Järvinen, Juha; Marttinen, Aleksi; Luoma, Marko; Peuhkuri, Markus; Manner, Jukka

Protecting Individuals by Hiding Flow Information on Last-mile Links

Published in:
2018 7th International Conference on Computer and Communication Engineering (ICCCE)

DOI:
10.1109/ICCCE.2018.8539299

Published: 19/09/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Järvinen, J., Marttinen, A., Luoma, M., Peuhkuri, M., & Manner, J. (2018). Protecting Individuals by Hiding Flow Information on Last-mile Links. In *2018 7th International Conference on Computer and Communication Engineering (ICCCE)* (pp. 141-146). IEEE. https://doi.org/10.1109/ICCCE.2018.8539299

2018 7th International Conference on Computer and Communication Engineering (ICCCE)

Protecting Individuals by Hiding Flow Information on Last-mile Links

Juha Järvinen, Aleksi Marttinen, Marko Luoma, Markus Peuhkuri and Jukka Manner
Department of Communications and Networking, Aalto University
Postal Address: PO Box 15600, FI-00076 AALTO
Email: {Juha.Tapio.Jarvinen, Aleksi.Marttinen, Marko.Luoma, Markus.Peuhkuri, Jukka.Manner}@aalto.fi

***Abstract*—Last-mile links connect a core network to the end-user's communication infrastructure. A single user's traffic can flow alone at a link, and thus an adversary can eavesdrop on the communication at the link, deducing critical information. Even if the traffic is encrypted, an aggressive enemy may perform statistical analysis based on the detected traffic. The results of this analysis may yield information in the communication that is private, characteristics of the communications, such as destination node or utilized networking protocols. With this information individual's privacy can be damaged.**

In this paper we introduce a system that enhances remarkably user privacy on last mile links by making it more difficult for an adversary to perform statistical analysis. The system is based on one of the onion routing realizations Tor that is improved with the Traffic Flow Confidentiality (TFC) technique. Either a network operator or a service operator can deliver this service to customers. Moreover, we have measured the performance of our system that conceals flow information on last-mile link and beyond enhancing individual's privacy.

Keywords—network security, privacy, last-mile networking

I. INTRODUCTION

Hiding traffic flows in core networks has been traditionally easier than in last-mile connections because of the nature of that specific traffic: there are more routes to choose between source and destination nodes, and there is a large amount of traffic – a spy must work more to find a target. In addition, routers are connected with fiber that is more difficult to eavesdrop and operators' core devices are located on unaccessible facilities. However, a last-mile connection is usually single-handed to a Internet Service Provider's (ISP) network with copper cables that can be located in easily accessible spaces. Conventionally the only way to hide traffic has been by encrypting the content inside a packet from a source to a destination similarly as in Virtual Private Networks (VPN). However, all other important information still remains accessible for statistical analysis: packet length, bit rate, frequency of packets, *src-dst* pair of VPN connections, or even information about whether a user is at home or not.

Typically only large companies have multi-homed network accesses but in typical last-mile cases there are only two connections to the same operator, or two different operators' networks that is almost the same situation to adversaries as a single connection: one connection can be in standby mode awaiting traffic from the user, and all the traffic uses just the other link.

On a last-mile connection it is easy to capture all the traffic of targeted people. At home, a single user is using the network at a specific time - not usually everyone in the family. Capturing the traffic is the easiest to do on last-mile networks – cables are located at easily accessible locations and they are often made of copper that is tapped relatively easily.

In this paper our contribution is to present a system, that enhances not only privacy issues on last-mile links but also in larger area. The system is based on an existent implementation of the Onion routing concept: Tor. The system can be exploited by both network and service operators.

This paper is structured as follows. In Sections II and III we introduce the current situation with regard to last-mile networking, and offer some real options for resolving the privacy vulnerability issues. In Section IV we present Last-mile prIVacy Enhancing System (LIVES) to improve the privacy of the last-mile links. Section V introduces the future work on the issue. In Section VI we present the measurements we conducted. Finally, Section VII concludes the article.

II. CURRENT SITUATION AND RELATED WORK

Currently there are no specific solutions that are intended for use with only last-mile links. All used data hiding mechanisms are focused on hiding the content of packets with VPNs or similar mechanisms. In other words, content is encrypted when using a VPN or, for example, application level security mechanism. Thus content cannot be read without first eavesdropping and decrypting the content, but other information, for example, source and destination IP addresses still remain accessible.

Encrypting the traffic helps with hiding the content but we still have important flow information left visible: It is easy to guess some of the used protocols in VPNs, e.g., VoIP and DNS, and also for example detect web browsing [1], [2]. Several protocols or services have a known fingerprint [3]. For example, if an adversary observes packets of fixed size around 200 bytes with intervals of tens of millisecond travelling encrypted in the last-mile connection, the adversary can conclude that someone is using a VoIP service. If some VPNs are leaking DNS queries systemically, the estimation of used services, protocols and connection

end-points becomes much easier. In addition, even the customer's end-point would be safe, the other end point may suffer some security problems. Finally, even though finding a target from a large data mass is difficult, it is not necessarily impossible. For example, at night time a small VPN provider's gateway may serve only a few customers.

Furthermore, if a service is using standard ports, even though they are encrypted with SSL/TLS and a user is using them over a plain Internet connection without any VPN, the purpose of the content can be deduced.

The smaller the targeted network is, the greater the reliability of the information an eavesdropper may get out of a connection: even a single user can be identified. The same also applies to VPN: the fewer number of users using the same VPN connection, the greater reliability of information is achieved. If a company has a VPN between sites and HQ, an adversary can identify only a few parameters, such as the direction of the connection etc. [4]

There has been on the market a robust next-generation method for network encryption and the concealment of data content: a Tor network that is an implementation of Onion routing [5]. First it was used by information security people, then by political dissidents. More recently, however, after information was leaked telling of massive eavesdropping networks and services by U.S. National Security Agency (NSA), even regular citizens have started to use the Tor network when newspapers began giving out instructions on how to use it. However, usually people are using it on a last-mile connection in proxy mode that creates only a VPN connection to the closest Tor node, leaving traffic parameters unhidden.

The Traffic Flow Confidentiality (TFC) technique for IPsec use was presented in RFC 4304 [6]. The TFC method is carried out with padding. As a limitation, the maximum padding of a packet length is 255 bytes. A more advanced TFC method for IPsec is presented by Kiraly et al. [7] with, for example, fragmentation and dummy packet generation. So far some TFC mechanisms are used in some commercial VPN solutions, e.g by Cisco. IPsec is a point-to-point tunnel on Layer 3, meaning it needs an L3 end-point usually carried out as a service. Carlén, for example, has performed TFC-related measurements in small networks where all the links use IPsec with TFC [4]. Respectively on Layer 2 there is a method to use TFC techniques. It is built into the IEEE 802.1AE Media Access Control (MAC) Security standard [8]. Since the mechanism is operating on L2, it needs actions from a network operator.

Both of the methods mentioned above are able to hide sensitive data on a single link; however, data can be visible immediately after the next node if we do not know what kind of network there is after the next hop. Of course, here is no such problem if the whole network is equipped with the above methods. Usually though, this is not a realistic assumption, even in military networks. On last-mile connections we can identify problems in flow confidentiality:

1) We know what the other end's IP address is and we can solve its location. If it is impossible to capture traffic at the customer site, it may be easier to eavesdrop traffic on the gateway site. Network flows from site A to Internet service are not distributed for different paths.

2) Usually a VPN is established from a computer at site A to a gateway at site B - meaning that only one user uses single VPN connection. Thus, it is easier to deduce what a user is doing, for example, calling a VoIP call. In addition, even just information indicating that a particular person is present at the site may be more than enough information to jeopardize a person's privacy, even the safety of their person.

3) There is not enough traffic and too few different user's traffic on the link simultaneously [4].

4) Possible DNS leaks on a VPN connection expose where a user is connecting from. Now even if an adversary is not able to hack a VPN connection on either of the last-mile or gateway site, he can capture traffic on a service site following, for example, usernames on VoIP calls. Connection between a gateway and a service site can be normal plain non-encrypted Internet traffic. The identifiers such as VoIP usernames can be found from social media, email addresses and location.

5) Currently there is no proper mechanism for concealing network and service operators that could be used after the VPN end-point. Respectively on the Onion Routing solution (Tor), the concealment of information is utilized in the rest of the network but on last-mile links there is still room for improving privacy issues.

6) Improving only one link's privacy is not necessarily enough. All the effort invested for privacy issues on a single link can be pointless, and sensitive data can be exposed after the next hop if we do not know the network topology or cannot rely on it.

Apthorpe et al. have earlier discussed a similar research problem [9]. Their point of view was primarily a user device in Wi-Fi networks. They proposed a solution to secure against eavesdropping between a user device and a VPN Exit point. Our work differs from their work since our system secures communication also after the VPN Exit point and we concentrate on wired last hop.

III. Alternatives to fix the problem

We have discovered some technologies that could help to protect last-mile connections and conceal, safeguard the privacy of user information. The mechanisms are the multipath technique, the Traffic Flow Confidentiality technique and Onion routing. In this section we introduce these techniques, analyze them and present the characteristics of each.

A. Multipath technique

In the multipath technique a single data packet can choose its route from origin to destination from several options if the device is connected outside using more than a single connection. During the previous years, different load balancing techniques have been studied more intensively, and now some large companies such as Apple have a support for multipath TCP in their products [10].

Currently home users typically have only one link to the Wide Area Network (WAN), and economically increasing the number of links obviously increases the costs. Even more exotic alternatives like satellite connections cost more. The benefits of these WAN technologies (in terms of

safeguarding the privacy of users) is that, although they are usually narrow band, they are more difficult to capture and thus analyze.

B. *Traffic Flow Confidentiality technique*

With multipath, the user information can be hidden if physically different paths are used: for example an ADSL and satellite link. With this method we can conceal, for example, the duration of a call, if the adversary is not capable of capturing data at each link. However, other attack possibilities still remain for an attacker, for instance, information about destination address.

We can state that multipath does not help in concealing user's information at all – we can briefly narrow down the attacker's attack vector, but this is not yet sufficient protection.

Traffic Flow Confidentiality (TFC) was already mentioned in 1989 in the ISO/OSI Security Architecture (ISO 7498-2) [11]. Kiraly et al. [12] have defined TFC as follows: "TFC mechanisms devised to alter or mask the statistical characteristics of the traffic patterns are necessary". Mechanisms for masking are, for example, padding, fragmentation, dummy packet generation, packet combining and adding packet forwarding delay. Masking mechanisms can be used in randomized, fixed or some other scheduling manner. There are some implementations on TFC by using IPsec [13] because of its built-in padding (255 bytes).

With TFC, it is possible to complicate an adversary's task in performing the statistical analysis. For example, by utilizing packet combination and extra delay mechanisms, it is nearly impossible to detect VoIP traffic by using statistical methods. Of course, delay must meet the protocol's maximum limitations. A principle of how the TFC mechanism affects the traffic pattern is shown in Figure 1. On the left of the figure three packet flows are sent on line (for clarity, they are not active simultaneously). On the right of the figure, the packet trains are masked: padded, combined, delayed: it is impossible anymore to say how original data is distributed among the packets on the path. However, the source and destination addresses are still visible.

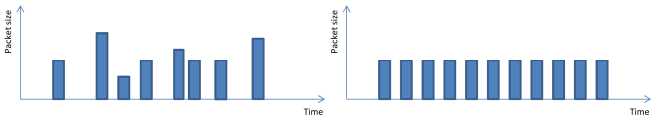

Fig. 1. An example of TFC mechanism affecting traffic pattern.

Two major problems remain:

1) The other end-point can be known. TFC cannot distribute traffic itself. It is similar to a point-to-point VPN where the content cannot be seen, but, in addition, statistical analysis is of no use to a potential adversary.

2) Secondly, what should the other TFC capable end-point be? Is it the nearest router/ Digital Subscriber Line Access Multiplexer (DSLAM) in the ISP network or is it a VPN gateway. At least, it should be a point where is a lot of other users' traffic too. If we trust an operator's network, where there is also a great deal of traffic from other, and the concern is the privacy of the last-mile connection only, then the end-point can be the nearest router/DSLAM.

As we can see, with TFC we manage to hide communications from the adversary's statistical analysis of the traffic, but information on a source-destination pair can be worked out by an adversary. Between larger sites this is not a problem but between smaller sites it is, since then there is knowledge of a relationship between a user and a site (for example, a company) — a spy can try to get information about a particular user from the company's side.

In addition, extra traffic is generated to the network depending on the selected TFC mechanism, thus bandwidth efficiency decreases. If TFC mechanisms can be used at the link level for all traffic, bandwidth efficiency is better than when used separately for each source-destination pair.

C. *Onion Routing*

Onion Routing (OR) enables anonymous communication in the Internet using an overlay network. Packets are encrypted between several network nodes (Onion routers). In the middle of the path there is no information who is a sender and a receiver, instead only the knowledge from where a packet is coming and where it is going next. Onion routing can act as concealment cloud, that may be a part of the path between the sender and receiver and conceal the real path of the packets. Goldschlag et al. presented the idea of OR in 1996 [5].

Tor is one of the most well known realization of OR. Tor has been discussed in the everyday media since NSA leaks by Edward Snowden because various sources and authorities say Tor is not a guaranteed way to act as fully anonymous user of Internet – messages in Tor can be decrypted [14]. The security of Tor can however be improved by using a Tor exit node also on a user's computer for encrypting and hiding source-destination pairs from the user's operator. Nonetheless, in doing so, the traffic of other Tor users may utilize the resources of the user's own network.

The Tor principle is to transfer different data streams over Internet using different paths. The user's traffic may return to the normal Internet anywhere in the world. Electronic Frontier Foundation (EFF) presents the working principle of Tor in [15]. There are two options to use Tor: 1) send traffic to a Tor cloud via a proxy software or 2) operate as an exit node. In the first option, the first hop will be encrypted similarly to SSL/TLS. This is for the use of a single user. In the second option the user runs an exit node and all data immediately access the Tor cloud. DNS leaks are reported also on Tor that reveals the service that you are using [16]. Tor is said to be low-speed communication system [17].

Tor can be thought as a perfect mechanism to conceal a user's generated traffic especially if the user is hosting a Tor relay node where all the traffic goes immediately to the Tor network. There remains however one problem on last-mile connections. Occasionally, the last-mile link only includes the secured Tor-traffic, meaning that the traffic of other Tor users does not traverse via user's own Tor node. In this case the situation is similar to a conventional VPN-connection - the user's content is encrypted but an adversary can identify what the user is doing, e.g., by observing Inter-Arrival Time (IAT). Combining this information with DNS leaks [16] we are again aware of the other end-point of traffic. Tor performs TFC- typed mangling by creating 512-byte long cells. However, cells of an equal size at the application layer

do not necessarily produce equal-sized packets at the network layer depending on the traffic conditions. [17]

IV. Last-mile privacy Enhancing System

As noticed above, currently there is no perfect solution for protecting comprehensively last-mile connections. Tor is quite near but there could be a better protection in cases when the traffic is not transmitted frequently on a link. In addition, if the Exit Node is not used in the home network, each user in the home network should install Tor Client software to ensure that the approach works. If we run the Exit Node at home, we can encounter problems as mentioned in the previous section. In this paper, we present a concept - Last-mile prIVacy Enhancing System (LIVES) - where we pick all the good features of all earlier mentioned methods for concealing user traffic, trying to eliminate their respective disadvantages. Next we go through the system architecture.

A. System Architecture

The presented system is based on the Tor system improved with TFC features. Basically we have a VPN connection to a Tor network. We have modified some parts of the normal Tor node, and also added new features, but they still have the ability to use and contact normal and ordinary Tor nodes.

This same idea can be offered for customers by both a network operator (NO) or a service operator (SO). Our system trusts the Tor infrastructure fully although there are articles describing how to break anonymity in a Tor network, e.g., by Wagener et al. [18]. Also U.S authorities have already managed to break either encryption of Tor or a node in some cases like in ”Silk Road” [19]. – execution has not been published. Tor is chosen here since it is currently the most developed open-source onion routing implementation.

Each customer’s ADSL modem or similar – the Customer Side Node (CSN) – is acting as a ”personal Tor exit” node. The node is an OpenVPN client (or similar) that takes all the user’s traffic and sends it in a VPN tube to a VPN gateway that is also a Tor Exit Node (TEN). The connection between them is not a part of a real Tor network. This is done since some operators have forbidden the existence of Tor exit nodes in their network due to the extra traffic, and this is one way to hide a connection to a Tor network especially in an SO environment. Of course, there is no need for this kind of hiding in NO mode.

The system architecture of our system is illustrated in Figure 2. It describes both the NO and SO cases. In the first one, VPN with continuous TFC is between CSN_A and TEN_1 (red line). In an SO environment VPN is between CSN_C and TEN_2. TENs are connected to the Tor network.

However, since there is not conventionally several traffic flows between the CSN and TEN caused by other Tor users, this traffic has to be protected somehow. It is done by mangling traffic by TFC methods to hide, for example, the amount of users, the direction of traffic. In addition, the queued traffic is continuous all the time regardless of whether the link is used.

This system architecture enables speeding up the rate of last-mile traffic rate if the source and destination are connected to the same TEN. Then the traffic is not destined to the Tor network at all. This speeds up their end-to-end connection since the Tor network is normally slower than a normal Internet connection by and ISP. Of course, the transmission rate between the CSN and TEN remains high in at least an NO network since the TFC parameters can be chosen to fulfill the link. In the case of an SO network this connection speed is usually lower, since network operator cannot waste the network resources; the use of TFC parameters have to be considered carefully, so that they are suitable for different types of traffic.

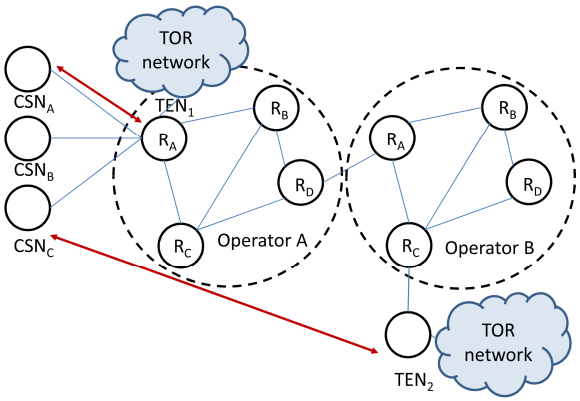

Fig. 2. Illustration of the system in both environments.

If the customer’s traffic is identity-free, or if an operator is hosting content services for their customer, this kind of traffic could also bypass the Tor network. This scenario is illustrated in Figure 3. Identity information is, for example, the user’s IP addresses or clear text account information. For a bypass in a TEN device there are Network Address Translation (NAT) mechanisms to conceal the real IP address of the customer. Each stream for this kind of fast lane has to be carefully considered, since each piece of identity information can expose the customer.

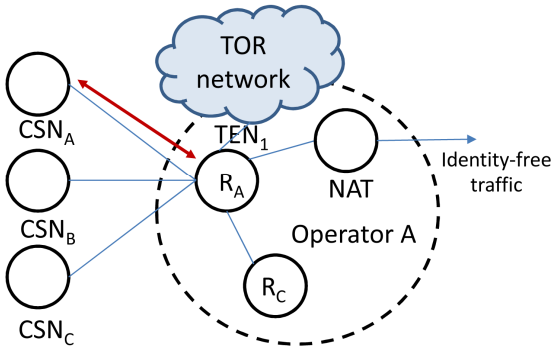

Fig. 3. A topology to speed up the speed of a connection speed.

In our system there is no problem if an operator is either a network or service operator since there are Layer 3 (IP) connections between the CSN and TEN. In this way our system can be used in many environments and no system modifications, for instance, with regards to topology, are required.

B. Other Options

Another option for topology is to use a Tor Exit Node as the CSN device. On the downside, this node also acts as a public exit node for Tor traffic that could slow down the connection speed for the user. Thus, quite seldomly private people would like to run an exit node in their own network even though they want to use the Tor network. The second

negative feature is that the random path or Tor traffic is shorter since also the second node has to be always the same, because of using TFC to mangle the packets.

The positive side of this option, however, is that it utilizes the entire path of the Tor cryptographical mechanism.

In our system, cryptographical mechanism is changing in a TEN device. This can be seen as a positive side when operators are not against our traffic in SO mode, since it does not look like Tor traffic. But it can also be seen as a negative side, because the layered security model is not end-to-end.

If the Tor network would already begin in the CSN device, traffic pairs that are in the same network should still use the Tor network. Now in our system these traffic pairs are able to use a faster network between the CSN-TEN-CSN.

If the system is used only in the NO environment, OpenVPN can be replaced by IEEE 802.1AE technology on L2 if the CSN and TEN are located on the same IP network.

C. *Difference between the NO and the SO systems*

The only difference between the NO and SO concepts is who is offering the service to the customer; is it the user's own network operator who provides the Internet connection, or is it a special operator offering it as a service. Since even in the NO mode securing takes a place on the IP Layer between the CSN and TEN (and not by using, for example, on L2 with IEEE 802.1AE), there is no need to change technique when moving to SO mode.

If a user's NO does not offer such service, a provider can act as an SO hosting this service. As we can see the situation is the same as above, now the path between the CSN and TEN is longer, locating also in another NO network. A customer can have a special device acting as a CSN or software is installed on the customer's real home router where some open-source operating system is installed, for example, DD-WRT or OpenWrt. The service operator is connected on a similar way from its other side to the normal Tor network. This is like VPN companies that provide VPN services. The major difference is that VPN encryption is not removed at the gateway, with traffic continuing as encrypted.

For the SO mode, some estimation of path parameters is needed between the CSN and TEN: the minimum and average values for path capacity, the maximum and average values for delay and delay variance, and also the Maximum Transfer Unit (MTU). In the NO mode similar measurements for path parameters are needed too, but they do not change usually much since it is the user's "private" line to the operator's edge (TEN) and the line can be fulfilled as the customer wants.

With regards to drawbacks for customers, they are unable to host any services at own location. In Tor, only reply packets find at their route back in the Tor network.

D. *Discussion*

In our opinion, this system should work effectively, and it improves last-mile privacy remarkably.

Improving only one link's privacy issue is not necessarily enough. All the effort concentrated on privacy issues on a single link can be pointless, and sensitive data can be exposed after a next hop if we do not know what kind of network there is. Since there is a need to build a larger security layer, not only for a certain single link, if the whole network is fulfilled with the same links. We have to remember that the system is for securing the privacy of users, then comes throughput. However, if the system offers very poor performance it is not intended for wider audiences in industry and the public sector; rather it is more suited to enthusiasts and for people who really want privacy.

Since the system is not only for a single link, it gives a positive sign to a customer from the perspective of privacy. Even though the TEN device is a possible weak link, since it changes the utilized crypto mechanism after that there is no end-to-end encryption like in pure Tor, user can utilize the end-to-end applications level security such as TLS.

As can be seen from this concept, we do not only improve the privacy of last-mile links, but we have also created an inexpensive way for operator to offer a secure overlay network and a safer connection to the Tor network. The effectiveness of bandwidth usage decreases at least minorly due to TFC in the NO mode: the last-mile can be full of traffic (maximum packet size of line and minimum IAT) continuously since the path is only in the customer's use. In other words, the customer can use TFC features on a line rate.

In the SO concept, the situation is completely reversed: there is a need for an all-way constant-bitrate path between the customer and the gateway. The other end-point cannot have economically many subscribers, that have a 20 Mbit/s connection, meaning that only about 45 subscribers can connect to a gateway if link speed is 1 Gbit/s. This also means a connection between a customer and a gateway should be 20 Mbits/s. If a connection to the gateway would be 2 Mbits/s, the gateway could have even 450 subscribers that sounds economically much feasible. However, will customers be satisfied? Will those who have been used to enjoying data speeds of 20 Mbits/s be content with speeds of only 2 Mbits/s? For the customers who respect the security issues this is easy, but not so for everyone. In the name of concealing customer behaviour, the improved security should be acceptable even though the transmission speed decreases.

End-user devices do not need any modifications in operation system etc. The CSN takes care of all the traffic concealment functionalities. Using a wireless network connecting to the CSN may endanger the whole concept since an eavesdropper can be so close to the CSN that it is possible to capture packets from the Wi-Fi connection.

All the traffic has to go through a VPN gateway all the time: there is no option that delay sensitive protocols or less important traffic could pass the tube. Passing would increase the bandwidth in the SO mode; however, in this scenario the customer reveals a lot of information of its network again, making the whole system almost unworkable. For instance, the information about the customer's entity could not be no longer concealed. Identity-free traffic can pass the Tor network in the TEN device to speed up the connection speed, if needed. Continuous VPN with TFC traffic stream makes it impossible to detect, for example, how many users use the network, whether somebody is at home, or what the network is being used for.

V. Validation

We measured our concept (Fig. 3) in a real network in two parts. First, we modeled TFC (padding / fragmentation) with TCP traffic, utilizing different packet sizes (46, 512, 1024 and 1500 bytes). The measurements were performed with the Iperf testing tool and each test lasted 10 seconds. In total, 4393 tests were run for each packet size. The link capacity was 100/10 Mbit/s and it was located from a CSN (located in a real home network) to a TEN (located in an operator's network). The results of the measurements are presented in Fig. 4. We can see that with all the packet sizes the achieved bandwidth was almost the same, i.e. the variation was very small.

Second, we tested the data speed from a TEN (located in an operator's network) to the Internet by measuring the average download speed of a 100 MB file from the same location 5103 times. After each download, the utilized Tor path changed. The NAT bypass mechanism was not tested. As a result, we can see that in the measurements the average download speed from the Internet to the home network was 6.625 Mbit/s. This also means that the link speed of 10 Mbit/s is enough between the CSN and TEN. Thus, roughly 100 users can use the network if the TEN has a 1 Gbit/s interface. Finally as a conclusion of the measurement, we can say that when Tor is used, bandwidth may significantly vary. Thus, a provider must implement a mechanism to ensure that the Tor path meets sufficient quality for bandwidth, delay, packet loss, etc. This can be verified by measuring Key Performance Indicators (KPIs).

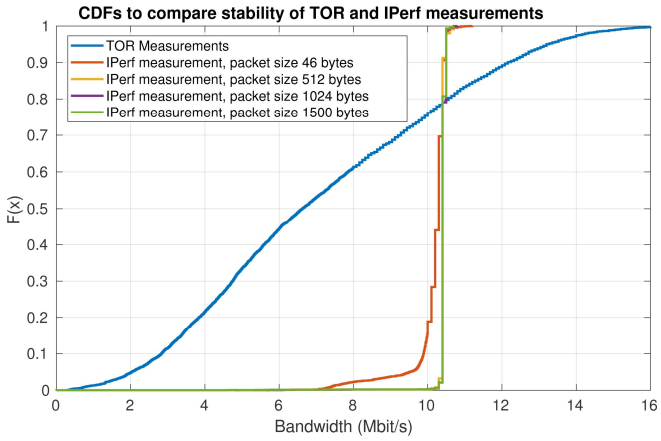

Fig. 4. Cumulative Distribution Function of bandwidth measurements.

VI. Conclusion

We have found that the current situation concerning privacy issues on last-mile connections is far from ideal. Traffic can be encrypted in VPNs but the other statistical information of traffic flows remain. In this paper we present a method - Last-mile prIVacy Enhancing System (LIVES) - to improve the level of privacy and conceal flow information on last-mile connection: we add the TFC method for a VPN connection and then connect this directly to the Tor network. This same architecture works for both network operators and for service operators. As the Tor measurement shows, bandwidth may significantly vary. Thus, a provider has to use some mechanism to ensure that the Tor path has sufficient capacity and quality.

A weakness of the presented system is the TEN device, since it changes the used crypto mechanism. However, the user can still improve the security of the connection for instance against man-in-the-middle attacks by using application level security. Our presented concept is based on Tor, however it can be adapted to work with another Onion routing implementation.

References

[1] B. Xu, M. Chen, C. Xing, and G. Zhang, "A network traffic identification method based on finite state machine," in 5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09., Sept 2009.

[2] J. V. Gomes, P. R. M. Inacio, M. Pereira, M. M. Freire, and P. P. Monteiro, "Identification of peer-to-peer voip sessions using entropy and codec properties," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 10, Oct. 2013.

[3] M. Crotti, M. Dusi, F. Gringoli, and L. Salgarelli, "Traffic classification through simple statistical fingerprinting," SIGCOMM Comput. Commun. Rev., vol. 37, no. 1, Jan. 2007.

[4] P. Carlén, "Traffic flow confidentiality mechanisms and their impact on traffic," in Military Communications and Information Systems Confer- ence (MCC), 2013, Oct 2013.

[5] D. M. Goldschlag, M. G. Reed, and P. F. Syverson, "Hiding routing information," in Proceedings of the First International Workshop on Information Hiding. London, UK, UK: Springer-Verlag, 1996.

[6] S. Kent, "IP Encapsulating Security Payload (ESP)," RFC 4303 (Proposed Standard), Internet Engineering Task Force, Dec. 2005.

[7] C. Kiraly, S. Teofili, G. Bianchi, R. Lo Cigno, M. Nardelli, and E. Delz- eri, Traffic Flow Confidentiality in IPsec: Protocol and Implementation, ser. IFIP International Federation for Information Processing. Springer Boston, 2008, vol. 262.

[8] A. Romanowi, "IEEE 802.1AE - media access control (MAC) security," 2006.

[9] N. Apthorpe, D. Reisman, and N. Feamster, "Closing the blinds: Four strategies for protecting smart home privacy from network observers," CoRR, vol. abs/1705.06809, 2017.

[10] A. Mäkelä, S. Siikavirta, and J. Manner, "Comparison of load-balancing approaches for multipath connectivity," Comput. Netw., vol. 56, no. 8, May 2012.

[11] ISO, "Information processing systems – open systems interconnection – basic reference model – part 2: Security architecture, iso 7498-2," 1989.

[12] G. Bianchi, C. Kiraly, R. L. Cigno, and S. Teofili, "Traffic flow confiden- tiality in ipsec: Protocol and implementation," in Pre-Proceedings Third IFIP / FIDIS Summer School,"The Future of Identity in the Information Society", August 2007.

[13] S. Frankel and S. Krishnan, "IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap," RFC 6071 (Informational), Internet Engineering Task Force, Feb. 2011.

[14] M. Suess. Breaking TOR Anonymity. [Online; accessed 13-June-2014]. Available: http://www.csnc.ch/misc/files/publications/the onion router v1.1.pdf

[15] Electronic Frontier Foundation, Tor and HTTPS. [Online; accessed 30-October-2017]. Available: https://www.eff.org/pages/tor-and-https

[16] Tor. Preventing Tor DNS Leaks Tor Bug Tracker. [Online; accessed 13-June-2017]. Available: https://trac.torproject.org/projects/tor/wiki/doc/Preventing_Tor_DNS_Leaks

[17] Z. Ling, J. Luo, W. Yu, and X. Fu, "Equal-sized cells mean equal-sized packets in tor?" in Proceedings of IEEE International Conference on Communications, ICC, 2011.

[18] C. Wagner, G. Wagener, R. State, T. Engel, and A. Dulaunoy, "Breaking tor anonymity with game theory and data mining," in 2010 4th Interna- tional Conference on Network and System Security (NSS), 2010.

[19] BBC News, FBI arrests Silk Road drugs site suspect. [Online; accessed 20-January-2018]. Available: https://www.bbc.com/news/technology-243737