APT: A Practical Transit-Mapping Service

Overview and Comparisons

draft-jen-apt

Dan Jen, Michael Meisel, Dan Massey, Lan Wang,
Beichuan Zhang, and Lixia Zhang

The Big Picture

- APT is similar to LISP at a very high level
  - Both separate routable addresses from endpoint addresses
  - Both use map-n-encap with UDP encapsulation
  - Both need a mapping service design
- Our design philosophy is "do no harm"
  - Avoid packet loss whenever possible
  - Minimize mapping service latency
  - Alignment of cost and performance

Thus, LISP and APT Differ in Significant Ways

- Distribution of mapping information
- Handling of transient failures
- Deployment scenarios

Outline

A. APT overview & major differences from LISP

1. Where mapping information is stored
2. Handling transient failures
3. Mapping dissemination
4. Incremental deployment

B. Comparison of APT and LISP

1. ISP-based vs. end-site-based deployment
2. Local vs. remote mapping pull
3. Flat vs. hierarchical mapping retrieval infrastructure

NOTE: Where LISP mapping designs differ, comparisons assume LISP-ALT

Part A1
Where Mapping Information is Stored

Terminology

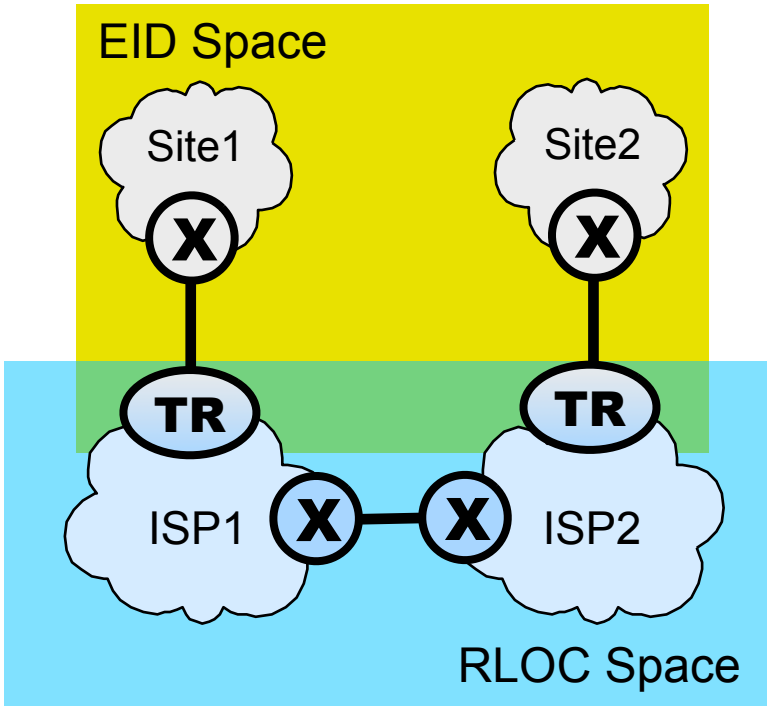
- EID and RLOC
  - We are using the LISP terminology for clarity
  - Note that EIDs are not really identifiers, just addresses
- MapSet
  - Maps an EID prefix to the entire ***set*** of ETR RLOCs through which it can be reached
  - Used by default mappers in APT
  - Used by TRs in LISP
- MapRec
  - Maps an EID prefix to a single ETR RLOC
  - Used by TRs in APT

Where Mapping Information is Stored

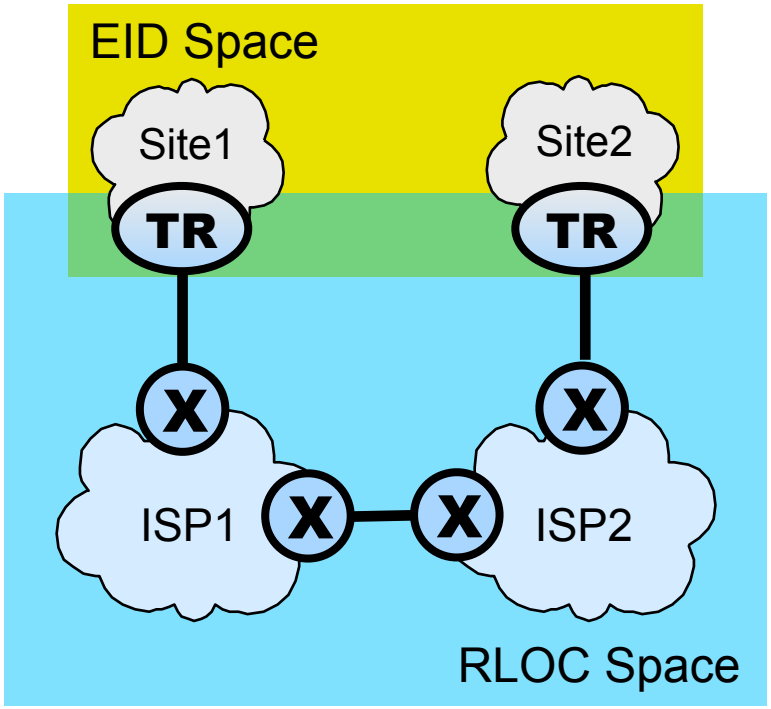
- APT and LISP ITRs both cache recently used mappings
- In LISP
  - LISP sites don't store the entire mapping table
  - Mapping information is retrieved via a remote pull
    - From the destination ETR
- In APT
  - Each AS stores a copy of the entire mapping table
    - In local devices called default mappers
  - Mapping information is retrieved via a local pull
    - Within the source AS

TR Placement: APT vs. LISP

- APT and LISP both claim to support CE or PE TRs
- APT recommends PE, LISP recommends CE



TRs at PEs (APT)



TRs at CEs (LISP)

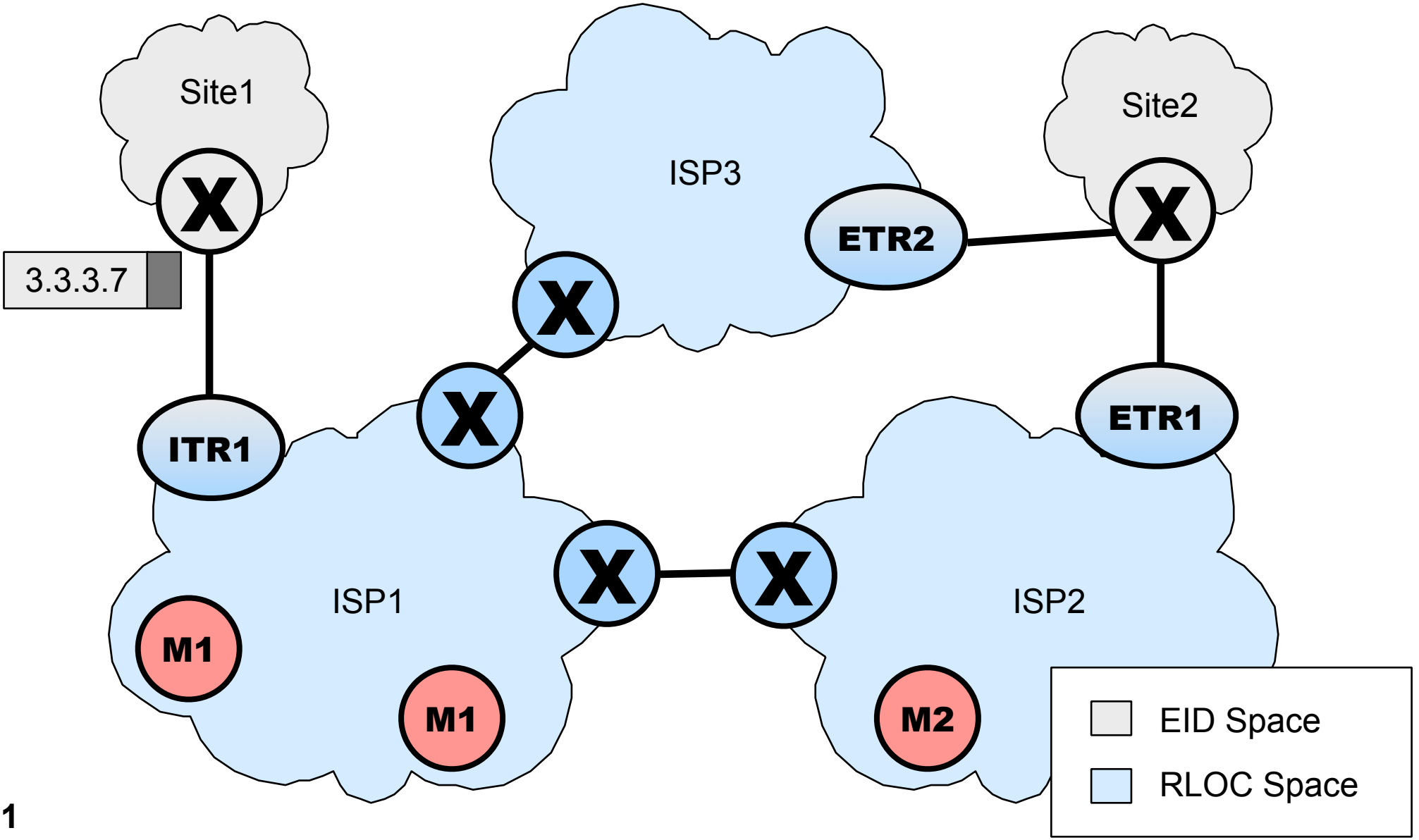
APT Tunnel Routers (TRs)

- One device with ITR and ETR functionality
- Cache only MapRecs
  - Delete unused MapRecs after some TTL
  - On a cache miss
    - Sends the packet to a default mapper
    - Default mapper encaps the packet with an ETR address
    - Default mapper sends the TR a MapRec for its cache

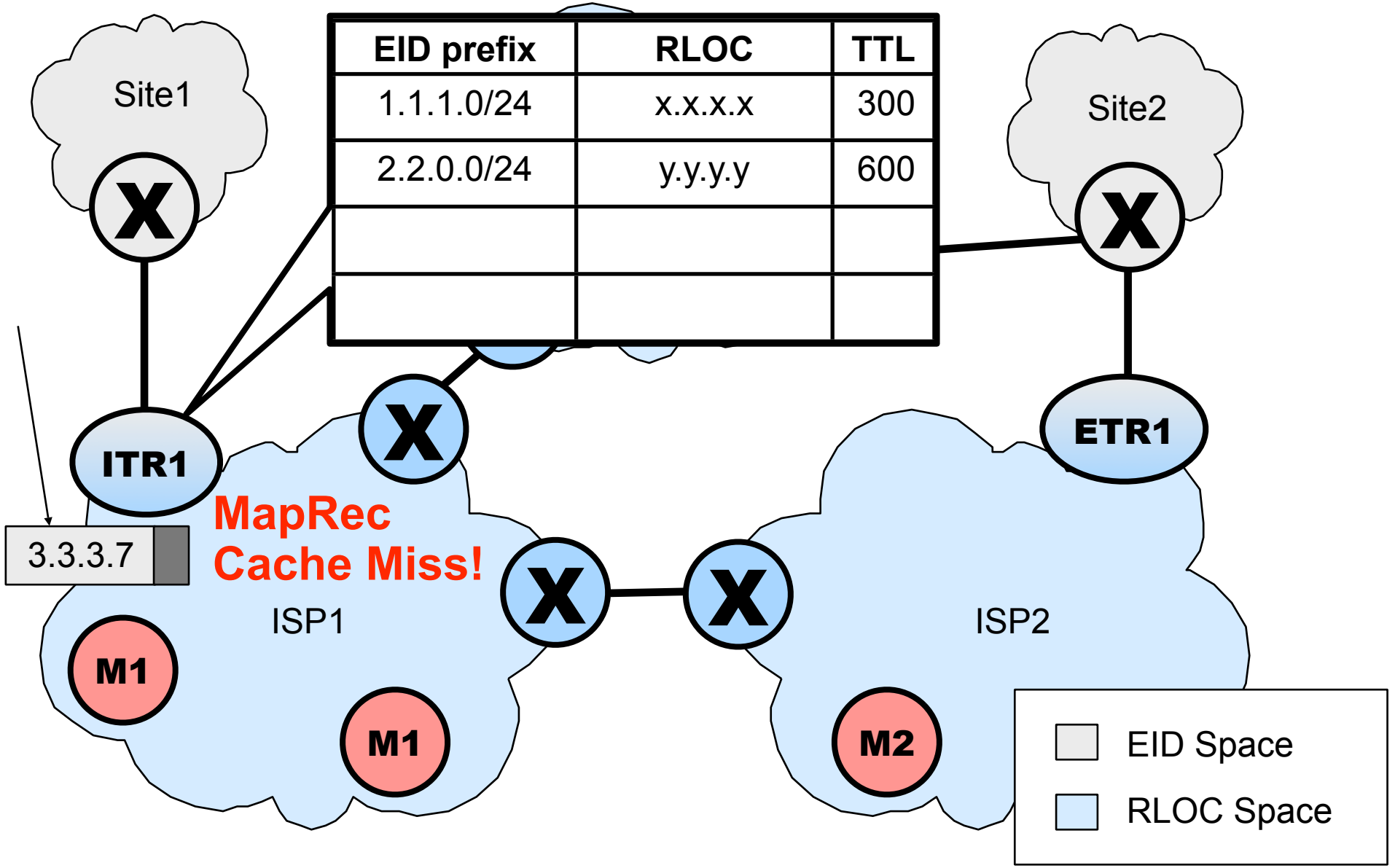
Default Mappers

- Store full MapSets
  - Each RLOC has a weight and priority for TE support
  - Handles ETR-selection policy so ITRs don't have to
- One or more default mappers per ISP
  - Each ITR can reach any of the default mappers in its ISP using the same anycast address for reliability

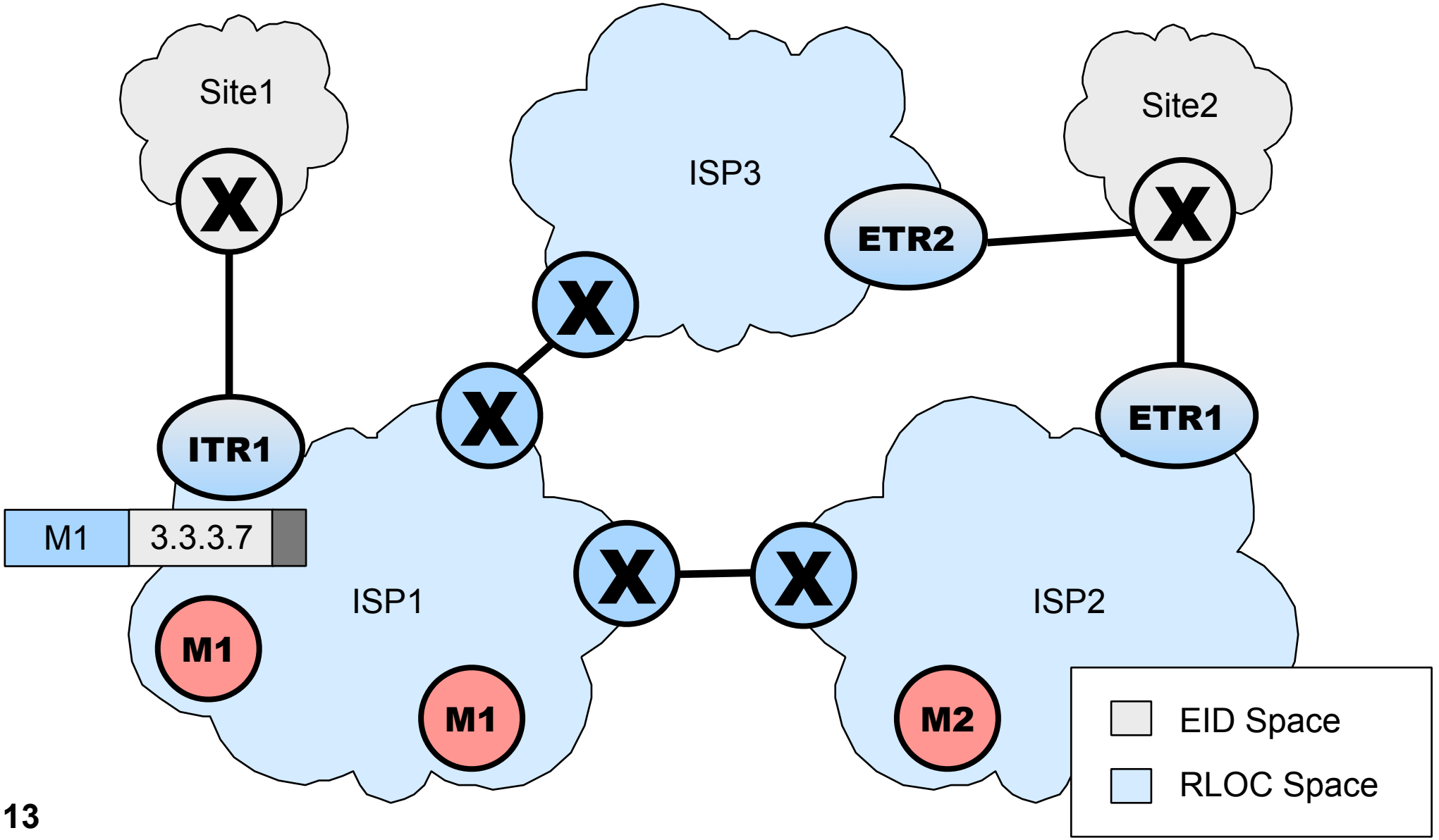
APT Example


Site1
3.3.3.7
ITR1
ISP3
ETR2
Site2
ETR1
ISP1
M1
M1
ISP2
M2
EID Space
RLOC Space

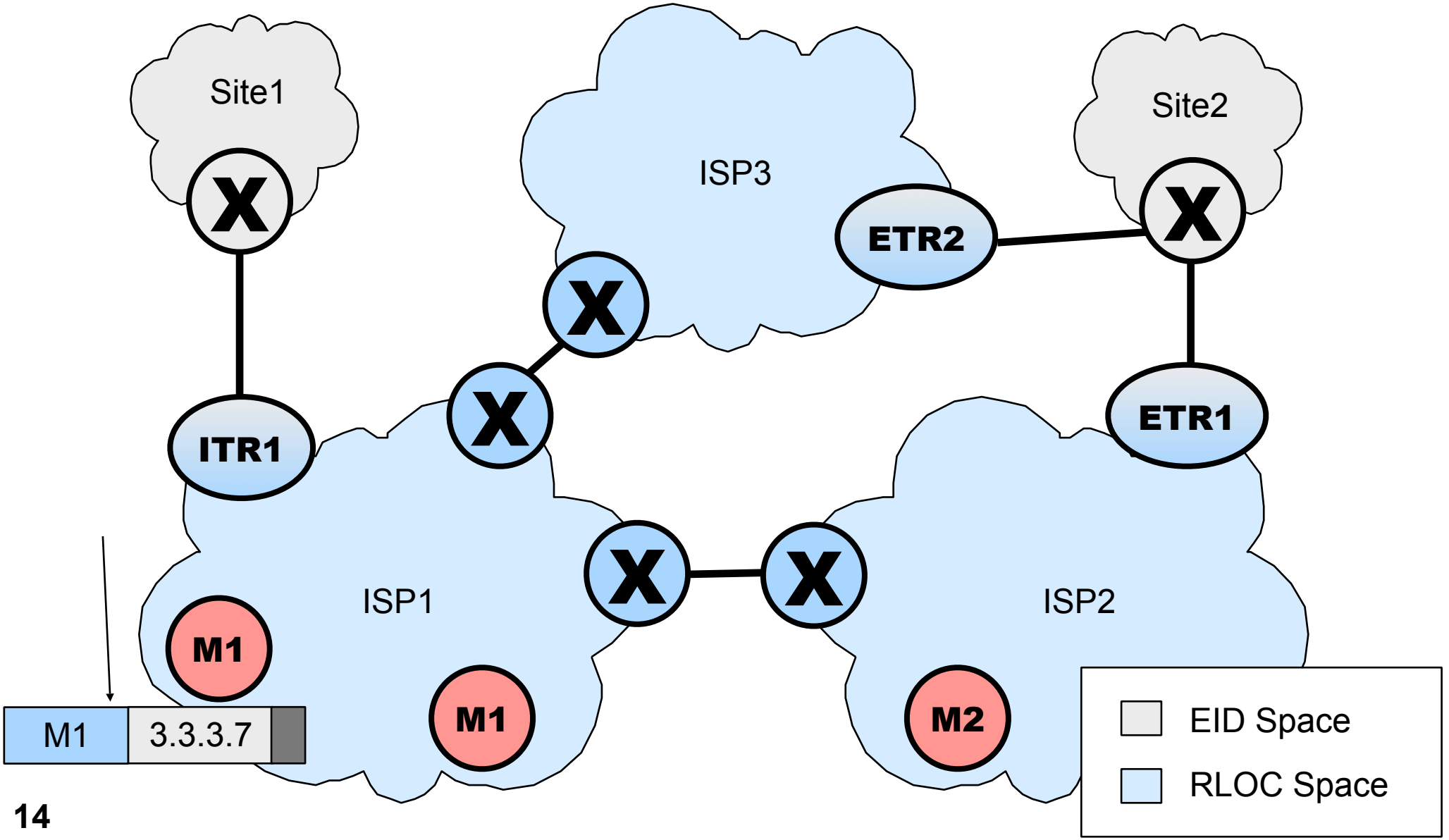
MapRec Not in Cache



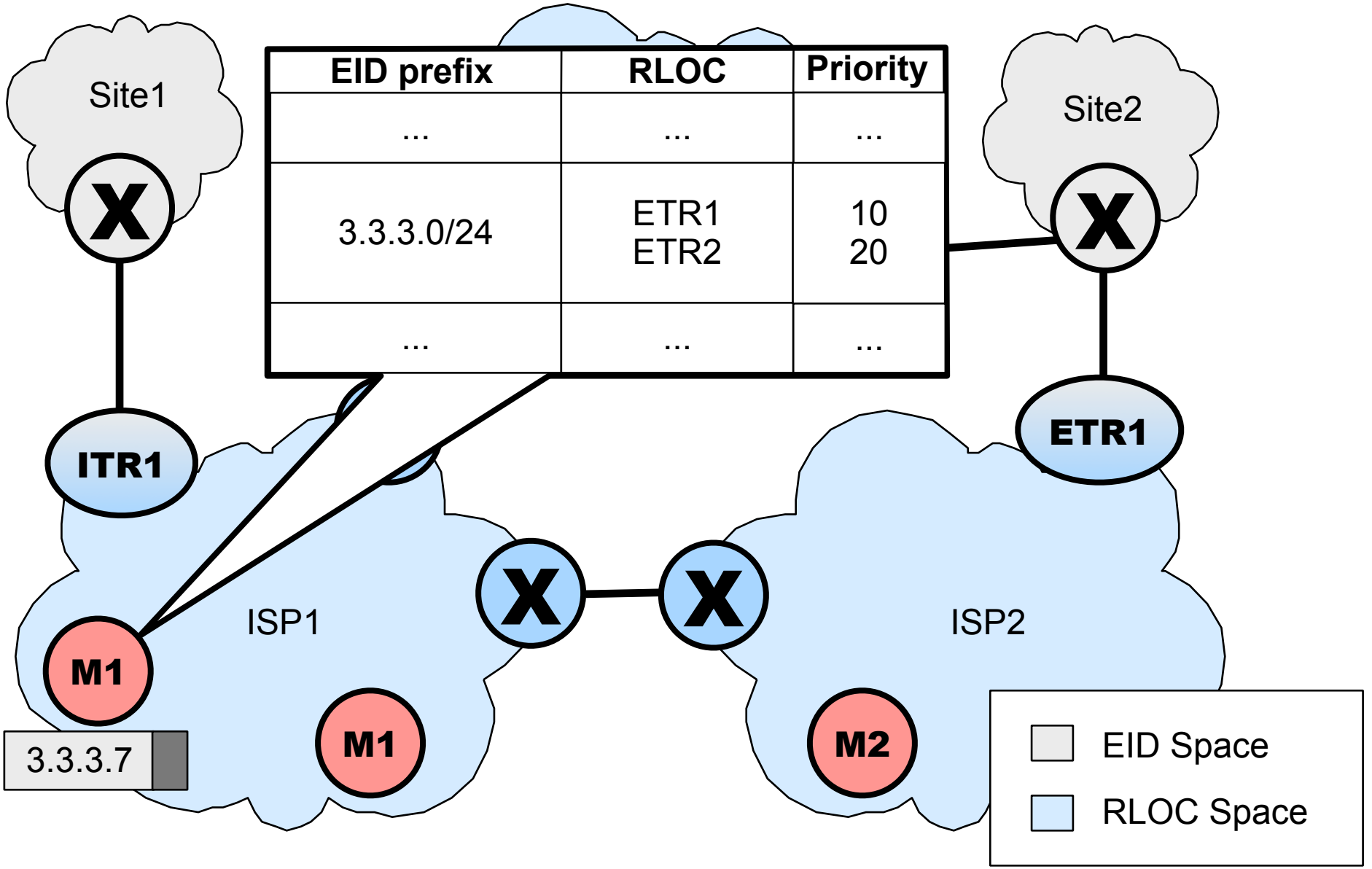
Encap with the Default Mapper Anycast Address



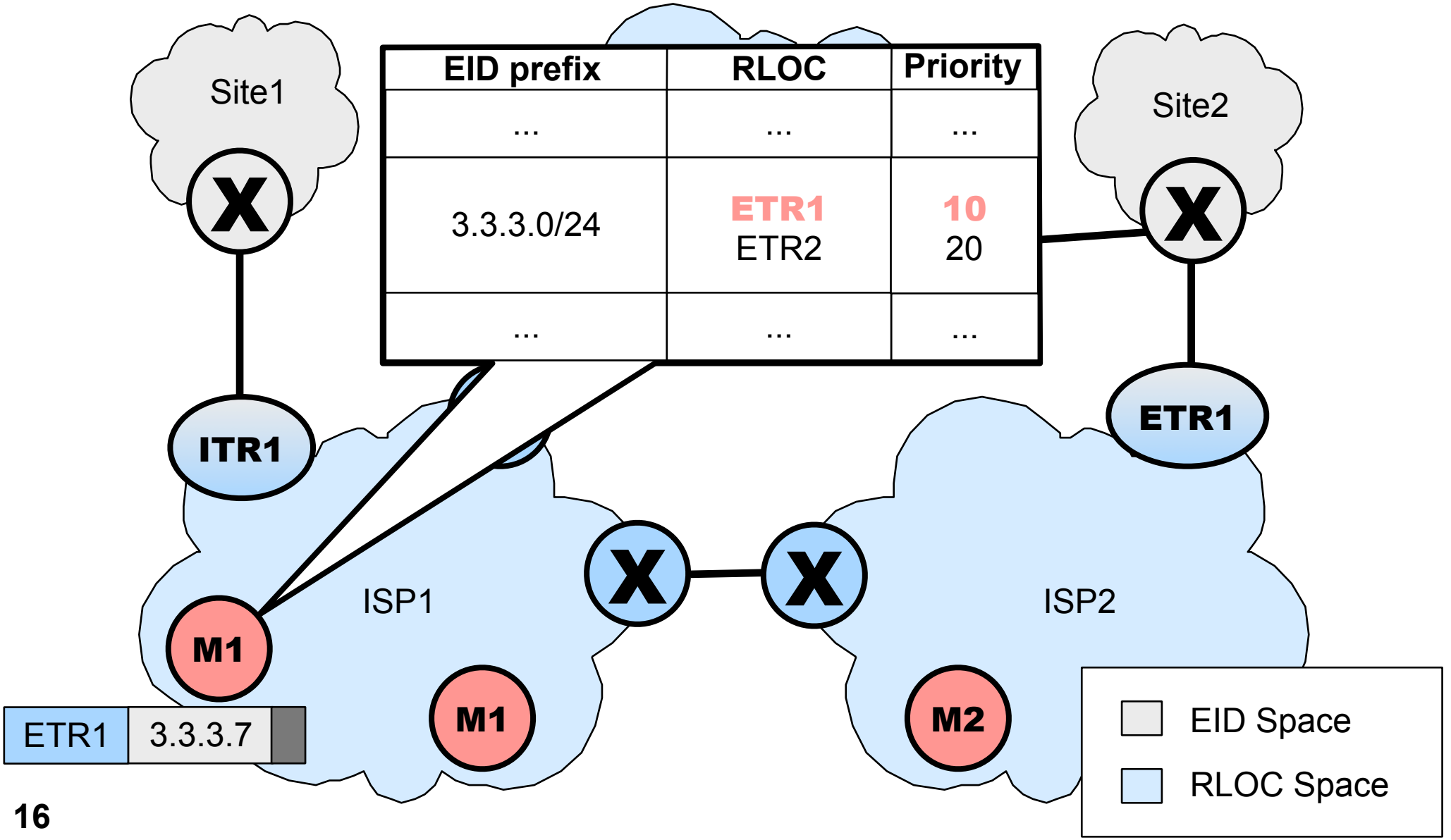
Default Mapper Decaps the Packet



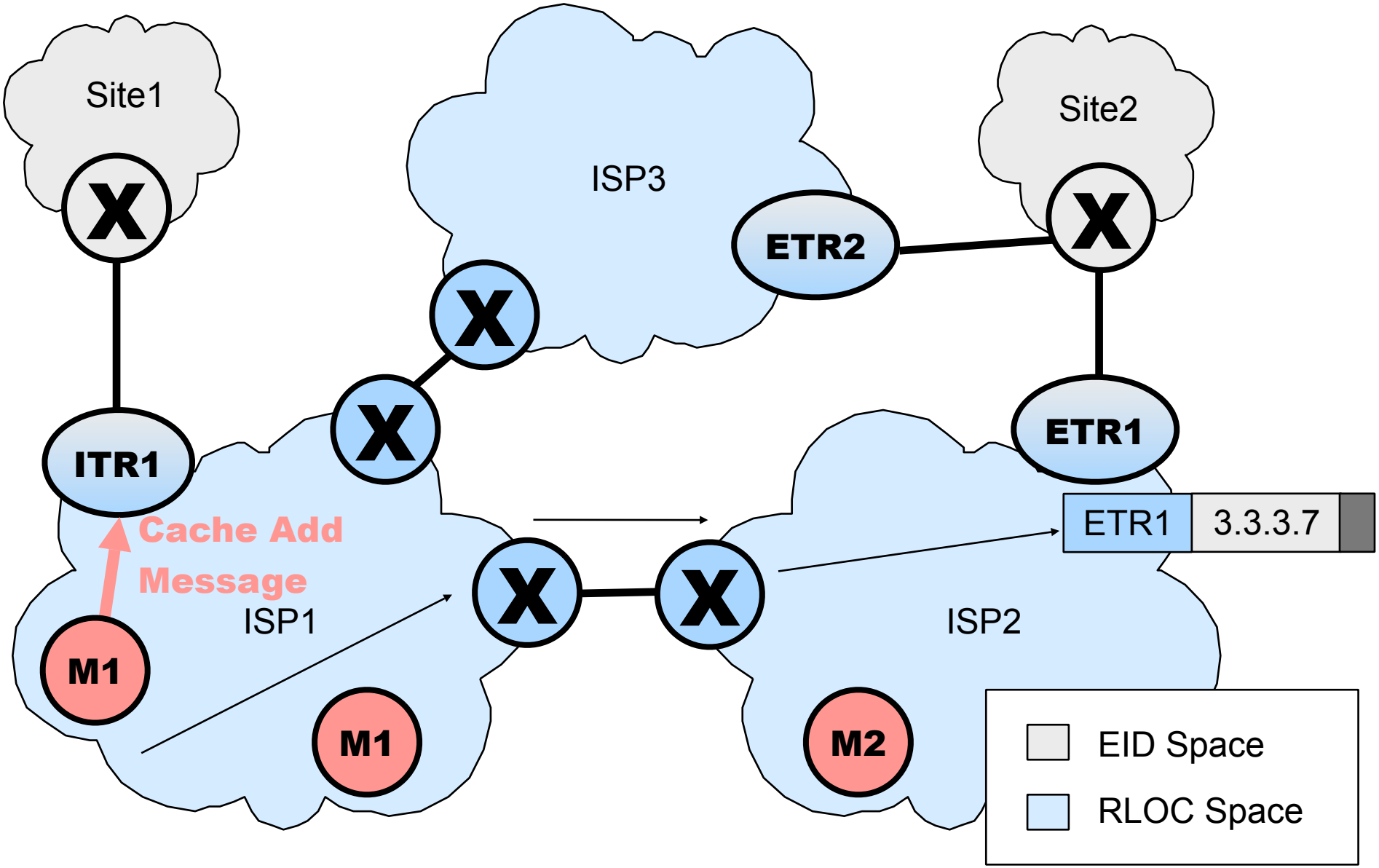
EID Prefix is Multihomed



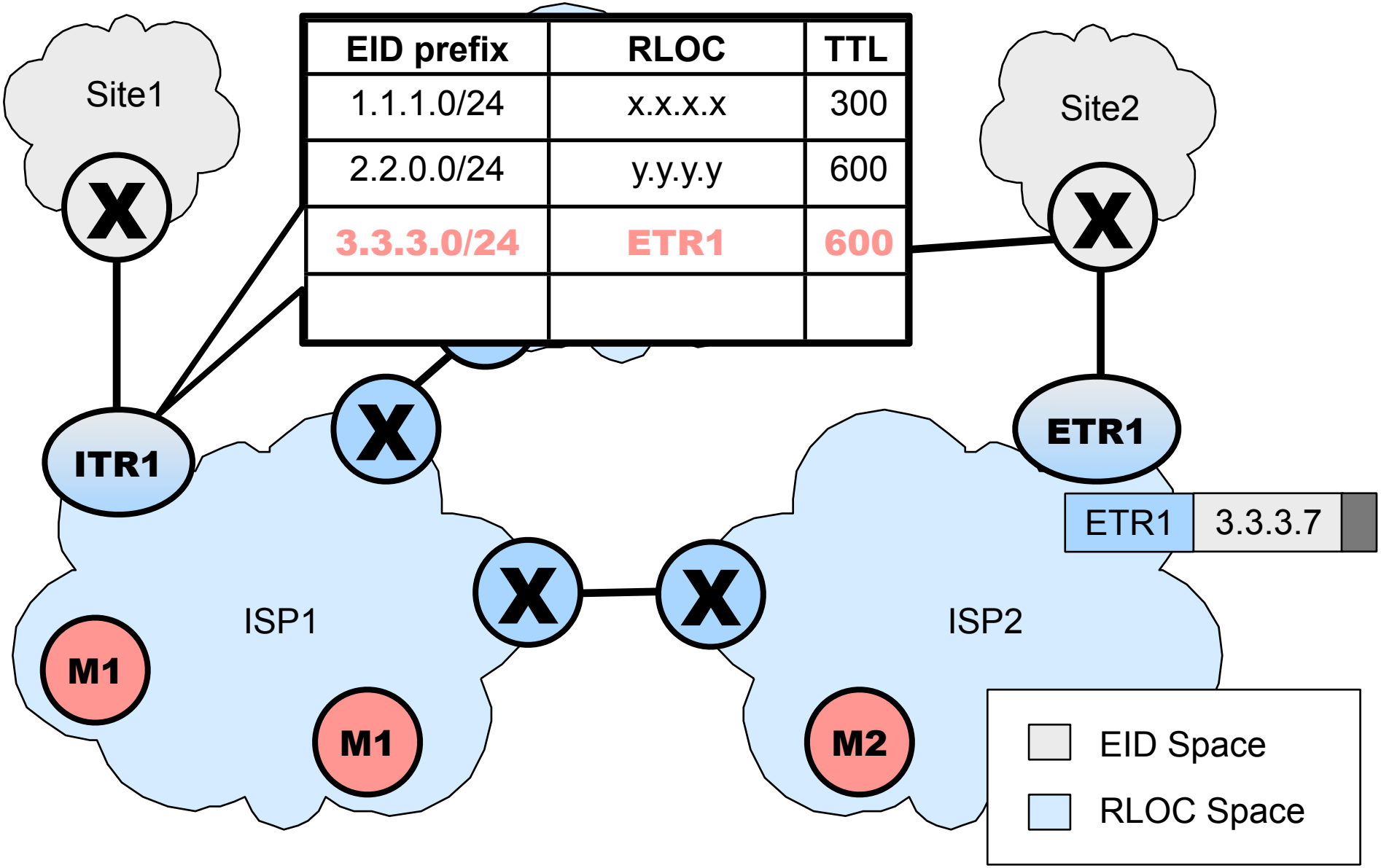
Default Mapper Selects a MapRec



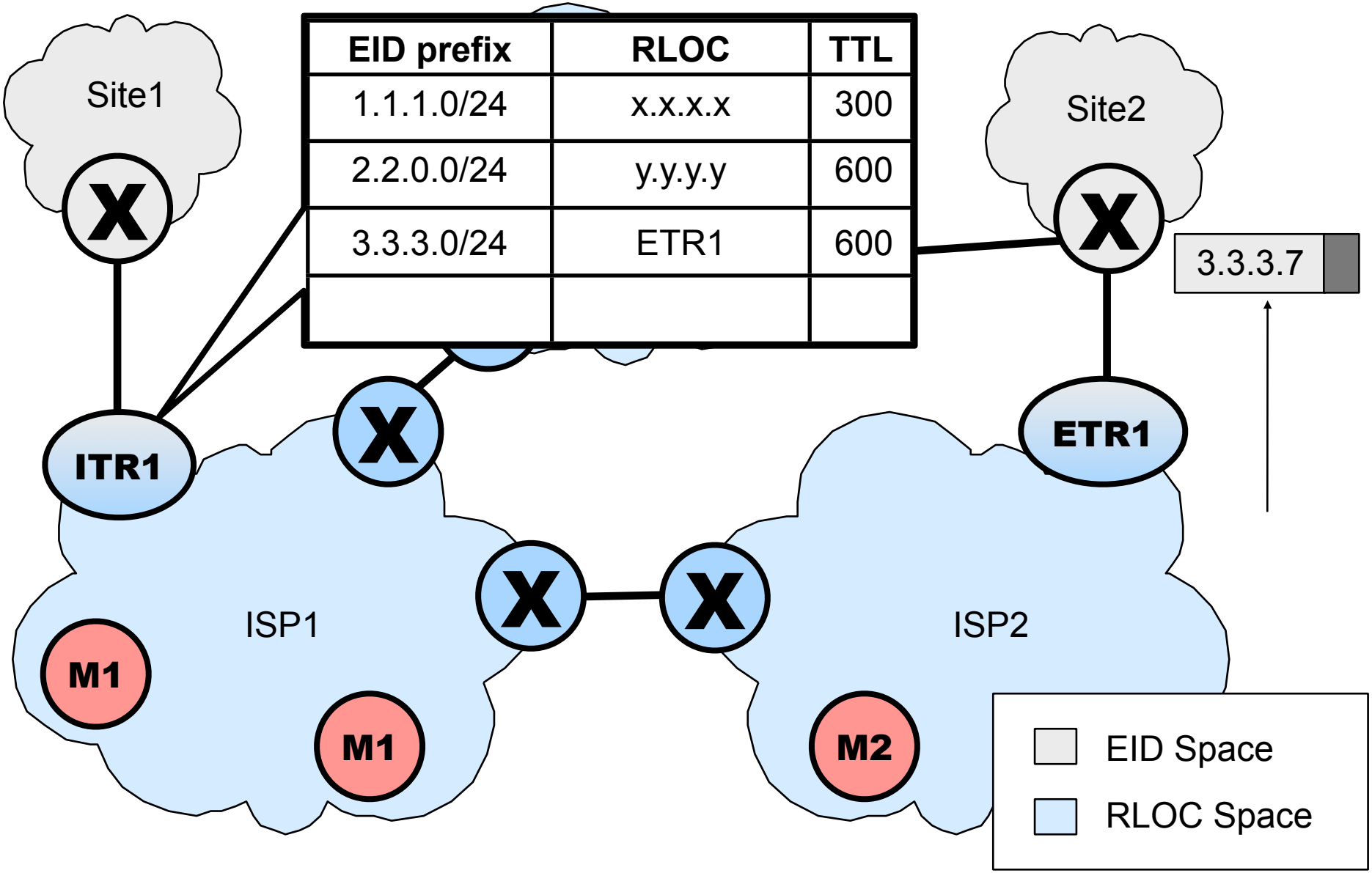
Default Mapper Responds with MapRec and Delivers Packet



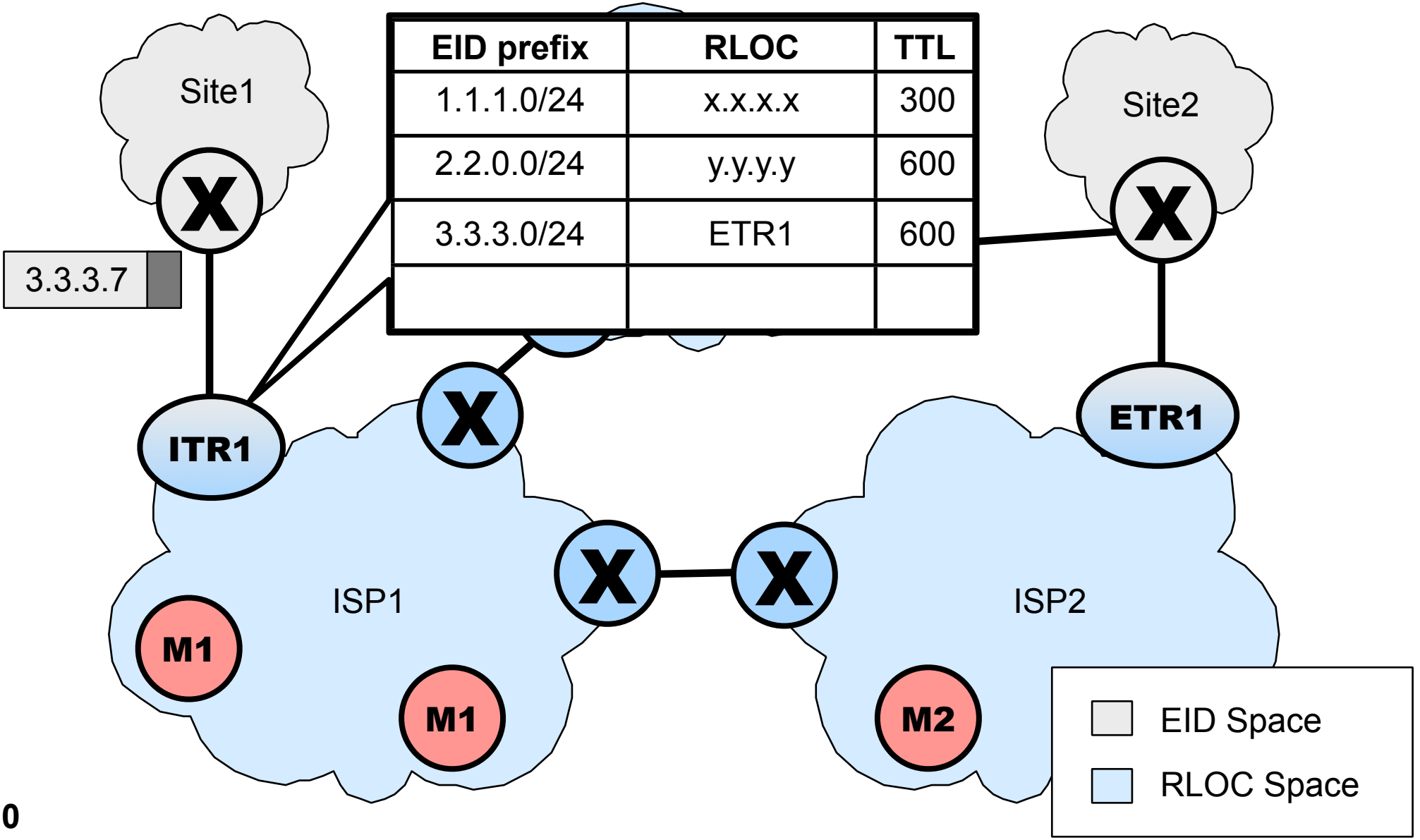
MapRec Added to Cache



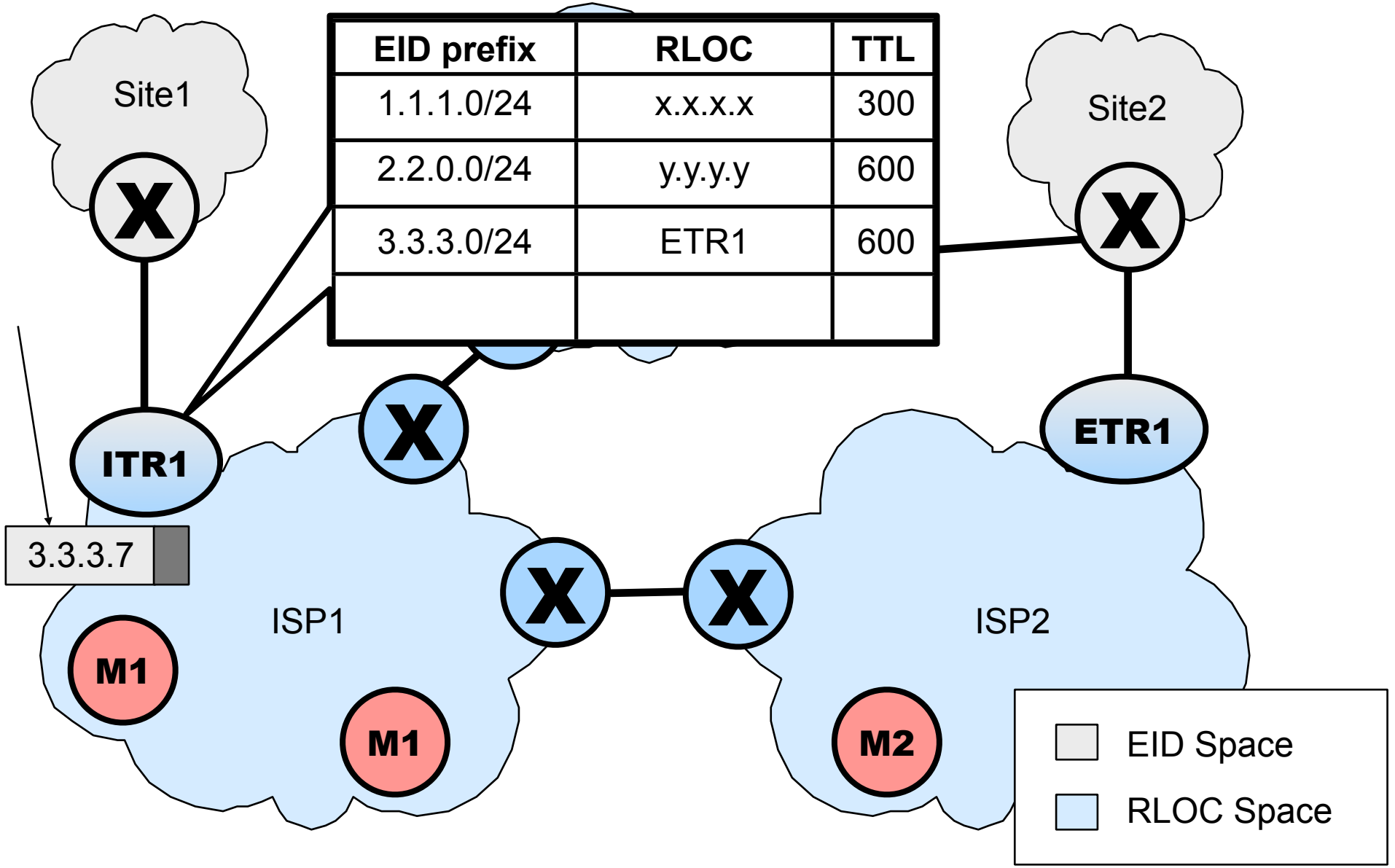
Packet Decapsulated and Delivered



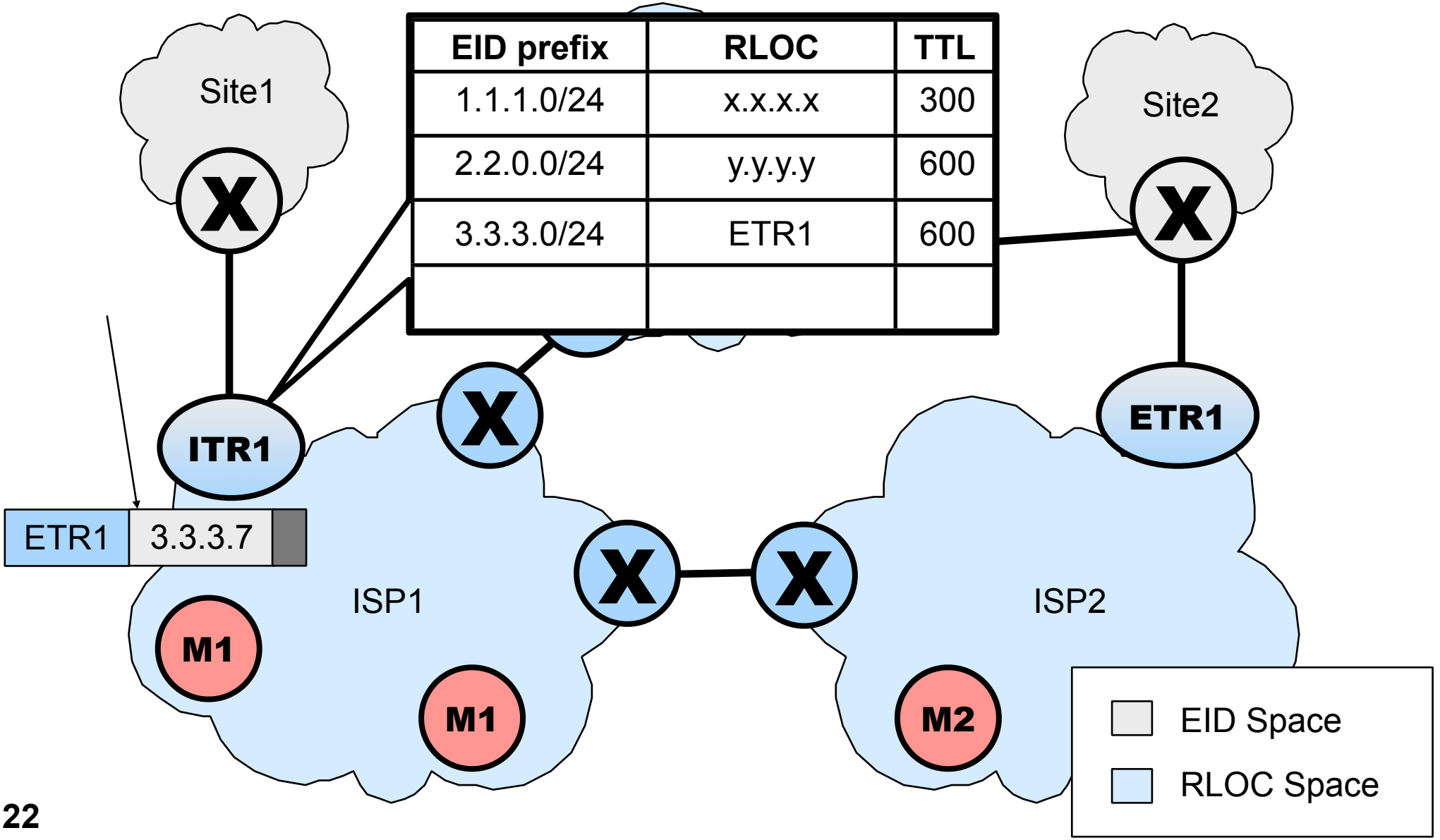
Next Packet


Site1
Site2
3.3.3.7
EID prefix
RLOC
TTL
1.1.1.0/24
x.x.x.x
300
2.2.0.0/24
y.y.y.y
600
3.3.3.0/24
ETR1
600
ITR1
ETR1
ISP1
ISP2
M1
M1
M2
EID Space
RLOC Space

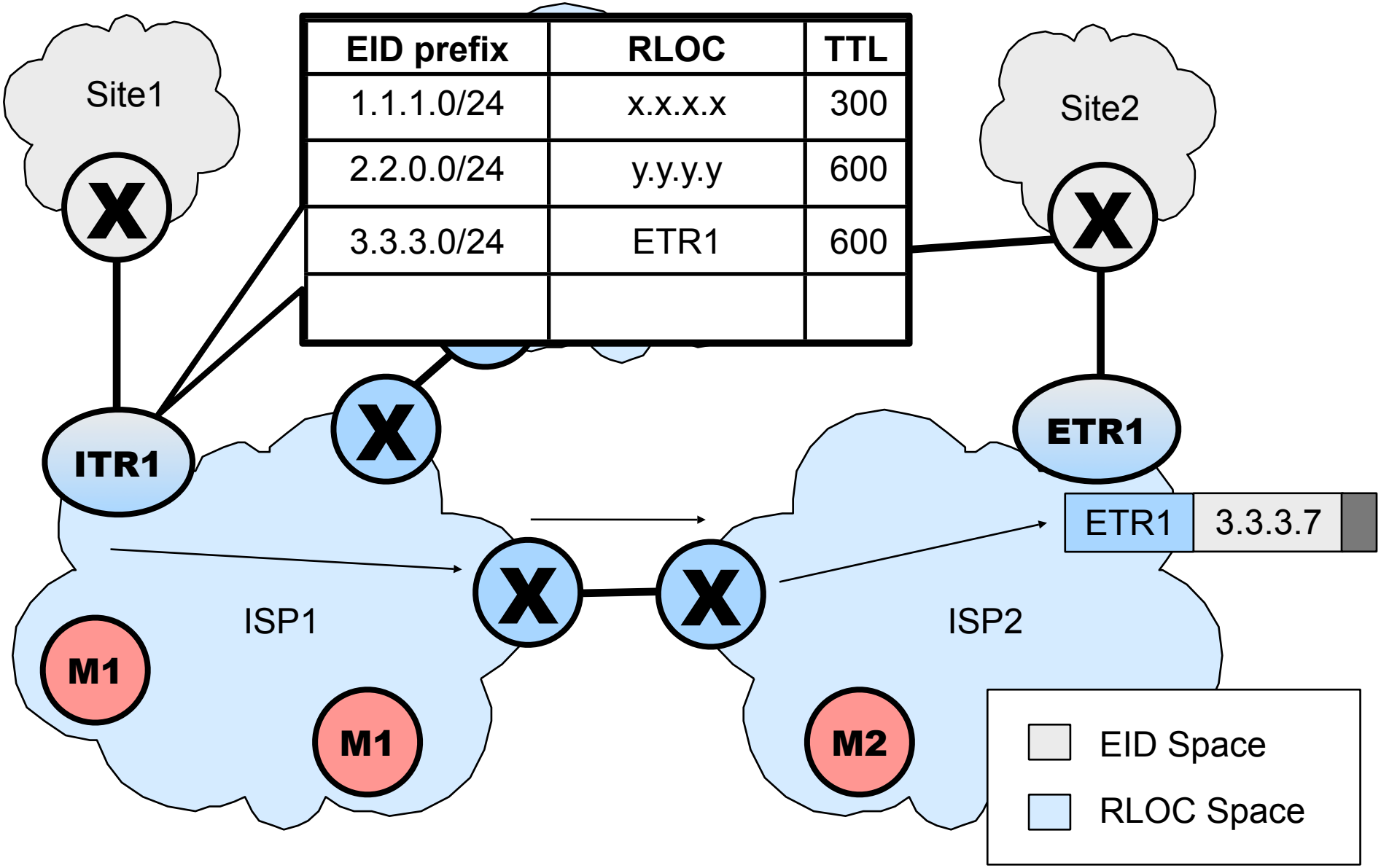
MapRec Already in Cache



Packet Encapsulated by ITR



Packet Delivered Directly to ETR



Part A2
Handling Transient Failures

Handling Transient Failures

- Three transient failures situations
  1. The PE fails
  2. The CE fails
  3. The CE-PE link fails
- Two parts to handling transient failures
  A. Handling packets in transit (to the unreachable destination)
  B. Notifying ITRs

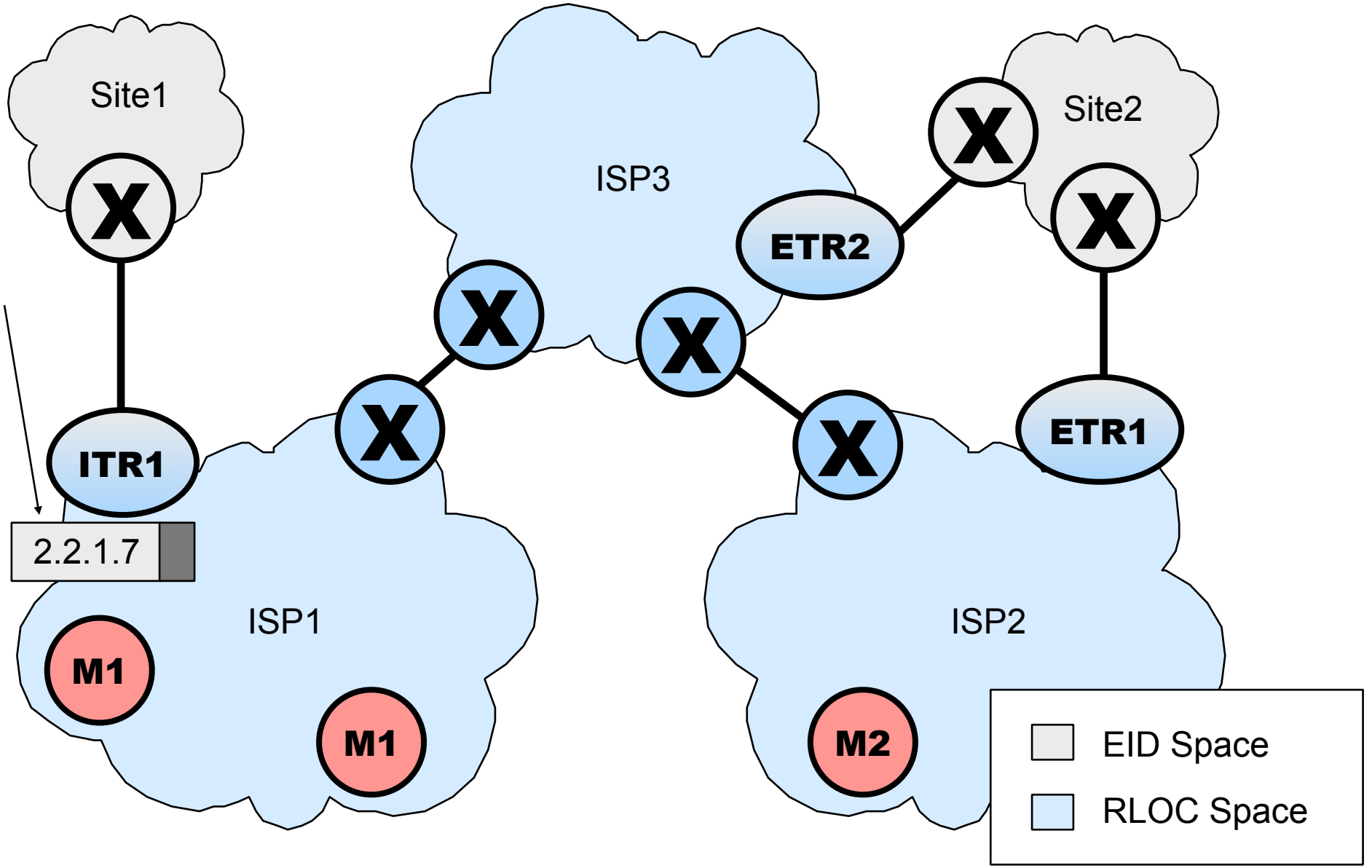
Part A: Handling Packets in Transit

- In LISP
  - Packets in transit are dropped
  - But reachability information is aggressively pushed to ITRs
- In APT
  - Recall our design philosophy: try to prevent packet loss
  - Packets in transit are rerouted by default mappers

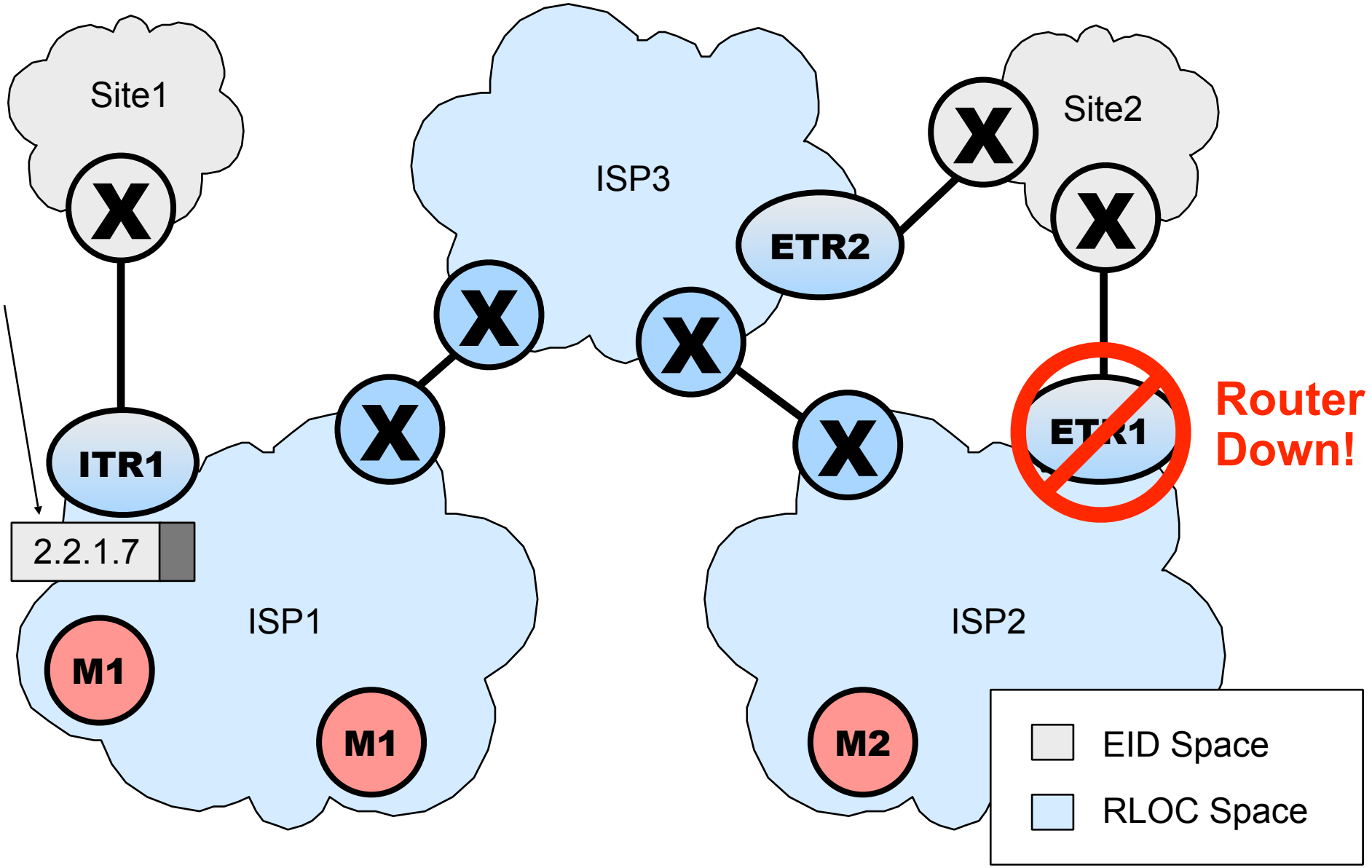
Part B: Notifying ITRs

- In LISP
  - Reachability state is stored in ITRs
  - Reachability is learned via the "Locator Reachability Bits" field (Loc-Reach-Bits) in all data and control packets
- In APT
  - Reachability state is managed by default mappers
  - Default mappers provide a reachable MapRec to TRs
  - Reachability is learned via data-triggered control messages

Example


Site1
ISP3
Site2
ETR2
ITR1
ETR1
2.2.1.7
ISP1
ISP2
M1
M1
M2
EID Space
RLOC Space

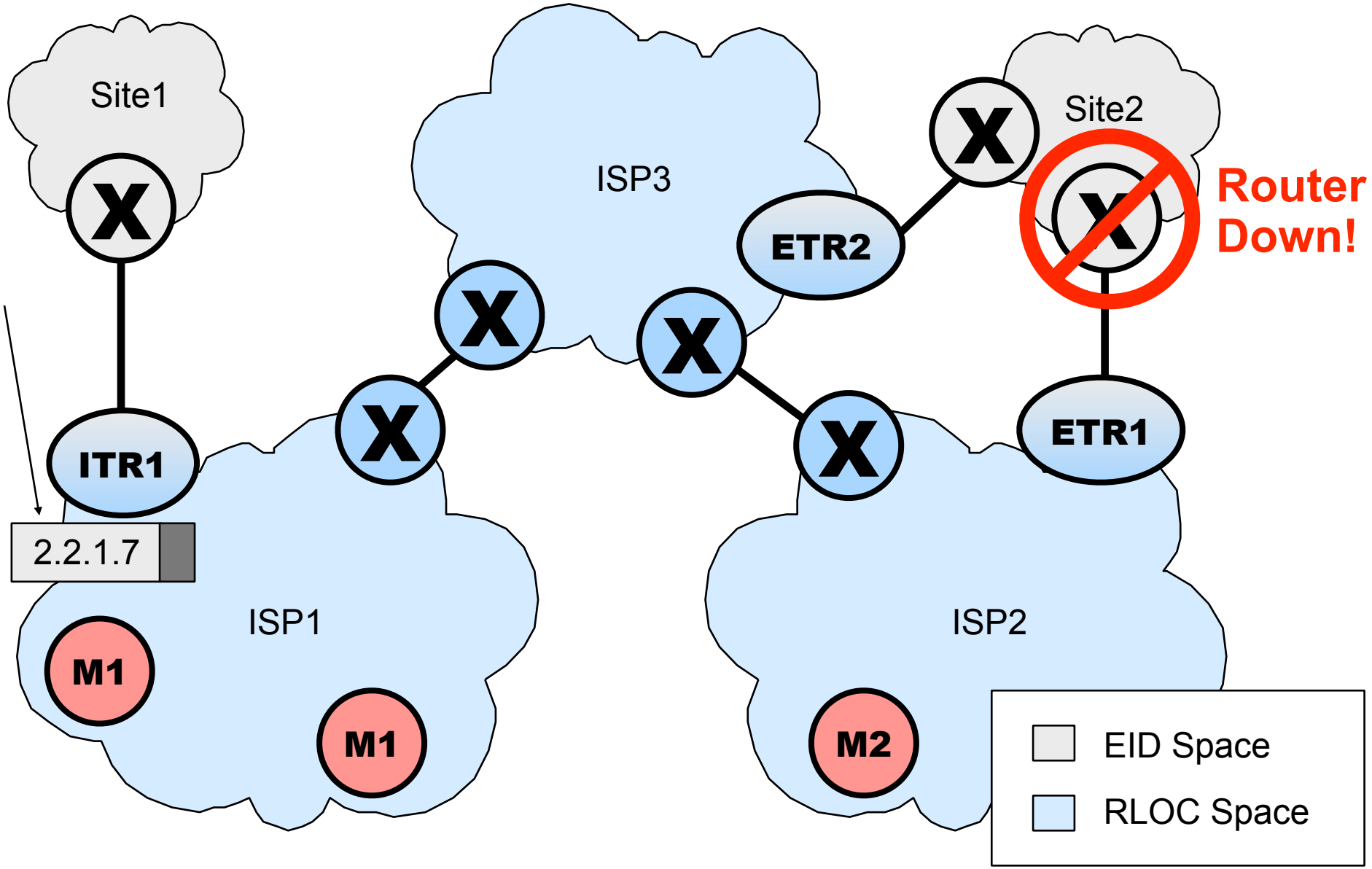
Situation 1 Example



Situation 1: PE Failure

- ETR1's default mapper (M2)
  - Has injected max-length paths to all ETRs into ISP2's IGP
  - Determines that the destination RLOC of the packet is a TR
  - Temporarily marks ETR1 as unreachable in Site2's MapSet
  - Sends the packet to an alternate ETR (ETR2)
  - Notifies one of ITR1's default mappers (M1)
- ITR1's default mapper (M1)
  - Also temporarily marks ETR1 as unreachable
  - Sends a Cache Drop Message to all of its TRs

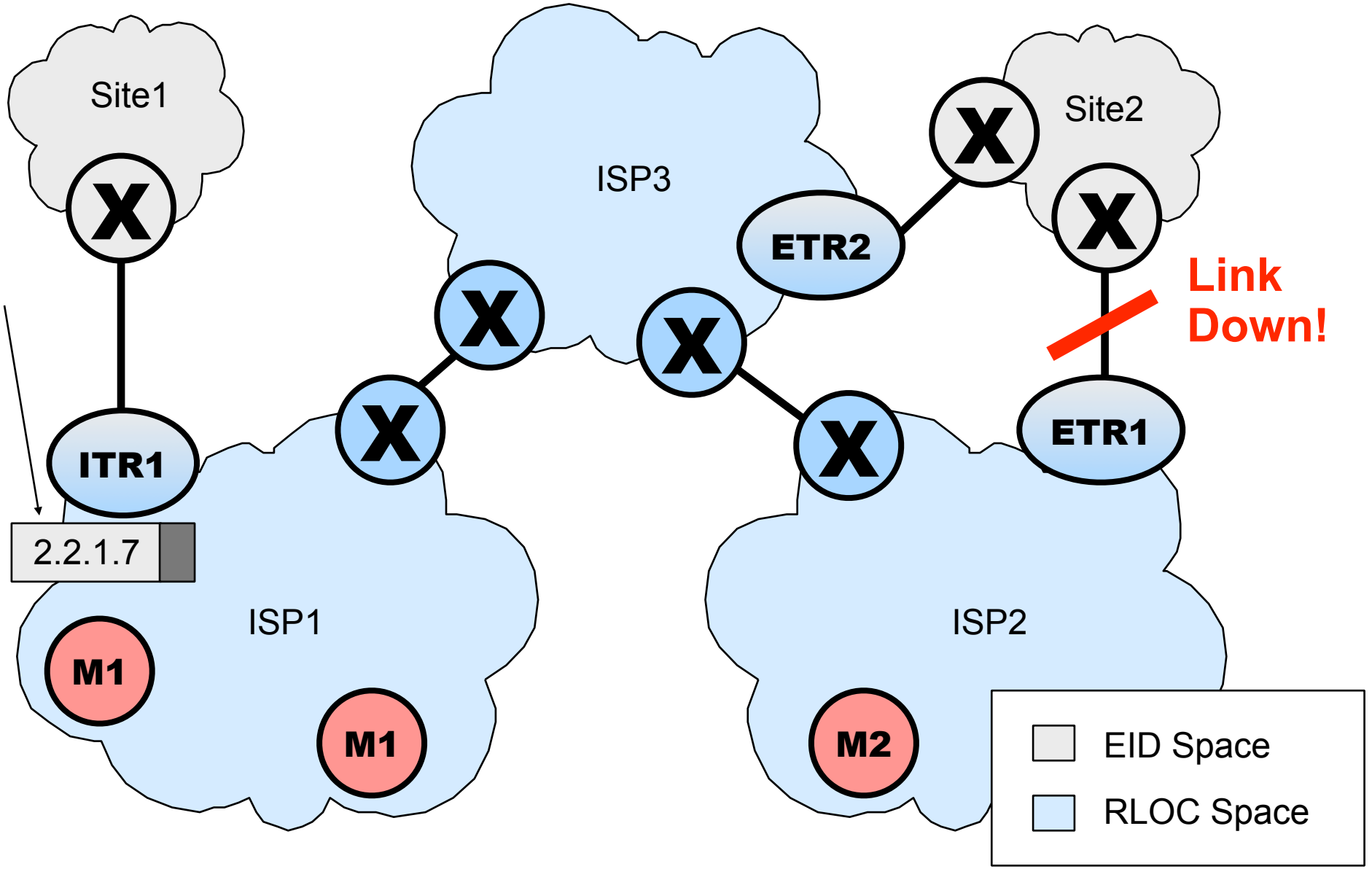
Situation 2 Example



Situation 2: CE Failure

- ETR1
  - Detects that the CE has failed
  - Forwards the packet to M2, setting a failure flag
- ETR1's default mapper (M2)
  - Same procedure as in Situation 1
- ITR1's default mapper (M1)
  - Same procedure as in Situation 1

Situation 3 Example



Situation 3: CE-PE Link Failure

- Looks just like a CE failure
- Handled in exactly the same way
  - See Situation 2

Part A3
Mapping Dissemination

Mapping Dissemination

- Default mappers need to learn other ASes' mapping information
- Mapping information is exchanged via DM-BGP
  - Separate BGP sessions running on a different TCP port
  - Only default mappers peer via DM-BGP
  - A new attribute carries one or more MapSets
  - DM-BGP does not create a routable topology
    - This is different from LISP-ALT

Security for Mapping Announcements

- Authentication of mapping information is critical
  - False MapSets could cause major problems
- Mapping announcements must be cryptographically signed by the originator
  - The signature must be verified at each DM-BGP hop
    - But not changed
  - Prevents spoofing, corruption, and alteration of mapping information
  - See our draft or ask us for details

Part A4
Incremental Deployment

Incremental Deployment

- Recall: APT is an ISP-centric design
- ISPs can become APT "islands"
  - ISPs can encap/decap right away within their AS
- Neighboring islands can merge to form larger ones
- Communication with non-APT sites
  - Packets are encapped/decapped as they pass through the island
  - The details are a work-in-progress

Part B1
ISP-based vs. End-site-based Deployment

ISP vs. End-site Deployment

- Potential incentives for ISP deployment
  - Smaller internal routing tables
  - Offer PI addressing to customers without affecting routing scalability
- Potential incentives for end-site deployment
  - PI addressing without depending on ISPs
  - Reliable, source-specific ingress traffic engineering (TE)

Benefits of Partial Deployment: APT

- First-mover APT ISPs can benefit
  - Can deploy unilaterally
  - Can remove customers' EIDs from internal routing tables
    - Though APT islands of one could also get this benefit from MPLS
- Partial deployment
  - ISPs join to form larger islands
  - ISPs can remove all EIDs serviced by their island from internal routing tables

Benefits of Partial Deployment: LISP

- First-mover LISP end sites
  - To get PI addressing without depending on ISPs
    - Will they lose connectivity to legacy networks?
    - Or need to depend on the ISP to provide a LISP proxy tunnel router?
- Partial deployment
  - Reliable, source-specific ingress TE
    - As long as the source is a LISP site

Part B2
Local vs. Remote Mapping Pull

Local vs. Remote Mapping Pull

- APT uses local pull to retrieve mappings
  - ITRs can obtain mappings quickly
  - Large overhead to distribute mapping changes
  - Default mapper storage requirements may be significant
- LISP uses remote pull to retrieve mappings
  - ITRs may obtain mappings only after a significant wait
  - No need to distribute mapping changes
  - Storage requirements should be minimal
- Empirical evaluation is needed to quantify these differences

Part B3
Flat vs. Hierarchical Mapping Retrieval Infrastructure

Flat vs. Hierarchical Mapping Retrieval Infrastructure

- The APT mapping retrieval infrastructure is flat
  - A copy of the global mapping table is stored at every default mapper
  - Changes must be replicated in all default mappers
- In LISP-ALT, the mapping retrieval infrastructure is hierarchical
  - Structurally similar to DNS
  - Higher-level ALT nodes don't maintain actual mapping information, just paths to the information

ALT Hierarchical Mapping Retrieval Infrastructure: Deployment Issues

- Assumes that EID prefixes are aggregatable
  - This means strict peering rules (unrelated to topology)
- Assumes that some sites will be willing to host higher-level ALT nodes
- Deployment depends on how realistic these assumptions are

Thank You!

- Questions?
- Comments?

APT TRs in CEs

- Only minor differences from PE TRs
- CE TRs get one provider-assigned RLOC per ISP
  - All of which appear in the MapSet for the site's EID prefix(es)
- For transient failures
  - ISPs have CE TRs' RLOCs in their IGP
  - Situation 3 (CE-PE link failure) becomes the same as Situation 2 (ETR RLOC unreachable)