A Survey of Optimizing the Performance of iSCSI

Shruti Kulkarni[1], Dr. K. N. Narasimha Murthy[2]

[1]PG Scholar, Department of Computer Science and Engineering
Vemana IT, Visvesvaraya Technological University, Belagavi, Karnataka, India.

[2]Professor, Department of Computer Science and Engineering
Vemana IT, Visvesvaraya Technological University, Belagavi, Karnataka, India

Abstract: ***Mobile devices are playing important role in everybody's life. Day to day mobile network, wireless network and mobile devices are becoming more dominant. So there is a requirement in building wireless storage system which can access information efficiently and correctly. Small Computer System Interface (SCSI) is a set of parallel interface standards used to connect scanners, printers, disk drives and other peripheral devices to a computer. iSCSI (internet SCSI) is an important protocol used for linking data storage facilities. iSCSI carries SCSI commands over IP network, and is also used to transmit data over internet and other networks and also enables location-independent storage and retrieval. Optimization of iSCSI parameters is an interesting research topic. Many researches have been made to analyse and implement iSCSI. This paper presents a survey on performance measurement and optimization of iSCSI parameters for mobile networks. The iSCSI parameters are the most suitable way of optimizing the performance, since changing kernel level TCP/IP parameters might negatively affect other applications' performance.***

Keywords: CDB (Command Descriptor Block), iSCSI, PDU (Protocol Data Unit), SCSI, Optimization.

1. Introduction

As the time goes by, mobile devices are going to be used in several areas. Mobile devices have challenges like limited storage capacity, unreliability in connections, data accessibility issues due to varying bandwidths and limited battery. Storage is a place where data is retained as an electromagnetic or optical form which is accessible by processor. Network is a collection of computers and devices which are interconnected by communication channel that facilitates communication and also allows for sharing of resources and information among interconnected devices. Storage Area Network (SAN) is a network which connects servers and storage devices together. Primary purpose of SAN is to transfer data between computer system and storage devices. SAN enhances the storage devices which are accessible to server so that device appears like locally attached device to the operating system. SAN is network of storage device which are not generally accessible through local area network by other devices. Most SAN uses SCSI protocol for communication between server and devices.

1.1 SCSI and iSCSI Concepts

SCSI stands for Small Computer System Interface. SCSI is a concept to put where computers and peripheral devices are physically connected by a set of standards for transferring data. Commands, protocols, electrical and optical interfaces are defined by SCSI standards. A wide range of devices including scanners and printers are connected by SCSI, but it most commonly used for hard disks and tape drives.

SCSI follows client server architecture. In SCSI terminology, clients of SCSI interfaces are called as initiators and logical units of server are called as target. Initiator communicates with the target by issuing a SCSI commands. SCSI commands are sent in the form of Command Descriptor Block (CDB). The CDB is to be composed of a one byte operation code followed by five or more bytes which are made up of command-specific parameters. Target responds to the initiator by returning a status code byte, such as 00h for success, 02h for an error (called a Check Condition), or 08h for busy at the end of command sequence. The initiator usually issues a SCSI request sense command in order to obtain a Key Code Qualifier (KCQ) from the target when the target returns a check condition in response to a command.

SCSI commands are categorised as follows:
1. N (non-data)
2. W(Writing data from initiator to target)
3. R (reading data)
4. B (bidirectional)

A unique SCSI identification number or ID is assigned to each device on the SCSI bus. Devices may include multiple logical units, which are addressed by Logical Unit Number (LUN). Simple devices have just one LUN; more complex devices may have multiple LUNs. In SCSI terminology, an individual I/O device is called a Logical Unit (LU). SCSI is a set of interfaces for requesting services from I/O devices, including hard drives, tape drives, CD and DVD drives, printers and scanners.

Each SCSI command concludes two phases. One is an optional data phase and other is required response phase. In the data phase, information can travel from the initiator to target (e.g., WRITE), target to initiator (e.g., READ), or in both directions. In the response phase, the final status of the operation is returned by the target, including any errors.

1.2Functional Overview of iSCSI

An iSCSI is a protocol which allows SCSI commands to be transferred over TCP/TP. iSCSI is also client-server architecture. Communication takes place between iSCSI initiator and iSCSI target. iSCSI request carries SCSI commands and iSCSI response carries SCSI responses and status. In iSCSI terminology, messages are sent in the form of protocol data unit (PDU). With respect to the initiator, the iSCSI transfer direction is defined. Inbound or incoming transfers are from a target to an initiator, while outbound or outgoing transfers are from an initiator to a target. An iSCSI task is an iSCSI request for which a response is expected.

1.3 Layers and Sessions

The initiator and target actions are specified by the following conceptual layering model and the way in which they relate to transmitted and received Protocol Data Units:

a) The SCSI layer builds/receives SCSI CDB's (Command Descriptor Blocks) and passes/receives them with the remaining command executes parameters to/from.
b) The iSCSI layer that builds/receives iSCSI PDUs and relays/receives them to/from one or more TCP connections; the group of connections form an initiator-target "session".

There can be one or more TCP connections over which communication between the initiator and target can take place. Control messages, SCSI commands, parameters, and data within iSCSI PDUs are carried by a TCP connection. A session is a group of TCP connections that link an initiator with a target. A session is defined by a session ID that is composed of an initiator part and a target part. TCP connections can be added and removed from a session [8]. Connection ID (CID) identifies each connection within a session.

The rest of the paper is arranged as follows. In the next section we first survey previous work in iSCSI performance improvement. The Section 3 details the design of iSCSI, section 4 discusses the comparison analysis and finally, conclusion of the paper is discussed in the section 5.

2. Related Work in iSCSI

As discussed in the previous chapter, iSCSI (internet SCSI) is an important protocol to enable location independent storage access through the TCP/IP network. Many researches and projects have been made to analyse and implement iSCSI. This section details about the previous work carried out on iSCSI. This section is divided into two parts. In the first part; previous work related to performance of iSCSI optimization is considered while in the second part; previous work related to iSCSI is drawn.

2.1 Previous work related to performance of iSCSI:

Yan Gao et al.[2] conducted experiments to optimize the performance of iSCSI on wireless network. Their paper presents a new iSCSl design with the concepts of multiple virtual TCP connections in an iSCSI session and parallel working mechanism in iSCSl layer over wireless LAN 802.11b. Conclusion of test results is that for small I/O request, 2K for example, multiple connection iSCS1 design can achieve 112% throughput improvement compared to normal iSCSI model. The maximum throughput can reach 0.62 MB/s for big I/O (128K), which is closed to the theoretical analysis result. Ja-Own Seo et al.[3]developed MobileStor; aniSCSI based remote storage system in order to address the problems of storage capacity for mobile devices. MobileStor is an iSCSI based remote storage system which provides the allocation of a mass storage space to each mobile client through networks. They conducted experiments which are performed to investigate the best performance values of iSCSI parameters for iSCSI based remote storage system are taken out in CDMA networks in order to realize the access to a remote storage system anytime and anywhere. And after the experiment, they suggested the optimal value of parameters. The experiment results from several test cases concluded that the best values are not the default values specified in the iSCSI standard.

Y Zang et al. [4]thoroughly tested the performance of a common open source component, the Open-iSCSI initiator. The result includes there is a drastic throughput degradation on 100 Mbps networks where the round trip time was more than about 40 ms. They restricted the problem to the TCP send buffer size and tested two methods of setting the TCP send buffer size appropriately. They propose a performance tuning scheme that enables users of Open-iSCSI to achieve significant throughput gains based on their results. Their scheme results in a dramatic throughput jump from 14 Mbps to 70 Mbps on a 100 Mbps link with an RTT of 100 ms. They also modified one of the data structures internal to Open-iSCSI to handle multiple memory pages in a single scatter-gather list entry. This modification resulted in an additional 20% throughput increase on a 100 Mbps link with an RTT of 200 ms.

Yudha Purwanto ST et al. [1] analysed optimization and evaluation of the internet Small Computer Standard Interface iSCSI Storage Area Network (SAN). They implemented test bed for iSCSI SAN optimization using Internet Protocol Multipathing. They analysed the I/O throughput as its write and read for performance improvements. Also they analysed ability of iSCSI SAN to handle broken link (Failover and Failback). Implementation result that internet protocol multipathing enhance performance and reliability of iSCSI SAN system.

Gauger et al. [4] [10] also carried out a simulation study on performance of iSCSI. Their study examines iSCSI throughput and total request write times while varying link RTT, link loss probability, process delays in the iSCSI layer, and I/O request characteristics. One major contribution of their paper is that their simulations are based on statistically realistic network and I/O request models. Although the maximum RTT studied in their model is only 10 ms, it is sufficiently descriptive as they set the link bandwidth to 1 Gbps. As for the results, Y Zang observation of iSCSI performance degradation over long delay links also conforms to their results in the scenario with a single iSCSI session over long RTT links.

Lu et al. [4] [12] accomplished a simulation study of an iSCSI-based storage system with the network simulator ns-2. Their study also includes the performance variance of iSCSI in a simulation setting while varying PDU size, TCP Maximum Segment Size (MSS), and TCP window size.

Aiken et al. [4] [9] carried out measurements of real iSCSI systems. They first determined the performance of two commercial iSCSI initiators - one software-based and one hardware-based. Then they compared these performance results against a storage subsystem based on Fibre Channel.

They noticed that the commercial software iSCSI initiator surpassed the hardware initiator and that the software initiator compared quite “favourably” to Fibre Channel. They also estimated the performance of the Intel iSCSI initiator and target pair in a SAN setting and in a Wide Area Network (WAN) setting.

2.2 Previous Work Related to iSCSI:

Rahul Bedarkar et al. [6] implemented RAM-DISKIO in iSCSI target. RAM-DISKIO is used to give high performance in iSCSI because, operations like read, write on RAM-DISK are less costly in terms of access times than that of Magnetic Disks, and ultimately performance gets increased. So their goal of the research work is to utilize the portion of RAM as a block device (disk), to export RAM-DISK from target to initiator as a Virtual SCSI device and to make use of this exported Virtual SCSI device for storing data. They also wrote the device driver to tackle their aim and also examined it using dd, iostat, bonnie++ tools. Finally, they compared results with FILEIO MODE of iSCSI and in each sample trail RAM-DISKIO MODE outperformed the FILEIO MODE.

Jiang Gua-song et al. [7] implemented the iSCSI out-of-band storage virtualization design and they detail the implementation process. Based on the iSCSI protocol, they established the out-of-band virtualization storage technology, the communication between the initiator and target devices in fully accordance with initiator and target model defined in iSCSI Specifications, but they made a change on the flow to meet the storage virtualization requirements.

3. Design of iSCSI

Figure 3.1 illustrates iSCSI protocol stack. Applications issues I/O requests which go through the SCSI layer, based on the I/O requests, the SCSI driver builds SCSI Command Description Blocks. Then SCSI CDBs are moved to the iSCSI layer. The iSCSI driver in the iSCSI layer assembles the CDBs into Protocol Data Units and sends them over to the TCP/IP layer. Thereafter, the PDUs travel across the network just like any other TCP payload. At the receiver, the iSCSI driver in the iSCSI layer disassembles the received PDUs. The encapsulated CDBs are passed to the SCSI layer. CDBs and the I/O requests are disassembled by the SCSI drive. The I/O requests are performed on the SCSI Logical Unit(s) at the destination.

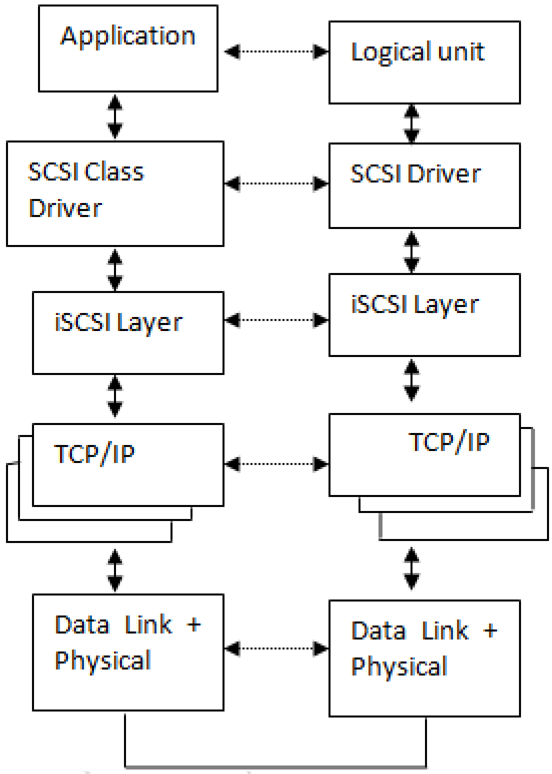

Figure 3.1: iSCSI Protocol Stack

4. Comparison Analysis of Previous Work

In this section, we are doing a comparative analysis of related previous work, depending on the type of network chosen, the area of optimization chosen and the results obtained with respect to iSCSI performance optimization. Since the work done is on different networks, comparing the results may not lead us in the correct direction. Also, the hardware used by each of the works may have profound effect on the performance of iSCSI. It might also be difficult to inculcate the optimizations done in one work into another since the optimizations carried out could be specific to the use case considered by respective works. Thus the objective of a comparative analysis is to understand the areas in which optimization can be achieved without negatively impacting other applications that use the same network interface and settings.

The following note summarizes the analysis: Firstly, with respect to network part, while Yan Gao et al. chose wireless network, Ja-Own Seo et al selected CDMA network. Y Zang et al and Yudha Purwanto preferred WAN and internet multipathing network respectively.

Second comparison parameter is the area of optimization. Yan Gao et al. conducted experiments to optimize the performance of iSCSI on wireless network and they considered iSCSI parameters (I/O request size) as an area of optimization. Ja-Own Seo et al conducted experiments to explore the best performance values of iSCSI parameters for Mobilestor.While Y Zang suggested the strategy for managing TCP send buffer size for open-iSCSI in which his preferred area of optimization was TCP send buffer size, Yudha Purwanto concentrated on IP protocol with multipathing.

The last comparison parameter is result part. The results given by Yan Gao are, for small I/O request, multiple connection iSCSI design can achieve high throughput improvement compared to normal iSCSl model. Ja-Own Seo concluded that best values of iSCSI parameters are not default value specified in the iSCSI standard, for the CDMA networks. The result part of Y Zang includes an approach for managing TCP send buffer size for open-iSCSI. Yudha Purwanto's conclusion includes improvement in iSCSI SAN performance at file read and at file write and that RAID 0 based iSCSI SAN system has significant improvement in multipathing.

5. Conclusion

This section details the results drawn in this survey paper. As discussed in the related work of iSCSI performance and from the discussion on comparison analysis, the inference is that the iSCSI parameters is most suitable way of optimizing the performance, since changing kernel level TCP/IP parameters might negatively affect other applications' performance. From the other related works of iSCSI, we infer that iSCSI performance optimization is best done on RAM-DISK, since RAM-DISK helps to identify performance without introducing the bottleneck of slower storage devices.

References

[1] Tody Ariefianto W, ST., MT, Hendra Wiratama, ST and Yudha Purwanto- "Storage Area Network Based-on Internet Small Computer Standard Interface Optimization Using Internet Protocol Multipathing",2013 International Conference of Information and Communication Technology (ICoICT)0-7803-8783-X/04/$20.00 0 2004 IEEE 303-307.
[2] Yan Gao, Yao-long Zhu, Hui Xiong, Renuga Kanagavelu, Jie Yan, Zhe-jie Liu –"An iSCSI Design Over Wireless Network",0-7803-8783-X/04/$20.00 0 2004 IEEE 60 -65
[3] Ja-Won Seo, Hae-Sun Shin, and Myong-Soon Park)-"Optimizing iSCSI Parameters for Improving the Performance of iSCSI based Mobile Appliance in Wireless Network",0-7695-2552-0/06 $20.00 © 2006 IEEE, 126-131.
[4] Y. Zhang and M.H. MacGregor "Tuning Open-iSCSI for Operation over WAN Links",-07803-8783-X/04/$20.00 0 2004 IEEE 85-97
[5] Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, iSCSI, InfiniBand and FCoE, 2nd Edition, Ulf Troppens, Rainer Erkens, Wolfgang Mueller-Friedt, Rainer Wolafka, Nils Haustein
[6] Chetan Paithane, Rahul Bedarkar, Sushrut Shirole and Satish Kumbhar"RAM-DISKIO Mode in iSCSI Target"2012 International Conference on Information and Network Technology (ICINT 2012), IPCSIT vol. 37 (2012) 1-4
[7] Jiang Guo-song and He Xiao-ling, "Design and implementation of iSCSI Out-of-band storage virtualization"2011 International Conference on Intelligence Science and Information Engineering978-0-7695-4480-9/11 $26.00 © 2011 IEEE 378-381.
[8] RFC3720 - http://www.ietf.org/rfc/rfc3720.txt
[9] S. Aiken, D. Grunwald, A. R. Pleszkun, and J. Willeke. A performance analysis of the iSCSI protocol.In MSS '03: Proceedings of the 20 thIEEE/11 th NASA Goddard Conference on Mass Storage Systems andTechnologies (MSS'03), page 123, Washington, DC, USA, 2003. IEEEComputer Society.
[10] C. Gauger, M. Kohn, S. Gunreben, D. Sass, and S. Perez. "Modeling and performance evaluation of iSCSI storage area networks over TCP/IP-based MAN and WAN networks".In Broadband Networks, 2005.BroadNets 2005. 2nd International Conference on, pages 850 –858 Vol.
[11] J. L. Hufferd. iSCSI: The Universal Storage Connection,volume 1. Addison-Wesley Professional, November 2002.
[12] Y. Lu, F. Noman, and D. H. C. Du. Simulation study of iSCSI-based storage system. In 12th NASA Goddard and 21st IEEE Conference on Mass Storage Systems and Technologies (MSST2004, pages 399–408,2004.

Author Profile

Dr K N Narasimha Murthy is presently working as a professor in the Department of Vemana Institute of Technology Banglore. He has published more than 20 research papers and guiding three phd students in the area of Computer Science and Engineering. His area of interests are Cloud computing, Image processing and Networking

Shruti Kulkarni is a presently a PG scholar at Vemana Institute of Technology. Her fields of interest are Computer Networking and Operating Systems.