Converting the MBone to Hierarchical Multicast

presented by Van Jacobson

(based on work of Deborah Estrin,
Steve Deering & Dino Farinacci)

31th IETF
San Jose, CA
5 Dec 1994

The MBone currently has 1500 nets. Routing is done by a single-level protocol very similar to RIP.

We tried something like this 10 years ago with < 200 nets on the original 56K Fuzzball NSFNet.

A 200 site transcontinental RIP net was a nightmare to keep running. It was very fragile and required a lot of careful, manual planning.

A 1500 site intercontinental RIP net is much worse.

Our highest priority is to move the MBone to a more managable, hierarchical, routing architecture.

- Confines problems.
- Makes interoperation of different protocols easier (N protocols have only $N + 1$ interactions instead of N^2)
- Allows orderly evolution (e.g., pruning deployment).

Motherhood (cont.)

- Allows experimentation.
- Reduces routing state and traffic burden.
- Allows better control of inter-domain traffic flow.

Constraints:

- Have to get something done soon.
- Has to be deployed incrementally.
- Has to work with many different multicast routing protocols.
- We're going to evolve *both* levels of the architecture, independently, over time.

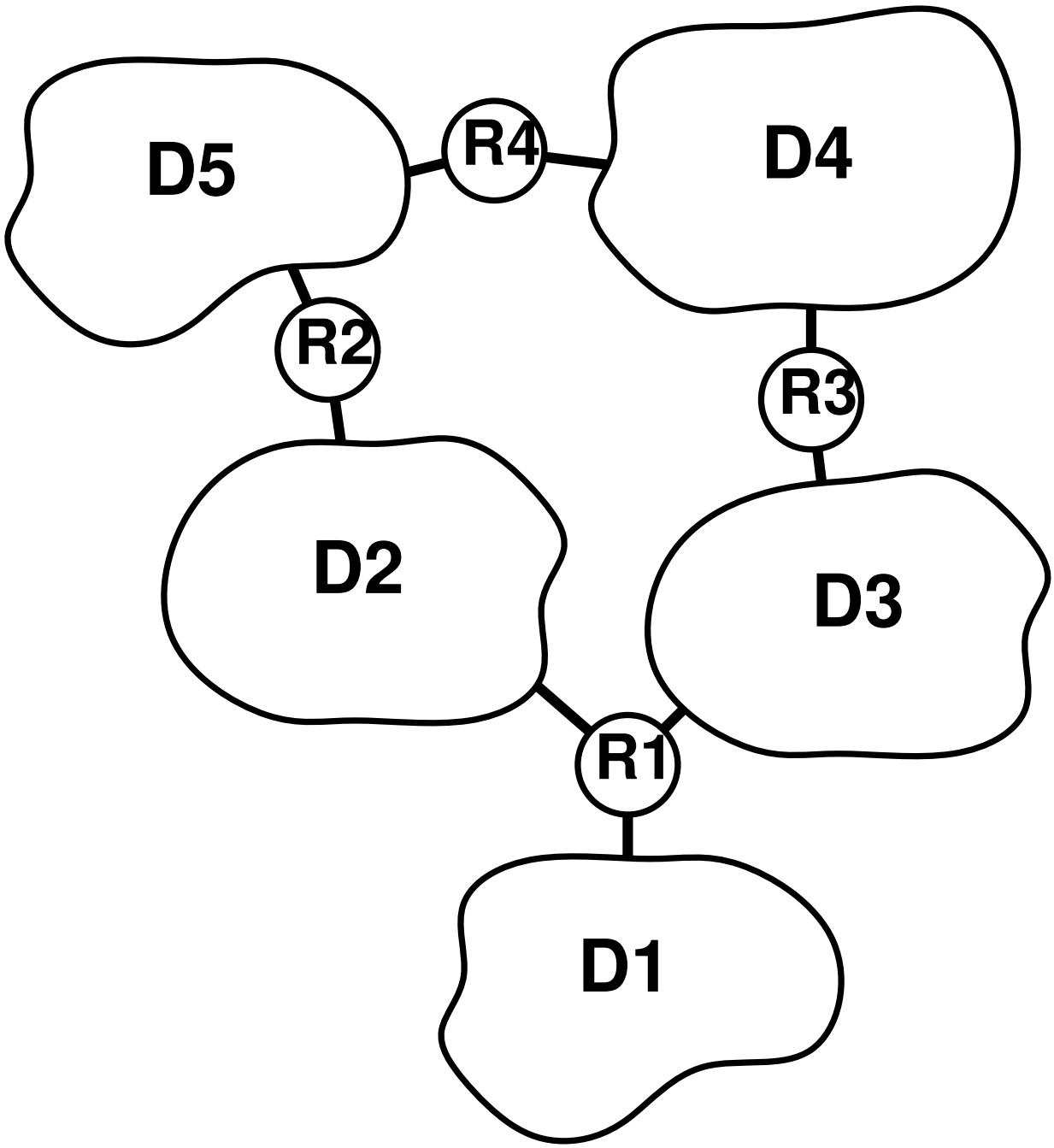
Plan:

- Draw domain boundaries at ‘natural’ places (i.e., at existing unicast intra-/inter-domain boundaries.

- Level-1 (intra-domain) continues to run whatever it’s running now.

- **No** level-1 peers across domain boundaries (boundaries are ‘hard’).

  (Hard to enforce this but can do loop detection).

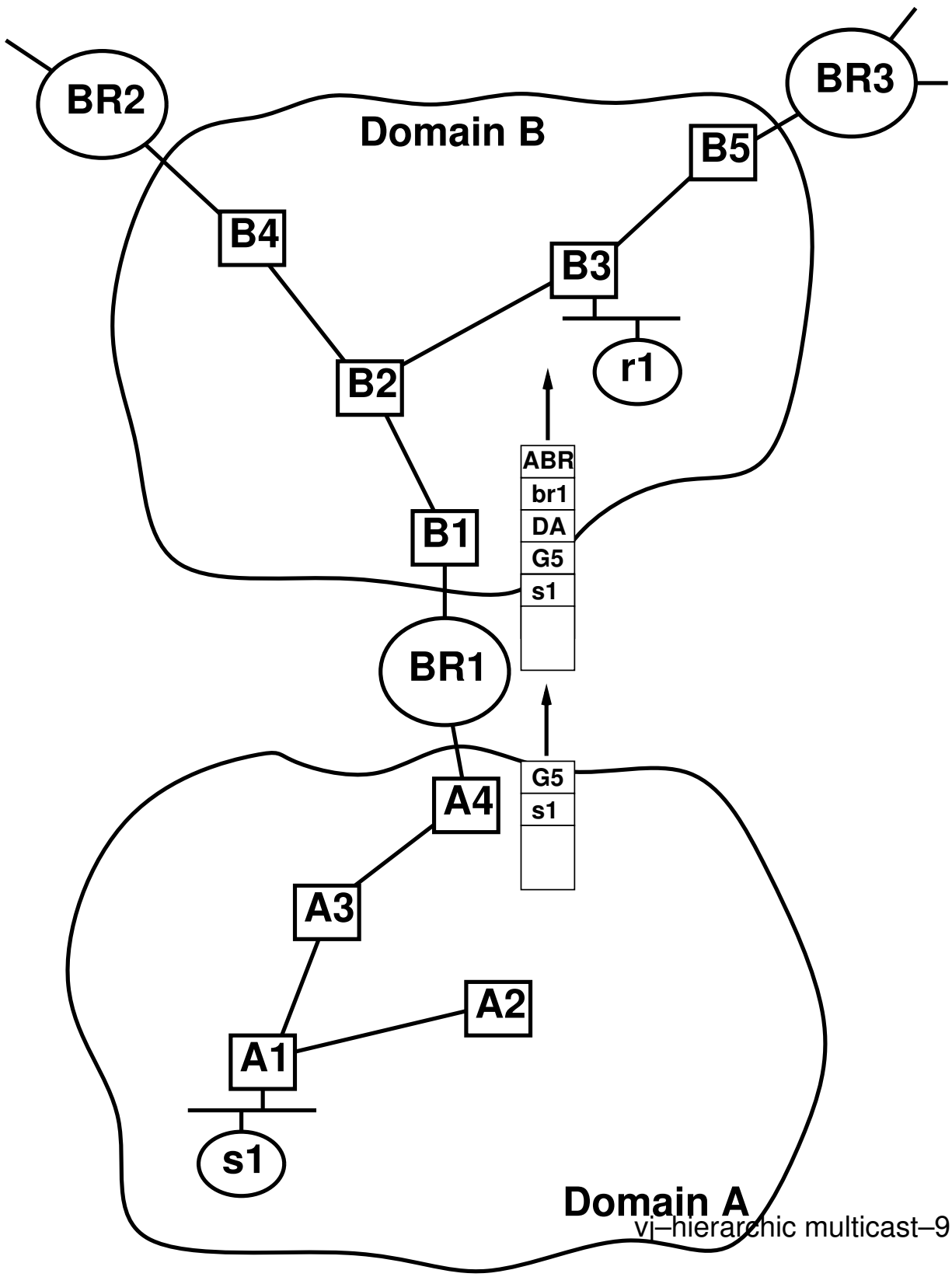

D5
R4
D4
R2
R3
D2
D3
R1
D1

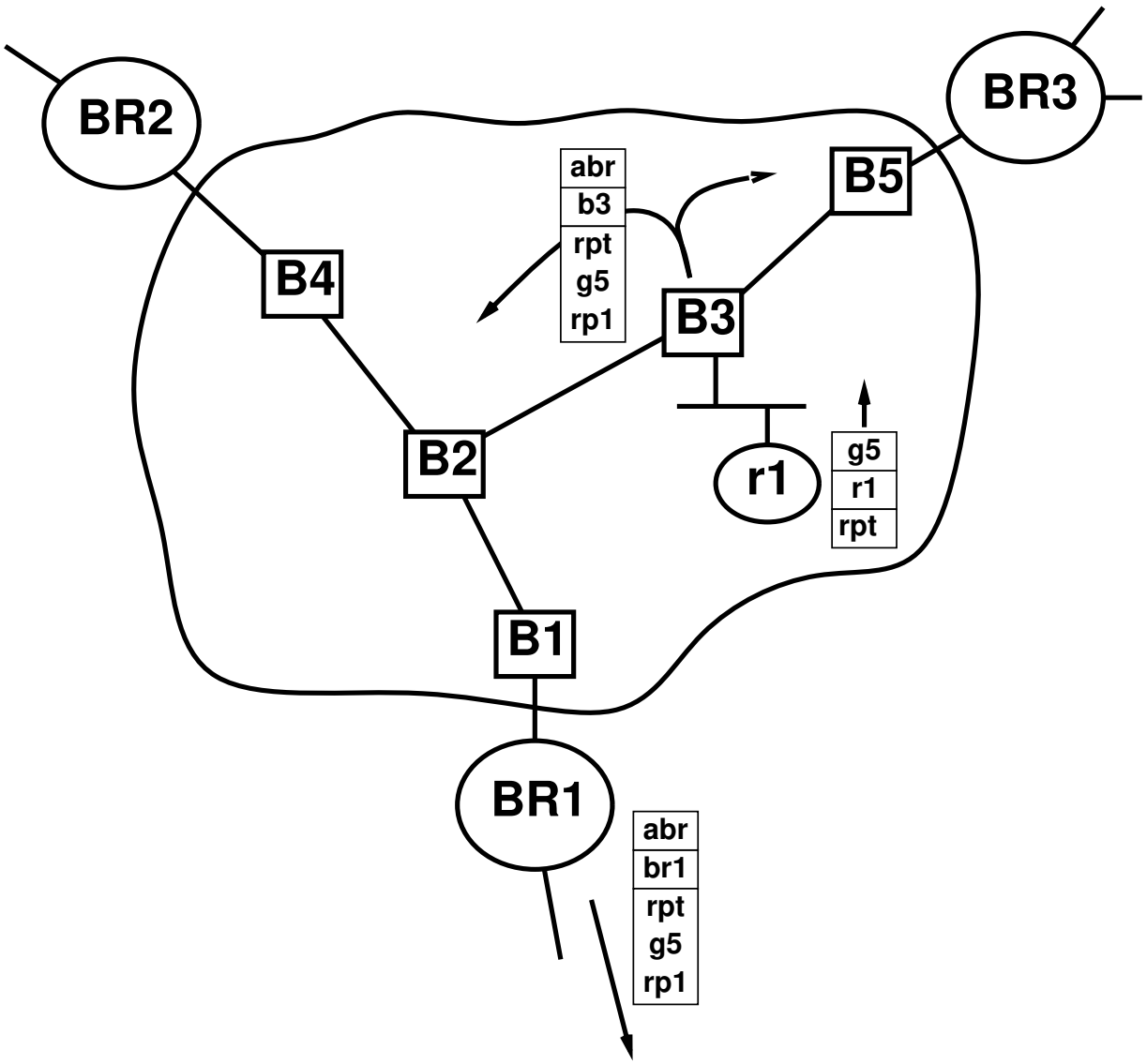
Plan (cont.):

- At border, level-2 (inter-domain) router peers with adjacent level-1 routers using level-1 protocols.

- Level-2 routers speak to each other via separate multicast routing protocol to build domain-level distribution trees.

- Level-2 routers use encapsulation and level-1 multicast to communicate.


BR2
BR3
Domain B
B5
B4
B3
r1
B2
ABR
br1
DA
G5
s1
B1
BR1
G5
s1
A4
A3
A2
A1
s1
Domain A


BR2
BR3
B4
B5
B3
B2
B1
r1
BR1
abr
b3
rpt
g5
rp1
g5
r1
rpt
abr
br1
rpt
g5
rp1

First level-2 protocol will be a variant of DVMRP because we have a lot of experience with it and there are well-tested, public domain implementations.

Will transition to a variant of PIM because we believe it has better scaling properties and gives better control of traffic flow (can use BGP unicast routes for interdomain distribution trees).

The fact that there's going to be a change at level-2 should be invisible and irrelevant to level-1 protocols.

To make this happen, we have to standardize the intra-to-inter protocol:

- Create all-border-router group
- define internal-router to all-border-router membership and RP reports
- (define border router to adjacent local router protocol)

Work to be done:

- fix existing DVMRP implementations to accept ‘default’ (all zero mask) route.
- upgrade existing level-1 protocols to generate ABR membership reports.
- make sure all new level-1 protocols or implementations generate these reports.

Work to be done (cont.):

- start engineering domain boundaries.
- work out a transition plan.
- design, implement and test first cut on inter-domain router code.