LISP Replication Engineering
coras-lisp-re-00

LISP WG, IETF-84, Vancouver

Florin Coras, Albert Cabellos and Jordi Domingo
Universitat Politècnica de Catalunya

Fabio Maino and Dino Farinacci
Cisco Systems

Context

- Need multicast for efficient one-to-many packet delivery
- Existing solutions
  - IP-multicast
  - Application layer multicast

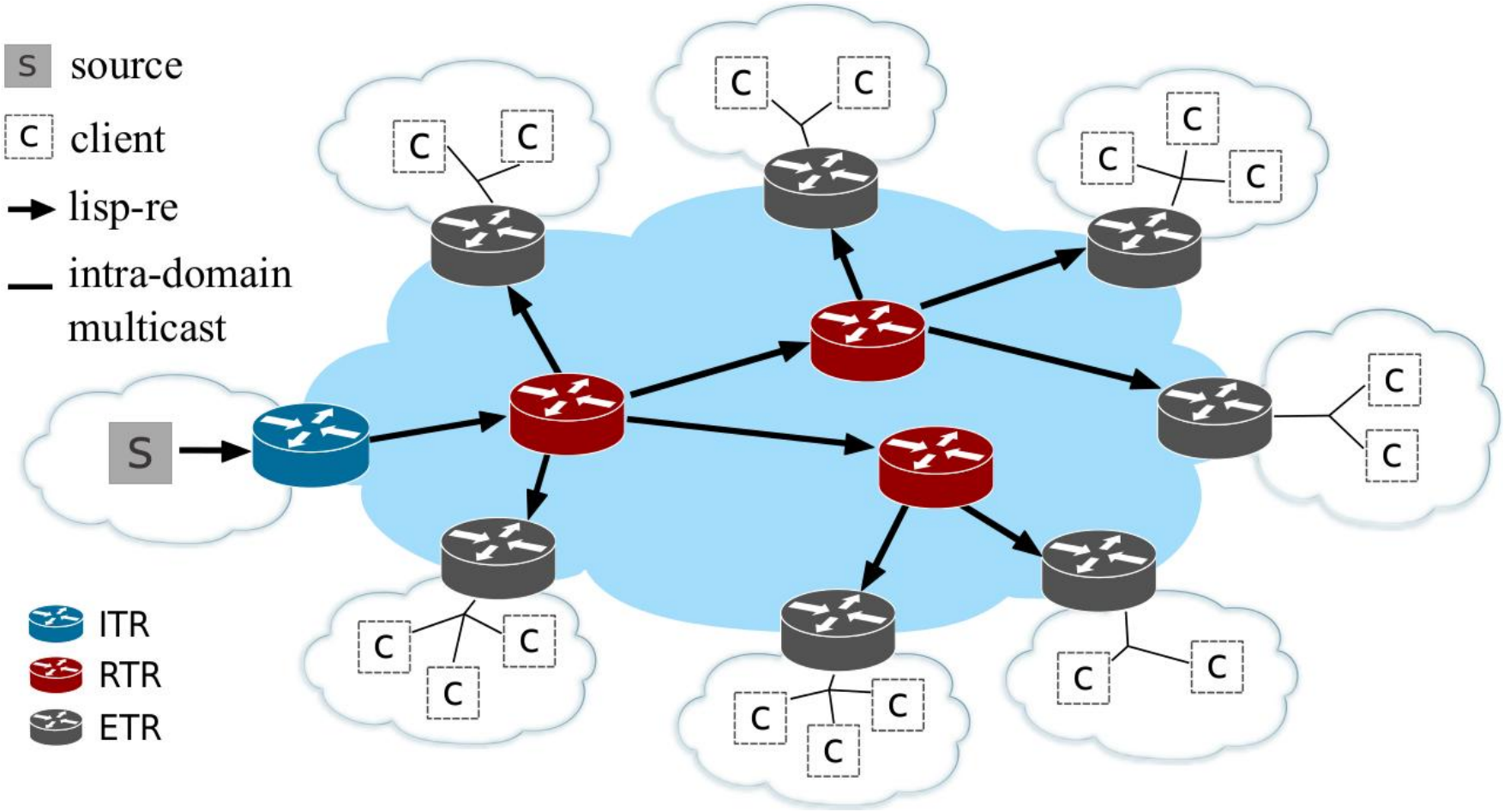
LISP and Multicast

- LISP-Multicast assumes the existence of inter-domain IP-multicast
- An alternative would be to perform unicast encapsulated replication
  - Scalability is a serious concern due to high head-end replication (at the ITR)

LISP Replication Engineering

- Use RTRs to offload the replication load of the ITR
  - Organize RTRs in a distribution tree
  - RTRs perform unicast or multicast RLOC encapsulated replication of EID multicast packets
- Group management functions are centralized in a Distribution Tree Coordinator (DTC)
  - It can either be the ITR or an external orchestration system

Data Plane Architecture



LISP Replication Engineering

- Group management
  - Procedures and objects needed for building (S-EID,G) map-cache state in ITR, RTR and ETR
  - Protocol mechanism used for communication are defined in LISP-Multicast (PIM) and farinacci-lisp-mr-signaling
- Distribution tree optimization
  - Optimization algorithm
  - Topology discovery

Distribution Tree Management (1)

- LISP Replication Node Database (LRND)
  - Maintained by the DTC per (S-EID,G) channel
  - Stores the state of the distribution tree
    - RLOCs of ITR, RTRs, ETRs
    - **Replication list** that specifies the child list of a member
    - Replication capacity of RTRs (out-of-band signaling)
    - Optionally, more information about the members needed for tree optimizations conveyed by means of out-of-band singnalling

Distribution Tree Management (2)

- Join Procedure
  1. The joining node sends a Map-Request/Join-Request for (S-EID,G) to the DTC
  2. If the request is for multicast replication LISP-multicast procedures are followed and no further steps are taken
  3. If unicast replication is requested, the RTR finds a distribution tree parent for the joining node, either randomly or by using an heuristic.
  4. The ITR updates the replication list of the selected parent to include the new child.
  5. The selected parent updates its mapping after it receives an SMR from the ITR and starts replicating content to the newcomer
  6. DTC Map-Replies with the destination EID-prefix set to (parent-RLOC, ETR-RLOC)

Distribution Tree Management (3)

- In case of graceful member departures
  - For ETR departure, DTC updates parent state
  - For RTR departure, DTC update parent and children state
    - Find new parents for ‘orphaned’ children
    - Use make-before-break procedure to avoid packet loss
- In case of member failure
  - Detect failure and inform DTC
  - After being informed, the DTC acts like in the event of a graceful departure

Distribution Tree Optimizations

- Optimization Algorithm
  - What it optimizes depends on operational requirements
  - The document provides an algorithm as example
- Topology Discovery
  - Active or passive measurement of the overlay topology
  - Precomputed network maps like the ones provided by iPlane [iplane] and/or out-of-band signaling

Optimization Algorithm (1)

- We set as goal the delivery of delay sensitive content
  - Minimize the distance (latency, AS-hops) between receiver end-hosts and the ITR
- The heuristic builds a spanning tree, starting at root
  - At each step it adds to the tree the member with the smallest distance to the ITR per multicast receiver

Optimization Algorithm (2)

- Simulation Results
  - Control overhead is easily manageable
  - Client churn slightly influences performance
    - Increases management overhead
  - Fan-out influences performance logarithmically
    - Fan-out values larger than 6 offer limited benefits

Still need discussion

- Topology discovery
  - The choice is administrator dependent but for now we don't have a protocol for conveying active measurements results
- How to use an orchestration system to program the mapping system with ELPs describing the distribution tree topology
- Avoid packet loss when
  - Optimizing the distribution tree
  - Member departs

Questions?

Backup Slides

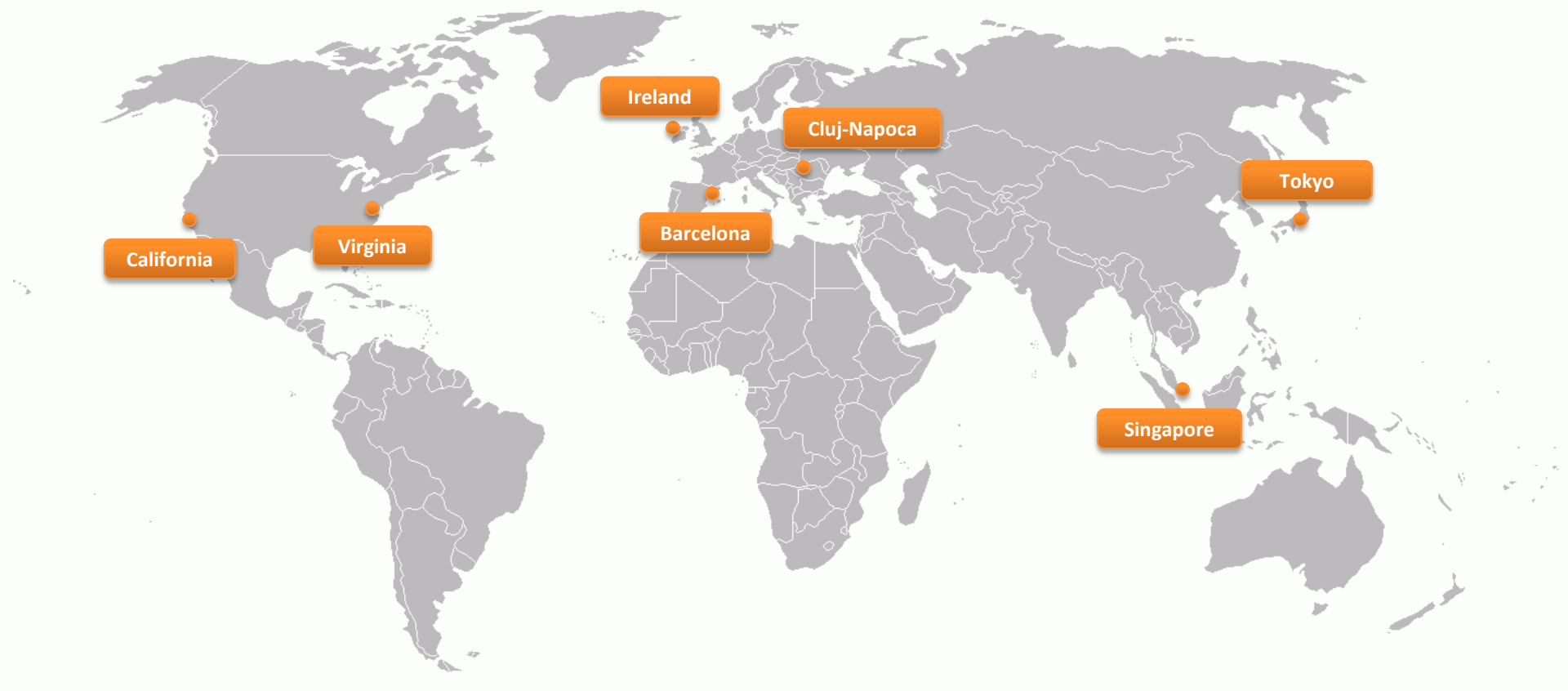
Evaluation (1)

- ITR simulator that uses an Internet-like AS topology and 3 generated traces of client arrivals and departures
- Internet-like AS topology
  - We aggregated topology information from: CAIDA, RouteViews, RIPE, iPlane
  - Latency information from iPlane

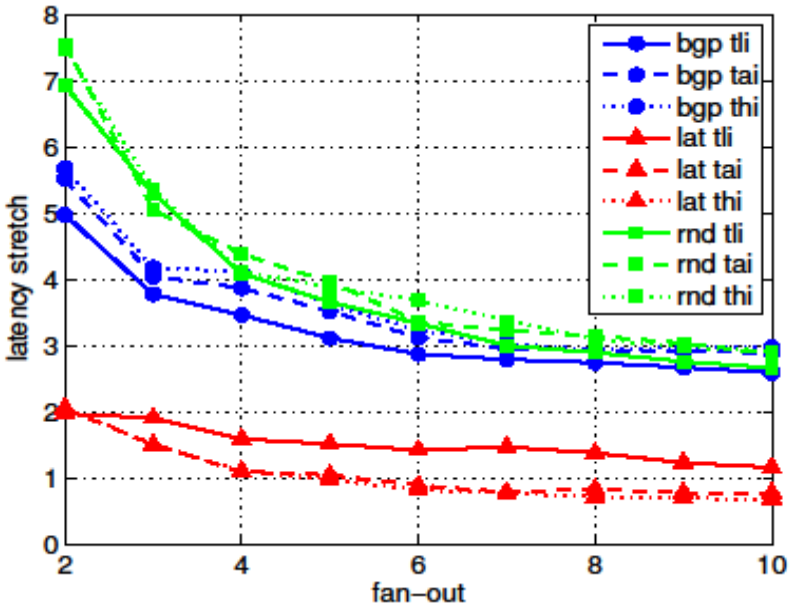
Evaluation (2)

- Generated traces of client arrivals and departures
  - Passively captured P2P TV traces to understand client clustering patterns
  - 146k unique IPs in 3.8k ASes
  - 3 traces of low, average and high client interest in streamed content: tli, tai, thi. They have respectively high, medium and large client churn.
- Topology discovery
  - **bgp**: ITR makes use of the BGP RIB at the source domain's xTR to infer the number of AS hops between members
  - **lat**: ITR requests nodes to measure their latencies to a subset of peers according to an heuristic

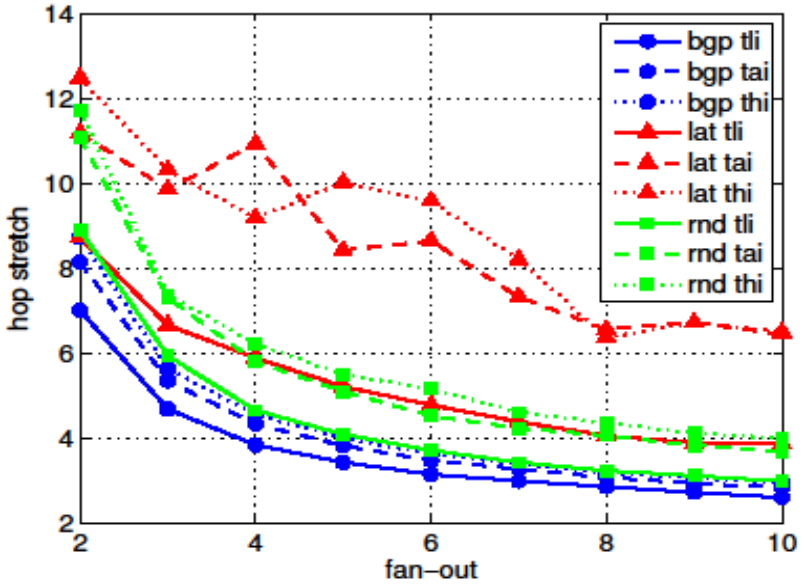
P2P TV Traces Capture Points



Results (1)

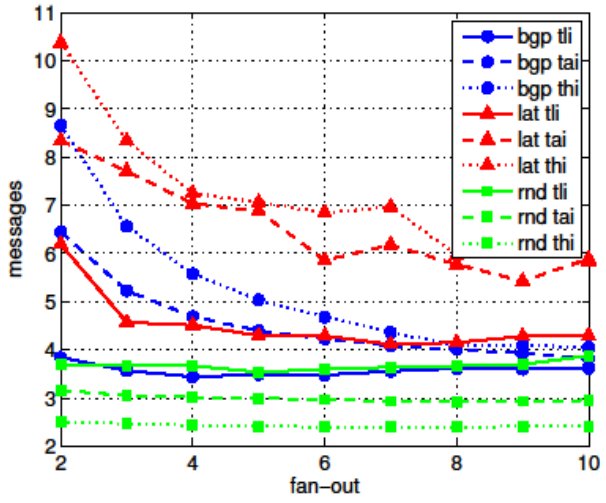


Latency stretch

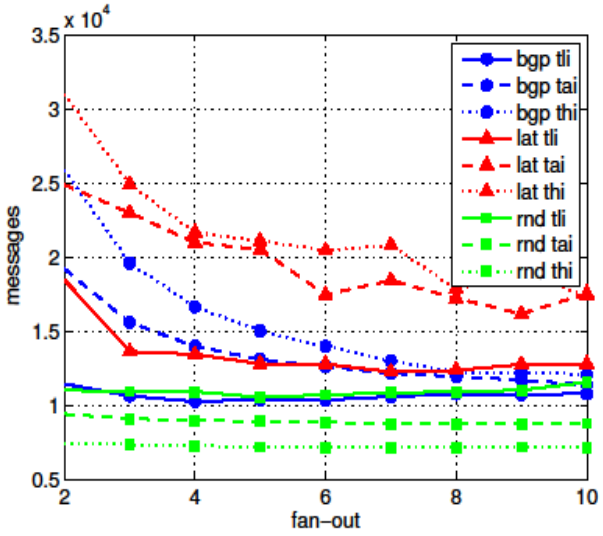


Hop stretch

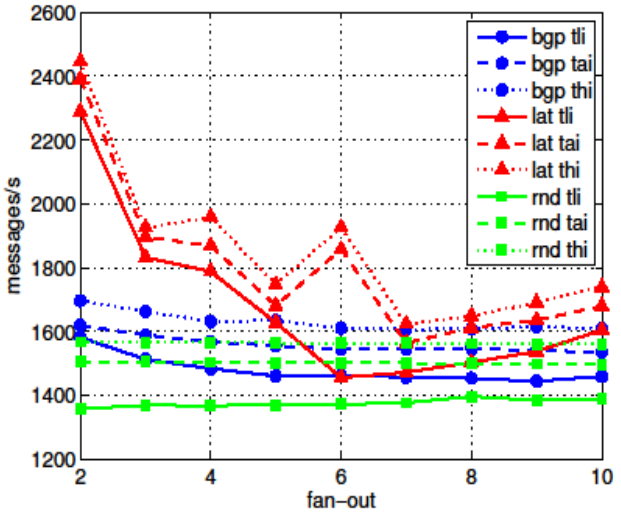
Results (2)



Av number of messages/ETR



Av number of messages/ITR



Peak messages/s for the ITR