Error Recovery Mechanism for iSCSI Protocol Based Mobile NAS Cluster System

Shaikh Muhammad Allayear[1], Sung Soon Park[2], Md. Nawab Yousuf Ali[3], and Ullah Mohammad Hasmat[4]

[1,2,3,4] Dept of Computer Science and Engineering
[1,3] East West University, Bangladesh
[2,4] Anyang University, Korea
{allayear,nawab}@ewubd.edu, {sspark,raju}@anyang.ac.kr

Abstract. In this paper, we proposed an error recovery module for iSCSI protocol based NAS (Network Attached Storage) system in wireless network environment, which ensures reliability and efficiency data transmission.

Keywords: iSCSI Protocol, Network Storage, wireless network.

1 Introduction

A NAS system is usually a specially designed device providing clients with files on a LAN; it is widely used to bridge the interoperability gap with many advantages.However, NAS supports only file I/O protocol such as NFS and CIFS, block-level storage applications are not available on NAS with the performance of NFS and CIFS is only a fraction of the exported storage system. To this day, they continue to have limited network, CPU, memory, and disk I/O resources due to their single server design, which binds one network endpoint to all files in a file system.Now, the iSCSI storage has emerged. iSCSI is an Internet Protocol-based storage standard for linking data storage facilities, and presents storage to servers as disk targets[1], which, from the perspective of the application, appear to be storage attached locally to the server. It presents storage space as virtual block-level devices, operating systems and applications have the ability to put their own file systems on them, which is something not possible with NAS.In current age mobile appliances are going to be used in more area as the time goes by. Due to their mobility, they should be small and use a flash memory to store the data. So it is very difficult to store large data and install large software [2],[3],[4],[5],[6],[7].To alleviate these problems and access mass storage we developed MNAS [6], an iSCSI based NAS Cluster system, for providing the allocation of a mass storage space to each mobile client through networks. The system offers to its users the possibility of keeping large size of data and database in a secure and safe space. But in wireless environment data may losses due to bad channel characteristics, interference or intermittent connectivity due to handoffs (mobile appliance). ISCSI over TCP performance in such networks suffers

V.V. Das, E. Ariwa, and S.B. Rahayu (Eds.): SPIT 2011, LNICST 62, pp. 151–156, 2012.

from significant throughput degradation and [1]very high interactive delays. Therefore, in this paper we propose error recovery module for iSCSI protocol based mobile NAS cluster system.The organization of this paper is as follows. In section 2, we described the motivations of iSCSI data transmission and parameters in wireless environment. In Section 3, error recovery method has described. Experiments and Analysis has described in section 4 and finally we conclude this paper in section 5.

2 Motivation

When we use iSCSI protocol in network with wireless and other lossy links in order to offer mobile NAS services with this situation, it suffers from significant non-congestion-related losses due to several reasons such as a bit errors and handoffs. So mobile NAS services in wireless networks needs several schemes designed to improve the performance of TCP as a try.Communication over wireless link is often characterized by sporadic high bit-error rates, and intermittent connectivity due to handoffs. TCP reacts to packet losses as it would in the wired environment and it drops its transmission window size before re-transmitting packets and resets its retransmission timer. These measures result in an unnecessary reduction in the link's bandwidth utilization, thereby causing a significant degradation in performance in the form of poor throughput and very high interactive delays [6].Since iSCSI runs on a TCP network, it requires new strategies to evade the forceful decrease of transmission rate from TCP congestion control mechanism without changing existing TCP.The iSCSI Data-In PDUs are passed to the TCP layer, since the iSCSI PDU is a SCSI transport protocol over TCP/IP. When the iSCSI Data-In PDU size is greater than the MSS (maximum segment size) of TCP layer, the PDU will be further fragmented into smaller segments. The iSCSI Data-In PDU is generally larger than the MSS in size. If some segments of parts of an iSCSI Data-In PDU are lost due to the high bit error rate in wireless networks, TCP layer would require re-transmitting the segments. At that time all the other segments of those parts of an iSCSI Data-In PDU must wait for being reassembled into an iSCSI Data-In PDU in TCP buffer. It decreases the performance of the system due to the reducing of the available capacity of TCP buffer. The sender can transmit data segments, which are allowed by the receiver using TCP flow control mechanism. In a wireless network with high bit error rate, the more size of iSCSI PDU is increased by the MRDSL (MaxRecvDataSegmentLength) parameter, the more segments would have to wait due to the lost of some segments of parts of an iSCSI PDU in TCP buffer. The performance of the system thus decreases more in both write and read operation. In addition, a wireless link generally becomes a bottleneck portion in an end-to-end TCP connection because of its narrowbandwidth, it also one the cause of performance falling to the storage system for mobile applications.

[1] This research (Grants No.2009-0075576) was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by Ministry of Education, Science and Technology.

[2] This work (Grants No. 47504) was supported by Business for Cooperative R&D between industry, Academy and Research Institute funded Korea small and Medium Business Administration 2011.

3 Error Recovery Module in iSCSI Protocol for Mobile NAS Cluster System

The error recovery module was implemented MNAS (Mobile NAS Cluster System)[6]. A storage management tool was placed on top of it. The storage management tool is used to allocate server's storage to each client to pre-defined extent. Figure 1 shows the architecture of error recovery module.

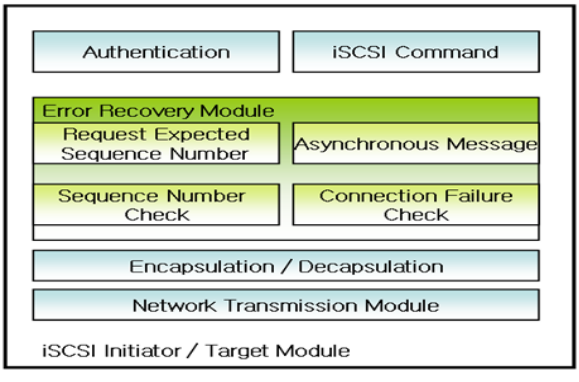

Fig. 1. Architecture of Error Recovery Module of MNAS (Mobile NAS Cluster System)

For the mobile appliances, it is very important because of high error rate in the wireless network. The following two considerations prompted the design of much of the error recovery functionality in iSCSI [7],[8].

- An iSCSI PDU may fail the digest check and be dropped, despite being received by the TCP layer. The iSCSI layer must optionally be allowed to recover such dropped PDUs.
- A TCP connection may fail at any time during the data transfer. All the active tasks must optionally be allowed to continue on a different TCP connection within the same session.

Considerable Errors. The iSCSI initiator and the iSCSI target organize the iSCSI system. iSCSI initiator is used on a mobile device, PDAs in this paper, iSCSI target has one or more mass storage devices.In wireless environment, many kinds of errors can be happened. However, iSCSI error recovery considers the errors on iSCSI protocol layer. iSCSI error recovery module considers following two errors in this paper.

- Sequence Number Error: When the iSCSI protocol transmits iSCSI Command PDU or Data PDU that has a sequence number, some PDU can be lost. So iSCSIPDU receiver (an initiator or a target) cannot get the valid PDU, it will get the PDU with an out of order.

– Connection Failure: If iSCSI target or initiator cannot communicate each other via a TCP connection, we define this situation, "connection failure".

3.1 Error Recovery Procedure

iSCSI protocol with error recovery checks the sequence number of every iSCSI PDU. If iSCSI target or initiator receives an iSCSI PDU with an out of order sequence number, then it requests an expected sequence number PDU again. In Connection failure case, when a connection has no data communication during the engaged time, iSCSI protocol with error recovery checks the connection status by the nop-command. We assume the multiple connections.

Sequence Number Error. When an initiator receives an iSCSI status PDU with an out of order or a SCSI response PDU with an expected data sequence number (ExpDataSN) that implies missing data PDU(s), it means that the initiator detected a header or payload digest error one or more earliest ready to transmission (R2T) PDUs or data PDUs. When a target receives a data PDU with an out of order data sequence number (DataSN), it means that the target must have hit a header or payload digest error on at least one of the earlier data PDUs. The target must discard the PDU and request retransmission with recovery R2T.

Connection Failure. At an iSCSI initiator, the following cases lend themselves to connection recovery

– TCP connection failure: The initiator must close the connection. It then must either implicitly or explicitly logout the failed connection with the reason code "remove the connection for recovery" and reassign connection allegiance for all commands still in progress associated with the failed connection on one or more connections. For an initiator, a command is in progress as long as it has not received a response or a Data-in PDU including status.
– Receiving an Asynchronous Message that indicates one or all connections in a session has been dropped. The initiator must handle it as a TCP connection failure for the connection(s) referred to in the Message.
  At an iSCSI target, the following case lend themselves to connection recovery

TCP Connection Failure. The target must close the connection and, if more than one connection is available, the target should send an Asynchronous Message that indicates it has dropped the connection. Then, the target will wait for the initiator to continue recovery.

4 Experiments and Analysis

We use the iSCSI initiator on Microsoft Windows CE.net 4.20 emulator version 4.0.2 for the convenience of experiments though it is available on real mobile devices like PDA. The iSCSI target is on the Linux kernel version 2.4.18 (Red hat 8.0). We

experiment with the 100Mbps Ethernet and the private network for no interference. Also we make the error for the experiments by ourselves.

Experiment for Sequence Error and Connection Failure. PDU loss is generated factitiously for this experiment and round trip time (RTT) of PDU is about 0.4 second. We use iSCSI write Command PDUs for 1 Mbyte data in this experiment. The experiment progress is:

- Initiator sends iSCSI PDUs to write 1Mbyte data.
- Several PDUs would not be sent by the PDU loss rate during the transmission.
- Target checks the PDU loss by the sequence number checking, and then it requests the PDU of the expected sequence number.

We consider an error recovery failure that target cannot receive the PDU of the expected sequence number during the time (RTT × 2 + α = 1 second).

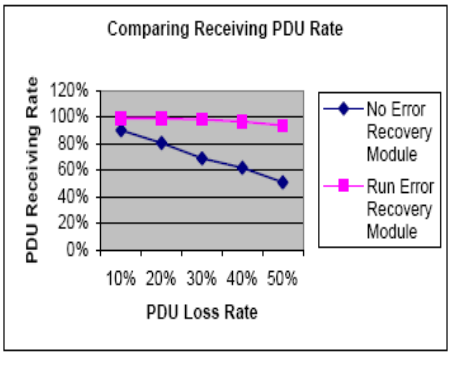

Fig. 2. Comparing the iSCSI PDU Receiving Rate. In this figure Y-axis is representing the PDU Receiving Rate and X-axis is representing the PDU Loss Rate.

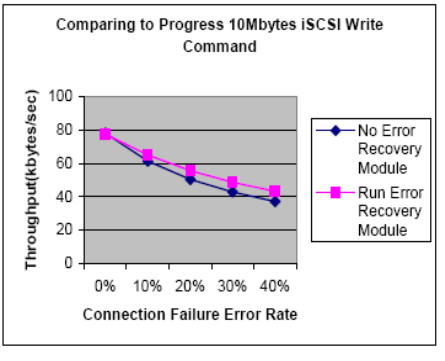

Fig. 3. Comparing to Progress 10Mbytes iSCSI Write Command. In this figure Y-axis is representing the throughput and X-axis is representing the connection failure error rate.

Figure 2 shows that PDU receiving rate is drastically decreasing without an error recovery module. However, the case of running our error recovery module just has about 8% PDU loss at 50% PDU loss rate. We assumed following factors for the experiments of connection failure:

- PDU loss is not considered.
- When iSCSI module without an error recovery has a connection failure, it reconnects after 60 seconds (Linux version 2.4.2 TCP Timeout). When iSCSI module has an error recovery, it reconnects by an error recovery module. We did not consider the time to reconnect by TCP or an error recovery module.
- Connection failure error rate is 0%, 10%, 20%, 30% and 40%. 10% Connection failure error rate is defined that 1/10 connections of all connections has connection failure during the time (DefaultTime2Wait + DefaultTime2Retain = 22 seconds) [8].

– We experimented with 4 connections and a connection failure can be happened concurrently.
– We experimented iSCSI write commands because iSCSI read commands have similar characters with iSCSI write commands.

Figure 3 shows that writing progress with our error recovery module has better throughput than writing progress without an error recovery module. Increasing the connection failure error rate, the difference of throughput becomes larger. Result showed that the difference of throughput on 40% error rate was above 6 Kbytes/sec.

5 Conclusion

In this paper, we describe the implementation of the error recovery module for iSCSI protocol based Mobile NAS Cluster system. According to experiments, iSCSI appliance with our error recovery module can transfer data stably and efficiently in wireless network which can have many errors. Consequently, our iSCSI appliance with the error recovery module has better reliability and efficiency than an iSCSI appliance without an error recovery module.

References

1. SAM-3 Information Technology – SCSI Architecture Model 3, Working Draft, T10 Project 1561-D, Revision 7 (2003)
2. Allayear, S.M., Park, S.S.: iSCSI Multi-connection and Error Recovery Method for Remote Storage System in Mobile Appliance. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3981, pp. 641–650. Springer, Heidelberg (2006)
3. Kim, D., Ok, M., Park, M.-S.: An Intermediate Target for Quick-Relay of Remote Storage to Mobile Devices. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3481, pp. 1035–1044. Springer, Heidelberg (2005)
4. Park, S., Moon, B.-S., Park, M.-S.: Design, Implement and Performance Analysis of the Remote Storage System in Mobile Environment. In: Proc. ICITA 2004 (2004)
5. Ok, M., Kim, D., Park, M.-S.: UbiqStor: A Remote Storage Service for Mobile Devices. In: Liew, K.-M., Shen, H., See, S., Cai, W. (eds.) PDCAT 2004. LNCS, vol. 3320, pp. 685–688. Springer, Heidelberg (2004)
6. Allayear, S.M., Park, S.S.: iSCSI Protocol Adaptation With NAS System Via Wireless Environment. In: International Conference on Consumer Electronics (ICCE), Las Vegus, USA (2008)
7. Allayear, S.M., Park, S.S., No, J.: iSCSI Protocol Adaptation with 2-way TCP Hand Shake Mechanism for an Embedded Multi-Agent Based Health Care Service. In: Proceedings of the 10th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems, Corfu, Greece (2008)
8. RFC 3270, `http://www.ietf.org/rfc/rfc3720.txt`