IPSJ Journal Vol. 51 No. 3 952–964 (Mar. 2010)

Regular Paper

A Scalable Network Architecture for a Large-scale Uni-directional Link

Kotaro Kataoka,[†1] Achmad Husni Thamrin[†1]
and Jun Murai[†2]

Effective bandwidth utilization and scalability are vital issues for IP networking over a large-scale uni-directional link (UDL), such as a wide-area wireless broadcast over satellite or terrestrial digital broadcasting. On a large-scale UDL, the current network architecture is not scalable to cover an extraordinary number of receivers that communicate using a Link-layer Tunneling Mechanism (LLTM). This paper proposes a network architecture for a large-scale UDL that: (1) decreases the traffic load of LLTM at the upstream network of the UDL, (2) coordinates the data link layer and network layer of receivers without communications via UDL, and (3) enables neighbor discovery for direct communication between receivers via a bi-directional link that is used as a return path for LLTM. Simulation results showed that our approach reduces by more than 90% the control messages to be sent via UDL compared with IPv6 stateless address autoconfiguration on the existing network architecture. Our proposal improves the UDL bandwidth consumption from $O(N)$ to $O(1)$, so that the bulk of the bandwidth can be utilized for delivering services, not for network configuration of receivers.

1. Introduction

Uni-directional Links (UDLs), such as satellite or terrestrial networks used for TV, radio or multimedia broadcast, are widely deployed around the world. Beyond the broadcasting services, UDLs have high potential as wide-area data links on the Internet, especially for backup lines in emergency situations and as shortcuts for Internet-wide multicast to a large number of subscribers.

The standardization of Link-layer Tunneling Mechanism (LLTM, RFC3077[1])) enabled emulation of Bi-directional Broadcast Multiple Access (B-BMA), where nodes in the transmitter system of UDL (upstream) and those in the receiver system (downstream) can communicate transparently as if the link layer supports bi-directional broadcast, like Ethernet using repeaters. LLTM allows network protocols, such as address resolution, routing and TCP, to work on UDL without any modification. However, the scalability of the performance of the network protocols is still an open issue.

In general, like digital TV or radio broadcast, UDL is not very high speed and its bandwidth is limited. It is natural to think that a large number of receivers exist on UDL due to the broadcast nature of the link. The typical usages of an operational B-BMA over a satellite UDL in Asia[2]) are, for example, (1) realtime video and audio communication for small multicast groups, (2) one to many data transfer or streaming, and (3) broadband Internet access. The user terminals tend to be large and fixed because they include a receive-only earth station. Meanwhile, for the case of UDL over terrestrial digital broadcast[3]), multicast is focused on serving large-scale receivers. We can expect services such as, for example, transferring the background traffic of a massively multiplayer online role-playing game (MMORPG) to its subscribers. The user terminals can be smaller, like laptop computers or smart phones, considering the size and availability of tuner devices. Here, effective link utilization is a vital issue to enable services via B-BMA emulated by LLTM. However, not all protocols are designed to serve such a large number of nodes on a single link.

We propose a network architecture that enables a large number of nodes to communicate globally using B-BMA as one of the Internet paths. This paper aims to achieve better utilization of UDL bandwidth for delivering services.

The rest of this paper is organized as follows: Section 2 describes the properties of UDL and B-BMA which is emulated using LLTM, and Section 3 discusses its issues. Section 4 proposes a network architecture for a large-scale B-BMA, and Sections 5, 6 and 7 describes its components respectively. Evaluation using simulations is shown in Section 8. Related work is shown in Section 9, and Section 10 concludes this paper.

2. Property of UDL and Emulated B-BMA Using LLTM

A UDL is a data link that works as a down link from a transmitter to one or

[†1] Graduate School of Media and Governance, Keio University
[†2] Faculty of Environment and Information Studies, Keio University

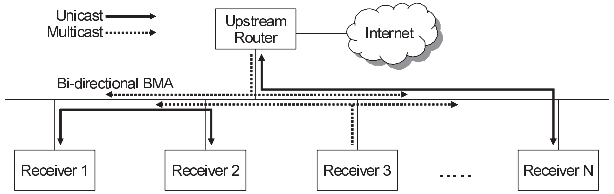


Fig. 1 Emulated network topology and expected traffic.

Table 1 Reference properties of UDLs.

Type of UDL	Bandwidth	Number of Receivers
Satellite TV (ISDB-S[4])	52 Mbps	4.2×10^6
Terrestrial TV (ISDB-T for Fixed[5])	24 Mbps	5.1×10^7
Terrestrial TV (ISDB-T for Mobile[6])	300 kbps	5.8×10^7
Terrestrial Radio (ISDB-TSB[7])	1 Mbps	N/A

more receivers. Receivers do not have the capability of direct transmission via the UDL. LLTM creates a B-BMA topology (**Fig. 1**) where an upstream router in the transmitter system of UDL and receivers (Receivers) are on an Ethernet link with long propagation delay.

Table 1 shows the properties of several types of UDLs. According to reports on the number of subscribers or receivers of satellite and terrestrial broadcasting[8,9], the expected number of receivers tends to be large on UDL, and its bandwidth is limited compared to other LAN or WAN technologies. Reception quality at receivers differs according to the receiver's physical distance from the transmitter, mobility or shielding attenuation.

Figure 2 shows the physical connections of nodes to B-BMA with a modified version of LLTM that uses Feed Bridge. **Table 2** describes the notations used in Fig. 2. The connectivity from Upstream Router to B-BMA is typically Ethernet LAN via Feed Bridge. On a Receiver a bi-directional link to B-BMA on RUI is emulated to enable transmission from the Receiver to B-BMA, where the actual transmission path is via RBI.

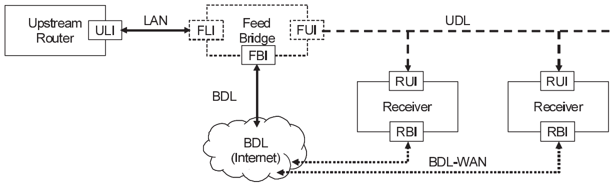


Fig. 2 Node connections on UDL using LLTM.

Table 2 Interfaces of nodes on bidirectional BMA.

Node	I/F	Description
Upstream Router	ULI	Upstream LAN Interface to connect to B-BMA
Feed Bridge	FLI	Feed LAN Interface to connect to Upstream Router
	FUI	Feed UDL Interface to transmit to UDL
	FBI	Feed BDL Interface to work as the end point of LLT
Receiver	RUI	Receiver UDL Interface to receive UDL, and connect to the emulated B-BMA using LLTM
	RBI	Receiver BDL Interface to work as the origin of LLT

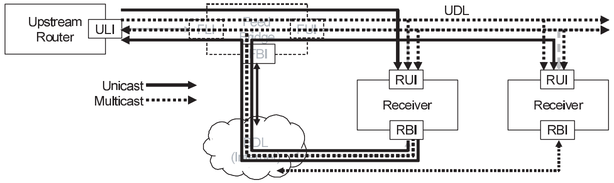


Fig. 3 Physical paths of Bi-directional BMA.

Unicast and multicast communications between Receivers are achieved by Broadcast Emulation, where Feed Bridge transmits a data link frame received from Receiver to UDL on behalf of the source Receiver as shown in **Fig. 3**.

3. Issues of B-BMA Using LLTM

Both the standard and modified LLTM have the following issues if we adopt the existing network architecture onto the emulated B-BMA:

P1 Lack of redundancy at upstream UDL,

P2 Long delay, caused by suboptimal path using LLTM, decreases protocol performance of communication between Receivers on B-BMA, and

P3 Control messages sent via UDL unnecessarily consume bandwidth.

3.1 Lack of Redundancy

On the topology using LLTM, the traffic from Receivers flocks to the upstream network of UDL because the LLTM end point resides there. However, both the standard and modified LLTM lack a mechanism to reduce the load and to provide redundancy at the LLTM end point. According to the specifications, Feed and Feed Bridge can advertise multiple LLTM end points using the Dynamic Tunnel Control Protocol (DTCP) HELLO message. These end points are prioritized and Receivers choose the preferred LLTM end point by default. This protocol specification may cause a bottle neck or a single point of failure.

3.2 Performance Reduction of Communication between Receivers

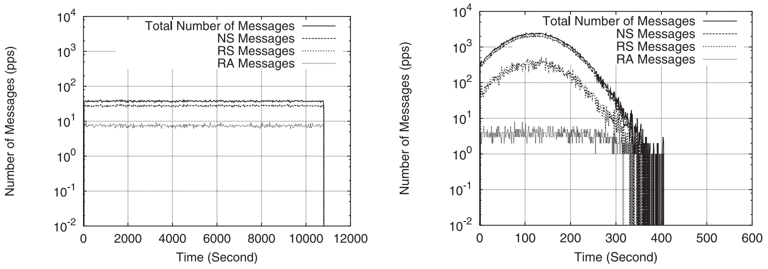
The communication path between Receivers on B-BMA using Link-layer Tunnel may not be optimal. Moreover, one-way delay on UDL tends to be long because of its physical link distance or FEC mechanism using time interleave. Even if the actual network distance between Receivers via BDL is very short, Receivers communicate using B-BMA because there is no mechanism to provide any information about alternative connectivity other than via B-BMA. In such a situation, TCP performance, for example, will decrease.

3.3 Unnecessary Consumption of UDL Bandwidth

IPv6 is a promising protocol for identifying a large number of receivers on B-BMA because of its large address space, thus this paper focuses the discussions on IPv6 as the network layer protocol. On IPv6 stateless address autoconfiguration (RFC4862 [12]), Broadcast Emulation in LLTM does not scale because multicast packets are used in configuring the network interface or the routing table at each receiver. This situation is undesirable as UDL bandwidth will be used for delivering control messages, not service traffic.

3.3.1 Duplicate Address Detection · Router and Prefix Discovery

Duplicate Address Detection (DAD [12]) and Router and Prefix Discovery (RPD [11]) are the basic procedures of IPv6 stateless address autoconfiguration. The scalability of DAD and RPD is both $O(N)$ in terms of the relationship between the number of nodes and the bandwidth consumption of UDL. DAD does not have a mechanism for message suppression. Meanwhile, in RPD the advertising router will respond to multiple Router Solicitation (RS) messages by a single Router Advertisement using multicast. Also, soliciting nodes may cancel sending messages if they receive Solicited or Unsolicited RA message. However, these mechanisms do not radically improve the scalability.



(a) Steady state (b) Flash crowd

Fig. 4 Number of IPv6 control messages sent via UDL.

Figure 4 shows the number of Neighbor Solicitation (NS), RS, and RA messages when 3×10^5 Receivers perform DAD and RPD in (a) a steady state scenario and in (b) a flash crowd scenario, as described in Section 8. The traffic on UDL was 73 packets per second (pps) at the busiest second, and 37.50 pps on average during the procedure of DAD and RPD in the steady state scenario. For the case of the flash crowd scenario, the traffic was 2,493 pps at the busiest second, and 871.07 pps on average. With the frame length of an NS, RS, and RA message being 86, 70, and 110 bytes respectively, up to 1.66 Mbps traffic was transferred on UDL in this simulation result.

The probability of an address conflict between two nodes on the same data link is very small [10]. Hence we expect very few Neighbor Advertisement (NA) messages to be sent via UDL even if a large number of Receivers connect to the B-BMA. We can see that it is not efficient to transmit $O(N)$ data on UDL to detect very few conflicts.

3.3.2 Address Resolution · Neighbor Unreachability Detection

Address Resolution (AR [11]) is performed to learn the MAC address of a neighbor node with a certain IPv6 address. A soliciting node sends an NS message to the data link using multicast, and the target node responds to the query by sending an NA message using unicast. Neighbor Unreachability Detection (NUD [11]) will be performed within a certain interval to confirm that the communicating neighbors are still reachable. In NUD, both NS and NA messages are exchanged using unicast.

When Receivers only communicate with the default gateway on the prefix, the order of the number of control messages sent via UDL will be $O(N)$. However, in the worst case when all nodes are communicating with all other neighbors the order becomes $O(N^2)$.

4. Architecture

This paper proposes a scalable network architecture for a large-scale B-BMA using IPv6. Our architecture achieves the optimal utilization of UDL bandwidth for services. Here we present the requirements for our system:

R1 Reduce the possibility of a single point of failure at the upstream UDL,
R2 Reduce the number of control messages sent via UDL,
R3 Provide an alternative path for communication between Receivers.

The basic ideas of our approach to meet these requirements are (1) extending the modified LLTM, and (2) introducing the concept of ***Adjacency*** between Feed Bridge and Receiver to enable IPv6 to operate in a scalable manner. Our architecture extends the modified LLTM as follows:

E1 Modify DTCP to periodically advertise a set of LLTM end points with no specified priority, from which Receivers will randomly choose one, and
E2 Make the Feed Bridge maintain Adjacency with Receivers.

Our approach introduces the following new mechanisms, detailed in Section 5 to 7:

M1 Extensible Upstream Backbone (EUB) to achieve R1: a network topology where additional Feed Bridges can be installed to create redundancy and to prevent a bottleneck,
M2 Adjacency Handler (AH) to achieve R1, R2 and R3: a mechanism where the Feed Bridge establishes two-way connectivity with each Receiver using BDL, thus enabling network configuration of a Receiver without using B-BMA, and
M3 De Facto Neighbor Discovery (DF-ND) to achieve R2 and R3: a new neighbor discovery mechanism that enables identification among Receivers via both B-BMA and BDL.

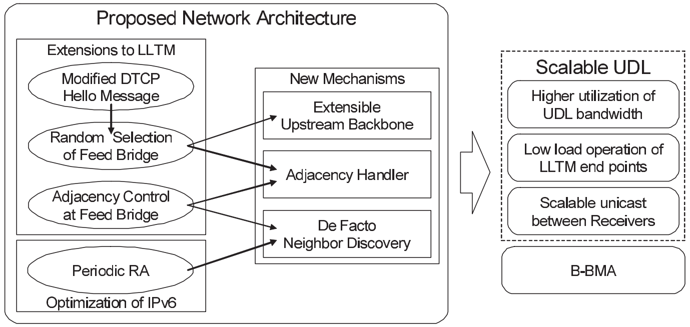

Proposed Network Architecture
Extensions to LLTM
Modified DTCP Hello Message
Random Selection of Feed Bridge
Adjacency Control at Feed Bridge
Periodic RA
Optimization of IPv6
New Mechanisms
Extensible Upstream Backbone
Adjacency Handler
De Facto Neighbor Discovery
Scalable UDL
Higher utilization of UDL bandwidth
Low load operation of LLTM end points
Scalable unicast between Receivers
B-BMA

Fig. 5 Components of proposed architecture.

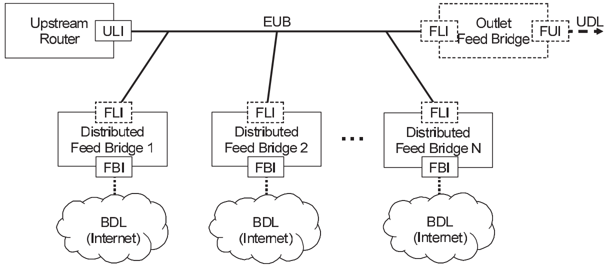


Fig. 6 Extensible upstream backbone (EUB).

Figure 5 depicts the components in our architecture. In addition to these components, DF-ND requires the transmissions of RA messages to be periodic rather than on-demand using a back-off random timer.

5. Extensible Upstream Backbone

The Extensible Upstream Backbone (EUB) is the upstream network of UDL where multiple Feed Bridges can be installed (**Fig. 6**). EUB corresponds to the LAN that appeared in Fig. 2.

On EUB, Feed Bridge is categorized into two types: Outlet Feed Bridge (OFB) and Distributed Feed Bridge (DFB). OFB is directly connected to UDL to bridge the traffic from Link-layer Tunnel or Upstream Router to the UDL. DFB load balances traffic from the Receiver as well as provides redundancy.

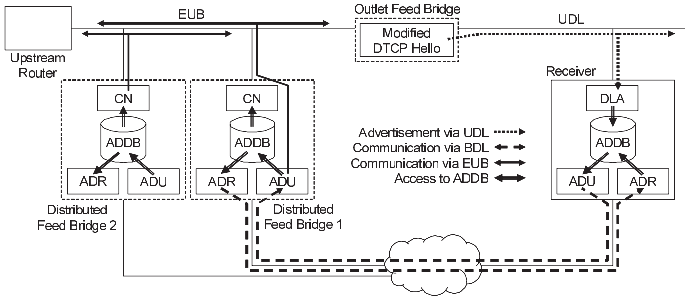


Fig. 7 Adjacency handler overview.

6. Adjacency Handler using Extended LLTM

Adjacency Handler (AH) is a mechanism to store and update the information about Receivers at the EUB, and then to enable network configuration at Receivers. Adjacency is a two-way connectivity between Receiver and DFB via BDL. A Receiver should always maintain adjacency with a DFB. As long as the Receiver is adjacent to a DFB, Receiver assumes that: (1) the uniqueness of Receiver's IPv6 address on B-BMA is assured and (2) DFB answers the queries that solicit information of the Receiver. **Figure 7** shows an overview of AH covering the following components:

- Adjacency Database (ADDB),
- Modified DTCP Hello, and
- Activity-based Feedback Suppression.

6.1 Adjacency Database

Receivers and DFBs maintain a database called *Adjacency Database* (ADDB) to record the active nodes on B-BMA. A Receiver and a DFB are adjacent to each other if both have the ADDB entry of the other. An ADDB entry contains the following parameters:

(1) IP Address of UDL I/F,
(2) MAC Address of UDL I/F,
(3) IP Address of BDL I/F,
(4) Type of Node, and
(5) Expiry.

Table 3 Type of node.

Type	Name	Description
1	Adjacent Receiver	DFB recognizes Adjacent Receiver as adjacent to maintain the state
2	Cache Receiver	DFB will establish an adjacency with the Cache Receiver when the Adjacent DFB becomes unavailable
3	Adjacent DFB	Receiver recognizes Adjacent DFB as adjacent
4	Backup DFB	Receiver will establish adjacency with Backup Receiver when Adjacent DFB is unavailable
5	Upstream Router	Receiver recognizes Upstream Router as the next hop to the Internet via B-BMA
6	Receiver	Receiver recognizes this Receiver as reachable on B-BMA

The types of nodes that Receivers or DFBs recognize are shown in **Table 3**. Each ADDB entry expires after a certain period of inactivity. Access to ADDB is triggered by AH, and DF-ND to refer, add, delete, or update its entry.

6.2 Modified DTCP Hello

OFB periodically transmits DTCP Hello messages via UDL to advertise connectivity to FBIs of DFBs. DTCP Hello messages are modified to contain the following parameters:

(1) Minimum Expiry of ADDB Entry ($E_{\min}$), and
(2) Maximum Expiry of ADDB Entry ($E_{\max}$).

Receiving the modified DTCP Hello, Receivers passively discover DFBs, and then randomly select one of the advertised DFBs as a candidate for Adjacent DFB.

6.3 Receiver Operation

Each Receiver transmits an Adjacency Update (ADU) to its chosen candidate Adjacent DFB to newly establish an adjacency. ADU contains the following parameters:

(1) Type (ADU),
(2) RUI IP Address,
(3) RUI MAC Address,
(4) RBI IP Address, and
(5) Expiry.

Here Receiver sends its IPv6 Link-local Address as the RUI IP Address.

After sending ADU, Receiver waits for Adjacency Reply (ADR) from a corresponding DFB that notifies the Receiver of the adjacency establishment. ADR contains the following parameters:

(1) Type (ADR),
(2) Registered RUI IP Address,
(3) Status of Adjacency,
(4) FBI IP Address of Backup DFB, and
(5) Expiry.

Receiver repeats sending ADU at a certain interval until it receives an ADR. If an ADR included the notification "Established", then Receiver updates its ADDB to activate the source of ADR as Adjacent DFB, and confirms its network configuration is unique and valid on B-BMA. Receiver also learns FBI IP Address as Backup DFB. If the ADR is marked "Discarded", Receiver re-configures its network interface and sends ADU again.

Receiver sends ADUs to Adjacent DFB as keep-alive of adjacency. When Receiver receives an ADR marking the adjacency state as "Established", Receiver confirms the continuation of adjacency and updates the entry of Adjacent DFB in ADDB.

When the adjacency is lost, Receiver may also send ADU to the Backup DFB. If Receiver does not have adjacency, then Receiver should refrain from sending any packet via B-BMA until it establishes a new adjacency.

6.4 Keep-alive Using Activity-based Feedback Suppression

Receiver reflects the status of unicast communication on B-BMA to dynamically control the Expiry set in the ADU. Using the Expiry, the life time of its entry in DFB's ADDB, and the inter-transmission time of ADUs are determined to control the frequency of sending ADUs.

Receiver calculates the average inter-arrival time (in seconds), I_{avg}, of unicast packets to itself, or to addresses that it has to forward. Receiver determines the Activity Factor α every E_{min} seconds as follows:

$$\alpha = e^{I_{\mathrm{avg}}}. \tag{1}$$

Receiver uses α to calculate a suitable Expiry to be set in the next ADU, E_{next}:

$$E_{\mathrm{next}} = \begin{cases} \alpha E_{\mathrm{min}} & \text{where } \alpha E_{\mathrm{min}} < E_{\mathrm{max}}, \\ E_{\mathrm{max}} & \text{otherwise,} \end{cases} \tag{2}$$

and then determines the time to transmit the next ADU. If E_{next} is shorter than the current expiry, Receiver sends an ADU with Expiry E_{next} in E_{next} second. If E_{next} is equal to or longer than the current Expiry, Receiver schedules the ADU to be sent with Expiry E_{next} before the current expiry.

If Receiver does not actively communicate using unicast on B-BMA, the expiry and inter-transmission time of ADUs increase exponentially. Hence, the Receiver can keep the adjacency for a long time by sending only a few ADUs.

6.5 DFB Operations

Every time a DFB receives an ADU, it searches the ADDB to confirm the status of the corresponding entry to RUI IP Address set in the ADU:

NEW Receiver's RUI IP Address is unique,
ACTIVE Receiver's RUI IP and MAC Address exist as a single entry, and
CONFLICT Receiver's RUI IP Address is already used by another Receiver.

When an ADU is received from BDL, the DFB works as Adjacent DFB that is responsible for sending ADR to the Receiver. If the search result in ADDB is not CONFLICT then the Adjacent DFB multicasts the ADU to let another DFB report CONFLICT if detected.

Other DFBs will receive ADU delivered using multicast via EUB, those DFBs examine the ADU in ADDB to detect CONFLICT. If CONFLICT is detected, the DFB sends a negative ADR as Conflict Notice (CN) using multicast on EUB, but otherwise it just discards the ADU.

If no CONFLICT is reported within the set period of time, Adjacent DFB sends a positive ADR to Receiver to notify the establishment of adjacency as "Established". In CONFLICT case, the entry corresponding to the ADU is overwritten by the information in the CN, and then the Adjacent DFB sends a negative ADR with status "Discarded" to the Receiver.

For garbage collection, an ADU that matches an expired entry is handled as NEW. Any other message that refers an expired entry is treated as that for a nonexistent Receiver. Then the expired entry is dropped from ADDB.

6.6 Redundancy of Adjacency on EUB

When a DFB establishes adjacency, the DFB unicasts the ADU to another DFB that is randomly selected on EUB. The purpose of this process is to make another DFB cache the entry of Adjacent Receiver for redundancy. When a DFB receives ADU using unicast via EUB, the DFB adds the ADU to its ADDB and then works as Backup DFB in case Adjacent DFB goes down.

Making multiple DFBs to partially share their ADDB helps to keep the size of ADDB small on each DFB compared with mirroring the ADDB on all DFBs. Let d be the number of DFBs that are active on EUB, and r be the number of total Receivers. If the number of Adjacent Receivers is well distributed to the available DFBs, the number of entries for Adjacent Receiver will be $\frac{r}{d}$ at each DFB. Here let R be the number of DFBs that share an ADDB entry for redundancy. Each Adjacent DFB sends ADU to $R-1$ DFBs, every time selected randomly, using unicast via EUB. Now each DFB is expected to receive $\frac{r}{d} \times \frac{1}{(d-1)}$ cache entries from $\{(R-1) \times (d-1)\}$ DFBs respectively. In the case of $R = 2$ (two DFBs share an ADDB entry), the total number of adjacent entries and cache entries in each ADDB is $\frac{2r}{d}$.

6.7 Adjacency Failover

When the adjacency is lost at Receiver, or FBI IP Address of Adjacent DFB does not appear in the modified DTCP Hello, it is possible that the Adjacent DFB has gone down. In such cases, Receiver sends an ADU to the Backup DFB to re-establish the Adjacency. This failover is quicker compared to newly establishing Adjacency with a non-Backup DFB because the procedure for failover is completed when the Backup DFB receives an ADU from a Cache Receiver.

When the Backup DFB receives an ADU from the Receiver, the DFB is now Adjacent DFB and sends ADU to another DFB that is randomly selected out of active DFBs on EUB. As long as Backup DFB is active on EUB, DF-ND should function seamlessly even when the Adjacent DFB becomes unavailable.

7. De Facto Neighbor Discovery

De Facto Neighbor Discovery (DF-ND) is a mechanism where neighbor discovery is proxied by DFBs based on adjacency without communication via UDL or BDL. Also, DF-ND provides an alternative connectivity via BDL for communication between Receivers. **Figure 8** shows an overview of DF-ND, which is composed of the following mechanisms:

- Periodic Router Advertisement,
- Downstream Neighbor Discovery, and
- Alternative Neighbor Discovery.

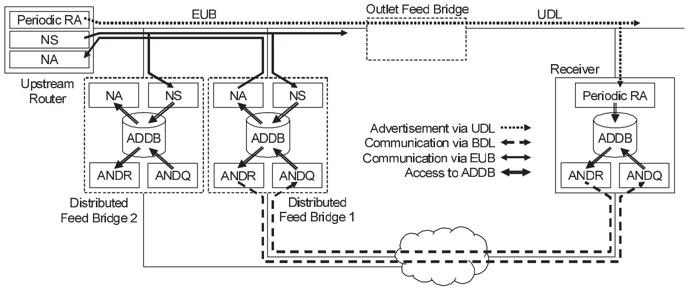


Fig. 8 Downstream neighbor discovery overview.

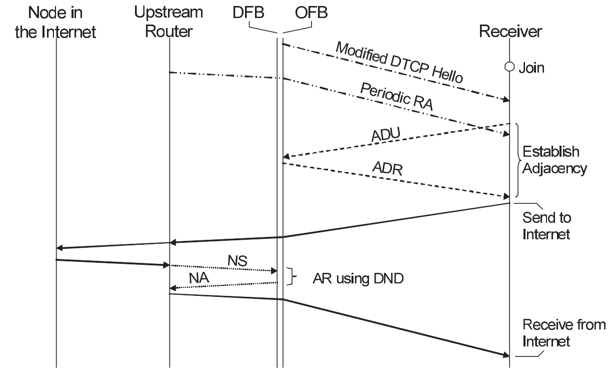


Fig. 9 Transactions in proposed approach.

7.1 Periodic Router Advertisement

Periodic RA configures Upstream Router to periodically transmit RAs more frequently than the current specification dictates. Periodic RA forces Upstream Router to advertise the following options without modifying the existing packet format of RA:

(1) Prefix Information Options for B-BMA Prefix, and

(2) Source Link-Layer Address Option for ULI MAC Address.

Receiving the RA, Receiver passively discovers Upstream Router, and then automatically configures its RUI IP Address. If Receiver has adjacency, Upstream Router and Receiver are deemed to be reachable, and Receiver can send to the Internet using Upstream Router as the gateway as shown in **Fig. 9**. On Up-

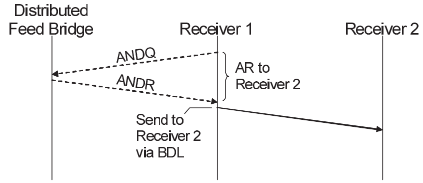


Fig. 10 AR between receivers using AND.

stream Router, the destination of a packet coming from the Internet may not be the neighbor. Neighbor Discovery to Receiver is performed as described in the following section.

7.2 Downstream Neighbor Discovery

Downstream Neighbor Discovery (DND) is a mechanism where DFBs proxy AR and NUD from Upstream Router to Receiver. DND is expected to extend Neighbor Discovery Proxies (RFC4389[13]) so that DFBs refer to ADDB. Each DFB captures any NS message from Upstream Router and searches for a corresponding entry in ADDB. If no entry matches the query, DFB discards the query without responding to Upstream Router.

If a DFB is Adjacent with the queried Receiver, the DFB immediately sends an NA message to answer using multicast. The Backup DFB delays sending an NA message even if the DFB has the corresponding entry, and cancels sending the response when it receives the response multicast on EUB. Upstream Router gains reachability to Receiver in data link layer and network layer on B-BMA by receiving the NA message.

7.3 Alternative Neighbor Discovery

7.3.1 AR between Receivers

DFBs have Receiver's RBI IP address in the ADDB entry to facilitate communication between Receivers via BDL. In Alternative Neighbor Discovery (AND), DFB provides RBI IP Address with Receiver that performs AR to another Receiver.

In **Fig. 10** Receiver sends AND Query (ANDQ) to its Adjacent DFB for AR to another Receiver. ANDQ contains the following parameters:
(1) Type (ANDQ), and
(2) Target RUI IP Address.

Receiving ANDQ, DFB looks up the target Receiver in ADDB. If the DFB finds the corresponding entry, the DFB responds to ANDQ by sending AND Reply (ANDR). ANDR includes:
(1) Type (ANDR),
(2) Target RUI IP Address,
(3) Target RUI MAC Address, and
(4) Target RBI IP Address.
If there is no corresponding entry in the DFB's ADDB, the DFB multicasts the ANDQ to ask another DFB for the information. If the DFB gets the answer on EUB, the DFB sends ANDR to the Receiver.

Receiving ANDR, Receiver acquires connectivity to the target Receiver via both B-BMA and BDL. For communication between Receivers, one encapsulates the data link frame that was to be sent via B-BMA in Generic Routing Encapsulation (GRE) header to transfer via BDL. The other, which receives a GRE encapsulated packet from BDL, decapsulates the original data link frame and process it like LLTM decapsulation at Feed Bridge. As long as Receivers transfer communication data via BDL, the connectivity should be maintained by NUD between Receivers described in the next section. If no ANDR is received at Receiver, that means "target does not exist", or "False Negative" because ANDR or ANDQ is lost. In both cases, receiver should retry sending ANDQ after a certain interval.

7.3.2 NUD between Receivers

To communicate between Receivers using UDL, a Receiver may need NUD to the target. Once Receiver successfully discovers the neighboring Receiver, Receiver performs NUD to the neighbor by communication via BDL. As shown in **Fig. 11**, Receiver1 sends ANDQ to the RBI IP Address of Receiver2, that is already known using AR to Receiver2 using AND.

When Receiver2 receives ANDQ from Receiver1 via BDL, it sends ANDR to the source of ANDQ. If Receiver1 receives ANDR from Receiver2, NUD to Receiver2 is successfully completed. Receiver1 may send ANDQ up to certain times when ANDR does not come. When Receiver1 determines that Receiver2 is not reachable via BDL, the Receiver terminates communication using B-BMA and the NUD process falls back to AR using AND.

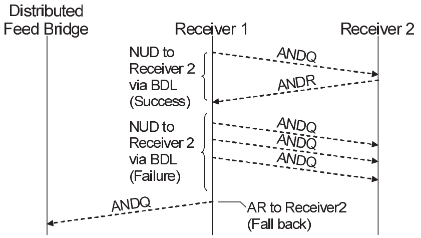

Distributed Feed Bridge
Receiver 1
Receiver 2
NUD to Receiver 2 via BDL (Success)
ANDQ
ANDR
NUD to Receiver 2 via BDL (Failure)
ANDQ
ANDQ
ANDQ
ANDQ
AR to Receiver2 (Fall back)

Fig. 11 NUD between receivers using AND.

8. Evaluation

8.1 Simulation Setup

We evaluate our system using two simulation scenarios with different models for node joins: a steady state scenario and a flash crowd scenario. In the steady state scenario, the inter-arrival time of Receivers to join the data link follows an exponential distribution as reported in the multicast experiment in MBone[14]. In the flash crowd scenario, the arrival time of Receivers is configured to join in a very short time period as reported in the WWW service analysis of World Cup football games[15]. The configuration of these scenarios is described in **Table 4**.

On each Receiver, I_{avg} is randomly calculated using exponential distribution at each calculation period as determined by $E_{\min}$. The λ of the random generator for each Receiver's I_{avg} follows a zipf distribution so that the majority of Receivers are listening to a multicast service via B-BMA, and the rest of the Receivers actively use B-BMA for a unicast service as described in Section 1.

Common parameters of the simulation are shown in **Table 5**. The UDL delay is based on the result of system evaluation of ISDB-Tsb system[3], and the BDL delay is based on the result of a preliminary test using ICMP echo and reply (ping) between a server connecting a high-speed Internet backbone and an iPhone 3G terminal connecting a 3G network in steady state.

Table 4 Configuration of simulation scenarios.

Steady Scenario	Model	Inter-arrival Time
	Random Generator	Exponential Distribution
	Parameters	$\lambda = 0.036$ (second)
Flash Crowd Scenario	Model	Arrival Time
	Random Generator	Normal Distribution
	Parameters	$\mu = 120$ (second), $\sigma^2 = 60$

Table 5 Simulation settings.

Parameter	Value
BDL Delay	$\mu = 35$ (ms), $\sigma^2 = 1$
UDL Delay	$\mu = 453$ (ms), $\sigma^2 = 1$
Periodic RA Interval	$I_{RA} = 1$ (second)
Modified DTCP Hello Interval	$I_H = 1$ (second)
Number of DFBs	$d = 10$
Redundancy	$R = 2$ (1 Backup DFB for each entry)
Active Duration	Min = 600, Max = 3,600 (second)
Average Packet Loss Rate	$(P_{UDL},\ P_{BDL}) = (0.05,\ 0.02),\ (0.01,\ 0.005)$

8.2 Bandwidth Consumption on UDL by One-way Advertisements

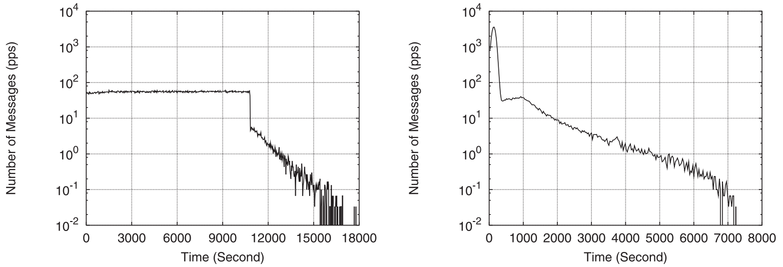
On our network architecture, the control messages that are sent via UDL are Modified DTCP Hello and Periodic RA. The total number of control messages sent via UDL within a certain time duration T is M_{ADV}:

$$M_{ADV} = \frac{T}{I_{RA}} + \frac{T}{I_H}. \tag{3}$$

M_{ADV} is independent of the number of Receivers. In our simulation the number of one-way advertisements via UDL is 2 packets per second (pps) because the inter-transmission time of Periodic RA and the modified DTCP Hello is 1 second for each. Compared to the simulation results shown in Fig. 4, the consumption of UDL bandwidth caused by control messages is drastically reduced in our architecture.

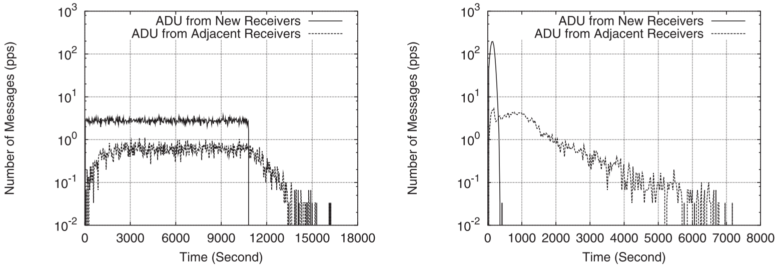
8.3 Performance of Adjacency Handler

The number of AH messages is $O(N)$ in the transient phase because AH does not have a mechanism to suppress transmission of control messages before each Receiver establishes Adjacency. In the keep-alive phase, the number of AH messages is expected to radically decrease in both scenarios. In this section we analyze the performance of AH based on the result of simulating 3×10^5 Receivers in the steady state and flash crowd scenarios. The average packet loss rate among simulated Receivers for these results is $(P_{UDL},\ P_{BDL}) = (0.05,\ 0.02)$. The other results of simulations did not exhibit a significant difference introduced by the



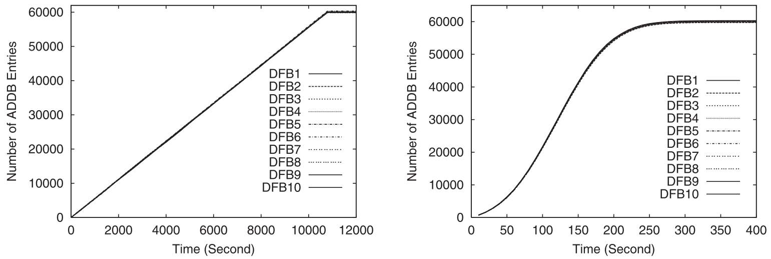
(a) Steady state (b) Flash crowd

Fig. 12 Number of ADU messages on EUB.



(a) Steady state (b) Flash crowd

Fig. 13 Number of incoming ADU messages at DFB1.



(a) Steady state (b) Flash crowd

Fig. 14 Growth of ADDB at each DFB.

configuration of packet loss rate.

Figure 12 (a) shows the number of ADUs sent via EUB averaged every 30 seconds through the simulation time of the steady scenario. The traffic on EUB was 100 pps at the busiest second, and 55.08 pps on average during the transient phase, whose duration is approximately 10,800 seconds. Figure 12 (b) shows the result of the flash crowd scenario, where 3×10^5 receivers joined the data link in approximately 330 seconds. The traffic by ADUs was approximately 3,826 pps in the busiest second on EUB, and 1,664.51 pps on average during the transient phase. For both scenarios, in the keep-alive phase Receivers exchange ADUs and ADRs with their Adjacent DFB until they leave the data link. The number of messages sent via EUB in the keep-alive phase is very small because of the activity-based feedback suppression of AH mechanism.

Focusing on a single DFB, **Fig. 13** (a) and (b) show the average number of incoming ADUs via BDL through the simulation time for the steady state and the flash crowd respectively. The ADUs are determined to be an ADU from a new Receiver or from an Adjacent Receiver. In both scenarios, the number of incoming ADUs is approximately one tenth compared to Fig. 12 (a) and (b) because of random selection of the Adjacent DFB at Receivers. As we can see from the flash crowd scenario, the message suppression functions effectively for the large number of Receivers because the number of ADUs from the Adjacent Receivers is kept low during and after the transient phase.

Figure 14 (a) and (b) show the growth of the number of ADDB entries for the steady state and the flash crowd respectively. After the transient phase in both cases, the total number of ADDB entries on each DFB is approximately 60,000 on average, which is the expected result according to the redundancy and the number of DFBs configured in our simulations.

8.4 Configuration of Redundancy

In our simulation, the redundancy R was set as 2 to let two DFBs keep an ADDB entry. One factor that determines the selection of R for the number of operational DFBs d is the probability of unreachability of an ADDB entry on EUB, P_L which is given by:

$$P_L = \begin{cases} 0 & \text{where } g < R, \\ \frac{{}_R C_R \times {}_{(d-R)} C_{(g-R)}}{{}_d C_g} & \text{otherwise,} \end{cases} \qquad (4)$$

where g denotes the number of DFBs that are down simultaneously. If an ADDB entry becomes unavailable on EUB, the failover of adjacency will not function.

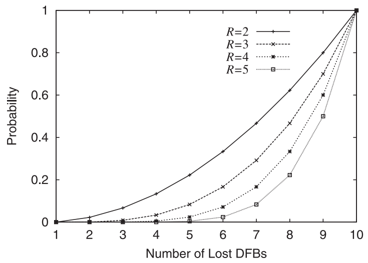


Fig. 15 Probability of loss of ADDB entry for 10 DFBs.

Also, DND and AND will be false negative even if the target Receiver is active on the data link. **Figure 15** shows the transition of P_L for ten operational DFBs, and P_L increases when g is equal to or bigger than R. In the case of $R = 2$, P_L marks approximately 0.12 when $g = 2$, and P_L increases up to 0.5 when $g = 7$. Meanwhile, P_L remains around 0.1 in the case of $R = 7$. Thus P_L will be the threshold to determine R by calculating the expected number of the unreachable entries in the case of simultaneous failures on EUB.

The other factor that determines R will be the trade-off between the memory consumption for keeping ADDB and the traffic load on EUB for including and searching ADDB entries. If R is big, the traffic on EUB will increase because a DFB will send the ADU from a Receiver to $(R-1)$ DFBs using unicast. Also, the small R reduces the memory consumption with the less redundancy. On the other hand, given R out of d DFBs are Adjacent or Backup for the queried Receiver, the probability where a query, namely an ANDQ message, will be immediately answered by a DFB without communication via EUB, P_q is $\frac{R}{d}$. Here the small P_q leads to the high frequency of communication on EUB to answer a certain number of queries. However, the benefits of the high-speed LAN technologies like 10 Gbps Ethernet facilitates the operation of DFBs with reduced memory consumption to handle an extraordinary number of Receivers.

8.5 Discussion regarding Impact of Mobility of Receivers

If Receiver is a mobile or multi-homed node, the Receiver might experience change of RBI IP Address, for example if the node switches from 3G to Wi-Fi. Here a factor that may decrease the reliability and scalability of our architecture is the mobility of Receivers. The neighborship between Receivers is maintained directly via BDL without interaction with DFB after AR is successfully completed. Hence the number of NUD messages will be $O(1)$ as long as Receivers know the RBI IP Address of each other to maintain the neighborship. However, the change of RBI IP Address may affect AR and NUD, because ANDR message may contain an RBI IP Address that is not reachable. Also, fault of NUD causes fall-back to AR to the neighboring Receiver that requires communication with the Adjacent DFB.

To avoid or reduce the impact of mobility of Receivers, a Receiver may send ADU to the Adjacent DFB, or send ANDR to the neighboring Receiver to inform the new RBI IP Address in a short time after the address change. However, if the mobility is high at many Receivers, the frequency of update from Receivers may exceed the effect of message suppression using Activity Factor. Another approach can be the use of Mobile IP [19] technologies that will hide the address change to the Adjacent DFB or the neighbor. However, the performance of communication via BDL will be subject to Mobile IP.

9. Related Work and Position of Our Architecture

Fujieda, et al. [16] extended Open Shortest Path First (OSPF) to improve scalability and completeness of protocol functions for the cases that OSPF is activated on a network topology that uses UDL. This extension was done by: defining UDL as a new link type, reducing the number of routers that synchronize LSDB with the designated router in a corresponding area, and introducing a mechanism to maintain consistency of LSDB among routers.

To avoid the feedback implosion that is expected when a large-scale multicast is activated, Thamrin, et al. [17] introduced a mechanism to record and select multicast listeners that transmit control messages to the link. Thamrin, et al. [18] also adapted a two-step random back-off timer to effectively delay transmission of control messages from multicast listeners. These approaches dramatically reduced the number of control messages required to handle large-scale multicast groups or sessions.

These efforts have been made to improve the scalability of network protocols on UDL. However, they do not address the coordination of data link layer and network layer of connecting nodes. Our architecture can be placed as the substrate for facilitating the above mentioned work.

10. Conclusion

This paper proposed a network architecture to use UDL as a scalable, broadcast capable, bi-directional data link for the Internet. The proposed architecture extends the existing LLTM, and introduces three new mechanisms based on the concept of *Adjacency*: (1) EUB to decrease traffic load of LLTM at the upstream of UDL, (2) AH and (3) DF-ND to configure data link layer and network layer of Receivers without communications via UDL.

Our approach minimizes the number of control messages sent via UDL regardless the number of connecting Receivers. The simulation results showed that our architecture reduces the control messages to be sent via UDL by more than 90% compared with enabling IPv6 stateless address autoconfiguration on the existing network architecture using the modified LLTM. The trade-off of our architecture is increased bandwidth consumption of LAN that can be heavily loaded for communication on AH and DF-ND. However, considering the capacity of the current LAN technologies and the benefit of multiple DFBs on EUB, such incremental load is still affordable and will not be the limiting factor in the scalability of the system.

References

1) Duros, E., Dabbous, W., Izumiyama, H., Fujii, N. and Zhang, Y.: A Link-Layer Tunneling Mechanism for Unidirectional Links, IETF, RFC3077 (2001).

2) Kataoka, K., Husni, T.A., Cho, K., Takei, J. and Murai, J.: Architecture of Satellite Internet for Asia-wide Digital Communications, *AINTEC 2007*, Springer LNCS, Vol.4866/2007, pp.242–255 (2007).

3) Kataoka, K., Kimura, H., Ishida, T., Kohara, H., Izawa, S., Kusumoto, H., Nakamura, O. and Murai, J.: An Architecture of IP Network over Broadcast Media, *IEICE Trans. Communications*, Vol.J91-B, No.12, pp.1669–1681 (2008).

4) ARIB: Transmission System for Digital Satellite Broadcasting, ARIB STD-B20, Ver.3.0 (2001).

5) ARIB: Transmission System for Digital Terrestrial Television Broadcasting, ARIB STD-B31, Ver.1.7 (2007).

6) ARIB: Operational Guidelines for Digital Terrestrial Television Broadcasting, ARIB TR-B14, Ver.3.8 (2008).

7) ARIB: Transmission System for Digital Terrestrial Sound Broadcasting, ARIB STD-B29, Ver.2.2 (2005).

8) SKY Perfect JSAT: Semiannual Report (2009). http://www.skyperfectjsat.co.jp/en/ir/library/2008_pdf/0809_SAR.pdf

9) JEITA Statistics Data (2009). http://www.jeita.or.jp/japanese/stat/digital/2009/pdf/200904digital.pdf

10) Moore N.: Optimistic Duplicate Address Detection (DAD) for IPv6, IETF, RFC4429 (2006).

11) Narten, T., Nordmark, E., Simpson, W. and Soliman, H.: Neighbor Discovery for IP version 6 (IPv6), IETF, RFC4861 (2007).

12) Thomson, S., Narten, T. and Jinmei, T.: IPv6 Stateless Address Autoconfiguration, IETF, RFC4862 (2007).

13) Thaler, D., Talwar, M. and Patel, C.: Neighbor Discovery Proxies (ND Proxy), IETF, RFC4389 (2006).

14) Almeroth, K.C. and Ammar, M.H.: Collecting and Modeling the Join/Leave Behavior of Multicast Group Members in the MBone, *5th IEEE International Symposium on High Performance Distributed Computing* (*HPDC-5 '96*), p.209 (1996).

15) Arlitt, M. and Jin, T.: A workload characterization study of the 1998 World Cup Web site, *IEEE Network*, Vol.14, No.3, pp.30–37 (2000).

16) Fujieda, S., Kusumoto, H. and Murai, J.: The Design of Extension for OSPF Handling Uni-Directional Links, *IEICE Trans. Communications*, Vol.J87-B, No.10, pp.1574–1585 (2004).

17) Thamrin, A.H., Kusumoto, H. and Murai, J.: Scaling Multicast Communications by Tracking Feedback Senders, *20th International Conference on Advanced Information Networking and Applications* (*AINA'06*), Vol.1, pp.459–464 (2006).

18) Thamrin, A.H., Izumiyama, H., Kusumoto, H. and Murai, J.: Delay Aware Two-Step Timers for Large Groups Scalability, *IEICE Trans. Communications*, Vol.E87-B, No.3, pp.437–444 (2004).

19) Johnson, D., Perkins, C. and Arkko, J.: Mobility Support in IPv6, IETF, RFC3775 (2004).

(Received June 16, 2009)
(Accepted December 17, 2009)

(Original version of this article can be found in the Journal of Information Processing Vol.18, pp.125–137.)

Kotaro Kataoka is a Ph.D. candidate in Media and Governance in Keio University. He graduated from the Faculty of Environment and Information Studies, Keio University in 2002. He received an M.M.G. from Graduate School of Media and Governance, Keio University in 2004. His research interests are Internet over broadcast media and post-disaster communications.

Achmad Husni Thamrin is Assistant Professor at Keio University. He is a graduate of Keio University, Graduate School of Media and Governance (Ph.D. in 2005, M.M.G. in 2002). His research interests include multicast, Internet over broadcast media, and peer-to-peer networks.

Jun Murai graduated from Keio University in 1979, Department of Mathematics, Faculty of Science and Technology. He received his Ph.D. and M.S. for Computer Science from Keio University in 1987 and 1981, and specialized in computer science, computer network and computer communication. He is currently Dean/Professor, Faculty of Environment and Information Studies, Keio University.