

TSVWG
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2011

D. Wing
P. Natarajan
Cisco
October 25, 2010

**Happy Eyeballs: Trending Towards Success with SCTP
draft-wing-tsvwg-happy-eyeballs-sctp-02**

Abstract

This document describes how to seamlessly migrate HTTP from running over TCP to running over SCTP, without negative impact to the user experience.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 28, 2011.

Copyright Notice

Copyright (c) 2010 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 (<http://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](#) [3](#)
- [2. Notational Conventions](#) [3](#)
- [3. Problem Statement](#) [3](#)
- [4. HTTP Client Recommendations](#) [4](#)
- [5. Additional Considerations](#) [6](#)
 - [5.1. Additional Network and Host Traffic](#) [6](#)
 - [5.2. Abandon Non-Winning Connections](#) [6](#)
 - [5.3. Flush or Expire Cache](#) [6](#)
 - [5.4. Multiple Interfaces](#) [7](#)
- [6. Security Considerations](#) [7](#)
- [7. Acknowledgements](#) [7](#)
- [8. IANA Considerations](#) [7](#)
- [9. References](#) [7](#)
 - [9.1. Normative References](#) [7](#)
 - [9.2. Informational References](#) [7](#)
- [Authors' Addresses](#) [8](#)

1. Introduction

HTTP ("The Web") is one of the most visible and time-critical applications that is used by nearly every Internet user. Research shows that web downloads over TCP suffer from head-of-line (HOL) blocking and poor response times on lossy networks. A multistreamed transport such as SCTP reduces HOL blocking and improves user response times as compared to TCP [[Natarajan](#)]. SCTP provides further improvements on networks with high latency, low bandwidth, or high loss -- typical of today's wireless and wide-area networks. For these reasons there is interest in running HTTP over SCTP [[I-D.natarajan-http-over-sctp](#)].

In order to successfully transition to SCTP, it is necessary that the SCTP connection establishment time is nearly identical (or lower) as compared to TCP. Due to IPv4 NAT and firewalls, lack of SCTP running on servers, and lack of a DNS-based mechanism indicating SCTP server support, SCTP cannot be tried, allowed to timeout, and then TCP tried -- because such a delay will be unacceptable to users.

This document suggests a mechanism for applications to quickly determine if SCTP or TCP is the most optimal transport protocol to use with a particular server.

Once SCTP (or TCP) is successful, the application trends towards preferring that transport protocol for subsequent connections to that server. Thus, repeated use of the application does not cause repeated SCTP (or TCP) probes.

The application recommendations in this document are primarily for HTTP clients ("web browsers") and may also be helpful for other time-sensitive applications.

2. Notational Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

3. Problem Statement

It is important that the same URI and hostname be used for SCTP and TCP. That is, "sctp.example.com" or "www.sctp.example.com" is not viable for successful deployment of SCTP. Using separate namespaces causes namespace fragmentation and reduces the ability for users to share URIs and hostnames, and complicates printed material (e.g.,

advertising) that include the URI or hostname.

Unlike IPv6 which has an AAAA record, there is no DNS query that indicates a host supports SCTP [[RFC4960](#)], and the HTTP URI scheme is not extensible to support an SRV query that could provide such support. Even if there was, it isn't possible to determine if a middlebox, such as a firewall or a NAT, would block the SCTP connection.

Thus, the same URI as used for TCP must be accessible via SCTP. The only way to accomplish that is to try connecting using SCTP.

4. HTTP Client Recommendations

To provide fast connections for users, HTTP clients should make connections quickly over various technologies, automatically tune itself to avoid flooding the network with unnecessary connections (i.e., for technologies that have not made successful connections), and occasionally flush its self-tuning.

Due to the proliferation of NATs on the IPv4 Internet the best success for SCTP can be achieved by attempting both native SCTP connections and SCTP-over-UDP [[I-D.tuexen-sctp-udp-encaps](#)] connections.

For SCTP the following parameters are used:

SWAIT: Application-wide wait time for an SCTP association attempt to complete. Default value of 50ms is RECOMMENDED.

PREF: This denotes per-destination transport preference. Possible values are "TCP", "SCTP", and "BOTH". Default value of "BOTH" is RECOMMENDED.

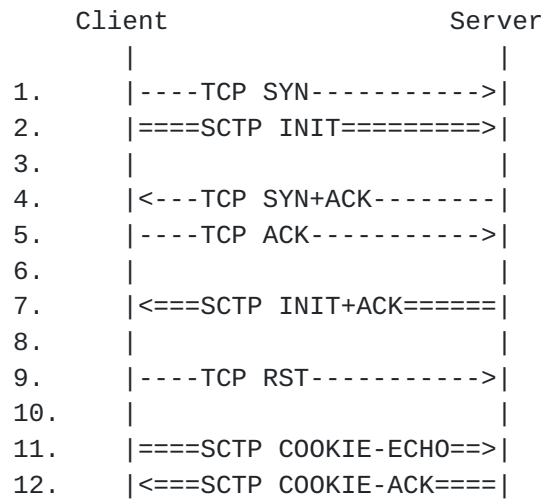


Figure 1: Message flow

In the above diagram, the client sends a TCP SYN and SCTP INIT at the same time (1). The client receives TCP SYN+ACK (4) before receiving SCTP INIT+ACK (7) which of course could arrive in the opposite order. Upon receiving the SCTP INIT+ACK (7), the client knows that SCTP is working and abandons the TCP connection before sending TCP user data (9). User data can be sent with the SCTP COOKIE-ECHO (11).

The HTTP client starts several threads in order to minimize the user-noticeable delay ("dead time") during the connection attempts. The client starts one or more threads based on the following logic:

If ((PREF == BOTH) or (PREF == SCTP)) start thread 1. If making a connection using IPv4 start thread 2.

If ((PREF == BOTH) or (PREF == TCP)) start thread 3.

thread 1 (SCTP):

- * Attempt to connect using SCTP (i.e., send SCTP INIT)

thread 2 (SCTP over UDP):

- * Attempt to connect using SCTP over UDP (i.e., send SCTP INIT over UDP)

thread 3 (TCP):

- * Attempt to connect using TCP

If an SCTP association attempt was made by a thread, the HTTP client waits for at least K ms; K = max(SWAIT, time taken for the TCP

connection to complete). If the TCP connection finishes during this wait period, the HTTP client MAY choose TCP for the current HTTP transfer but MUST wait until K ms to figure if the SCTP association can be completed.

If the HTTP client did not choose TCP during the wait period and the SCTP association completes successfully, the HTTP client prefers SCTP over TCP connections and abandons the TCP connection.

After a connection is successful, we want to adjust the per-destination preference for this destination. It is not recommended to dynamically adjust the application-wide default value for SWAIT. If the SCTP association was successful, set destination's `PREF="SCTP"`, else set `PREF="TCP"`.

5. Additional Considerations

This section discusses considerations and requirements that are common to new technology deployment.

5.1. Additional Network and Host Traffic

Additional network traffic and additional server load is created due to these recommendations and mitigated by application-wide and per-destination timer adjustments. The procedures described in this document retain a quality user experience while transitioning from TCP to SCTP. The quality user experience benefits the user but to the detriment of the network and server that are serving the user.

5.2. Abandon Non-Winning Connections

It is RECOMMENDED that the non-winning connections be abandoned, even though they could be used to download content.

5.3. Flush or Expire Cache

Because every network has different characteristics (e.g., a middlebox that permits or blocks SCTP) the SCTP parameters (SWAIT and PREF) SHOULD be reset to their default whenever the host is connected to a new network ([[cx-osx](#)], [[cx-win](#)]). However, in some instances the application and the host are unaware the network connectivity has changed (e.g., when behind a NAT) so it is RECOMMENDED that per-destination values expire after 10 minutes of inactivity.

5.4. Multiple Interfaces

Interaction of the suggestions in this document with multiple interfaces is for further study.

6. Security Considerations

[[Placeholder.]]

7. Acknowledgements

The mechanism described in this paper was inspired by Stuart Cheshire's discussion at the IAB Plenary at IETF72, and how ICE [[RFC5245](#)] tests connectivity to different transports.

Thanks to Fred Baker, Jeff Kinzli, Christian Kuhtz, and Iljitsch van Beijnum for fostering the creation of this document.

Thanks to Scott Brim, Stig Venaas, and Andrew Yourtchenko for providing feedback on the document.

8. IANA Considerations

This document has no IANA actions.

9. References

9.1. Normative References

- [I-D.tuexen-sctp-udp-encaps]
Tuexen, M. and R. Stewart, "UDP Encapsulation of SCTP Packets", [draft-tuexen-sctp-udp-encaps-05](#) (work in progress), July 2010.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4960] Stewart, R., "Stream Control Transmission Protocol", [RFC 4960](#), September 2007.

9.2. Informational References

- [I-D.natarajan-http-over-sctp]
Natarajan, P., Amer, P., Leighton, J., and F. Baker,

"Using SCTP as a Transport Layer Protocol for HTTP",
[draft-natarajan-http-over-sctp-02](#) (work in progress),
July 2009.

[Natarajan]

Natarajan, P., "Leveraging Innovative Transport Layer Services for Improved Application Performance", February 2009, <<http://www.cis.udel.edu/~amer/PEL/poc/pdf/NatarajanPhDdissertation.pdf>>.

[RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", [RFC 5245](#), April 2010.

[cx-osx] Adium, "AIHostReachabilityMonitor", June 2009, <https://bugzilla.redhat.com/show_bug.cgi?id=505105>.

[cx-win] Microsoft, "NetworkChange.NetworkAvailabilityChanged Event", June 2009, <<http://msdn.microsoft.com/en-us/library/system.net.networkinformation.networkchange.networkavailabilitychanged.aspx>>.

Authors' Addresses

Dan Wing
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
USA

Email: dwing@cisco.com

Preethi Natarajan
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
USA

Email: prenatar@cisco.com

