predicts that by the end of 2016, global Intenet traffic will be more than 70% encrypted, with some networks surpassing the 80% threshold.

Main drivers of this will be major video providers, like Netflix, completing the transition towards encrypted traffic. Additionally, programs such as the Electronic Frontier Foundation's "Let's Encrypt" program which launched in 2015[11], will help drive the adoption of encryption by helping developers avoid the complexity, bureaucracy, and cost of using HTTPS by providing a free, automated, and open certificate authority that anyone can use.

Moving forward, the tools used by operators to create intelligent broadband networks will focus on traffic application and volume characteristics, but not content, since encryption will render most content optimization and caching techniques ineffective. Additionally, accurately classifying and sub-classifying traffic will continue to and increasingly rely on systems that are capable of resource demanding machine learning and heuristics methods because more simplistic methods used by systems with little computing power will only work on the small amount of unencrypted traffic remaining on the Internet.

Additional Information

This Global Internet Phenomena Spotlight aims to provide an introductory overview on encryption through the sharing of real network data and a high-level overview of the terms and technology used.

For those interested in learning more on this topic Sandvine has also published a Technology Showcase entitled "Traffic Classification: Identifying and Measuring Internet Traffic" which provides additional real-world examples on how to identify encrypted and obfuscated Internet traffic with the flexibility and versatility of SandScript; Sandvine's unique policy definition language.

This Technology Showcase is ideal for operators, or interested subscribers who want to:

1. **Gain a Technical Foundation of Traffic Classification**
   Understand how Sandvine's Traffic Classification technology addresses additional technical considerations including, identifying encapsulated and/or tunneled traffic, overcoming routing asymmetry, achieving stateful awareness and correlating across flows and session.

2. **Gain Insight into Powerful Traffic Identification Techniques**
   Learn more about Sandvine's three main traffic identification techniques that empower our unique SandScript policy definition language: signatures, trackers and analyzers.

3. **Learn How to Address Encryption, Obfuscation and Proxies**
   Understand how Sandvine's SandScript policy definition language allows CSPs to combine and apply a wide range of traffic identification techniques to effectively identify encrypted and obfuscated traffic.

The Technology Showcase can be downloaded here: https://www.sandvine.com/trends/encryption.html

11. "Let's Encrypt" website: https://letsencrypt.org/

Headquarters
Sandvine Incorporated ULC
Waterloo, Ontario Canada
Phone: +1 519 880 2600
Email: sales@sandvine.com

European Offices
Sandvine Limited
Swindon, UK
Phone: +44 (0) 1256 698021
Email: sales@sandvine.co.uk

sandvine®
Intelligent Broadband Networks