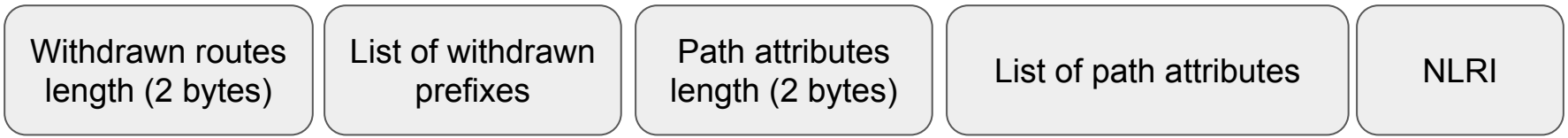
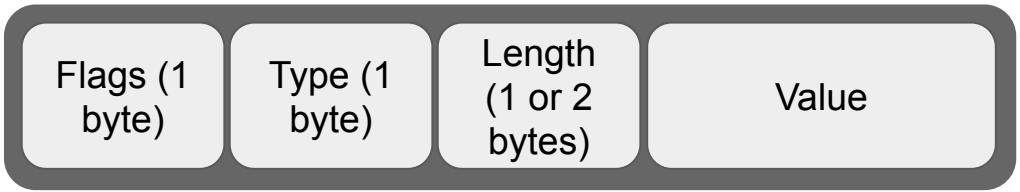
BGP routing policies in ISP networks

BGP - An Overview

- Protocol to allow Autonomous Systems to communicate routing information for IP prefixes
- After TCP connection between routers is established, information is updated with the following message format:



Attributes include: Origin (IGP, EGP, or incomplete), AS path, Next hop (IP address), MED, LocalPref, Communities



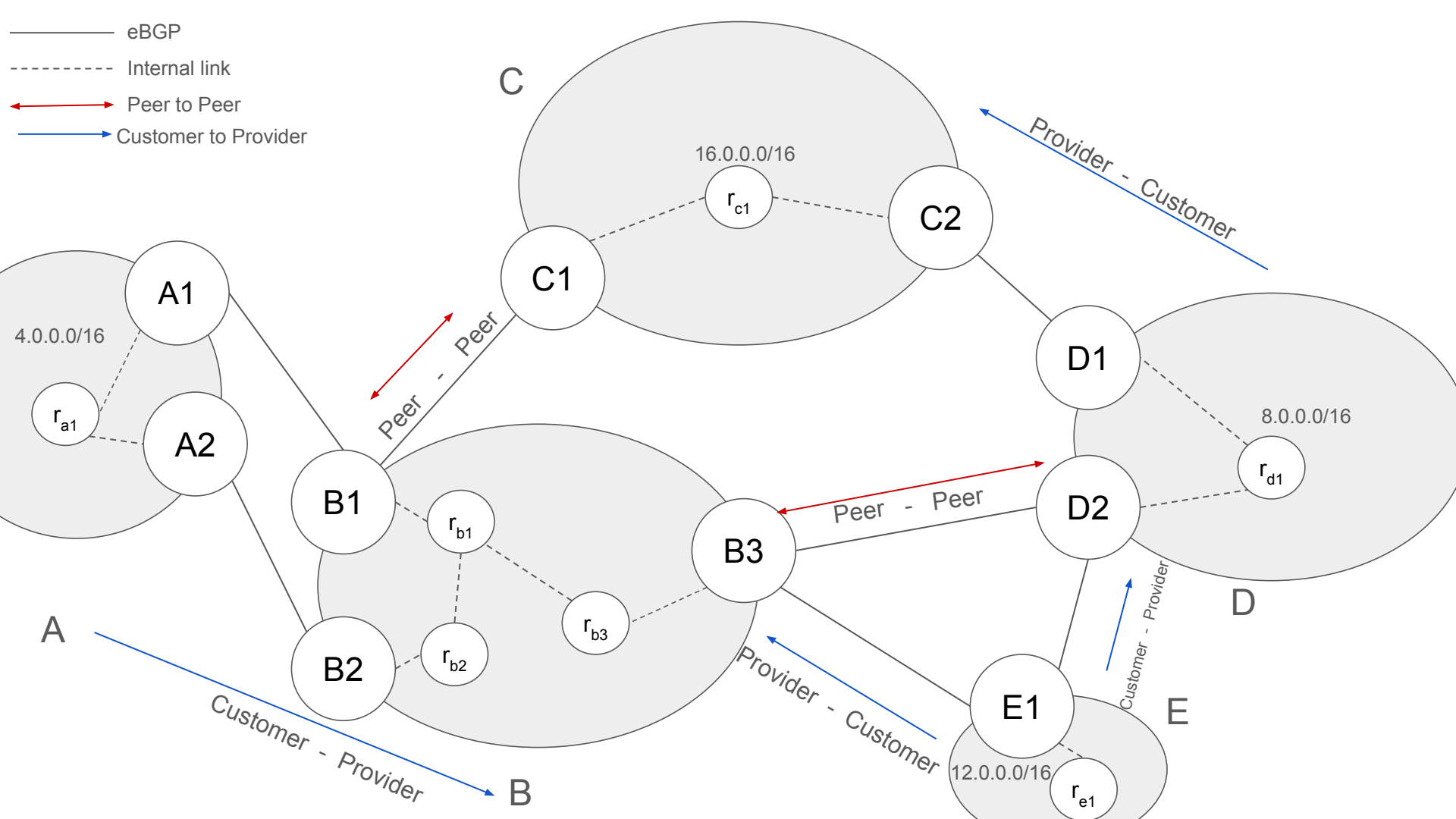
BGP - An Overview

A route advertisement is processed in three steps:

1. Import policy - filters routes / adds information as necessary
2. Decision process:

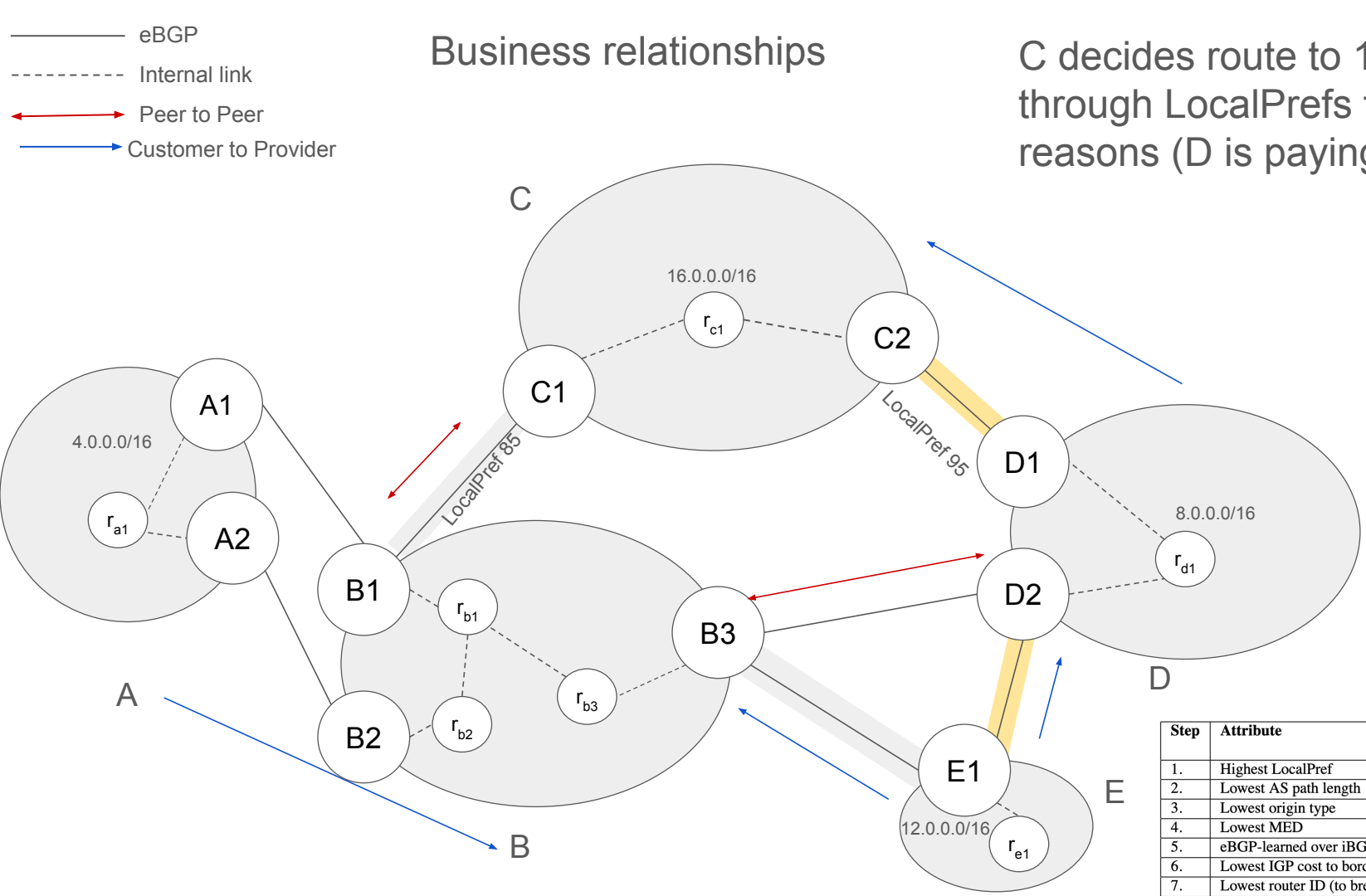
Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

3. Export policy - decides which routers to advertise to


eBGP
Internal link
Peer to Peer
Customer to Provider
C
16.0.0.0/16
r_{c1}
C2
C1
Provider - Customer
A1
4.0.0.0/16
r_{a1}
A2
Peer - Peer
Peer - Peer
D1
8.0.0.0/16
r_{d1}
D2
B1
r_{b1}
B3
Peer - Peer
A
r_{b3}
B2
r_{b2}
Provider - Customer
Customer - Provider
E1
E
Customer - Provider
12.0.0.0/16
r_{e1}
D
B

Business relationships

C decides route to 12.0.0.0/16 through LocalPrefs for business reasons (D is paying customer)



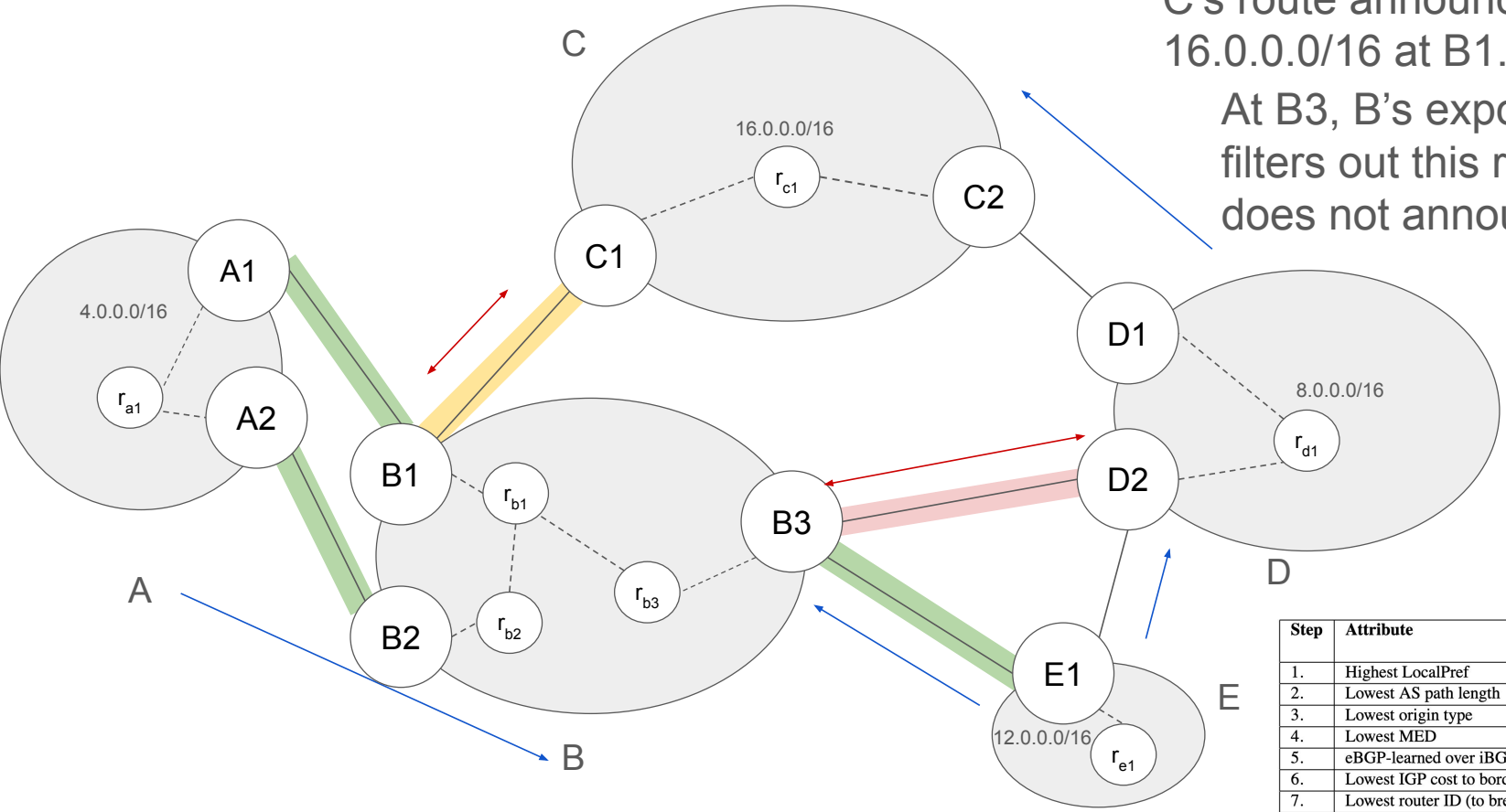
Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

Business relationships

eBGP
Internal link
Peer to Peer
Customer to Provider

B's import policy adds a peer-to-peer community tag to C's route announcement for 16.0.0.0/16 at B1.

At B3, B's export policy filters out this route and does not announce it to D

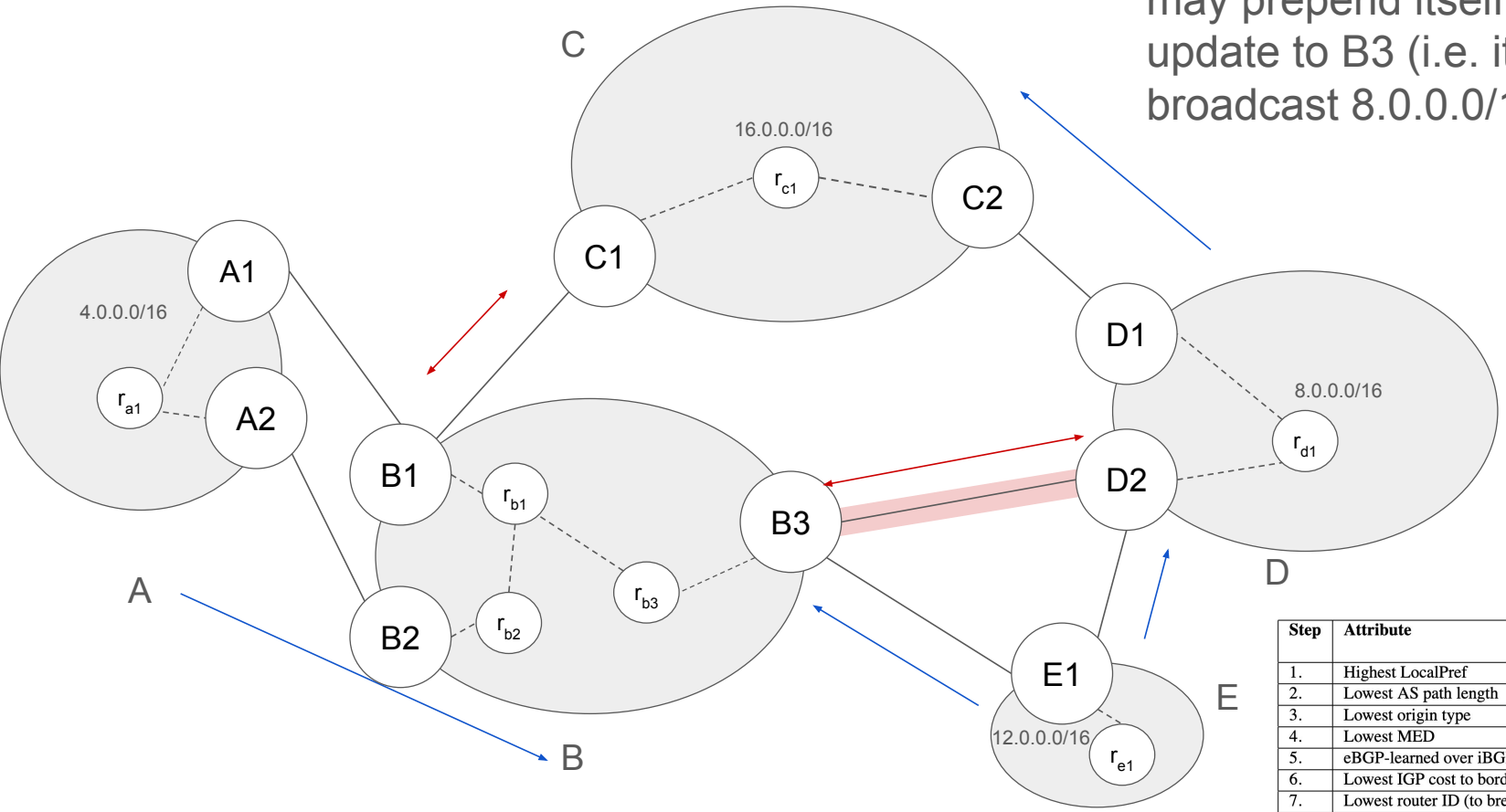


Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

Traffic Engineering

eBGP
Internal link
Peer to Peer
Customer to Provider

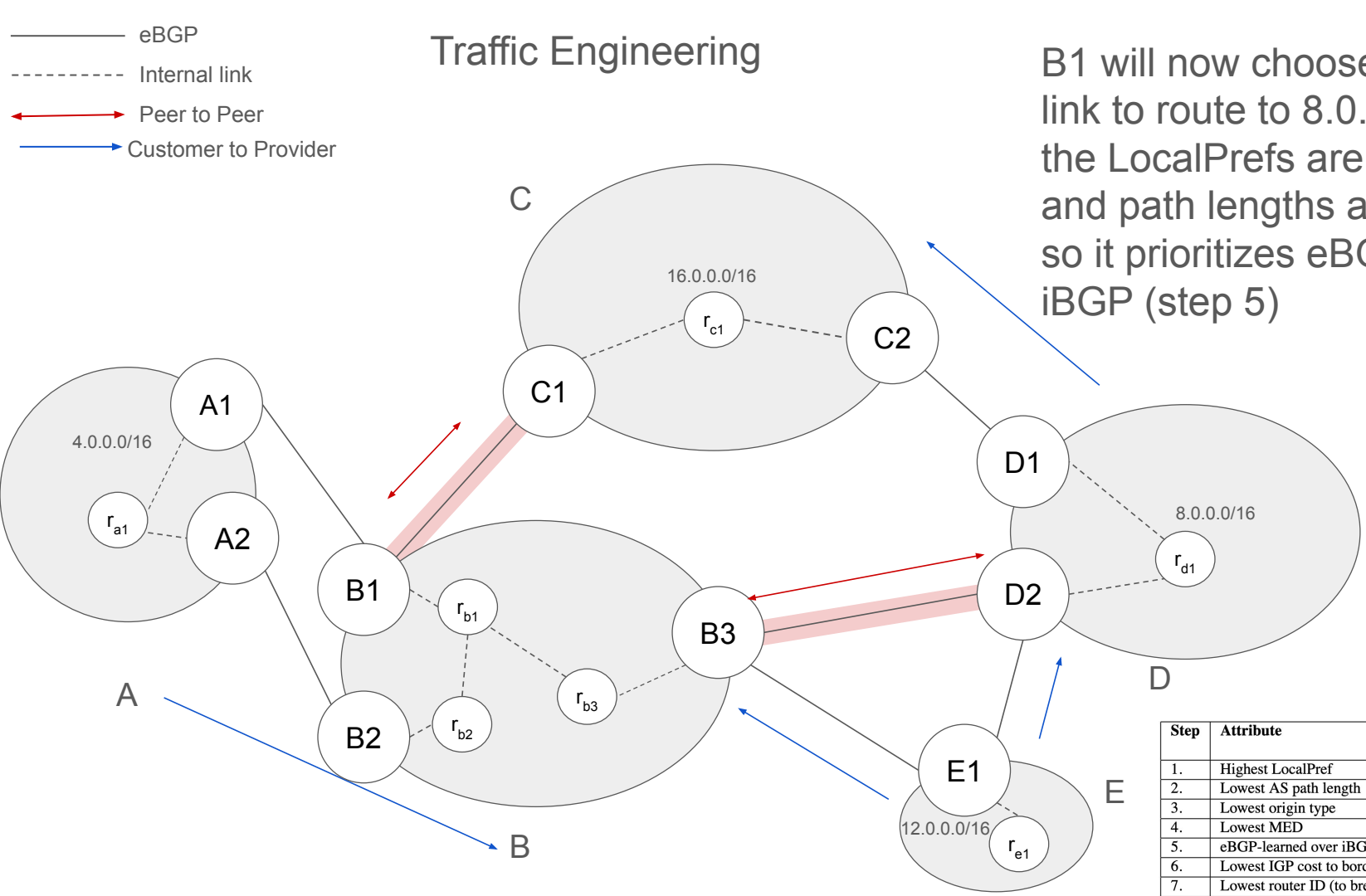
Say D needs to lower the traffix on the B3 - D2 link. Then, D may prepend itself on its update to B3 (i.e. it will broadcast 8.0.0.0/16 {D, D})



Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

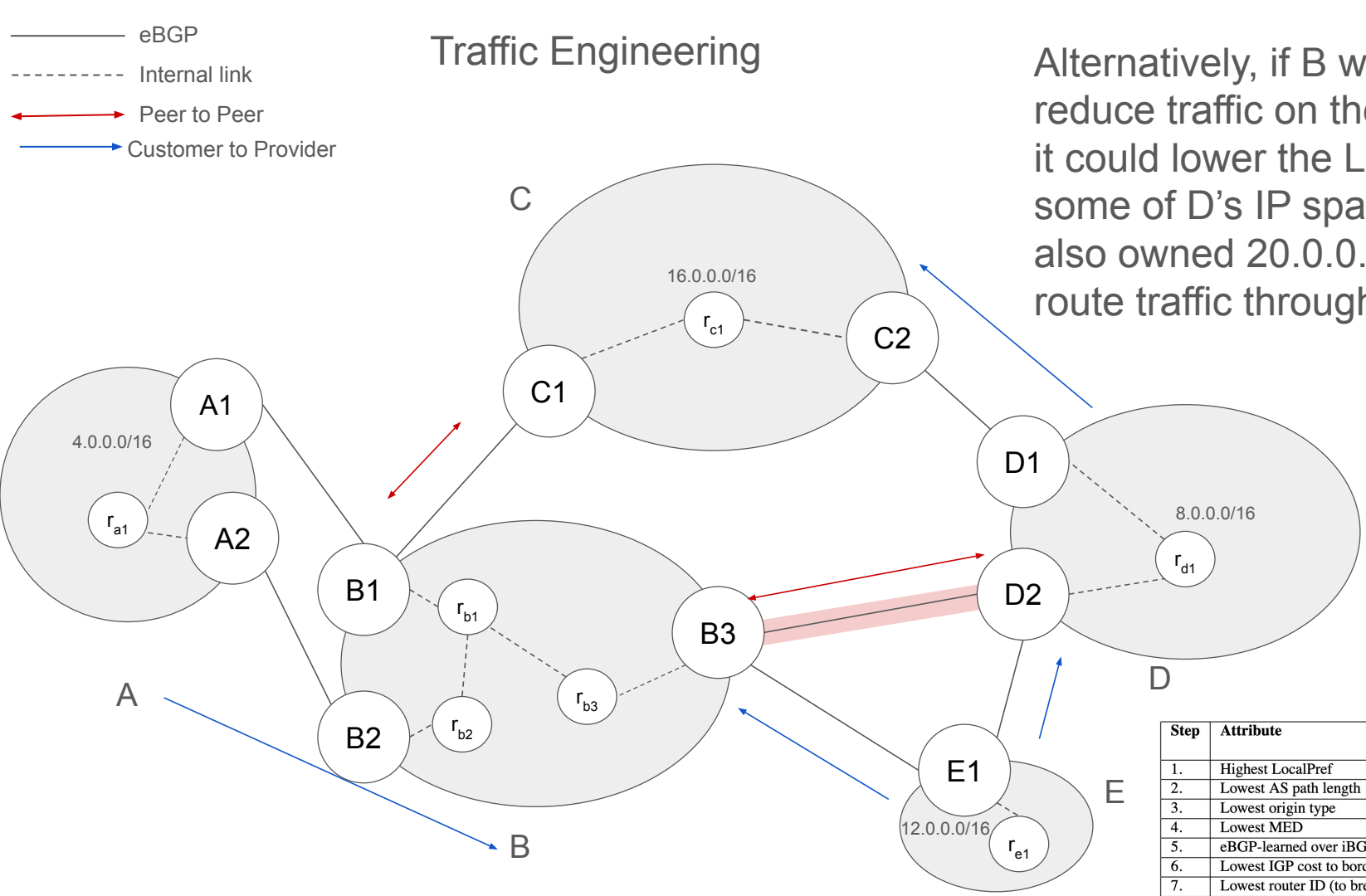
Traffic Engineering

B1 will now choose the B1 - C1 link to route to 8.0.0.0/16, as the LocalPrefs are equivalent, and path lengths are the same, so it prioritizes eBGP over iBGP (step 5)



Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

Traffic Engineering

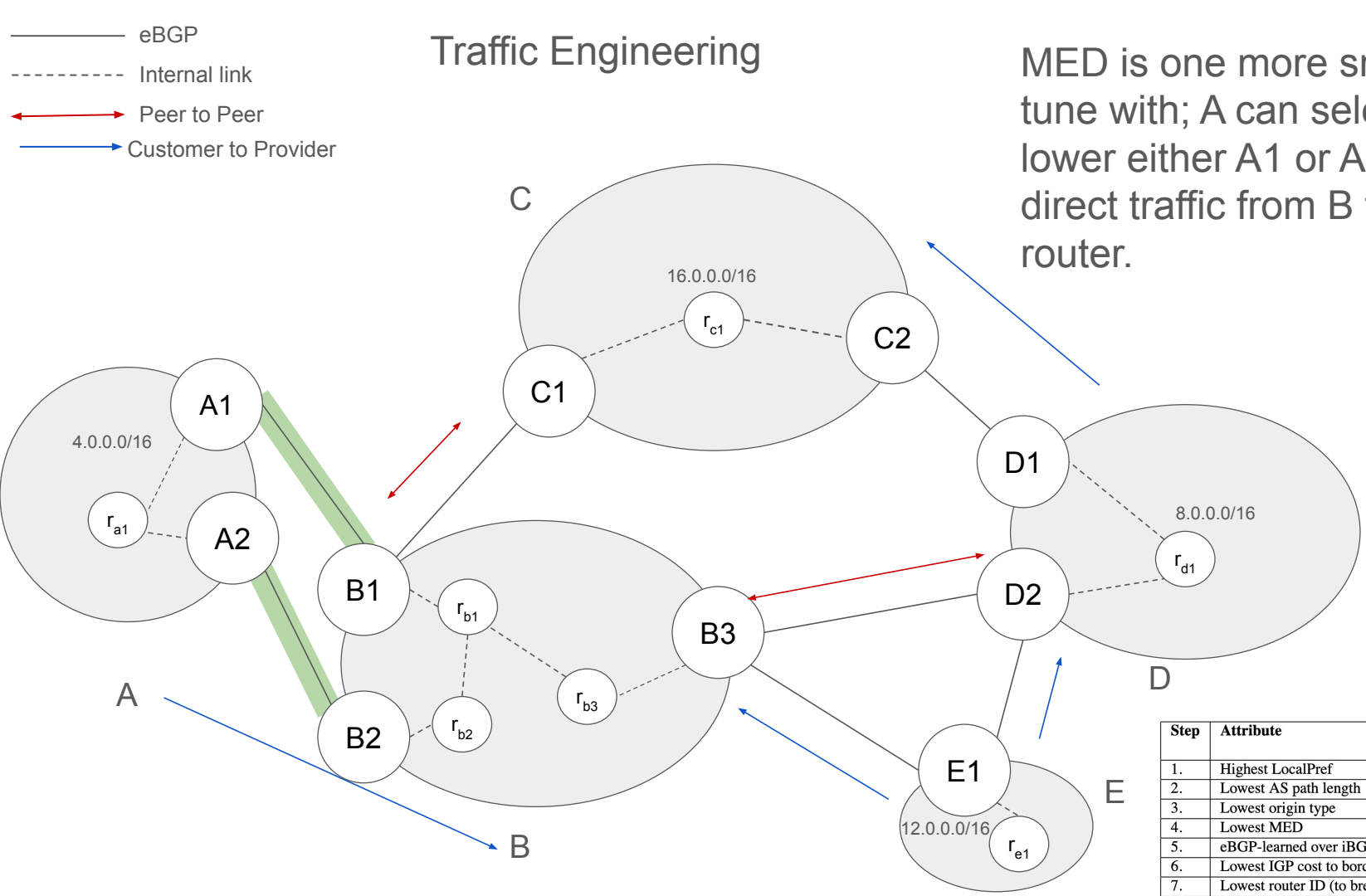


Alternatively, if B wanted to reduce traffic on the B3-D2 link, it could lower the LocalPref on some of D's IP space (e.g. if D also owned 20.0.0.0/16), to route traffic through C.

Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

Traffic Engineering

MED is one more small knob to tune with; A can selectively lower either A1 or A2's MED to direct traffic from B to a specific router.



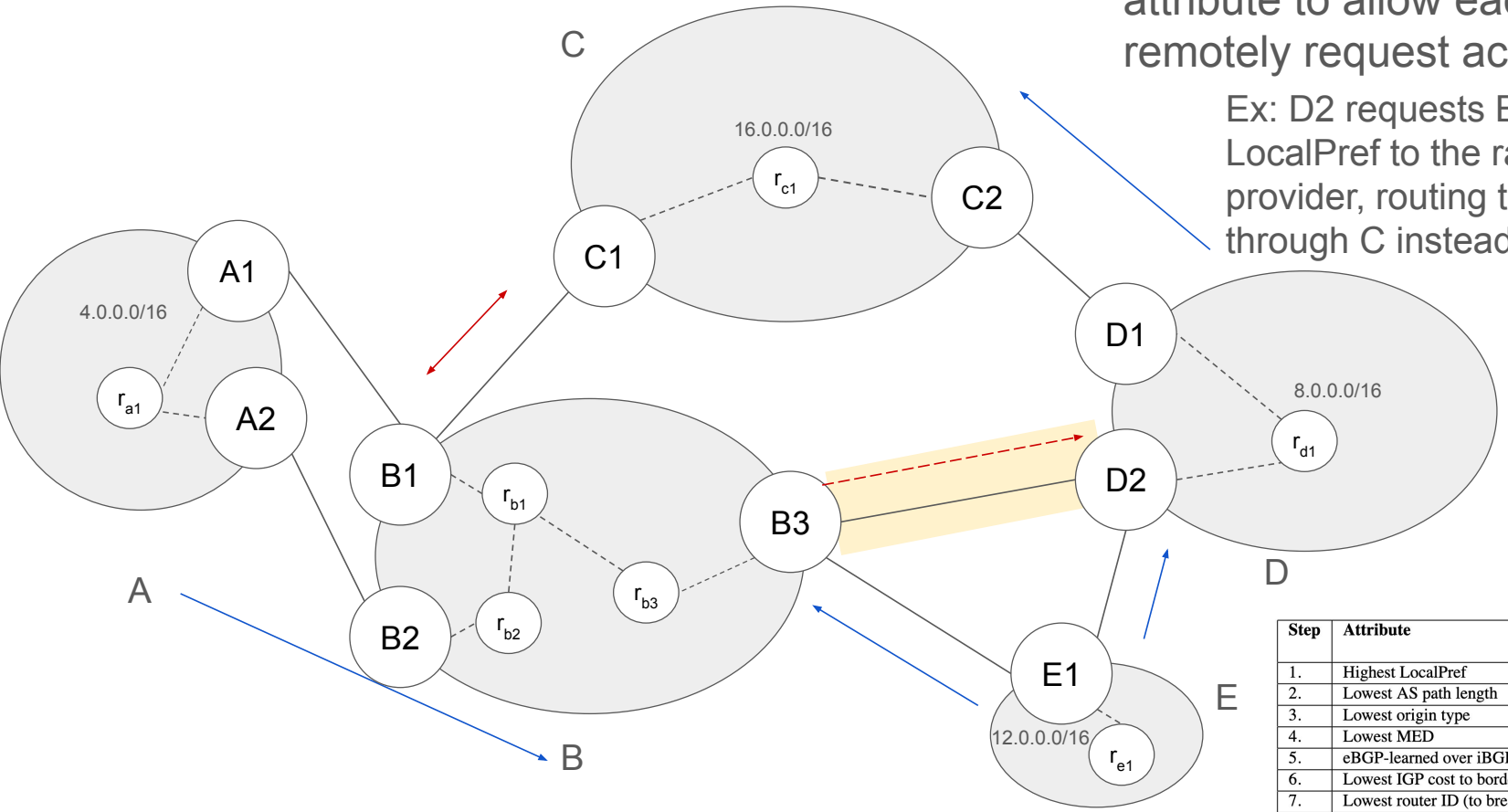
Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

Traffic Engineering

eBGP
Internal link
Peer to Peer
Customer to Provider

With prior agreement, ASes can use the Community attribute to allow each other to remotely request actions.

Ex: D2 requests B3 lower its LocalPref to the range of a provider, routing traffic through C instead



Step	Attribute	Controlled by local or neighbor AS?
1.	Highest LocalPref	local
2.	Lowest AS path length	neighbor
3.	Lowest origin type	neither
4.	Lowest MED	neighbor
5.	eBGP-learned over iBGP-learned	neither
6.	Lowest IGP cost to border router	local
7.	Lowest router ID (to break ties)	neither

Scalability

To support scalability, regardless of cause:

- Aggregate routes outgoing routes, limit specificity of incoming route prefix
- Limit the number of prefixes accepted from a neighbor during one BGP session
- Implement flap damping - have a counter for each time a route is updated, and exclude a route from the decision process if the counter gets too high

Security

Specifically against malicious, non-misconfigured ASes or routers:

- Filter prefixes offered by customers to only accept or export ones that they correctly own
- RPKI

Using BGP to protect against DoS attacks:

- Set the next hop for a victim's IP addresses to a null route that drops all packets, or to a monitoring system to study attacker behavior

Using BGP to protect against spam:

- Set the next hop for an attacker's IP addresses to a null route to prevent two-way communication

Security protocols - RPKI

- Regional Internet Registries issue certificates to organizations, connecting their owned IP prefixes to public keys
- Prefix owners sign Route Origin Authorizations, certifying certain AS numbers as valid originators of prefix routes (up to a configurable max prefix length)

Doesn’t protect against malicious black hole - ing / setting bad next hops; the only check is that the originator AS is valid for the given prefix and prefix length

Security protocols - BGPsec

- An extension of RPKI - instead of just the originator AS signing the prefix, each AS signs a hash of the prefix, its own AS number, the recipient's AS number, and the previous signature
- Receiving routers verify each signature in the list
- Not implemented across the internet …
  - Cryptography is computationally expensive, especially for routers
  - If a single AS in the chain doesn't have BGPsec capability, the path becomes unsigned from that point onwards

Hybrid Link-state Path-vector (HLP)

- Hierarchical - top-level providers are partitioned with all of their customers into groups
  - Customers of multiple top-level providers can belong to multiple groups
- Each group distributes all the information about the ASes contained within for faster route-finding (iterative propagation not necessary)
- Between groups, path vectors are formed similar to BGP

Why is this specifically not adopted?

- No LocalPref equivalent
- Less privacy; all members of a group learn about all other customers of that top-level provider

Discussion

- The knobs to control incoming traffic (MED, route prepending) are below the sender's LocalPref in the decision process. Are there any receiver controls that should be prioritized above it?
- More generally, the paper advocates for the fixed decision process for making predictable decisions. Would a programmable decision process's benefits outweigh the loss of this predictability?