Moving the Internet Beyond Best-effort

Van Jacobson

van@ee.lbl.gov

Network Research Group
Berkeley National Laboratory
Berkeley, CA 94720

DOE LSN Research Workshop

Reston, VA

January 5, 1998

What problem are we solving?

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

Despite ATM QoS fantasies to the contrary, it’s a zero-sum game.

Do users or institutions control what traffic is treated specially?

If users, then solution is trivial – users individually decide which of their traffic is most important and attach labels to convey this to the network. There are no trust or coordination issues.

Unfortunately this doesn't work (it assumes there will always be enough bandwidth to handle the sum of all users' special needs).

In a world of finite bandwidth, institutions have to control sharing.

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

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

Is the control end-to-end, hop-by-hop, intra-domain, inter-domain, per-path, or per-boundary?

Yes.

(The Internet is big, there must be a win to all forms of incremental deployment or the service will never get deployed.)

What is the service: “Better best effort” or “Virtual leased line”?

Yes.

(Current demand for the former seems to be mostly intra-domain traffic control while the latter seems to be inter-domain service offerings. 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.

Design Constraints — Scaling

A Differentiated Service 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.

Design Constraints — Scaling

⇒ 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., one or two bits of special/normal.

⇒ No state in center means everything but edge sees only aggregates (potential fairness problems).

Dave Clark (MIT LCS) has proposed “**edge-tagging**” as a scalable way of offering differentiated services.

- Leaf router adjacent to the source(s) has traffic signature for “special” traffic and “profile meter” giving its characteristics.

- That router “marks” (sets IP precedence field) in all special traffic that conforms to profile meter.

- All routers unmark all other traffic.

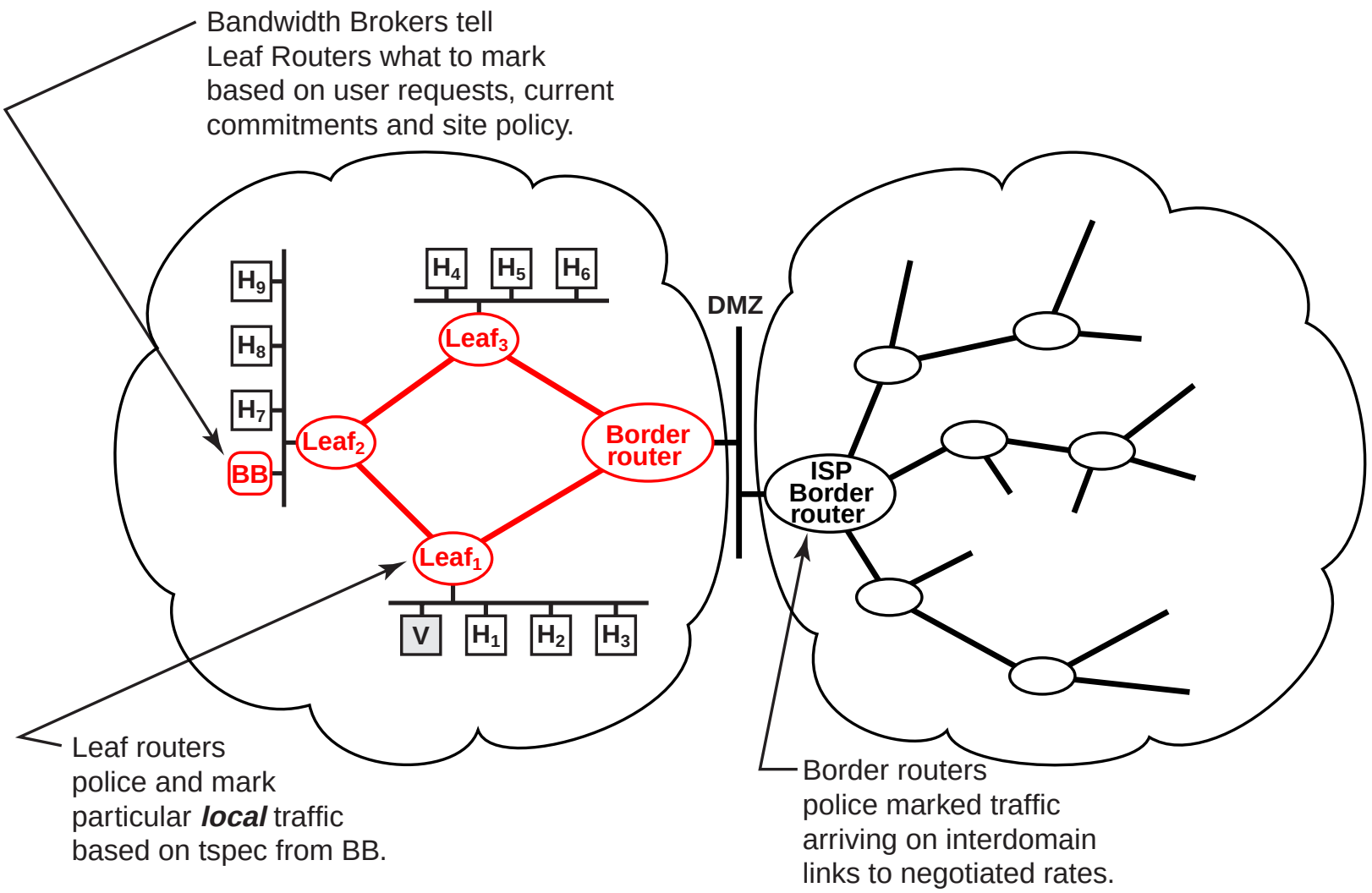
But there are still problems:

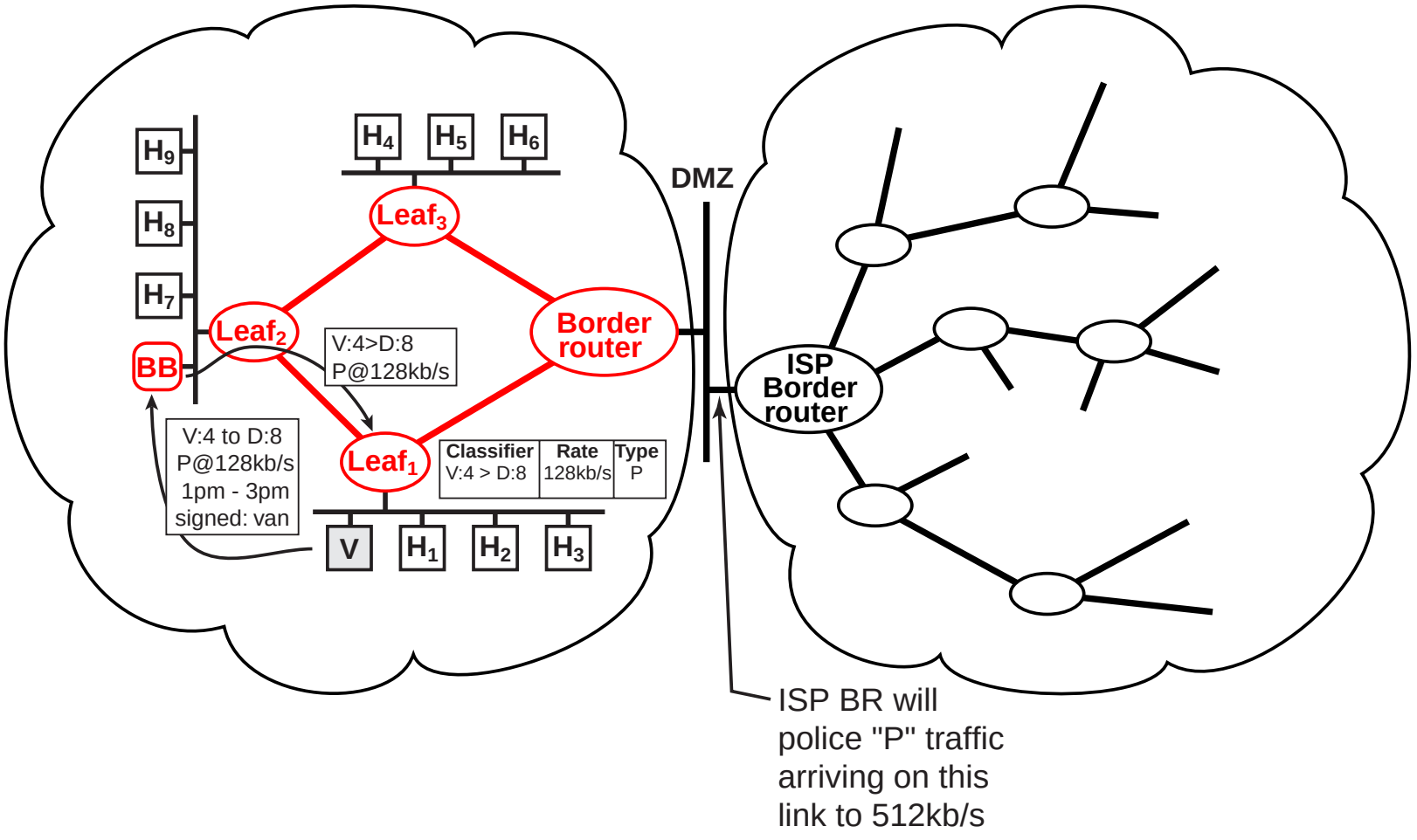
- Who decides what users get to request special service?
- Where is organizational policy on use of limited bandwidth implemented?
- Who tells the edge router what to tag?
- Who makes sure that simultaneous uses of special service fit within allocation?

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