A LaTeX Template for SIGCOMM 18

Paper # XXX, XXX pages

ABSTRACT

1 INTRODUCTION

The Transmission Control Protocol (TCP) [55] is one of the most critical protocols in today's Internet. A wide range of applications that require reliable delivery use it. During the last four decades, TCP evolved under the pressure of competing protocols. During the 1980s, software-based TCP implementations were considered too slow. Researchers proposed new transport protocols such as XTP [61] which could be implemented in hardware. TCP implementations got a considerable speed boost [16], and XTP did not succeed. However, the TCP speed boost and usage triggered the development of various important TCP extensions, including, timestamps and large windows [10] to scale to the gigabit link speed or Selective Acknowledgments [25].

During the late nineties, early 2000s, transport protocol researchers explored other alternatives to TCP . Two of these approaches were adopted and standardized within the IETF: DCCP [43] and SCTP [65]. We rarely use DCCP today. Despite its benefits (support for multihoming, better design, and extensibility), only a few niche applications use SCTP [12]. This limited deployment is probably due to two different factors. First, SCTP required changes to the applications to replace TCP . Second, operators have deployed middleboxes (NAT, firewalls) that often block packets that do not carry TCP or UDP [35].

SCTP initially supported multihoming by switching from one path to another. It was later extended to be able to use different paths continuously [40]. Multipath TCP [26, 57] brought similar multihoming capabilitiy to TCP , and included a coupled congestion control scheme [75], later brought to SCTP as well. This particular succession of events shows how different designs can collobarate to advance each others. Multipath TCP is now deployed, notably on smartphones [8]. Other recent TCP extensions include TCP Fast Open [15] or TCPCrypt [7].

In the mid-nineties, the Secure Socket Layer protocol was proposed to secure emerging e-commerce websites [22]. This protocol evolved in different versions of the Transport Layer Security (TLS) protocol, the most recent one being version 1.3 [58]. Many details of the TLS protocol have changed since the first version of SSL [44]. Nowadays, TLS is almost ubiquitous on web servers [34] thanks to the availability of various TLS implementations and automated certificate authorities [1]. Furthermore, many non-web applications also rely on TLS [2].

Transport protocols continued to evolve in parallel. QUIC started as a proprietary protocol used by Google to speed up web transfers [45, 60]. During the last years, it evolved into a complete transport protocol whose standardization is being finalized within the IETF [39]. QUIC combines the functions that are usually found in TCP , TLS , and HTTP/2. A key characteristic of QUIC is that it encrypts almost all the packets, including most of their headers. Although QUIC is essentially a new transport protocol, it does not run directly above IP in contrast with SCTP, TCP , or DCCP. QUIC runs above UDP. This choice is mainly motivated by the desire to avoid as much as possible middlebox interference. QUIC's clean architecture has attracted researchers who have already proposed various extensions to the protocol [17, 18, 37, 50, 54, 66, 72].

Does the finalization of version 1 of the QUIC specification mark the end of the TCP era and move all transport research on this new protocol? We do not think so. History tells us that TCP has evolved with competing transport protocols. QUIC is today's competitor, but there is still plenty of room to improve TCP .

In this paper, we take a step back. As QUIC benefits from a closer integration between the reliability and the security mechanisms, we reconsider the separation between TCP and TLS . TLS brings security features, but TLS 1.3 can do much more. Thanks to the TLS 1.3 messages and records' extensibility, TLS can provide a secondary channel that enables hosts to exchange more control information and structured data. Furthermore, since TLS records are encrypted, middleboxes cannot easily interfere with the data exchanged over this new channel.

We combine TCP and TLS in a protocol that we call **TCPLS** . We describe in Section 2 a first design for TCPLS with the goals of *(i)* solving extensibility issues in TCP . *(ii)* Exporting complex transport features to the application and *(iii)* drawing a path to make TCP /TLS a good challenger to QUIC with modern appications. Then, we discuss how TLS ' flexible record layer can be used to provide a new channel to exchange information between TCPLS implementations. The design presentation concludes with an overview of the API to interact with the application. Our second contribution is the ongoing implementation of a TCPLS prototype on Linux by extending `picotls`, a TLS 1.3 implementation. We use it in Section 3 to illustrate the benefits of TCPLS with a multihoming connection migration use case. Finally, we analyze in Section 4 some of the research questions that TCPLS opens.

2 TCPLS DESIGN

TCPLS offers a cross-layer interface to TLS and TCP with the motivation to do more than securing the transport layer. Merging the stacks benefits both protocols and the application using this new approach. First, TCP suffers from a lack of extensibility due to size restrictions in its header and due to potential middlebox interferences [35]. TCPLS aims to solve TCP 's extensibility issue in the long run by offering a secure control channel to exchange TCP options without suffering from middlebox interferences and size restrictions in TCP headers. Second, TLS does not have a clear view of the transport protocol, and offering one with TCPLS brings opportunity for performance improvement (e.g., avoiding records fragmentation by matching the record size to the congestion window), and for connection reliability (e.g., failover). Third, applications are becoming more complex, which appeals to exposing transport-level functionalities and letting them tune the underlying transport to their use case. Essentially, this last motivation discusses a novel manner to expose transport-level functionalities that are encrypted, authenticated, reliable, extensible and adapted to complex application-level requirements.

2.1 Overview

TCP separates control information and data by placing the control information in the packet header and the data in the payload. This separation worked well until middleboxes started to interfere with TCP [19, 35, 49]. On a fraction of Internet paths, including e.g., some enterprise and cellular networks, some middleboxes interfere by adding, removing, or changing TCP options [35, 73, 76] and, in some cases, also transparently terminating TCP connections. These middleboxes have slowed down the evolution of TCP in recent years. TCPLS also uses the packet header to exchange TCP control information, but it leverages TLS to create a second and secure control channel. In a nutshell, TCPLS leverages the extensibility of TLS 1.3 to place control information such as TCP options inside the TLS handshake messages and new TLS records. Since this information is encrypted and authenticated, the communicating hosts can exchange new control information without encountering middlebox interference. We describe several examples of these new types of control information in Section 2.2 and Section 3.2.

In our current prototype, a TCPLS session starts with a classic TCP handshake. Immediately after, the client sends the ClientHello TLS message. The server replies with a ServerHello message which can contain encrypted data but also encrypted control information. For example, a dual-stack server may advertise its IPv6 address in the encrypted ServerHello message when contacted over its IPv4 address. We highlight one of our roadmap features in Section 4 to enable a 0-RTT TCPLS which would enable TCP to catchup the QUIC design regarding fast connection establishment. We also describe in Section 3 how our current prototype uses this information to support connection migration, failover, and other features.

Once the TCPLS session has been established, TCPLS sends TLS records. Most of these records contain application data transmitted by the client or the server. The control channel between the client and the server enables TCPLS to support new features, such as streams. Indeed, applications such as HTTP/2 support multiple streams mapped to a single TCP connection. However, there are situations, e.g., to prevent head-of-line blocking, where different streams should be mapped over other underlying TCP connections. With TCPLS , the client and the server can establish different datastreams over a single TCPLS session. The data from all these streams is encrypted using TLS . Furthermore, thanks to the TCPLS API, the client and the server can map each data stream to an underlying TCP connection. Thus, a TCPLS session can be composed of one or more TCP connections similarly as a Multipath TCP connection gathers subflows. Note that, if the multipath mode is enabled, then a lost packet over one TCP connection may create HOL blocking since packets received on other streams may have to wait for the lost packet to be properly reordered and delivered to the application. In the case of the multipath mode not enabled, this problem would not happen, but the application has to be careful to map application objects per stream, and not to mix these objects among several streams since the ordering would not be guaranteed.

To support data from a given datastream to be exchanged over several TCP connections, TCPLS includes its sequence numbers. A client and server can also enable acknowledgments. Thanks to these TCPLS acknowledgments, a TCPLS session can react to the failure of the underlying TCP connection by reestablishing a new TCP connection to continue the transfer of data and replay the records that have been lost.

A TCP connection ends with the exchange of FIN or RST packets. However, some middleboxes force the termination of TCP connections by sending RST packets [24, 74]. TCPLS can preserve established connections by automatically restarting the underlying TCP connection upon reception of a spurious reset. TCPLS defines the connection termination at the stream level: closing the last stream attached to a TCP connection allows clients and servers to securely terminate the TCPLS session.

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39

Type	Version	Length	} Header
Option Type	User Timeout	TType	} Payload

Figure 1: A new type of TLS Record containing a TCP option.

2.2 The Secure Control Channel

TLS 1.3 [58] has been designed with careful consideration for potential extensions. It supports the EncryptedExtensions message sent by the server alongside the ServerHello. Any extension sent with the ServerHello message is encrypted with the handshake key, and is not part of the context used to derive the eventual application key.

A reasonable approach to designing extensibility mechanisms in today's Internet is to avoid leaking any information that could help an on-path attacker recognize specific users or applications. Indeed, censorship [13, 29, 51] can be easily implemented when protocol messages can be distinguished, and avoiding trivial opportunities to implement censorship should become the bare minimum in designing a new protocol. TCPLS 's control protocol considers those problems by avoiding unencrypted data within the ClientHello.

In our design, the client indicates its willingness to use TCPLS with a transport parameter in the ClientHello. Upon reception of this parameter, the server can opportunistically send lightweight TCPLS data and TCP options as EncryptedExtensions. If the client does not support some extension, it echoes back an alert with the value of the option it does not recognize, but the connection continues.

The server or the client can also send TCPLS control messages after the handshake. These control messages take advantage of the TLS 1.3 content-type extensibility feature to avoid middlebox interference. Indeed, in TLS 1.3, the Record Protocol ensures that any new message appears as an APPDATA message type while the true content type (TType) is stored at the end of the encrypted payload. As an example, Figure 1 shows the TCPLS control message structure that carries the TCP User Timeout [28] option. *TType* is the true type of this record (TCP_OPTION), while its Type is set to *APPDATA*.

2.3 Datastreams and TCP Connections

In TCPLS , each stream has its own cryptographic context. They use the same key but derive the blockcipher IV such that nonce-misuse cannot happen while the record sequence number within each stream starts at 0. Only one application-level key is used for N streams, for each direction. The reason behind this design choice is to avoid security degradation with the usage of multiple keys (by a factor k with k keys) [14].

Having a separate cryptographic context means that TCPLS can do concurrent encryption and decryption between streams while maintaining decryption correctness and security, and potentially also use this capability to process streams over multiple cores. Finally, if we have multiple streams over the same TCP connection, TCPLS does not explicitely know which received data belongs to which stream. To obtain this information, we either require to modify the associated information within TLS records to add a stream id (these associated data are not encrypted but the AEAD cipher authenticates them). This choice means potential middlebox interference, which we chose to avoid. The other option is to leverage the AEAD cipher to check the authentication tag of the incoming record until we find the stream that properly verifies the tag). This operation is lightweight: it does not require full decryption of the record because TLS 1.3 uses AEAD ciphers doing Encrypt then MAC (and MAC then Decrypt), and looking for the right stream needs to be performed once each time the application writes to another stream over the same TCP connection.

Note that, security-wise, each failed decryption is considered as a forgery attempt. However, we have large limits on the confidentiality and integrity with all AEAD ciphers [31, 46] before a successful iorgery may be considered as a non-negligeable probability.

TCPLS enables the client or the server to associate new TCP connections to an existing TCPLS connection. This is similar to what Mutipath TCP does [26, 57], but with some differences. First, Multipath TCP supports only one bytestream. Second, TCPLS does not suffer from the same security limitations as Multipath TCP. To secure the attachment of additional subflows, Multipath TCP hosts exchange keys in plaintext during the handshake [26, 27]. These keys are then used later to authenticate the attachment of subflows to a connection. An attacker that has observed the initial handshake can attach any subflow to an existing Multipath TCP connection [3].

TCPLS securely solves this "connection join" problem. For example, consider a client connecting to a dual-stack server. Figure 2 depicts the TLS messages exchanged. The client starts with a ClientHello. This includes the TCPLS extension to negotiate TCPLS . The server replies with a ServerHello containing several important and encrypted control information α. First, the server announces its IPv4 and IPv6 addresses. Second, it associates one connection identifier. This identifier uniquely identifies the connection on the server. Third, the server provides a list of cookies that enable the client to attach additional TCP connections to the TCPLS connection. To attach a new connection, e.g., using the server's IPv6 address, the client opens a TCP connection and sends a ClientHello message containing the connection identifier

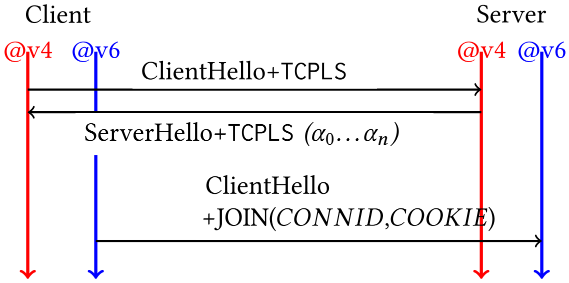

Figure 2: TCPLS supports the attachment of additional TCP connections to a TCPLS connection. Each α_i is encrypted with the handshake key.

(*CONNID*) and one of the cookies supplied (*COOKIE*) by the server.

The Connection identifier allows the server to attach the new TCP connection to the right TCPLS session, assuming the received cookie is valid. The Connection identifier and the cookie play that same role as Multipath TCP's token. However, the cookie is longer, encrypted in the ServerHello message, and one-time use (i.e., when the server receives a valid cookie, it accepts the connection, attaches it to the right TCPLS session, and discard the cookie). Thanks to the cookies, the server can limit the number of TCP connections that a client can attach to a TCPLS connection. This prevents some denial of service attacks that are possible with Multipath TCP.

2.4 The TCPLS API

The API that applications use to interact with a protocol plays an important role in enabling them to leverage all the protocol features. The most popular API to interact with the transport layer remains the BSD socket API. Researchers and the IETF have explored new ways to expose a transport API [32, 36, 53, 63, 71].

In this spirit, application-level developers would only be required to configure a TCPLS context and register function callbacks. To illustrate TCPLS API's flexibility, we consider a simple use case inspired by Happy Eyeballs [62]. This technique is used by web browsers when interacting with dual-stack servers. They try to establish TCP connections using IPv4 and IPv6 and prefer the one that offers the lowest latency. This avoids problems when an address family is broken on a path but not the other and sometimes results in lower latency [4].

Figure 3 shows an example of our current API workflow. The API can handle explicit multipath techniques such as Happy Eyeball by chaining `tcpls_connect()` with an appropriate timeout of 50ms, as shown in the Figure. TCPLS lets the application explicitly choose the multipath mesh by calling several times `tcpls_connect(src, dest, timeout);`. The application may configure callbacks to connection events that would occur within TCPLS , such as a connection establishment, a stream attachment, a multipath join, the reception of a TCP option to tune TCP, and more. When multiple streams are attached to multiple TCP connections, the application may configure various TCPLS behaviours. Among them, we support HOL-blocking avoidance, aggregation of bandwidth with multipathing, connection failover, and connection migration. Note that, HOL-blocking avoidance is incompatible with the aggregation of bandwidth with multipathing (the application can do either one but not both at the same time).

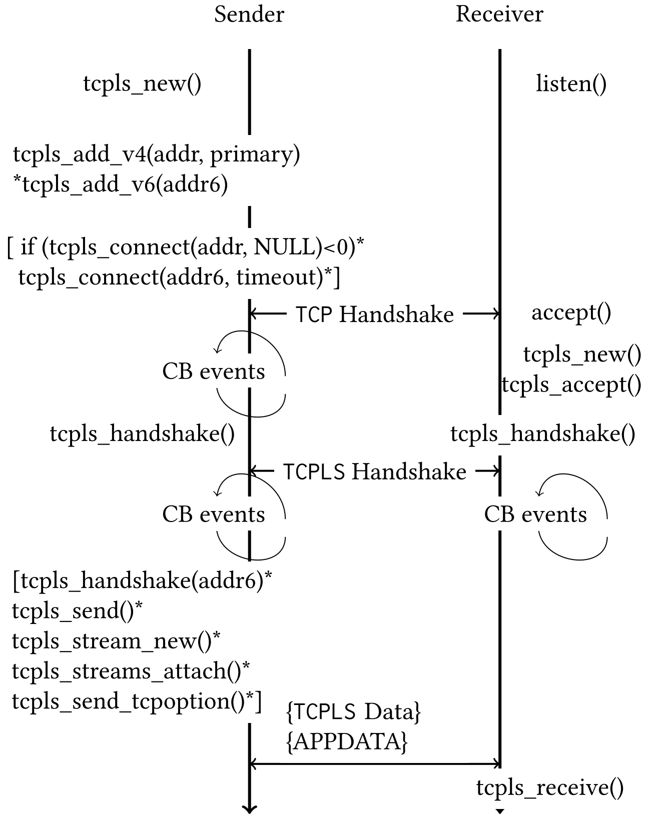

Figure 3: API Workflow example. * means optional call, [ ] means optional call flow, and { } means encrypted.

2.5 Comparison to QUIC, TCP and TLS /TCP

Table 1 compares the features supported by TCP , TLS /TCP , QUIC and TCPLS . QUIC and TCPLS are very similar in their capabilities. They mainly differ in their semantic. TCPLS 's semantic is to let the applications make the decision, and we design its API to fulfill this goal. That is, the meaning of

	TCP	TLS /TCP	QUIC	TCPLS
Transport reliability	✓	✓	✓	✓
Message conf. and auth.	✗	✓	✓	✓
Connection reliability	✗	✗	✓	(✓)
0-RTT	✓	(✗)	✓	✓
Session Resumption	✗	✓	✓	✓
Connection Migration	✗	✗	✓	✓
Application-exposed features				
Streams	✗	✗	✓	✓
Happy eyeballs	✗	✗	✗	✓
Explicit Multipath	✗	✗	✗	✓
App-level Con. migration	✗	✗	✗	✓
Pluginization	✗	✗	✗	(✓)
Resilience to HOL blocking	✗	✗	✓	(✓)
Secure Connection Closing	✗	✗	✓	(✓)

Table 1: Protocol features comparison. (✗) means that the feature is available, but not straightforward to use. (✓) means that the feature is partially available and under development.

TCPLS is to offer advanced, extensible and secure transport-layer functionalities on top of TCP , while exposing a simple but powerful API to let the application composes the properties its transport should have. One example is further demonstrated in Section 3.2, in which TCPLS 's simple API allows the application to take advantage of path aggregation (in multipath mode) and connection migration to obtain a smooth handover between networks.

Note that several of the features suggested by TCPLS are also suggested on TCP or QUIC via research works such as a new socket API for explicit multipath for TCP [32], or eBPF plugins in QUIC [18].

3 TCPLS PROTOTYPE

This section describes several of the possible benefits of TCPLS compared to keeping TCP and TLS isolated. We provide some use cases and experiment with the connection application-level connection migration offered by our API. Other user cases described in Section 4 are flagged to the roadmap and we expect them to further demonstrate the strength of a more intertwined TLS /TCP transport protocol.

Our current implementation offers: *(i)* An experimental API that wraps TLS and TCP and enables applications to handle multihoming, multipathing, and various transport layer mechanisms. *(ii)* An improved TCP extensibility mechanism that sends TCP options through the secure TCPLS channel. We currently support the TCP User Timeout option. Supporting another TCP option is only a matter of extending the sender's API and processing the option on the receiver side. TCPLS 's internal machinery can already send any TCP option during or after the handshake. *(iii)* The ability for the server to send eBPF bytecode over the secure channel to upgrade the client's TCP congestion control scheme or tune other TCP mechanisms [11, 69]. *(iv)* The support of parallel streams and multiplexing over TCP connections with different cryptographic context.

3.1 More Space for TCP Options

The TCP specification limits the size of the entire TCP header (including options) to 64 bytes. Unfortunately, the TCP designers did not foresee that so many TCP extensions would be standardized. Today, the size of the TCP header becomes a constraint. For example, it severely limits the number of gaps that can be covered by selective acknowledgments. This gets worse with extensions such as Multipath TCP [26] that consume more space in the TCP header. The IETF has discussed this problem for several years, but the latest attempt to solve it [68] has not yet been implemented by major TCP stacks.

TCPLS provides more space for some TCP options. First, with TCPLS , TCP options can be negotiated during the TLS handshake. Since the TLS messages are included in the TCP payload, there is more space to carry them. Another advantage of this approach is that the TCP options are secured by TLS . This implies that they cannot be modified by middleboxes. This could be an advantage, but could also prevent TCPLS from correctly working through some types of transparent TCP proxies.

Second, we can also carry TCP options inside TLS records. For example, we used this feature to implement the TCP User Time Out option [28]. A client can use this option to set the maximum value of the retransmission timer on a server. Linux TCP has a socket option that allows setting this timer locally, but it does not implement the option. With TCPLS , the client sends the option inside a TLS record, the server extracts it and performs the required `setsockopt`.

3.2 Application-level Connection Migration

Given the availability of multiple IP paths, connection migration might be a powerful tool to improve the application connection's reliability. We implement Connection migration and Failover as two distinct measures to handle two different inquiries: 1) The application expects to take advantage of multiple IP paths. 2) The application expects to be resilient to a network outage. In the first case, we implement connection migration and multipathing from a protocol viewpoint, as the same exchange of messages and API calls from TCPLS . It is left for the application to decide and program through the API calls whether it wants to move all the traffic from one path to another or split the traffic among the available paths figures/migration.png

Figure 4: Application-level connection migration during a 60MB file download

according. The second inquiry focuses on simply configuring TCPLS to automatically move the traffic to another available IP-level path if a network outage is detected.

Figure 4 shows the result of an Application-level connection migration demo using the API (i.e., it is left to the application to decide when to migrate, and we expose a simplistic code flow to perform it). In this experiment, we use an IPMininet network [41, 67] composed of a client and a server with a dual-stack of IPs. One path within the network is composed of OSPF routers with IPv4 only, and one path is composed of OSPF6 routers IPv6 only. We configure the bandwidth to 30Mbps, the lowest delay to the v4 link. Our application downloads a 60 MB file from a server and migrates to the v6 connection in the middle of the download.

Triggering the connection migration involves chaining 5 API calls: first, `tcpls_handshake()` configured with handshake properties announcing a JOIN over the v6 connection id. Then, the creation of a new stream `tcpls_stream_new()` for the v6 connection id, finally followed by the attachment of this new stream `tcpls_streams_attach()` and the secure closing of the v4 TCP connection using `tcpls_stream_close()`. Following these events, the server seamlessly switches the path while looping over `tcpls_send` to send the file content. Note that all the events trigger callbacks on the server side, to let the server react appropriately if other requirements need to be fulfilled.

TCPLS 's application connection migration takes advantage of multipath to offer a smooth handover to applications, which QUIC cannot do at the moment.

4 RESEARCH AGENDA

By closely integrating TCP and TLS , TCPLS opens new research questions in the transport layer and above. We highlight some of these in this section.

4.1 A More Secure Multipath TCP

Multipath TCP [26, 27] is a recent TCP extension that allows a connection to send data over different paths. It defines several TCP options, including ADD_ADDR to advertise addresses and RM_ADDR to remove addresses. Thanks to the ADD_ADDR option, a dual-stack server can advertise its IPv6 address over an IPv4 connection initiated by the client. The client can then use this address to create an IPv6 subflow that is part of the same connection.

One of the major deployments of Multipath TCP is on Apple's iPhones [8]. This implementation has decided not to support the ADD_ADDR option for security reasons. Since the Multipath TCP options are sent in clear, an attacker or a malicious middlebox could try to hijack connections. With TCPLS , this security concern can be addressed elegantly. First, Multipath TCPLS would not need to exchange a key in clear. It uses cookies (random 128-bits bitstrings) sent as Encrypted Extensions in the ServerHello during the handshake, and utilized in TCPLS JOIN handshakes. Second, in the case of Multipath TCP, the ADD_ADDR and RM_ADDR option could be sent inside TLS records that are encrypted and authenticated. The information would then reach MPTCP using a new `setsockopt`. Furthermore, since the TLS records are part of the bytestream, they are reliably delivered in contrast with the new ADD_ADDR option that is transmitted unreliably, like all TCP options, and thus needs to be echoed.

4.2 A More Secure TCP Fast Open

TCP Fast Open (TFO) [15, 56] is another recent TCP extension that allows sending data inside the SYN. TFO defines a TCP option that encodes a cookie. This cookie is used to prevent attacks from spoofed IP addresses. When a client connects first to a server, it sends an empty cookie but no data in the SYN packet. The server computes a cookie bound (e.g. using a hash) to the client's IP address and returns it in the SYN+ACK. For subsequent connections, the client sends its cookie in the SYN and places data in the payload. The server validates the cookie and processes the data since it comes from a legitimate IP address. However, the TCP header length limits the size of the TFO cookie. TCPLS could easily include a longer cookie inside the TLS ClientHello within the SYN payload. This solution would reduce the number of options in the TCP header and provide stronger protections against attackers. With this change, TCPLS would support a 0-RTT connection establishment similar to QUIC. In datacenters and controlled environments, this would work well. However, measurements are required before deploying that approach in enterprise and wireless networks as some of them contain middleboxes that block TCP Fast Open [52]. This middlebox interference has also affected QUIC [45] and is thus not specific to TCP .

4.3 Pluginizing TCPLS

PQUIC [18] proposed an elegant approach to deploy QUIC extensions. Instead of waiting for new client and server implementations, PQUIC includes an eBPF virtual machine to implement new features as bytecode that can be exchanged over the QUIC connection.

A TCPLS implementation could also be pluginized. The Linux kernel already includes an eBPF virtual machine [64].

It has already been used to develop several types of TCP extensions [11, 69, 70] and recent versions of the mainline kernel allow loading congestion control schemes implemented in eBPF bytecode. TCPLS can transport eBPF bytecode using TLS records as a second non-data stream. An interesting research question would be to evaluate the limits of such a dynamic extensibility? A first intuition to make TCP 's extensibility mechanism independent from the TCPLS version would be to let TCPLS communicating plugins to handle new TCP options and control behaviours, such that the supported TCP extensibility capability is not frozen by a given TCPLS version, but rather dependent on the set of plugins exchanged.

4.4 Limits to Cross-Layer Integration

QUIC and TCPLS show that there are benefits in integrating protocols at adjacent layers. QUIC already integrates HTTP/3 [6], but will likely be used by other applications [5, 38] in the future. TCP and TLS are already used by a wide range of applications [2]. Should TCPLS also be tuned for specific applications such as HTTP/3? What are the critical differences between QUIC and TCPLS from a functionality viewpoint?

4.5 Middlebox Interference

As explained earlier, the deployment of recent TCP extensions has been hindered by various types of middleboxes [20, 21, 33, 48]. QUIC already suffered from such problems [45] and many implementations fall back to TLS over TCP when blocked by a firewall. Researchers and application developers could include middlebox detection techniques inside TCPLS . Consider a TCPLS client that copies its SYN header within a TCPLS message alongside the early data, or makes it part of the zero-rtt resumption ticket message. By comparing the received TCP header with the original one, the server would immediately and reliably detect the presence of NAT, transparent proxies or other types of middleboxes. It could then inform the client and adjust the configuration of their stacks accordingly or fall back to regular TCP to preserve connectivity. If deployed, such a protocol would enable researchers to understand the impact of middleboxes more accurately.

4.6 Efficient TCPLS Implementations

During the last years, researchers have proposed to move transport protocol implementations to user-space [42, 47] to kernel bypass techniques such as netmap [59] or DPDK. In parallel, parts of TLS moved to the Linux kernel [9] for performance reasons. With new results on SmartNICs [23], it would be interesting to analyze the best architecture for new TCPLS implementations. Given the enormous efforts on implementing QUIC[30], it would be exciting to compare QUIC and TCPLS from a performance viewpoint. From a protocol standpoint, performance advantages of combining those two layers may be achieved from, for example, adjusting the size of TLS records based on the current TCP congestion window to avoid fragmented records (non-fragmented records makes TCPLS ' design having a zero-copy code path). More generally, we expect many performance benefits from a more intertwined TLS /TCP transport protocol at the cost of design complexity.

SOFTWARE ARTEFACTS

A TCPLS reference documentation and implementation is under active development. The current specifications and code are available on https://pluginized-protocols.org/tcpls, forked from a fast and full TLS 1.3 implementation written in C. Our TCPLS prototype adds about 5k lines of C code to `picotls` latest version based on the latest specification of TLS 1.3.

ACKNOWLEDGMENTS

We thanks the anonymous reviewers for their helpful feedback, and Mathieu Jadin for its helpful guidance with IP-Mininet. This research is supported by the Walloon Region through the “Programme de recherche d'interet general WALINNOV" - MQUIC project (convention number 1810018).

REFERENCES

[1] Josh Aas, Richard Barnes, Benton Case, Zakir Durumeric, Peter Eckersley, Alan Flores-López, J Alex Halderman, Jacob Hoffman-Andrews, James Kasten, Eric Rescorla, et al. 2019. Let's Encrypt: An Automated Certificate Authority to Encrypt the Entire Web. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 2473–2487.

[2] Blake Anderson and David McGrew. 2019. TLS Beyond the Browser: Combining End Host and Network Data to Understand Application Behavior. In *Proceedings of the Internet Measurement Conference*. 379–392.

[3] M. Bagnulo. [n.d.]. Threat Analysis for TCP Extensions for Multipath Operation with Multiple Addresses. RFC 6181. https://tools.ietf.org/html/rfc6181

[4] Vaibhav Bajpai and Jürgen Schönwälder. 2019. A longitudinal view of dual-stacked websites—failures, latency and happy eyeballs. *IEEE/ACM Transactions on Networking* 27, 2 (2019), 577–590.

[5] D Bider. [n.d.]. QUIC-based UDP Transport for Secure Shell (SSH). Internet-Draft. https://datatracker.ietf.org/doc/draft-bider-ssh-quic Work in Progress.

[6] Mike Bishop. 2020. *Hypertext Transfer Protocol Version 3 (HTTP/3)*. Internet-Draft draft-ietf-quic-http-29. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-quic-http-29 Work in Progress.

[7] Andrea Bittau, Daniel B. Giffin, Mark J. Handley, David Mazieres, Quinn Slack, and Eric W. Smith. 2019. Cryptographic Protection of TCP Streams (tcpcrypt). RFC 8548. https://doi.org/10.17487/RFC8548

[8] Olivier Bonaventure and S Seo. 2016. Multipath TCP deployments. *IETF Journal* 12, 2 (2016), 24–27.

[9] Daniel Borkmann and John Fastabend. 2018. Combining kTLS and BPF for Introspection and Policy Enforcement. In *Linux Plumbers Conference'18*.

[10] David A. Borman, Robert T. Braden, and Van Jacobson. 1992. TCP Extensions for High Performance. RFC 1323. https://doi.org/10.17487/RFC1323

[11] Lawrence Brakmo. 2017. Tcp-bpf: Programmatically tuning tcp behavior through bpf. *NetDev 2.2* (2017).

[12] Łukasz Budzisz, Johan Garcia, Anna Brunstrom, and Ramon Ferrús. 2012. A taxonomy and survey of SCTP research. *ACM Computing Surveys (CSUR)* 44, 4 (2012), 1–36.

[13] Zimo Chai, Amirhossein Ghafari, and Amir Houmansadr. 2019. On the Importance of Encrypted-SNI (ESNI) to Censorship Circumvention. In *Free and Open Communications on the Internet*. USENIX. https://www.usenix.org/system/files/foci19-paper_chai_update.pdf

[14] Sanjit Chatterjee, Alfred Menezes, and Palash Sarkar. 2011. Another look at tightness. In *International Workshop on Selected Areas in Cryptography*. Springer, 293–319.

[15] Yuchung Cheng, Jerry Chu, Sivasankar Radhakrishnan, and Arvind Jain. 2014. TCP Fast Open. RFC 7413. https://doi.org/10.17487/RFC7413

[16] David D Clark, Van Jacobson, John Romkey, and Howard Salwen. 1989. An analysis of TCP processing overhead. *IEEE Communications magazine* 27, 6 (1989), 23–29.

[17] Yong Cui, Zhiwen Liu, Hang Shi, Jie Zhang, Kai Zheng, and Wei Wang. 2020. *Deadline-aware Transport Protocol*. Internet-Draft draft-shi-quic-dtp-01. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-shi-quic-dtp-01 Work in Progress.

[18] Quentin De Coninck, François Michel, Maxime Piraux, Florentin Rochet, Thomas Given-Wilson, Axel Legay, Olivier Pereira, and Olivier Bonaventure. 2019. Pluginizing quic. In *Proceedings of the ACM Special Interest Group on Data Communication*. 59–74.

[19] Gregory Detal, Benjamin Hesmans, Olivier Bonaventure, Yves Vanaubel, and Benoit Donnet. 2013. Revealing Middlebox Interference with Tracebox. In *Proceedings of the 2013 ACM SIGCOMM conference on Internet measurement conference*. ACM.

[20] Gregory Detal, Benjamin Hesmans, Olivier Bonaventure, Yves Vanaubel, and Benoit Donnet. 2013. Revealing middlebox interference with tracebox. In *Proceedings of the 2013 conference on Internet measurement conference*. 1–8.

[21] Korian Edeline and Benoit Donnet. 2017. A first look at the prevalence and persistence of middleboxes in the wild. In *2017 29th International Teletraffic Congress (ITC 29)*, Vol. 1. IEEE, 161–168.

[22] Dr. Taher Elgamal and Kipp E.B. Hickman. 1995. *The SSL Protocol*. Internet-Draft draft-hickman-netscape-ssl-00. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-hickman-netscape-ssl-00 Work in Progress.

[23] Daniel Firestone, Andrew Putnam, Sambhrama Mundkur, Derek Chiou, Alireza Dabagh, Mike Andrewartha, Hari Angepat, Vivek Bhanu, Adrian Caulfield, Eric Chung, et al. 2018. Azure accelerated networking: SmartNICs in the public cloud. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*. 51–66.

[24] Sally Floyd. 2002. Inappropriate TCP Resets Considered Harmful. RFC 3360. https://doi.org/10.17487/RFC3360

[25] Sally Floyd, Jamshid Mahdavi, Matt Mathis, and Dr. Allyn Romanow. 1996. TCP Selective Acknowledgment Options. RFC 2018. https://doi.org/10.17487/RFC2018

[26] Alan Ford, Costin Raiciu, Mark J. Handley, and Olivier Bonaventure. 2013. TCP Extensions for Multipath Operation with Multiple Addresses. RFC 6824. https://doi.org/10.17487/RFC6824

[27] Alan Ford, Costin Raiciu, Mark J. Handley, Olivier Bonaventure, and Christoph Paasch. 2020. TCP Extensions for Multipath Operation with Multiple Addresses. RFC 8684. https://doi.org/10.17487/RFC8684

[28] Fernando Gont and Lars Eggert. 2009. TCP User Timeout Option. RFC 5482. https://doi.org/10.17487/RFC5482

[29] Devashish Gosain, Anshika Agarwal, Sahil Shekhawat, H. B. Acharya, and Sambuddho Chakravarty. 2017. Mending Wall: On the Implementation of Censorship in India. In *SecureComm*. Springer. https://censorbib.nymity.ch/pdf/Gosain2017a.pdf

[30] QUIC Working Group. [n.d.]. Available QUIC Implementations. ([n. d.]). https://github.com/quicwg/base-drafts/wiki/Implementations.

[31] F Günther, M. Thomson, and C.A. Wood. [n.d.]. Usage Limits on AEAD Algorithms. Internet-Draft. https://tools.ietf.org/html/draft-wood-cfrg-aead-limits-00 Work in Progress.

[32] Benjamin Hesmans and Olivier Bonaventure. 2016. An enhanced socket API for Multipath TCP. In *Proceedings of the 2016 applied networking research workshop*. 1–6.

[33] Benjamin Hesmans, Fabien Duchene, Christoph Paasch, Gregory Detal, and Olivier Bonaventure. 2013. Are TCP extensions middlebox-proof?. In *Proceedings of the 2013 workshop on Hot topics in middleboxes and network function virtualization*. 37–42.

[34] Ralph Holz, Johanna Amann, Abbas Razaghpanah, and Narseo Vallina-Rodriguez. 2019. The Era of TLS 1.3: Measuring Deployment and Use with Active and Passive Methods. *arXiv preprint arXiv:1907.12762* (2019).

[35] Michio Honda, Yoshifumi Nishida, Costin Raiciu, Adam Greenhalgh, Mark Handley, and Hideyuki Tokuda. 2011. Is it still possible to extend TCP?. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. 181–194.

[36] Tomas Hruby, Teodor Crivat, Herbert Bos, and Andrew S Tanenbaum. 2014. On Sockets and System Calls: Minimizing Context Switches for the Socket API. In *2014 Conference on Timely Results in Operating Systems (TRIOS 14)*.

[37] Christian Huitema. 2020. *Quic Timestamps For Measuring One-Way Delays*. Internet-Draft draft-huitema-quic-ts-02. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-huitema-quic-ts-02 Work in Progress.

[38] Christian Huitema, Allison Mankin, and Sara Dickinson. 2020. *Specification of DNS over Dedicated QUIC Connections*. Internet-Draft draft-ietf-dprive-dnsoquic-00. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-dprive-dnsoquic-00 Work in Progress.

[39] Jana Iyengar and Martin Thomson. 2020. *QUIC: A UDP-Based Multiplexed and Secure Transport*. Internet-Draft draft-ietf-quic-transport-28. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-quic-transport-28 Work in Progress.

[40] Janardhan R Iyengar, Paul D Amer, and Randall Stewart. 2006. Concurrent multipath transfer using SCTP multihoming over independent end-to-end paths. *IEEE/ACM Transactions on networking* 14, 5 (2006), 951–964.

[41] Mathieu Jadin, Olivier Tilmans, Maxime Mawait, and Olivier Bonaventure. 2020. Educational Virtual Routing Labs with IPMininet. In *ACM SIGCOMM Education Workshop 2020*.

[42] EunYoung Jeong, Shinae Wood, Muhammad Jamshed, Haewon Jeong, Sunghwan Ihm, Dongsu Han, and KyoungSoo Park. 2014. mtcp: a highly scalable user-level TCP stack for multicore systems. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. 489–502.

[43] Eddie Kohler, Mark Handley, and Sally Floyd. 2006. Designing DCCP: Congestion control without reliability. *ACM SIGCOMM Computer Communication Review* 36, 4 (2006), 27–38.

[44] Platon Kotzias, Abbas Razaghpanah, Johanna Amann, Kenneth G Paterson, Narseo Vallina-Rodriguez, and Juan Caballero. 2018. Coming of age: A longitudinal study of tls deployment. In *Proceedings of the Internet Measurement Conference 2018*. 415–428.

[45] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, et al. 2017. The quic transport protocol: Design and internet-scale deployment. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. 183–196.

[46] Atul Luykx and Kenneth G Paterson. 2015. Limits on authenticated encryption use in TLS. *Personal webpage: http://www. isg. rhul. ac. uk/˜kp/TLS-AEbounds. pdf* (2015).

[47] Ilias Marinos, Robert NM Watson, and Mark Handley. 2014. Network stack specialization for performance. *ACM SIGCOMM Computer Communication Review* 44, 4 (2014), 175–186.

[48] Alberto Medina, Mark Allman, and Sally Floyd. 2004. Measuring interactions between transport protocols and middleboxes. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. 336–341.

[49] Alberto Medina, Mark Allman, and Sally Floyd. 2005. Measuring the Evolution of Transport Protocols in the Internet. *SIGCOMM Comput. Commun. Rev.* (April 2005), 37–52. https://doi.org/10.1145/1064413.1064418

[50] François Michel, Quentin De Coninck, and Olivier Bonaventure. 2019. QUIC-FEC: Bringing the benefits of Forward Erasure Correction to QUIC. In *2019 IFIP Networking Conference (IFIP Networking)*. IEEE, 1–9.

[51] Mehrab Bin Morshed, Michaelanne Dye, Syed Ishtiaque Ahmed, and Neha Kumar. 2017. When the Internet Goes Down in Bangladesh. In *Computer-Supported Cooperative Work and Social Computing*. ACM. https://nehakumardotorg.files.wordpress.com/2014/03/p1591-bin-morshed.pdf

[52] Christoph Paasch. 2016. Network support for TCP Fast Open. *Presentation at NANOG* 67 (2016).

[53] Tommy Pauly, Brian Trammell, Anna Brunstrom, Gorry Fairhurst, Colin Perkins, Philipp S. Tiesel, and Christopher A. Wood. 2020. *An Architecture for Transport Services*. Internet-Draft draft-ietf-taps-arch-07. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-taps-arch-07 Work in Progress.

[54] Michele Polese, Federico Chiariotti, Elia Bonetto, Filippo Rigotto, Andrea Zanella, and Michele Zorzi. 2019. A survey on recent advances in transport layer protocols. *IEEE Communications Surveys & Tutorials* 21, 4 (2019), 3584–3608.

[55] Jon Postel. 1981. Transmission Control Protocol. RFC 793. https://doi.org/10.17487/RFC0793

[56] Sivasankar Radhakrishnan, Yuchung Cheng, Jerry Chu, Arvind Jain, and Barath Raghavan. 2011. TCP Fast Open. In *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*. 1–12.

[57] Costin Raiciu, Christoph Paasch, Sebastien Barre, Alan Ford, Michio Honda, Fabien Duchene, Olivier Bonaventure, and Mark Handley. 2012. How Hard Can It Be? Designing and Implementing a Deployable Multipath TCP. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*. 399–412.

[58] Eric Rescorla. 2018. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446. https://doi.org/10.17487/RFC8446

[59] Luigi Rizzo. 2012. Revisiting network I/O APIs: the netmap framework. *Queue* 10, 1 (2012), 30–39.

[60] Jim Roskind. 2013. QUIC, Quick UDP Internet Connections. https://docs.google.com/document/d/1RNHkx_VvKWyWg6Lr8SZ-saqsQx7rFV-ev2jRFUoVD34/preview.

[61] Robert M Sanders and Alfred C Weaver. 1990. The Xpress transfer protocol (XTP)— a tutorial. *ACM SIGCOMM Computer Communication Review* 20, 5 (1990), 67–80.

[62] David Schinazi and Tommy Pauly. 2017. Happy Eyeballs Version 2: Better Connectivity Using Concurrency. RFC 8305. https://doi.org/10.17487/RFC8305

[63] Philipp S Schmidt, Theresa Enghardt, Ramin Khalili, and Anja Feldmann. 2013. Socket intents: Leveraging application awareness for multi-access connectivity. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*. 295–300.

[64] Alexei Starovoitov. 2014. BPF syscall, maps, verifier, sample. (June 2014). Linux kernel patch https://lkml.org/lkml/2014/6/27/545.

[65] Randall R. Stewart. 2007. Stream Control Transmission Protocol. RFC 4960. https://doi.org/10.17487/RFC4960

[66] Ian Swett, Marie-Jose Montpetit, Vincent Roca, and François Michel. 2020. *Coding for QUIC*. Internet-Draft draft-swett-nwcrg-coding-for-quic-04. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-swett-nwcrg-coding-for-quic-04 Work in Progress.

[67] Olivier Tilmans and Mathieu Jadin. [n.d.]. IPMininet. ([n. d.]). https://github.com/cnp3/ipmininet, Accessed Feb-20-2020.

[68] Joseph D. Touch and Wesley Eddy. 2018. *TCP Extended Data Offset Option*. Internet-Draft draft-ietf-tcpm-tcp-edo-10. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-tcpm-tcp-edo-10 Work in Progress.

[69] Viet-Hoang Tran and Olivier Bonaventure. 2019. Beyond socket options: making the Linux TCP stack truly extensible. In *2019 IFIP Networking Conference (IFIP Networking)*. IEEE, 1–9.

[70] Viet-Hoang Tran and Olivier Bonaventure. 2020. Beyond socket options: Towards fully extensible Linux transport stacks. *Computer Communications* 162 (2020), 118–138.

[71] Michael Tüxen, Vladislav Yasevich, Peter Lei, Randall R. Stewart, and Kacheong Poon. 2011. Sockets API Extensions for the Stream Control Transmission Protocol (SCTP). RFC 6458. https://doi.org/10.17487/RFC6458

[72] Tobias Viernickel, Alexander Froemmgen, Amr Rizk, Boris Koldehofe, and Ralf Steinmetz. 2018. Multipath QUIC: A deployable multipath transport protocol. In *2018 IEEE International Conference on Communications (ICC)*. IEEE, 1–7.

[73] Zhaoguang Wang, Zhiyun Qian, Qiang Xu, Zhuoqing Mao, and Ming Zhang. 2011. An untold story of middleboxes in cellular networks. *ACM SIGCOMM Computer Communication Review* 41, 4 (2011), 374–385.

[74] Nicholas Weaver, Robin Sommer, and Vern Paxson. 2009. Detecting Forged TCP Reset Packets.. In *NDSS*.

[75] Damon Wischik, Costin Raiciu, Adam Greenhalgh, and Mark Handley. 2011. Design, Implementation and Evaluation of Congestion Control for Multipath TCP.. In *NSDI*, Vol. 11. 8–8.

[76] Xing Xu, Yurong Jiang, Tobias Flach, Ethan Katz-Bassett, David Choffnes, and Ramesh Govindan. 2015. Investigating transparent web proxies in cellular networks. In *International Conference on Passive and Active Network Measurement*. Springer, 262–276.