Mangrove: Network Visibility Across the Internet

Joseph Zhang, Bradley Lewis, Alex Shi, Y. Richard Yang, Guo Dong, Daniel Mertus

IETF 119
March 19, 2024
Australia

Overview and Current Challenges

Goal: Create an indexing engine for the control plane of the Internet

What path do packets take from some address A to some address B?
How can I visualize connections between networks across the internet, at varying granularities?
How may I analyze network performance **at the edge** with limited monitoring capabilities?
How can I get highly dynamic updates on latency and jitter between two devices that communicate across the internet?

Challenge	**Techniques to solve**
Heterogeneous data sources	Unified data representation format (MLFR)
Data scale	Prefix-based compression and database sharding
Partial information	Prolog-style logical conflict resolution with business logic, differing levels of granularity
Real time, dynamic updates	Pre-aggregated/historical data, update bus

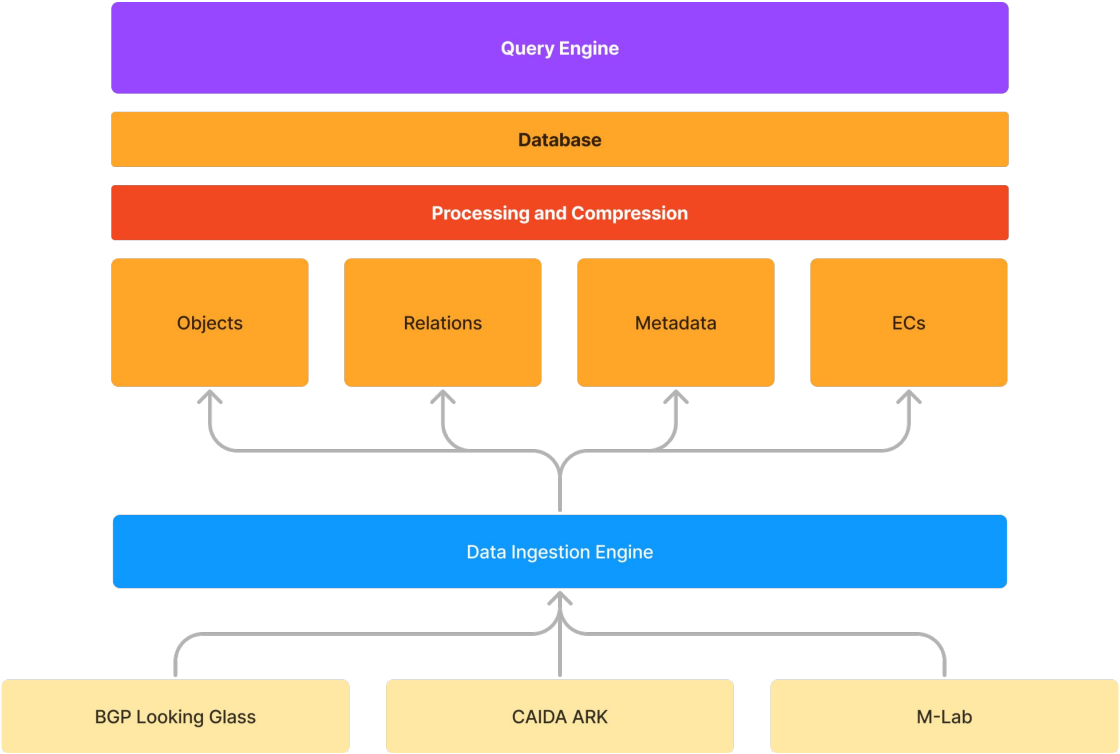
Existing Data Sources

Source	Description
CAIDA Ark	Hosts measurement infrastructure across distributed Ark nodes and publishes traceroute data across CAIDA domain.
RIPE Atlas	Deploys measurement probes to perform continuous measurements across RIPE administrative domain.
BGP streams	Traceroutes of AS's through BGP routing dumps, looking glasses.
Ookla, M-Lab, perfSONAR	Provides broadband measurements and metadata about connections such as traceroute data, connection data...

Challenge 1: Data is decentralized and not indexable.

Challenge 2: Data has differing representations with no logical mapping with topology.

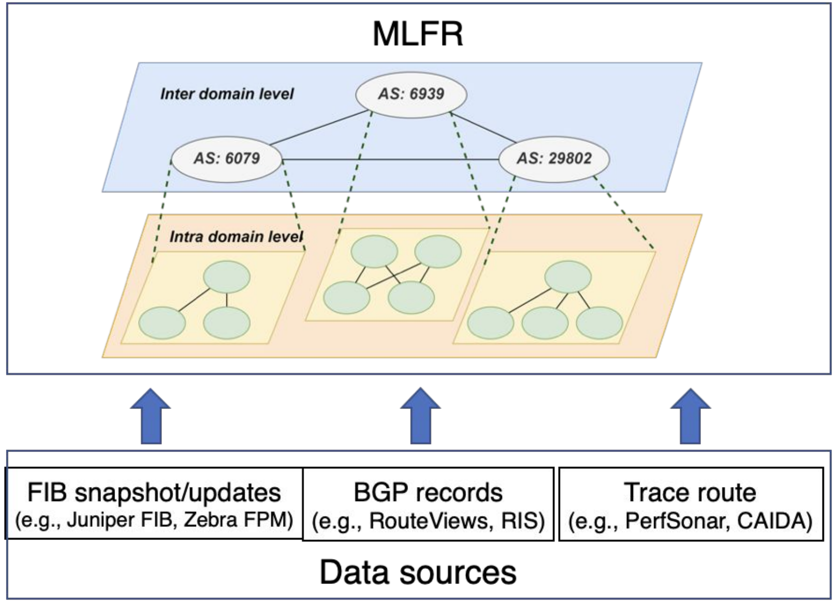
High level solution



Heterogeneous Data Sources: Multilevel FIB Representation (MLFR)

MLFR Intermediate Representation

- Network: <{v},{e},{FIB-rule}>
- v: networkNode | stubNode
- networkNode: router | AS
- e: <port, port> | <port, stubNode> | <v, v>
- FIB-rule: <ingress e, pkt-match> -> action/nexthop



Generating MLFR

BGP: rib|R|route-views.eqix|40.183.224.0/19|206.126.236.25|6079 6939 29802 293|

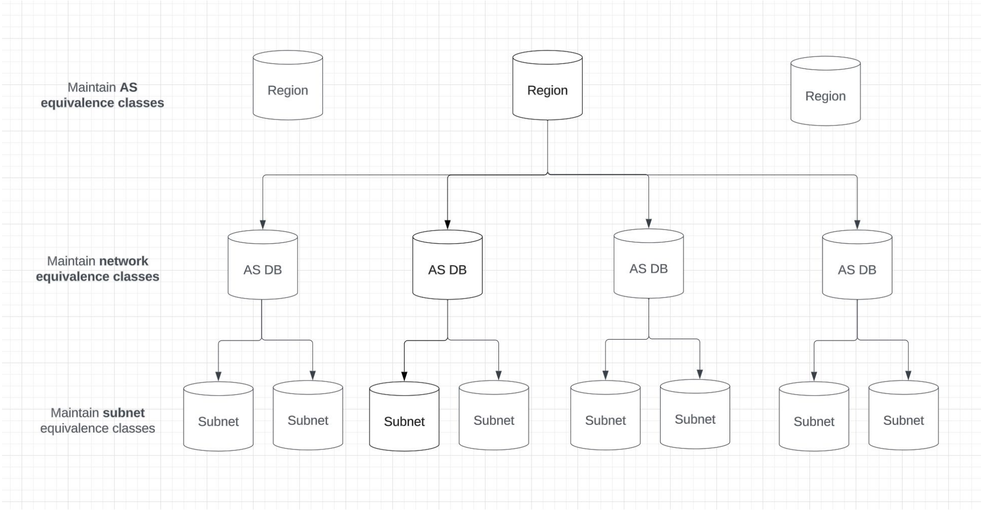
MLFR:
<(AS:6079, AS:6939), dip=40.183.224.0/19> -> AS:29802
<(AS:6939, AS:29802), dip=40.183.224.0/19> -> AS:293

Trace: [srcIP, IP1, IP2, IP3,..., dstIP]

MLFR:
<(srcIP, IP1), dip=dstIP> -> IP2
<(IP1, IP2), dip=dstIP> -> IP3
...
<(IPn, IP_{n-1}), dip=dstIP> -> dstIP

Data Scale

- Equivalence classes
  - Using flash, we can calculate **equivalence classes** for different levels of detail granularity (AS ⇒ Subnets ⇒ IPs)
  - Store the equivalence class on each shard
- Matching
  - Match the IP pair with the longest prefix we have information on
  - Reconstruct paths using that level of equivalence class granularity
- Compression
  - Equivalence classes at different granularities allow for more compressed storage
  - Sharding based on ASs allows for distributed, horizontal scaling



Effect 1: Scale down from billions of devices → millions of routers, thousands of ASs
Effect 2: Scale down storage of N^2 relationships to << N equivalence classes

Filling in the gaps: Partial information and real time data streams

Partial Information

- Using computational deduction based on a set of **business rules**, we are able to infer and consolidate partial routing/topology information within the heterogeneous data sources:
  - Make inferences of missing steps in ie. traceroutes
  - Reconcile multiple data sources e.g. bandwidth of smaller network ≤ bigger network
  - Business logic based topology reconstruction e.g. AS provider relationships
- Perform queries against the **closest known** solution:
  - Exact path match
  - Closest geographical path match
  - Longest prefix match

Real time data streams/updates

- Topology
  - Use Flash this update our inverse model of the internet easily
- Metadata
  - Map Ookla, M-Lab, perfSonar and other streams to network nodes
  - Data enrichment using IP to device resolution from CAIDA
  - Aggregate data to compute average/median data for lower levels of granularity using Kafka streams (as per the efforts with the Geant backbone vis a vis Telefonica)
- Use pre-aggregated data to verify measurements in real-time

Contact Us

Joseph Zhang
joseph.zhang@yale.edu

Bradley Lewis
bradley.lewis@yale.edu

Alex Shi
alex.shi@yale.edu

Daniel Mertus
daniel.mertus@yale.edu

Richard Yang
yry@cs.yale.edu

Guo Dong
gd@tongji.edu.cn