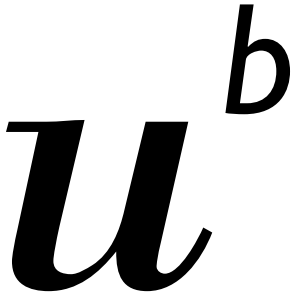


b

UNIVERSITÄT
BERN

UNIVERSITÄT BERN

Advanced Networking and Future Internet

VII. Multicast

Prof. Dr. Torsten Braun, Institut für Informatik

Bern, 26.10.2020

Advanced Networking and Future Internet: Multicast

Table of Contents

1. Introduction

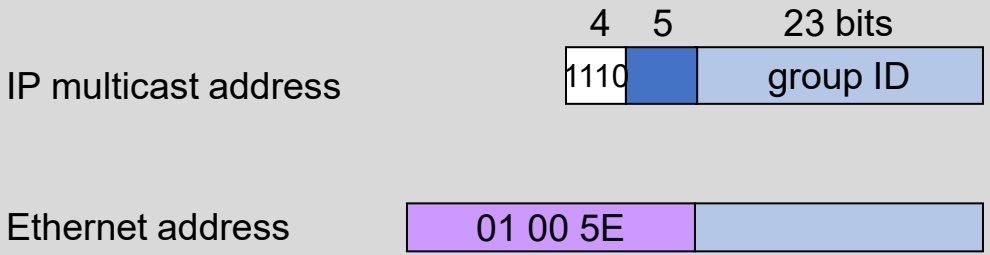
1. Multicast Application

- Audio/Video Conferencing
- Computer Supported Cooperative Work
- Push technologies (software and information distribution)
- Parallel computing
- Games
- TV
- …

UNIVERSITÄT BERN

1. Introduction

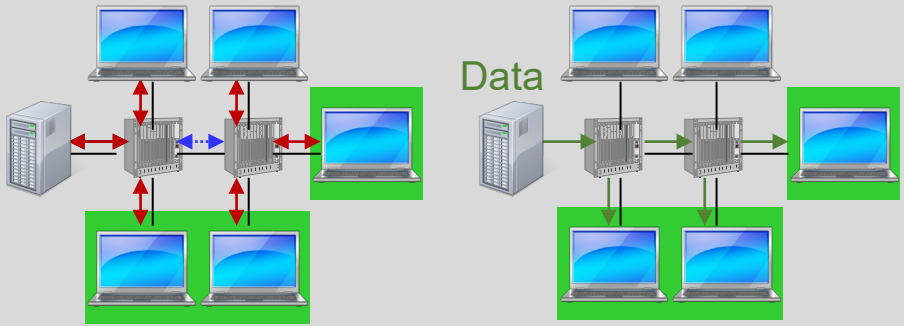
2. IPv4 / Ethernet Multicast Addressing



1. Introduction

3. Multicast Model

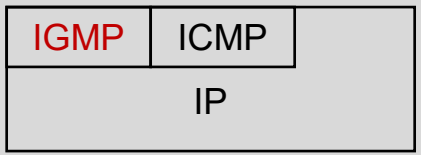
- Receiver group is identified by IP multicast address.
- Sender sends a multicast packet to IP multicast address.
- Packet is sent along a multicast distribution treeto be established using Internet Group Management Protocol and multicast routing protocols.
- Anonymous receivers
- Each end system can join a multicast group.
- Transport protocols: UDP is multicast capable, but TCP is not.
  ⇒ Datagram oriented multicast communication

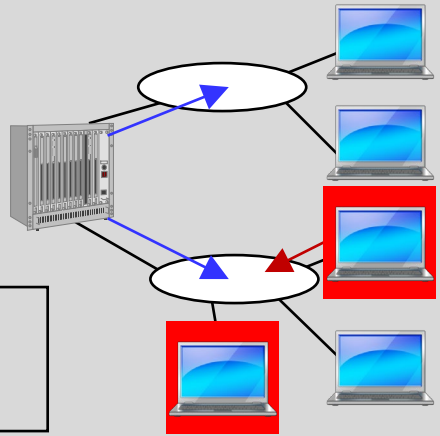


2. Internet Group Management Protocol

1. IGMP Version 1

- Routers need to know about group memberships.
- Messages
  - Membership query
    - Periodic, e.g., 1/min, by router (querier) to „all hosts“ group
    - 1 querier per physical network (router with lowest IP address)
  - Membership report
    - Response of a host to indicate group membership
    - should be sent immediately after joining a group
- Service primitives
  - **`JoinHostGroup (address, interface)`**
    ⇒ transmission of reports
  - **`LeaveHostGroup (address, interface)`**
    ⇒ no transmission of reports
  - Implementation via setsocketopt





IGMP Membership Query
IGMP Membership Response

2. Internet Group Management Protocol

2. IGMP Versions 2 and 3

IGMPv2

- End system sets flag, if it has answered to the last seen Membership Query.
- A leaving end system sends Leave Group message, if flag is set.
- Router sends then a group-specific Membership Query allowing the router to detect quickly, whether there are further group members.

IGMPv3

- Source filtering
  - allows to request multicast packets from one or more senders
- New primitive
  - `IPMulticastListen (socket, interface, multicast-address, filter-mode, source-list)`

  replaces JoinHostGroup/LeaveHostGroup.
- Group-(and-source)-specific Membership Query messages
  - Report messages might contain selected sources.

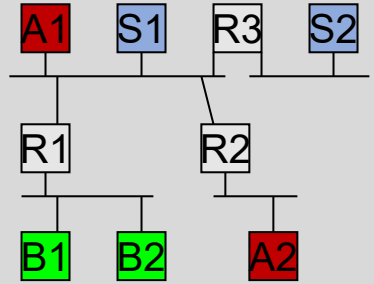
3. Intra-Domain Multicast Routing

- Link state routing
  - Routers have information about each link and the complete topology of a domain.
  - Dijkstra‘s Shortest Path algorithm for route calculation
  - Example:
    Multicast Open Shortest Path First
- Distance Vector routing
  - Flood and Prune
  - Routers know for each tuple [sender, group] whether and over which interface multicast data must be forwarded.
  - Examples:
    - Distance Vector Multicast Routing Protocol
    - Protocol Independent Multicast – Dense Mode: based on underlying unicast routing protocol, does not require exchange of multicast routing information
- Core-Based Trees
  - no source-specific trees
  - Example: PIM - Sparse Mode (SM)

3. Intra-Domain Multicast Routing

1. Multicast OSPF

- Multicast Open Shortest Path First
- Group membership link state advertisements:
  [group, attached_network]
- Routers extract information from IGMP.
- Multicast data distribution via Dijkstra's Shortest Path algorithm



R: router
S: sender
A: member of group A
B: member of group B

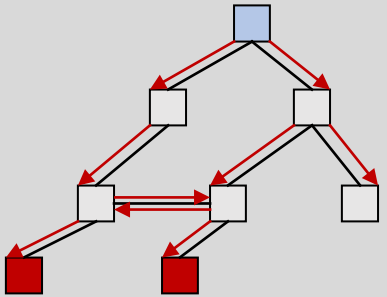
S2→B: R3 forwards packet
S1→B: R3 does not forward packet

S1→B: R1 forwards packet
S1→A: R1 does not forward packet

3. Intra-Domain Multicast Routing

2.1.1 Protocol Independent Multicast - Dense Mode

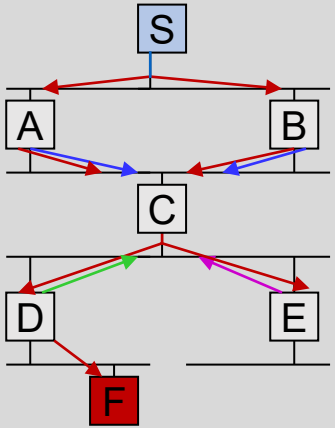
- Protocol operation when a router receives multicast packets from S to G
  - If input interface = unicast output interface for S:
    - Forward packet via all interfaces (except incoming interface)
  - Otherwise:
    - Prune (S,G) message to input interface
- Protocol messages
  - Prune (S,G)
    - is also sent if no group members exist on a leaf link
  - Graft (S,G)
    - Invalidation of pruning (also periodically)
    - Acknowledgement by Graft Ack
  - Join (S,G)
    - Invalidation of Prune (S,G) by neighbor router on the same link.
  - Assert (S,G)
    - Router indicates distance to sender (required for collision detection)



3. Intra-Domain Multicast Routing

2.1.2 PIM-DM Protocol Operation

- Routers A and B receive multicast packet at output interface and send Assert indicating distance to sender to reserved multicast address. Subsequent packets will only be forwarded from A to C, but not from B to C.
- Prune, Join: C forwards data packets to D, but E does not forward.



3. Intra-Domain Multicast Routing

2.2.1 Core-Based Trees

PIM-DM, DVMRP and MOSPF

- Calculation of multicast distribution trees for each tuple [sender, multicast group]
- Advantage
  - Optimized path from source to each receiver
- Disadvantages
  - Increased resource usage
  - High number of routing entries

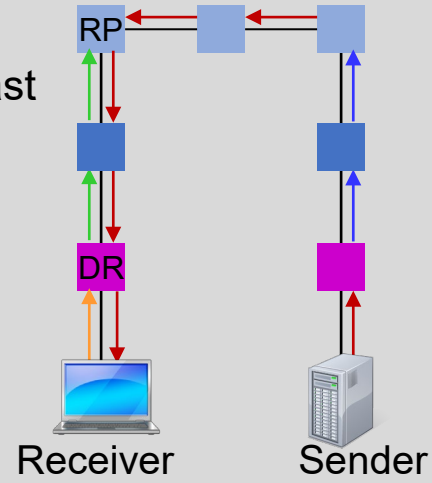
Core-Based Trees, e.g., PIM-SM

- All senders use a single tree for a multicast group.
- Advantages
  - Less routing table entries
  - Lower network overhead
- Disadvantage
  - Paths between source and receiver may not be optimal.

3. Intra-Domain Multicast Routing

2.2.2 PIM-Sparse Mode: Protocol Operation

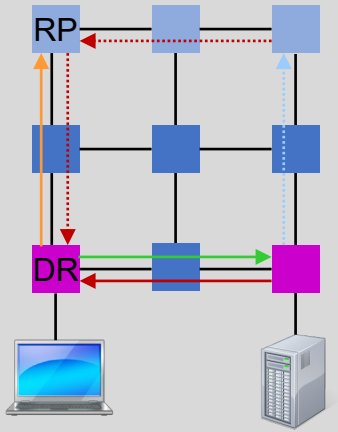
- PIM-SM is based on core-based trees.
- Rendezvous Points (RP) for each multicast group form core of multicast distribution tree.
- Designated Routers (DR) connect group members to RPs.
- Receiver joins multicast group.
  - IGMP Membership Report
  - Periodic PIM-Join/Prune from DR to RP
- Data transmission to multicast group
  - DR encapsulates data in PIM-Register and forwards it to RP.
  - RP decapsulates data and distributes it along multicast tree.
  - Routers forward data via interfaces, from which PIM-Join/Prune has been received.



3. Intra-Domain Multicast Routing

2.2.3 PIM-SM: Source-Specific Tree

- DR with group members or RP can initiate a source-specific tree by sending a source-specific Join to source, e.g., in case of high data rate
- PIM router at source forwards data directly to DR or RP.
- Prune message from DR to RP



3. Intra-Domain Multicast Routing

2.2.4 PIM Source-Specific Multicast (PIM-SSM)

- Support of one-to-many model, e.g., 1 speaker, TV
- Channel = combination of source address S and group address G
- Receiver specifies source and group in IGMP Join.
- Router sends PIM-SM source specific Join towards source.
- Establishment of source-specific tree in routers
- Only source can send to the source-specific tree.

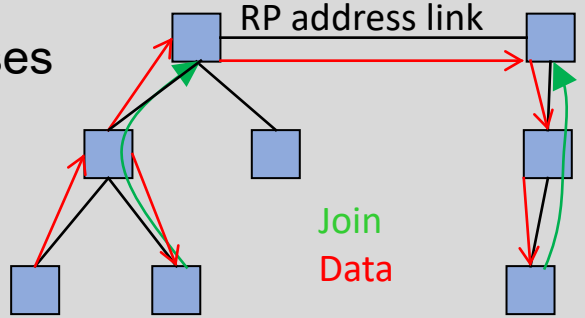
Advantages

- More direct connections to receivers
- Lower risk of misuse by malicious senders
- Reuse of multicast address in various domains without conflicts if sources of own domains
- Usage across domains possible

3. Intra-Domain Multicast Routing

2.2.5 BIDIRectional-PIM

- Variant of PIM-SM for many-to-many multicasting, if senders and receivers are the same, e.g., A/V conferencing
- BIDIR-PIM has not only branches to receivers, but also to sources.
- Advantages
  - No need for RPs, but only (routable) RP addresses
  - No source registration process
  - Bidirectional trees use less state (no source-specific states)



4. Inter-Domain Multicast Routing

Multicast Source Discovery Protocol

- Mechanism to interconnect PIM-SM domains

Border Gateway Multicast Protocol

- Core-based tree protocol

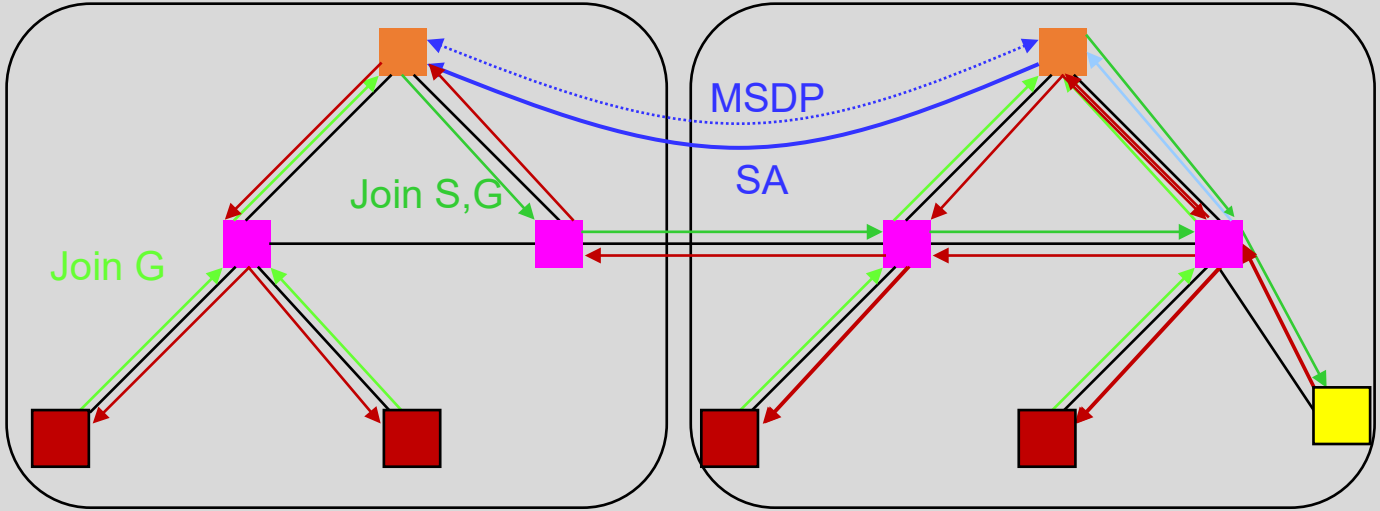
4. Inter-Domain Multicast Routing

1.1 Multicast Source Discovery Protocol

- Goal: Robust interconnection of domains without central RP in a single domain
- Establishment of TCP connections among MSDP capable RPs in different domains
- Exchange of Source-Active (SA) messages for active multicast sources
  - Source address
  - Destination multicast address
  - RP address
- For a received SA message the RP transmits a source-specific Join towards the source, if there are group members in its own domain.
- Intermediate solution, but does not scale, because each domain must be notified about new sources. Data need first to be encapsulated in SA messages.

4. Inter-Domain Multicast Routing

1.2 MSDP Example



4. Inter-Domain Multicast Routing

2.1 Border Gateway Multicast Protocol

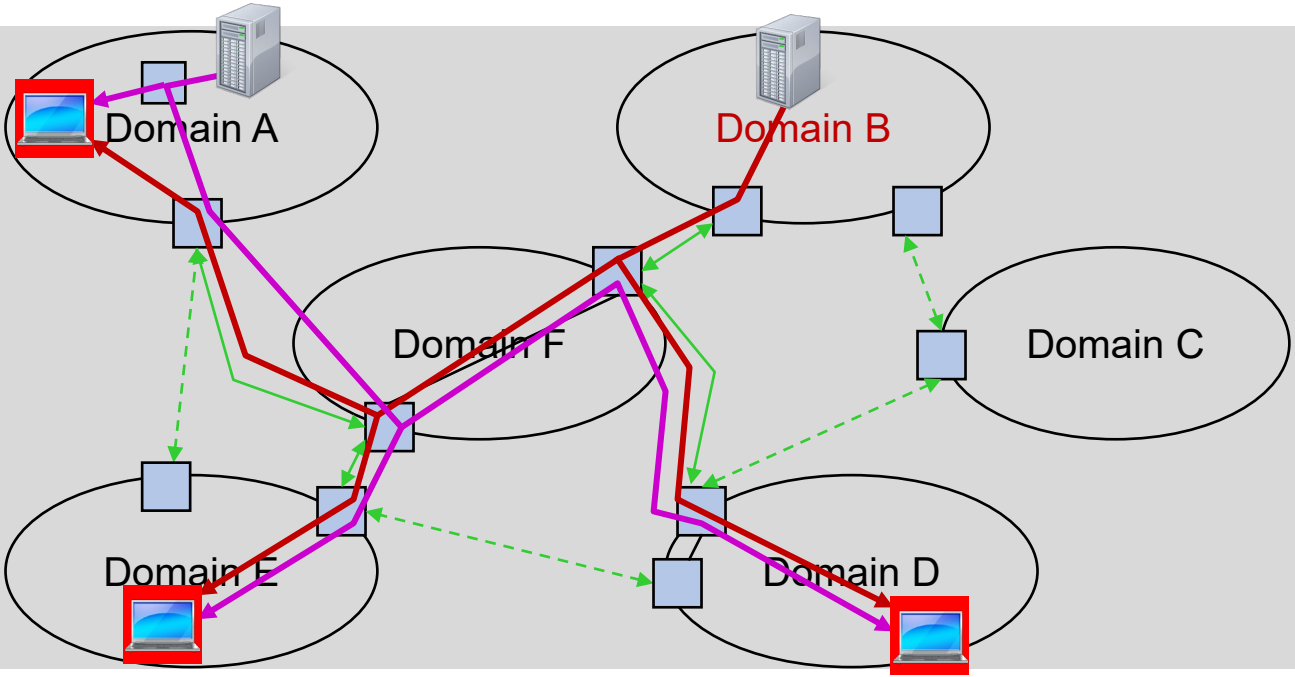
- Shared trees for active multicast groups
- Uni-directional and bi-directional trees
- Routing information exchange over TCP connections between border routers
- Shared trees are rooted at an autonomous system (domain) that allocated a multicast address, e.g., using MALLOC architecture
- Information about address assignments is distributed by Multiprotocol BGP.

Protocol Operation

- Receiver joins multicast group.
- Border router sends Join towards root domain using intra-domain routing protocol.
- Creation of group-specific forwarding entries in border routers

4. Inter-Domain Multicast Routing

2.2 BGMP Example



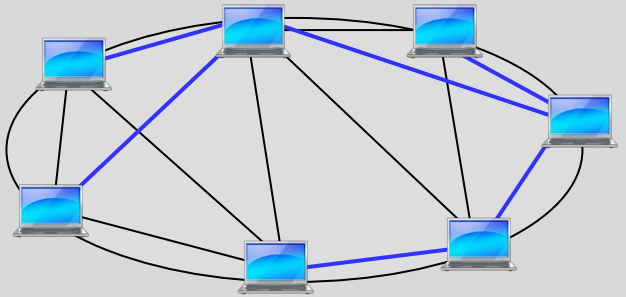
5. IP Multicast Problems

- Security
  - Any receiver can join a multicast group and receive traffic.
    → encryption
  - Any sender can send traffic to global multicast addresses:
    risk of denial-of-service attacks
    → distribution by a single source
- Scalability
  - Multicast routers have routing entries of the following form:
    "[source, multicast address]
    → output interfaces"
- Deployment and management overhead
  - Multicast requires support of both unicast and multicast routing protocols.
  - Multicast address management
- Reliable multicast transport is still an open issue.
- Billing is difficult to achieve.
- Multicast makes only sense if (bandwidth savings > management costs)

6. Application Layer Multicast

- Implementation of
  - Group Management
  - Packet replication

  in end systems
- Self-organizing group of end systems
- Protocol components
  - Group management
  - Construction of distribution tree from knowledge about neighbors

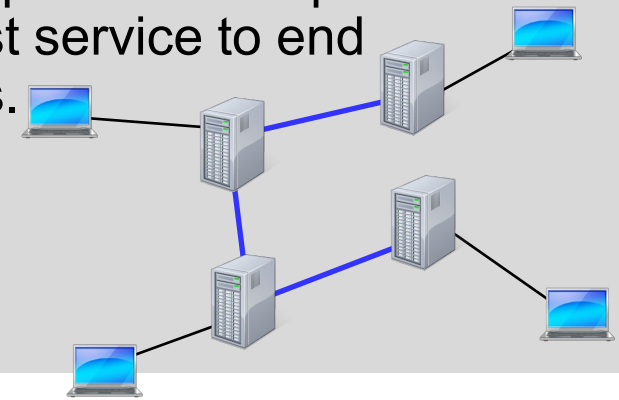


6. Application Layer Multicast

1. Deployment

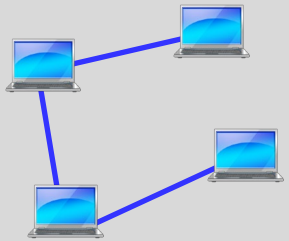
Infrastructure-level multicast

- Multicast overlay network between proxies
- Proxies provide transparent multicast service to end systems.



End system multicast

- Overlay established between end systems using unicast network service
- Option: IP multicast in backbone and end system multicast in local network



6. Application Layer Multicast

2. Group Management

Problem

– Users must find, join, and leave multicast sessions.

Solutions

– Rendezvous point

– P2P mechanisms

– Flooding

6. Application Layer Multicast

3. Construction of Distribution Trees

Mesh First

- Members keep connected mesh topology.
- Source is chosen as root.
- Routing to root for building the tree
- Tree formation depends on mesh
- Examples:
  - Scribe
  - Narada
  - CAN
  - NICE

Tree First

- Tree building without mesh
- Members select parents from known members in tree.
- Direct control over tree
- Examples:
  - ALMI
  - ESM for Internet Broadcast

6. Application Layer Multicast

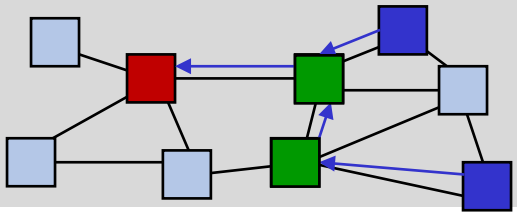
4. Routing Mechanisms

- Shortest Path
  - Construct minimum diameter spanning tree, e.g., using RTTs between source and end systems
  - Example: Narada
- Minimum Spanning Tree
  - Construct spanning tree with lowest costs, not necessarily minimizing diameter of tree
  - Example: ALMI
- Clustering
  - Hierarchical cluster of nodes
  - Example: Nice
- Peer-to-Peer
  - Reverse or forward path forwarding
  - Example: Scribe

6. Application Layer Multicast

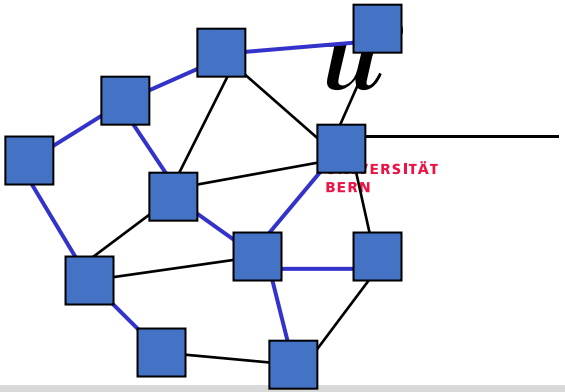
5.1 Scribe

- Scribe is based on Pastry (Plaxton routing)
- Each group has a unique group ID (topic ID).
- The node with the ID closest to the group ID is the rendezvous point of the group (= root of multicast tree).
- Creation of a group: send create message to rendezvous point
- Nodes that are part of the multicast tree are called forwarders.
  They may or may not be group members.
  Forwarders maintain children table.
- Joining nodes send join message towards RP
  - Nodes not already being forwarders become forwarders and add child to children table for that group.
  - Join message is terminated by an already active forwarder.
- SplitStream runs on top of Scribe, establishes several (node-disjoint) trees for a group and stripes the content over these trees.
  → robustness, bandwidth



6. Application Layer Multicast

5.2 End System Multicast (Narada)

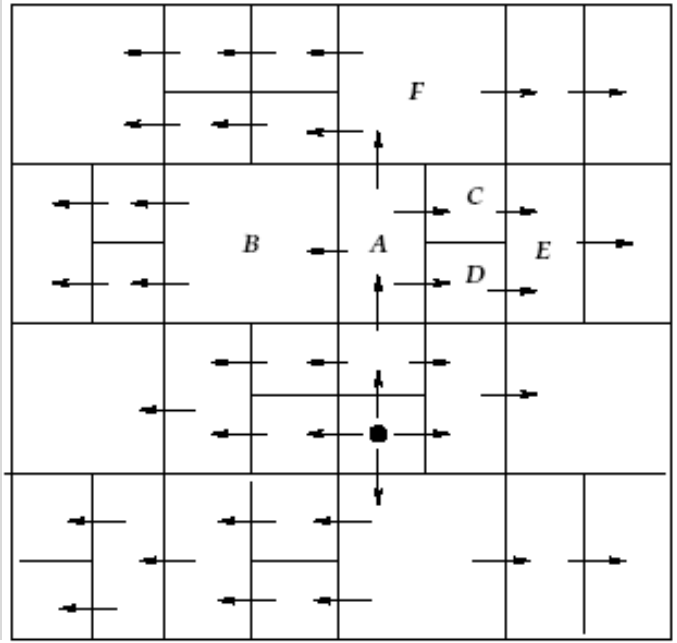


- Self-organizing overlay network
- Every member maintains a list of all other members.
  - Updates for joins and leaves !
  - Exchange of refresh messages between neighbor nodes
- Mesh establishment between nodes
- Distance vector routing algorithm is running on top of the mesh.
- Incremental improvement of mesh quality by adding and dropping overlay links
- Mean delay (simulation):
  - 2-3 times the delay achieved with DVMRP
  - Factor is increasing with group size !
- More sophisticated (shortest widest path) algorithm can significantly improve delay, sometimes even better than DVMRP.

6. Application Layer Multicast

5.3 Application Layer Multicast with CAN

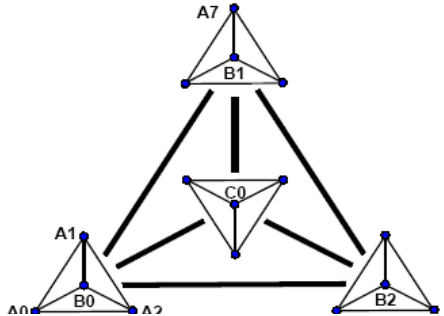
- given: CAN C with subset of nodes wishing to form multicast group G
- Creation of mini CAN C_G, made up of only members of G
- Mapping of group address of G to point (x, y) in CAN C.
- The node owning (x, y) serves as bootstrap node for G.
- Joining group G is reduced to joining C_G.
- Multicast forwarding: directed flooding

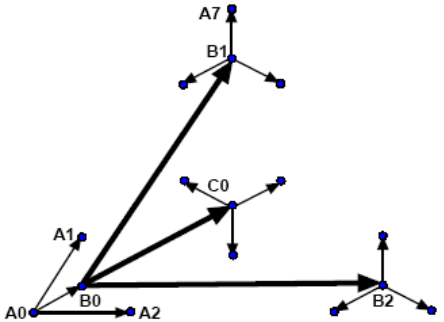


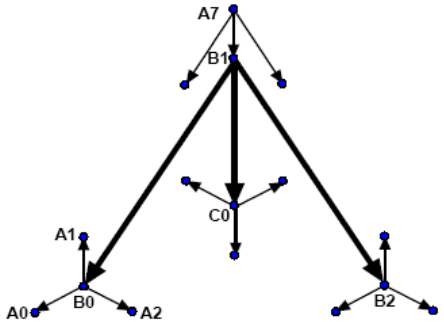
6. Application Layer Multicast

5.4 NICE is the Internet Cooperative Environment

- Nodes are partitioned into clusters with cluster leader = node with minimal distance to cluster members
- k ≤ cluster size ≤ 3k-1
- Cluster leaders form cluster on higher level. → hierarchy
- Cluster leader on level i is member of level i+1.
- Different overlay structures for control (cliques) and data messages (trees for sources A0 and A7)

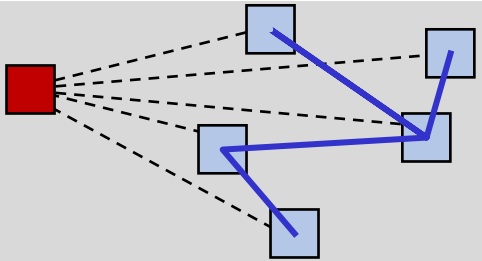






6. Application Layer Multicast

5.5 Application Level Multicast Infrastructure



- for small groups
- centralized group management
- Session controller
  - handles registration
  - maintains a multicast tree using a control protocol with group members
  - maintains point-to-point connections with each peer
- Performance monitoring by members, reporting to session controller
- Backup session controller

6. Application Layer Multicast

5.6 ESM for Internet Broadcast

- Revised design of Narada, based on tree-first approach
- Single-source video broadcast application
- Group Management
  - Joining host contacts source and gets a partial list of member nodes including nodes between source and joining node as well as some random nodes.
  - Parent selection algorithm
  - Members learn about others by gossip protocol: Each member periodically (e.g., every 2s) picks a member and sends a subgroup of known members. Member list entries time out (e.g., 5 min).
  - Leaving nodes continue forwarding for 5s to allow children looking for new parents.
  - Monitoring of loss, bandwidth, delay; new parent selection if observed performance < 90 % of source rate

Parent selection

- Probing of random subset of nodes
- Returned information: observed bandwidth, delay, degree saturation, descendant (→ loop avoidance)

Thanks

for Your Attention

Prof. Dr. Torsten Braun, Institut für Informatik

Bern, 26.10.2020