Why is SCTP needed given TCP and UDP are widely available?

ISOC MEMBER BRIEFING #17

June, 2004

Authors: Randall Stewart, Paul D. Amer

Definition

Stream Control Transmission Protocol (SCTP) [RFC2960] is an end-to-end transport protocol that provides services heretofore unavailable from either of the workhorse transport protocols that have supported the Internet for more than twenty years: reliable, connection-oriented TCP [RFC793], or unreliable, connectionless UDP [RFC768].

Why is SCTP needed given TCP and UDP are widely available?

In 1998 an IETF working group (SIGTRAN) was formed to design a mechanism for reliably transporting call control signaling over the Internet. SIGTRAN's goal was to create an IP complement to the telephone switching's SS7 network. During SIGTRAN's work, two key problems surfaced in the use of TCP:

- Head-of-line blocking - a problem where sending independent messages over an order-preserving TCP connection causes delivery of messages sent later to be delayed within a receiver's transport layer buffers until an earlier lost message is retransmitted and arrives. These later messages often establish independent telephone calls. For call control signaling, the delay on later messages caused critical call control timers to expire thus resulting in undesirable call setup failures.

- Multihoming - where a host with multiple points of attachment to the Internet, for redundancy purposes, does not want to wait for a routing convergence (often on the order of minutes) to communicate critical messages to its peer communication endpoint. For call control signaling, such delay is unacceptable when an alternate available path exists. A TCP connection only binds a single point of attachment at either end point.

Expanded Coverage from ISOC

In-depth articles, papers, links and other resources on a variety of topics are available from the ISOC site at:
www.isoc.org/internet/issues

Relevant IETF RFCs

RFC 3257 SCTP Applicability Statement
RFC 3309 Stream Control Transmission Protocol (SCTP) Checksum Change
RFC 2960 Stream Control Transmission Protocol

For More Information

S. Ladha, P. Amer. "Improving multiple file transfers using SCTP multistreaming," *Proceedings IPCCC* 2004, 4/04 (to appear)

A. Caro, J. Iyengar, P. Amer, G. Heinz, S. Ladha, K. Shah. "SCTP: A standard for robust Internet data transport," *IEEE Computer*, 36(11), 11/03, 56-63

A. Caro, P. Amer, R. Stewart. "Retransmission policies with transport layer multihoming," *Proceedings ICON* 2003, Sydney, 9/03, 255-260

Related Organizations

www.ietf.org

SCTP vs TCP vs UDP

Services/Features	SCTP	TCP	UDP
Connection-oriented	yes	yes	no
Full duplex	yes	yes	yes
Reliable data transfer	yes	yes	no
Partial-reliable data transfer	optional	no	no
Ordered data delivery	yes	yes	no
Unordered data delivery	yes	no	yes
Flow control	yes	yes	no
Congestion control	yes	yes	no
ECN capable	yes	yes	no
Selective ACKs	yes	optional	no
Preservation of message boundaries	yes	no	yes
Path MTU discovery	yes	yes	no
Application PDU fragmentation	yes	yes	no
Application PDU bundling	yes	yes	no
Multistreaming	yes	no	no
Multihoming	yes	no	no
Protection against SYN flooding attacks	yes	no	n/a
Allows half-closed connections	no	yes	n/a
Reachability check	yes	yes	no
Psuedo-header for checksum	no (uses vtags)	yes	yes
Time wait state	for vtags	for 4-tuple	n/a

Considering these problems, SIGTRAN began work on a new transport protocol to carry its call control signaling over IP. Simultaneously, the IETF transport Area Directors (Scott Bradner and Vern Paxson) recognized the value of solving these problems for a wider audience. They expanded the scope of the work from a small, dedicated protocol for a specific task (SIGTRAN) to a general-purpose transport protocol that other applications could use as well. Within this larger scope, SCTP was born. SCTP also includes the following enhancements:

www.rfc-editor.org

About the Authors

Paul D. Amer received the BS degree summa cum laude in Mathematics from SUNY Albany in 1974, and the MS and PhD degrees in CIS in 1976 and 1979, respectively, from The Ohio State University. Since 1979, Prof. Amer has been at the University of Delaware's CIS Department. As Director of UD's Protocol Engineering Laboratory (PEL), Prof. Amer and his students currently focus on (1) innovative transport layer services and protocols to improve Internet performance, and (2) data compression in multimedia.

Randall Stewart joined Cisco's Internet Technologies Division (ITD) in July of 2000. He is involved in various product solution architectures and core IOS technology. Currently focusing on:

- Cisco IOS integration of SCTP
- Cisco IOS integration

Multistreaming - SCTP supports multiple, independent logical streams of messages within an SCTP association[1]. Each message sent over an SCTP association is assigned to a particular stream. All data within a stream is delivered in order with respect to other data in that stream. Data in different streams have no order constraints. SCTP's resulting parallel ordered streams provide a specific instance of 'partial ordered' delivery. It is SCTP's multistreaming service that circumvents the head-of-line blocking problem discussed above. Multistreaming has been found to facilitate FTP'ing multiple files, say for system backups or mirror site downloads. Multistreaming also is appropriate for applications that wish to multiplex related, yet independent data streams (e.g., voice, text, video) over a single end-to-end association rather than open multiple TCP connections, one for each stream.

- SCTP's solution for Multihoming - each of the two endpoints during an SCTP association setup can specify multiple points of attachment. Having multiple interfaces allows data to be automatically sent to alternate addresses when failures occur, and most importantly, without the application even knowing a lower level failure occurred. Such fault tolerance is unavailable for TCP, which binds each endpoint to a single interface. Should either endpoint interface or the link to the interface fail, all TCP connections bound to that interface would need to timeout and abort, thus forcing the application(s) to reestablish the connection(s), and accurately pick up from where the aborted connection(s) left off.

- Message Orientation - In TCP, data sent between two endpoints is a stream of bytes. If needed, an application must provide message framing. In SCTP, message boundaries are preserved. If an application sends a 100-byte message, the peer application will receive all 100 bytes in a single read: no more, no less. UDP provides a message-oriented service, but without SCTP's reliability.

- Un-Ordered Service - In TCP, all messages are reliably delivered to a receiving application in exactly the order used by the sending application. In addition to ordered message service (and parallel ordered service discussed above), SCTP offers the reliable delivery of messages with no order constraints. UDP provides unordered service, but again without SCTP's reliability. Unordered reliable delivery will be useful for many applications, in particular disk over LAN services (iSCSI, RDMA, etc.) where the application already provides ordering, and does not want TCP's overhead (and performance penalty) to preserve order.

- Extensibility - a TCP packet is limited to 40 bytes for options. In contrast, SCTP packets can be expanded through the use of Tag-Length-Value (TLV) fields. Embedded within SCTP's TLV structures are compatibility handling procedures so that implementations remain interoperable even when one supports a more advanced feature set than another.

- Heartbeat/Keep-alive - SCTP has a default keep-alive function. Regular heartbeats validate reachability of peer addresses, and help maintain a Round Trip Time (RTT) estimate for each alternate

of RBSCP (Rate Based Satellite Control Protocol)

Previous to Cisco, Mr. Stewart worked at Motorola, NYNEX S&T, Nortel and AT&T Communication focusing on fault tolerance and call control signaling protocols, with special focus on integrating SS7 and IP.

Acknowledgments
The ISOC Member Briefing series is made possible through the generous assistance of ISOC's Platinum Program Sponsors: Afilias, APNIC, ARIN, Microsoft, Ripe NCC, and SIDA. *More information on the Platinum Sponsorship Program*: http://www.isoc.org/isoc/membership/platinum.shtml

About the Background Paper Series

Published by:
The Internet Society
1775 Wiehle Avenue, Suite 102
Reston, Virginia 20190 USA
Tel: +1 703 326 9880
Fax: +1 703 326 9881
4, rue des Falaises
CH-1205 Geneva
Switzerland
Tel: +41 22 807 1444
Fax: +41 22 807 1445
Email: info@isoc.org
Web: http://www.isoc.org/
Series Editor: Martin Kupres

address.

- Message Time-to-Live - SCTP has an option to specify a message's time-to-live. This feature allows a sending application to specify how long a message is useful. If this time expires before the message can be reliably communicated to the receiver, the sending SCTP entity can stop trying, effectively 'undo'ing or skipping the message. This type of reliability is called "partial reliability". Partial reliability is of interest, for example, to the mobile community (military and commercial) and the online gaming community where the current "location or environment" status is only good for a short period, and then outdated by a revised status. In this situation, being able to discard the stale data will save on bandwidth utilization, while providing the application with better service in the face of loss or congestion.

- Syn cookies - SCTP uses a four-way handshake with a signed cookie. A receiver of a new SCTP association setup message maintains no state (i.e., reserves no resources) until after the initiator proves it is at the IP address from which it claims to be setting up the association. Syn cookies thus prevent using IP spoofing for the known denial of service attack termed SYN flooding [RFC2827].

- · Stronger checksum - SCTP uses a 32-bit end-to-end checksum (CRC32C) proven to be mathematically stronger than the 16-bit ones-complement sum used by TCP and UDP. SCTP's better checksum provides stronger verification that a message passes end-to-end without bit errors going undetected.

- Advanced TCP services - Recent advances in TCP, such as SACK [RFC2018], Appropriate Byte Counting [RFC3465], and Explicit Congestion Notification [RFC3168], have been built into SCTP by design instead of retrofitting. The user of SCTP is assured these features are present as compared to TCP versions where one must validate that both sides of a connection have implemented all appropriate RFCs.

Deployment

SCTP became an RFC in October 2000. As with any new Internet technology, ease of deployment becomes the barrier to widespread use. Currently, all major operating systems have an SCTP implementation. For some, such as BSD Unix, Linux, and Solaris, SCTP comes as part of the OS, or has a patch (or package) that can be added. For other platforms, such as Windows, a third party package must be purchased to add SCTP's advanced feature set. A comprehensive list of SCTP implementations can be found at www.sctp.org. Six workshops since June 2000 have been organized to test interoperability of these implementations.

The authors predict increased demand and use of SCTP as more applications find benefits from SCTP's richer set of services. Eventually all OS vendors will respond to customer demands, and supply SCTP out of the box. Until such time, many applications may hesitate to move to SCTP.

Summary

SCTP offers several advantages. Most important at the application level are SCTP's multistreaming of boundary-preserved messages, and multihoming's added fault tolerance. For example, SCTP is a natural candidate to support telephone signaling over the Internet as well as other message-oriented applications. Applications that require high degrees of fault tolerance, such as online banking and stock market transaction exchanges, will benefit from SCTP's multihoming.

A subtle advantage of SCTP's multihoming is that, if widely used by applications where redundancy is critical, multihoming can reduce the widespread routing table growth discussed in ISOC brief #3. Organizations that move their mission critical applications to SCTP can take advantage of multihoming, and not cross-advertise routes (i.e., effectively creating hosts routes in IP routing tables). Using SCTP would thereby yield significant reduction to current Internet routing table sizes.
Multihoming within an SCTP association has another interesting side effect in that an association simultaneously spans both IPv6 and IPv4 addresses. Thus SCTP facilitates the transition of the Internet from IPv4 to IPv6. The wider implications of SCTP's "dual stack" nature are undetermined.

By its design, SCTP provides for extensibility, which when used judiciously, will provide SCTP with years of usefulness to the Internet community. Enhancements will be easier to introduce while still maintaining inter-operability amongst prior implementations. The authors expect SCTP's many new features will attract current and as yet undiscovered applications as wider scale deployment occurs. Any hesitancy to use SCTP will be overcome by market pressure and the profitable engineering practice of choosing the right transport layer tool for the job.

[1] SCTP uses the term association to refer to a connection between two endpoints. A new term was introduced because an SCTP association is broader in scope than a TCP connection. An SCTP association may bind multiple IP addresses at each endpoint whereas TCP binds precisely one address for each endpoint.