INTERNET DRAFT   C. Hsu, E.Muramoto, J. Buford
<draft-hsu-xcast-bcp-2004-01.txt>   Panasonic
                                                             Y. Imai
                                             WIDE Project/Fujitsu Lab
                                                         Ali Boudani
                                                        IRISA-France
                                                           R. Boivie
                                                                 IBM
                                                          July, 2005
                                             Expires January, 2006

     Best Current Practices of XCAST (Explicit Multi-Unicast) by 2004

Status of this Memo

   This document is an Internet-Draft and is subject to all provisions
   of section 3 of RFC 3667. By submitting this Internet-Draft, each
   author represents that any applicable patent or other IPR claims of
   which he or she is aware have been or will be disclosed, and any of
   which he or she becomes aware will be disclosed, in accordance with
   Section 6 of BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF), its areas, and its working groups. Note that
   other groups may also distribute working documents as Internet-
   Drafts.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time. It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."

   The list of current Internet-Drafts can be accessed at
   http://www.ietf.org/ietf/1id-abstracts.txt

   The list of Internet-Draft Shadow Directories can be accessed at
   http://www.ietf.org/shadow.html.

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in RFC 2119 [rfc2119].
   Allocation policy names Specification Required, IETF Consensus
   Action, and Designated Expert are to be interpreted as described in
   RFC 2434 [rfc2434]

Copyright Notice

Abstract

   In 2000, XCAST (Explicit Multi-Unicast) was first proposed as a
   protocol intended for small group multicasting. It combined three
   similar datagram simulcasting protocols that were submitted as
   earlier Internet Drafts by a number of different research groups.
   After that, researchers in those groups worked together to further
   develop, experiment and conduct standardization in IETF as well as in
   the open source community. This resulted in several implementations
   of protocol stacks, systems of multi-party video conference and group
   management, and mechanisms for harmonizing XCAST with ASM and SSM. In
   addition, an XCAST backbone for various experiments has been launched
   to operate inter-continentally.

   One of the main purposes of this document is to summarize the efforts
   undertaken by XCAST community as "best current practices" that would
   then be formally introduced to the IETF/IRTF community. In addition,
   we raise an issue concerning IANA resource allocation, which prevents
   us from widely expanding our experimental networks as well as
   distributing our software. Finally, we request IESG and other experts
   to check the validity of our proposed protocol and make any
   suggestions about how IANA can assign appropriate resources
   accordingly.

Table of Contents

10. Full Copyright Statement

11. IPR Notices

Changes:
00->01: added the section "Research Topics and Unsolved Problems".

1. **Standardization Efforts**

   XCAST (Explicit Multi-Unicast) is a datagram delivery protocol for
   efficient small group communications. It is based on the idea that a
   datagram is transmitted with an "explicit list of unicast addresses
   of multiple destinations" and each intermediate node (e.g., router),
   according to the information stored in the explicit list, forwards
   and branches (if needed) the datagram subsequently to the appropriate
   next hop by referring to its unicast routing table.

   In 1999, three documents were submitted as Internet-Drafts [CLM, SGM,
   MDO6]. Since the targeted area and mechanisms proposed in those I-Ds
   were similar, Ooms et al. combined the ideas of those I-Ds and came
   out with an Internet-Draft called "XCAST Basic Specification" [basic-
   spec] in 2000. At the same time, the authors of XCAST Basic
   Specification noticed an earlier proposal similar to their approach
   has been published by Aguilar [Aguilar].

   They started an introduction to XCAST technology at the MADDOGS, one
   of the Routing Area meetings in IETF, and received suggestions about
   moving forward to hold an IRTF RG or a BOF in 2001. Given the early
   stage of the activity, substantial running systems and strong demand
   from users, operators and service providers were not available, so
   when the first XCAST BOF was held in IETF, a consensus could not be
   reached at that time.

   After that, with increasing demands from users, the XCAST community
   has become ever larger with more nodes and running networks. Status
   of the XCAST community has been periodically reported to and
   discussed with IESG members, routing area directors (Abha, Bill and
   Alex) and transport area director Allison. The I-D, XCAST Basic
   Specification, has been updated to version 8. The differences among
   each version are trivial, showing that the proposed protocol is
   relatively stable.

2. **Development**

   To prove the concept, validity, interoperability and effectiveness of

this protocol, the XCAST community has collaborated to develop a
variety of XCAST-based software ranging from protocol stacks to
applications.

2.1 Protocol Stacks

2.1.1 XCAST4

IBM and Alcatel have implemented an XCAST protocol stack for IPv4 on
Linux and FreeBSD respectively. To validate their implementations,
they have conducted an experiment via inter-continental Internet
where the exchange of text messages was successful. The IBM version
is publicly available.

2.1.2 XCAST6

WIDE project/Fujitsu Laboratories and ETRI/Soongsil University have
implemented XCAST protocol stacks for IPv6 on NetBSD/FreeBSD and
Linux respectively. To verify multi-platform interoperability, they
have conducted conformance tests using XCAST with modified MBone
tools (vic & rat). A demonstration of multipoint video conferencing
and collaborative working among participants in Japan and Korea was
held at IETF 54 Yokohama. The ETRI/Soongsil University version is
publicly available [SNU-XCAST6]. The original release of WIDE
project/Fujitsu Laboratories is being maintained by the XCAST fan
club in Japan, and modified versions of the stable distributions of
NetBSD/FreeBSD are released on sourceforge.net accordingly
[XCAST6-kit].

2.2 Applications

2.2.1 MBone Tools

Well-known Mbone tools VIC and RAT have been modified and integrated
with XCAST. On the sender side, instead of binding the multicast
group address to the socket, the modified tools call XCASTAddMember()
to append each unicast address of the participated node. On the
receiver side, no modification is necessary since the payload of an
XCAST datagram can be acquired by calling ordinary recv() function.
The total amount of additional code is less than 200 lines for vic
and 150 lines for rat.

2.2.2 xcgroup

To use the above-mentioned Mbone tools, operators need to specify a
long list of IP addresses before conducting an experiment. This,
however, might not always be appropriate, since it is difficult to
maintain consistent membership information, i.e., the set of unicast

   addresses of XCAST nodes, among the participants.

   To solve this problem, we created a CGI script for httpd and a client
   program called "xcgroup". Operators invoke xcgroup with a URL of CGI
   for managing the information of group name and its membership (i.e.,
   participating nodes). Client xcgroup periodically sends a query to
   the CGI script by http and the CGI script acquires the IPv6 source
   address of the query and records it in a list. Then it retrieves all
   the IPv6 source addresses recorded in the list and provides the
   client program that information. As a result, consistent membership
   information among the participants can be provided. Xcgroup announces
   the list of acquired unicast addresses for Mbone tools via mbus
   [[rfc3259](rfc3259)].

[2.2.3](2.2.3) ping6x, traceroute6x

   Semi-permeable tunneling, one of the incremental deployment
   technologies adopted in XCAST, is utilized to forward the received
   XCAST packets to the next XCAST router by referring to the routing
   information obtained via any of the unicast routing protocols. This
   approach, however, lacks a mechanism to check the reachability of an
   overlay XCAST delivery tree. For this purpose, ping6x and
   traceroute6x were developed by modifying existing ping6 and
   traceroute6.

   To identify an overlay XCAST delivery tree, operators simply send a
   probing packet to one of the specific destinations listed in the
   XCAST header, since an XCAST datagram is treated as an ordinary
   packet by non-XCAST routers, ICMP response similar to those in
   unicast ping6 or traceroute6 can be obtained. As a result, the
   reachability of the overlay XCAST delivery tree can be identified by
   referring to the information of all the ICMP responses.

[2.2.4](2.2.4) Live CDs

   To use XCAST6, it is necessary to re-compile the kernel and install a
   new library. This prevents a wider acceptance and use of XCAST
   technology. To alleviate this problem, some open source communities
   based on NetBSD and FreeBSD, have collaborated and developed a Live
   CD solution for XCAST6. These Live CDs are called Ebifurya CD and
   XCAST6 enable FreeSBIE. The Live CD solution includes a pre-installed
   bootable image containing XCAST-based Mbone tools, USB camera tools,
   sound card drivers with utilities and an xcgroup client program.

[3](3). Experimental Networks

   To prove the concept of XCAST, WIDE XCAST WG started operating a
   weekly video conference meeting using an early version of XCAST,

i.e., Multiple Destination Option of IPv6, among seven Japanese
organizations and one university in the USA in 2000 via the WIDE
6Bone. Since the delivery tree was constructed using semi-permeable
tunneling technology at the beginning stage of the experiment,
improper daisy-chained connections always resulted in poor
performance. To overcome this problem, they launched another project
called X6Bone to construct a virtual network with v6/v4 tunnels
connecting many SOHO routers to a HUB branch XCAST router via ADSL.
With this, the transmission of XCAST datagrams could be branched at
the HUB router, thus eliminating the problem of inefficient daisy-
chained connections. In addition, a bi-monthly event called "Midnight
XCAST party" where a number of SOHO users were connected via
commercial ADSL had been held successfully 15 times.

By July of 2005, the number of groups joining the weekly meeting has
increased to nine organizations including BSD User groups, Linux User
groups, XCAST protocol user groups and IRISA/France. This is achieved
by allocating a delegated /40 prefix of IPv6 address space for each
group and then each group redistributes /48 for its members
respectively.

These results demonstrate that the concept of XCAST is feasible.
Since XCAST only requires the operation of current existing unicast
routing controls, it is extremely easy to operate compared to PIM.

4. **Harmonization with ASM and SSM**

For mass distribution services, such as live event broadcastings in
IETF, on-demand streaming videos and large-scale distance learning
courses, ASM/SSM is the most efficient mechanism to realize such
services since it is scalable with respect to the number of
receivers. On the other hand, for personal usage such as conducting a
small scale business meeting, chatting with a group of friends and
playing a multi-party on-line game, XCAST is the preferred means
since it does not require any routing state to be maintained in the
intermediate routers, thus it is scalable with respect to the number
of simultaneous small group formed.

We believe the combined use of ASM, unicast and XCAST further enables
the provision of wider service offerings. This section summarizes
current approaches to the harmonization of ASM, SSM and XCAST.

4.1 **XCAST+**

XCAST+ is an extension of XCAST intended for a more efficient
delivery of multicast packets [XCAST+]. Every source or destination
is associated to a Designated Router (DR). Instead of encoding in the
XCAST packet header the set of group members, XCAST+ encodes the set

of their DRs. When a new member wants to join the group G of source
S, it sends an IGMP-join message to its DR. The DR will then send a
join-request message to the source S. The DR of the source intercepts
this message and analyzes it in order to keep track of all concerned
DR addresses. When the source S wants to send a message to the group
G, it sends a multicast packet. This packet is received by its DR and
converted to an XCAST packet using the Multicast-to-XCAST algorithm
(M2X). The packet is then forwarded as in XCAST to the DRs, since the
destination list in the XCAST header contains the DR addresses
instead of the member addresses. After that, each DR converts the
XCAST packet to a multicast packet using the XCAST-to-Multicast
protocol (X2M) and sends it in its subnetworks.

4.2 **SEM**

As mentioned above, XCAST+ is designed to support a large number of
medium-sized groups. However, since the delivery mechanism used from
source DR to destination DRs is similar to that of the original
XCAST, the problem of the limited number of DRs that can be included
in the list of destinations still remains.

To solve this problem, SEM (Simple Explicit Multicast) introduces the
idea of deploying routers at branching point of the XCAST delivery
tree. This idea is similar to that proposed in REUNITE [REUNITE].
Both protocols require branching routers to keep state of (S,G) and
next branching router. The major difference between SEM and REUNITE
is how to maintain the state in each router. REUNITE periodically
sends probing packets back from DRs to S and checks the status of the
branch points. On the other hand, SEM forwards the probing packets
from S to DRs using XCAST+. With this, the SEM delivery tree is not
affected even if asymmetry exists in the unicast routing paths.

5. **Unsolved Problems**

As described in previous sections, XCAST experimentation is quite
active, and multiple implementations have been developed by several
research groups around the world. However, there remain some issues
that need to be resolved. This section aims to identify those issues
and provide guidance for current practices and further directions.

5.1 **Group management**

To allow application developers to utilize their own application-
specific group management functions with XCAST without thinking about
group address assignment or specific XCAST capabilities, the group
management function of XCAST should be completely separated from the
function of XCAST-based forwarding or routing. There are several
research topics that focus on how to integrate the group formation

and management functions of social networking or other interactive
communication applications with XCAST-based forwarding or routing
technology.

5.2 Transport Layer Support

Transport layer support plays a crucial role in achieving effective
date exchange among the participating nodes in an XCAST group.
Adequate congestion control, flow control and reliable end-to-end
delivery are examples of transport layer protocols. There are several
research topics that focus on how to avoid the congestion by
observing the network conditions with a light-weight protocol, how to
quickly isolate the congestion and how to fairly allocate sharing
network resource among connections between sender and multiple
receivers.

5.3 Harmonization with Broadcast

In addition to the harmonization models introduced in section 4,
there are several research topics that focus on how to effectively
utilize the broadcast capability of existing access systems, such as
satellite or wireless, with XCAST.

5.4 XCAST Tunneling

It must be clarified that XCAST tunneling has the scope of
applicability intended for inter-domain use. Although, as stated in
section 11.1 of [Basic], XCAST routers could exchange and maintain
routing information using any of standard protocols such as RIP,
OSPF, ISIS and BGP, this, however, needs to be further investigated.
In the case of intra-domain use, it might be enough to append
capability information to current intra-domain routing protocols and
facilitate the exchange of XCAST routing information, however in the
case of inter-domain use, injecting the entire BGP routing table into
the XCAST routing table might not be applicable.

In addition, the actual tunnel encapsulation, either IP in IP, GRE or
simply using the next XCAST router as the IP destination in the
regular IP header (i.e., without wrapper), is currently not concrete.

Furthermore, if multiple tunneling methods are to be adopted in
XCAST, mechanisms of how the participating nodes learn about what
modes they and their peer nodes are using, e.g, via manually
configured or automatically detected, need to be further discussed.

5.5 DSCP Operation

Although the usage of DSCP (DiffServ Code Point) option was described

   in section 8.2 of [Basic], the feasibility of its inter-domain use,
   such as re-writing the DSCP when packets cross domain boundaries or
   re-mapping DSCP when packets cross domain using XCAST tunneling,
   needs to be further investigated.

5.6 NAT Traversal and Firewall Penetration

   Solutions and approaches to the potential problems of NAT traversal
   and firewall penetration that arise in the usage of a different
   carried protocol type for IPv4, need to be further discussed.

5.7 BITMAP Handling

   To simplify the processing required in handling BITMAP in the list of
   destination of XCAST router, the type of the length of the BITMAP
   (i.e., fixed sized or variable) and its effects need to be further
   determined.

5.8 Network Deployment

   We believe that the total amount of traffic generated by XCAST-based
   group communication in the Internet backbone can be dramatically
   decreased if XCAST routers could be appropriately deployed in
   strategic points throughout the network. A deeper understanding of
   the requirement of the above-mentioned network deployment and an in-
   depth discussion of how to strategically places XCAST routers in the
   current or future Internet backbone (e.g., instead of placing XCAST
   routers into the core part of the backbone, it might be enough to
   deploy them at the edge) are needed.

6. IANA Considerations

   We had requested the IANA resource for XCAST in 2000 but the request
   was rejected due to the lack of suggestions from IESG and protocol
   experts.

   Given the growing interest in XCAST and the need to conduct more
   experimentation and testing between research groups, we would like to
   accelerate the deployment of XCAST aware nodes and routers. However,
   we are currently using protocol option values which are not
   officially assigned by IANA.

   To avoid confusion, we need experimental IANA resources described in
   a separate Internet Draft for Informational RFC [xc-iana]. In
   addition, we request IESG or other protocol experts to examine the
   validity of our proposal and make any suggestion to IANA accordingly.

7. Security Considerations

This document has no known security implications.

8. References

[rfc2026]   S. Bradner, "The Internet Standards Process -- Revision 3",
             BCP 9, RFC 2026, October 1996.

[rfc2434]   Narten, T., and Alvestrand, H., "Guidelines for Writing an
             IANA Considerations Section in RFCs", RFC 2434, October,
             1998.

[Aguilar]   L. Aguilar. Datagram Routing for Internet Multicasting. In
             ACM SIGCOMM '84 Communications Architectures and Protocols,
             pages 58-63. ACM, June, 1984.

[basic]   R. Boivie, et al., "Explicit Multicast (XCAST) Basic Speci-
             fication",draft-ooms-xcast-basic-spec-07.txt. January 2005
             Submitted for Informational RFC

[CLM]   D. Ooms, W. Livens, CONNECTIONLESS MULTICAST: A novel and
             scaleable method for multipoint-to-multipoint communication
             in IP networks, In "Proceedings of XVII World Telecommunica-
             tions Congress (WTC/ ISS 2000)", Birmingham, England, May,
             2000.

[SGM]   R. Boivie, et al., "Small Group Multicast: A New Solution
             for Multicasting on the Internet.", IEEE Internet Computing
             4(3): 75-79 (2000) http://csdl.com-
             puter.org/dl/mags/ic/2000/03/w3075.htm

[MDO6]   Y. Imai, et al., "MDO6: Multicast delivery system by the
             list of destinations." in Japanese, Internet Conference '99,
             http://www.internetconference.org/ic99/papers/demo05.pdf

[TEST-KRJP] "Xcast6 Interoperability Successfully Tested in Multicast
             Linking Korea and Japan", Press Release, August 2002
             http://pr.fujitsu.com/en/news/2002/08/1.html

[XCAST6-kit]
          http://sourceforge.net/projects/xcast6/

[Xcast+]   M. Shin, et al., "Multicast Datagram Delivery Using Xcast in
            IPv6 Networks.", ICOIN 2003: 263-272

[REUNITE]   I. Stoica, et al., "REUNITE: A recursive Unicast Approach to
            Multicast", http://www.ieee-infocom.org/2000/papers/613.ps.

[rfc3259]   J. Ott, et al., "A Message Bus for Local Coordination",
            RFC3259, April 2002

[xc-iana]   C. Hsu, et al., "IANA Considerations for XCAST (Explicit
            Multi-Unicast)", internet draft, http://www.ietf.org/inter-
            net-drafts/draft-hsu-xcast-iana-considerations-00.txt, April
            2005.

[rfc2119]   Bradner, S., "Key words for use in RFCs to Indicate require-
            ment Levels", BCP 14, RFC 2119, March 1997.

9. Authors Addresses

   Chih-Chang Hsu
   Matsushita Electric Industrial Co., Ltd.
   4-5-15 Higashi-shinagawa, Shinagawa-ku, Tokyo 140-8632, Japan
   Phone : +81-3-5460-2734
   E-mail: hsu.mike@jp.panasonic.com

   Eiichi Muramoto
   Matsushita Electric Industrial Co., Ltd.
   4-5-15 Higashi-shinagawa, Shinagawa-ku, Tokyo 140-8632, Japan
   Phone : +81-3-5460-2734
   E-mail: muramoto.eiichi@jp.panasonic.com

   Yuji Imai
   Fujitsu LABORATORIES Ltd.
   1-1, Kamikodanaka 4-Chome, Nakahara-ku, Kawasaki 211-8588, Japan
   Phone : +81-44-754-2628
   Fax   : +81-44-754-2793
   E-mail: kimai@flab.fujitsu.co.jp

John F. Buford
Panasonic Digital Networking Laboratory
Two Research Way, Princeton, NJ 08540, USA
Phone : +1-609-734-7342
Fax   : +1-609-987-8827
E-mail: buford@research.panasonic.com

Ali Boudani
IRISA/INRIA Rennes
Campus Universitaire de Beaulieu
Avenue du General Leclerc 35042 Rennes France
Phone : (33) 02 99 84 25 37
Fax : (33) 02 99 84 25 29
E-mail : aboudani@irisa.fr

Rick Boivie
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
Phone: 914-784-3251
Email: rhboivie@us.ibm.com

[10](10). Full Copyright Statement

[11](11). IPR Notices