Multicast Inter-domain Mini Workshop @Networkshop 2006

Stig Venaas (UNINETT)
stig.venaas@uninett.no
Tim Chown (Univ. Southampton)
tjc@ecs.soton.ac.uk

Interdomain multicast

- Introduction
- Multicast addressing
- MSDP and MBGP
- SSM
- Multicast in JANET

Interdomain Multicast

- We will look at how multicast domains can be interconnected to get multicast connectivity throughout the Internet
- Each domain typically uses their own RP for all groups
- A PIM domain might be defined as a set of routers that use the same RP set
  - This definition is not accurate
- A multicast domain might be a PIM domain as defined above
  - I feel it's better to think of it as a multicast network inside an administrative domain, e.g. a university or a back-bone network
- When we talk about interdomain multicast, it's how to do multicast between such domains

IPv4 Multicast Addressing

- IPv4 multicast addresses: 224.0.0.0 – 239.255.255.255 (224/4, class D)
- These are subdivided in rather complicated ways,
  - see http://www.iana.org/assignments/multicast-addresses/ for details
- Examples:
  - 224.0.0.0 – 224.0.0.255 (224.0.0/24) – Local network control block, never forwarded
    - 224.0.0.1 - All local hosts
    - 224.0.0.2 - All local routers
    - 224.0.0.5 - OSPF
    - 224.0.0.13 - PIM
    - 224.0.0.22 - IGMP
  - 224.0.1.0 – 224.0.1.255 (224.0.1/24) – Internetwork control block, forwarded
  - 224.2/16 – SAP
  - 232/8 – SSM (only to be used for Source Specific Multicast)
  - 233/8 – GLOP
  - 234.0.0.0 – 238.255.255.255 – Reserved
  - 239/8 – Administrative scoping

Address Assignment, SAP and GLOP

- Knowing what addresses to use when creating a session seems rather complicated
- SAP (Session Announcement Protocol, RFC 2974)
  - Announces a session
  - SAP applications also help you pick what addresses to use
  - Uses dynamic groups in range 224.2.128.0 – 224.2.255.255 for global sessions
  - Global announcements sent to 224.2.127.254
  - sdr is the most common SAP application, but not used so much these days
- GLOP (not an acronym)
  - Assignment based on AS numbers, RFC 3180
  - 233.x.x/24 where x.x is an officially assigned AS number
  - For private AS space there is EGLOP (RFC 3138)
    managed by registries, e.g. RIPE (still 233.x.x/24, but with private AS numbers)

Administrative Scoping – 239/8

- Addresses in the range 239/8 are used for administrative scoping
  - Private address space, not to be used globally
  - Different networks can use the same addresses
- 239.255/16 is the smallest administrative scope
  - Sometimes used for site-local
- 239.192/14 is organization-local scope
  - These addresses should work throughout JANET
  - All but 239.194/16 are restricted to JANET
    - 239.192/16, 239.193/16 and 239.195/16 used for sessions visible throughout JANET, but not outside
  - 239.194/16 is used for GÉANT
    - i.e. sessions using these groups are available throughout GÉANT (European academic networks), but not outside
- Multicast distribution can be restricted by specifying a small TTL value for packets
  - Limited use. With routing protocols like PIM-SM and MSDP, packets may travel very far even if TTL is small

Interdomain Multicast Routing

- Many networks don't support multicast, so multicast often needs to be routed differently from unicast
  - BGP is used for setting up peerings to route unicast between networks
  - If you use BGP, you may need MBGP (Multiprotocol BGP)
  - If you use only static unicast routes, you are fine with just static multicast routes
- Each domain typically uses their own RP for all groups
  - Each RP only knows about sources in its own domain
  - To get connectivity between domains, MSDP is used so that RPs learn about sources in other domains
  - This is not needed for SSM to work between domains (no RPs)
- You may wish to configure boundaries to restrict some multicast groups to be local
  - Filtering what can flow across the domain's boundary

Multiprotocol BGP (MBGP)

- Originally BGP supported just IPv4 unicast
- Multiprotocol BGP (RFC 2858) may have AFI (Address Family Identifier) IPv4/IPv6 and for those, SAFI (Subsequent AFI)
  - SAFI = 1 for unicast, 2 for multicast, 3 for both
  - Seems IETF wants to deprecate 3
- For a peering one may configure IPv4/IPv6 unicast/multicast separately with different policies
- Multicast routes used for RPF
  - Sometimes in addition to unicast routes
  - One may also sometimes translate unicast routes into multicast
- Recommend only using multiprotocol BGP
  - Whenever unicast BGP between multicast networks exists; enable multicast peering if multicast connectivity is desired
  - Try to avoid tricks like translation or unicast routes for RPF
- Note that you only need to worry about this if you have BGP peerings to networks you want to have multicast connectivity with
- We sometimes talk about multicast BGP, meaning multiprotocol BGP for exchanging multicast routes

IPv4 Multiprotocol BGP on IOS

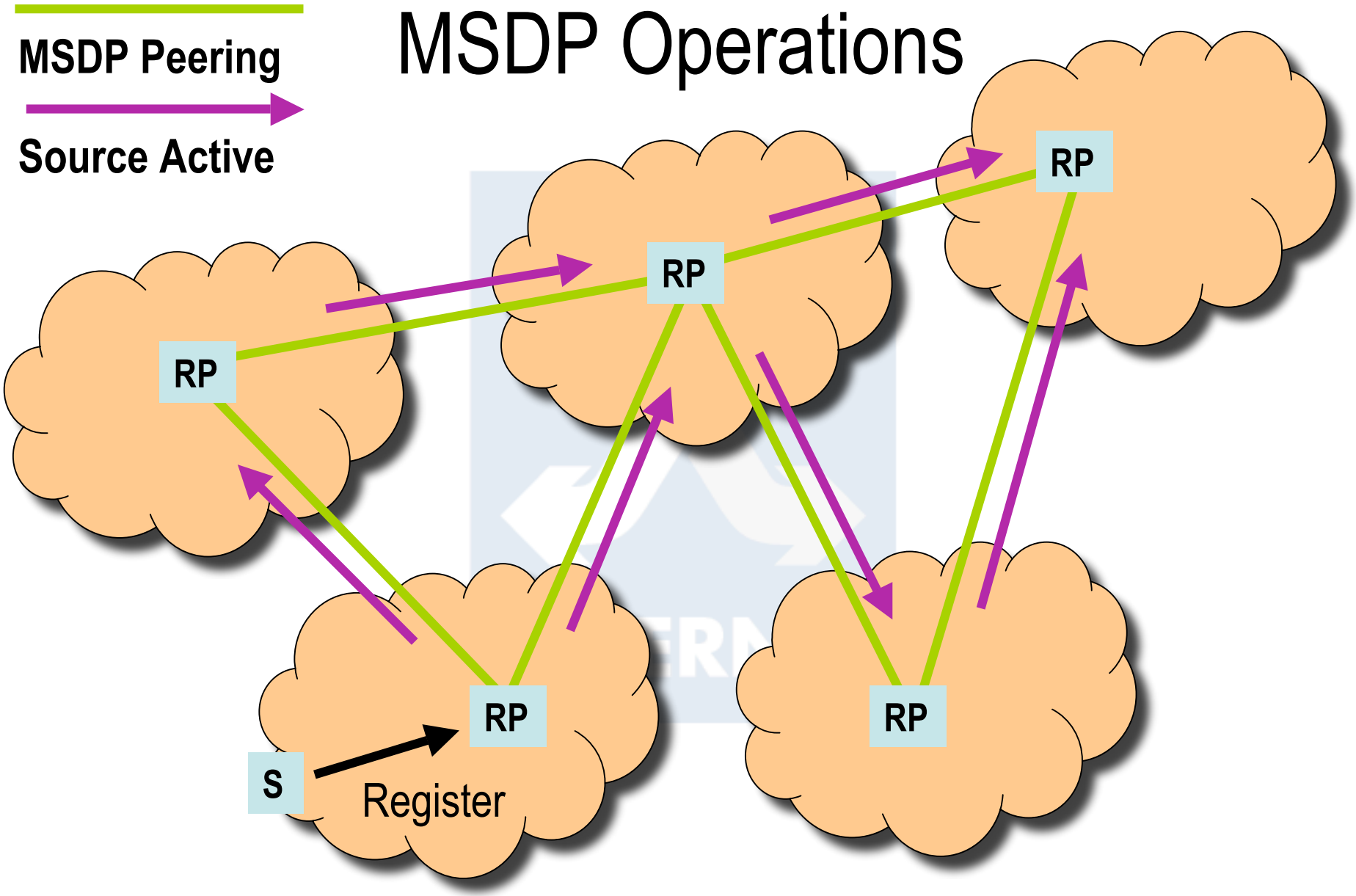
```
router bgp 224
 no bgp default ipv4-unicast
 neighbor 192.0.2.1 remote-as 64001
 neighbor 192.0.2.129 remote-as 64002
 !
 address-family ipv4
 …
 !
 address-family ipv4 multicast
 neighbor 192.0.2.1 activate
 neighbor 192.0.2.65 activate
 network 192.0.2.128 mask 255.255.255.128
 exit-address-family
```


- Basically you configure multiprotocol BGP like you do for unicast
- If the policies are the same you can, for each multicast peer, copy from that peer's unicast config

MSDP – Multicast Source Discovery Protocol

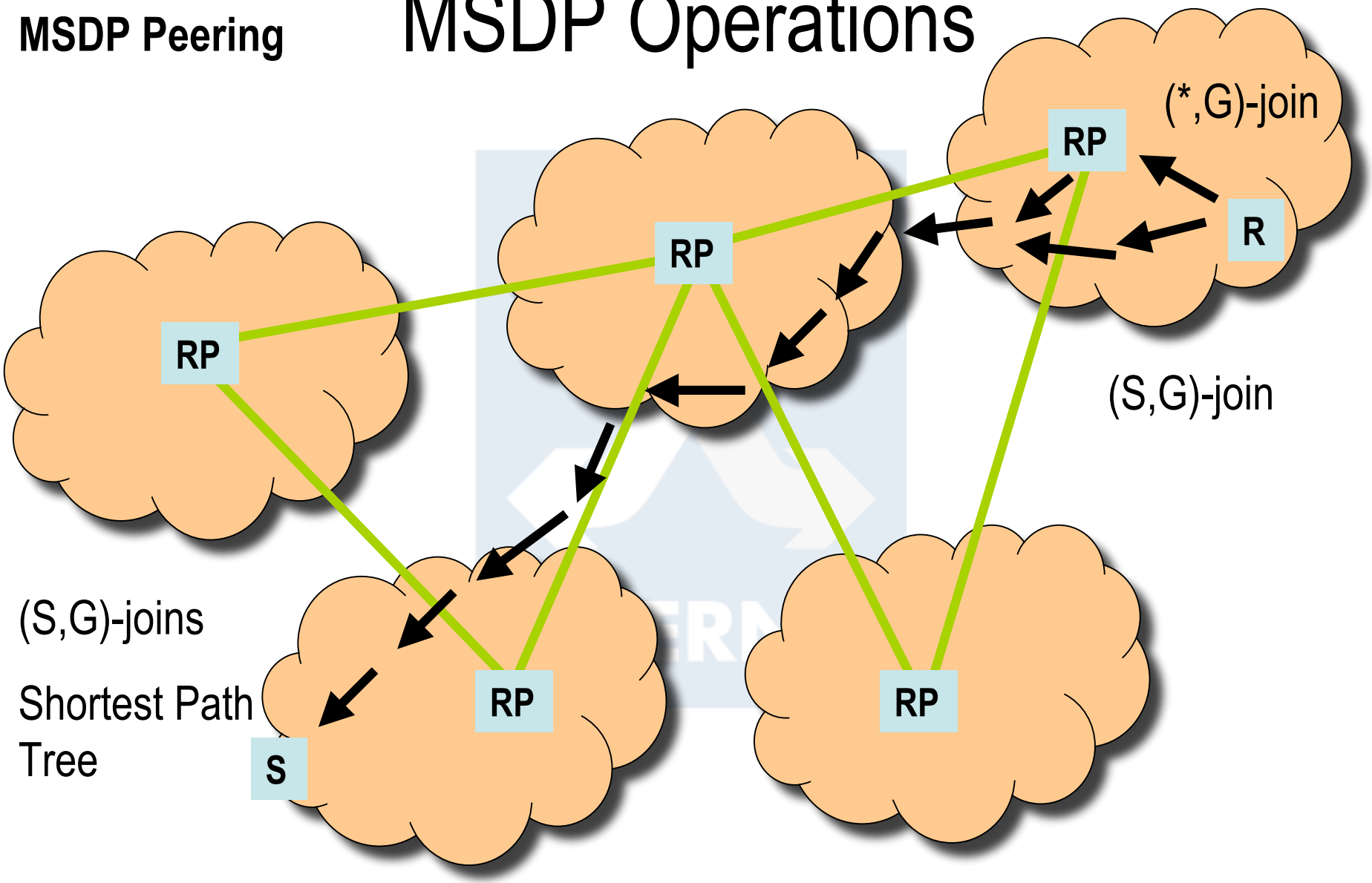
- MSDP (RFC 3618) sets up a mesh of MSDP peerings between domains
  - Usually between MBGP pairs using the same addresses for the end-points
  - Internal MSDP peerings can be set up if the RP is not where the external peerings are terminated
  - Internal peerings also allow multiple RPs for the same group
  - Used for anycast-RP
- When an RP learns a new local source:
  - Sends a Source Active (SA) advertisement
  - SAs are flooded through the mesh of MSDP peerings to other RPs
  - SAs may optionally contain data packet
  - SAs are cached
  - RPF is used to prevent loops
- If an RP knows of sources within other domains, it can construct Shortest Path Trees (SPTs) to them when a local host joins the group

MSDP Operations



MSDP Operations

MSDP Peering



MSDP configuration on IOS

```
ip msdp peer 192.0.2.1 connect-source Loopback0
ip msdp sa-filter in 192.0.2.1 list 21
ip msdp sa-filter out 192.0.2.1 list 21
ip msdp peer 192.0.2.65 connect-source Loopback0
ip msdp cache-sa-state
ip msdp originator-id Loopback0
ip msdp mesh-group example 192.0.2.65
ip msdp mesh-group example 192.0.2.66
!
access-list 21 deny 239.255.0.0 0.0.255.255
access-list 21 deny 229.55.150.208
access-list 21 deny 224.0.1.0 0.0.0.255
access-list 21 permit any
```


- Here we have one external MSDP peer
- We also do anycast RP between 3 RPs, consisting of this router plus 2 others in mesh-group example

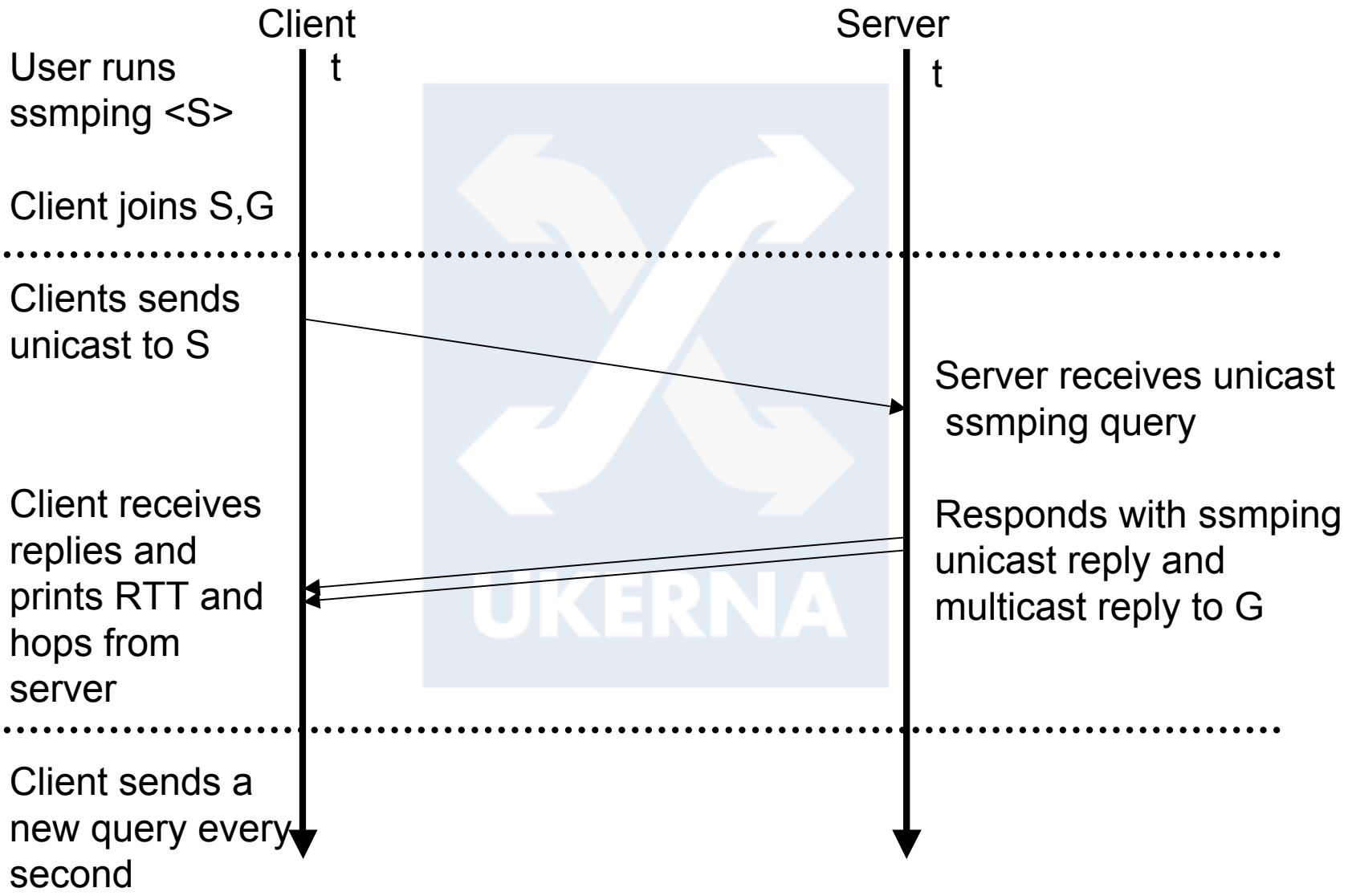
SSM – Source Specific Multicast

- SSM is a new multicast service model
  - Receivers specify the source address(es) in addition to the group
- This avoids rogue sources sending to the group
  - Imagine watching one video stream and someone else also sends their video, or just random data at high rates
- The main benefit is hugely simplified routing
  - PIM-SM is still used, but no RPs involved
  - No MSDP needed
  - You still need multicast BGP (if you use unicast BGP)
- SSM requires source discovery at the application layer
  - Very easy for streaming with one fixed source (maybe most important use of multicast)
  - May be difficult for some multi-party or discovery applications

ssmping

- A tool for testing multicast connectivity
- Behavior is a bit like normal ping
- A server must run ssmpingd
- A client can ping a server by sending unicast ssmping query
- Server replies with both unicast and multicast ssmping replies
- In this way a client can check that it receives SSM from the server
  - also parameters like delay, number of router hops etc.
- JANET is running a server at ssmping.beacon.ja.net
- There is a similar tool called asmping for checking ASM connectivity
- See http://www.venaas.no/multicast/ssmping/

How ssmping works



Example IPv4 ssmping output (v6 supported)

```
$ ssmping -4 -c 5 ssmping.beacon.ja.net
ssmping joined (S,G) = (193.60.199.162,232.43.211.234)
pinging S from 158.38.63.22
  unicast from 193.60.199.162, seq=1 dist=16 time=39.331 ms
  unicast from 193.60.199.162, seq=2 dist=16 time=39.394 ms
multicast from 193.60.199.162, seq=2 dist=16 time=43.905 ms
  unicast from 193.60.199.162, seq=3 dist=16 time=39.542 ms
multicast from 193.60.199.162, seq=3 dist=16 time=39.547 ms
  unicast from 193.60.199.162, seq=4 dist=16 time=39.137 ms
multicast from 193.60.199.162, seq=4 dist=16 time=39.142 ms
  unicast from 193.60.199.162, seq=5 dist=16 time=39.535 ms
multicast from 193.60.199.162, seq=5 dist=16 time=39.539 ms

--- 193.60.199.162 ssmping statistics ---
5 packets transmitted, time 5000 ms
unicast:
   5 packets received, 0% packet loss
   rtt min/avg/max/std-dev = 39.137/39.387/39.542/0.292 ms
multicast:
   4 packets received, 0% packet loss since first mc packet (seq 2) recvd
   rtt min/avg/max/std-dev = 39.142/40.533/43.905/1.958 ms
$
```

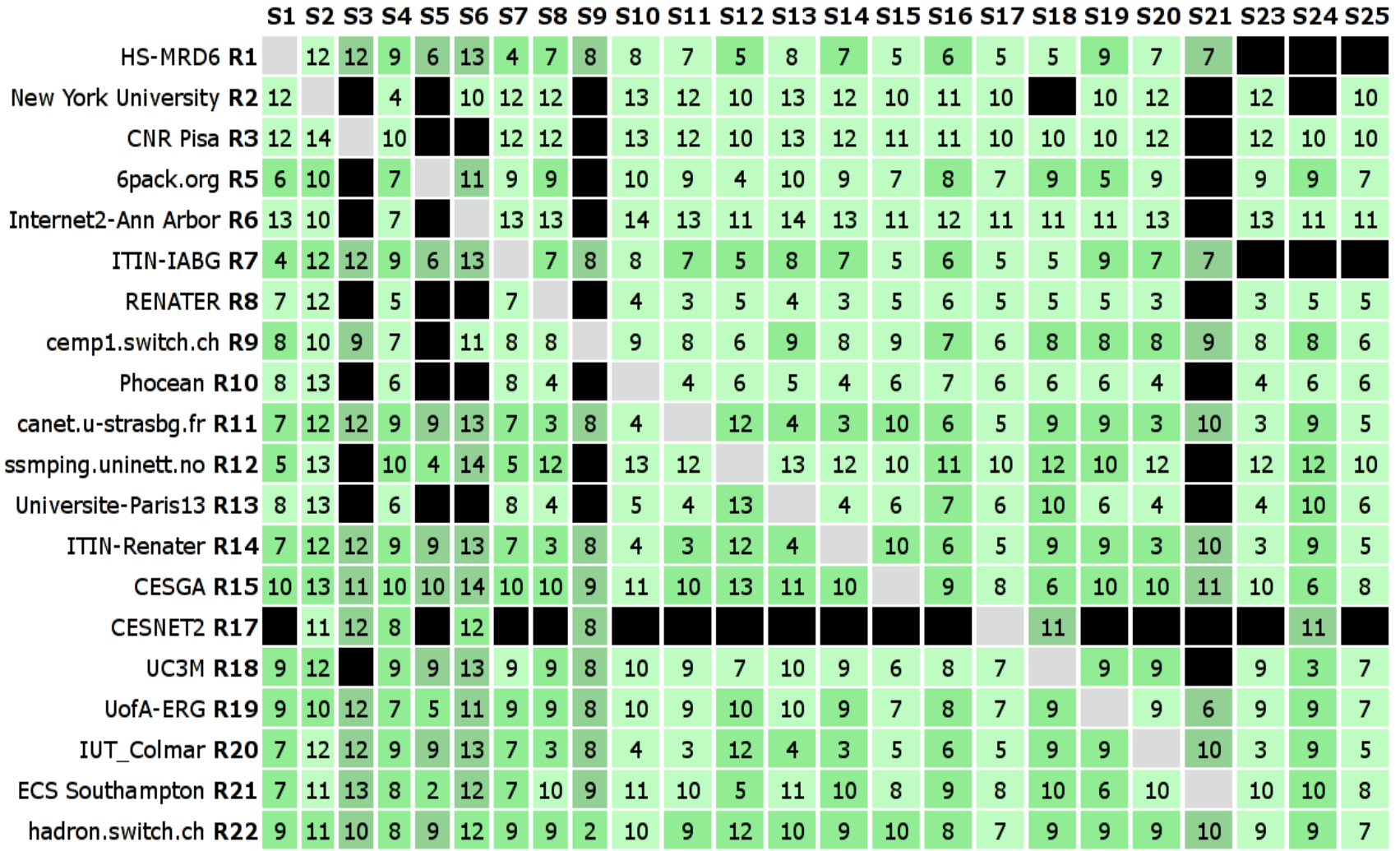
What does ssmping output tell us?

- 16 unicast hops from source, also 16 for multicast, might indicate that unicast and multicast follow the same path
- Multicast RTTs are about same for unicast and multicast
  - However, the delay for the first multicast packet is large, would need to send more queries for a proper test
  - Note the difference in unicast and multicast RTT shows one way difference for unicast and multicast replies, since they are replies to the same request packet
- Multicast tree not ready for first multicast reply, ok for 2nd so the tree was in place after about one second when the second packet was sent
- No unicast loss, no multicast loss after tree established

dbeacon

- dbeacon is a new multicast beacon
  - http://dbeacon.innerghost.net/
  - Alternative to NLANR beacon
  - IPv4 and IPv6, ASM and SSM
  - Written in C, light and easy to install
  - No central server, ASM used for signalling
  - Any beacon client can be configured to provide a matrix
  - Beacon options, e.g.
    - dbeacon –b ff7e:a30:2001:db8:10::beac –S –a admin@email
  - With Apache:
    - ScriptAlias /matrix/ /usr/share/dbeacon-matrix/matrix.pl
    - Edit script for path to matrix $dumpfile = '/var/lib/dbeacon/dump.xml';

Example dbeacon Matrix



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S23	S24	S25
HS-MRD6 **R1**		12	12	9	6	13	4	7	8	8	7	5	8	7	5	6	5	5	9	7	7			
New York University **R2**	12			4		10	12	12		13	12	10	13	12	10	11	10		10	12		12		10
CNR Pisa **R3**	12	14		10			12	12		13	12	10	13	12	11	11	10	10	10	12		12	10	10
6pack.org **R5**	6	10		7		11	9	9		10	9	4	10	9	7	8	7	9	5	9		9	9	7
Internet2-Ann Arbor **R6**	13	10		7			13	13		14	13	11	14	13	11	12	11	11	11	13		13	11	11
ITIN-IABG **R7**	4	12	12	9	6	13		7	8	8	7	5	8	7	5	6	5	5	9	7	7			
RENATER **R8**	7	12		5			7			4	3	5	4	3	5	6	5	5	5	3		3	5	5
cemp1.switch.ch **R9**	8	10	9	7		11	8	8		9	8	6	9	8	9	7	6	8	8	8	9	8	8	6
Phocean **R10**	8	13		6			8	4			4	6	5	4	6	7	6	6	6	4		4	6	6
canet.u-strasbg.fr **R11**	7	12	12	9	9	13	7	3	8	4		12	4	3	10	6	5	9	9	3	10	3	9	5
ssmping.uninett.no **R12**	5	13		10	4	14	5	12		13	12		13	12	10	11	10	12	10	12		12	12	10
Universite-Paris13 **R13**	8	13		6			8	4		5	4	13		4	6	7	6	10	6	4		4	10	6
ITIN-Renater **R14**	7	12	12	9	9	13	7	3	8	4	3	12	4		10	6	5	9	9	3	10	3	9	5
CESGA **R15**	10	13	11	10	10	14	10	10	9	11	10	13	11	10		9	8	6	10	10	11	10	6	8
CESNET2 **R17**		11	12	8		12			8									11					11	
UC3M **R18**	9	12		9	9	13	9	9	8	10	9	7	10	9	6	8	7		9	9		9	3	7
UofA-ERG **R19**	9	10	12	7	5	11	9	9	8	10	9	10	10	9	7	8	7	9		9	6	9	9	7
IUT_Colmar **R20**	7	12	12	9	9	13	7	3	8	4	3	12	4	3	5	6	5	9	9		10	3	9	5
ECS Southampton **R21**	7	11	13	8	2	12	7	10	9	11	10	5	11	10	8	9	8	10	6	10		10	10	8
hadron.switch.ch **R22**	9	11	10	8	9	12	9	9	2	10	9	12	10	9	10	8	7	9	9	9	10	9	9	7

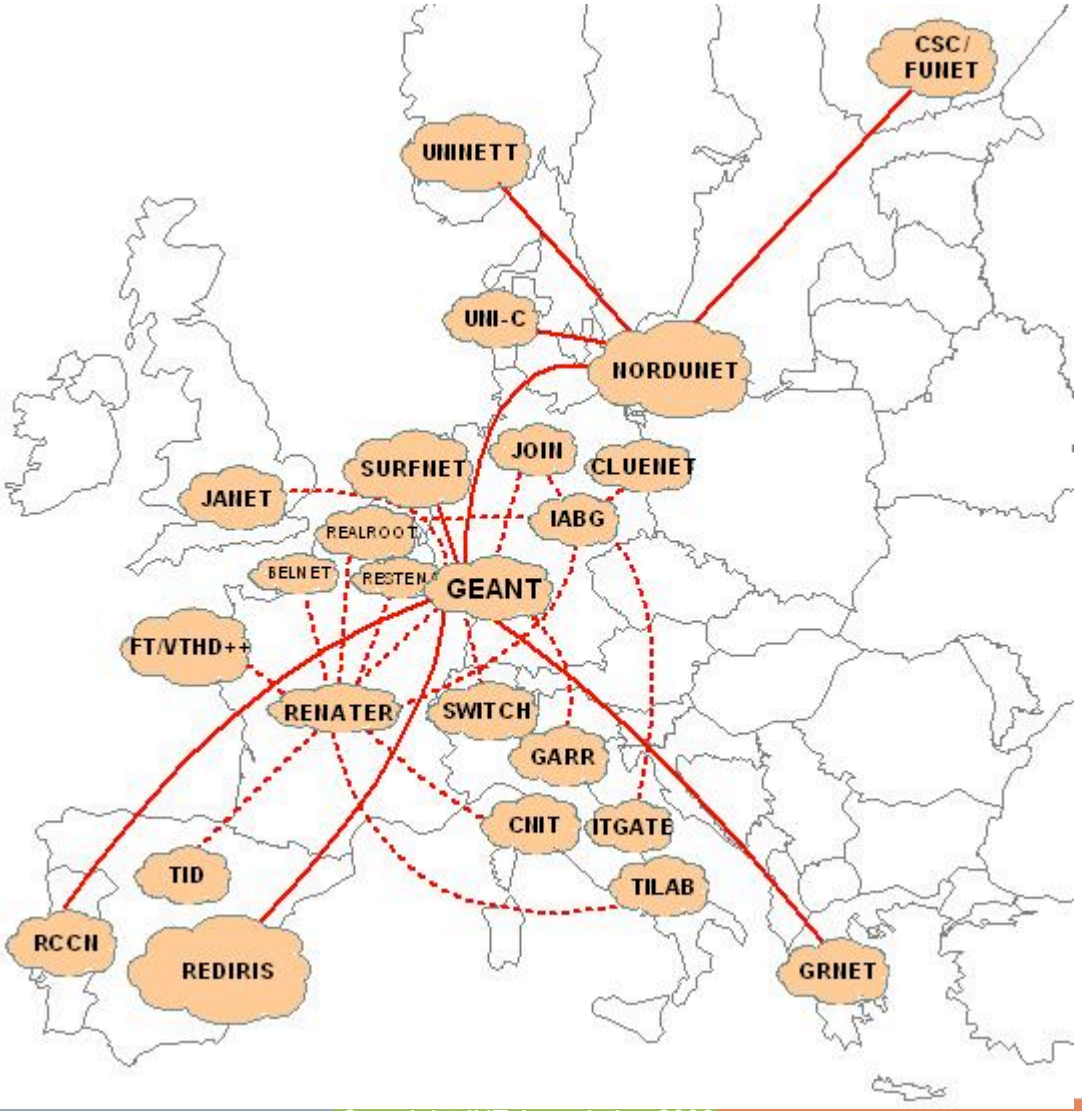
Multicast Applications

- Mbone tools, vic/rat etc
  - IPv6 multicast conferencing applications
  - http://www-mice.cs.ucl.ac.uk/multimedia/software/
- AccessGrid
  - Uses vic and rat for high quality room-based conferencing
  - http://www.accessgrid.org
- VideoLAN (vlc)
  - Video streaming, also IPv6 multicast. Server and client
  - http://www.videolan.org/
- DVTS
  - Streaming DV over RTP over IPv4/IPv6
  - DV devices using Firewire can be connected to two different machines and you can stream video between them over the Internet
  - http://www.sfc.wide.ad.jp/DVTS/
- Mad flute
  - Streaming of files using multicast (IPv4/IPv6 ASM/SSM)
  - http://www.atm.tut.fi/mad/

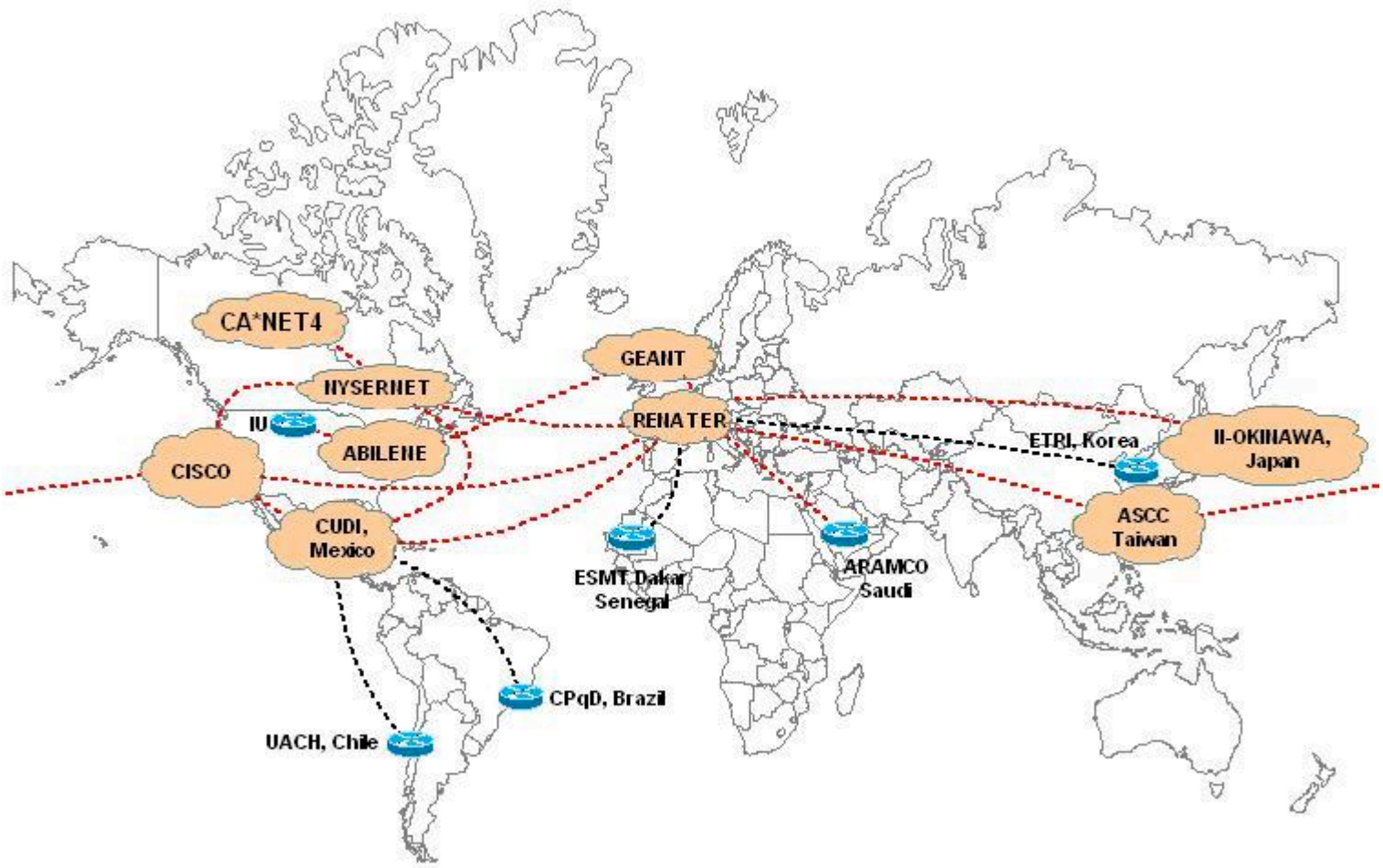
IPv6 Multicast Deployment

- There are also interdomain IPv6 multicast deployments
- A tunneled multicast network called M6Bone was established in France in 2001
  - Today more than 50 sites from four continents are connected
- Networks like Abilene, GEANT and NORDUnet have native IPv6 multicast in their production networks and native peerings
- JANET may provide IPv6 tunnels for experimenting with IPv6 multicast
- The M6Bone has both a web site and a mailing list for information and discussion on IPv6 multicast
  - Web site at http://www.m6bone.net/
  - Mailing list: m6bone@ml.renater.fr
- The largest router vendors support IPv6 multicast on most platforms. Many other vendors have some support, or plan to offer support soon

M6Bone – Europe



M6Bone – The World



Interdomain IPv6 Multicast

- For IPv4, each site typically has their own RP for all global groups. RPs in different sites use MSDP to learn of remote sources
- This avoids relying on a 3rd party to host some central RP
- MSDP does not scale, and there is no MSDP for IPv6
- The lack of MSDP means that for a given global group there can be only one single RP on the Internet
- IPv6 has something called embedded-RP, where the RP address can be encoded into the multicast group address
  - This makes it easy for everyone to have their own RP
  - Everyone else on the internet will use their RP for that group
  - E.g. the group ff7e:140:2001:700:f000:100:1234:beac uses the RP 2001:700:f000:100::1

Networkshop 34 - Multicast Workshop
Status of Multicast on JANET

Duncan Rogerson
<d.rogerson@ukerna.ac.uk>

April, 2006

Status of IP Multicast on JANET

- No changes
- Architecture based on
  - PIM Sparse Mode
  - MSDP and MBGP in the backbone
  - MSDP between backbone and Regional Networks for distribution of active source information
  - MBGP between backbone and Regional Networks for populating the multicast forwarding table
  - Direct backbone connects – PIM SP
  - Regional Network to site – models may vary

Multicast Monitoring

- Still problematic
  - Esp troubleshooting tools
  - Incidents still usually not reported until after the event
- New version of multicast beacon deployed
  - Retains global, Regional and AccessGrid views

SuperJANET 4 to SuperJANET 5

- SuperJANET 5 Backbone Routers (SBR)
  - Will not be configured as RPs
  - Router hardware specification does not support RP functionality
- SuperJANET 4 Backbone Access Routers (BAR)
  - Were configured as RPs
  - Use as RPs was never recommended or supported

SuperJANET 4 to SuperJANET 5

- Dual links to Regional Networks
  - More complex routing configuration
  - However, no more so than tuning the unicast routing config for dual links
- IPv6 multicast?
  - Most likely implemented at start
  - Mature JunOS and IOS code available