

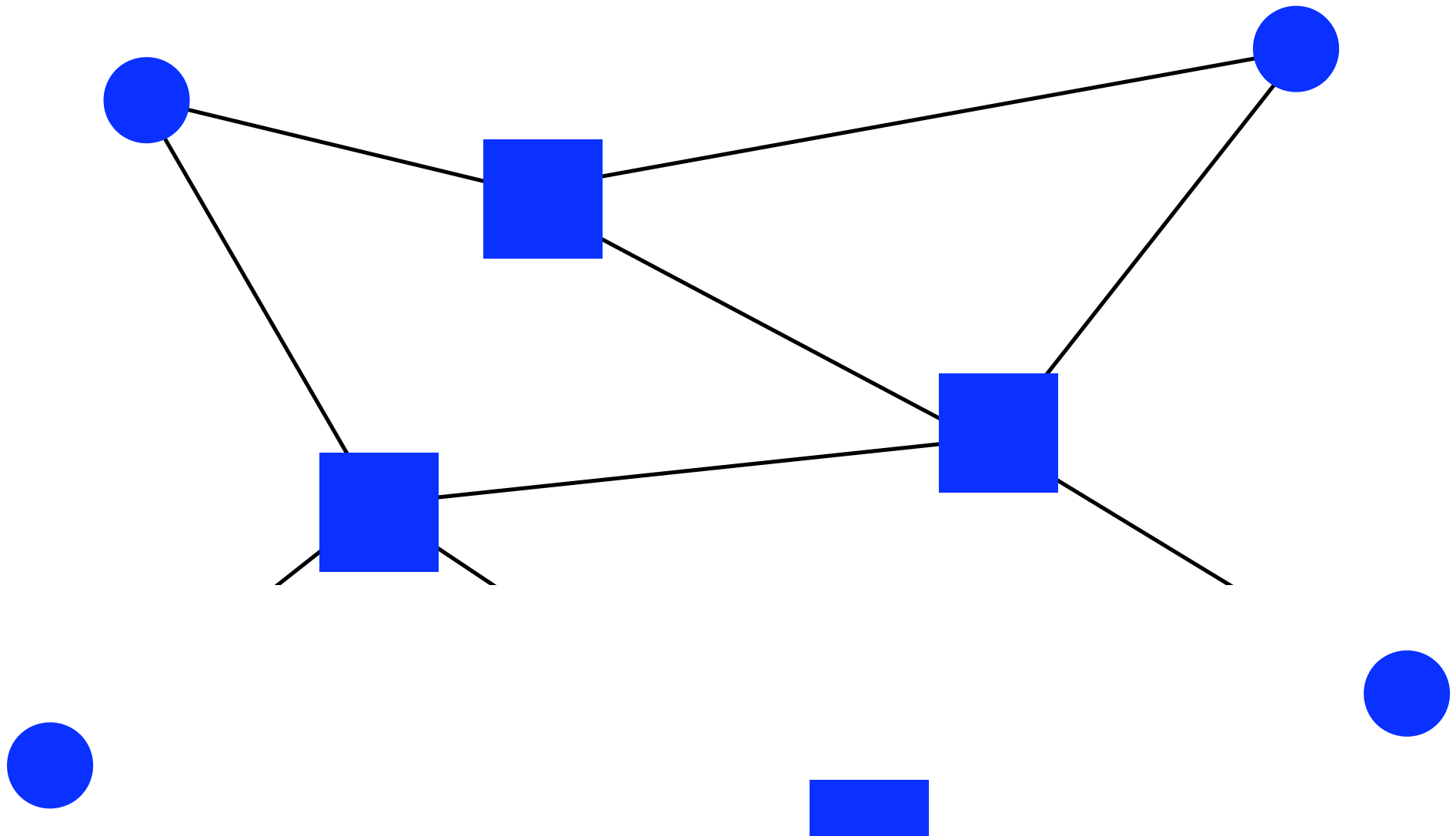
Better IP Routing

Brussels, January 14th, 2005

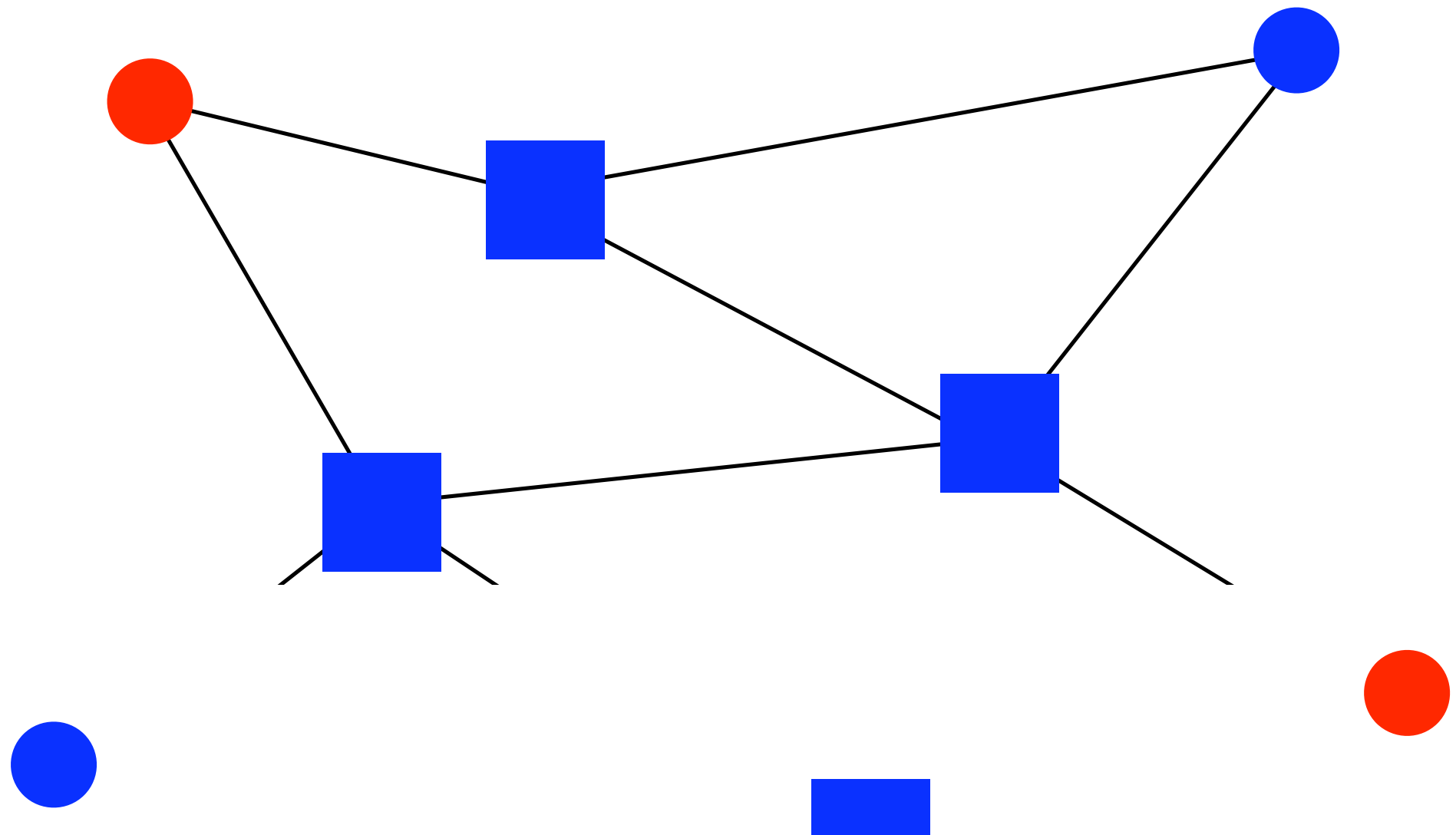
Ijitsch van Beijnum

e · x · p · e · r · t
BGP

Multicast Routing

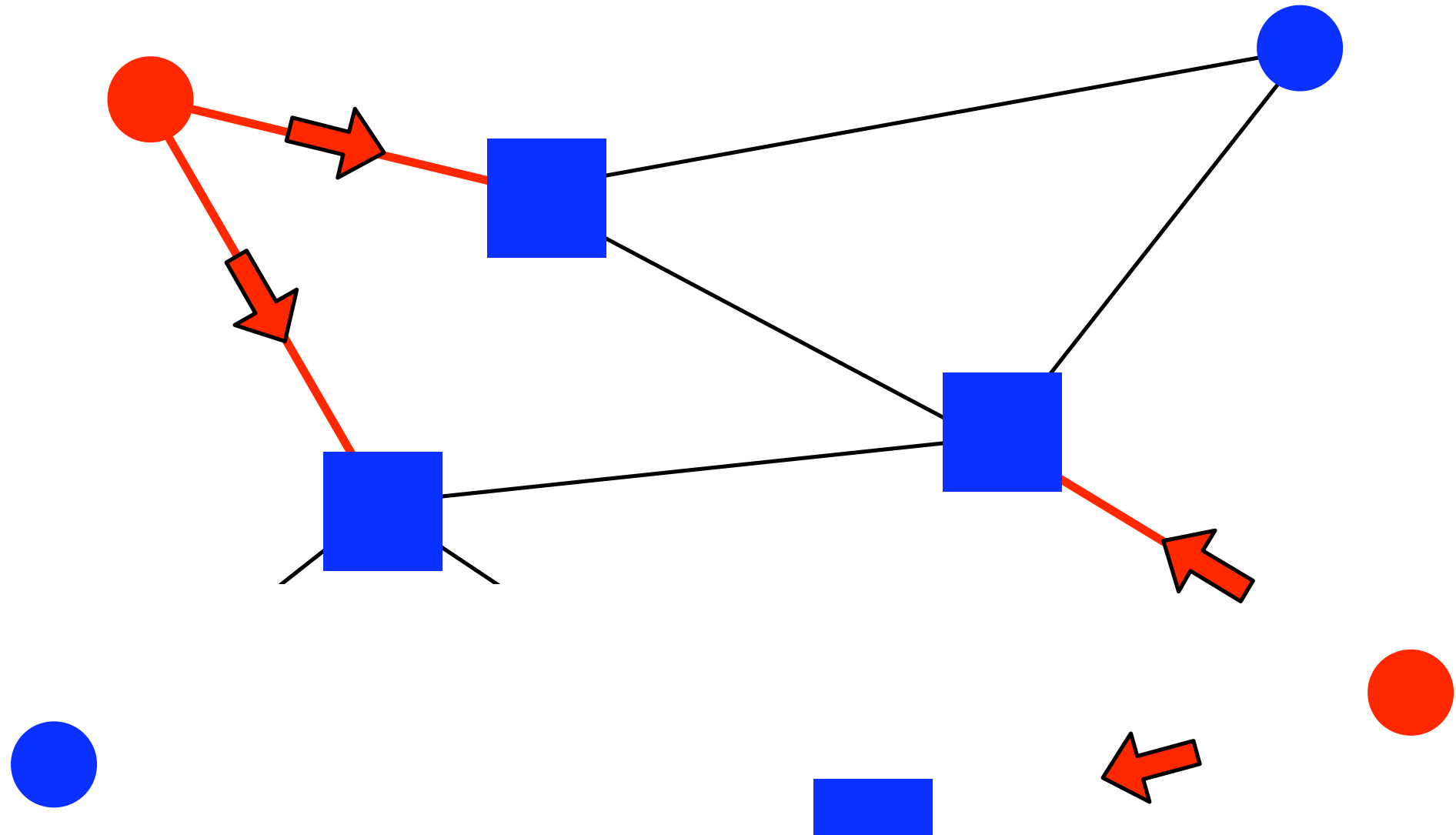


Multicast Hosts

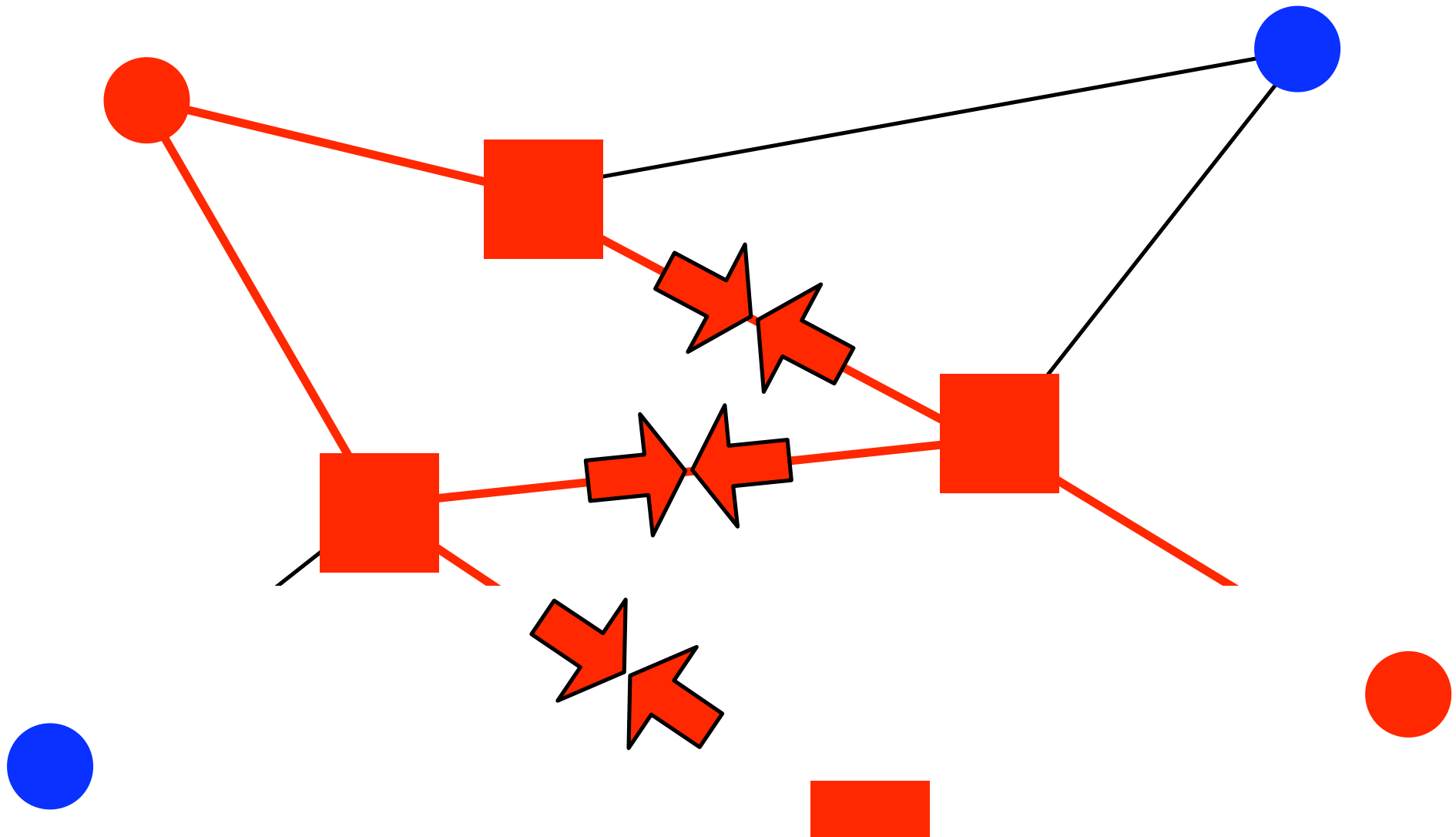


Multicast Basics

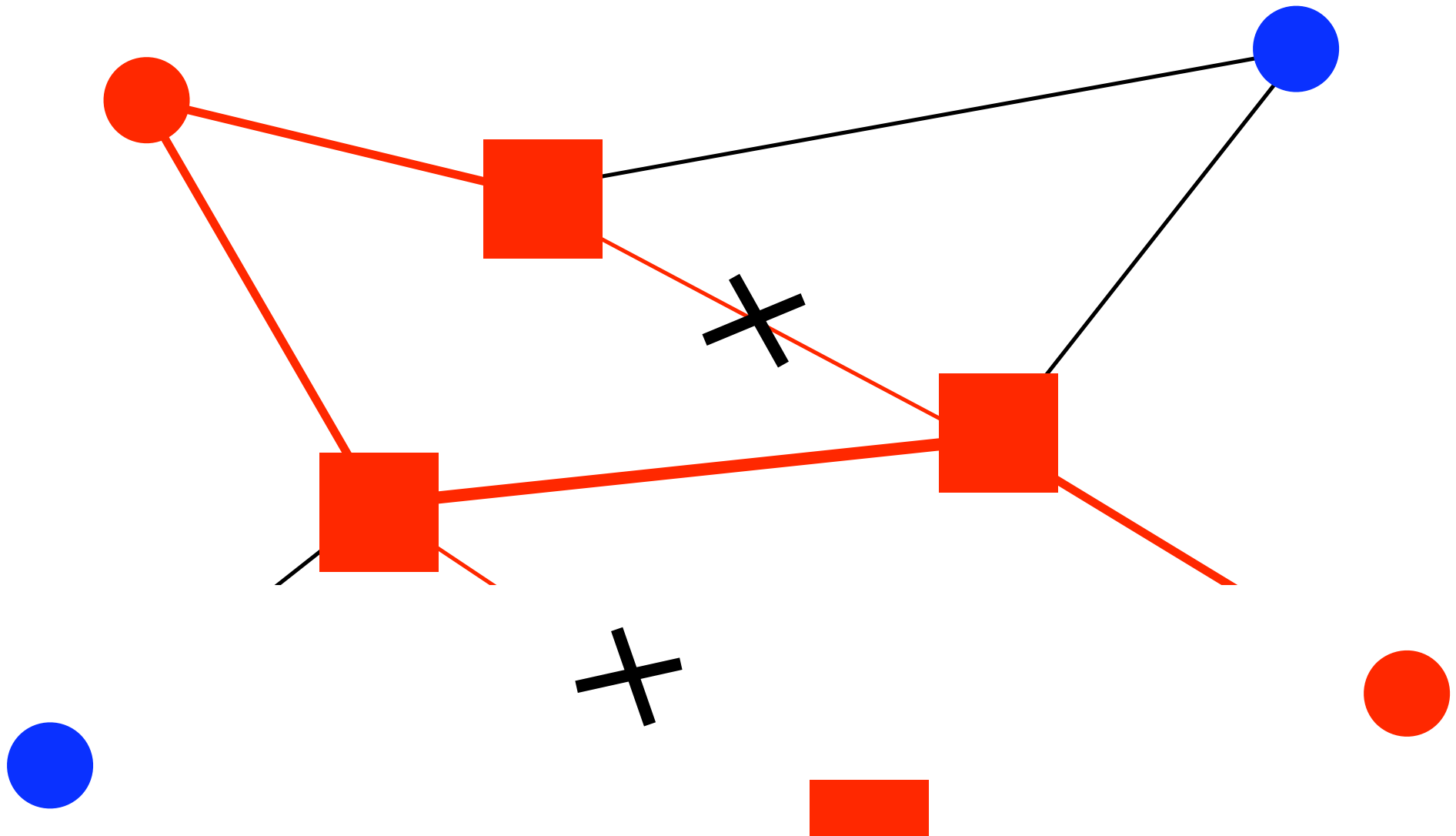
Hosts send IGMP/MLD to routers to indicate they're listening



Multicast Routing



Pruning; RPF Check



Multicast

- Very different from IPv4 unicast
- Need to build forwarding tree
 - Any Source Multicast (ASM): a tree per group member (yuck!)
 - Source Specific Multicast (SSM): single tree per multicast group
- Prune trees, but still multiple copies of one packet, unless it conforms to tree (PBF)

MC Routing Protocols

- Distribute group addresses: up to apps
- Host to router (and switches): IGMP/MLD
- Distribute source addresses between ASes: MSDP, MBGP
- Create forwarding trees within the AS: usually PIM Sparse Mode

IPv6

- Routing IPv6: same as v4, larger addresses
- Intra-AS: RIPng, OSPFv3, IS-IS
- Multiprotocol extensions for BGP
 - Carry IPv6 routing information over IPv4 or IPv6 TCP sessions

Address Policies

- Very different between v4 and v6!
- Policy: conservation, registration, routability
- In v4 conservation: give out small blocks
- In v6 routability: give out large blocks, no real provider independent space (so far)
- In v6 ISPs /32 or bigger, all endusers /48

Global Routing Tables

- IPv4:
 - 18667 active ASes
 - 149521 prefixes, 8 per AS
- IPv6:
 - 513 active ASes (2.7%)
 - 705 prefixes, 1.4 per AS (0.5%)

IPv6 Table Explosion?

- IANA and RIRs say: ok to filter at /32
- (but some micro allocations: root DNS etc)
- Some are concerned about unique site locals showing up in global routing table
- Can we avoid or limit PI?
- Can't accept /48s in v6 like /24s in v4: everyone qualifies

IPv6 Developments

- Active work on host-based multihoming (invisible in interdomain routing)
- Nevertheless pressure for provider independent addressing
- Likely that RIRs will be getting VERY large blocks (/12 or even /6)

BGP Security

- IETF RPSEC wg in requirements phase
- S-BGP (Secure BGP) proposed by BBN
- soBGP (secure origin BGP) proposed by Cisco
- Relatively easy to secure prefix/AS mapping
- Unwanted propagation of legitimate announcement much harder to fix

S-BGP

- Draft by BBN around for some time now, proof-of-concept implementation available
- Sign every update (including next hop AS, so no more peer group optimization)
- Carry authentication data in path attribute
- Heavy: 4 x the memory, signature check for every AS in every path, delays startup

soBGP

- Newer than S-BGP, no code AFAIK
- Mostly tie prefix to source AS
- But can be extended with additional checks
- Authentication data in new BGP message
- Architecture allows offloading to special purpose box, not as heavy as S-BGP

Other BGP Security

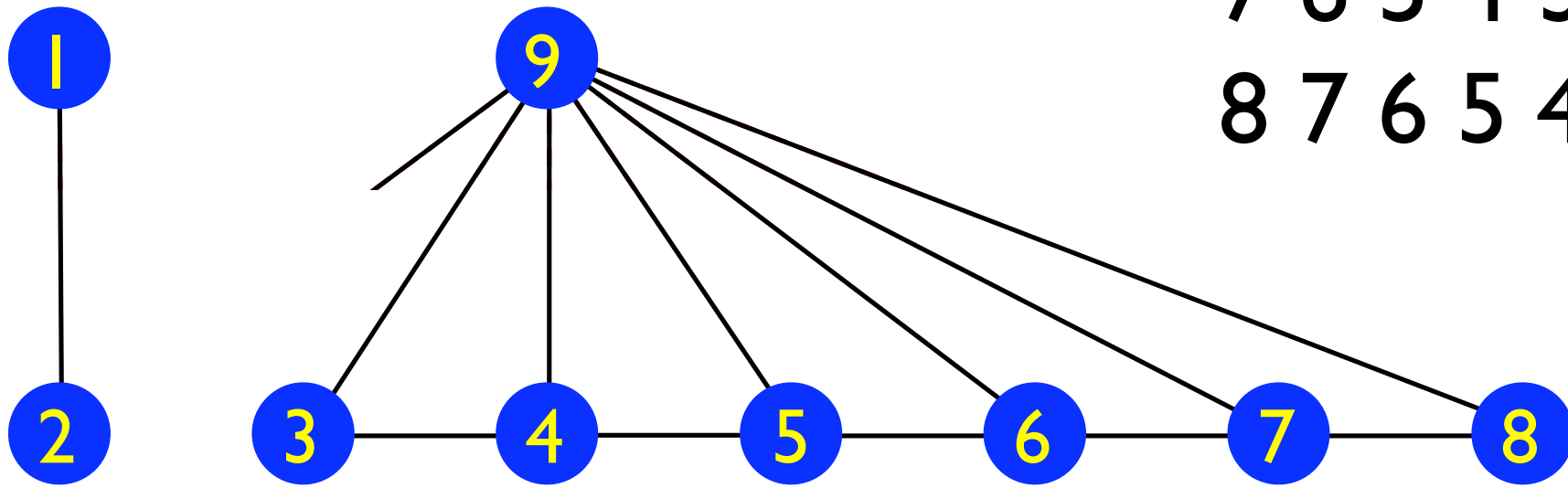
- TCP MD5 option more widely used, but not great: too much CPU, kernel hacks, open to crypto DoS
- IPsec not BGP-specific and much better
- Can run BGP over loopbacks to avoid management plane exposure
- Separate data/management contrary to IP view of the world, new risks, do it anyway?

General BGP Problems

- AS path length only real end-to-end metric, AS hierarchy too flat to be very useful
- "Count to infinity" and flap amplification
- Can only do hop-by-hop
- Doesn't detect end-to-end reachability problems (black holes)
- Global table size: not enough aggregation

"Count to Infinity"

2 1
3 2 1
4 3 2 1
5 4 3 2 1
6 5 4 3 2 1
7 6 5 4 3 2 1
8 7 6 5 4 3 2 1



Non-Problems

- Work per-prefix rather than per-AS: no longer an issue in IPv6?
- iBGP scalability and interaction with IGP?
- AS# depletion: 32 bit AS in IETF pipelines (but seems to be staying there...)

Research

- Do we want to keep BGP or rebuild from scratch?
- Better metrics (delay, bandwidth?)
- Introduce link-state mechanisms
- Support routing on more than just destination address?
- Automatic aggregation (geography...)

Last Minute

- More dynamic environment for BGP because of on-demand L2 or L1 paths
- Not try to reserve or discover bandwidth: just blast packets at full speed
 - (need to prioritize "blast" and "regular" differently of course)

Good Points BGP

- Leverages proven transport, easy to adopt
IPsec
- Distributed computation
- Policy support

- References:
 - <http://www.irtf.org/charters/routing.html>
 - draft-irtf-routing-history
 - RFC 3569
 - IETF multi6 and RPSEC wgs
- iljitsch@muada.com
- <http://www.bgpexpert.com/>