APT: An Architecture for Practical Transit Core Separation

Michael Meisel

work with Dan Jen, Dan Massey, Lan Wang, Beichuan Zhang, and Lixia Zhang

APT is a Map & Encap Scheme

- Map & encap is one category of separation scheme
- Encapsulation is used in the transit core
- Needs a mapping service
- (LISP is a map & encap scheme, too)

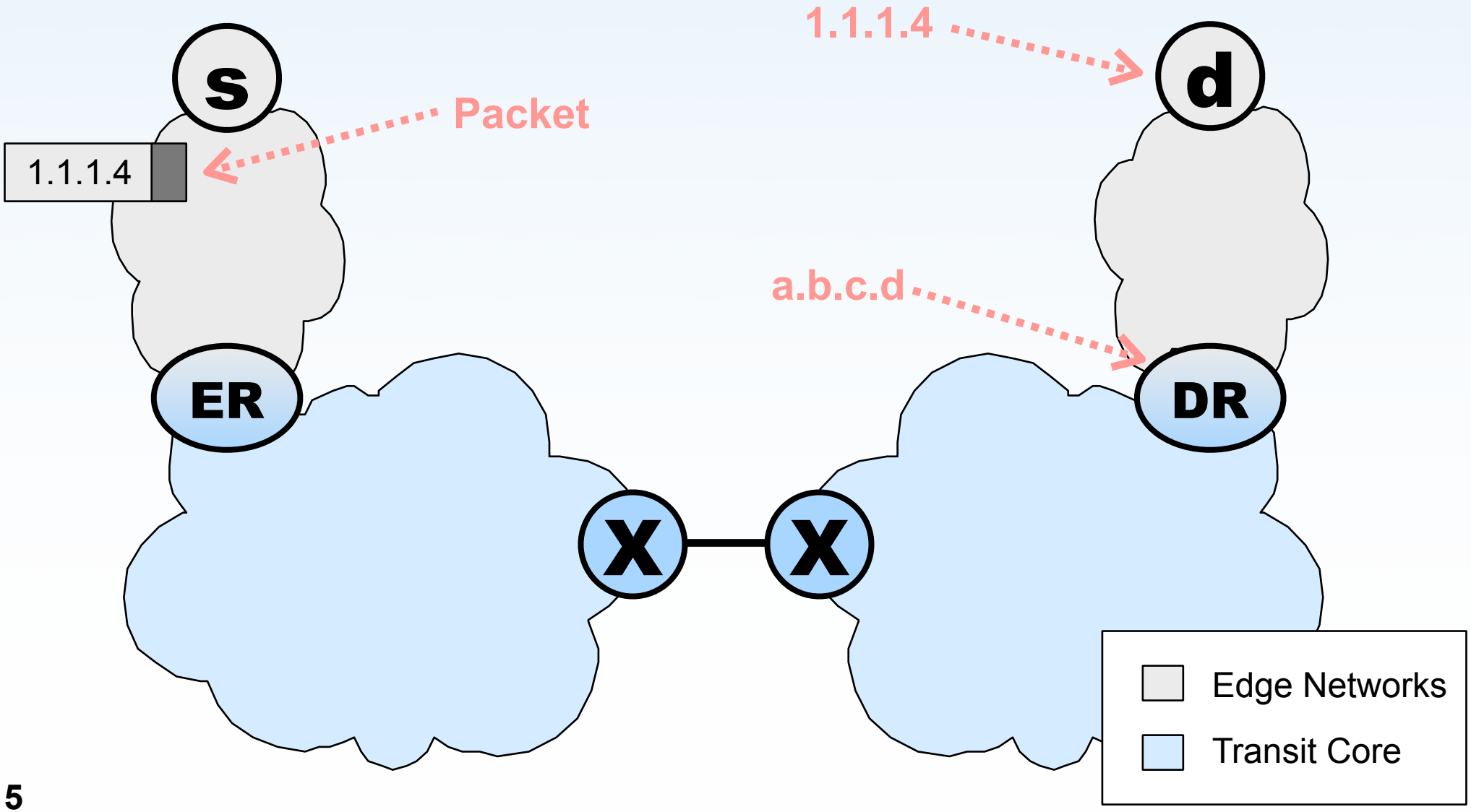
Encap and Decap at Border Routers

- Encapsulating Routers (ERs)
  - Use mappings to encapsulate packets
  - Caches recently used mappings
- Decapsulating Routers (DRs)
  - Strip encapsulation header
- Generally both functions in one device (EDR)

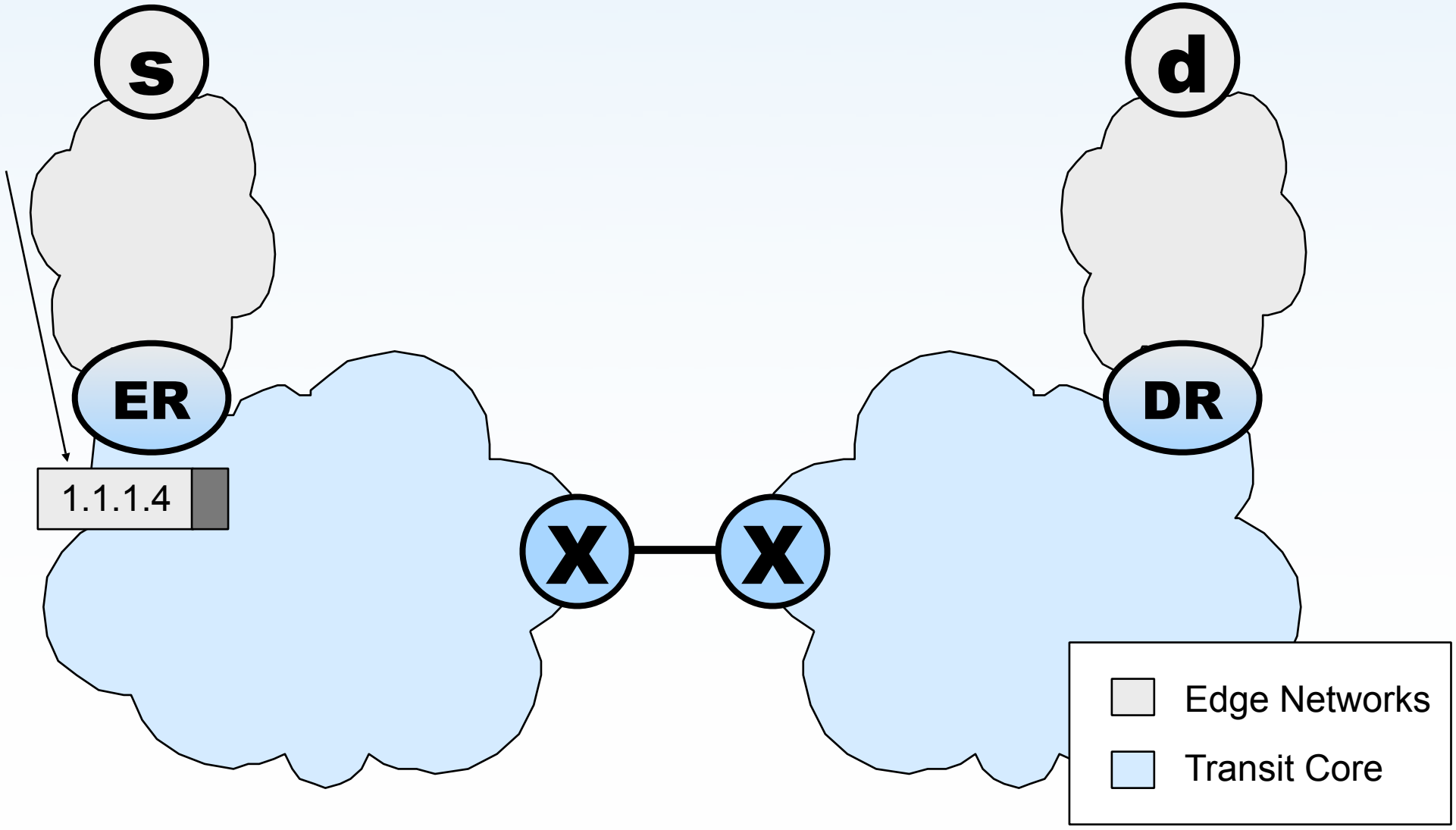
What's a Mapping?

- Maps an edge network's prefix(es) to its transit core attachment points
  - Attachment points are providers' border routers
  - This is a one-to-many mapping per prefix
- No reachability info, just topological info

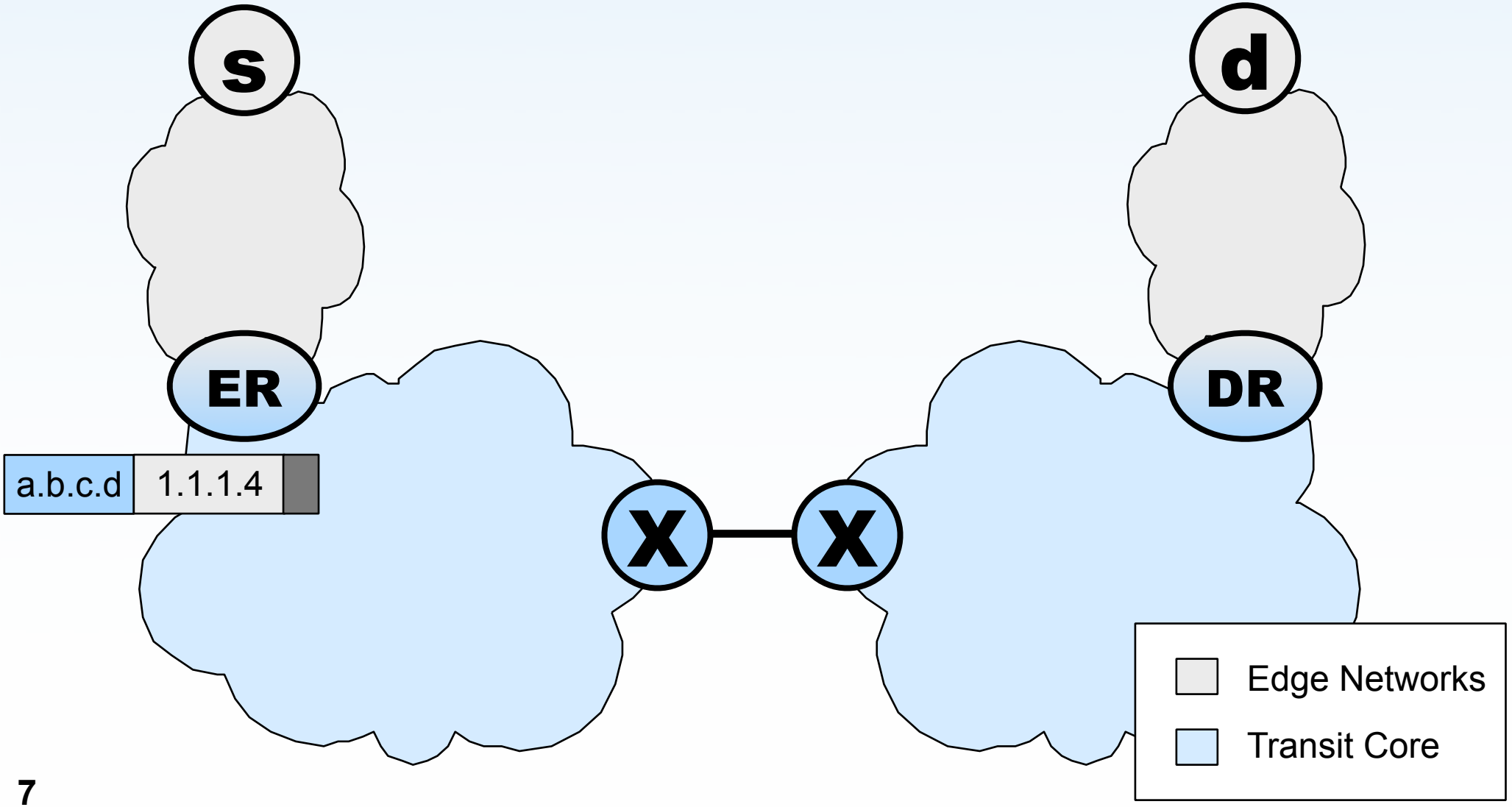
Map & Encap Example



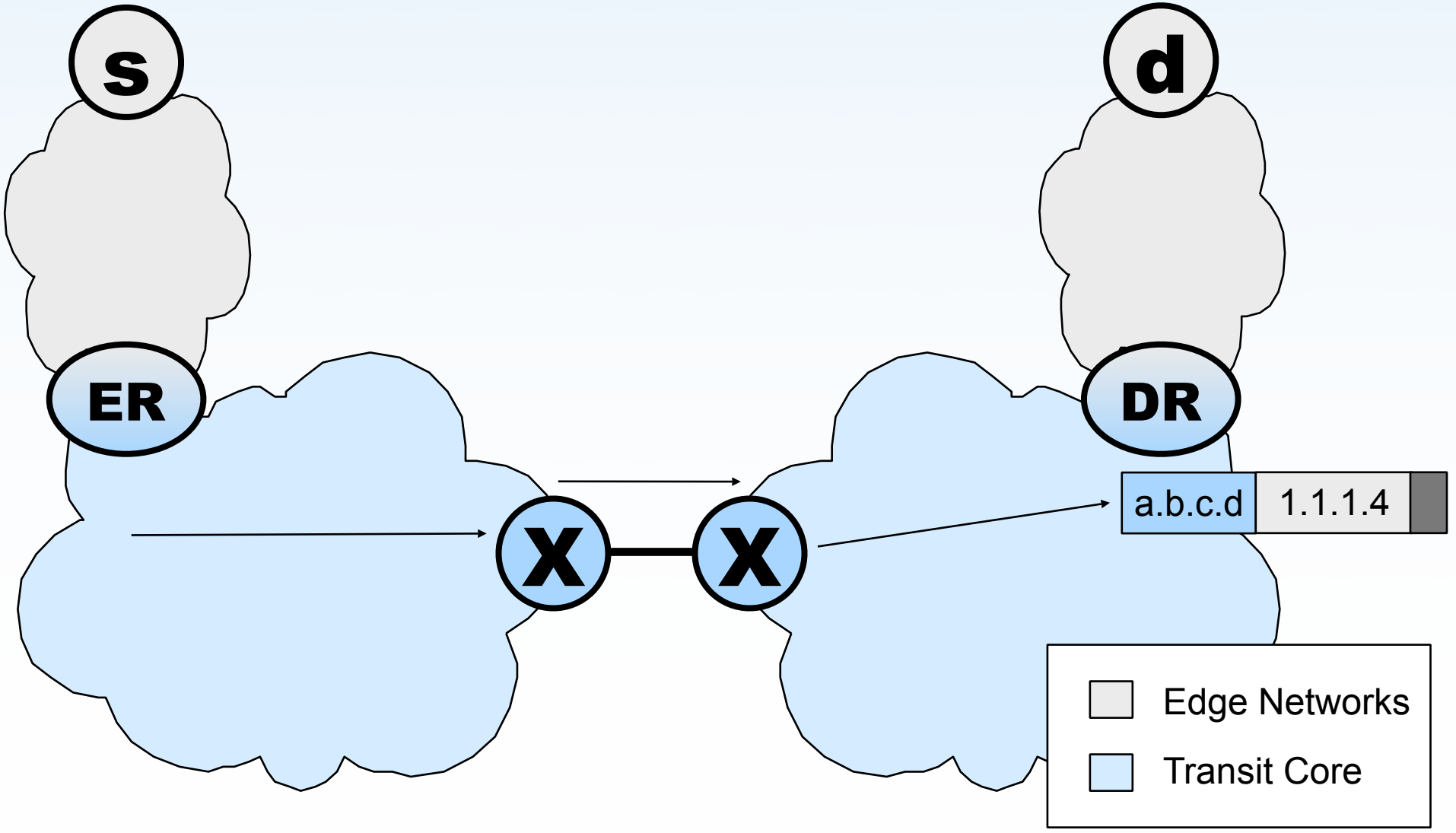
Packet Arrives at ISP



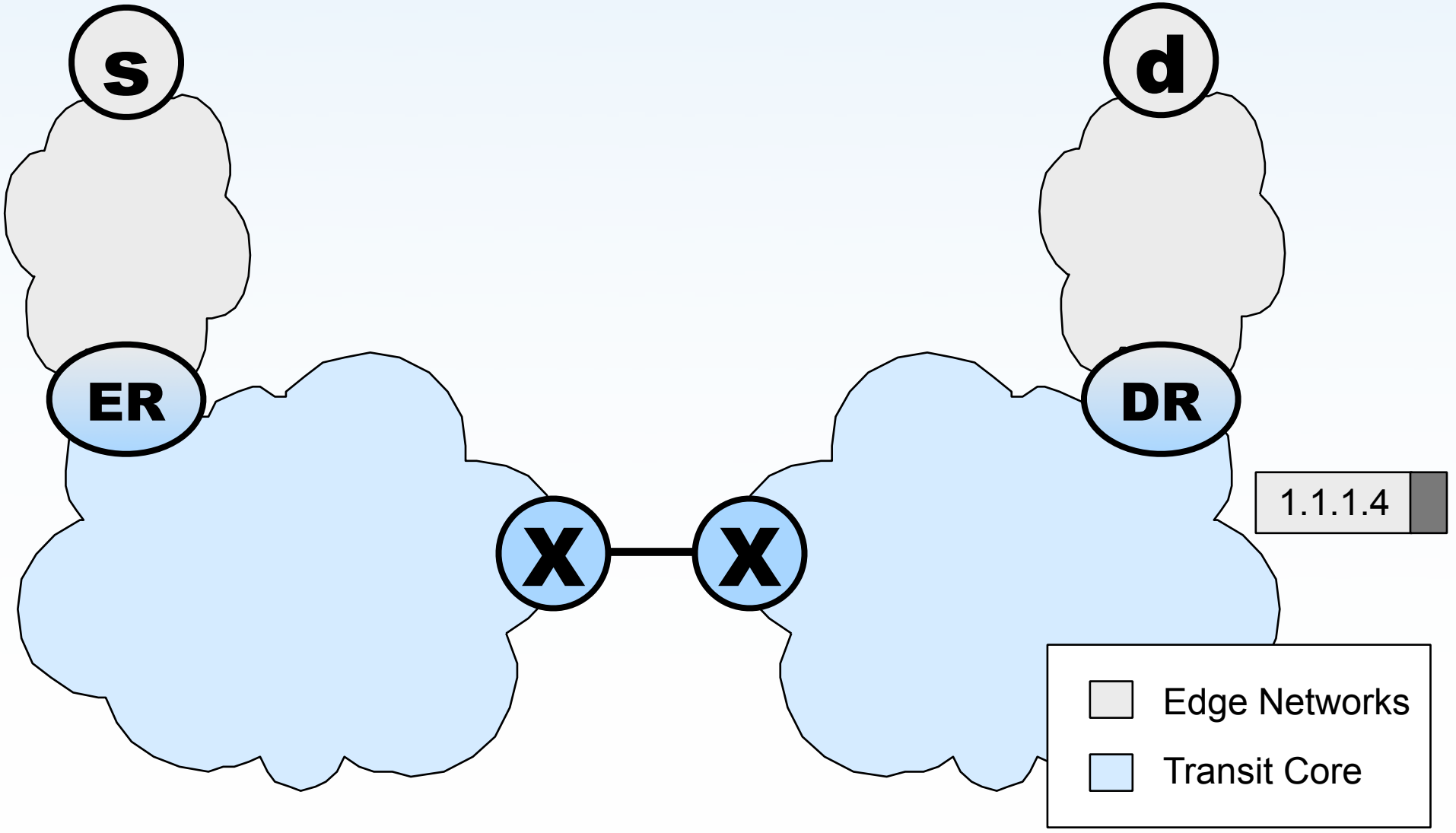
Packet Encapsulated in Transit Core Header



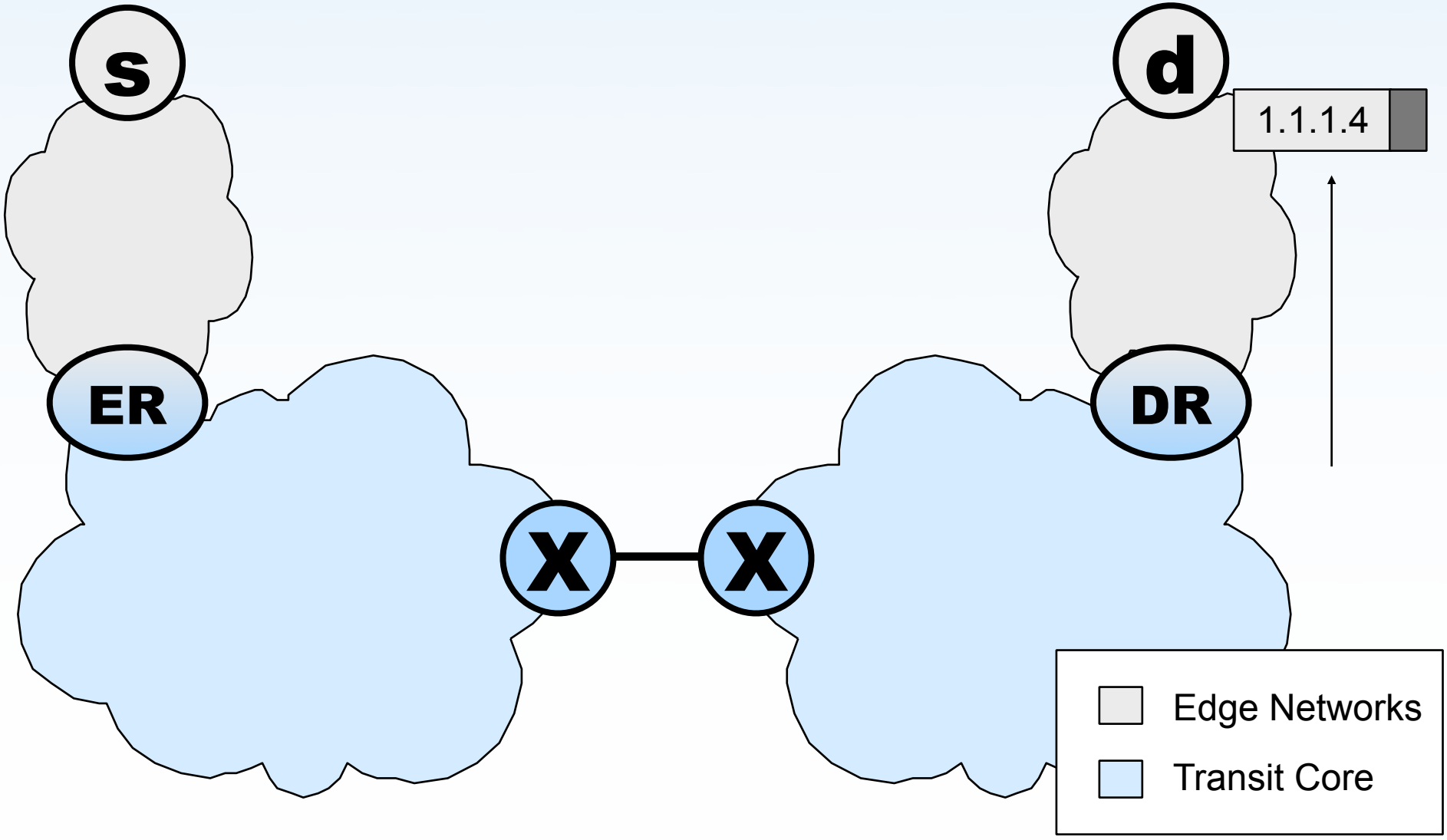
Packet Delivered across Transit Core



Packet Decapsulated



Packet Delivered

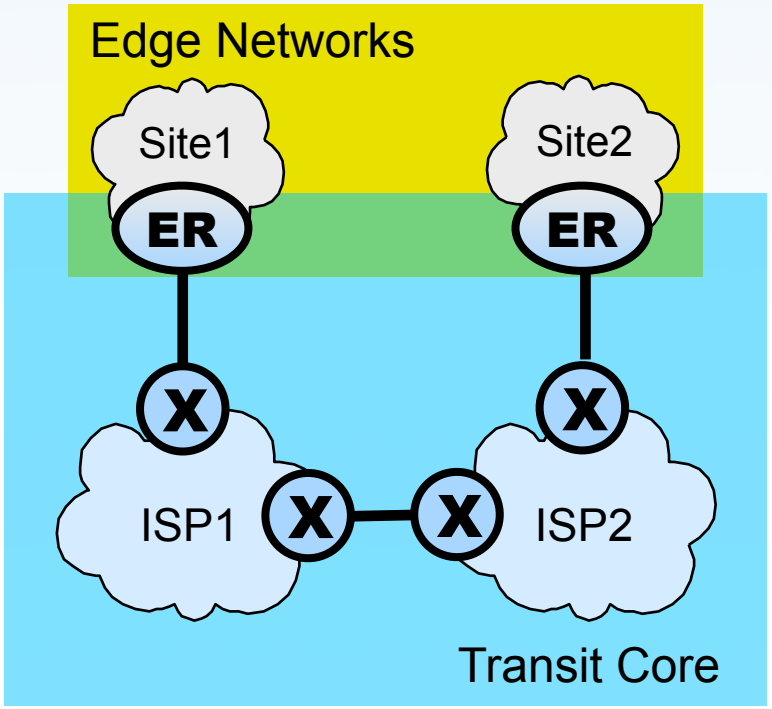


The APT Philosophy: “Do No Harm”

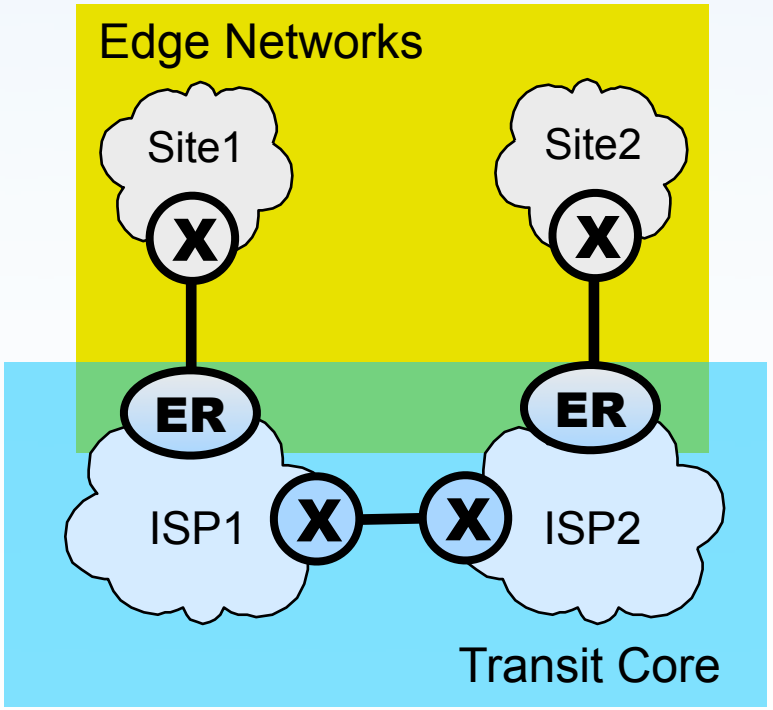
- Make no changes at edge sites
- Maintain current network performance
- Minimize the amount of new infrastructure
- End result
  - ISPs can deploy APT unilaterally
  - Cost is aligned with benefit

Make No Changes at Edge Sites

- LISP encapsulates packets at edge sites
- APT encapsulates packets at ISPs



Encap at edge (LISP)

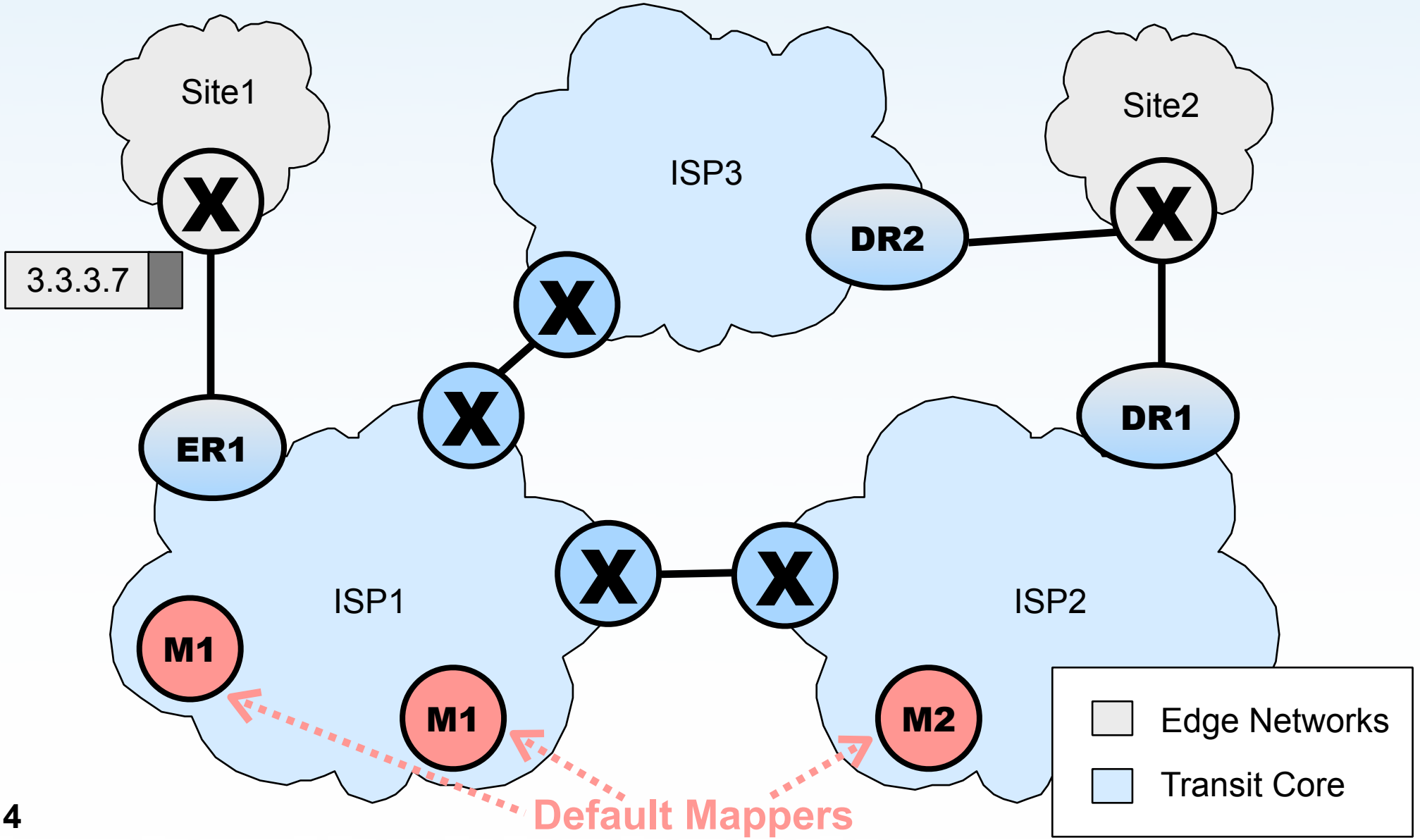


Encap in core (APT)

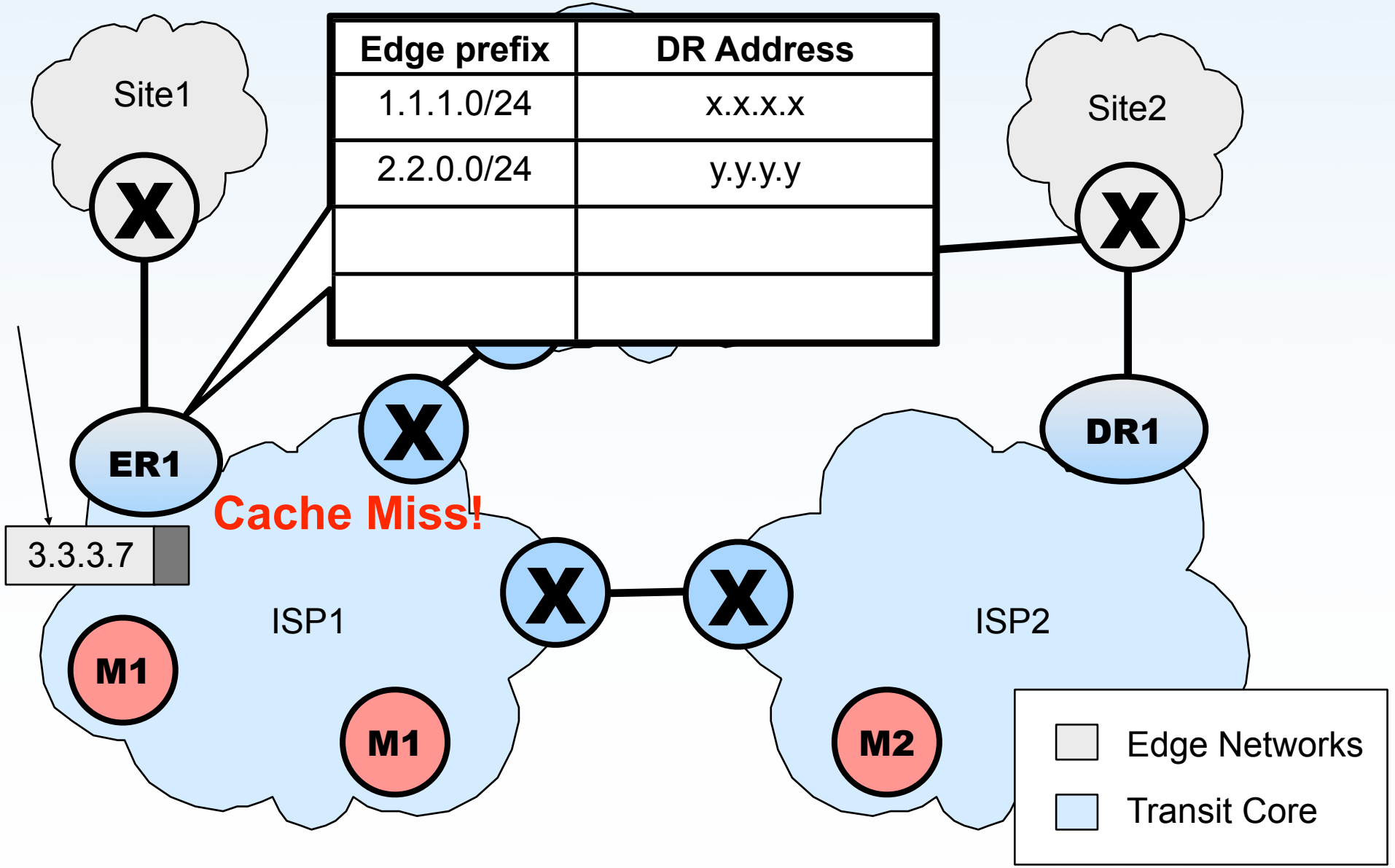
Maintaing Current Network Performance with Default Mappers

- Default mappers store the full mapping table
- Each APT ISP has at least one
  - But no more than a few
- ERs cache only recently used mappings
  - Cache miss? Ask your DM
  - All mapping lookups occur intra-domain

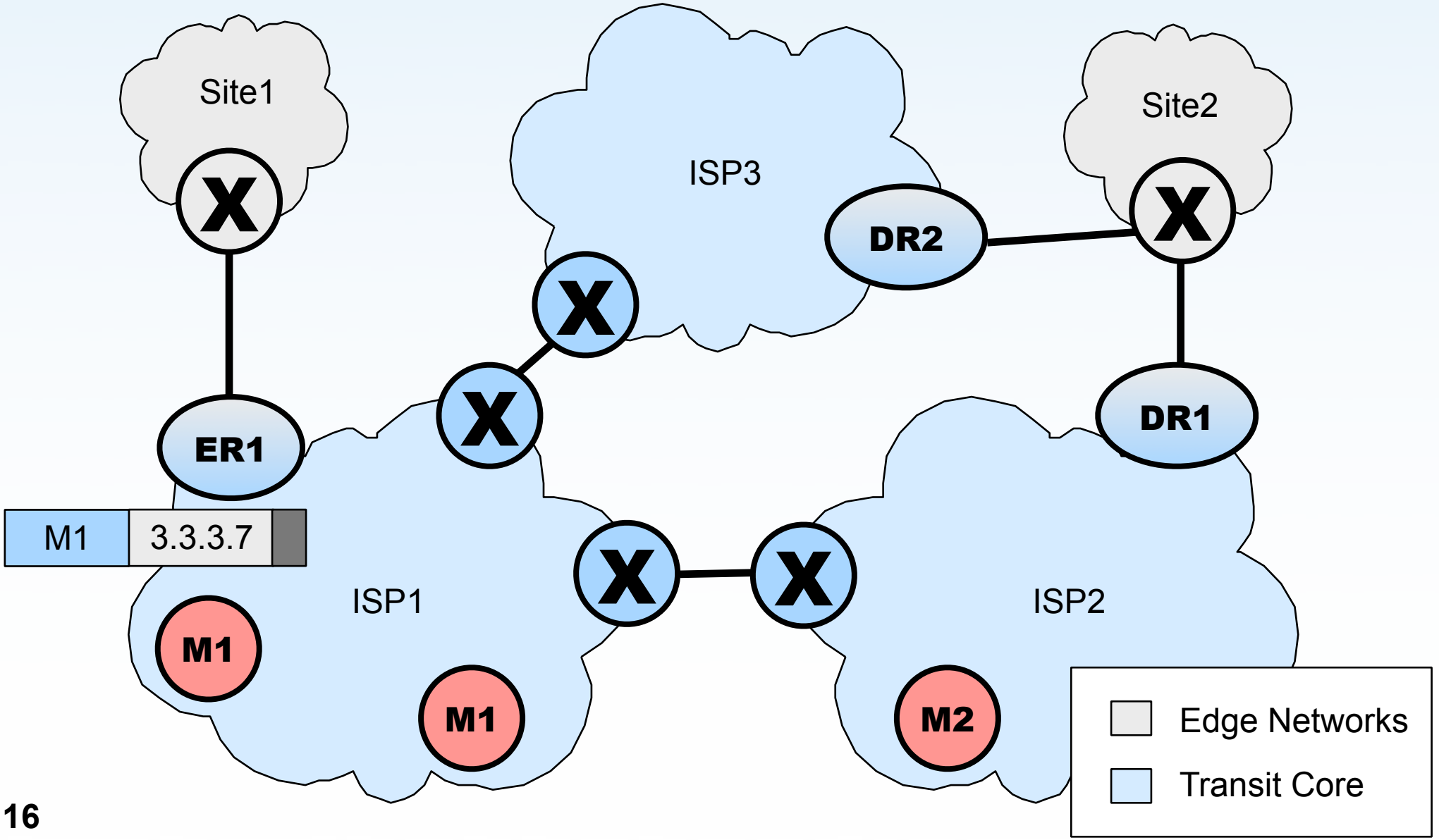
APT Example


Site1
ISP3
Site2
3.3.3.7
DR2
ER1
DR1
ISP1
ISP2
M1
M1
M2
Default Mappers
Edge Networks
Transit Core

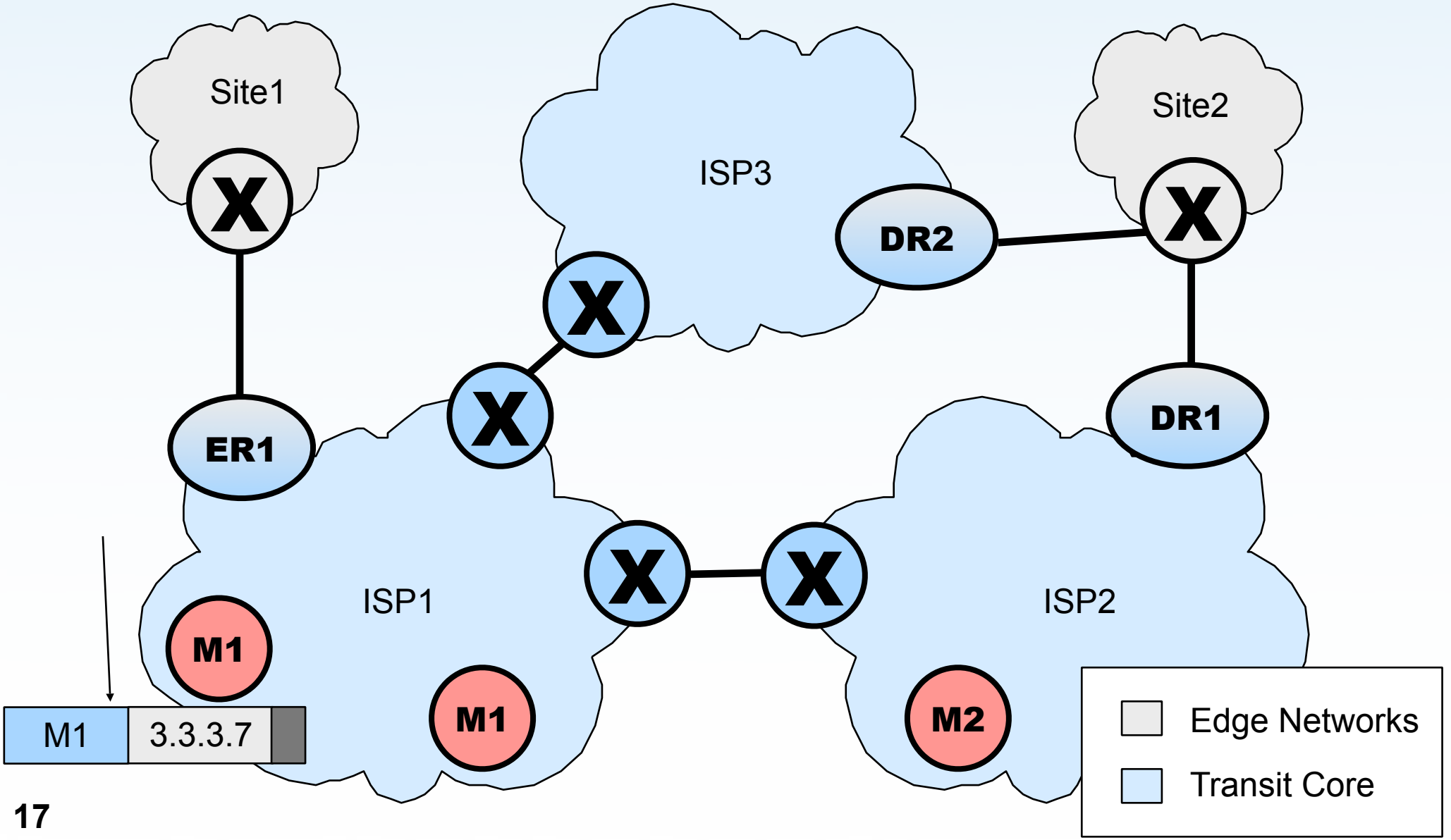
Mapping Not in Cache



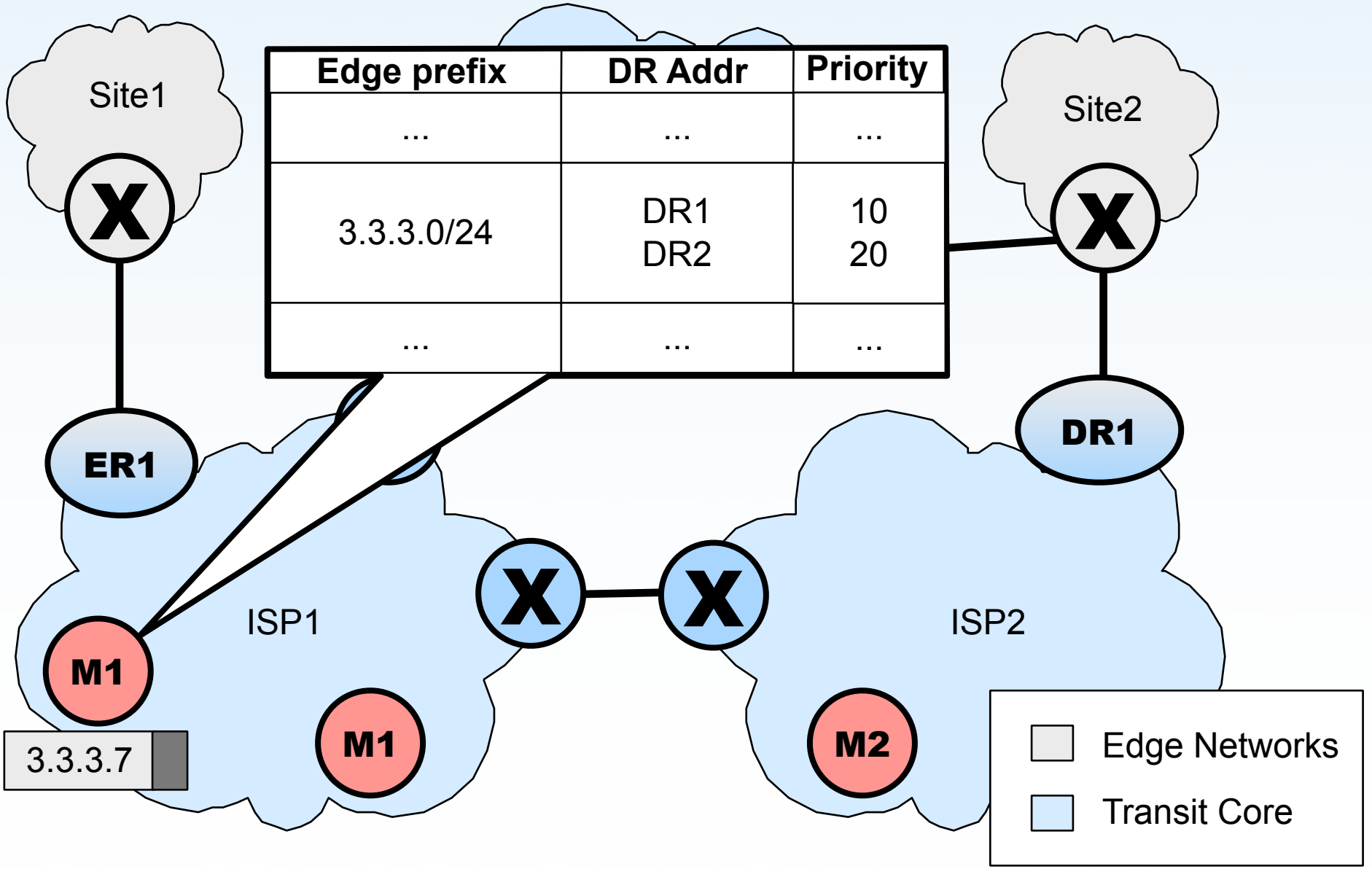
Encap with the Default Mapper Anycast Address



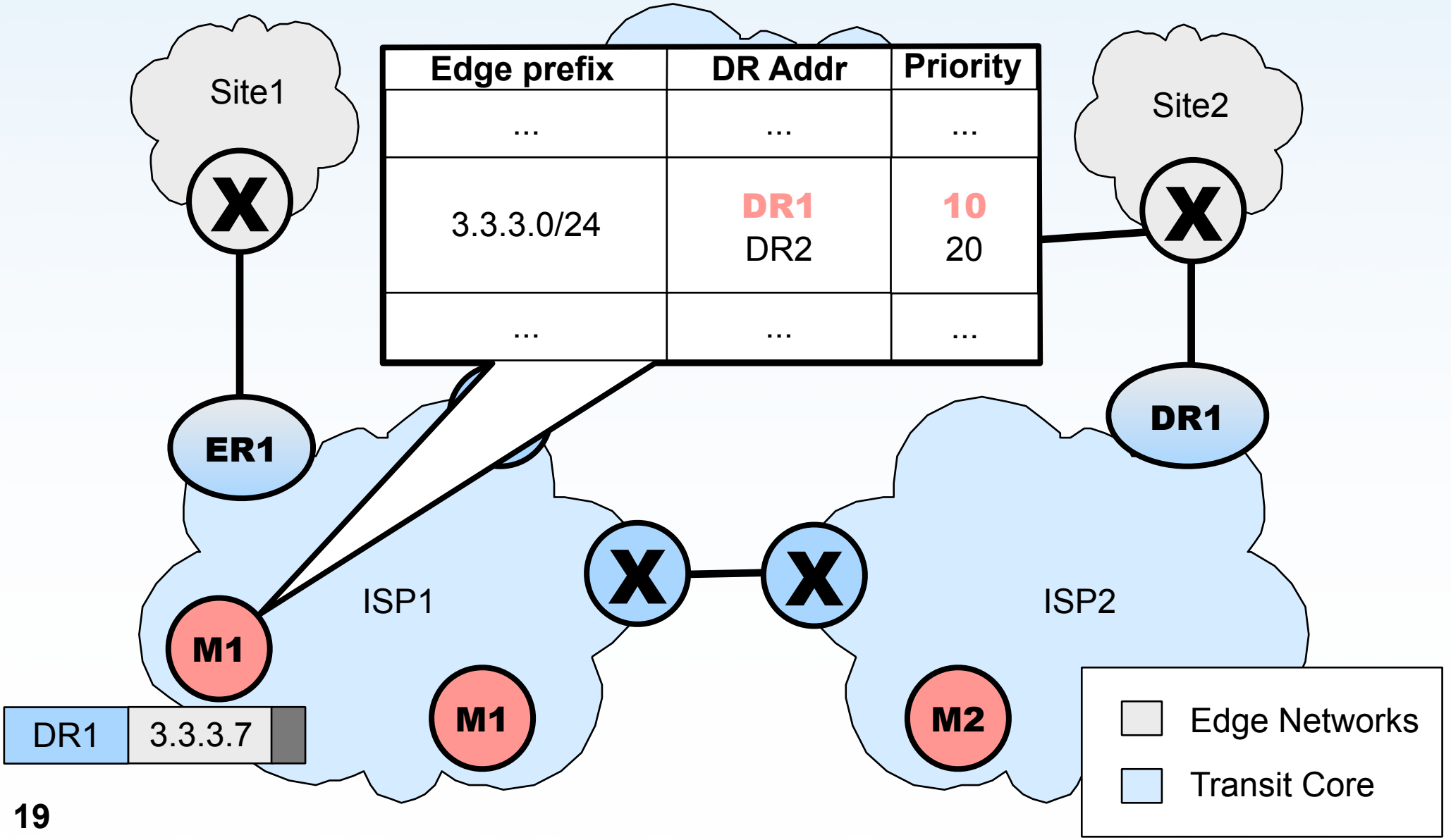
Default Mapper Decaps the Packet



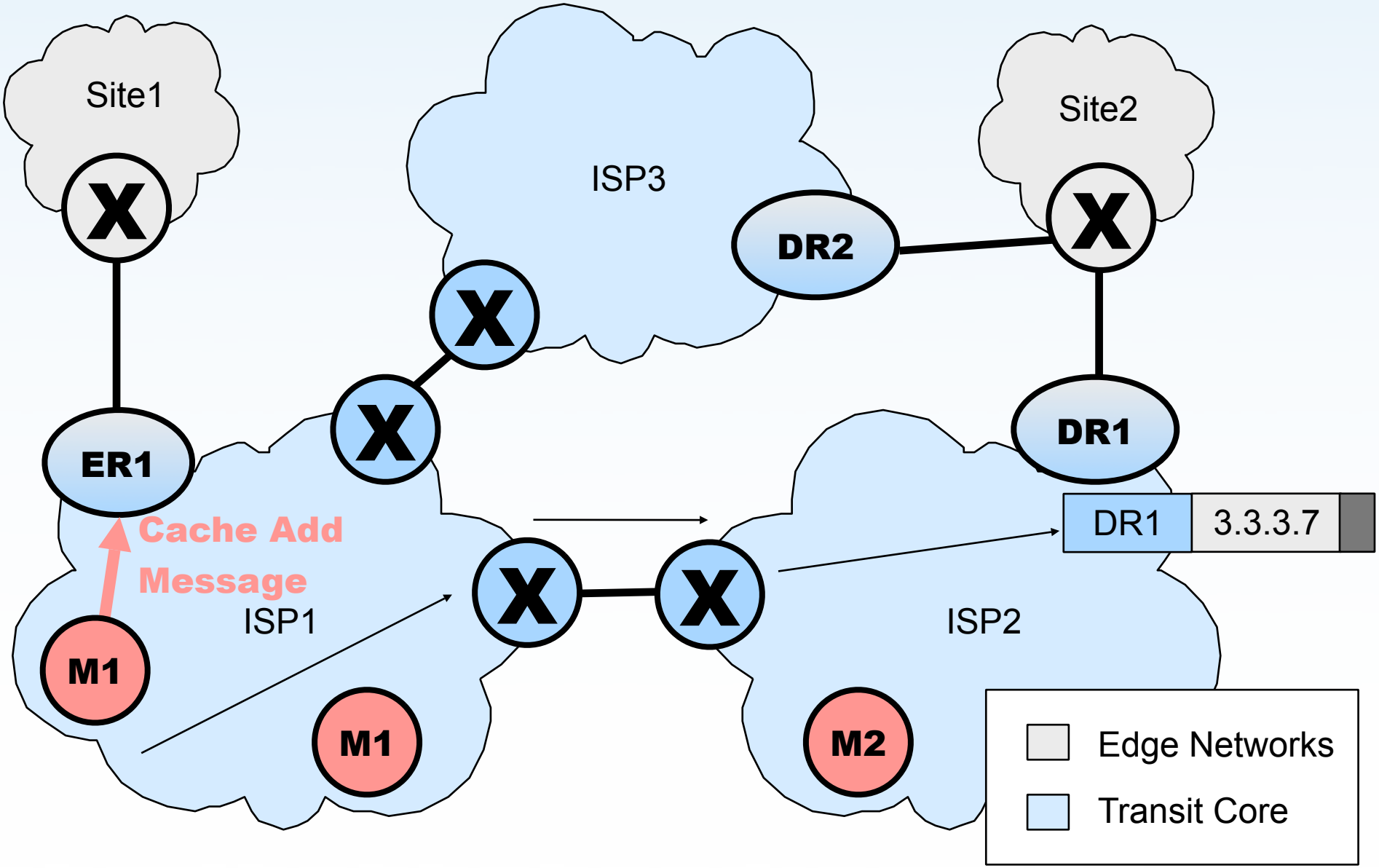
Edge prefix is Multihomed



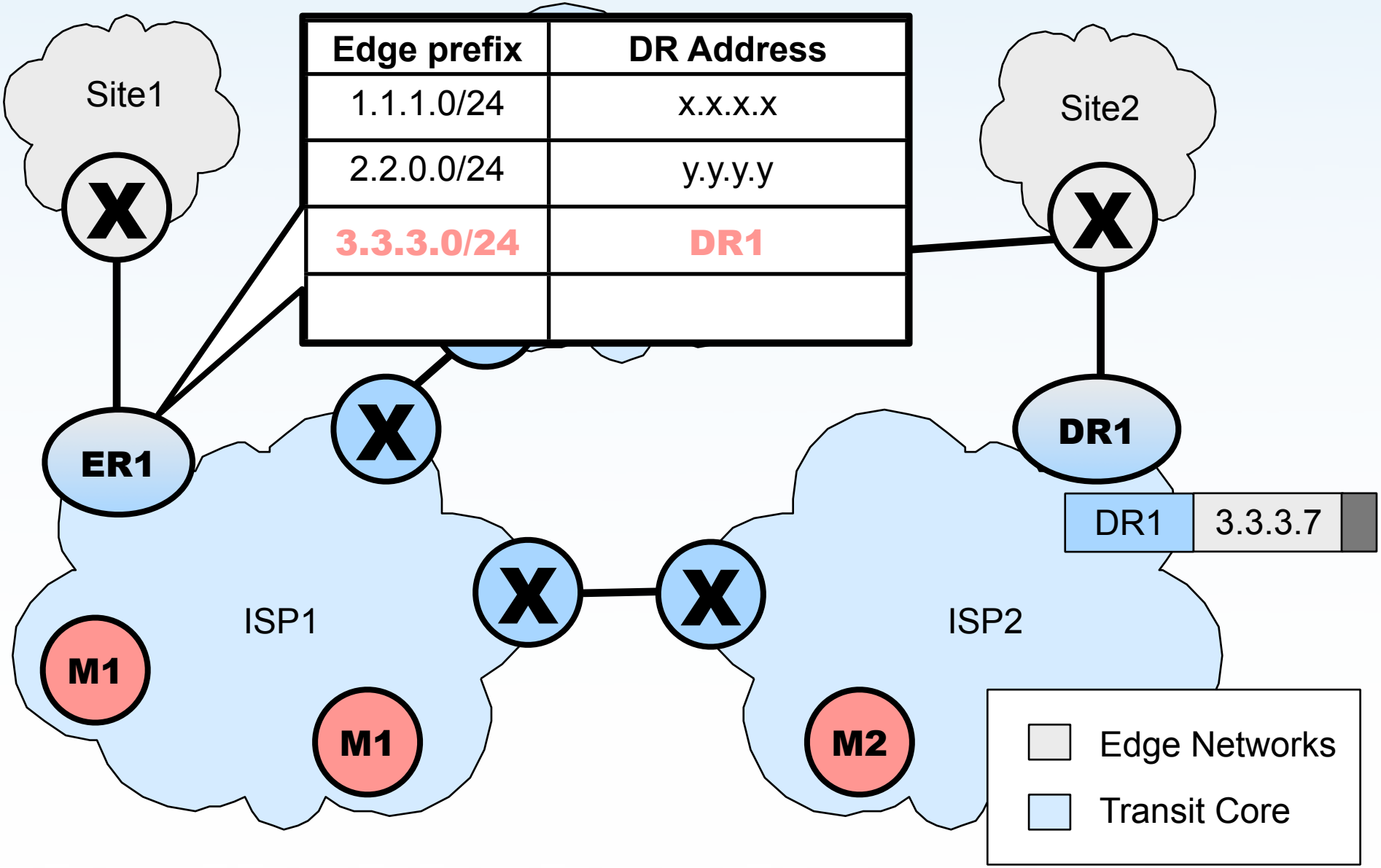
Default Mapper Selects a Mapping



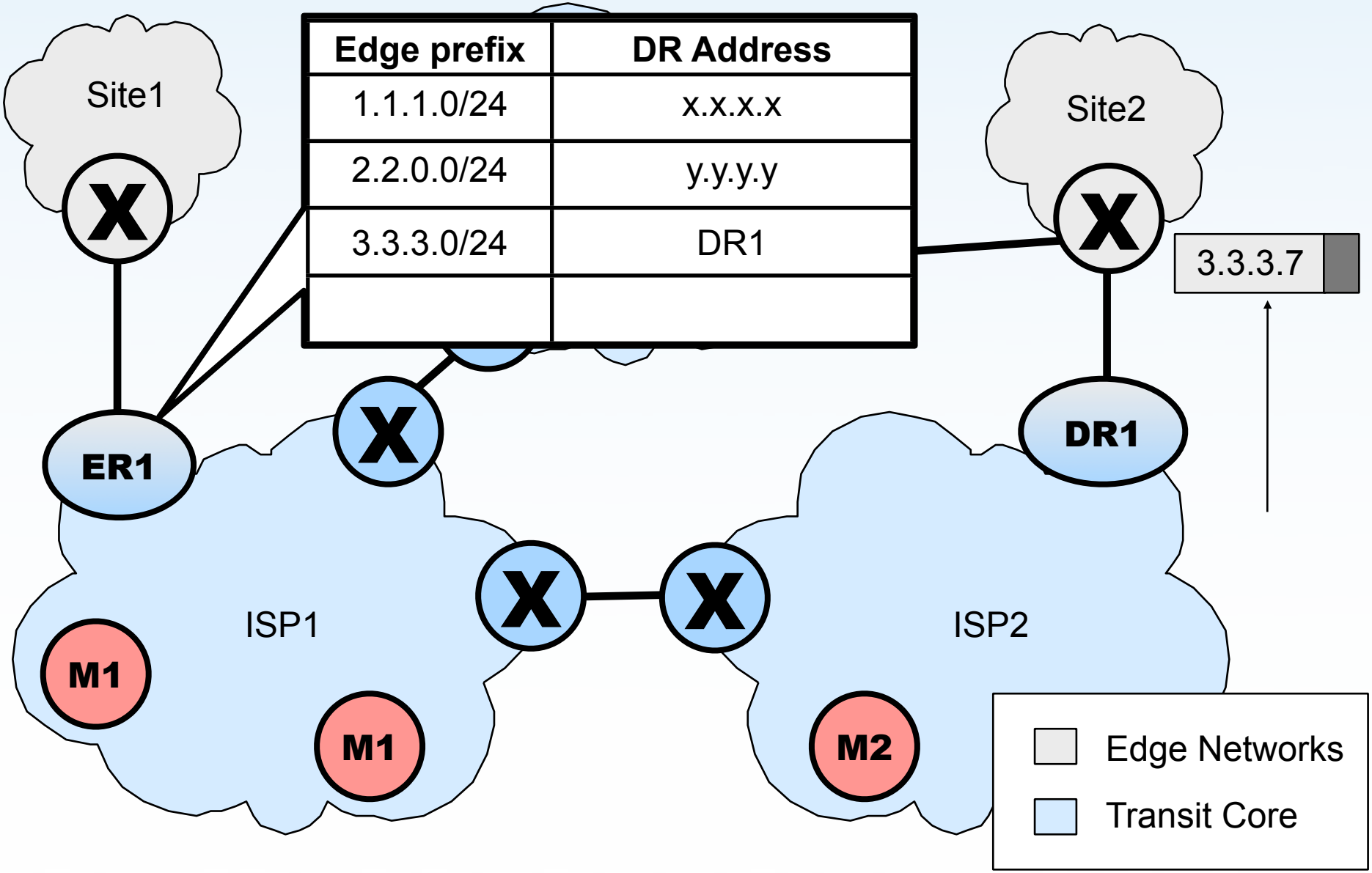
Default Mapper Responds with Mapping and Delivers Packet



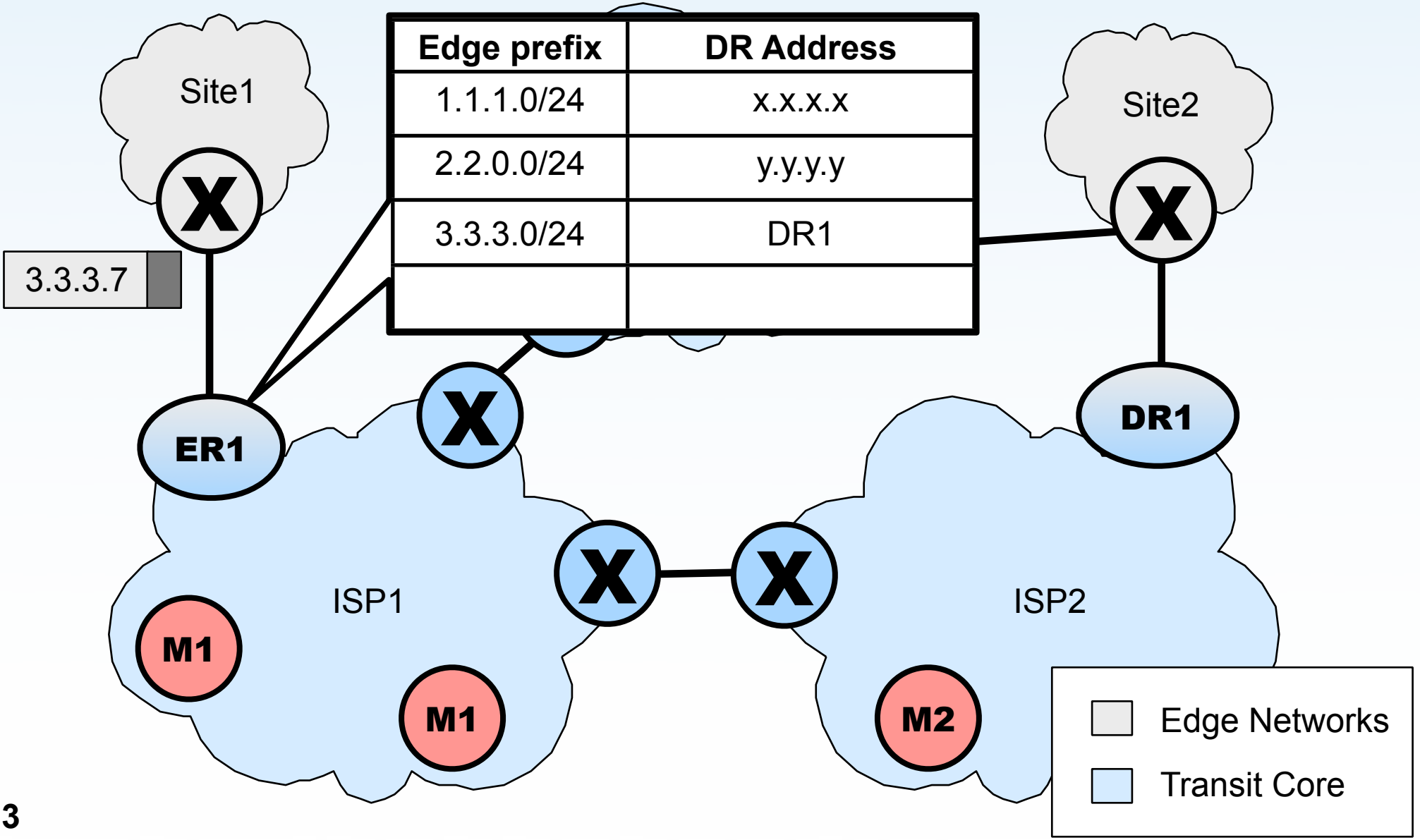
Mapping Added to Cache



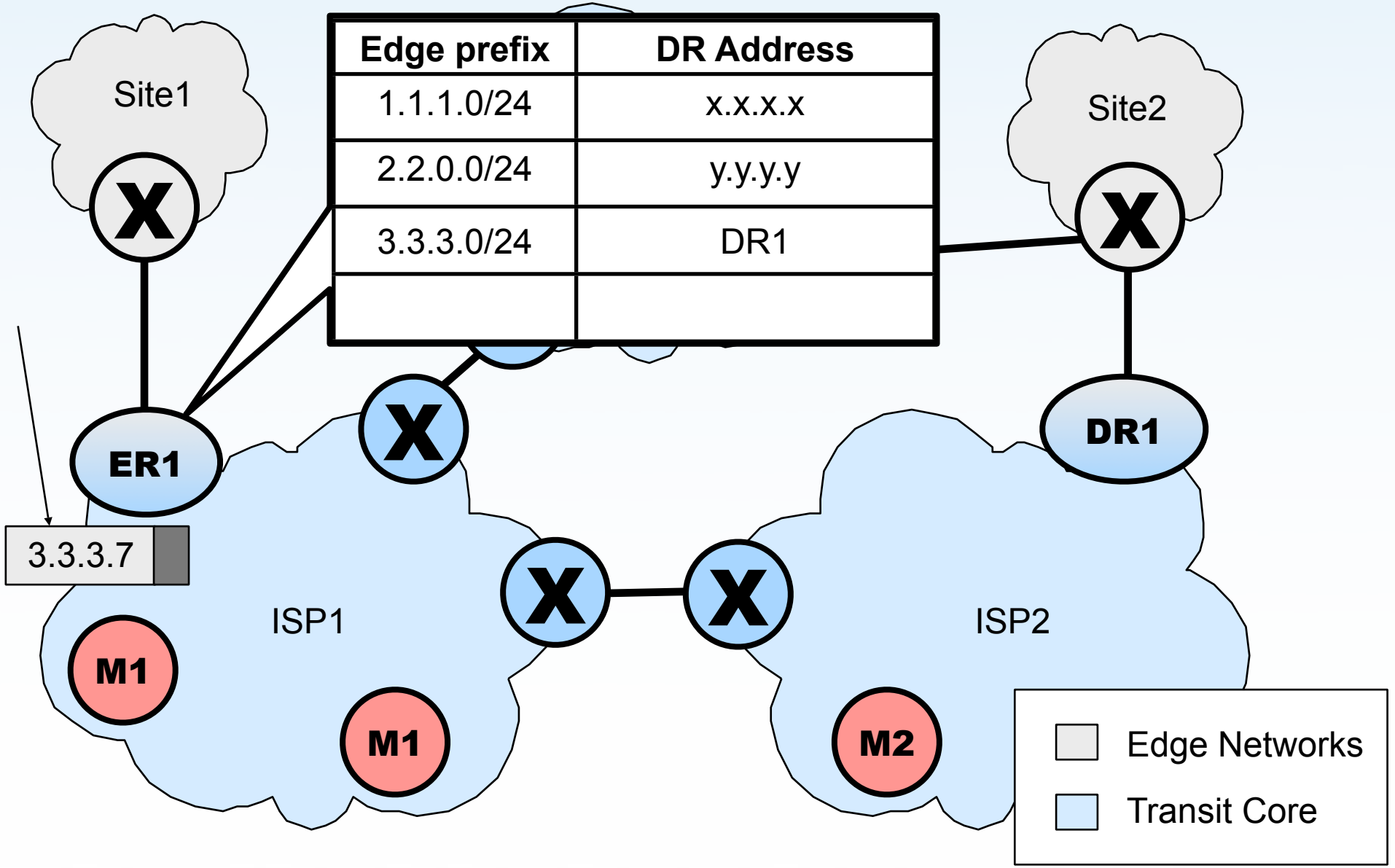
Packet Decapsulated and Delivered



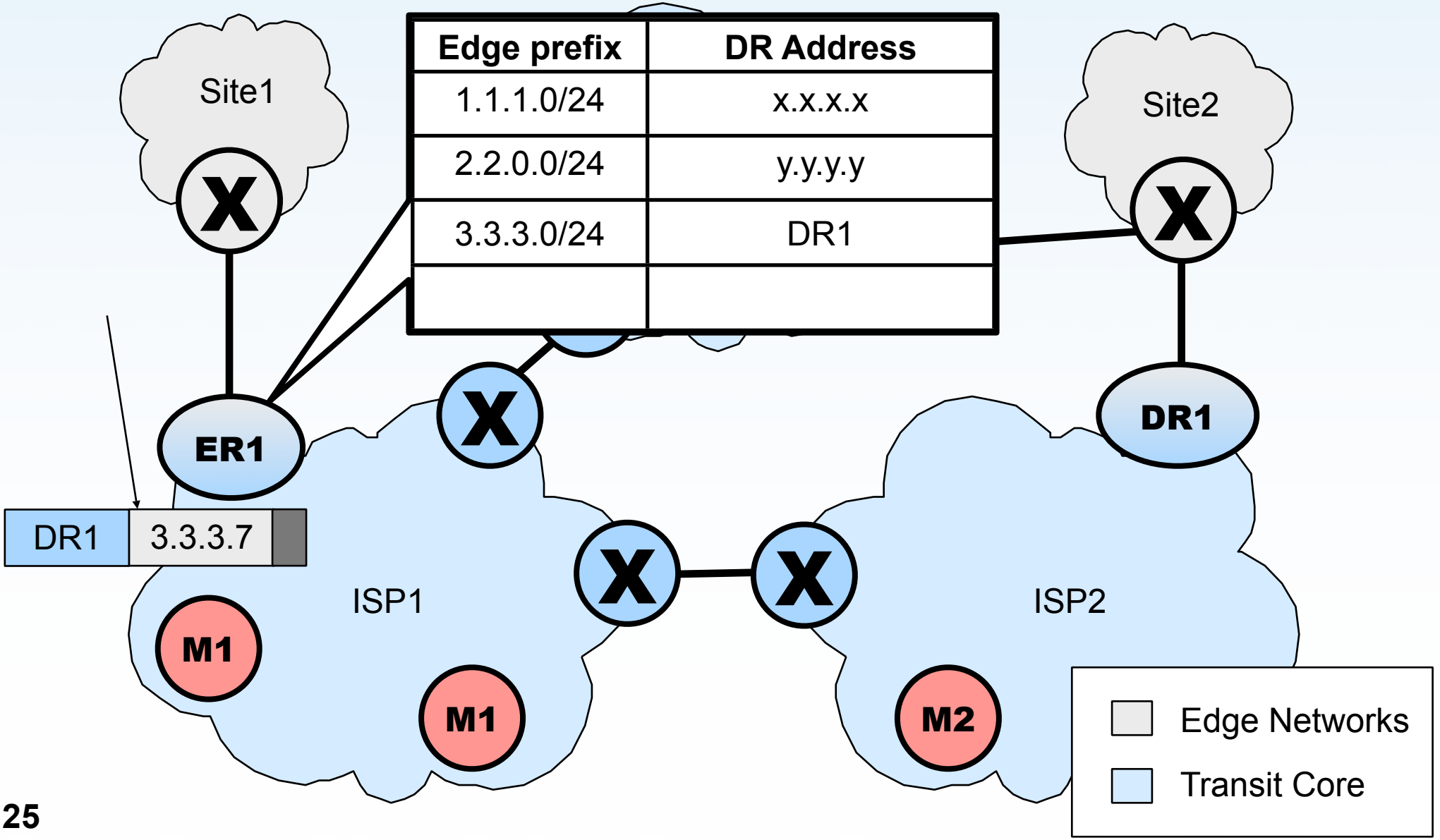
Next Packet


Site1
Site2
3.3.3.7
Edge prefix
DR Address
1.1.1.0/24
x.x.x.x
2.2.0.0/24
y.y.y.y
3.3.3.0/24
DR1
ER1
DR1
ISP1
ISP2
M1
M1
M2
Edge Networks
Transit Core

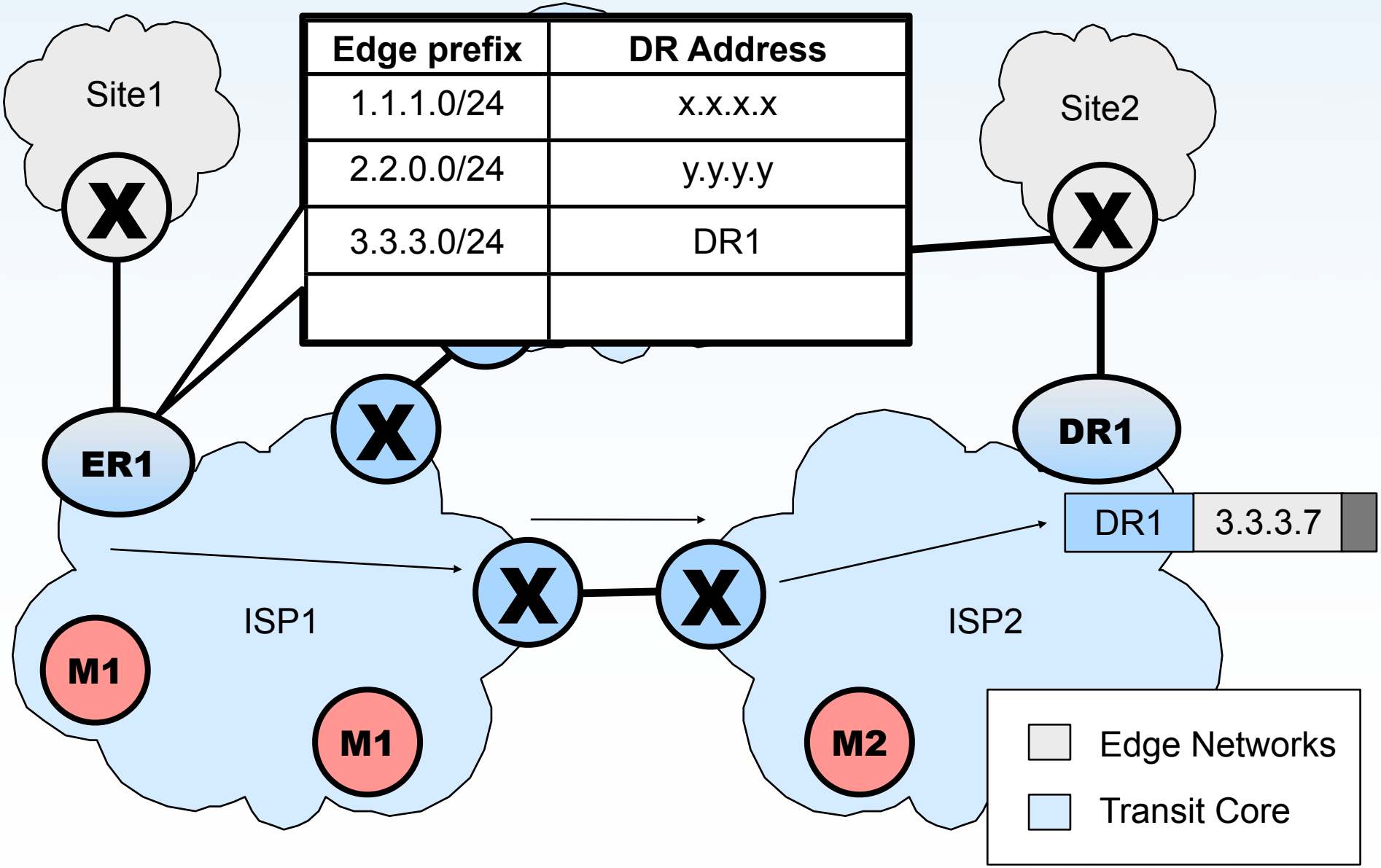
Mapping Already in Cache



Packet Encapsulated by ITR



Packet Delivered Directly to ETR



Minimizing New Infrastructure

- DMs store the entire mapping table locally
- No third party infrastructure for mapping lookups
- So how do DMs get the table?

Mapping Dissemination

- DMs need to learn other ISPs' mapping info
- DMs form an overlay
- ISPs flood their mapping info through the overlay

In Defense of Flooding

- Extremely low latency for mapping lookups
- “Aren’t you just moving the problem from the routing system to the mapping system?”
  - No!
  - Mapping info doesn’t contain reachability info
  - Only changes when customer-provider relationship changes
  - The dissemination path isn’t important

Aligning Cost with Benefit

- EDRs and DMs are both deployed at ISPs
- An ISP can have as many or few DMs as it likes
- Each ISP maintains its own mapping table
- **One ISP can turn on APT on its own**

APT Incremental Deployment

By Dan Jen

APT Deployment Incentives

- APT *can* be deployed unilaterally by an ISP
  - But *why* would an ISP want to do this?

- Answer: Routing Table Reduction!

Routing Table Size is a Problem

- Currently, all routers in an ISP store every prefix in the Internet(currently 300k) in memory.
- Router hardware improvements can't keep up with growing number of prefixes.
  - http://www.vaf.net/~vaf/apricot-plenary.pdf

APT Can Help!

Router Storage Under APT

- 3 types of routers
  - Regular, DRs, and Default Mapper nodes
- All routers store 1 prefix per APT ISP
  - Less than 100 even under universal deployment
- DRs also store a small cache of mappings, and prefixes from directly-connected peers
- Default mapper SYSTEM must also store full mapping table as well as full routing table.
  - But DM system can consist of many nodes

DM Distributed Storage = Memory Savings

- Assume M prefixes on the Internet, mapped and unmapped.
- Assume N nodes make up the DM system
- Each DM node stores M/N prefixes total, mapped and unmapped.
  - Savings!
- All other routers only store about 100 prefixes.
  - Savings!

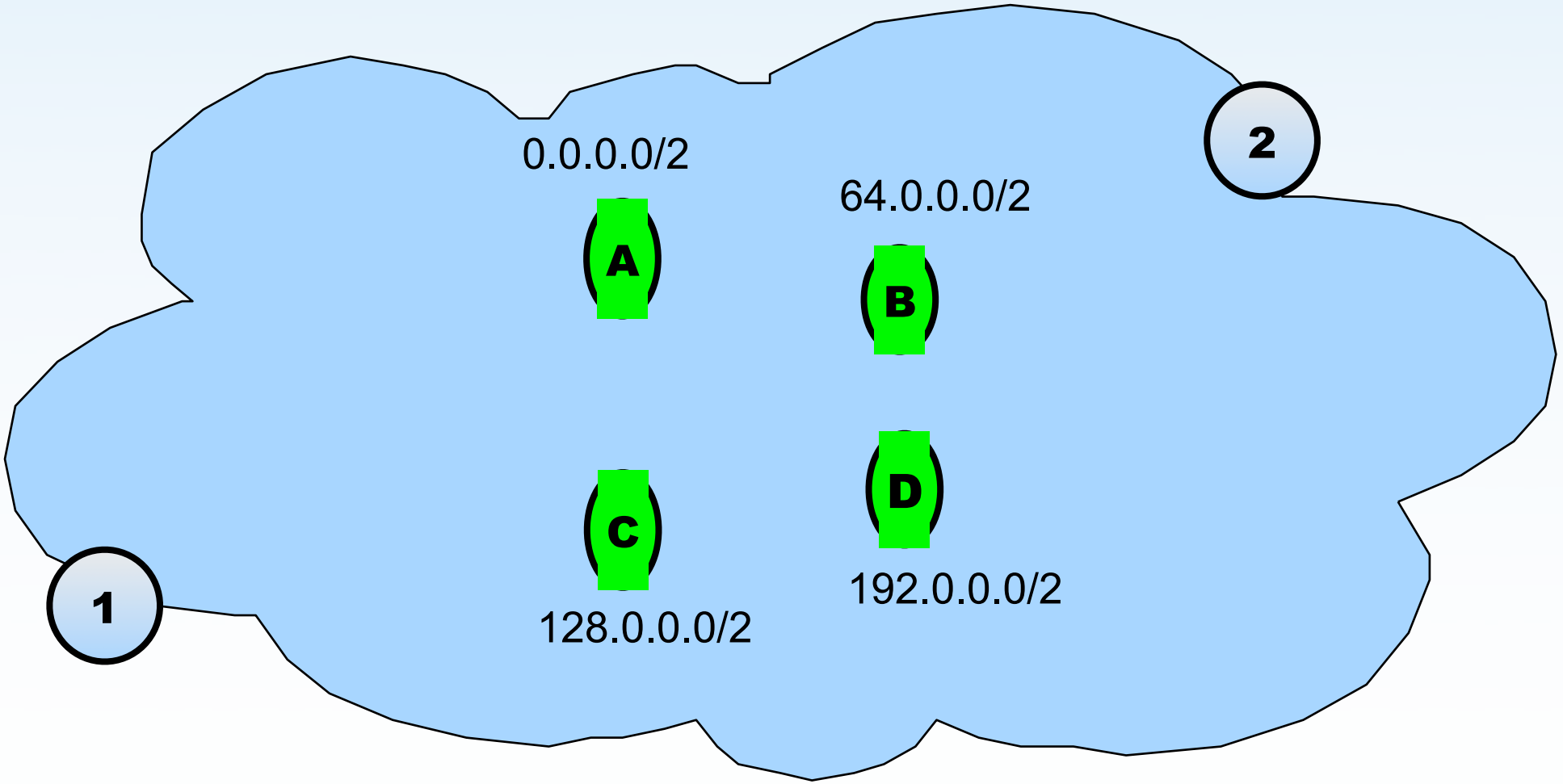
IP addressing primer

- 32-bit addresses
- 4 sections, separated by '.'
- Each section represents 8 bits, from 0-255
- Ex: 5.255.8.24
  - 00000101 11111111 00001000 00011000

IP addressing primer

- To represent addresses from 1.2.3.0 to 1.2.3.255, we use 1.2.3.0/24, meaning:
  - All addresses whose first 24 bits are 1.2.3
- Thus we can divide the entire IP address range into 4 equal parts:
  - 0.0.0.0/2
  - 64.0.0.0/2
  - 128.0.0.0/2
  - 192.0.0.0/2

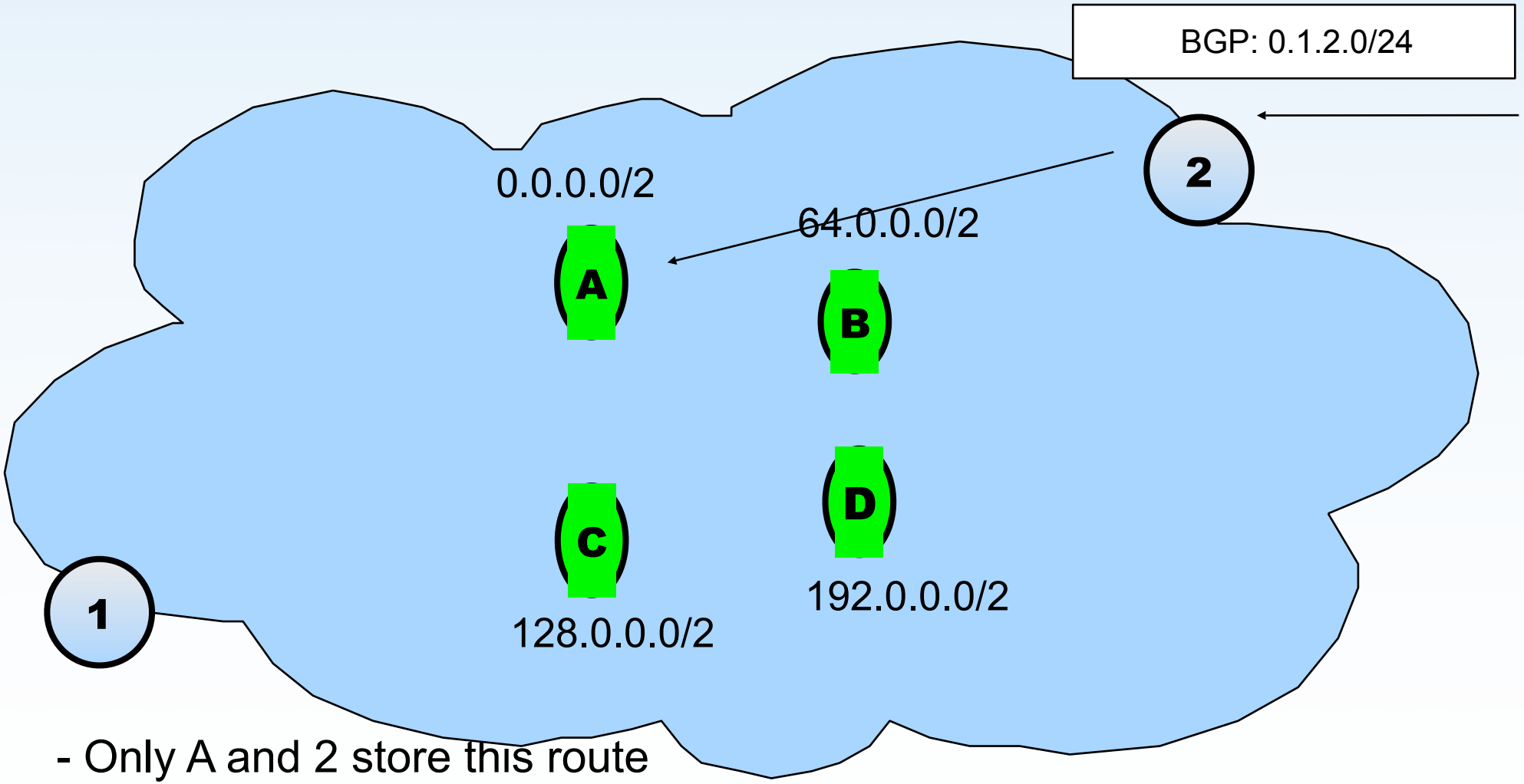
APT 1st Mover Savings



- numbered routers are DRs

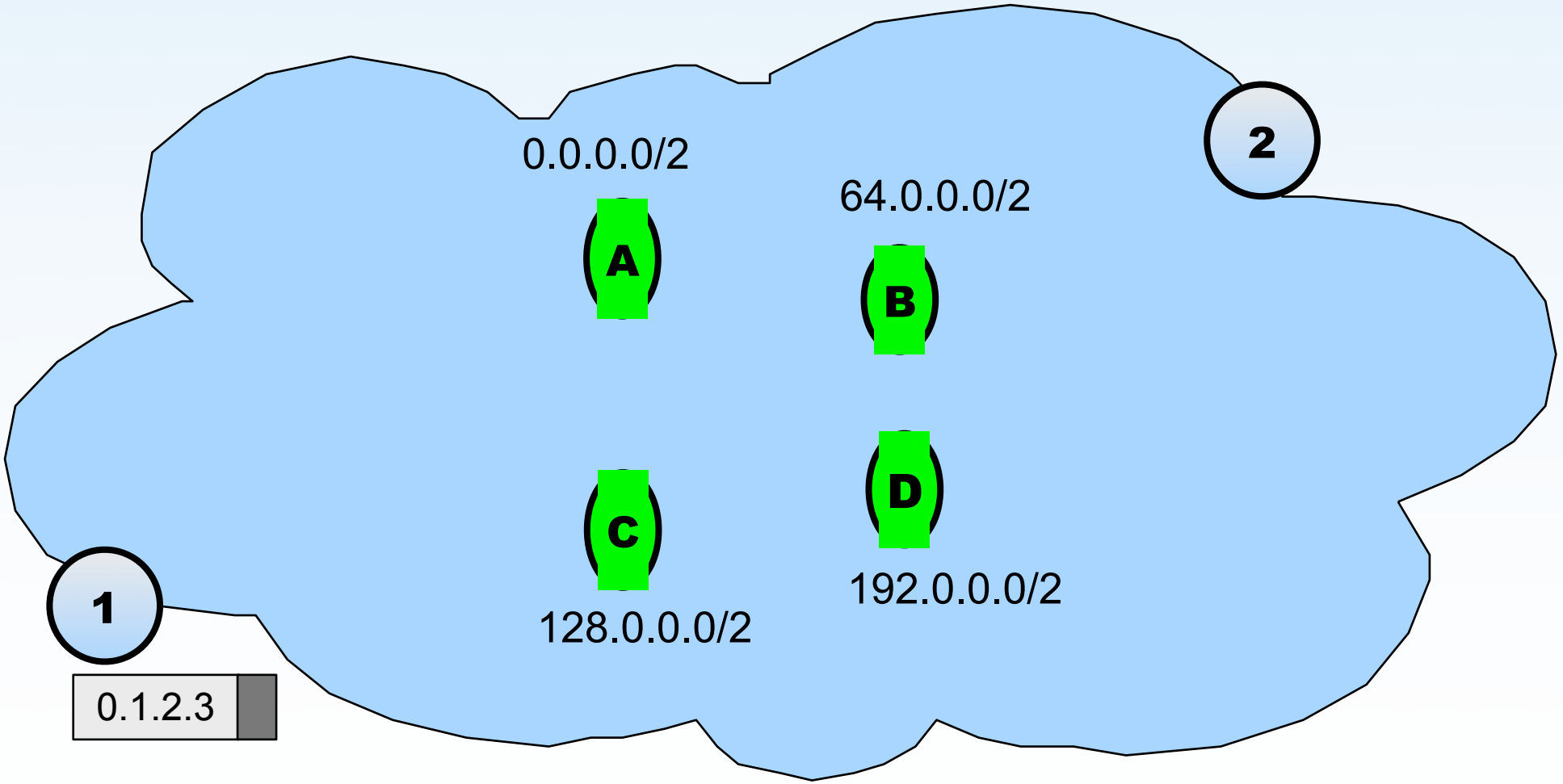
- A,B,C,D make up the DM system

APT 1st Mover Savings

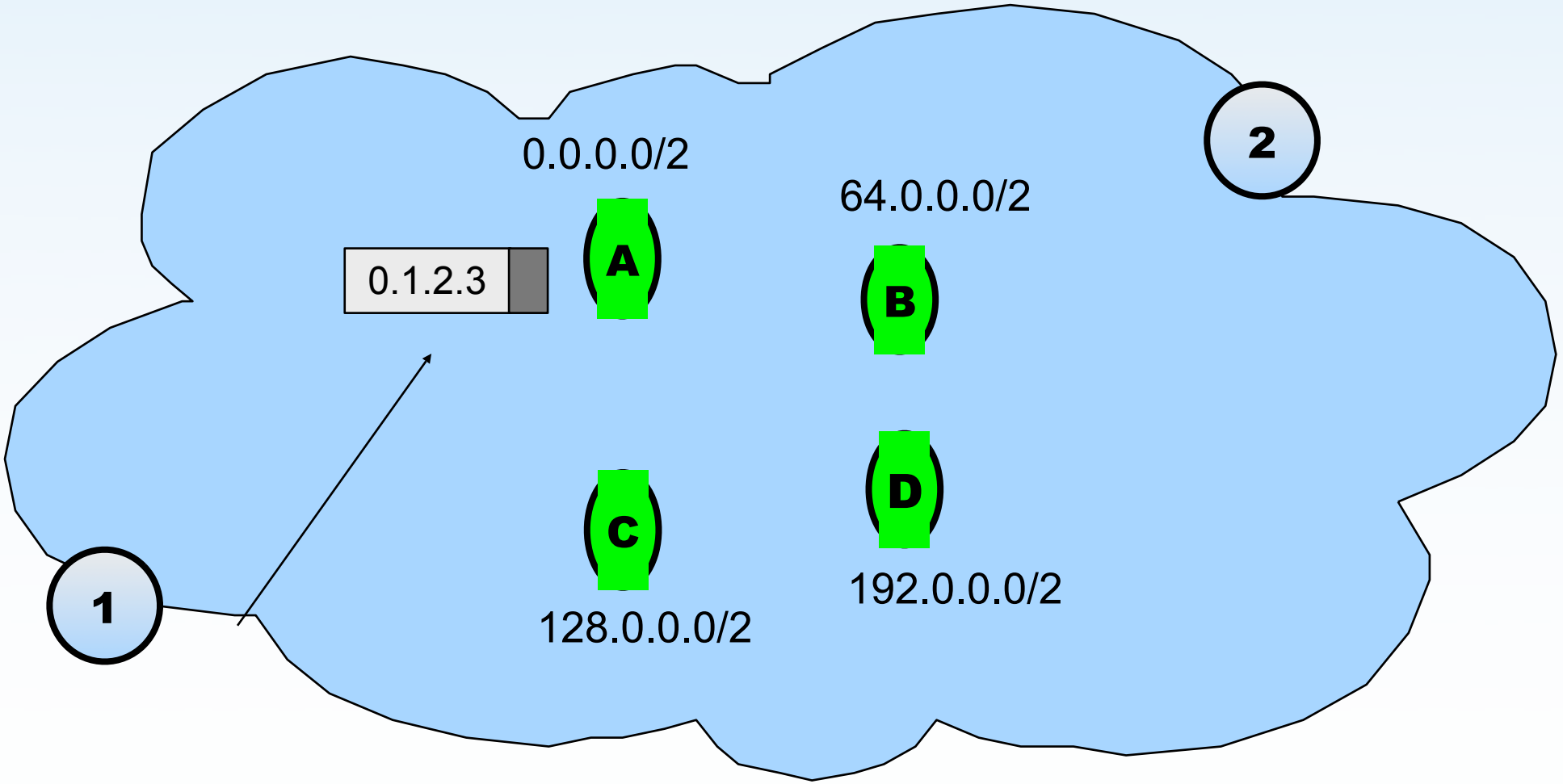


- Only A and 2 store this route
- A knows 2 is the egress point for address range
- 2 knows the real external next hop for the prefix

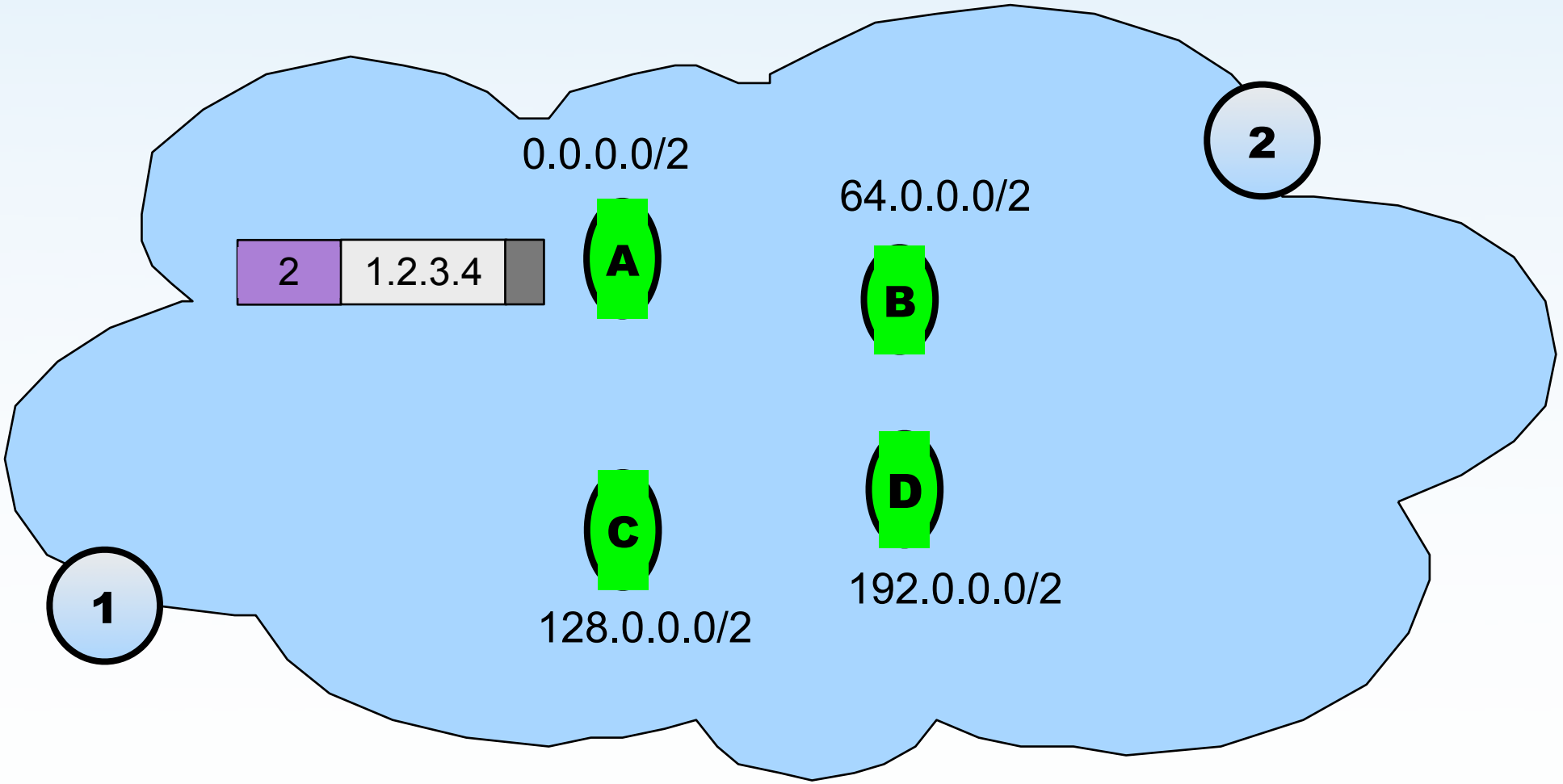
APT 1st Mover Packet Delivery



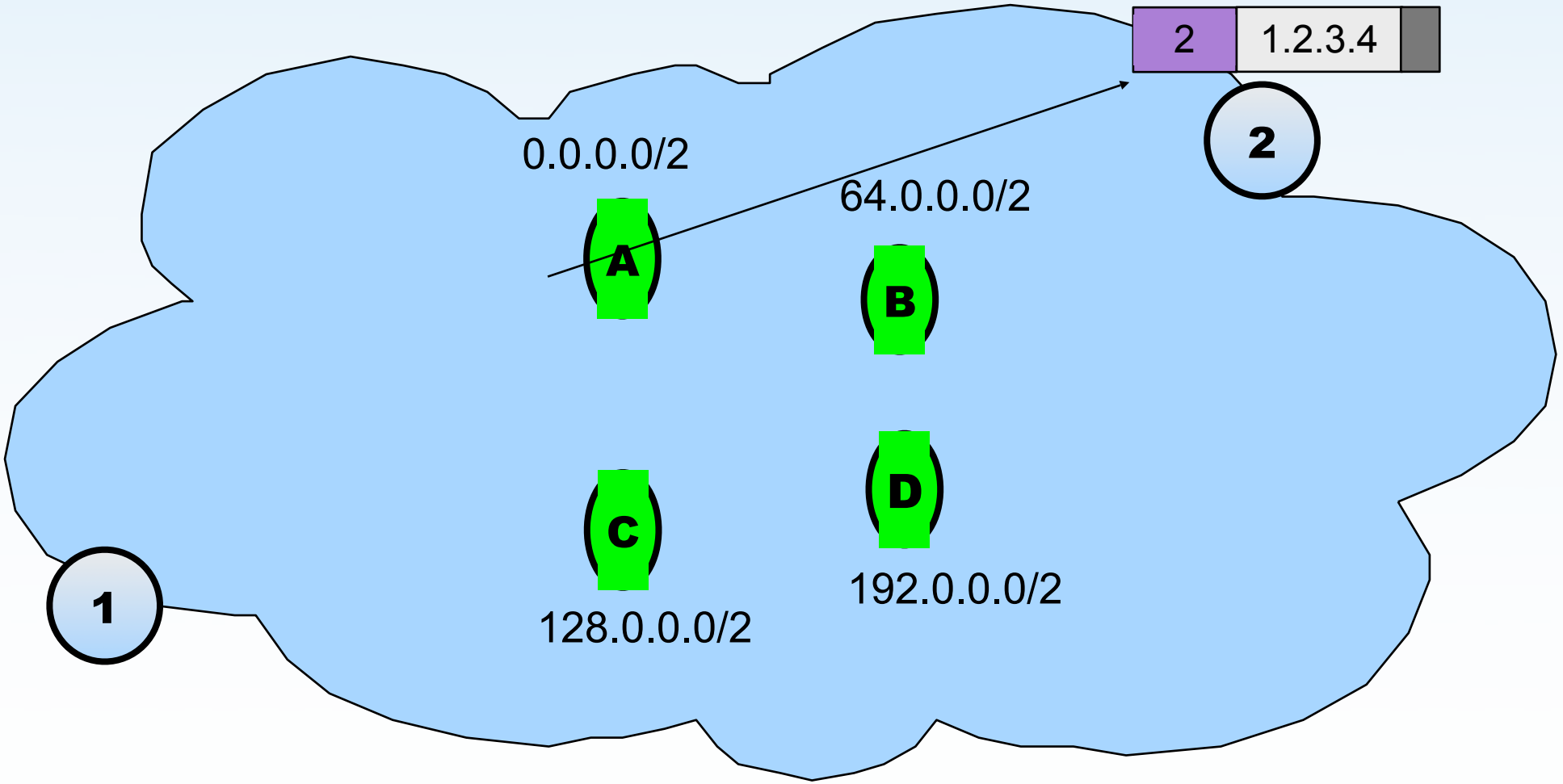
APT 1st Mover Packet Delivery



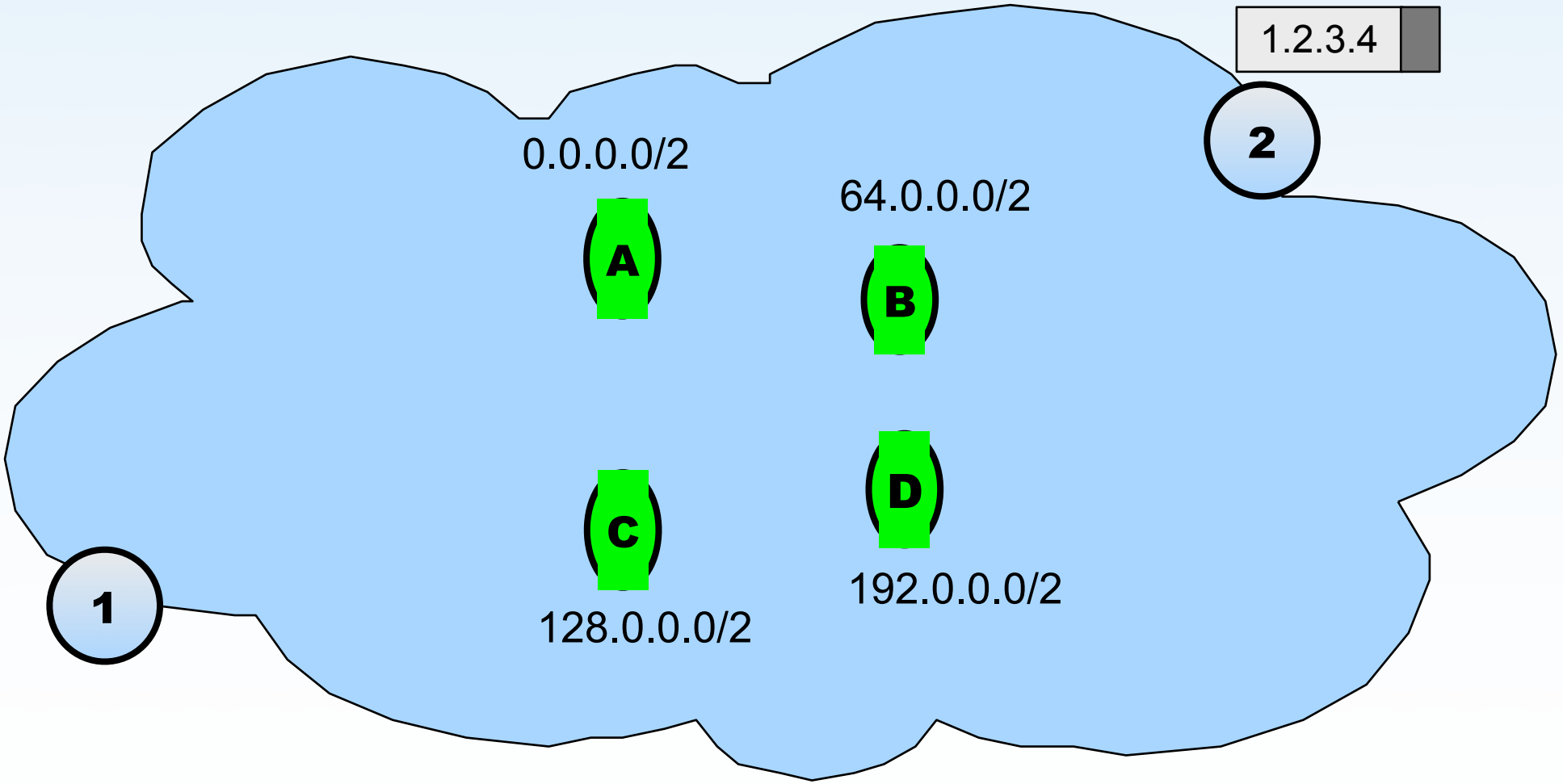
APT 1st Mover Packet Delivery



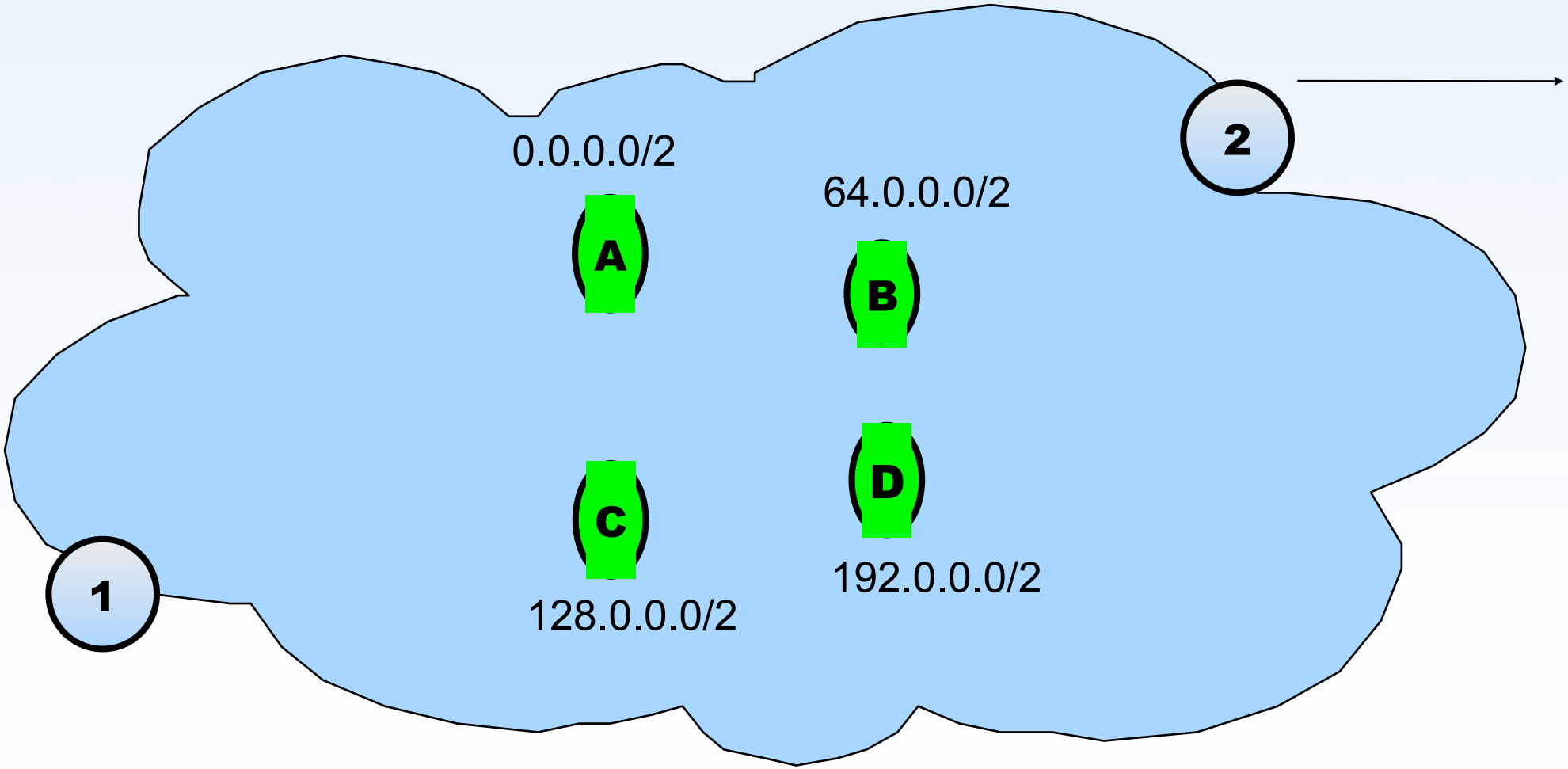
APT 1st Mover Packet Delivery



APT 1st Mover Packet Delivery



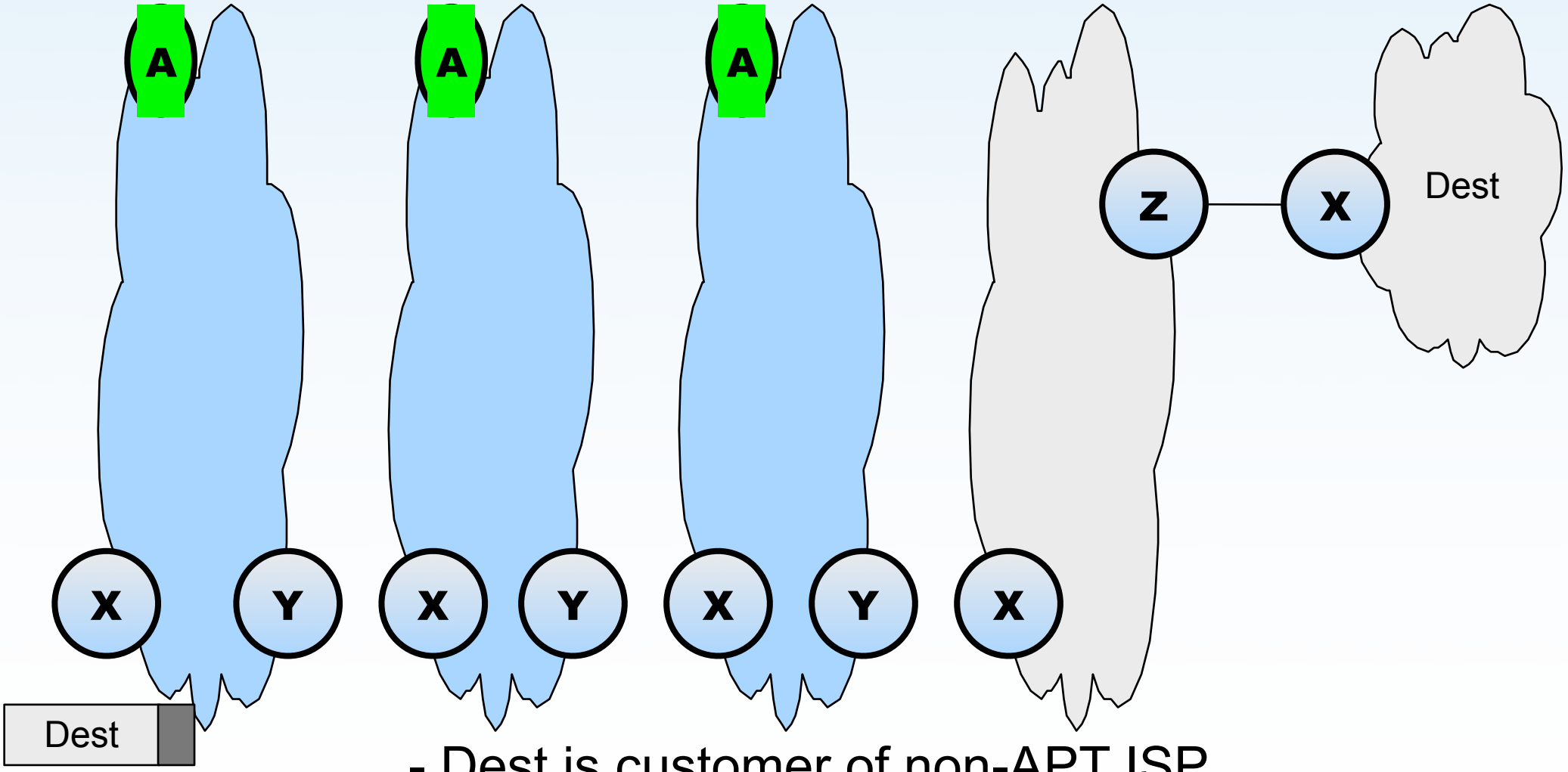
APT 1st Mover Packet Delivery



The Stretch Issue

- Note that packets traveling through an APT ISP will travel extra hops before exiting the AS
- Every APT ISP will stretch packets traversing through it. Stretch can add up.

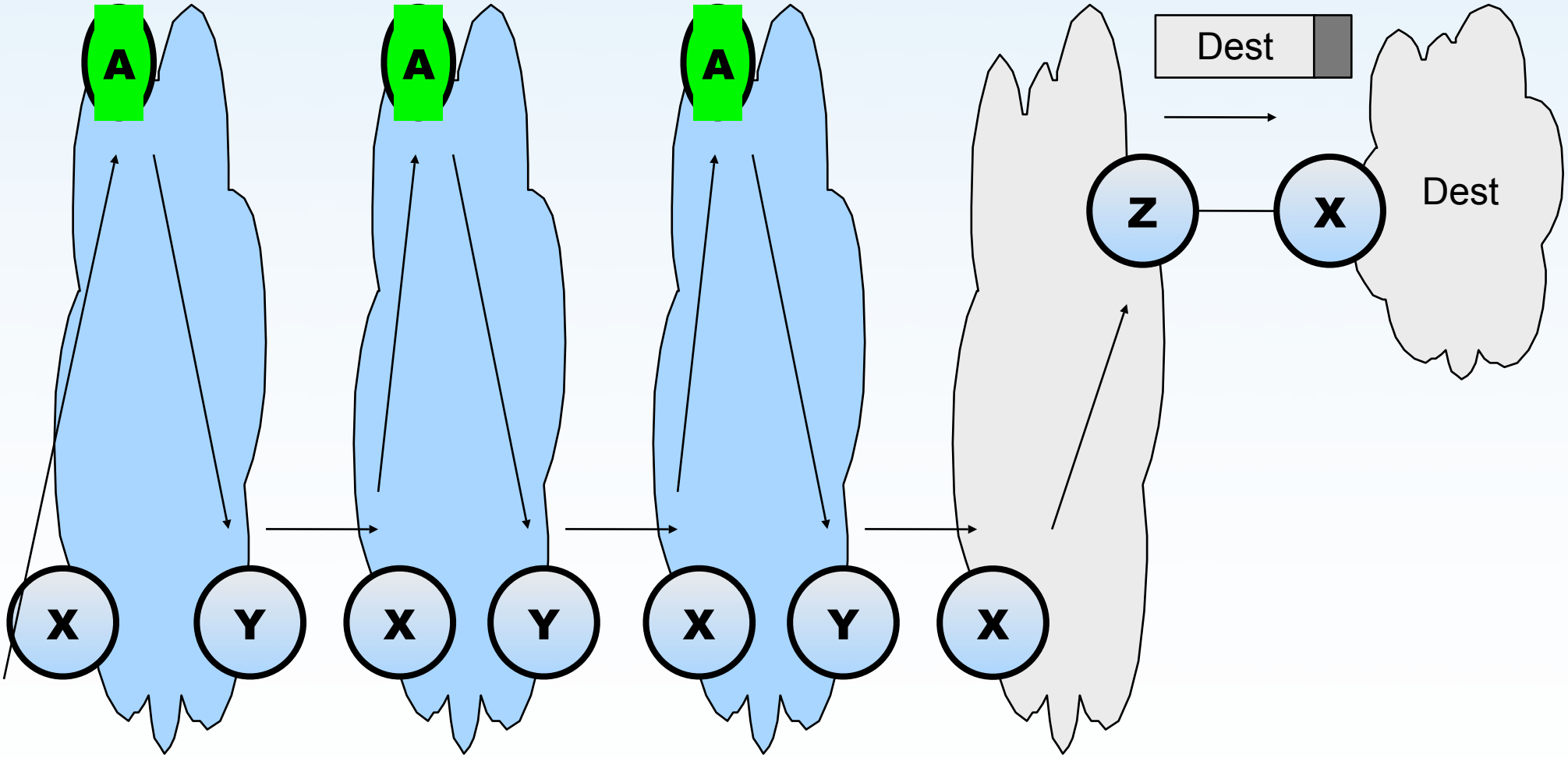
The Stretch Issue



- Dest is customer of non-APT ISP

- All blue ISPs are APT ISPs

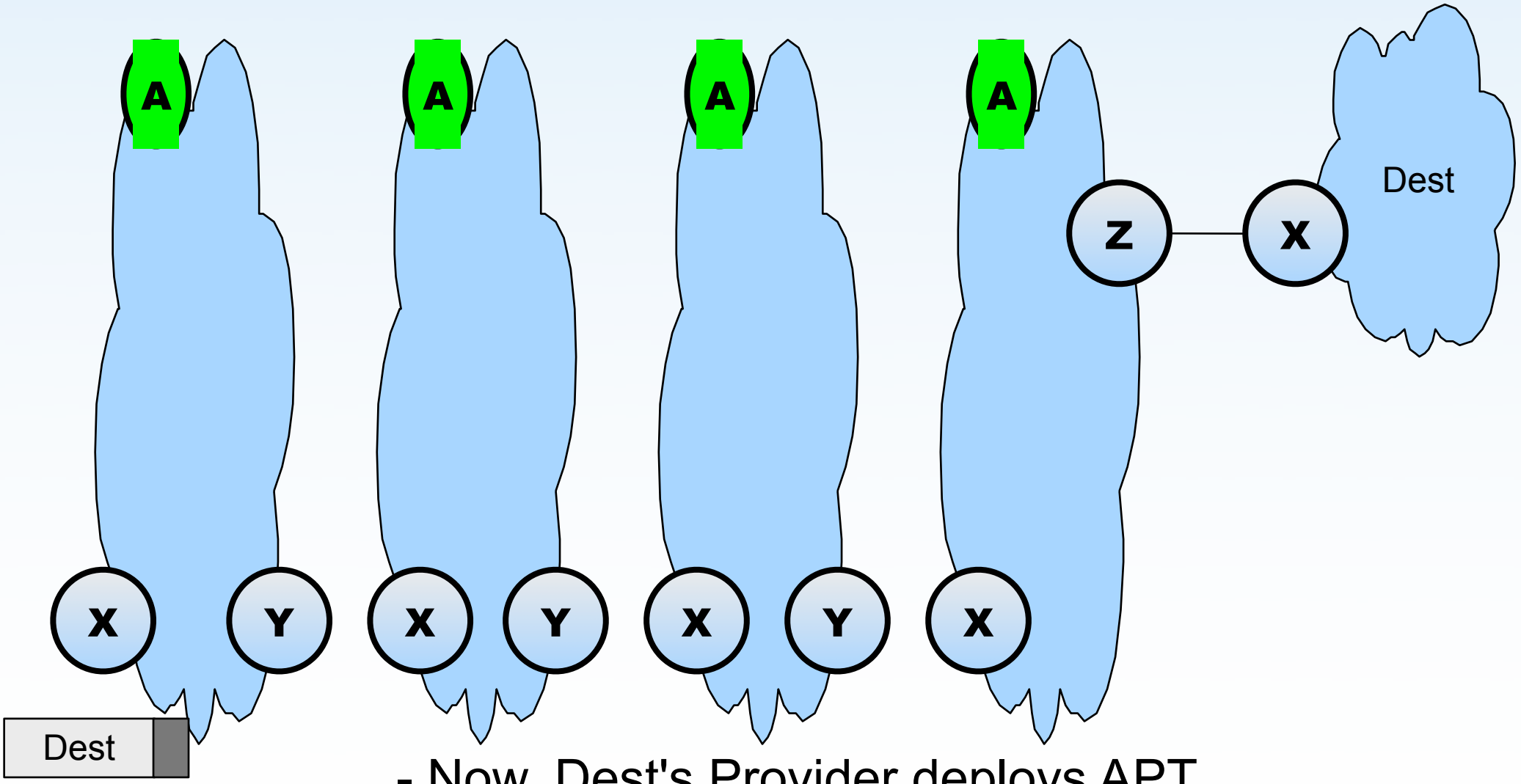
The Stretch Issue



2nd Mover Incentive

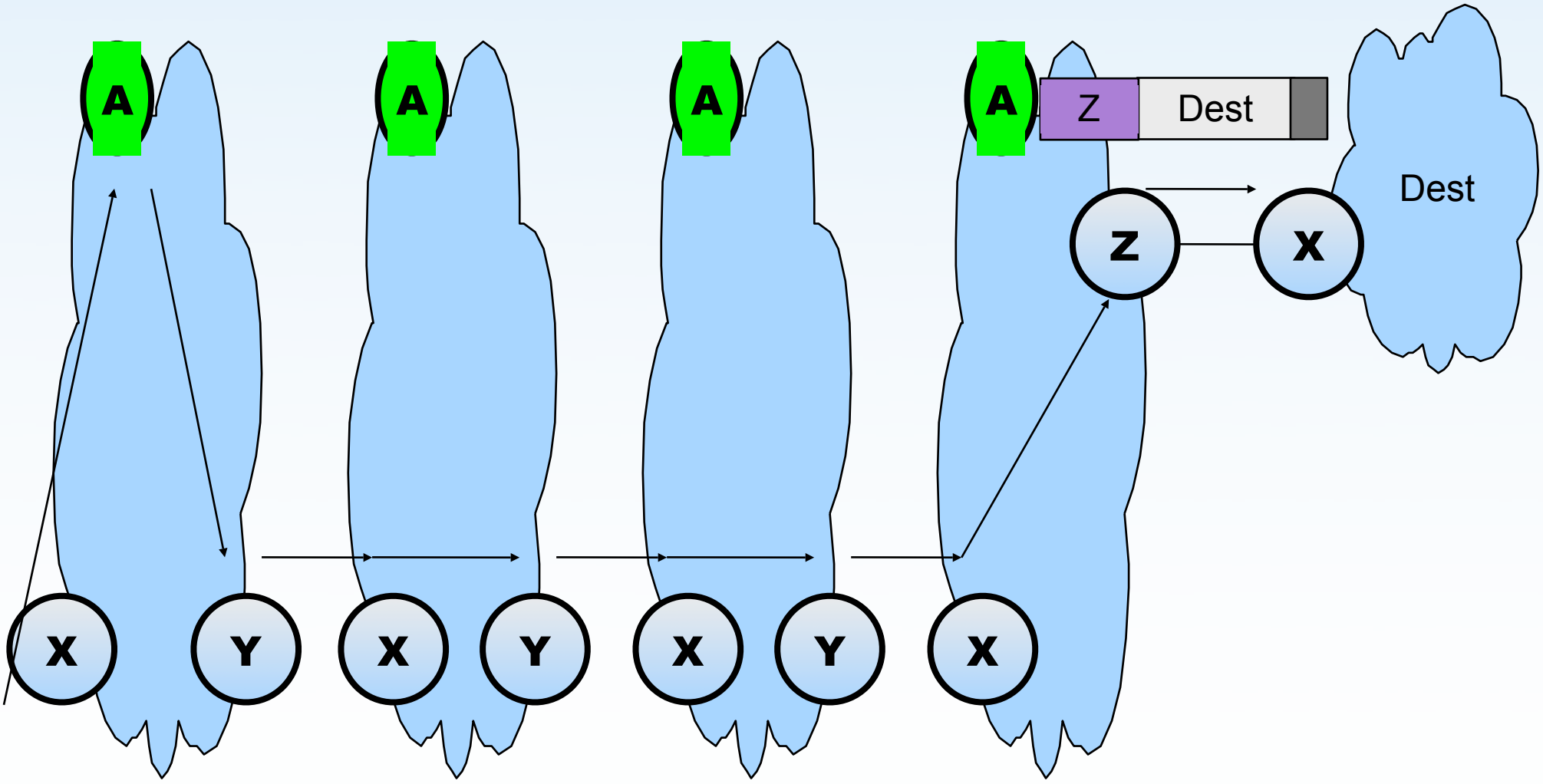
- A 2nd ISP might want to adopt APT for all the same benefits as 1st mover
- But also, ISP can now avoid “the stretch issue”
  - APT ISP can join overlay
  - Mappings ensure that packets stretched at most once.
  - No stretch on cache hits

2nd Mover Incentive



- Now, Dest's Provider deploys APT

2nd Mover Incentive



- This assumes a cache miss.

- Cache hit would add no stretch at all.

Thank You!

- Questions?
- Comments?