Differentiated Services for the Internet

Van Jacobson
van@ee.lbl.gov
Network Research Group
Berkeley National Laboratory
Berkeley, CA 94720

Internet2 Joint Applications/Engineering
QoS Workshop
Santa Clara, CA

May 21, 1998

IETF standards work in this area

- the long-running **int-serv** working group
  `http://diffserv.lcs.mit.edu/`

- the brand new **diff-serv** working group
  `http://www-nrg.ee.lbl.gov/diff-serv-arch/`

What problem are we solving?

Give “better” service to some traffic (at the expense of giving worse service to the rest).

ATM marketing fantasies to the contrary, QoS is a zero-sum game:

- it does not create bandwidth.
- it does not guarantee that you get better service.

What is the service?

There are two camps:

- **“Better best effort”** — ISPs want finer control of <u>relative</u> bandwidth allocation, particularly under heavy load (Implementation in terms of drop-preference or weighted-round-robin).

- **“Virtual leased line”** — Users want an end-to-end <u>absolute</u> bandwidth allocation, independent of other traffic (Implementation in terms of priority queuing and strict policing).

The IETF is more vocal on the former but there is demand for both.

What are the target applications / protocols?

Bad question. In 1978, the answer was RJE. In 1988, email/ftp. In 1998, probably web. This too will change.

IP/TCP/UDP/IGMP/OSPF/BGP work for any application. Differentiated services must too.

Control of Sharing

QoS gives an institution control of how it's bandwidth is shared among different users.

⇒ There must be some way for the institution to communicate its sharing policy to the network.

⇒ Since users can't get whatever they want, there's incentive to steal and architecture must include security and workable incentive and trust models.

(This is especially important if design attempts to limit state in the network.)

Scaling

A differentiated services mechanism must work at the scale of the Internet (e.g., millions of networks) and at the full range of speeds of the Internet (e.g., Gb/s links). To get that kind of scaling the design must:

- push all the state to the edges, and
- force all per-conversation work (e.g., shaping, policing) to the edges.

Scaling (cont.)

⇒ Edge-only state suggests that special/normal service indication must be carried in the packet.

⇒ Administrative diversity and high speed forwarding both argue for very simple semantics on that indication. E.g., a few bits of special/normal.

⇒ No state in center means it sees only aggregates (potential fairness problems).

Interdomain service

- Almost all Internet traffic crosses *many* administrative boundaries. End-to-end service implies that all those independent units agree to treat the traffic as special. Multilateral agreements rarely work.

- ISPs are competitive enterprises. They act in their own best interests and, where necessary, against the interest of their competitors.

Interdomain service (cont.)

- ⇒ End-to-end service should be constructed from bilateral agreements.
- ⇒ Service must not require extending trust or control across administrative boundaries.
- ⇒ There must be fault isolation (a customer shouldn’t be able to trash another but is welcome to trash himself).

Service characterization

For economically viable service, customers have to know what they're buying.

- Delivered service *can't* depend on other people's traffic.

- The customer *must* be able to measure conformance.

$\Rightarrow$ ISP must not "over sell" premium service (simple queuing theory: if aggregate inflow $>$ outflow, delivered bandwidth, drop rate and/or delay can be *arbitrarily bad*).

$\Rightarrow$ Even if service is limitted to capacity, aggregation can cause some users to get poor or non-existant service unless **traffic shaping** done at borders.

Are we standardizing “Services” or Packet Forwarding “Behaviors”?

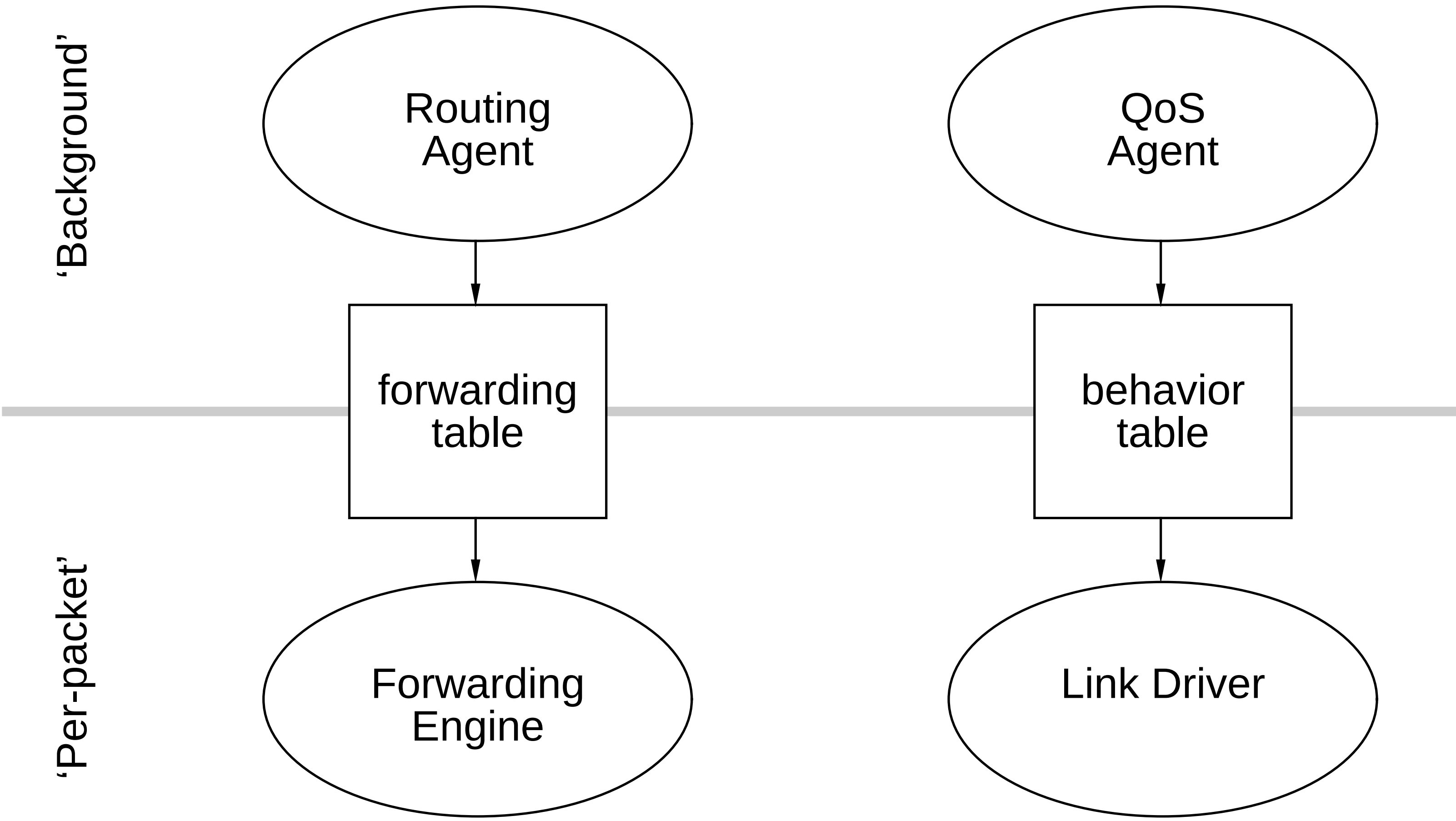
If the answer is “services”, everyone has to agree on what constitutes a useful service and every router has to implement the machinery for it (i.e., to deploy a new service, you have to upgrade the world).

Since a router can’t actually do that many different things to a packet, it makes more sense to standardize forwarding behavior (e.g., “send this packet first” or “drop this packet last”).

$$\text{Behaviors} + \text{Rules} = \text{Services}$$

(think in terms of IP Forwarding / Routing architectural separation)

IP Gateway Architecture



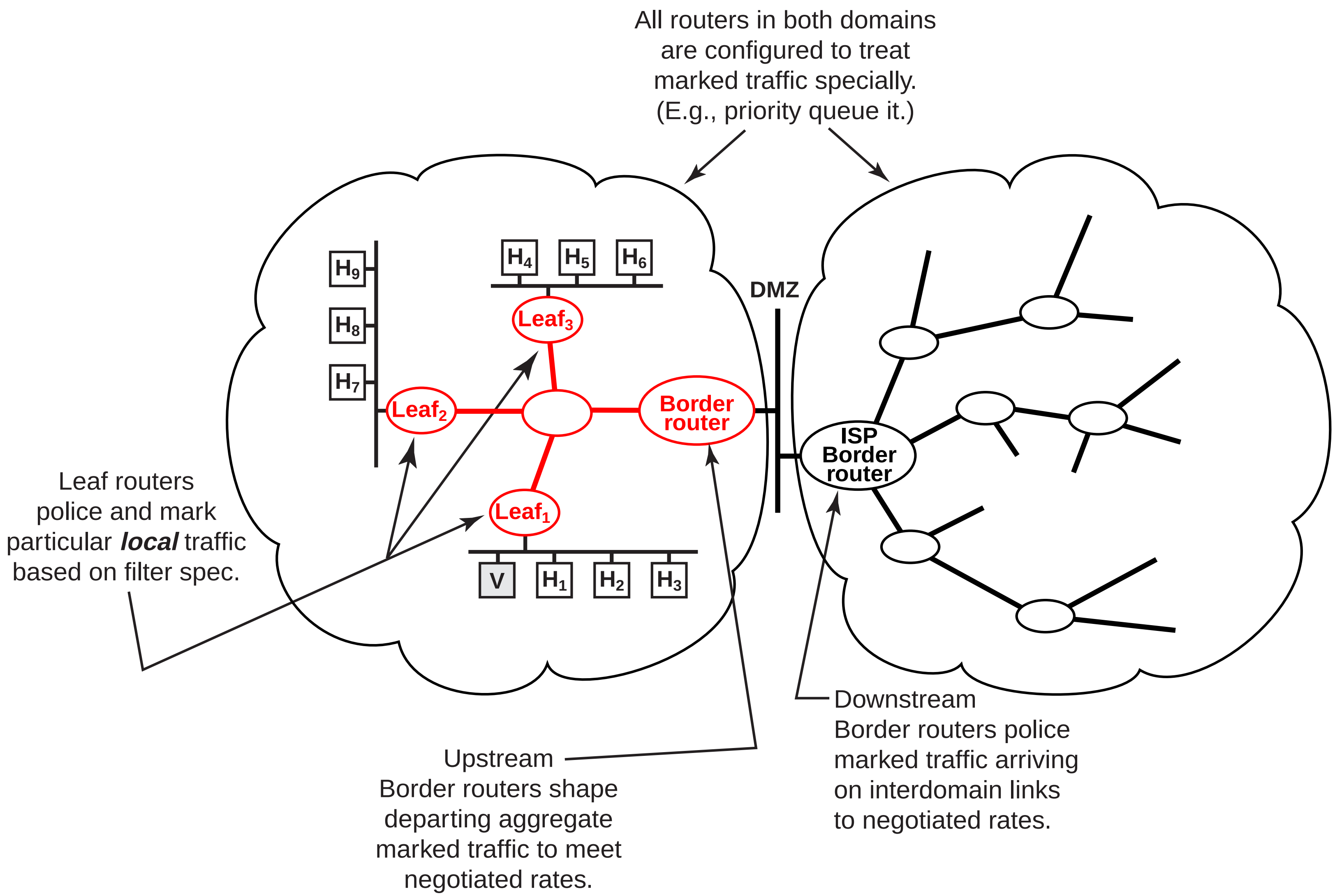
So how do we do it?

Ten years ago, Dave Clark (MIT LCS) and I proposed "**edge-tagging**" as a scalable way of offering differentiated services.

- Leaf router adjacent to the source has traffic signature for "special" traffic and "profile meter" giving its characteristics.

- That router "marks" (sets IP ToS field) in all special traffic that conforms to profile meter.

- Leaf routers unmark all other traffic.


All routers in both domains are configured to treat marked traffic specially. (E.g., priority queue it.)
DMZ
H9
H8
H7
H4
H5
H6
Leaf3
Leaf2
Border router
Leaf1
V
H1
H2
H3
ISP Border router
Leaf routers police and mark particular *local* traffic based on filter spec.
Upstream Border routers shape departing aggregate marked traffic to meet negotiated rates.
Downstream Border routers police marked traffic arriving on interdomain links to negotiated rates.

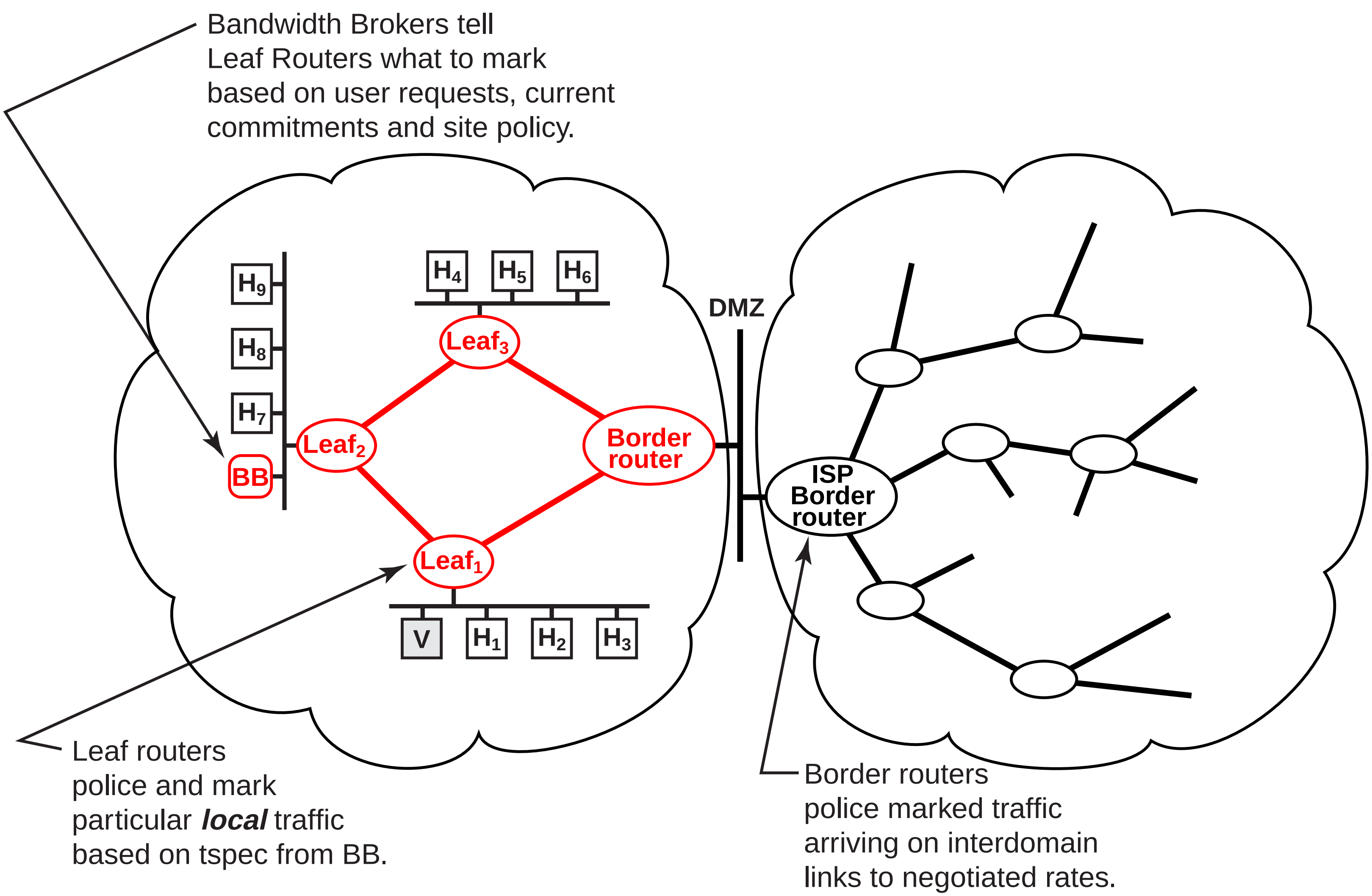
But there are still problems:

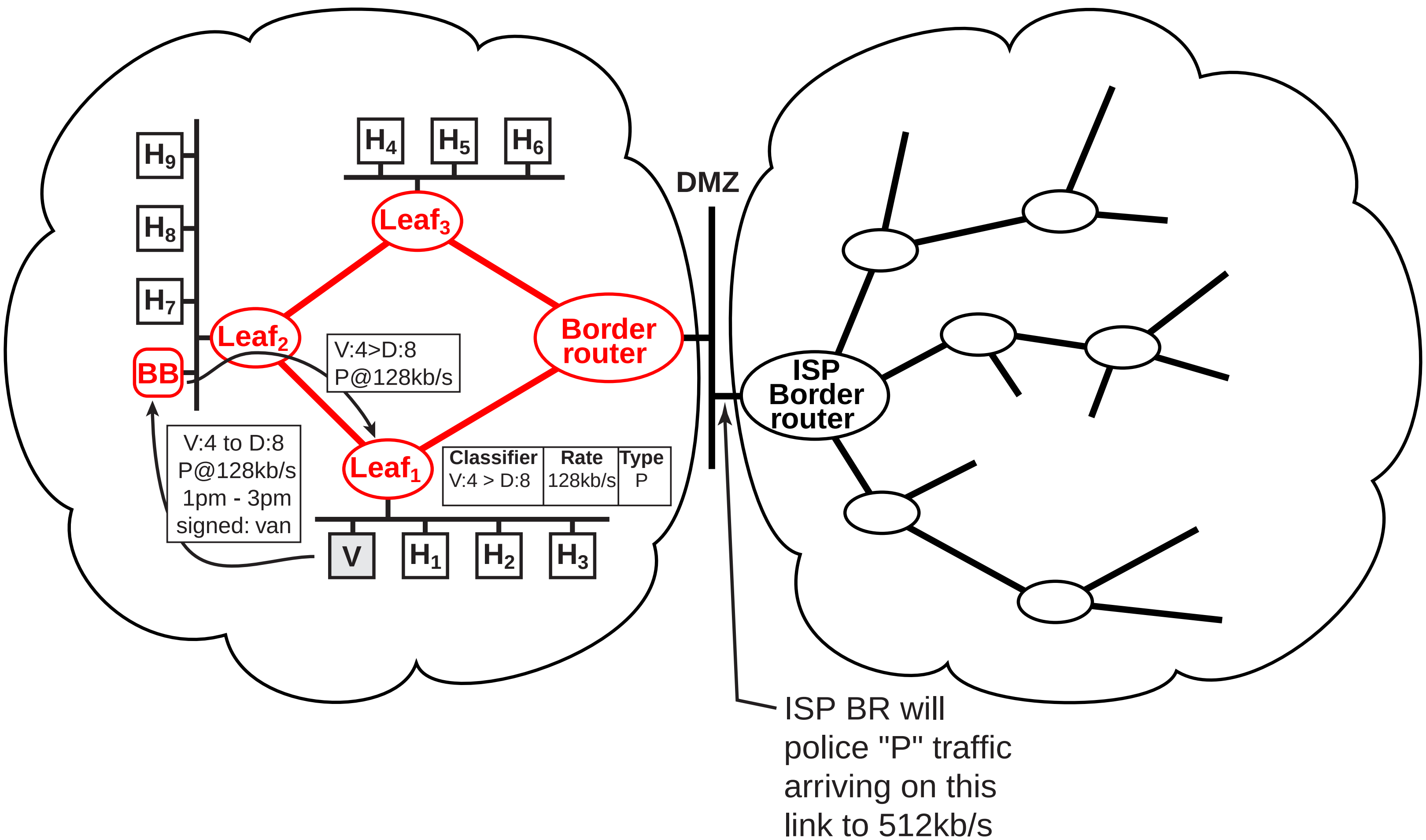
- Who decides what users get to request special service?
- Where is organizational policy on use of limited bandwidth implemented?
- Who tells the edge router what to tag?
- Who makes sure that simultaneous uses of special service fit within allocation?

Answer: Introduce a **Bandwidth Broker** (BB) to be repository of policy database of priority and limits for user & project access to special bandwidth. Repository includes user credentials so requests can be authenticated.

BB is part of network infrastructure so can have trusted, secure association with all routers.

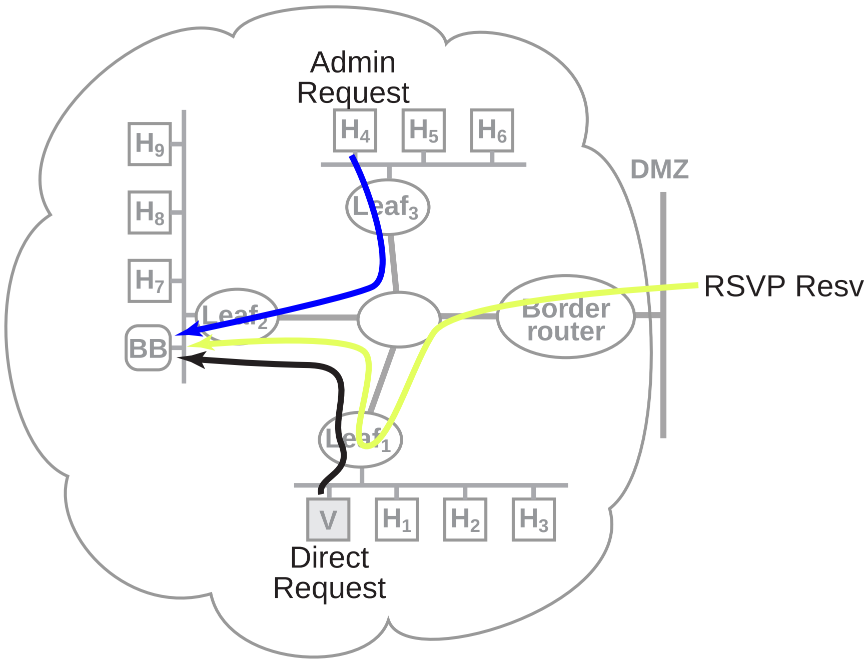
Requests go from user to BB (so it can record use and resolve conflicts) then to appropriate router so security model is well-founded.


Bandwidth Brokers tell
Leaf Routers what to mark
based on user requests, current
commitments and site policy.
H9
H8
H7
BB
H4
H5
H6
Leaf3
Leaf2
Border
router
Leaf1
V
H1
H2
H3
DMZ
ISP
Border
router
Leaf routers
police and mark
particular *local* traffic
based on tspec from BB.
Border routers
police marked traffic
arriving on interdomain
links to negotiated rates.



H9
H8
H7
BB
Leaf2
H4
H5
H6
Leaf3
Border router
DMZ
V:4>D:8
P@128kb/s
V:4 to D:8
P@128kb/s
1pm - 3pm
signed: van
Leaf1
Classifier | Rate | Type
V:4 > D:8 | 128kb/s | P
V
H1
H2
H3
ISP Border router
ISP BR will police "P" traffic arriving on this link to 512kb/s

Who talks to the BB? How?



Whatever works.

Certainly site network admins but also suitably enhanced (and authorized) applications, either via direct messages or indirectly via RSVP.

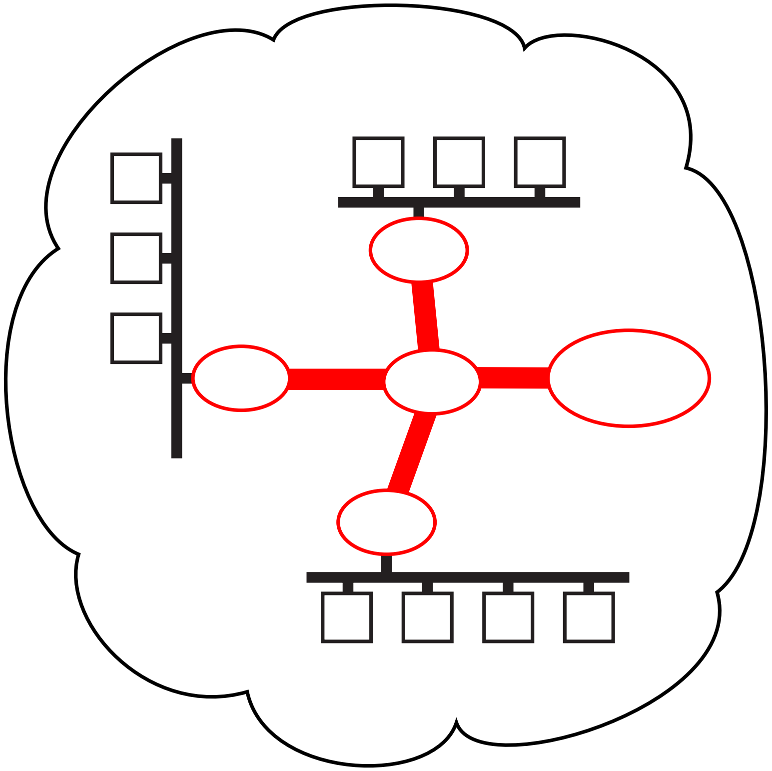
Paths or Boundaries?

Is the bandwidth being allocated across a boundary or along a path?

There's no one right answer but path allocation is <u>much</u> harder:

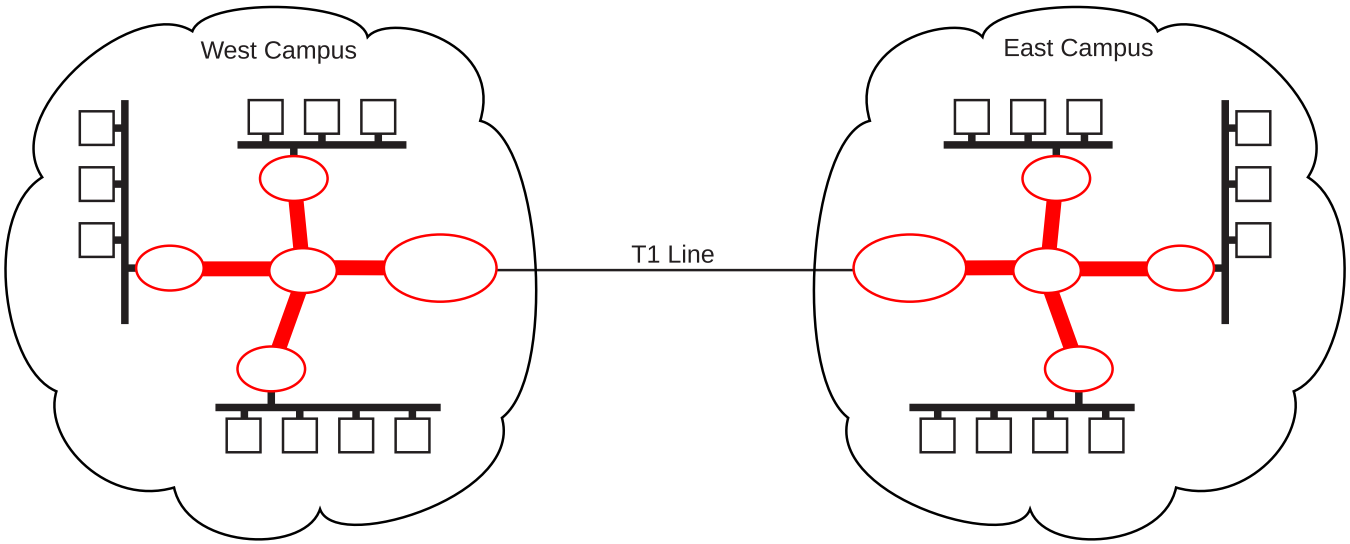
- forces allocator to know full topology.
- number of paths grows combinatorily with diameter.
- may have to subvert routing to “pin” path.

Why solve a hard problem if most topologies allow easy solutions?



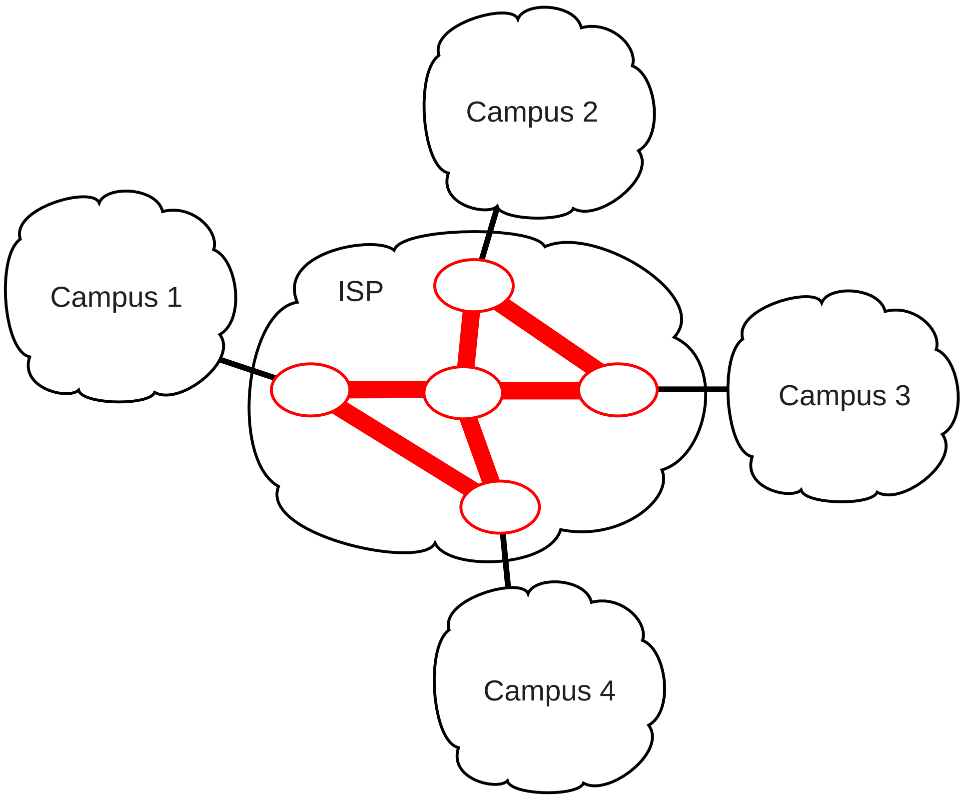
Example: Campus with at least 10Mb/s everywhere internally.

Allocation from a single 10Mb/s "premium" bandwidth pool allows 300 simultaneous voice/video sessions with <u>no</u> topological knowledge in allocator.



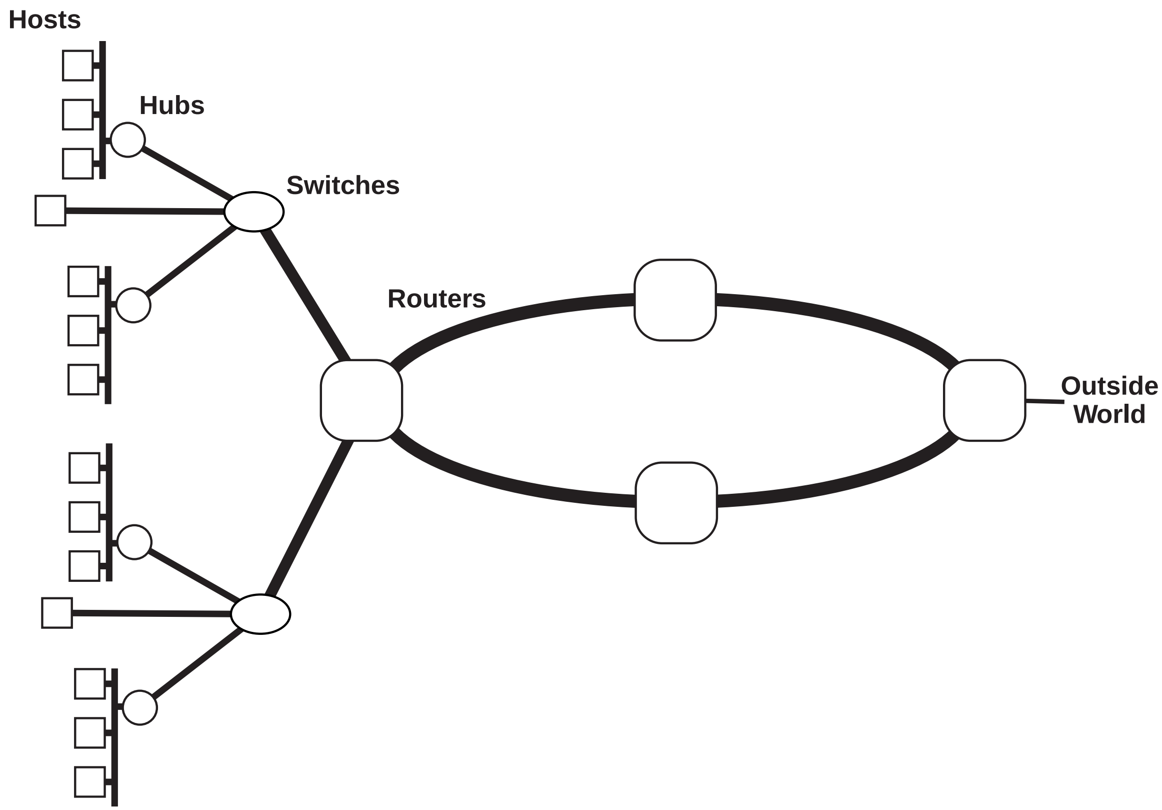
Example: Two campuses connected by a low-speed WAN link.

Add a T1-sized pool of bandwidth for “external” conversations to previous allocator.



Example: Multiple campuses connected to a common ISP by low-speed tails.

Each campus runs a simple allocator for its campus$\rightarrow$ ISP link and ISP runs a simple allocator for each of its ISP$\rightarrow$ campus links.



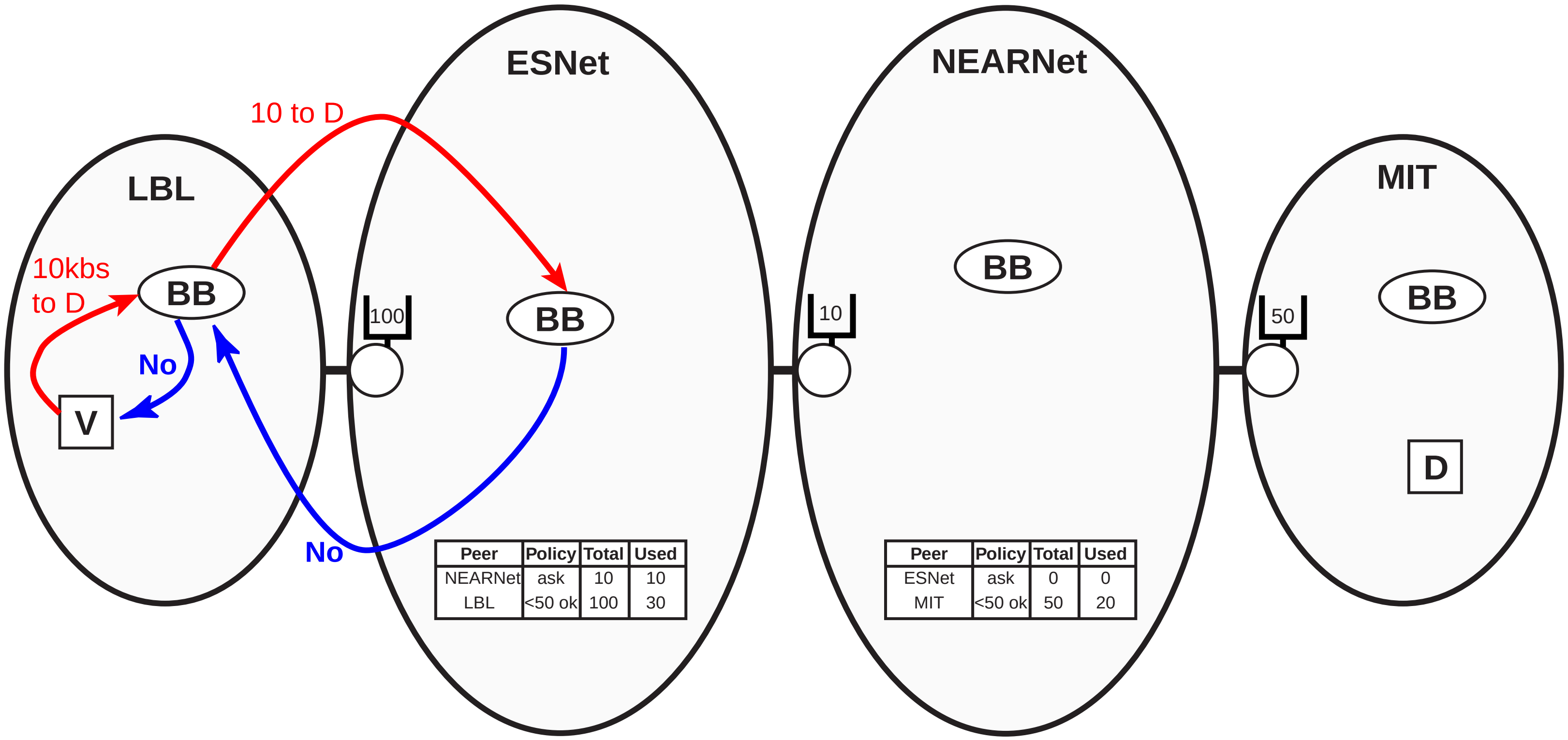
Example: The elements of a campus network have very different capabilities. If resources are not assigned to a path you can account for hardware realities and put the tagging/shaping/policing machinery only where you need it).

Interdomain setup and control

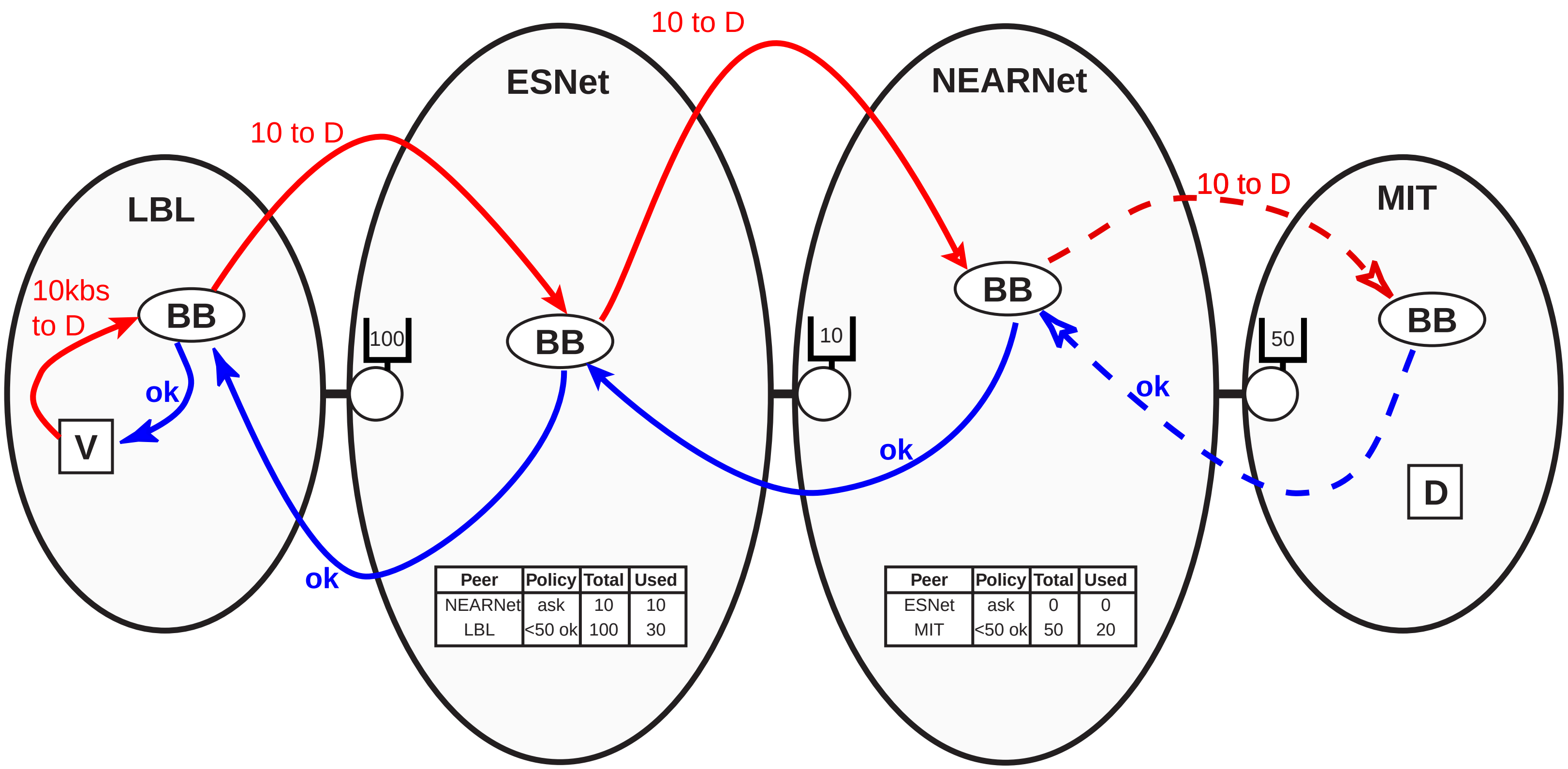
The Bandwidth Broker must exist so bandwidth sharing decisions will reflect institutional priorities and policies.

But it also provides a natural vehicle for dynamic, interdomain control.

Example: Interdomain request with all resources in use.



Example: End-to-end interdomain request.



Peer	Policy	Total	Used
NEARNet	ask	10	10
LBL	<50 ok	100	30

Peer	Policy	Total	Used
ESNet	ask	0	0
MIT	<50 ok	50	20