Stardust Technologies, Inc
1901 S. Bascom Ave, #333
Campbell, CA 95008 USA
Tel: 408-879-8080
Fax: 408-879-8081
Web: www.ipmulticast.com

IP Multicast Initiative (IPMI)

How IP Multicast Works

An IP Multicast Initiative White Paper

A technical overview of IP Multicast concepts, addressing, group management and approaches to routing

Inside…

Stardust Technologies, Inc., 1901 S. Bascom Ave., Suite 333, Campbell, CA 95008

How IP Multicast Works

A technical overview of IP Multicast concepts, addressing, group management and approaches to routing.

Scope Of This Document

This document provides a technical introduction to IP Multicast concepts and technical features. It discusses the requirements for IP Multicast delivery, addressing and host group management, and approaches to multicast routing. Some familiarity with IP is assumed. If you are an engineer interested in evaluating or implementing IP Multicast, an understanding of the concepts in this document will help you. You may also be interested in other documents in this white paper series which are available from http://www.ipmulticast.com.

IP Multicast Mechanisms

Advantages of IP Multicast

Many emerging Internet applications are one-to-many or many-to-many, where one or multiple sources are sending to multiple receivers. Examples are the transmission of corporate messages to employees, communication of stock quotes to brokers, video and audio conferencing for remote meetings and telecommuting, and replicating databases and web site information. IP Multicast efficiently supports this type of transmission by enabling sources to send a single copy of a message to multiple recipients who explicitly want to receive the information. This is far more efficient than requiring the source to send an individual copy of a message to each requester (referred to as point-to-point *unicast*), in which case the number of receivers is limited by the bandwidth available to the sender. It is also more efficient than broadcasting one copy of the message to all nodes (*broadcast*) on the network, since many nodes may not want the message, and because broadcasts are limited to a single subnet.

Multicast is a receiver-based concept: receivers join a particular multicast session group and traffic is delivered to all members of that group by the network infrastructure. The sender does not need to maintain a list of receivers. Only one copy of a multicast message will pass over any link in the network, and copies of the message will be made only where paths diverge at a router. Thus IP Multicast yields many performance improvements and conserves bandwidth end-to-end.

IP Multicast Delivery and Groups

IP Multicast is an extension to the standard IP network-level protocol. RFC 1112, Host Extensions for IP Multicasting, authored by Steve Deering in 1989, describes IP Multicasting as: "the transmission of an IP datagram to a 'host group', a set of zero or more hosts identified by a single IP destination address. A multicast datagram is delivered to all members of its destination host group with the same 'best-efforts' reliability as regular unicast IP datagrams. The membership of a host group is dynamic; that is, hosts may join and leave groups at any time. There is no restriction on the location or number of members in a host group. A host may be a member of more than one group at a time." In addition, at the application level, a single group address may have multiple data streams on different port numbers, on different sockets, in one or more applications. Multiple applications may share a single group address on a host.

An Overview of What's Needed for IP Multicast

To support native IP Multicast, the sending and receiving nodes and network infrastructure between them must be multicast-enabled, including intermediate routers. Requirements for native IP Multicast at the end node hosts are:

- Support for IP Multicast transmission and reception in the TCP/IP protocol stack.
- Software supporting IGMP (see IGMP section below) to communicate requests to join a multicast group(s) and receive multicast traffic.
- Network interface cards and drives which efficiently filter for LAN data link layer addresses mapped from network layer IP Multicast addresses.
- IP Multicast application software, such as for video conferencing.

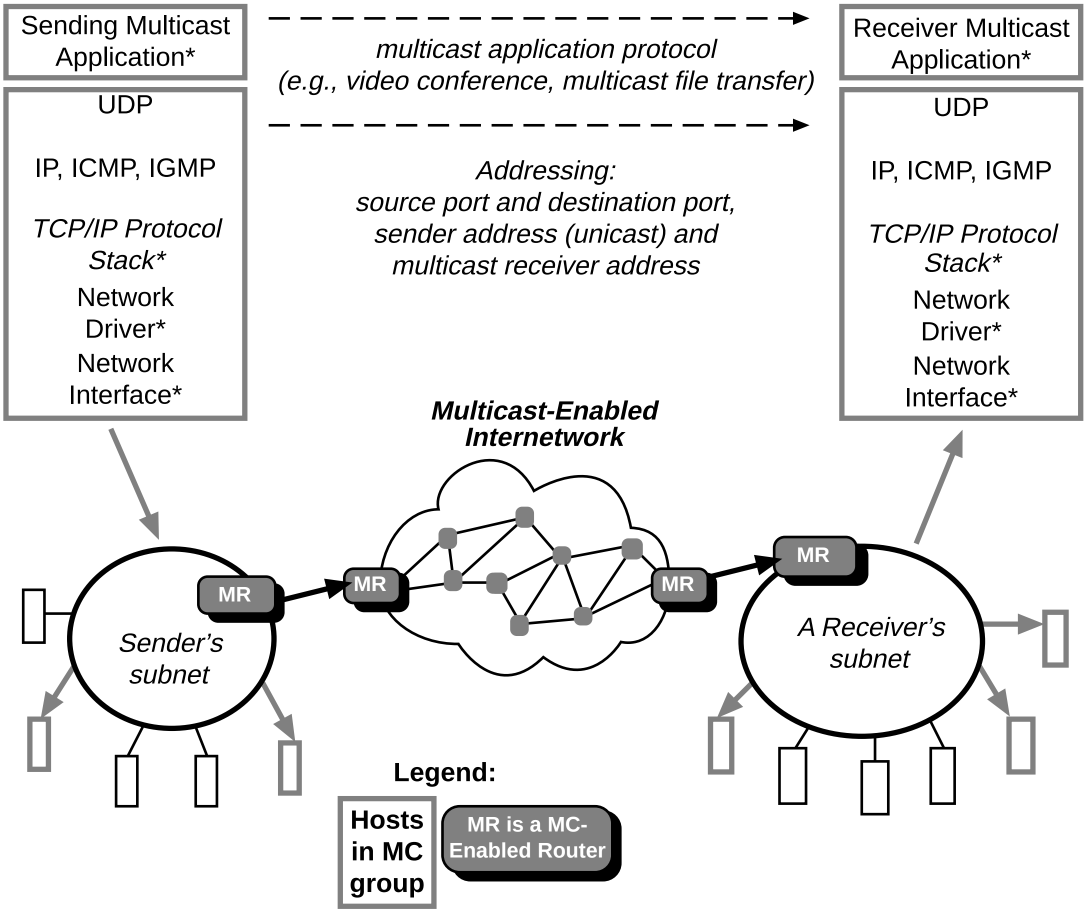
To run or evaluate IP Multicast on a LAN, only the above are needed. No routers need be involved for a host's adapter to create or join a multicast group and share multicast data with other hosts on that LAN segment. To expand native IP Multicast traffic to a WAN requires:

- All intermediate routers between the sender(s) and receiver(s) must be IP Multicast-capable. Many new routers have support for IP Multicast. Older ones may require more memory before they can be upgraded.
- Firewalls may need to be reconfigured to permit IP Multicast traffic.

IP Multicast has broad and growing industry backing, and is supported by many vendors of network infrastructure elements such as routers, switches, TCP/IP stacks, network interface cards, desktop operating systems and application software. Your vendors can help you select appropriate hardware and software.

The following diagram depicts, at a high level, components that must be multicast-enabled. The direction of traffic shown is for multicast datagrams. Traffic needed to communicate host group membership and routing information is **not** shown.

Figure 1
Multicast-enabled components



Sending Multicast Application*
UDP
IP, ICMP, IGMP
*TCP/IP Protocol Stack**
Network Driver*
Network Interface*
multicast application protocol
(e.g., video conference, multicast file transfer)
Addressing:
source port and destination port,
sender address (unicast) and
multicast receiver address
Receiver Multicast Application*
UDP
IP, ICMP, IGMP
*TCP/IP Protocol Stack**
Network Driver*
Network Interface*
Multicast-Enabled Internetwork
MR
MR
MR
MR
Sender's subnet
A Receiver's subnet
Legend:
Hosts in MC group
MR is a MC-Enabled Router

A transitional deployment technique has been developed to connect islands of multicast routers separated by links which do not support IP Multicast. With this approach, called IP tunneling, multicast datagrams are encapsulated in a standard point-to-point unicast datagram. Tunneling is used extensively in the MBONE. This paper addresses only native IP Multicast. Tunneling is discussed in the IP Multicast Initiative white paper "Introduction to IP Multicast Routing."

Multicast Filtering Switches

IP Multicast can be optimized in a LAN by using multicast filtering switches. An IP Multicast-aware switch provides the same benefits as a multicast router, but in the local area. Without one, the multicast traffic is sent to all segments on the local subnet. An IP Multicast aware switch can automatically set up multicast filters so the multicast traffic is only directed to the participating end nodes.

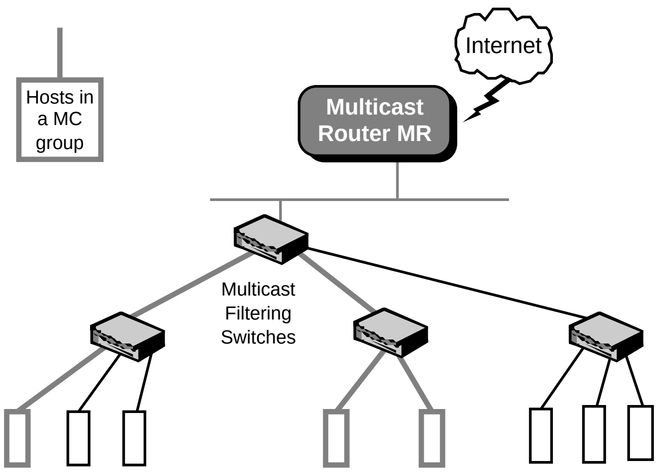


Figure 2
Multicast Filtering Switches

IP Multicast Addresses and Host Group Management

IP Multicast Addressing

IP Multicast uses Class D Internet Protocol addresses—those with 1110 as their high-order four bits—to specify multicast host groups. In Internet standard "dotted decimal" notation, host group addresses range from 224.0.0.0 to 239.255.255.255. Two types of group addresses are supported: permanent and temporary. Examples of permanent addresses, as assigned by the Internet Assigned Numbers Authority (IANA), are 224.0.0.1, the "all-hosts group" used to address all IP Multicast hosts on the directly connected network, and 224.0.0.2, which addresses all routers on a LAN. The range of addresses between 224.0.0.0 and 224.0.0.255 is reserved for routing protocols and other low-level topology discovery or maintenance protocols. Other addresses and ranges have been reserved for applications, such as 224.0.13.000 to 224.0.13.255 for Net News. These reserved IP Multicast addresses are listed in RFC 1700, "Assigned Numbers". The Session Announcement Protocol and Session Description Protocol Internet drafts describe how to create and detect MBONE session address/port assignments.

To send an IP Multicast datagram, the sender specifies an appropriate destination address, which represents a host group. IP Multicast datagrams are sent using the same "Send IP" operation used for unicast datagrams.

Compared to sending of IP Multicast datagrams, reception of IP Multicast datagrams is much more complex, particularly over a WAN. To receive datagrams, a user's host application requests membership in the multicast host group associated with a particular multicast (e.g. "I want to view today's live press conference with the President"). This membership request is communicated to the LAN router and, if necessary, on to intermediate routers between the sender and the receiver. As another consequence of its group membership request, the receiving host's network interface starts filtering for the LAN-specific hardware (data-link layer) address associated with the new multicast group address. WAN routers deliver the requested incoming multicast datagrams to the LAN router, which maps the host group address to its associated hardware address and builds the message (for example, an Ethernet frame) using this address. The receiving host's network interface card and network driver, listening for these addresses, pass the multicast messages to the TCP/IP protocol stack, which makes them available as input to the user's application, such as a video viewer.

Whereas an IP unicast address is statically bound to a single local network interface on a single IP network, an IP host group address is dynamically bound to a set of local network interfaces on a set of IP networks. An IP host group address is not bound to a set of IP unicast addresses. Multicast routers don't need to know the list of member hosts for each group - only the groups for which there is one member on the subnetwork. A multicast router attached to an Ethernet need associate only a single Ethernet multicast address with each host group having a local member.

Time To Live (TTL)

Each IP Multicast packet uses the time-to-live (TTL) field of the IP header as a scope-limiting parameter. The TTL field controls the number of hops that an IP Multicast packet is allowed to propagate. Each time a router forwards a packet, its TTL is decremented. A multicast packet whose TTL has expired (is 0) is dropped, without an error notification to the sender. This mechanism prevents messages from needless transmission to regions of the worldwide Internet that lie beyond the subnets containing the multicast group members.

A local network multicast reaches all immediately-neighboring members of the destination host group (the IP TTL is 1 by default). If a multicast datagram has a TTL greater than 1, the multicast router(s) attached to the local network take responsibility for internetwork forwarding. The datagram is forwarded to other networks that have members of the destination group. On those other member networks that are reachable within the IP time-to-live, an attached multicast router completes delivery by transmitting the datagram as a local multicast. TTL thresholds in multicast routers prevent datagrams with less than a certain TTL from traversing certain subnets. This can provide a convenient mechanism for confining multicast traffic to within campus or enterprise networks. Several standard settings for TTL are specified for the MBONE: 1 for local net, 15 for site, 63 for region and 127 for world.

Internet Group Management Protocol (IGMP)

Multicast packets from remote sources must be relayed by routers, which should only forward them on to the local network if there is a recipient for the multicast host group on the LAN. The Internet Group Management Protocol (IGMP) is used by multicast routers to learn the existence of host group members on their directly attached subnets. It does so by sending IGMP queries and having IP hosts report their host group memberships. The basic version of IGMP dates from 1988 and is now a full Internet standard. It is described in RFC 1112.

IGMP is loosely analogous to ICMP and is implemented over IP. IGMP messages are encapsulated in IP datagrams. IGMP has only two kinds of packets: Host Membership Query and Host Membership Report, with the same simple fixed format containing some control information in the first word of the payload field and a class D address in the second word:

<table>
<tr><td>Version
(bits 0-3)</td><td>Type
(bits 4-7)</td><td>Code
(bits 8-15)</td><td>Checksum
(bits 16-31)</td></tr>
<tr><td colspan="4">Multicast Group Address (Class D)</td></tr>
</table>

Other types are used by extensions to this protocol for use by routing protocols.

To determine if any hosts on a local subnet belong to a multicast group, one multicast router per subnet periodically sends a hardware (data link layer) multicast IGMP Host Membership Query to all IP end nodes on its LAN, asking them to report back on the host groups memberships of their processes. This query is sent to the all-hosts group (network address 224.0.0.1) and a TTL of 1 is used so that these queries are not propagated outside of the LAN. Each host sends back one IGMP Host Membership Report message per host group, sent to the group address, so all group members see it (thus only one member reports membership).

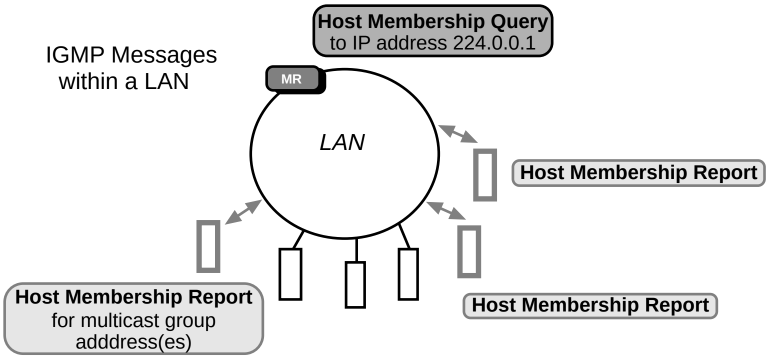


Figure 3
IGMP

When a process asks its host to join a new multicast host group, the driver creates a hardware multicast address, and an IGMP Host Membership Report with the group address is immediately sent. The host's network interface is expected to map the IP host group addresses to local network addresses as required to update its multicast reception filter. Each host keeps track of its host group memberships, and when the last process on a host leaves a group, that group is no longer reported by the host.

Periodically the local multicast router sends an IGMP Host Membership Query to the "all-hosts" group, to verify current memberships. If all member hosts reported memberships at the same time frequent traffic congestion might result. This is avoided by having each host delay their report by a random interval if it has not seen a report for the same group from another host. As a result, only one membership report is sent in response for each active group address, although many hosts may have memberships.

IGMP updates are used by multicast routing protocols to communicate host group memberships to neighboring routers, propagating group information through the internetwork. IGMP is used to identify a designated router in the LAN for this purpose. The bandwidth needed to transmit host group information is usually slight compared to the multicast application traffic, so this propagation method is workable. More sophisticated methods enable routers to determine dynamically how to best forward the multicast application traffic, as discussed in the next section.

Multicast Routing Concepts

IP Routing

The Internet is composed of a myriad of subnetworks connected by routers. When the source of a message is located on one subnet and the destination is located on a different subnet, there must be some way of determining how to get from the source to the destination. This is the function of the IP Protocol. Each host on the Internet has an address that identifies its physical location; part of the address identifies the subnet on which it resides and part identifies the particular host on that subnet. Routers periodically send routing update messages to adjacent routers, conveying the state of the network as perceived by that particular router. This data is recorded in routing tables that are then used to determine optimal transmission paths for forwarding messages across the network.

Unicast transmission involves transmission from a single source to a single destination. Thus, the transmission is directed towards a single physical location that is specified by the host address. The routing procedure, as described above, is relatively straightforward because of the binding of a single address to a single host.

Multicast Routing

Routing multicast traffic is a more complex problem. A multicast address identifies a particular transmission session, rather than a specific physical destination. An individual host is able to join an ongoing multicast session, by using IGMP to communicate this desire to its subnet router. A naive approach to sending data to multiple receivers would be for the source to maintain a table identifying all the receiving subnets participating in the session and to send a separate copy of the data to each receiving subnet. However, this would be an extremely inefficient use of bandwidth, since many of the data streams would follow the same path throughout much of the network.

New techniques have been developed to address the problem of efficiently routing multicast traffic. Since the number of receivers for a multicast session can potentially be quite large, the source should not need to know all the relevant addresses. Instead the network routers must somehow be able to translate multicast addresses into host addresses. The basic principal involved in multicast routing is that routers interact with each other to exchange information about neighboring routers. To avoid duplication of effort, a single router is selected (via IGMP) as the Designated Router for each physical network.

Spanning Trees

For efficient transmission, Designated Routers construct a *spanning tree* that connects all members of an IP Multicast group.

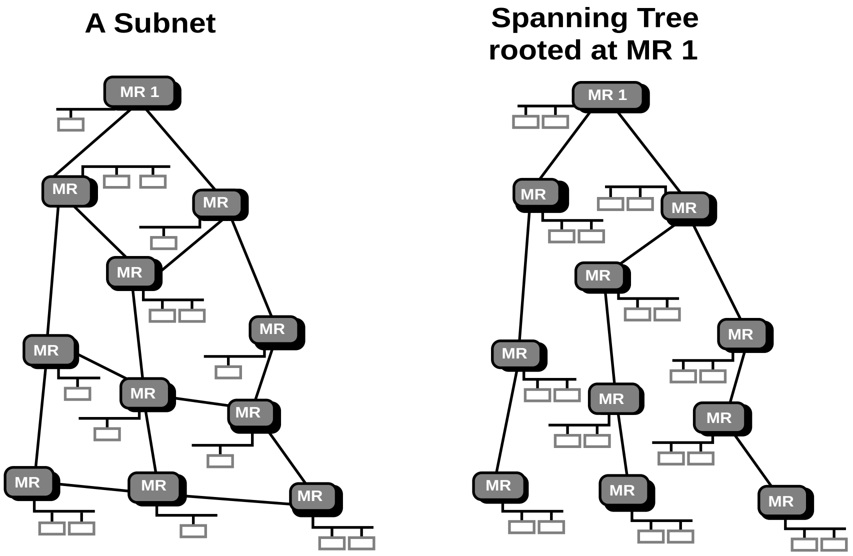


Figure 4
Spanning Trees

A spanning tree has just enough connectivity so that there is only one path between every pair of routers, and it is loop-free. If each router knows which of its lines belong to the spanning tree, it can copy an incoming multicast datagram onto all of its outgoing branches, generating only the minimum needed number of copies. Messages are replicated only when the tree branches, thus minimizing the number of copies of the messages that are transmitted through the network.

Since multicast groups are dynamic, with members joining or leaving a group at any time, the spanning tree must be dynamically updated. Branches in which no listeners exist must be discarded (pruned). A router selects a spanning tree based on the network layer source address of a multicast packet, and prunes that spanning tree based on the network layer destination address.

The spanning algorithm used and how multicast routers interact depends on the objectives of the routing protocol. Several IP Multicast routing algorithms and protocols have been designed with different objectives and features.

Two Basic Approaches to IP Multicast Routing

IP Multicast routing algorithms and protocols generally follow one of two basic approaches, depending on the distribution of multicast group members throughout the network. The first approach is based on the assumption that the multicast group members are densely distributed throughout the network and bandwidth is plentiful, i.e., almost all hosts on the network belong to the group. So-called "dense-mode" multicast routing protocols rely on periodic flooding of the network with multicast traffic to set up and maintain the spanning tree. Dense-mode routing protocols include Distance Vector Multicast Routing Protocol (DVMRP), Multicast Open Shortest Path First (MOSPF), and Protocol-Independent Multicast - Dense Mode (PIM-DM).

The second approach to multicast routing is based on the assumption that the multicast group members are sparsely distributed throughout the network and bandwidth is not necessarily widely available, for example across many regions of the Internet. It is important to note that sparse-mode does not imply that the group has a few members, just that they are widely dispersed. In this case, flooding would unnecessarily waste network bandwidth and hence could cause serious performance problems. Hence, "sparse-mode" multicast routing protocols must rely on more selective techniques to set up and maintain multicast trees. Sparse-mode routing protocols include Core Based Trees (CBT) and Protocol-Independent Multicast - Sparse Mode (PIM-SM). See the IP Multicast Initiative white paper "Introduction to IP Multicast Routing" for more information.

Other Protocols which use IP Multicast

There are a number of exciting protocols presently being developed by the Internet community, IETF working groups and industry vendors to support new applications of IP Multicast. Only a brief introduction is possible here. RTP, the Real-Time Transport Protocol, provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network services. RSVP, the ReSerVation Protocol, enhances the current Internet architecture with support requests for a specific quality of service (QoS) from the network for particular data streams or flows. RTSP, the Real-Time Streaming Protocol is an application-level protocol for control over the delivery of data with real-time properties to enable controlled, on-demand delivery of real-time data, such as audio and video. Reliable multicast protocols are being developed to overcome the limitations of unreliable multicast datagram delivery and expand the uses of IP Multicast.

Conclusion

IP Multicast enables many new types of applications and reduces network congestion and server loads. IP Multicast products and services are receiving widespread industry attention because of their potential benefits. Advances are being made in areas such as reliable multicasting, real-time applications support, and network management and diagnosis. This paper has introduced the technical concepts and mechanisms of IP Multicast. To learn more, the following references are recommended. Many were used in the preparation of this document. The authors are gratefully acknowledged. We also invite you to review the IP Multicast Initiative white paper series at www.ipmulticast.com.

More information

IP Multicast Initiative (IPMI)

There are many more aspects of IP Multicast that were not discussed here. The IP Multicast Initiative web site at www.ipmulticast.com has a technical resource center which can provides more background in-depth information including white papers and relevant RFC's. The web site also offers a product and services directory and lists members of the IP Multicast Initiative who can be contacted for information and assistance.

The IP Multicast Initiative provides marketing and educational services to promote the creation, use and deployment of multicast products and solutions. Supported by a growing number of the most important vendors in the IP Multicast arena, the Initiative and its services are managed and provided by Stardust Technologies, Inc.

For more information about the Initiative and membership, contact Stardust Technologies at 408-879-8080 or visit the Initiative web site.

Useful References

Books

Christian Huitema, ***Routing in the Internet***, Prentice Hall, 1995

Vinay Kumar, ***MBONE: Interactive Media on the Internet***, New Riders, 1996

Craig Partridge, ***Gigabit Networking***, Addison-Wesley, 1994

Radia Perlman, ***Interconnections***, Addison-Wesley, 1992

Bob Quinn and Dave Shute, ***Windows Sockets Network Programming***, Addison-Wesley, 1996

Andrew Tanenbaum, ***Computer Networks***, Prentice Hall, 1996

IETF RFCs

[http://ds.internic.net/rfc/rfcnnnn.txt, nnnn is the RFC number]

RFC 1112 Host Extensions for IP Multicasting

RFC 1700 Assigned Numbers

IETF Internet Drafts

[ftp://ietf.org/internet-drafts/name-of-file]

Session Description Protocol draft-ietf-mmusic-sap-00.txt, .ps

Session Announcement Protocol draft-ietf-mmusic-sdp-02.txt, .ps