



# sandvine<sup>®</sup>

Intelligent Broadband Networks

## Global Internet Phenomena Spotlight Encrypted Internet Traffic



# Global Internet Phenomena Spotlight: Encrypted Internet Traffic

## Introduction

There is a growing trend on the Internet, with more and more applications beginning to encrypt their traffic in order to protect a subscriber's content from prying eyes.

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<sup>1</sup>.
- **Encryption as a means of obfuscation:** Some applications apply encryption in an attempt to evade detection and the application of traffic management. For instance, BitTorrent clients have added increasing levels of encryption over the years<sup>2</sup>.

It is important for subscribers and operators 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.<sup>3</sup>

This paper aims to 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.

---

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. An overview is available at [http://en.wikipedia.org/wiki/BitTorrent\\_protocol\\_encryption](http://en.wikipedia.org/wiki/BitTorrent_protocol_encryption)

3. 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.

## Current State of Encryption Adoption

Sandvine worked with a North American fixed access network in April 2015 with the goal to demonstrate just how much traffic is encrypted currently.

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 in those reports 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 SSL-based 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.

Figure 1 below shows a breakdown of our research and how 29.1% of total downstream traffic is now encrypted, with 65% remaining unencrypted.

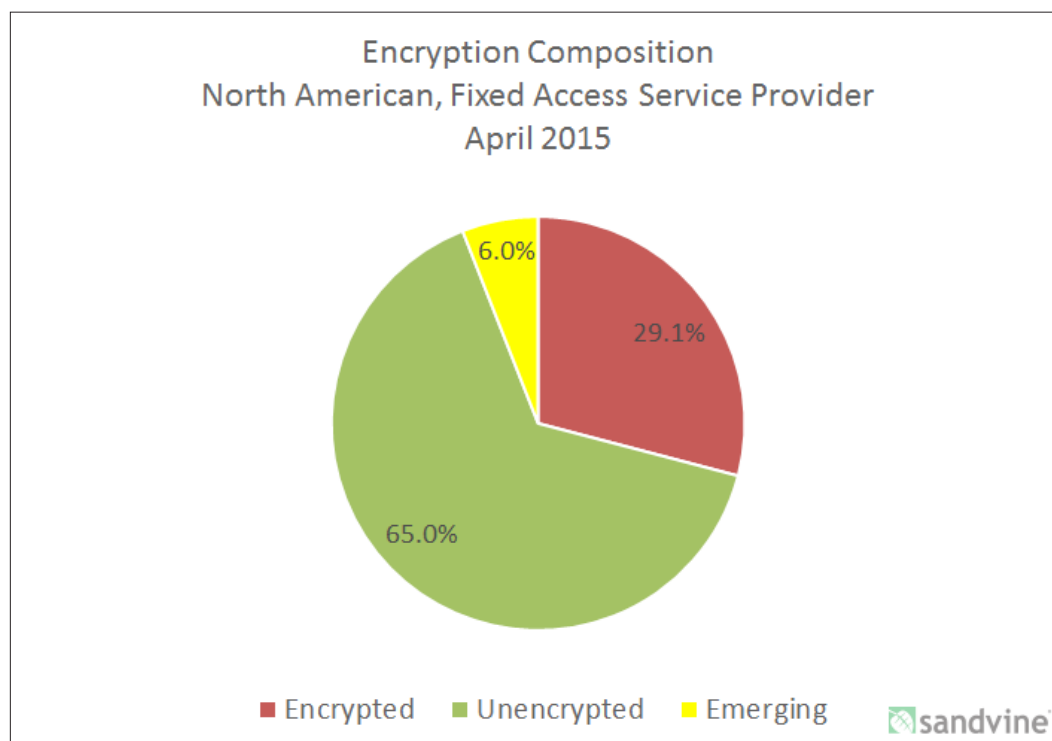


Figure 1 - Encryption Composition - North America, Fixed Access - April 2015

The 6% emerging traffic refers to traffic yet to be classified by Sandvine, so a determination of whether it is encrypted or unencrypted was not possible. Sandvine’s philosophy towards traffic identification is to focus on accuracy first, and completeness second. That is, we will not sacrifice accuracy (i.e., we will not accept false positives) to reduce the amount of traffic that is unrecognized. Simply put, false positives are unacceptable, as they can have a disastrous impact across a range of use cases for both subscriber (e.g., billed incorrectly) and operator (e.g., harm to reputation from mis-managing or incorrectly charging for certain traffic). That said, we routinely see traffic recognition rates upward of 95%<sup>4</sup>.

As for the leading sources of encrypted traffic, YouTube, which is typically the second largest source of total traffic on North American fixed access networks, is also the largest source of encrypted traffic. YouTube as a whole accounts for 13.69% of total downstream traffic on the one network in this study, with the encrypted portion (HTTPS) of YouTube accounting for 11.45% of traffic. Based on observations in other markets, the exact ratio of encrypted to unencrypted YouTube traffic actually varies by country, but on this particular North American network 83.6% of YouTube traffic is

4. Our Global Internet Phenomena program lets us carefully monitor the rates of unrecognized traffic at our deployments around the world, and these rates are consistently less than 5% of traffic volume. The precise level in any network will vary, of course, based on local characteristics, management policies, and the frequency with which customers update their Loadable Traffic Identification Packs. That said, this average is based upon a mix of some customers - some of whom update frequently, and some of whom don't - and it's not uncommon to see a recognition rate of more than 97%.



encrypted.

The second largest source of encryption comes from BitTorrent traffic, which prior to YouTube making the transition towards encrypting their traffic, served as the largest source of encrypted traffic on the web. It should also be noted that due to the fact Google encrypts virtually all of their services, the company as a whole generates a significant amount of encrypted traffic. For the purposes of reporting and billing in our products, Sandvine breaks these applications (Gmail, Search, Maps, etc.) out individually, where they each represent a small portion of traffic.

Daily Downstream Traffic Share - Encrypted Applications	
Application	Traffic Share
YouTube	11.45%
BitTorrent	7.20%
Facebook	2.31%




Table 1 - Daily Downstream Traffic Share - Encrypted Applications

Table 2 below shows a list of the top three unencrypted traffic generating applications. Netflix, the long-time dominant source of all traffic, is the unsurprising leader at 35.7% of daily traffic. Netflix technically has an encrypted traffic payload today due to the use of DRM, but for the purposes of this paper, the use of encryption during transport was used to define encrypted traffic.

After that, Apple's iTunes is a distant second on this fixed network, accounting for 2.67% of total downstream traffic. A mobile network could potentially produce different results depending on the mix of Android and iOS devices. Interestingly, Google Play, Google's marketplace, and arguably iTunes' biggest competitor, actually encrypts all of its traffic, preventing third-parties from detecting the music, movies or apps being downloaded by subscribers.

As highlighted above, a small portion of YouTube traffic remains unencrypted, but this small portion is still significant in terms of volume. Unencrypted YouTube accounts for 2.24% of total downstream traffic on this network. For the purposes of this study we did not delve into exploring why a small portion of YouTube remains unencrypted. It could simply be a decision by YouTube to slowly migrate to full encryption, or could be tied to something like individual client devices which may not support HTTPS.

Daily Downstream Traffic Share - Unencrypted Applications	
Application	Traffic Share
Netflix	35.65%
iTunes	2.67%
YouTube	2.24%




Table 2 - Daily Downstream Traffic Share - Encrypted Applications

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."<sup>5</sup>

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

So what does this mean for traffic composition in North America, where Netflix is the largest source of traffic? On the network examined for this study, Netflix accounted for 35.7% of total daily downstream traffic. Ignoring the potential for Netflix traffic share to grow or decline, Figure 2 shows that in 2016, almost two-thirds of traffic on North American fixed access networks will be encrypted, and the reality is it will likely be over two-thirds as additional applications make the switch to HTTPS via programs such as the Electronic Frontier Foundation’s “Let’s Encrypt” program.

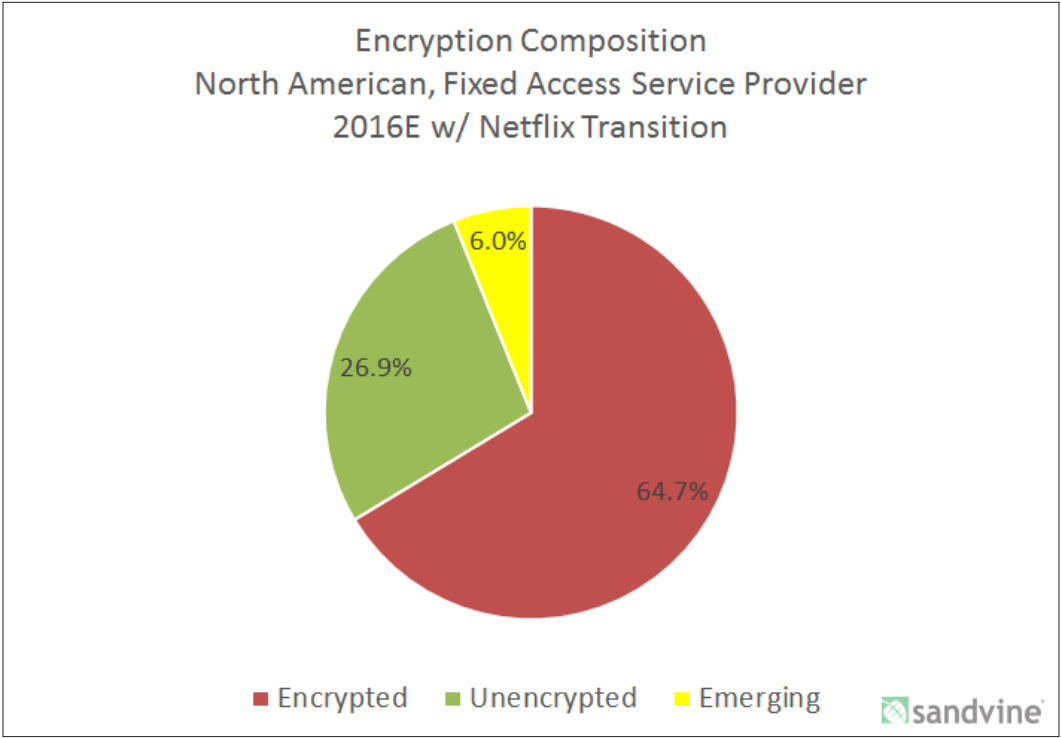


Figure 2 - Encryption Composition - North America, Fixed Access - 2016 Estimate

## Common Encryption and Obfuscation Techniques

Due to the prevalence of encryption and obfuscation 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)**<sup>6</sup>: 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<sup>7</sup> 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<sup>8</sup>, 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/plugin-ins or configurations can instruct web browsers to use certain optimization protocols or techniques. For instance, Windows Phone has a Browser Optimization Service<sup>9</sup> that compresses data.

---

6. An overview is available at: [http://en.wikipedia.org/wiki/Transport\\_Layer\\_Security](http://en.wikipedia.org/wiki/Transport_Layer_Security); the IETF RFC can be found here: <http://tools.ietf.org/html/rfc5246>

7. You can learn more about this initiative here: <https://www.eff.org/https-everywhere>

8. Learn more about this service here: <https://developer.chrome.com/multidevice/data-compression>

9. More information is available at [https://dev.windowsphone.com/en-US/OEM/docs/Driver\\_Components/Browser\\_Optimization\\_Service](https://dev.windowsphone.com/en-US/OEM/docs/Driver_Components/Browser_Optimization_Service)

## Emerging Encryption and Obfuscation Techniques

While the above four encryption and obfuscation methods are relatively common on the Internet today, there are a number of emerging technologies aimed at improving web quality that also provide encryption or obfuscation.

- **SPDY**<sup>10</sup>: “Speedy” is an open networking protocol, developed primarily by Google, that modifies the way HTTP requests and responses are sent in the data path. In draft form, it was used by several large players before eventually being merged with the IETF HTTP/2 standard. The stated goals of SPDY are to reduce web page load latency and improve web security, and SPDY achieves these objectives via compression, multiplexing, and prioritization of HTTP traffic. Practically, the result of these measures is the same as if encryption were the intent, as content is obscured. Both the client (e.g., web browser) and web server need to support SPDY in order for it to be used; generally, SPDY is supported by the major web browsers and many major web services (e.g., Google, Twitter, Facebook, and WordPress). SPDY receives a great deal of attention, but in reality will only ever account for a small percentage of Internet traffic as it is now standardized as HTTP/2.<sup>11</sup>
- **QUIC (Quick UDP Internet Connections)**<sup>12</sup>: “Quick” is an experimental transport protocol developed by Google that has the same practical impact as TLS encryption. QUIC is designed to provide security protection equivalent to TLS/SSL, with the added benefits of reduced latency. Like SPDY, QUIC requires both client and server support; at the time of this writing, QUIC is supported by Google Chrome and Google servers, and in an April 2015 blog post Google claimed that roughly half of all requests from Chrome to Google servers are served over QUIC.<sup>13</sup>
- **HTTP/2 (HTTP 2.0)**<sup>14</sup>: This next planned version of HTTP is based on SPDY (the draft of HTTP 2.0 published in November 2012 is based on a straight copy of SPDY). HTTP/2 is largely an effort to standardize SPDY implementations and to ensure backwards compatibility with HTTP 1.1 (the most recent standard, in use since 1999). That said, there are other differences between SPDY and HTTP/2; the main difference is that HTTP/2 allows multiplexing to happen at different hosts at the same time, to expedite downloading multiple web pages or content from multiple sources.

## Respecting Content Privacy

While many of Sandvine’s competitors have attempted to delve into traffic content (i.e., going beyond simply identifying “this is Netflix HD” and measuring its characteristics to instead say “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<sup>15</sup>: 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<sup>16</sup>. 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.

10. General information about SPDY, including specific client and server support, is available at <http://en.wikipedia.org/wiki/SPDY>; protocol documentation can be found at <http://www.chromium.org/spdy>

11. Wondering why? SPDY primarily exists to eliminate TCP round-trip time latency on mobile networks; in reality, this latency is only a problem in 3G environments. Further, SPDY only applies to browser traffic, and even then only to non-SSL traffic. So, ultimately, SPDY will be used for non-SSL browser-based traffic on 3G networks.

12. More information can be found here: <http://en.wikipedia.org/wiki/QUIC>

13. More information can be found here: <http://blog.chromium.org/2015/04/a-quic-update-on-googles-experimental.html>

14. The latest IETF draft (at time of this writing) is available here: <https://tools.ietf.org/html/draft-ietf-httpbis-http2-14>; the Wikipedia summary can be found here: [http://en.wikipedia.org/wiki/HTTP\\_2.0](http://en.wikipedia.org/wiki/HTTP_2.0)

15. This is also why no network-based DRM enforcement system has ever been successful.

16. This has already happened to at least one vendor’s product, impacting all CSPs who had purchased it. Interestingly, the vendor’s website continues to promote capabilities that are no longer available, even though the lack of availability is publicly acknowledged and received widespread coverage.

## Encryption's Impacts on Service Creation and Subscriber Experience

Even with the majority of the Internet traffic soon to be encrypted, it should not negatively impact a subscriber's ability to be offered innovative service plans as long as their service provider 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 recent Infonetics report, "the practice of third-party content encryption has created challenges in managing third-party video content."<sup>17</sup>

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.

For operators with these solutions in place (and the subscribers who are currently experiencing the benefits of them), the trade-off for increased privacy that encrypted video provides could result in a reduced quality of experience from some video sites.

17. Infonetics Service Provider Deep Packet Inspection Products Report (subscription required) - <http://www.infonetics.com/pr/2014/1H14-Service-Provider-DPI-Market-Highlights.asp>



## 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 2015, the majority of the world's markets will see Internet traffic that is more than 50% encrypted, and that by the end of 2016 65-70% of traffic will be encrypted in most markets.

While this report focuses on North American fixed access traffic where Netflix dominates, in many other markets YouTube (an already encrypted application) is the leading source of traffic in fixed and mobile, often accounting for over 20-30% of traffic by itself.

Additionally, programs such as the Electronic Frontier Foundation's "Let's Encrypt" program due to launch in mid-2015<sup>18</sup>, 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 offer intelligent broadband will focus on the application and volume, but not the content, since content optimization and caching techniques currently in use will no longer be effective.

## 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:

- **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.
- **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.
- **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>

---

18. "Let's Encrypt" website: <https://letsencrypt.org/>