

TAPS Working Group
Internet-Draft
Intended status: Informational
Expires: August 30, 2018

T. Pauly, Ed.
Apple Inc.
B. Trammell, Ed.
ETH Zurich
A. Brunstrom
Karlstad University
G. Fairhurst
University of Aberdeen
C. Perkins
University of Glasgow
P. Tiesel
TU Berlin
C. Wood
Apple Inc.
February 26, 2018

An Architecture for Transport Services
draft-pauly-taps-arch-00

Abstract

This document provides an overview of the architecture of Transport Services, a system for exposing the features of transport protocols to applications. This architecture serves as a basis for Application Programming Interfaces (APIs) and implementations that provide flexible transport networking services. It defines the common set of terminology and concepts to be used in more detailed discussion of Transport Services.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 30, 2018.

Copyright Notice

Copyright (c) 2018 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Background	3
3. Design Principles	4
3.1. Common APIs for Common Features	4
3.2. Access to Specialized Features	4
3.3. Scope for API and Implementation Definitions	5
4. Transport Services Architecture and Concepts	6
4.1. Transport Services API Concepts	7
4.1.1. Basic Objects	9
4.1.2. Pre-Establishment	10
4.1.3. Establishment Actions	11
4.1.4. Data Transfer Objects and Actions	11
4.1.5. Event Handling	12
4.1.6. Termination Actions	12
4.2. Transport System Implementation Concepts	13
4.2.1. Gathering	14
4.2.2. Racing	14
5. IANA Considerations	14
6. Security Considerations	14
7. Acknowledgements	15
8. Informative References	15
Authors' Addresses	16

1. Introduction

Many APIs to perform transport networking have been deployed, perhaps the most widely known and imitated being the BSD `socket()` interface. The names and functions between these APIs are not consistent, and vary depending on the protocol being used. For example, sending and receiving on a stream of data is conceptually the same between operating on an unencrypted TCP stream and operating on an encrypted

TLS stream over TCP, but applications cannot use the same socket send() and recv() calls on top of both kinds of connections. Similarly, terminology for the implementation of protocols offering transport services vary based on the context of the protocols themselves. This variety can lead to confusion when trying to understand the similarities and differences between protocols, and how applications can use them effectively.

The goal of the Transport Services architecture is to provide a common, flexible, and reusable interface for transport protocols. As applications adopt this interface, they will benefit from a wide set of transport features that can evolve over time, and ensure that the system providing the interface can optimize its behavior based on the application requirements and network conditions.

This document is developed in parallel with the specification of the Transport Services API [draft-trammell-taps-interface] and Implementation [draft-brunstrom-taps-impl] documents.

2. Background

The Transport Services architecture is based on the survey of Services Provided by IETF Transport Protocols and Congestion Control Mechanisms [RFC8095], and the distilled minimal set of the features offered by transport protocols [I-D.ietf-taps-minset]. This work has identified common features and patterns across all transport protocols developed thus far in the IETF.

Since transport security is an increasingly relevant aspect of using transport protocols on the Internet, this architecture also considers the impact of transport security protocols on the feature set exposed by transport services [I-D.pauly-taps-transport-security].

One of the key insights to come from identifying the minimal set of features provided by transport protocols [I-D.ietf-taps-minset] was that features either require application interaction and guidance (referred to as Functional Features), or else can be handled automatically by a system implementing Transport Services (referred to as Automatable Features). Among the Functional Features, some were common across all or nearly all transport protocols, while others could be seen as features that, if specified, would only be useful with a subset of protocols, or perhaps even a single transport protocol, but would not harm the functionality of other protocols. For example, some protocols can deliver messages faster for applications that do not require them to arrive in the order in which they were sent. However, this functionality must be explicitly allowed by the application, since reordering messages would be undesirable in many cases.

3. Design Principles

The goal of the Transport Services architecture is to redefine the interface between applications and transports in a way that allows the transport layer to evolve and improve without fundamentally changing the contract with the application. This requires a careful consideration of how to expose the capabilities of protocols.

There are several degrees in which a Transport Services system can offer flexibility to an application: it can provide access to multiple sets of protocols and protocol features, it can use these protocols across multiple paths that may have different performance and functional characteristics, and it can communicate with different Remote Endpoints to optimize performance. Beyond these, if the API for the system remains the same over time, new protocols and features may be added to the system's implementation without requiring significant changes in applications for adoption.

The following considerations were used in the design of this architecture.

3.1. Common APIs for Common Features

Functionality that is common across multiple transport protocols should be accessible through a unified set of API calls. An application should be able to implement logic for its basic use of transport networking (establishing the transport, and sending and receiving data) once, and expect that implementation to continue to function as the transports change.

Any Transport Services API must allow access to the distilled minimal set of features offered by transport protocols [I-D.ietf-taps-minset].

3.2. Access to Specialized Features

Since applications will often need to control fine-grained details of transport protocols to optimize their behavior and ensure compatibility with remote peers, a Transport Services system also needs to allow more specialized protocol features to be used. The interface for these specialized options should be exposed differently from the common options to ensure flexibility.

A specialized feature may be required by an application only when using a specific protocol, and not when using others. For example, if an application is using UDP, it may require control over the checksum or fragmentation behavior for UDP; if it used a protocol to frame its data over a byte stream like TCP, it would not need these

options. In such cases, the API should expose the features in such a way that they take effect when a particular protocol is selected, but do not imply that only that protocol may be used if there are equivalent options.

Other specialized features, however, may be strictly required by an application and thus constrain the set of protocols that can be used. For example, if an application requires encryption of its transport data, only protocol stacks that include some transport security protocol are eligible to be used. A Transport Services API must allow applications to define such requirements and constrain the system's options. Since such options are not part of the core/common features, it should be simple for an application to modify its set of constraints and change the set of allowable protocol features without changing the core implementation.

3.3. Scope for API and Implementation Definitions

The Transport Services API is envisioned as the abstract model for a family of APIs that share a common way to expose transport features and encourage flexibility. The abstract API definition [draft-trammell-taps-interface] describes this interface and is aimed at application developers.

Implementations that provide the Transport Services API [draft-brunstrom-taps-impl] will vary due to system-specific support and the needs of the deployment scenario. It is expected that all implementations of Transport Services will offer the entire mandatory API, but that some features will not be functional in certain implementations. All implementations must offer sufficient APIs to use the distilled minimal set of features offered by transport protocols [I-D.ietf-taps-minset], including API support for TCP and UDP transport, but it is possible that some very constrained devices might not have, for example, a full TCP implementation.

In order to preserve flexibility and compatibility with future protocols, top-level features in the Transport Services API should avoid referencing particular transport protocols. Mappings of these API features in the Implementation document, on the other hand, must explain the ramifications of each feature on existing protocols. It is expected that the Implementation document will be updated and supplemented as new protocols and protocol features are developed.

It is important to note that neither the Transport Services API nor the Implementation document defines new protocols that require any changes on remote hosts. The Transport Services system must be deployable on one side only, as a way to allow an application to make

better use of available capabilities on a system and protocol features that may be supported by peers across the network.

4. Transport Services Architecture and Concepts

The concepts defined in this document are intended primarily for use in the documents and specifications that describe the Transport Services architecture and API. While the specific terminology may be used in some implementations, it is expected that there will remain a variety of terms used by running code.

The architecture divides the concepts for Transport Services into two categories:

1. API concepts, which are meant to be exposed to applications; and
2. System-implementation concepts, which are meant to be internally used when building systems that implement Transport Services.

The following diagram summarizes the top-level concepts in the architecture and how they relate to one another.

any system that provides generic Transport Services, these properties should primarily offer knobs that are applicable to multiple transports. Properties may have a large impact on the rest of the aspects of the interface: they can modify how establishment occurs, they can influence the expectations around data transfer, and they determine the set of events that will be supported.

- o Establishment (Section 4.1.3) focuses on the actions that an application takes on the basic objects to prepare for data transfer.
- o Data Transfer (Section 4.1.4) consists of how an application represents data to be sent and received, the functions required to send and receive that data, and how the application is notified of the status of its data transfer.
- o Event Handling (Section 4.1.5) defines the set of properties about which an application can receive notifications during the lifetime of transport objects. Events can also provide opportunities for the application to interact with the underlying transport by querying state or updating maintenance options.
- o Termination (Section 4.1.6) focuses on the methods by which data transmission is ceased, and state is torn down in the transport.

The diagram below provides a high-level view of the actions taken during the lifetime of a connection.

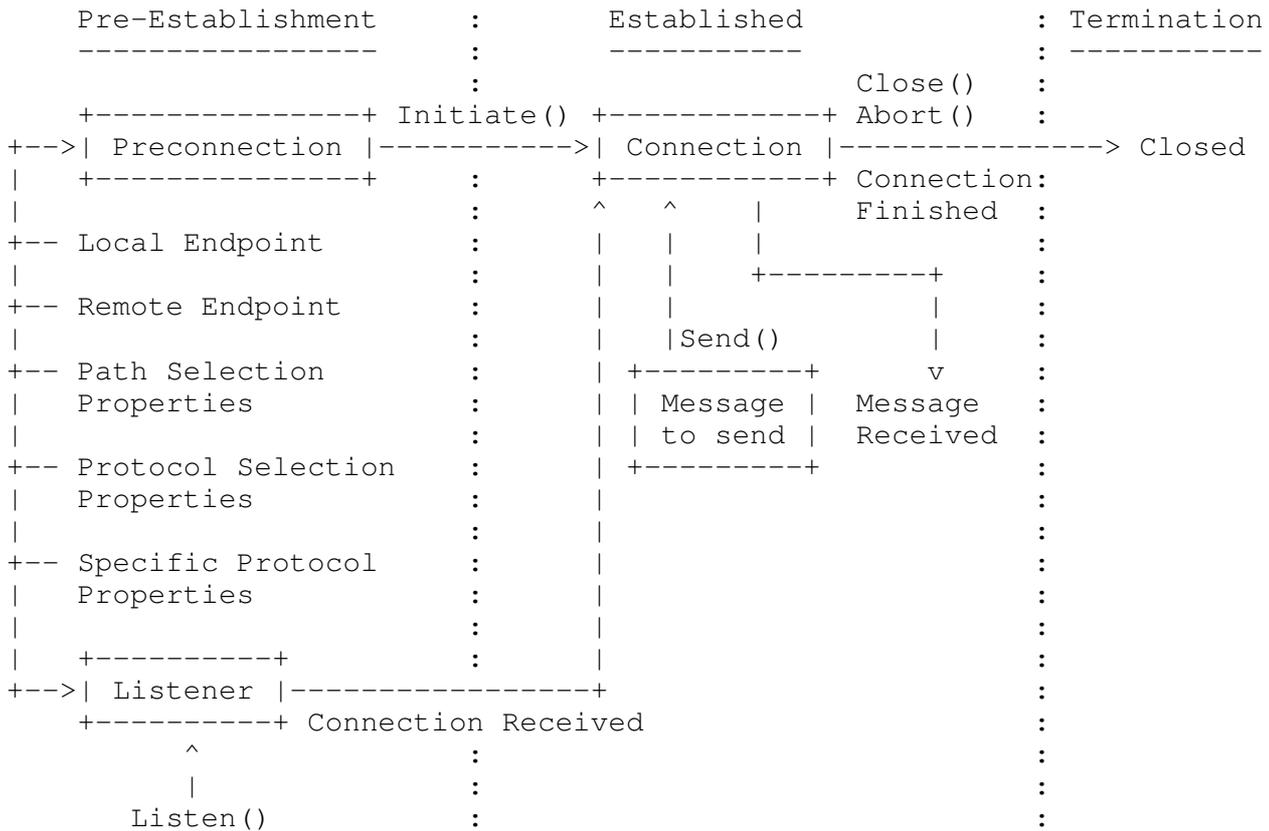


Figure 2: The lifetime of a connection

4.1.1.1. Basic Objects

- o Preconnection: A Preconnection object is a representation of a potential connection. It has state that describes parameters of a Connection that might exist in the future: the Local Endpoint from which that Connection will be established, the Remote Endpoint to which it will connect, and Path Selection Properties, Protocol Selection Properties, and Specific Protocol Properties that influence the choice of transport that a Connection will use. A Preconnection can be fully specified and represent a single possible Connection, or it can be partially specified such that it represents a family of possible Connections. The Local Endpoint must be specified if the Preconnection is used to Listen for incoming connections, but is optional if it is used to Initiate connections. The Remote Endpoint must be specified in the Preconnection is used to Initiate connections, but is optional if it is used to Listen for incoming connections.

- o **Connection:** A Connection object represents an active transport protocol instance that can send and/or receive Messages between a Local Endpoint and a Remote Endpoint. It holds state pertaining to the underlying transport protocol instance and any ongoing data transfer. This represents, for example, an active connection in a connection-oriented protocol such as TCP, or a fully-specified 5-tuple for a connectionless protocol such as UDP.
- o **Listener:** A Listener object accepts incoming transport protocol connections from Remote Endpoints and generates corresponding Connection objects. It is created from a Preconnection object that specifies the type of incoming connections it will accept.

4.1.2. Pre-Establishment

- o **Endpoint:** An Endpoint represents one side of a transport connection. Endpoints can be Local Endpoints or Remote Endpoints, and respectively represent an identity that the application uses for the source or destination of a connection. Endpoint can vary in levels of specificity, and can be resolved to more concrete identities.
- o **Remote Endpoint:** The Remote Endpoint represents the application's name for a peer that can participate in a transport connection. For example, the combination of a DNS name for the peer and a service name/port.
- o **Local Endpoint:** The Local Endpoint represents the application's name for itself that it wants to use for transport connections. For example, a local IP address and port.
- o **Path Selection Properties:** The Path Selection Properties consist of the options that an application may set to influence the selection of paths between itself and the Remote Endpoint. These options can come in the form of requirements, prohibitions, or preferences. Examples of options which may influence path selection include the interface type (such as a Wi-Fi Ethernet connection, or a Cellular LTE connection), characteristics of the path that are locally known like Maximum Transmission Unit (MTU) or discovered like Path MTU (PMTU), or predicted based on cached information like expected throughput or latency.
- o **Protocol Selection Properties:** The Protocol Selection Properties consist of the options that an application may set to influence the selection of transport protocol, or to configure the behavior of generic transport protocol features. These options come in the form of requirements, prohibitions, and preferences. Examples

include reliability, service class, multipath support, and fast open support.

- o **Specific Protocol Properties:** The Specific Protocol Properties refer to the subset of Protocol Properties options that apply to a single protocol (transport protocol, IP, or security protocol). The presence of such Properties does not necessarily require that a specific protocol must be used when a Connection is established, but that if this protocol is employed, a particular set of options should be used.

4.1.3. Establishment Actions

- o **Initiate** is the primary action that an application can take to create a Connection to a remote endpoint, and prepare any required local or remote state to be able to send and/or receive Messages. For some protocols, this may initiate a server-to-client style handshake; for other protocols, this may just establish local state; and for peer-to-peer protocols, this may begin the process of a simultaneous open. The process of identifying options for connecting, such as resolution of the Remote Endpoint, occurs in response the Initiate call.
- o **Listen** is the action of marking a Listener as willing to accept incoming Connections. The Listener will then create Connection objects as incoming connections are accepted (Section 4.1.5).

4.1.4. Data Transfer Objects and Actions

- o **Message:** A Message object is a unit of data that can be represented as bytes that can be transferred between two endpoints over a transport connection. The bytes within a Message are assumed to be ordered within the Message. If an application does not care about the order in which a peer receives two distinct spans of bytes, those spans of bytes are considered independent Messages. Messages may or may not be usable if incomplete or corrupted. Boundaries of a Message may or may not be understood or transmitted by transport protocols. Specifically, what one application considers to be two Messages sent on a stream-based transport may be treated as a single Message by the application on the other side.
- o **Send** is the action to transmit a Message or partial Message over a Connection to a Remote Endpoint. The interface to Send may include options specific to how the Message's content is to be sent. Status of the Send operation may be delivered back to the application in an event (Section 4.1.5).

- o Receive is an action that indicates that the application is ready to asynchronously accept a Message over a Connection from a Remote Endpoint, while the Message content itself will be delivered in an event (Section 4.1.5). The interface to Receive may include options specific to the Message that is to be delivered to the application.

4.1.5. Event Handling

This list of events that can be delivered to an application is not exhaustive, but gives the top-level categories of events. The API may expand this list.

- o **Connection Ready:** Signals to an application that a given Connection is ready to send and/or receive Messages. If the Connection relies on handshakes to establish state between peers, then it is assumed that these steps have been taken.
- o **Connection Finished:** Signals to an application that a given Connection is no longer usable for sending or receiving Messages. This should deliver an error to the application that describes the nature of the termination.
- o **Connection Received:** Signals to an application that a given Listener has passively received a Connection.
- o **Message Received:** Delivers received Message content to the application, based on a Receive action. This may include an error if the Receive action cannot be satisfied due to the Connection being closed.
- o **Message Sent:** Notifies the application of the status of its Send action. This may be an error if the Message cannot be sent, or an indication that Message has been processed by the protocol stack.
- o **Path Properties Changed:** Notifies the application that some property of the Connection has changed that may influence how and where data is sent and/or received.

4.1.6. Termination Actions

- o **Close** is the action an application may take on a Connection to indicate that it no longer intends to send data, is no longer willing to receive data, and that the protocol should signal this state to the remote endpoint if applicable.
- o **Abort** is an action the application may take on a Connection to indicate a Close, but with the additional indication that the

transport system should not attempt to deliver any outstanding data.

4.2. Transport System Implementation Concepts

The Transport System Implementation Concepts define the set of objects used internally to a system or library to provide the functionality of transport networking, as required by the abstract interface.

- o **Connection Group:** A Connections Group is a set of Connections that share properties. For multiplexing transport protocols, the Connection Group defines the set of Connections that can be multiplexed together.
- o **Path:** A Path represents an available set of properties of a network route on which packets may be sent or received.
- o **Protocol Instance:** A Protocol Instance is a single instance of one protocol, including any state it has necessary to establish connectivity or send and receive Messages.
- o **Protocol Stack:** A Protocol Stack is a set of Protocol Instances (including relevant application, security, transport, or Internet protocols) that are used together to establish connectivity or send and receive Messages. A single stack may be simple (a single transport protocol instance over IP), or complex (multiple application protocol streams going through a single security and transport protocol, over IP; or, a multi-path transport protocol over multiple transport sub-flows).
- o **System Policy:** System Policy represents the input from an operating system or other global preferences that can constrain or influence how an implementation will gather candidate paths and protocols (Section 4.2.1) and race the candidates during establishment (Section 4.2.2). Specific aspects of the System Policy may apply to all Connections, or only certain ones depending on the runtime context and properties of the Connection.
- o **Cached State:** Cached State is the state and history that the implementation keeps for each set of associated endpoints that have been used previously. This can include DNS results, TLS session state, previous success and quality of transport protocols over certain paths.

4.2.1. Gathering

- o **Path Selection:** Path Selection represents the act of choosing one or more paths that are available to use based on the Path Selection Properties provided by the application, and a Transport Services system's policies and heuristics.
- o **Protocol Selection:** Protocol Selection represents the act of choosing one or more sets of protocol options that are available to use based on the Protocol Properties provided by the application, and a Transport Services system's policies and heuristics.

4.2.2. Racing

- o **Protocol Option Racing:** Protocol Racing is the act of attempting to establish, or scheduling attempts to establish, multiple Protocol Stacks that differ based on the composition of protocols or the options used for protocols.
- o **Path Racing:** Path Racing is the act of attempting to establish, or scheduling attempts to establish, multiple Protocol Stacks that differ based on a selection from the available Paths.
- o **Endpoint Racing:** Endpoint Racing is the act of attempting to establish, or scheduling attempts to establish, multiple Protocol Stacks that differ based on the specific representation of the Remote Endpoint and the Local Endpoint, such as IP addresses resolved from a DNS hostname.

5. IANA Considerations

RFC-EDITOR: Please remove this section before publication.

This document has no actions for IANA.

6. Security Considerations

The Transport Services architecture does not recommend use of specific security protocols or algorithms. Its goal is to offer ease of use for existing protocols by providing a generic security-related interface. Each provided interface mimics an existing protocol-specific interface provided by supported security protocols. For example, trust verification callbacks are common parts of TLS APIs. Transport Services APIs will expose similar functionality. Clients must take care to use security APIs appropriately. In cases where clients use said interface to provide sensitive keying material, e.g., access to private keys or copies of pre-shared keys (PSKs), key

use must be validated. For example, clients should not use PSK material created for ESP with IETF-QUIC, and clients must not use private keys intended for server authentication as a key for client authentication. Moreover, unlike certain transport features such as TFO or ECN which can fall back to standard configurations, Transport Services systems must not permit fallback for security protocols. For example, if a client requests TLS, yet TLS or the desired version are not available, its connection must fail. Clients are responsible for implementing protocol or version fallback using a Transport Services API if so desired.

7. Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644334 (NEAT).

This work has been supported by Leibniz Prize project funds of DFG - German Research Foundation: Gottfried Wilhelm Leibniz-Preis 2011 (FKZ FE 570/4-1).

Thanks to Stuart Cheshire, Josh Graessley, David Schinazi, and Eric Kinnear for their implementation and design efforts, including Happy Eyeballs, that heavily influenced this work.

8. Informative References

[draft-brunstrom-taps-impl]

"Implementing Interfaces to Transport Services", n.d..

[draft-trammell-taps-interface]

"An Abstract Application Layer Interface to Transport Services", n.d..

[I-D.ietf-taps-minset]

Welzl, M. and S. Gjessing, "A Minimal Set of Transport Services for TAPS Systems", draft-ietf-taps-minset-01 (work in progress), February 2018.

[I-D.pauly-taps-transport-security]

Pauly, T., Rose, K., and C. Wood, "A Survey of Transport Security Protocols", draft-pauly-taps-transport-security-01 (work in progress), January 2018.

[RFC8095] Fairhurst, G., Ed., Trammell, B., Ed., and M. Kuehlewind, Ed., "Services Provided by IETF Transport Protocols and Congestion Control Mechanisms", RFC 8095, DOI 10.17487/RFC8095, March 2017, <<https://www.rfc-editor.org/info/rfc8095>>.

Authors' Addresses

Tommy Pauly (editor)
Apple Inc.
One Apple Park Way
Cupertino, California 95014
United States of America

Email: tpauly@apple.com

Brian Trammell (editor)
ETH Zurich
Gloriastrasse 35
8092 Zurich
Switzerland

Email: ietf@trammell.ch

Anna Brunstrom
Karlstad University

Email: anna.brunstrom@kau.se

Godred Fairhurst
University of Aberdeen
Fraser Noble Building
Aberdeen, AB24 3UE
Scotland

Email: gorry@erg.abdn.ac.uk
URI: <http://www.erg.abdn.ac.uk/>

Colin Perkins
University of Glasgow
School of Computing Science
Glasgow G12 8QQ
United Kingdom

Email: csp@csperkins.org

Philipp S. Tiesel
TU Berlin
Marchstrasse 23
10587 Berlin
Germany

Email: philipp@inet.tu-berlin.de

Chris Wood
Apple Inc.
One Apple Park Way
Cupertino, California 95014
United States of America

Email: cawood@apple.com