

Stream Control Transmission Protocol Performance Analysis in Disk Local Area Network Services

Lynet O. Sibiyia¹, Japheth O. Ondiek²

¹Computer Science Department, Faculty of Science, Egerton University,
P.O Box 536, Egerton, 20115, Kenya.

²Mathematics and Computer Science Department, Kabanga University,
P. O. BOX 20301, Kericho, 20200, Kenya.

Abstract – Performance is the driving force for effective computer network utilization in the current and future trend in the network communication. SCTP is a new transport protocol defined by the IETF intended for the transport of voice signaling data (SS7) system over IP networks. SCTP with very recent defined extensions is growingly considered for other application scenarios too. In this paper, performance analysis of Disk Local Area Network Service – iSCSI, running over SCTP is presented. SCTP is believed to have many advantages in comparison to TCP in the TCP/IP transport layer. Based on our experimental results, recommendations with regard to SCTP usage and its ability to function as a TCP alternative are made.

Keywords: LAN, SCTP, TCP, UDP, iSCSI, Initiator, Target, SAN, VLAN

1. INTRODUCTION

This work intent to obtain experimental results indicating SCTP's benefits by comparing SCTP with existing TCP protocol and analyze which gains SCTP brings. It is important to understand why it is necessary to implement SCTP in various services, instead of relying on the current TCP protocol. To achieve this, one must have a deeper knowledge of the protocols, the weaknesses inherent in TCP and the new features found in SCTP. PSTN signaling over the IP network is an example for which deficiencies of Transmission Control Protocol are relevant. Whilst this application prompted the SCTP development, it is found to be suitable for other computer applications as well. Hence the need to investigate disk Local Area Network services running over SCTP. This section begins with an overview of TCP and its major weaknesses followed by an overview of SCTP and the core features available then followed by an overview of iSCSI as a disk Local Area Network service.

2. RESEARCH QUESTIONS

1. Why SCTP for LAN Disk Services?

The expected contributions of the research to scientific knowledge are as follows:

- a) As the Internet develops rapidly the demands for the lower-layer transport protocols is high. TCP cannot measure up to the demands of new applications such as Migration from

PSTN to Next Generation Network, Telephony signaling messages #7 for PSTN on the Internet in reference to its shortcomings.

- b) Storage-interconnect hardware (SCSI, Fiber Channel) is about to be replaced by Ethernet-based hardware and upper-level protocols (ULPs) i.e. iSCSI. These protocols can take advantage of RDMA feature provided by network interface.
- c) The Next Generation Network which entails voice, data, and media such as video encapsulated into packets, is converging to all IP-based networks. ISDN sets, Cellular mobile phones, SIP terminals, can be configured in multi-homed environments by installing two or more network interfaces to support different protocols. SCTP support of overlay multicast due to its multi-streaming and multi-homing features, is ideal for these environments.
- d) Support multiple network interfaces to boost throughput.

3.0 RELATED WORK

TCP achieves reliability and strict ordered byte-stream abstraction in communicating end-points by increasing flow of bytes at a pace corresponding in size to RTTs. Flow and control algorithm employed by TCP AIMD in which the window increases slowly while the network investigates the availability of extra bandwidth then reduce quickly due to congestion (*Jani et al, Intel Corporation*). The robustness of TCP's flow/congestion control is due to continued patching of TCP for the last 10 years. However, TCP has deficiencies in various areas such as integrity or robustness checks, susceptibility to DoS attacks, and extra.

3.1 TCP/IP

TCP/IP being a successful protocol has come through exponential increase on the Internet without substantial changes, which is a strong evidence of its robustness and ability to scale. Despite anything to the contrary, various groups found TCP unsuitable. Multimedia users found the congestion control procedures

employed by TCP too limiting. In Wireless and mobile communication TCP reacts poorly to jitter and packet losses (www.cs.ucd.ie). This has resulted to the creation of modern communication protocols and measures for computer networking.

3.2 SCTP

SCTP was developed by IETF SIGTRAN working group to handle the deficiencies of Transport Control Protocol for VOIP over Internet Protocol networks (Stewart, Xie 2001). SCTP (RFC2960) provides services not found in transport protocols (TCP - RFC793 and UDP - RFC768) that supported the Internet for over two decades. SCTP is defined in (Stewart et al., 2000, 2003; Stone et al., 2002), as a connection-oriented, message-based reliable communication protocol running on a connectionless, datagram-oriented service i.e. Internet Protocol, and supportive of multi-homed end-points. It provides an end-user with flexible data delivery methods such as ordered or un-ordered delivery services. SCTP services in reference to their usefulness in an iSCSI are cited below.

- i. Multi-streaming - support multi-homed hosts with multiple IP addresses
- ii. Streams can be used to a greater advantage by allotting high priority to inter-process communication by use of flow and congestion control per association. (Heinz 2004).
- iii. A stream can either be assigned an in-order or immediate delivery to the upper layer. Un-ordered delivery reduces jitter. Remote Direct Memory Access communication doesn't require ordering from the transport layer being used (Stewart 2004).
- iv. Multi-homing - support multiple streams of packets
- v. Association establishment robustness (Jani et al, Intel Corporation) is a feature that iSCSI can take advantage of. The association establishes a verification tag used for all subsequent data transfer. The 32-bit CRC and heartbeat mechanism contribute to SCTP robustness.
- vi. DoS attack is avoided by SCTP's four-way handshake mechanism.

3.4 Internet Small Computer System Interface (iSCSI)

iSCSI is a Storage Area Network who's popularity has increased as a low-cost option to Fibre Channel. The iSCSI protocol is a new architecture standardized by the IETF (Satran et al 2004) and was designed to transport SCSI application data over TCP/IP networks. It maps SCSI family of protocols onto TCP/IP enabling convergence of storage traffic on to standard TCP/IP fabrics (Chadalapaka, 2003). Various scholars have

realized that disk LAN services can take on an important role for users accessing data storage on the Internet. Services like iSCSI play a dramatic change in the evolution of storage area networks (SANs). With 10GigE, iSCSI can scale to 10 Gigabit for business applications that require high business dealings (<http://www.Starwindsoftware.com>).

3.5 iSCSI Protocol Model

Figure 3.1 depicts the iSCSI layer interfacing to the operating system's SCSI. iSCSI layer contains encapsulated SCSI commands, accompanying data and reporting status capability. When the operating system or application requires a read/write operation, the SCSI CDB is encapsulated for transport across a gigabit link and delivered to the target (www.agilent.co.jp)

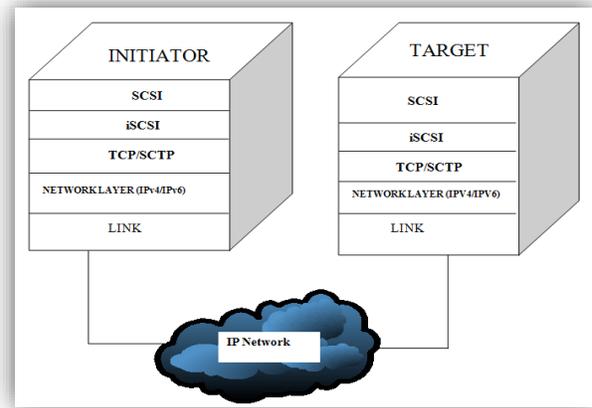


Figure 3.1: iSCSI protocol layering model

4.0 METHODOLOGY APPROACH

4.1 Proposed Methodological Approach

The scientific approach is chosen for this research because the subject under study is theory-oriented whereby tests have to be carried out involving an experiment. Characteristics of this approach (tools, Model, etc.) are shown below:

4.1.1 Experimental Design: Tools employed

- a) Ubuntu v13.04, lksctp-tools (sctp_darn utility), Wireshark Simulator v1.10.1, Open-iSCSI (Target and Initiator).
- b) Toshiba Laptop: Intel Core 2 Duo T8100 2.1GHz processor, Intel X3100 Integrated Graphics, 2GB RAM DDR2, 160GB hard drive.

- c) Desktop - Compaq Presario: CPU type : Pentium ® Dual Core , CPU speed : 3.20GHz/800MHz, Cache RAM(L2): 2048KB, RAM: 2GB, Hard disk : 350GB
- d) VLAN Switch: *ProCurve Switch 4104gl*: Address table size of 8,000 entries, Routing/switching capacity of 18.3Gbps, Supports a maximum of 96 10/100 or 80 Gigabit External I/O ports
- e) CISCO Router 2900 series:
- f) Microsoft Visio Professional 2007 for diagramming

4.1.2 Parameters

1. iSCSI: Performance Factors

- a) Workload characteristics: Sequential streaming vs random access, Read/write, large/small transfers
- b) Network characteristics: Speed (100, 1000, 10000 Mbps), Distance (LAN, WAN).

2. iSCSI Performance Metrics

- a) Throughput/Bandwidth utilization - Application bytes transferred in seconds.
- b) CPU utilization (low Central Processing Unit utilization is desirable).
- c) Transaction rate (high processing time for read/write operations is desirable).

4.2 Network Prototype

Figure 4.1 shows a full-scale working model of the experiment setup

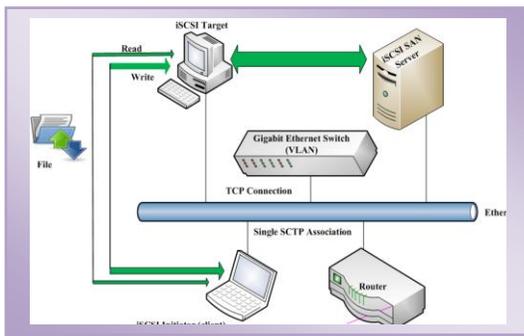


Figure 4.1.: Network Prototype diagram

4.3 CONCEPTUAL DESIGN

The conceptual design of this investigation is theory-oriented in the sense that we want to test the performance of the SCTP protocol as a medium for the iSCSI protocol being a disk LAN service which essentially was designed to run over TCP/IP. The main objective involves experimenting iSCSI protocol on these two transmission protocol layers at different instances with similar data and hardware.

By carrying out this experiment we would like to answer the question “is it possible that the iSCSI running over SCTP has better performance compared to iSCSI running over TCP/IP?” How this is done is shown in Figure 4.2.

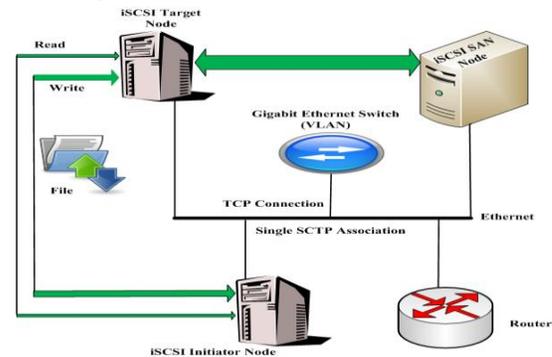


Figure 4.2: Conceptual Model

4.3.1 Data/Input

We have used Expect script in Figure 4.3 for a small data sample (Bytes) streamed from the iSCSI initiator to the iSCSI target over the two protocols at at different times and observe what happens at both ends for each protocol. According to script 1, data is written and read as it is streamed after every 0.5 seconds a hundred times and thereafter the transaction ends. For a large data sample, data is written and read as it is streamed after every 0.1 seconds a 10000 times.

```
#!/usr/bin/expect
set lhost 172.16.85.15
set lport 132
set rhost 172.16.85.19
set rport 3260
set count 100
spawn sctp_darn -H $lhost -P $lport -h $rhost -p $rport -s
expect ">"
for {set i 1} {$i < $count} {incr i 1} {
send -- "lynet\r"
expect ">"
}
```

Figure 4.3: Expect Script - Small data sample

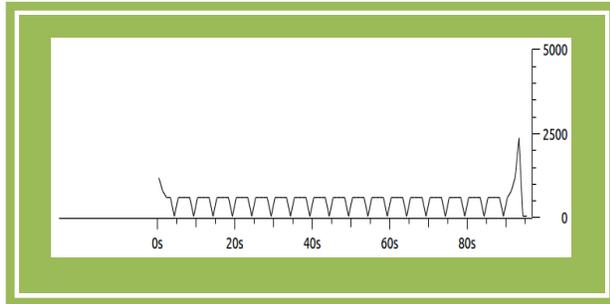


Figure 5.1.3 (b) TCP isograph bits/tick

Throughput Measurement

Throughput is the most important concept to be understood when examining TCP. It refers to an average amount of output relative to input transferred through a system from input to output. Throughput is typically expressed as bits per second “bps” or bytes per second “B/s” with a proper prefix. In this dissertation “kbps” or “Mbps” which is equivalent to 1000 bits or 1000000 bits per second is used.

SCTP and TCP Conversations A→B

Throughput generated for, say, A→B or B→A is measured. Table 5.1.1 depicts SCTP conversations A→B with throughput of 969.01 bps in duration of 59.6237s whereas Table 5.1.2 depicts TCP Conversations A→B with throughput of 303.51 bps in duration of 90.1442s. This indicates that the SCTP has a higher throughput.

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A→B	Bytes A→B	Packets A←B	Bytes A←B	Rel Start	Duration	bps A→B
172.16.85.19	3260	172.16.85.15	132	103	7222	103	7222	0	0	0.000000000	59.6237	969.01

Table 5.1.1: SCTP Conversations A→B

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A→B	Bytes A→B	Packets A←B	Bytes A←B	Rel Start	Duration	bps A→B
172.16.85.19	51852	172.16.85.15	iscsi-target	38	3420	38	3420	0	0	1.170810000	90.1442	303.51

Table 5.1.2: TCP Conversations A→B

TSNs and SACK – chunk s Analysis

The course of the TNS is important in analyzing congestion control issues. Clicking on any of the TSNs in Figure 5.1.4 will show the TSN coordinates and pressing twice consecutively selects the represented main window frame plus the Acks that have been advertised in the Gap Ack Blocks. The zoomed graph of the same is shown in Figure 5.1.5. Here, the graph depicts leaving TSN packets (black arrow) and

Cumulative TSNs (red arrow), arriving as a function of time.

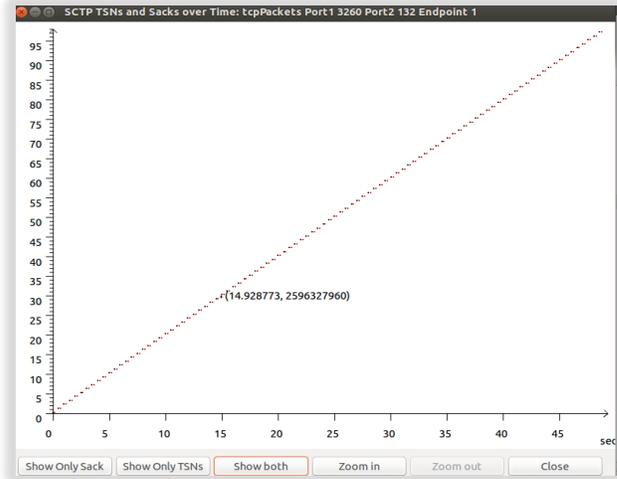


Figure 5.1.4: TSNs and SACKs graph over time

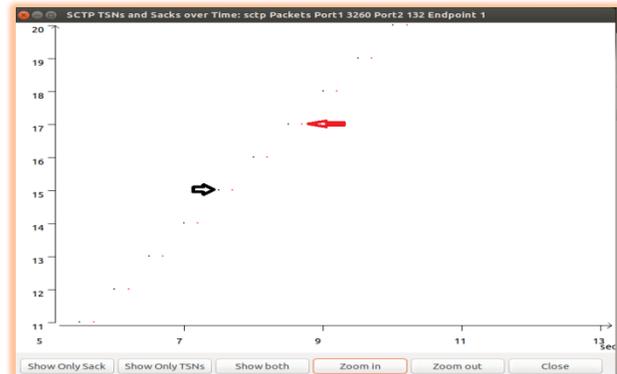


Figure 5.1.5: Magnified TSNs and SACKs g graph

The Advertised Receiver Window and Byte Transmission

To analyze the flow control feature of SCTP, the advertised receiver window size is necessary. In Figure 5.1.6 the DATA – chunks movement is observed with the vertical axis showing the number of bytes while the horizontal axis, showing time in seconds. Figure 5.1.7 shows a magnified SCTP Data and Advertised Receiver Window in which the dark gray lines represent the a_rwnd, showing relationship between the a_rwnd moving upward and the stored data at the receiver.

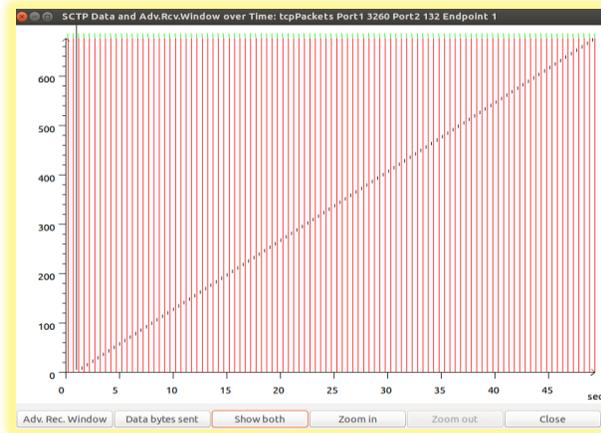


Figure 5.1.6: Sctp Data and Advertised Receiver Window

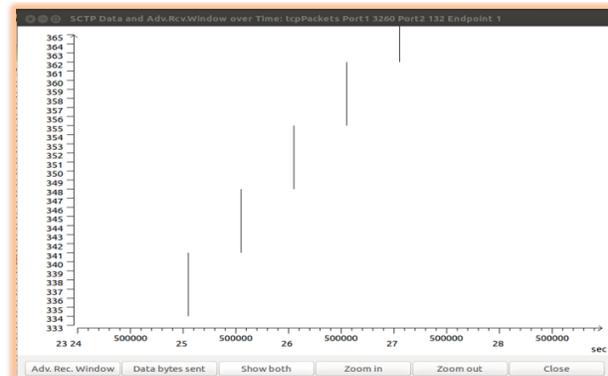


Figure 5.1.7: Magnified Sctp Data and Adv. Receiver Window

SCTP Associations

An association is the relationship between two SCTP endpoints extend across multiple IP addresses. SCTP has a specific variable referred to as the verification tag (VTag) unique for each end-point. An association is identified by the following: source address to remote VTag and the destination address VTag. The SCTP handshake forms basis for an association. Table 5.1.3 shows SCTP associations.

Port 1	Port 2	No of Packets	Checksum	No of Errors	Data Chunks	Data Bytes	VTag 1	VTag 2
3260	132	215	CRC32C	0	97	679	2382808270	3121242772

Table 5.1.3: SCTP Associations

When the iSCSI client streams to the iSCSI server, the following results for TCP protocol are realized:

TCP Throughput graph

The throughput graph shows the throughput of the TCP stream vs. time. The graph in Figure 5.1.8 indicates time slot (shown with orange arrow) where nothing is being sent from the initiator node to the target node. Fast-Retransmission and Retransmission areas are circled in brown.

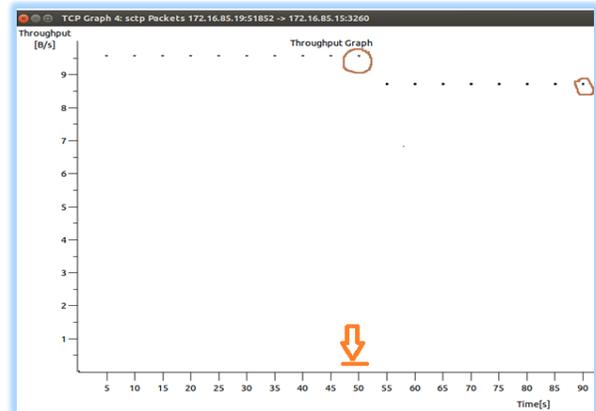


Figure 5.1.8: TCP Throughput graph

The sequence graph Figure 5.1.9 shows the maximum and minimum throughput. The space in between the two is the receiver window size. The Time-Sequence for iSCSI client to iSCSI server connection reveals step-ladder like pattern after 5 seconds. Transmission Control Protocol's timers increase exponentially. This results to the graph increasing in steps (Seggelmann, 2012). Data in packets sent from the initiator node don't show any error messages during these periods. The connection at all times remained established.

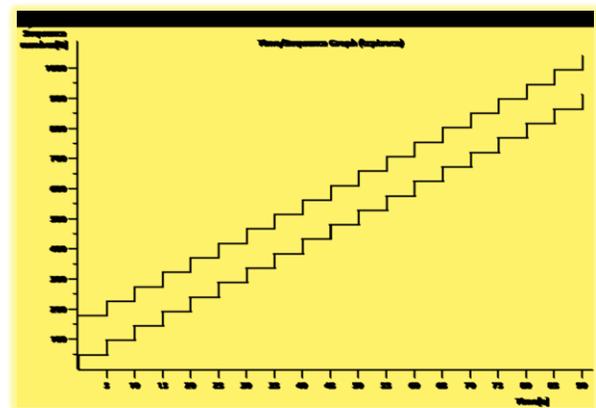


Figure 5.1.9: TCP Sequence Graph B->A

TCP Round Trip Time Graph

In Figure 5.1.10, the Round Trip Time graph for ACKs depicts many dots that are clustered closer towards the x-axis indicating a consistent response time, but there are a few dots that are steeply climbing towards the

top. The dots circled in green are during the time the Fast Retransmits occur, and show maximum length in Round Trip Time graph.

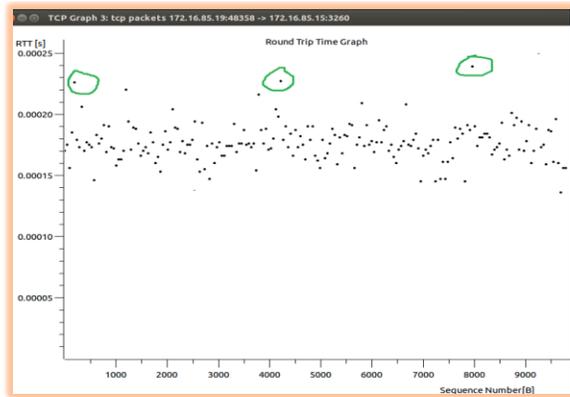


Figure 5.1.10: Round Trip Time Graph

5.2 Scenario B – Large Data Sample (2,657,184 bytes)

SCTP and TCP iographs – Measuring Application Utilization

Figures 5.2.1a and 5.2.1b show the iSCSI application bandwidth usages for SCTP and TCP. For the TCP protocol, this is the receiver advertised window size. The y-axis displays bits/tick while the x-axis displays time in seconds. This shows the bit-rate that is used by the iSCSI application. From the SCTP iograph, the read/write operation transfers 2.6Kbps whereas the TCP iograph transfers 1.25Kbps. The blue arrow in SCTP iograph shows where very little traffic moving from the initiator node to the target node. SCTP protocol performs better.

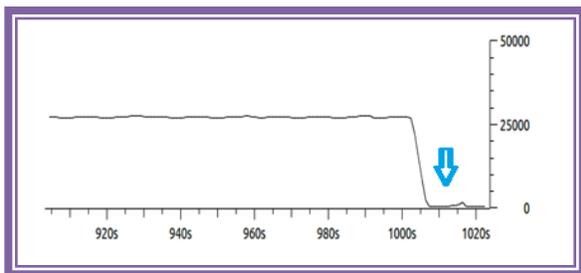


Figure 5.2.1 (a) SCTP iograph bits/tick

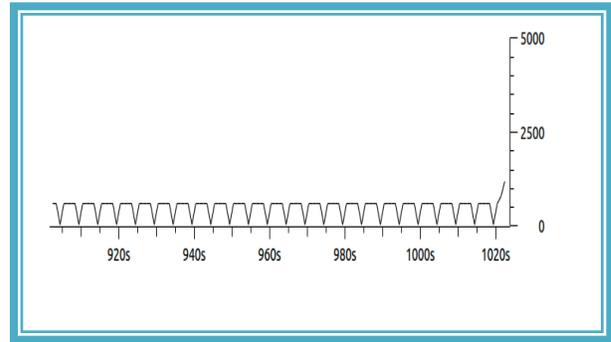


Figure 5.2.1 (b) TCP iograph bits/tick

SCTP and TCP Conversations A→B

Table 5.2.1 shows SCTP conversations A→B with a throughput of 24141.03 bps in duration of 1014.2041s while Table 5.2.2 depicts TCP Conversations A→B with a throughput of 288.95 bps in duration of 1021.6322s. In this scenario the TCP protocol totally under-performed in terms of throughput.

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A→B	Bytes A→B	Packets A←B	Bytes A←B	Rel Start	Duration	bps A→B	bps A←B
172.16.85.19	3260	172.16.85.15	132	5056	3060492	5056	3060492	0	0	0.000003000	1014.2041	24141.03	

Table 5.2.1: SCTP Conversations A→B

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A→B	Bytes A→B	Packets A←B	Bytes A←B	Rel Start	Duration	bps A→B	bps A←B
172.16.85.19	48358	172.16.85.15	iscsi-target	410	36900	410	36900	0	0	0.184488000	1021.6322	288.95	

Table 5.2.2: TCP Conversations A→B

SCTP and TCP Conversations A←B

In Table 5.2.3 the SCTP conversations A←B has a throughput of 2481.20 bps in duration of 1014.2042s while Table 5.2.4 the TCP Conversations A→B has a throughput of 183.00 bps in duration of 1021.6320s. In this scenario the TCP protocol totally under-performed in terms of throughput.

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A→B	Bytes A→B	Packets A←B	Bytes A←B	Rel Start	Duration	bps A→B	bps A←B
172.16.85.19	48358	172.16.85.15	iscsi-target	265	23370	0	0	205	23370	0.184488000	1021.6320	N/A	183.00

Table 5.2.3: SCTP Conversations A←B

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A? B	Bytes A? B	Packets A? B	Bytes A? B	Rel Start	Duration	bps A? B	bps A? B
172.16.85.19	3260	172.16.85.15	13260	5057	314556	0	0	5057	314556	0.000000	1014.2042	N/A	2481.20

Table 5.2.4: TCP Conversations A←B

Figure 5.2.2 plots the TSNs of packets leaving indicated by a **black arrow** and Cumulative TSN of the ACKs indicated by a **red arrow**; arriving at the source.

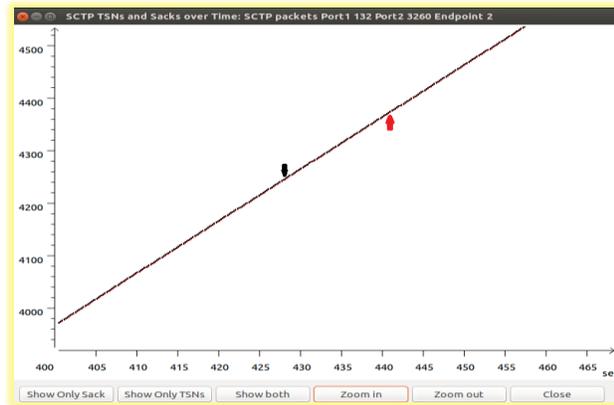
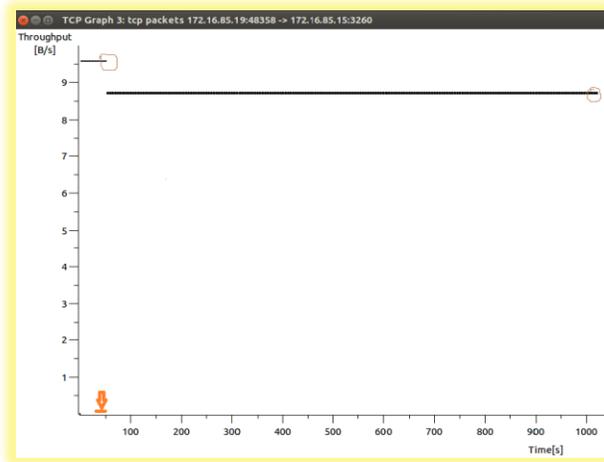


Figure 5.2.2: Magnified TSNs and SACKs

The Advertised (Adv) Receiver (Rcv) Window and Bytes Transmission

Figure 5.2.3: shows SCTP Data and Advertised Receiver Window in which the **purple** line represent the *a_rwnd*.

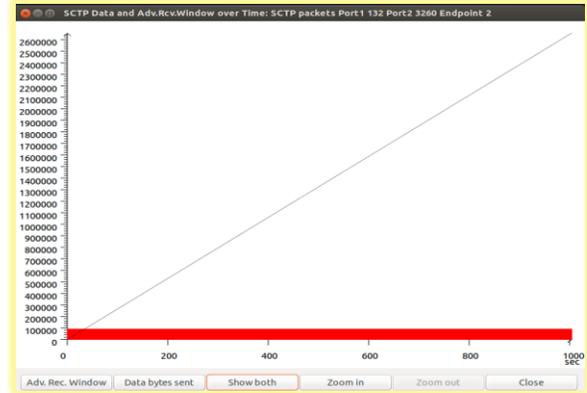


Figure 5.2.3: SCTP Data and Advertised Receiver Window

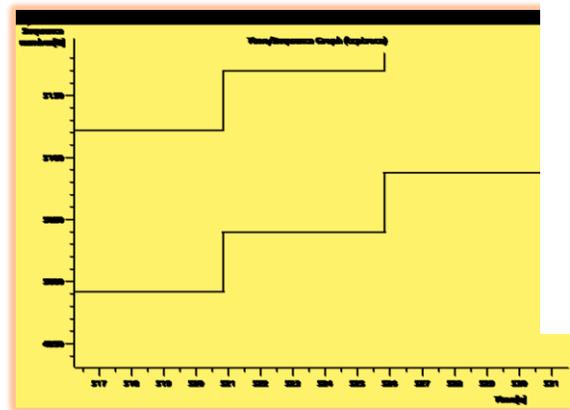


Figure 5.2.4: Magnified Time Sequence tcptrace graph

Figure 5.2.5: Time Sequence tcptrace graph

The magnified Time-Sequence graph figure 5.2.4 for iSCSI initiator to iSCSI target connection is depicted. This time reveals wider step-ladder like pattern as a result of TCP’s timers increasing exponentially. In Figure 5.2.5 the time slot (**orange** arrow) indicates where nothing is being sent from the iSCSI initiator to the iSCSI target. Fast-Retransmission and Retransmission areas are circled in **brown**.

The Round Trip Time graph for ACKs when transferring large data sample indicates many dots clustered closer towards the x-axis indicating a consistent response time (Figure 5.2.6). Climbing upwards are some of the dots circled in **green** occurring during the Fast Retransmission showing maximum length in Round Trip Time.

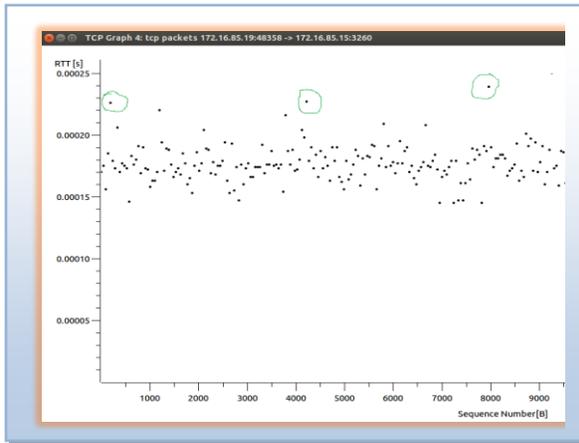


Figure 5.2.6: TCP RTT (Round Trip Time) graph

6.0 DISCUSSION OF FINDINGS

6.1 Discussion of Results

In this research, a comparative study of iSCSI over SCTP and TCP on an Ethernet interface is presented. Two scenarios are simulated, that of small and large data samples streamed from the iSCSI initiator to the iSCSI target at different instances, and observing what happens on each of the two protocols.

The simulations show transmission of different data sizes in the network with no lost packet. With small data (693 bytes) sample the results show that SCTP has better turnout application utilization, although the margin between the two protocols isn't that big. With large data chunks (2,657,184 bytes) transmission, results show that SCTP provides better turnout. From (Ha et al, 2005), their results, show SCTP having better turnout over TCP for a larger data sample. This is also found to be true (Chang et. al, 2007) obtaining similar results. Again, experiment results and analysis obtained by (Nurul et. al. 2006) show better turnout for SCTP in the case of large data chunks transmitted across the network with no lost packet. It can also be seen that application utilization rate is high as well for SCTP.

The bit-rate that is used by the iSCSI application is much higher on the SCTP transport layer compared to that of the TCP transport layer. This is the equivalent of one read/write operation moving across the Ethernet wire. The retransmission mechanisms of SCTP are responsible for the higher throughput achieved when the two protocols share a common network link.

Going by the experiments presented in this research, we concluded that SCTP has the potential to become the successor of TCP, involving the best congestion control algorithms available and offering solutions to TCP problems.

FUTURE WORK

Possible future work will involve investigating the feasibility of testing other disk LAN services on the SCTP transport layer. These will include:

- 1) NAS (Network - Attached Storage);
- 2) AoE (ATA Over Ethernet);
- 3) DAS (Direct-Attached Storage).

REFERENCES

- [1.] N. Jani and K. Kant, "SCTP Performance in Data Center Environments", Intel Corporation
- [2.] R. Stewart, Q. Xie, 2001, "Stream Control Transmission Protocol (SCTP)", a Reference Guide, Addison Wesley, 2001, ISBN: 0-201-72186-4.
- [3.] Stewart et al, 2000, "Stream Control Transmission Protocol".
- [4.] R. Stewart et al. 2003, "Stream control transmission protocol (SCTP) implementer's guide", IETF, Network Working Group.
- [5.] J. Stone, R. Stewart, and D. Otis, 2002, "Stream Control Transmission Protocol (SCTP) Checksum Change".
- [6.] R. Stewart et al. 2004, "Stream Control Transmission Protocol (SCTP) partial reliability extension", RFC 3758.
- [7.] G.J. Heinz & P.D. Amer, 2004, "Priorities in stream transmission control protocol Multi-streaming", SCI 2004.
- [8.] R. Stewart, 2004, "Stream Control Transmission Protocol (SCTP) RDMA DDP Adaptation".
- [9.] J. Satran et al. 2004 "Internet Small Computer Systems Interface (iSCSI)," RFC 3720
- [10.] Mallikarjun Chadalapaka, 2003, "A Study of iSCSI Extensions for RDMA (iSER)", ACM SIGCOMM 2003 Workshops.
- [11.] Seggelmann Robin, 2012, "SCTP: Strategies to Secure End-To-End Communication".
- [12.] Ha et al. 2005, "Performance Comparison of SCTP and TCP over Linux Platform", Springer-Verlag Berlin Heidelberg, Volume 3645, 2005, pp 396-404
- [13.] Chang et al. 2007, "Analysis of FTP over SCTP and TCP in congested network", AIT
- [14.] Nurul Islam MD and Kara, A., 2006, "Throughput Analysis of SCTP over a Multi- homed Association", CIT'06 IEEE International Conference.
- [15.] Noonan. (n.d). www.cs.ucd.ie. Retrieved September 4, 2013, from www.cs.ucd.ie Web site: <http://www.cs.ucd.ie>
- [16.] [Starwindsoftware.com](http://www.Starwindsoftware.com). (n.d.). Retrieved September 1, 2013, from [Starwindsoftware Website: http://www.Starwindsoftware.com](http://www.Starwindsoftware.com)
- [17.] www.agilent.co.jp. (n.d.). Retrieved September 5, 2013, from <http://www.agilent.co.jp>