Advanced FCoE: Extension of Fibre Channel over Ethernet

Satoshi Kamiya, Kiyohisa Ichino, Masato Yasuda, Noriaki Kobayashi, Norio Yamagaki and Akira Tsuji
System IP Core Research Laboratories, NEC Corporation, Kawasaki, Kanagawa, Japan
Email: {kamiya@ak, k-ichino@bk, m-yasuda@ct, n-kobayashi@iw, n-yamagaki@cj, a-tsuji@bq}.jp.nec.com

***Abstract*—Recently, Fibre Channel over Ethernet (FCoE) has been gaining attention as a solution of LAN/SAN convergence in Data Centers. In FCoE, it is necessary to introduce "FCoE Forwarder (FCF)" and to replace legacy Ethernet switches with lossless Ethernet switches. Current FCoE, specified by FC-BB-5, has a scalability issue due to concentrating all operations of User-plane and Control-plane into FCF. Although FC-BB-6 project aims to solve the issue, it requires "FC/FCoE Data Forwarder (FDF)", which raises DC cost. In this paper, we propose a novel FCoE system, called Advanced FCoE (AFCoE), to realize large-scale FCoE system at low-cost. In AFCoE, network edge devices have mechanisms to separate FCoE frames into User-plane frames or Control-plane frames. The User-plane traffic is directly forwarded by legacy Ethernet switches to remote edge devices, while the Control-plane traffic is forwarded to a controlling server and processed there. In addition, reliable data transport technology is introduced for frame loss caused in legacy Ethernet switches. We also show the validation results of a prototype AFCoE system.**

I. Introduction

Today, cloud computing has been attracting attention as a new Information Technology (IT) environment. Cloud computing realizes improvement of utilization efficiency of IT services, cutting down operating cost and emerging new IT service rapidly. It composes Data Center (DC) that operates huge number of servers and storages in one place. Recent rapid growing of cloud computing makes a scale of DCs get larger, which increases DC cost in various aspects. One of approaches for the cost savings is to reduce CAPEX and OPEX by I/O consolidation on servers. Many servers have multiple interfaces, such as Local Area Network (LAN) and Storage Area Network (SAN). By converging interfaces of LAN and SAN, equipment cost and electric power cost can be reduced.

Fibre Channel over Ethernet (FCoE)[1], [2], [3], [4], [5] is one of the LAN/SAN converged system technologies. FCoE transports Ethernet frames which encapsulate Fibre Channel (FC)[10] frames. FC does not allow frame loss by network congestion and FCoE possesses the same property by inheritance. FCoE is lack of retransmission capability to compensate frame loss, so it requires "Lossless Ethernet". Data Center Bridging (DCB), an Ethernet extension, has been proposed in IEEE[6], providing lossless Ethernet. FCoE requires that servers, storages and Ethernet switches support DCB (henceforth, we call DCB-supported Ethernet switches "DCB switches").

There are two problems about FCoE. One is that FCoE system has scalability limitation. In FC-BB-5[7], the latest specification of FCoE, FCoE traffic must go through FCoE Forwarder (FCF). This behavior could make FCF a bottleneck which limits a scale of the system. The other problem is a cost of FCoE system. FCoE requires replacing existing switches with DCB switches and FCFs, causing increased network cost. FC-BB-5 specifies a configuration with multiple FCFs, which solves the scalability problem. However, such a system leads to high cost and management complexity. Although FC-BB-6[8], the next project of FC-BB-5, introduces "FC/FCoE Data Forwarder (FDF)" in order to solve the scalability issue of the FCoE, the cost problem remains at the expense of FDF.

In this paper, we propose an enhanced FCoE system, called *Advanced FCoE (AFCoE)*, to solve the scalability issue of the FCoE and to realize large-scale FCoE system at low-cost. AFCoE system behaves as a large-scale FCoE switch by separating frames into User-plane (U-plane) frames or Control-plane (C-plane) frames (U/C separation) at network edge devices such as Converged Network Adapters (CNAs). U-plane frames can be transferred by Ethernet switches. C-plane frames are sent to a controlling server which executes control and management function of AFCoE. This approach improves the efficiency of a future wide variety of information transports over commodity Ethernet including FCoE. The amount of C-plane traffic is generally much less than that of U-plane. AFCoE is more scalable than FC-BB-5 because it does not cause traffic concentrate problem. In addition, AFCoE introduces functions of retransmission and reordering into network edge devices in order to use legacy Ethernet switches instead of DCB switches. These functions provide reliable data transport in Ethernet layer. The cost of AFCoE system is lower than that of other methods because AFCoE does not need FC-BB-6's FDF, multiple FCFs, nor DCB switches. Moreover, due to the U/C separation, AFCoE is easy to interwork with OpenFlow[9] that has received attention recently.

The rest of the paper is organized as follows. In Section II, we briefly review the FCoE and its object. In Section III, we describe the proposed AFCoE system. In Section IV, we show our prototype AFCoE system and validation results. In Section V, we discuss interwork between AFCoE and OpenFlow. Finally, we conclude in Section VI.

978-0-9836283-2-3

This paper was peer reviewed by subject matter experts for publication in the Proceedings of DC–CaVES 2011

II. Fibre Channel over Ethernet (FCoE)

In this section, we briefly show the FCoE, and point out its technical issues. Then, we discuss the requirements to address the issues for an enhanced FCoE system.

A. SAN protocols

The protocols for SAN include Fiber Channel (FC)[10], Internet Small Computer System Interface (iSCSI)[11], Fibre Channel over IP (FCIP)[12], Internet Fibre Channel protocol (iFCP)[13], and FCoE. FC can hardly achieve LAN/SAN convergence in Ethernet because it depends on credit-based flow control for lossless communication. IP-SAN protocols such as iSCSI, FCIP, and iFCP rely on TCP/IP as reliable transport and easily accomplish LAN/SAN convergence in existing Ethernet. However, TCP/IP processing could be significant overhead in 40 Gbps or 100 Gbps Ethernet.

On the other hand, FCoE does not need TCP/IP, but it requires "Lossless Ethernet" provided by Data Center Bridging (DCB)[6]. DCB consists of Priority-based Flow Control (PFC)[14], Enhanced Transmission Selection (ETS)[15], Data Center Bridging eXchange (DCBX) protocol[15], Congestion Notification (CN)[16], and TRansparent Interconnection of Lots of Links (TRILL)[17]. FCoE encapsulates FC frames directly into Ethernet frames, having the advantage of performance over IP-SAN.

FCoE specification is defined by FC-BB-5 project in INCITS T11. FC-BB-6 is under development and currently is not fixed. In the following, the contents are described according to FC-BB-5.

B. Components

Figure 1 shows the storage network system with FCoE. FCoE system consists of storages and servers as FCoE Nodes (ENodes), FCFs (also called FCoE switches), and DCB switches. ENodes are classified into initiators and targets. An initiator is the ENode which sends Small Computer System Interface (SCSI) requests, and a target is the ENode which receives them and replies. ENodes and FCFs are connected with virtual FC links which are lossless. ENodes are connected to FCF directly or via DCB switches by physical Ethernet links.

FCoE and FC may be used together in initial FCoE deployment. As Ethernet grows up to 40 Gbps or 100 Gbps, "Native FCoE system" will be dominant where all ENodes and switches are connected in FCoE. In this paper, we focus on "Native FCoE system".

C. Protocol Sequence

Figure 2 shows FCoE protocol sequence (see Refs. [7], [18] in details of each procedure). FCoE sequences are the following;

(1) Exchange messages between adjacent devices (DCB Exchange (DCBX))
(2) Connection establishment between ENode and FCF (FCoE Initialize Protocol (FIP) VLAN Discovery, FIP Discovery, FIP Fabric Login (FLOGI), name resolution (dNS), etc.)
(3) Connection establishment between ENodes via FCF (Port Login (PLOGI) and Process Login (PRLI))
(4) Data transfer between ENodes via FCF (SCSI-FCP)

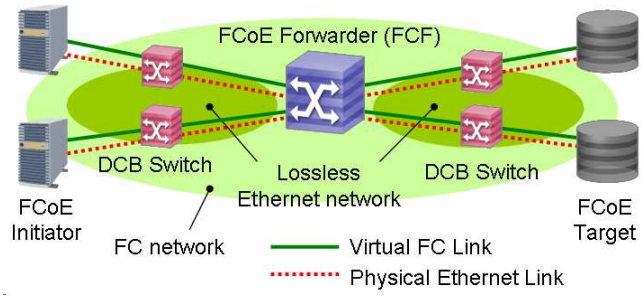

Fig. 1. The storage network system with FCoE (Native FCoE).

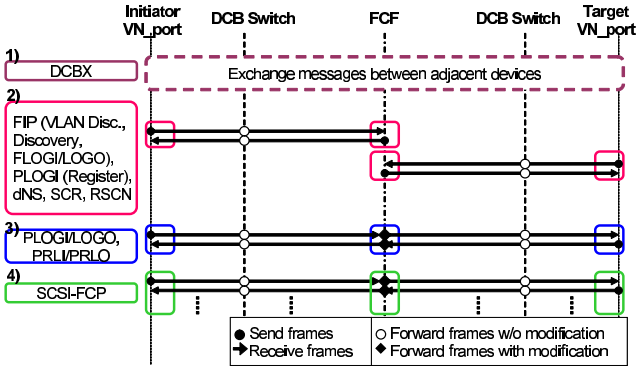

Fig. 2. FCoE protocol sequence.

In FCoE system, all data traffic between initiators and targets (from (2) to (4)) go through FCF since all FCoE frames are routed according to FC address (FC-ID).

D. Technical Issues for Realizing Large-scale FCoE system

To realize large-scale FCoE system at low-cost, this section describes two technical issues; 1) Scalability Limitation, and 2) Performance Degradation by Data Loss.

1) Scalability Limitation: In the current FCoE, all traffic between ENodes goes through FCF. FCF must perform management and control processing of FCoE system. Therefore, the scale of FCoE system is limited by the performance of the FCF. As cloud computing is growing, the scale of DC will be larger. The poor scalability could be one of technical issues in a large-scale DC.

2) Performance Degradation by Data Loss: Since the transport layer of FCoE must keep lossless quality equivalent to FC-2 layer[1], CNAs on ENodes, FCF and switches must support

[1]FC-2 layer consists of three sub-layers, FC-2V (FC-2 - Virtual), FC-2M (FC-2 - Multiplexer), and FC-2P (FC-2 -Physical). FC-2P of which has flow controls to prevent data loss when the service class 2 of FC is used. However, the flow control is not provided because FCoE includes only FC-2V.

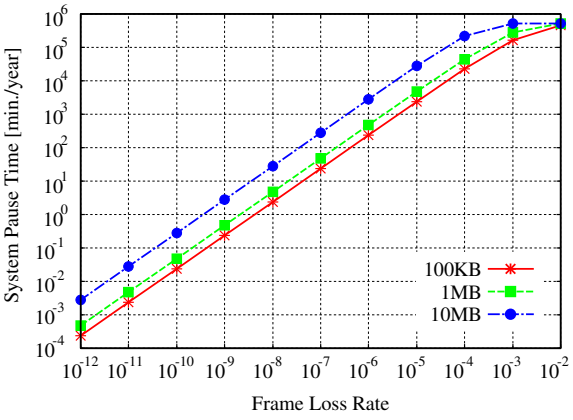

Fig. 3. The estimated pause time of disk access.

DCB. It means that it costs to replace legacy Ethernet switches with DCB switches. Without DCB switches, performance degradation occurs because a legacy Ethernet switch may discard frames in network congestion.

Figure 3 shows the estimated pause time of disk access in case of increasing frame loss rate. The disk accesses are executed repeatedly and the accessed size is 100 KB, 1 MB or 10 MB respectively. In our estimations, when a frame is lost during a disk access, SCSI layer detects the loss after timeout, and retries disk access. In Figure 3, the SCSI timeout value and the disk seek time are set to 2 seconds and 5 milliseconds, respectively.

As shown in Figure 3, the pause time increases as frame loss rate is increased. If a frame loss is occurred over probability of 10^{-8}, the pause time becomes over one minute per year. In particular, when using legacy Ethernet switches with small frame buffer, it becomes more important to conceal the frame loss. That is, Ethernet layer requires the function to ensure data integrity for FC layer to prevent the impact of frame losses. In this paper, we call Ethernet with this functionality *Reliable Ethernet*.

E. Requirements for Realizing Large-scale FCoE System

In this section, we discuss the requirements to overcome the above mentioned issues. As the details are discussed below, we consider that the mechanisms of U/C separation and reliable Ethernet transport which is different from DCB are required to realize large-scale FCoE system at low-cost.

1) Scalability Limitation: There are two approaches to address the scalability issues; (1) clustering FCFs and (2) revising FCF structure. For the approach (1), it is necessary to implement routing protocol between FCFs and to manage multiple FCFs. This approach incurs higher equipment cost and management complexity. We consider that the approach (2), revising FCF structure is more effective. Here, focusing on FCoE protocol sequence (Figure 2), it comes to see that the sequence consists of control frames (i.e. C-plane frames), which are information of FCoE and FC for the system management and control, and FCoE data frames (i.e. U-plane frames). From this observation, "U/C separation" is one of effective solutions to revise FCF structure.

U/C separation is the framework which separates U-plane and C-plane, and enables to avoid the problem that one plane limits the another plane's scalability. In this case, C-plane frames must be processed at the network device with management and control functionality such as a server. There is no need to transfer U-plane frames via FCF, which removes the scale limitation of FCF.

If the reliable Ethernet transport is provided at CNA on each ENode, legacy Ethernet switches can be used instead of DCB switches. That is, the large-scale FCoE system is realized by using CNAs with the functions of reliable Ethernet transport and U/C separation, legacy low-cost Ethernet switches, and a server for C-plane processing. Its total cost is not expensive compared to that of DCB-supported CNAs, DCB switches and FCF from the following considerations;

- Legacy Ethernet switch which permits frame loss will have smaller buffer than DCB switch, and
- C-plane operations can be implemented with software on a server unlike FCF because C-plane traffic must be much less than U-plane traffic.

2) Performance Degradation by Data Loss: To use legacy low-cost Ethernet switches, it is necessary to achieve the reliable Ethernet transport. Generally, there are two methods which compensate for data loss; (1) Retransmission, and (2) Data recovery by Forward Error Collection (FEC). We consider that the retransmission is suitable since it enables to shorten processing delay at normal operations, and has the prospect of realizing high-speed processing in the simple circuit. Also, FEC can not deal with loss due to congestion. A reordering function is needed in combination with retransmission because the retransmission causes out-of-order frames at receiver.

III. Advanced FCoE (AFCoE)

In this section, we propose "Advanced FCoE (AFCoE)" which satisfies the requirements discussed in Section II-E.

A. Overview

Figure 4 shows the system of our proposed AFCoE.

In AFCoE, U/C separation and fast retransmission/reordering functions for the reliable Ethernet transport are equipped at only the network edge devices.

U-plane frames are forwarded by legacy Ethernet switches with possible frame loss, and the control and management processing of FC and FCoE are performed by a controlling server. We call this controlling server the "FCoE Fabric Controller (FCC)".

Concretely, Figure 5 shows the protocol sequence of AFCoE. The differences between AFCoE and FCoE (Figure 2) are newly introduced FCC and switches which correspond to FCF, and forwarding of SCSI-FCP (SCSI commands and these reply data) frames without address modifications within the network. In AFCoE, SCSI-FCP frames, which are majority of

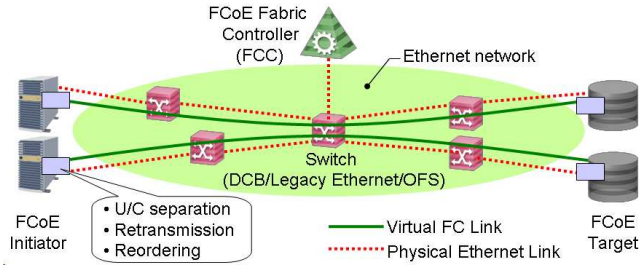

Fig. 4. Advanced FCoE system.

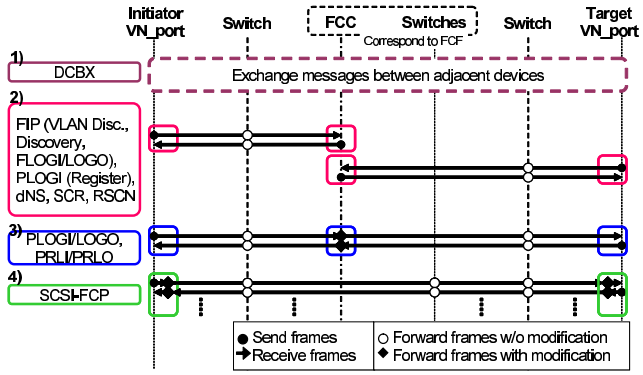

Fig. 5. AFCoE protocol sequence.

frames among traffic between ENodes, are forwarded in cut-through manner in Ethernet while other FCoE frames and FIP frames are transmitted to FCC.

By U/C separation, whole AFCoE system can behave as a large-scale FCoE switch virtually. Its switching capacity can be freely increased by adding legacy Ethernet switches, and the management and control processing can be also scalable by clustering of FCCs and load balancing.

In the following, we explain the features of AFCoE, the system architecture, U/C separation and the reliable Ethernet transport.

B. Features

The main features of AFCoE include the following three functions;

1) U/C Separation: In FCoE, because the amount of C-plane traffic is less than that of U-plane traffic (SCSI-FCP frames), C-plane operations are removed from FCF and performed in the dedicated controlling server (FCC). U-plane frames are transferred on legacy Ethernet.

2) Reliable Ethernet Transport: Instead of lossless Ethernet provided by DCB, reliable Ethernet is realized by introducing fast retransmission function and reordering function into Ethernet layer of ENodes.

3) Flat Data Transport Network by Using L2 Address: An FCoE frame is forwarded according to Ethernet MAC address, where MAC address is configured by using Fabric Provider MAC Address (FPMA) specified in FC-BB-5. This achieves FCoE frame transport via multiple legacy Ethernet switches.

With the above features, the AFCoE system can be regarded as a large-scale switch virtually, unlike the FCoE system with FCF. In AFCoE, U/C separation can increase the number of covered SANs by clustering FCCs, and configure the non-blocking switching network by adding adequate number of legacy Ethernet switches when requiring more SAN capacity (i.e. the number of ENodes). Clustering FCCs also could avoid single point of failure. In addition, AFCoE can use legacy Ethernet switches instead of FCF because the function of interpreting FC is not needed for each switch. Moreover, the reliable Ethernet transport can enable to use of legacy Ethernet switches instead of DCB switches.

From the above discussions, AFCoE has advantages of scalability and cost over the traditional FCoE.

C. System Architecture

The AFCoE system is composed of four components; (1) initiator, (2) target, (3) FCC, and (4) switch as shown in Figure 4. Note that an initiator and a target are called ENode. An ENode performs U/C separation and modifies MAC address of each FCoE frame so that the frame can be forwarded by FC-unaware switches. FCC is a controlling server with management and control functions, instead of FCF. FCC processes protocol sequence of FIP and FC, such as FIP FLOGI, PLOGI, PRLI and dNS between an initiator and a target. FCC also manages topology information of the system, routing information for frame forwarding, and so on (see Section III-D3). A switch forwards FIP/FCoE frames according to MAC address.

D. U/C Separation

1) Classification of Frame for Separation: In AFCoE, a SCSI-FCP frame is separated from the others for cut-through forwarding in the switch. This function is equipped in NIC or CNA of ENode. U/C separation is easily implementable because it just separates SCSI-FCP frame from the others.

Figure 6 shows the FCoE frame format (see Ref.[3] in details). Concretely, a SCSI-FCP frame can be identified by referring to Ether type field and FC type field within FCoE frame, where Ether type = 0x8906 (FCoE) and FC type = 0x08 (FCP) represent SCSI-FCP frame.

With U/C separation, a SCSI-FCP frame is forwarded via only the switches, which leads to lower delay and improvement of system performance.

2) MAC Address Translation for Frame Separation: Figure 7 shows the MAC address translation of AFCoE. When an initiator send a frame to a target, the initiator separates SCSI-FCP frames from the others (U/C separation), and translates the destination MAC address from FCC MAC address into Virtual N_Port (VN_Port)[2] MAC address of the target (Address Translation), as shown in Figure 7 (a). The SCSI-FCP frame is transferred to the target via legacy Ethernet switches. Then,

[2]VN_Port is a virtual port used in FC-2V sub-layer on the ENode.

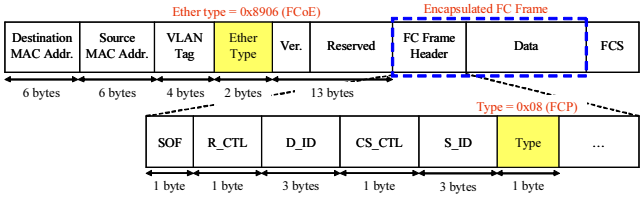

Fig. 6. FCoE frame format.

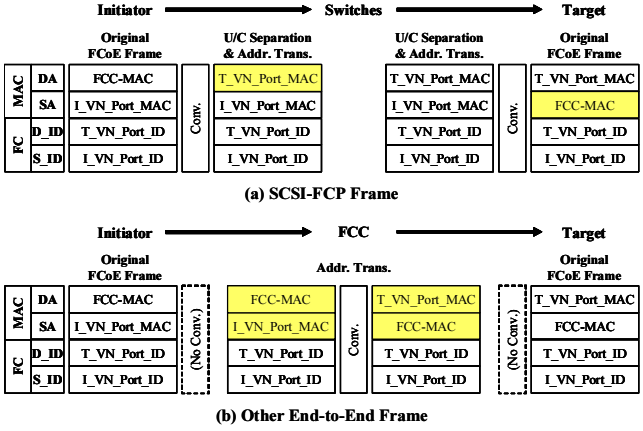

Fig. 7. The MAC address translation of AFCoE for (a) a SCSI-FCP frame and (b) other End-to-End frame.

the target translates the source MAC address from VN_Port MAC address of the initiator into FCC MAC address. On the other hand, the other frames from ENode are forwarded to FCC without MAC address translations.

When FCC sends a frame to ENode, FCC modifies the source and destination MAC addresses as FCF does, as shown in Figure 7 (b). In this way, SCSI-FCP frames can be transferred by only legacy Ethernet switches in the network.

3) Components of FCC: Figure 8 shows the components of FCC. FCC consists of Control Frame Broker (CFB), Processing Module Unit (PMU), Fabric Information Database (FI-DB) and Node Information Database (NI-DB).

CFB identifies arrived FCoE/FIP frames, and transfers them to PMU. PMU processes the frames and generates new frames to ENode by using information of FI-DB or NI-DB.

PMU consists of the following modules;

(1) Fabric VLAN ID responder responds to FIP VLAN Discovery from ENode and sends back fabric VLAN ID as FIP VLAN Discovery Advertisement to the ENode.
(2) FCC MAC responder responds to FIP Discovery from ENode and sends back FCC MAC as FIP Discovery Advertisement to the ENode.
(3) Login Server responds to FIP FLOGI/LOGO from ENode, registers/unregisters the ENode to/from MAC-FC ID table in NI-DB, and sends back virtual MAC address as FIP FLOGI accept (ACC) / reject (RJT) to the ENode.
(4) Directory Server responds to dNS from ENode, registers the ENode information to registered node table in NI-DB, or sends back the specified information to the ENode.

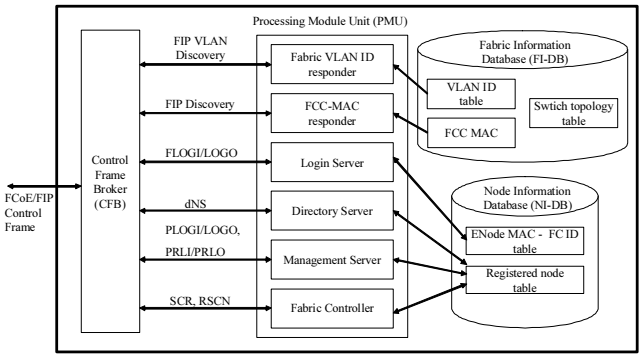

Fig. 8. Components of FCC.

(5) Management Server responds to PLOGI/LOGO or PRLI/PRLO from ENode, registers/unregisters the ENode to/from registered node table in NI-DB, and forwards the frame to another ENode.
(6) Fabric Controller responds to SCR from ENode, registers the ENode to registered node table in NI-DB and sends back SCR ACC/RJT. When major fabric changes occur, Fabric Controller sends RSCN to all ENodes in the registered node table.

FI-DB consists of the following tables;

(1) VLAN ID table stores fabric VLAN ID.
(2) FCC MAC stores its own MAC address.
(3) Switch topology table stores the connection information between switches.

NI-DB consists of the following tables;

(1) ENode MAC-FC ID table stores the relation between an ENode MAC address and an FC ID.
(2) Registered node table stores ENode information such as a virtual MAC address and an ENode connection state.

In AFCoE system, C-plane operations and FC/FCoE/FIP management function are executed on FCC. Essentially, the processing load of FCC is not so high because the amount of C-plane traffic is much less than that of U-plane traffic. This means that single FCC covers large network including many ENodes.

E. Reliable Ethernet

In AFCoE, fast retransmission function and reordering function are required to achieve reliable Ethernet transport. These functions are introduced into the intermediate layer between Ethernet and FCoE layers, as shown in Figure 9. Both functions are deployed in only ENodes, which enables use of legacy Ethernet switch.

We consider that they should have (1) capability of low-latency suitable for DC environment, (2) simplicity to keep up with throughput growth of Ethernet (e.g. 40 Gbps and

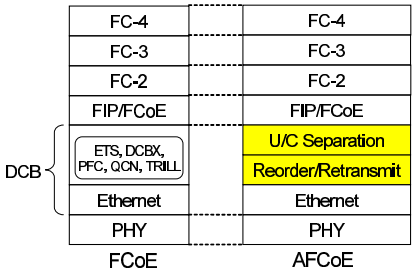

Fig. 9. AFCoE protocol stack.

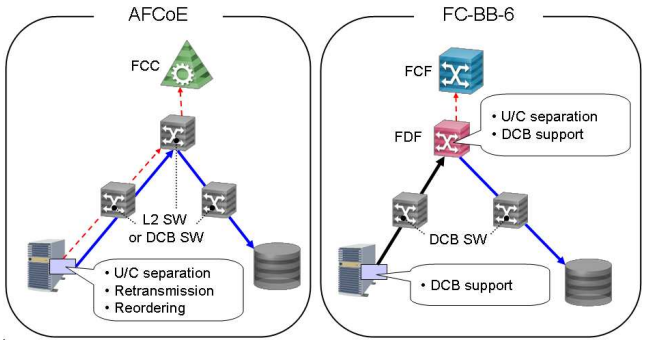

Fig. 10. The difference of between AFCoE and FC-BB-6.

100 Gbps), (3) efficiency of the retransmission, and (4) ease of integration with the reordering function.

As a retransmission method which meets the above requirements in Ethernet layer, Rapid and Reliable Data Delivery (R2D2)[19] has been proposed. It is shown that the software-based R2D2 works well in GbE network environment in Ref. [19]. In higher-speed network environment like 10 GbE and 40 GbE, the hardware-based R2D2 is desirable.

We consider that R2D2 and reordering are easily implementable as firmware because of their simple mechanism. Hence, the cost of CNA with their functions will be reasonable.

F. The advantage of AFCoE over FC-BB-6

Currently FC-BB-6 project[8] is working for a system which solves the issues of FC-BB-5. FC-BB-6 performs U/C separation and makes scalable FCoE network.

In FC-BB-6, a new switch called FC/FCoE Data Forwarder (FDF) is introduced and deployed at between ENode and FCF. FDF performs U/C separation and shares information with FCF which its superior switch of the FDF, reducing FCF load and improving scalability of FCoE system.

Figure 10 shows comparison between AFCoE and FC-BB-6. The difference is the location where U/C separation is executed. In FC-BB-6, FDF is additionally required. The number of FDFs required increases as the network grows. On the other hand, AFCoE does not need FC-aware switches in network at all as mentioned before. Figure 11 shows the topology overview of FC-BB-6 and AFCoE. AFCoE makes network simple and flat, compared to FC-BB-6.

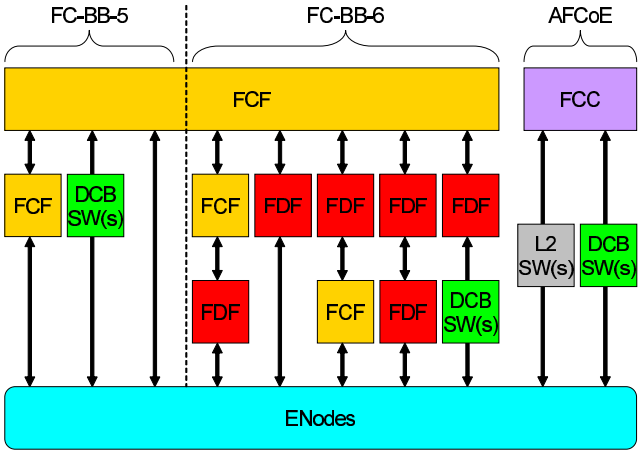

Fig. 11. Topology overview. (referred FDF Requirements[20])

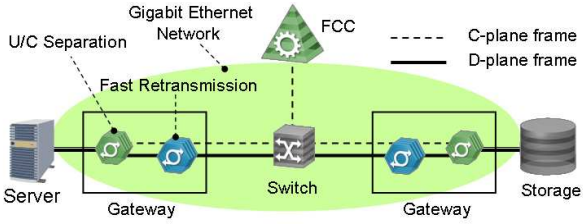

Fig. 12. Prototype AFCoE system.

IV. Implementation and Validation

We implemented a prototype of the proposed AFCoE system and validated it. The main purpose of this prototype system is functional verification.

A. Prototype System

In AFCoE, functions for U/C separation and reliable Ethernet are implemented in CNAs. However, in our AFCoE prototype system, we implemented them as software on server for early validation. We call the server "gateway (GW)". Figure 12 shows our prototype AFCoE system. The prototype system consists of a server (initiator), a storage (target), an FCC, a legacy Ethernet switch, and two GWs. All links are Gigabit Ethernet (GbE). We place each GW beside ENode.

The fast retransmission function is based on R2D2. The retransmission and reordering functions use Selective Repeat Automatic Repeat-Request (ARQ) with sliding window. This algorithm is similar to Link Access Procedure, Balanced (LAPB) [21] but more simplified because the retransmission is kicked by only timeout instead of reception of explicit retransmission request from the receiver. In many DC networks, Round-trip delay Time (RTT) is shorter than several hundreds microseconds [22], and the retransmission timeout value can be set to the same order of time. Therefore, the retransmission is done so quickly that the elimination of

TABLE I
EXPERIMENTAL CONDITION.

Packet loss rate	1 % (Random loss)
FC link timeout	10 seconds
Retransmission timeout	100 micro seconds

TABLE II
EXPERIMENTAL EQUIPMENT.

Equipment	Model	OS
Server	IBM x3650M2 (CNA: Qlogic QLE8142-SR)	RHEL5.3 (32bit)
Storage	NetApp FAS3140	ONTAP 7.3.2P5
Switch	NEC QX-S5828T	-
GW	NEC Mate MY33A/E7	CentOS5.5 (64bit)
FCC	NEC Mate MY24A/B4	Fedora 13 (32bit)

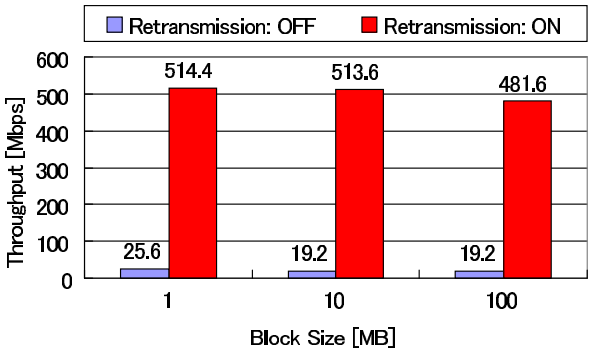

Fig. 13. U-plane throughput under packet loss environment.

explicit retransmission requests will not cause performance degradation in AFCoE system.

B. Validation

We validated functions of AFCoE which described in Section III.

1) U/C separation: We confirmed that the three following FCoE operations work correctly by using the AFCoE prototype system; (1) Fabric Login from server to storage, (2) Port Login and Process Login from server to storage, and (3) Disk access from server to storage. U-plane frames did not relay FCC but were directly transferred between ENodes. AFCoE can provide the same functionality as "Native FCoE system".

2) Reliable Ethernet: We evaluated reliable Ethernet function to investigate how presence or absence of retransmission affects performance under the environment where frame loss may occur. We measured throughput by sequential write using *dd* command. Tables I and II show our experimental condition and Figure 13 shows the throughput of disk access when varying data block size.

As shown in Figure 13, throughput without retransmission is seriously low, 1.9 - 2.6 % of physical link rate (1Gbps). The cause of low throughput is due to FC link timeouts. Throughput with retransmission is improved to 48 - 51 % of physical link rate and 20 - 27 times higher than that without retransmission because the retransmission prevents the timeouts by concealing frame loss from FC layer. From the above results, it was confirmed that fast retransmission is effective in frame loss.

V. OPENFLOW-BASED AFCOE SYSTEM

In this section, we discuss OpenFlow-based AFCoE system. OpenFlow[9] is one of frameworks to realize U/C separation. OpenFlow consists of OpenFlow Switch (OFS) as U-plane and OpenFlow Controller (OFC) as C-plane, where each OFS is controlled by OFC. OFC and OFS are communicated each other by secure channel on OpenFlow protocol.

The current OFS can not interpret the contents of FIP and FCoE frame header because OpenFlow supports only TCP/UDP, IP and Ethernet frames and not perform U/C separation which is key solution of scalability problem. On the other hand, AFCoE can interpret FIP and FCoE frames and translate between FC-ID and MAC address at network edge devices. Both OpenFlow and AFCoE exploit U/C separation. They are interworked easily.

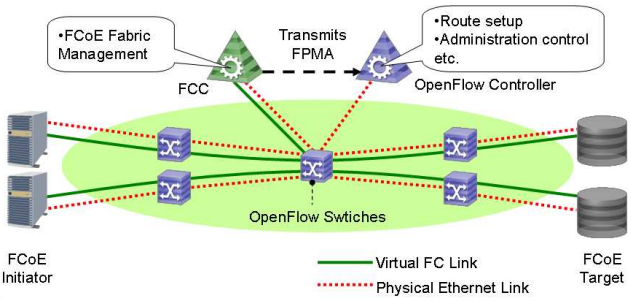

Fig. 14. OpenFlow-based AFCoE system.

The advantages of OpenFlow-based AFCoE system are (1) to achieve LAN/SAN unified management in combination with LAN management capability of OpenFlow, and (2) to set up path of SAN traffic in addition to LAN traffic in LAN/SAN converged network by using OpenFlow's routing mechanism in order to guarantee each own QoS.

Figure 14 shows the OpenFlow-based AFCoE system. OFC can calculate and set up a route when end-to-end MAC address pairs are given. AFCoE uses this feature. FCC notifies OFC of FPMA assigned to ENode when FCC accepts PLOGI. After that, OFC prepares a route for U-plane traffic between the ENodes. In this way, OFC enables to manage FCoE traffic as well as LAN traffic. FIP and FCoE frames can be transferred via OFS.

VI. CONCLUSION

This paper proposed Advanced FCoE (AFCoE) as an enhanced FCoE system to address the scalability issues of FCoE which is a solution of LAN/SAN convergence, and to realize large-scale FCoE system using legacy Ethernet switches. The main features of AFCoE are U/C separation and

reliable Ethernet transport at the network edge. In AFCoE, U-plane frames are separated from the others at network edges, and transferred by legacy Ethernet switches. The operations for C-plane frames are performed at a controlling server. Besides, to achieve the reliable data transport in Ethernet layer, the retransmission function and the reordering function are introduced into network edge devices. With these mechanisms, AFCoE can realize the large-scale FCoE system at low-cost.

AFCoE can conceal the differences in the characteristics of FC and Ethernet at the network edge. This approach improves the efficiency of a future wide variety of information transports over Ethernet including FCoE. We suggest that OpenFlow-based AFCoE, combination of OpenFlow and FCoE will provide LAN/SAN unified management capability to enable LAN/SAN convergence in C-plane as well as U-plane.

Our future works include system validation and performance evaluation in more complex network, implementation and evaluation of OpenFlow-based AFCoE system.

ACKNOWLEDGMENT

This work was partly supported by Ministry of Internal Affairs and Communications (MIC).

REFERENCES

[1] FCoE (Fibre Channel over Ethernet), "http://www.fcoe.com/."
[2] J. Jiang and C. DeSanti, "The Role of FCoE in I/O Consolidation," *Proceedigs of International Conference on Advanced Infocomm Technology*, July 2008.
[3] C. DeSanti and J. Jiang, "FCoE in Perspective," *Proceedigs of International Conference on Advanced Infocomm Technology*, July 2008.
[4] S. Gai, "Data Center Networks and Fibre Channel over Ethernet (FCoE)," *Lulu.Com*, 2008.
[5] S. Gai and C. DeSanti, "I/O Consolidation in the Data Center - A Complete Guide to Data Center Etherent and Fibre Channel over Ethernet-," *Cisco Press*, 2009.
[6] IEEE 802.1 Data Center Bridging Task Group, "http://www.ieee802.org/1/pages/dcbridges.html."
[7] Fibre Channel - Backbone - 5, June 2009, "http://www.fcoe.com/09-056v5.pdf."
[8] Fibre Channel - Backbone - 6, October 2010, "http://www.t11.org/ftp/t11/pub/fc/bb-6/10-211v2.pdf."
[9] The OpenFlow Switch Consortium, "http://www.openflow.org/."
[10] Home Page for Technical Committee T11 "http://www.t11.org/index.html."
[11] Internet Small Computer Systems Interface (iSCSI) "http://tools.ietf.org/html/rfc3720."
[12] Fibre Channel Over TCP/IP (FCIP) "http://tools.ietf.org/html/rfc3821."
[13] iFCP - A Protocol for Internet Fibre Channel Storage Networking "http://tools.ietf.org/html/rfc4172."
[14] IEEE 802.1: 802.1Qbb - Priority-based Flow Control, "http://www.ieee802.org/1/pages/802.1bb.html."
[15] IEEE 802.1: 802.1Qaz - Enhanced Transmission Selection, "http://www.ieee802.org/1/pages/802.1az.html."
[16] IEEE 802.1: 802.1Qau - Congestion Notification, "http://www.ieee802.org/1/pages/802.1au.html."
[17] Transparent Interconnection of Lots of Links (trill), "http://datatracker.ietf.org/wg/trill/."
[18] FC-FS-3, July 2010, "http://www.t11.org/ftp/t11/pub/fc/fs-3/10-010v2.pdf."
[19] B. Atikoglu, M. Alizadeh, J. S. Yue, B. Prabhakar and M. Rosenblum, R2D2: Rapid and Reliable Data Delivery in Data Centers, April 2010, "http://forum.stanford.edu/events/posterslides/R2D2RapidandReliableDataDeliveryinDataCenters.pdf."
[20] R. Hathorn, B. Maskas and E. Smith, FC-BB-6 FDF Requirements, July 2010, "http://www.t11.org/ftp/t11/pub/fc/bb-6/10-343v0.pdf."
[21] LAPB (Link Access Procedure, Balanced), ISO Standard ISO/IEC 7776.
[22] V. Vasudevan, A. Phanishayee, H. Shah, E. Krevat, H. Shah, H. Shah, E. Krevat, D. G. Andersen, G. R. Ganger, G. A. Gibson and B. Mueller, "Safe and Effective Fine-grained TCP Retransmissions for Datacenter Communication," *SIGCOMM' 09*, August 2009.