YANG Model for Transmission Control Protocol (TCP) Configuration
draft-scharf-tcpm-yang-tcp-03

Abstract

This document specifies a YANG model for TCP on devices that are
configured by network management protocols. The YANG model defines
groupings for fundamental parameters that can be modified in many TCP
implementations. The model includes definitions from YANG Groupings
for TCP Client and TCP Servers (I-D.ietf-netconf-tcp-client-server).
The model is NMDA (RFC 8342) compliant.

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 May 7, 2020.

Copyright Notice

Copyright (c) 2019 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
The Transmission Control Protocol (TCP) [RFC0793] is used by many applications in the Internet, including control and management protocols. Therefore, TCP is implemented on network elements that can be configured via network management protocols such as NETCONF [RFC6241] or RESTCONF [RFC8040]. This document specifies a YANG [RFC7950] 1.1 model for configuring TCP on network elements that support YANG data models, and is Network Management Datastore Architecture (NMDA) [RFC8342] compliant. This document includes definitions from YANG Groupings for TCP Clients and TCP Servers [I-D.ietf-netconf-tcp-client-server]. The model focuses on fundamental and standard TCP functions that are widely implemented. The model can be augmented to address more advanced or implementation-specific TCP features.

Many protocol stacks on internet hosts use other methods to configure TCP, such as operating system configuration or policies. Many TCP/IP stacks cannot be configured by network management protocols such as NETCONF [RFC6241] or RESTCONF [RFC8040] and they do not use YANG data models. Yet, such TCP implementations often also have means to
configure the parameters listed in this document. All parameters
defined in this document are optional.

This specification is orthogonal to Management Information Base (MIB)
for the Transmission Control Protocol (TCP) [RFC4022]. The TCP
Extended Statistics MIB [RFC4898] is also available, and there are
MIBs for UDP Management Information Base for the User Datagram
Protocol (UDP) [RFC4113] and Stream Control Transmission Protocol
(SCTP) Management Information Base (MIB) [RFC3873]. It is possible
to translate a MIB into a YANG model, for instance using Translation
of Structure of Management Information Version 2 (SMIv2) MIB Modules
to YANG Modules [RFC6643]. However, this approach is not used in
this document, as such a translated model would not be up-to-date.

There are also other related YANG models. Examples are:

- Application protocol models may include TCP parameters, for
  example in case of BGP YANG Model for Service Provider Networks
  [I-D.ietf-idr-bgp-model].

- TCP header attributes are modeled in other models, such as YANG
  Data Model for Network Access Control Lists (ACLs) [RFC8519] and
  Distributed Denial-of-Service Open Thread Signaling (DOTS) Data
  Channel Specification [I-D.ietf-dots-data-channel].

- TCP-related configuration of a NAT (e.g., NAT44, NAT64,
  Destination NAT, ...) is defined in A YANG Module for Network
  Address Translation (NAT) and Network Prefix Translation (NPT)
  [RFC8512] and A YANG Data Model for Dual-Stack Lite (DS-Lite)
  [RFC8513].

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and
"OPTIONAL" in this document are to be interpreted as described in BCP
14 [RFC2119] [RFC8174] when, and only when, they appear in all
capitals, as shown here.

3. Model Overview

3.1. Modeling Scope

TCP is implemented on many different system architectures. As a
result, there are many different and often implementation-specific
ways to configure parameters of the TCP protocol engine. In
addition, in many TCP/IP stacks configuration exists for different
scopes:
- Global configuration: Many TCP implementations have configuration parameters that affect all TCP connections. Typical examples include enabling or disabling optional protocol features.

- Interface configuration: It can be useful to use different TCP parameters on different interfaces, e.g., different device ports or IP interfaces. In that case, TCP parameters can be part of the interface configuration. Typical examples are the Maximum Segment Size (MSS) or configuration related to hardware offloading.

- Connection parameters: Many implementations have means to influence the behavior of each TCP connection, e.g., on the programming interface used by applications. A typical example are socket options in the socket API, such as disabling the Nagle algorithm by TCP_NODELAY. If an application uses such an interface, it is possible that the configuration of the application or application protocol includes TCP-related parameters. An example is the BGP YANG Model for Service Provider Networks [I-D.ietf-idr-bgp-model].

- Policies: Setting of TCP parameters can also be part of system policies, templates, or profiles. An example would be the preferences defined in the TAPS interface An Abstract Application Layer Interface to Transport Services [I-D.ietf-taps-interface].

There is no ground truth for setting certain TCP parameters, and traditionally different implementation have used different modeling approaches. For instance, one implementation may define a given configuration parameter globally, while another one uses per-interface settings, and both approaches work well for the corresponding use cases. Also, different systems may use different default values.

The YANG model defined in this document includes definitions from the YANG Groupings for TCP Clients and TCP Servers [I-D.ietf-netconf-tcp-client-server]. Similar to the base model, this specification defines YANG groupings. This allows reuse of these groupings in different YANG data models. It is intended that these groupings will be used either standalone or for TCP-based protocols as part of a stack of protocol-specific configuration models.

In addition to configuration of the TCP protocol engine, a TCP implementation typically also offers access to operational state and statistics. This includes amongst others:
o Statistics: Counters for the number of active/passive opens, sent and received segments, errors, and possibly other detailed debugging information

o TCP connection table: Access to status information for all TCP connections

o TCP listener table: Information about all TCP listening endpoints

Similar to the TCP MIB [RFC4022], this document also specifies a TCP connection table.

TODO: A future version of this document may include statistics equivalent to the TCP MIB [RFC4022]:

o active-opens
o passive-opens
o attempt-fails
o establish-resets
o currently-established
o in-segments
o out-segments
o retransmitted-segments
o in-errors
o out-resets

3.2. Basic TCP Configuration Parameters

There are a number of basic system parameters that are configurable on many TCP implementations, even if not all TCP implementations may indeed have exactly all these settings. Also, the syntax, semantics and scope (e.g., global or interface-specific) can be different in different system architectures.

The following list of fundamental parameters considers both TCP implementations on hosts and on routers:

o Keepalives (see also [I-D.ietf-netconf-tcp-client-server])
* Idle-time (in seconds): integer
  * Probe-interval (in seconds): integer
  * Max-probes: integer
  - Maximum MSS (in byte): integer
  - FIN timeout (in seconds): integer
  - SACK (disable/enable): boolean
  - Timestamps (disable/enable): boolean
  - Path MTU Discovery (disable/enable): boolean
  - ECN
    - Enabling (disable/passive/active): enumeration

TCP-AO is increasingly supported on routers and also requires
configuration.

Some other parameters are also common but not ubiquitously supported,
or modeled in very different ways:

  - Delayed ACK timeout (in ms)
  - Initial RTO value (in ms)
  - Maximum number of retransmissions
  - Window scaling
  - Maximum number of connections

TCP can be implemented in different ways and design choices by the
protocol engine often affect configuration options. In a number of
areas there are major differences between different software
architectures. As a result, there are not many commonalities in the
corresponding configuration parameters:

  - Window size: TCP stacks can either store window state variables
    (such as the congestion window) in segments or in bytes.
  - Buffer sizes: The memory management depends on the operating
    system. As the size of buffers can vary over several orders of
magnitude, very different implementations exist. This typically influences TCP flow control.

- Timers: Timer implementation is another area in which TCP stacks may differ.

- Congestion control algorithms: Many congestion control algorithms have configuration parameters, but except for fundamental properties they often tie into the specific implementation.

This document only models fundamental system parameters that are configurable on many TCP implementations, and for which the configuration is reasonably similar.

3.3. Model Design

[[Editor's node: This section requires further work.]]

This document extends the YANG model "ietf-tcp-common" defined in [I-D.ietf-netconf-tcp-client-server]. The intention is to define YANG groupings for TCP parameters so that they can be used in different YANG models.

As an example for the configuration of SACK, a YANG model could import the YANG model "ietf-tcp-common" as well as the model defined in this document as follows:

```yaml
module example-tcp {
  namespace "http://example.com/tcp";
  prefix tcp;

  import ietf-tcp {
    prefix tcp;
  }

  import ietf-tcp-common {
    prefix tcpcmn;
  }

  container example-tcp-config {
    description "Example TCP stack configuration";

    uses tcpcmn:tcp-common-grouping;
    uses tcp:tcp-sack-grouping;
  }
}
```
3.4. Tree Diagram

This section provides a abridged tree diagram for the YANG module defined in this document. Annotations used in the diagram are defined in YANG Tree Diagrams [RFC8340].

```
module: ietf-tcp
   +--rw tcp!
      +--rw connections
         ...
```

3.5. Example Usage

[[Editor's note: This section is TBD.]]

4. TCP Configuration YANG Model

Editor's note: How to use ietf-tcp-common as basis? For instance, is the tcp-system-grouping therein needed?

<CODE BEGINS> file "ietf-tcp@2019-11-04.yang"

```yaml
module ietf-tcp {
   yang-version "1.1";
   namespace "urn:ietf:params:xml:ns:yang:ietf-tcp";
   prefix "tcp";

   import ietf-tcp-client {
      prefix "tcpc";
   }
   import ietf-tcp-server {
      prefix "tcps";
   }
   import ietf-tcp-common {
      prefix "tcpcmn";
   }
   import ietf-inet-types {
      prefix "inet";
   }

   organization
      "IETF TCPM Working Group";

   contact
      "WG Web:  <http://tools.ietf.org/wg/tcpm>
      WG List:  <tcpm@ietf.org>

      Authors: Michael Scharf (michael.scharf at hs-esslingen dot de)
```
This module focuses on fundamental and standard TCP functions that widely implemented. The model can be augmented to address more advanced or implementation specific TCP features.

revision "2019-11-04" {

description
"Initial Version";
reference
"RFC XXX, TCP Configuration.";
}

// Features
feature server {

description
"TCP Server configuration supported.";
}

feature client {

description
"TCP Client configuration supported.";
}

// TCP-AO Groupings

grouping mkt {
  leaf options {
    type binary;
    description
"This flag indicates whether TCP options other than TCP-AO are included in the MAC calculation. When options are included, the content of all options, in the order present, is included in the MAC, with TCP-AO’s MAC field zeroed out. When the options are not included, all options other than TCP-AO are excluded from all MAC calculations (skipped over, not zeroed).

Note that TCP-AO, with its MAC field zeroed out, is always included in the MAC calculation, regardless of the setting of this flag; this protects the indication of the MAC length as well as the key ID fields (KeyID, RNextKeyID). The option flag applies to TCP options in both directions (incoming and outgoing segments).";
    reference
"RFC 5925: The TCP Authentication Option.";
  }
}
leaf key-id {
  type uint8;
  description "TBD";
}

leaf rnext-key-id {
  type uint8;
  description "TBD";
}

description
"A Master Key Tuple (MKT) describes TCP-AO properties to be associated with one or more connections.";
}

grouping ao {
  leaf enable {
    type boolean;
    default "false";
    description
      "Enable support of TCP-Authentication Option (TCP-AO).";
  }

  leaf current-key {
    type binary;
    description
      "The Master Key Tuple (MKT) currently used to authenticate outgoing segments, whose SendID is inserted in outgoing segments as KeyID. Incoming segments are authenticated using the MKT corresponding to the segment and its TCP-AO KeyID as matched against the MKT TCP connection identifier and the MKT RecvID. There is only one current-key at any given time on a particular connection.

      Every TCP connection in a non-IDLE state MUST have at most one current_key specified.";
    reference
      "RFC 5925: The TCP Authentication Option.";
  }

  leaf rnext-key {
    type binary;
    description
      "The MKT currently preferred for incoming (received) segments, whose RecvID is inserted in outgoing segments as
RNExtKeyID.

Each TCP connection in a non-IDLE state MUST have at most one rnext_key specified.

reference
"RFC 5925: The TCP Authentication Option.";

leaf-list sne {
  type uint32;
  min-elements 1;
  max-elements 2;
  description
  "A pair of Sequence Number Extensions (SNEs). SNEs are used to prevent replay attacks. Each SNE is initialized to zero upon connection establishment.";
  reference
  "RFC 5925: The TCP Authentication Option.";
}

leaf-list mkt {
  type binary;
  min-elements 1;
  description
  "One or more MKTs. These are the MKTs that match this connection's socket pair.";
  reference
  "RFC 5925: The TCP Authentication Option.";
  description
  "Authentication Option (AO) for TCP.";
  reference
  "RFC 5925: The TCP Authentication Option.";
}

// TCP general configuration groupings

grouping tcp-mss-grouping {
  description "Maximum Segment Size (MSS) parameters";

  leaf tcp-mss {
    type uint16;
    description
    "Sets the max segment size for TCP connections.";
  }
}

// grouping tcp-mss-grouping

grouping tcp-mtu-grouping {

description "Maximum Transfer Unit (MTU) discovery parameters";

leaf tcp-mtu-discovery {
    type boolean;
    default false;
    description
        "Turns path mtu discovery for TCP connections on (true) or
        off (false)";
}
} // grouping tcp-mtu-grouping

grouping tcp-sack-grouping {
    description "Selective Acknowledgements (SACK) parameters";

    leaf sack-enable {
        type boolean;
        description
            "Enable support of Selective Acknowledgements (SACK)";
    }
} // grouping tcp-sack-grouping

grouping tcp-timestamps-grouping {
    description "Timestamp parameters";

    leaf timestamps-enable {
        type boolean;
        description
            "Enable support of timestamps";
    }
} // grouping tcp-timestamps-grouping

grouping tcp-fin-timeout-grouping {
    description "TIME WAIT timeout parameters";

    leaf fin-timeout {
        type uint16;
        units "seconds";
        description
            "When a connection is closed actively, it must linger in
            TIME-WAIT state for a time 2xMSL (Maximum Segment Lifetime).
            This parameter sets the TIME-WAIT timeout duration in
            seconds.";
    }
} // grouping tcp-fin-timeout-grouping

} // grouping tcp-ecn-grouping
description "Explicit Congestion Notification (ECN) parameters";

leaf ecn-enable {
  type enumeration {
    enum disable;
    enum passive;
    enum active;
  }
  description
    "Enabling of ECN.";
}
} // grouping tcp-ecn-grouping

// augment statements

container tcp {
  presence "The container for TCP configuration.";
  description
    "TCP container.";
}

container connections {
  list connection {
    key "local-address remote-address local-port remote-port";

    leaf local-address {
      type inet:ip-address;
      description
        "Local address that forms the connection identifier.";
    }

    leaf remote-address {
      type inet:ip-address;
      description
        "Remote address that forms the connection identifier.";
    }

    leaf local-port {
      type inet:port-number;
      description
        "Local TCP port that forms the connection identifier.";
    }

    leaf remote-port {
      type inet:port-number;
      description
        "Remote TCP port that forms the connection identifier.";
    }
}

5. IANA Considerations

5.1. The IETF XML Registry

This document registers two URIs in the "ns" subregistry of the IETF XML Registry [RFC3688]. Following the format in IETF XML Registry [RFC3688], the following registrations are requested:
5.2. The YANG Module Names Registry

This document registers a YANG module in the YANG Module Names registry YANG - A Data Modeling Language [RFC6020]. Following the format in YANG - A Data Modeling Language [RFC6020], the following registrations are requested:

- **name:** ietf-tcp
- **namespace:** urn:ietf:params:xml:ns:yang:ietf-tcp
- **prefix:** tcp
- **reference:** RFC XXXX

6. Security Considerations

The YANG module specified in this document defines a schema for data that is designed to be accessed via network management protocols such as NETCONF [RFC6241] or RESTCONF [RFC8040]. The lowest NETCONF layer is the secure transport layer, and the mandatory-to-implement secure transport is Secure Shell (SSH) [RFC6242]. The lowest RESTCONF layer is HTTPS, and the mandatory-to-implement secure transport is TLS [RFC8446].

The Network Configuration Access Control Model (NACM) [RFC8341] provides the means to restrict access for particular NETCONF or RESTCONF users to a preconfigured subset of all available NETCONF or RESTCONF protocol operations and content.

7. References

7.1. Normative References

[I-D.ietf-netconf-tcp-client-server]


[RFC9388]  Mealling, M., "The IETF XML Registry", BCP 81, RFC 3688, 
DOI 10.17487/RFC3688, January 2004, 

the Network Configuration Protocol (NETCONF)", RFC 6020, 
DOI 10.17487/RFC6020, October 2010, 

and A. Bierman, Ed., "Network Configuration Protocol 
(NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, 

[RFC6242]  Wasserman, M., "Using the NETCONF Protocol over Secure 
Shell (SSH)", RFC 6242, DOI 10.17487/RFC6242, June 2011, 

RFC 7950, DOI 10.17487/RFC7950, August 2016, 

Protocol", RFC 8040, DOI 10.17487/RFC8040, January 2017, 

[RFC8174]  Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 
2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, 

BCP 215, RFC 8340, DOI 10.17487/RFC8340, March 2018, 

Access Control Model", STD 91, RFC 8341, 
DOI 10.17487/RFC8341, March 2018, 

[RFC8342]  Bjorklund, M., Schoenwaelder, J., Shafer, P., Watsen, K., 
and R. Wilton, "Network Management Datastore Architecture 
(NMDA)", RFC 8342, DOI 10.17487/RFC8342, March 2018, 

Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018, 

7.2. Informative References

[I-D.ietf-dots-data-channel]
Boucadair, M. and R. K, "Distributed Denial-of-Service
Open Threat Signaling (DOTS) Data Channel Specification",
draft-ietf-dots-data-channel-31 (work in progress), July
2019.

[I-D.ietf-idr-bgp-model]
Jethanandani, M., Patel, K., Hares, S., and J. Haas, "BGP
YANG Model for Service Provider Networks",
draft-ietf-idr-bgp-model-07 (work in progress), October 2019.

[I-D.ietf-taps-interface]
Trammell, B., Welzl, M., Enghardt, T., Fairhurst, G.,
Pauly, "An Abstract Application Layer Interface to
Transport Services",
draft-ietf-taps-interface-04 (work in progress), July 2019.

[RFC3873] Pastor, J. and M. Belinchon, "Stream Control Transmission
Protocol (SCTP) Management Information Base (MIB)",
RFC 3873, DOI 10.17487/RFC3873, September 2004,

the Transmission Control Protocol (TCP)", RFC 4022,
DOI 10.17487/RFC4022, March 2005,

[RFC4113] Fenner, B. and J. Flick, "Management Information Base for
the User Datagram Protocol (UDP)", RFC 4113,
DOI 10.17487/RFC4113, June 2005,

Extended Statistics MIB", RFC 4898, DOI 10.17487/RFC4898,

[RFC6643] Schoenwaelder, J., "Translation of Structure of Management
Information Version 2 (SMIv2) MIB Modules to YANG
Modules", RFC 6643, DOI 10.17487/RFC6643, July 2012,
internet-draft              tcp configuration              november 2019


Appendix A. Acknowledgements

Michael Scharf was supported by the StandICT.eu project, which is funded by the European Commission under the Horizon 2020 Programme.

The following persons have contributed to this document by reviews:
Mohamed Boucadair

Appendix B. Changes compared to previous versions

Changes compared to draft-scharf-tcpm-yang-tcp-02

o Initial proposal of a YANG model including base configuration parameters, TCP-AO configuration, and a connection list

o Editorial bugfixes and outdated references reported by Mohamed Boucadair

o Additional co-author Mahesh Jethanandani

Changes compared to draft-scharf-tcpm-yang-tcp-01

o Alignment with [I-D.ietf-netconf-tcp-client-server]

o Removing backward-compatibility to the TCP MIB

o Additional co-author Vishal Murgai

Changes compared to draft-scharf-tcpm-yang-tcp-00

o Editorial improvements
Authors’ Addresses

Michael Scharf
Hochschule Esslingen - University of Applied Sciences
Flandernstr. 101
Esslingen  73732
Germany

Email: michael.scharf@hs-esslingen.de

Vishal Murgai
Cisco Systems Inc

Email: vmurgai@cisco.com

Mahesh Jethanandani
VMware

Email: mjethanandani@gmail.com