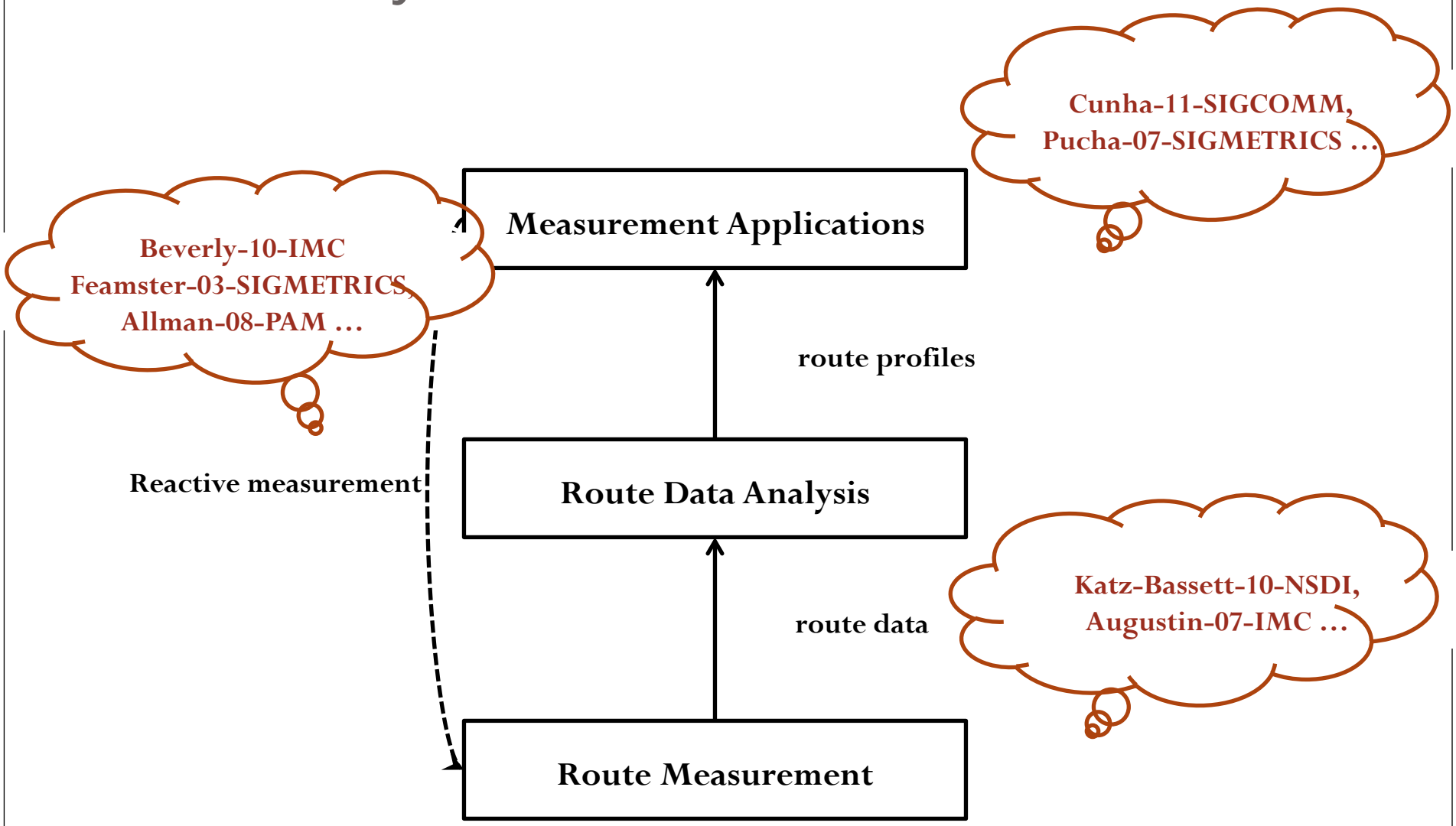
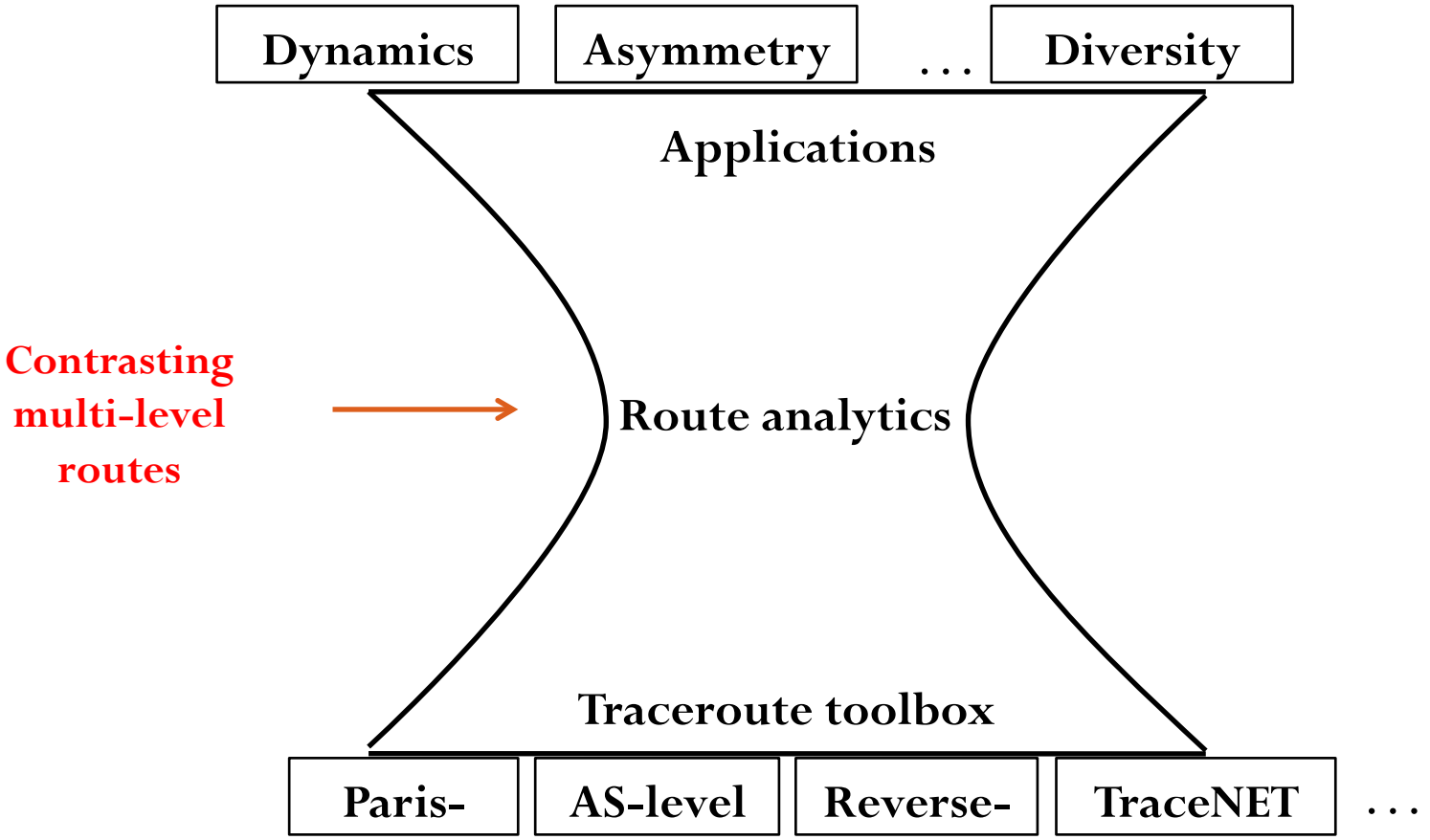
Efficient Analytics over Route Data

Ang Chen
University of Pennsylvania
AIMS-2013 @ CAIDA

Anatomy of route measurement tasks



Route analytics: the "narrow waist"



Route contrasting: an analytic primitive

- Route asymmetry
  - **AS-level asymmetry**
  - **IP-level asymmetry**
  - **…**
- Route diversity
  - **Due to multi-homed ASes**
  - **Due to overlay routing**
  - **…**
- Route dynamics
  - **History-based prediction of path changes**
  - **Correlation between path changes & delay variations**
  - **…**
- …

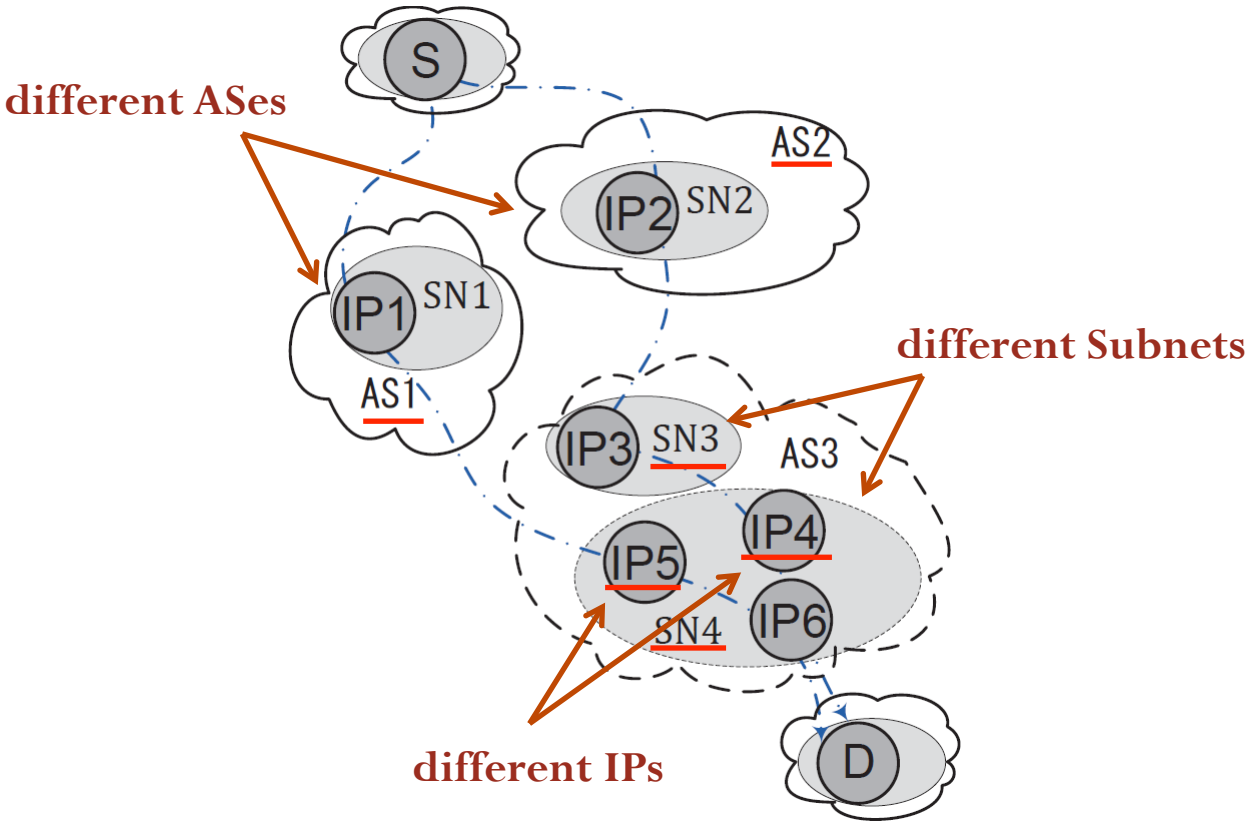
Route contrasting: challenges

- Big route data:
- Ark:
  - **500 million IPv4 traceroutes per round**
  - **10 billion IPv4 traceroutes in total**
  - **4 TB of data**
  - **10,269 routed IPv6 prefixes**
  - **wide deployment of IPv6 – many more destinations to probe …**
- ***Multi-level*** route analytics ***multiplies*** the amount of computation
  - **Contrasting IP-level routes**
  - **Contrasting Subnet-level routes**
  - **Contrasting AS-level routes**
  - **Contrasting Organizational-, PoP-, Geographic- levels of routes?**

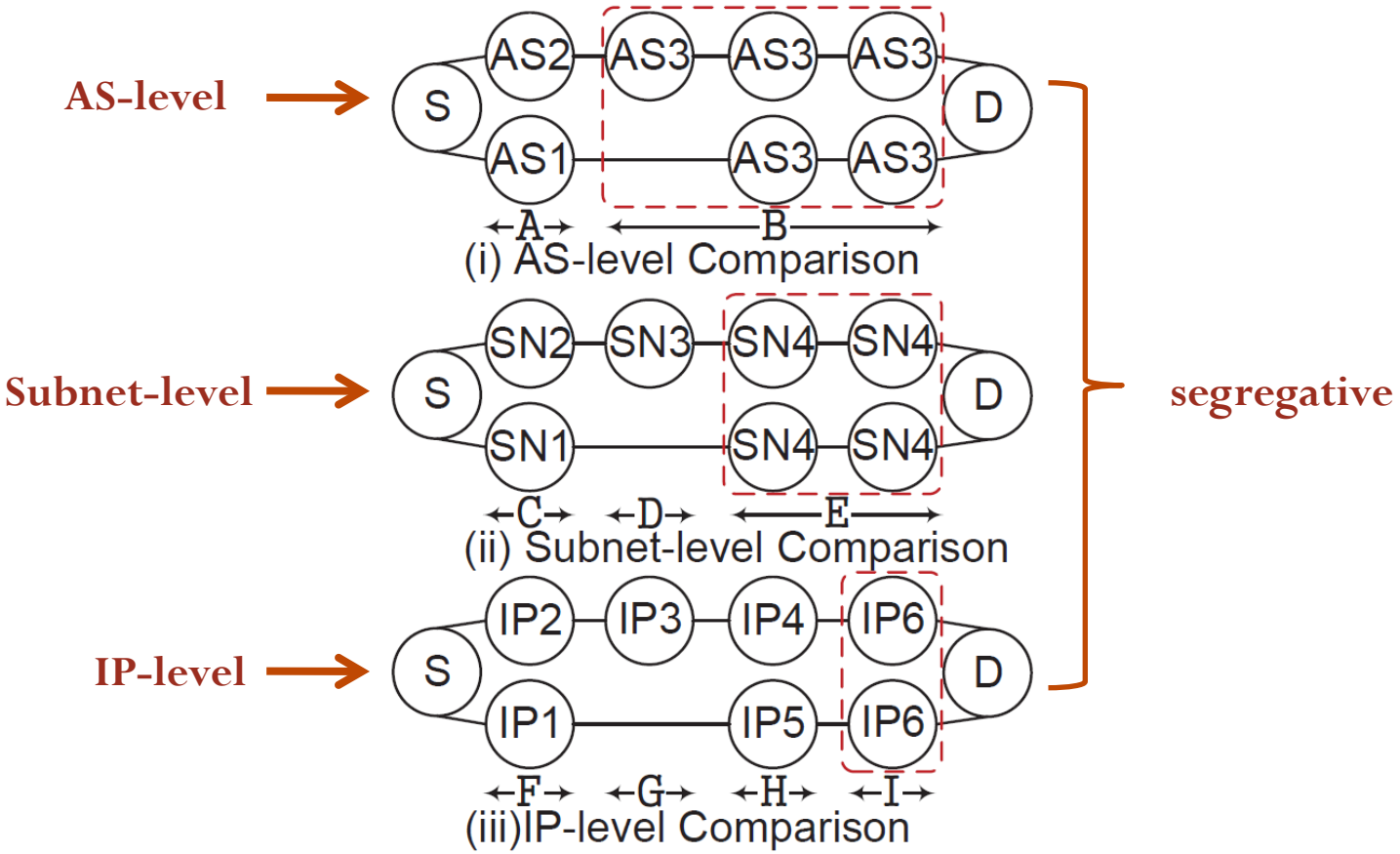
  [Cai-10-IMC, Spring-02-SIGCOMM, Katz-Bassett-06-IMC]

Multi-level route analytics

- Route differences on various levels of granularities:

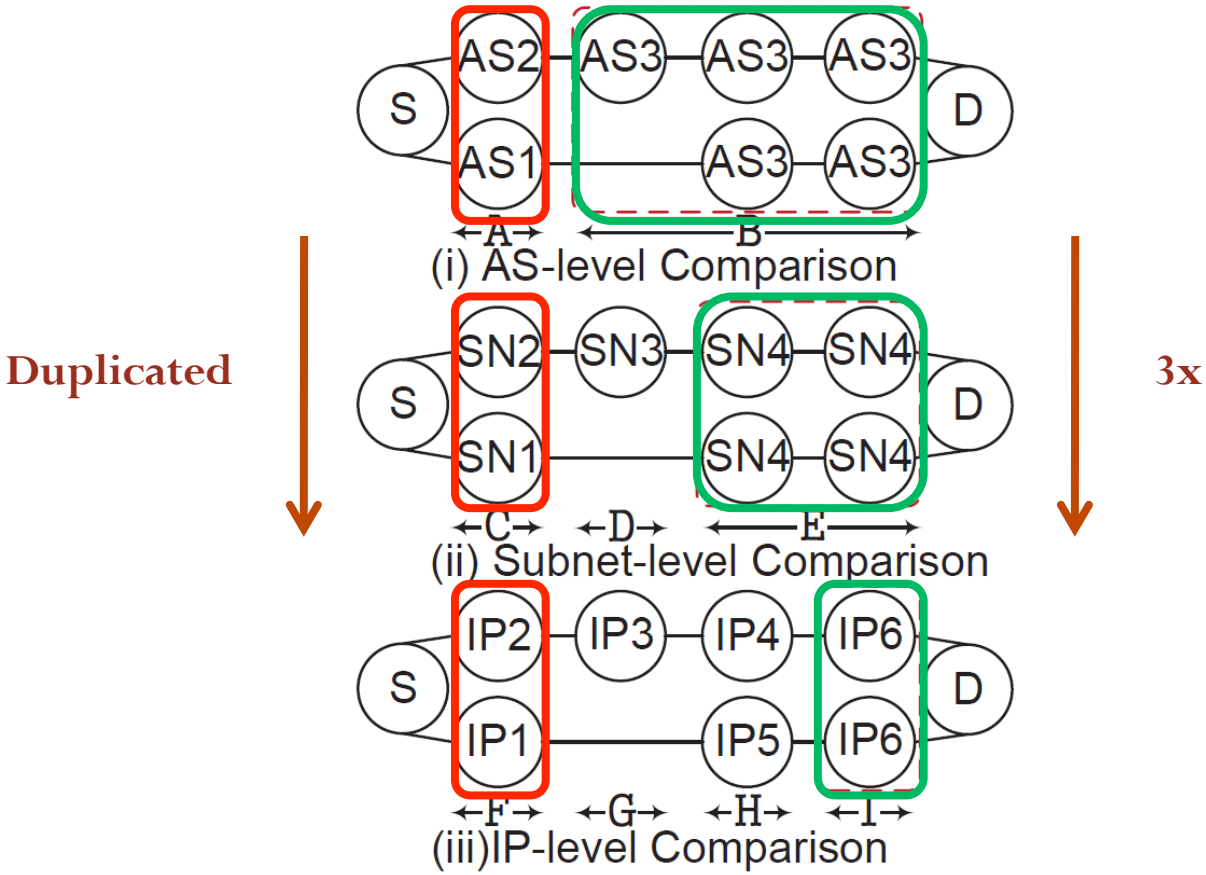


Existing approaches



Duplicated node comparisons

- Existing approaches result in duplicated computation

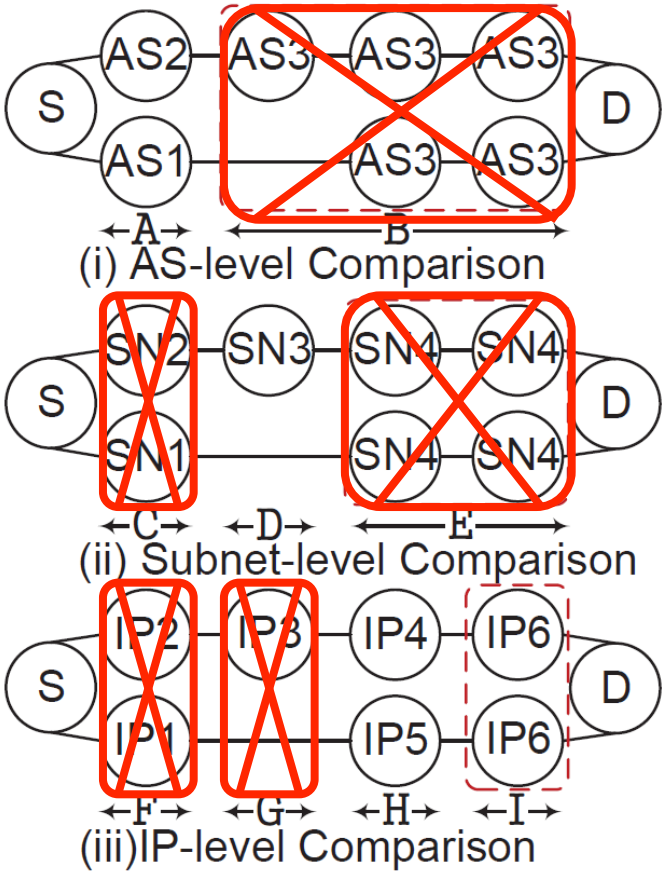


Can we do better?

- Eliminate redundant comparisons
- Reduced computational load
- Faster route analytics
- …

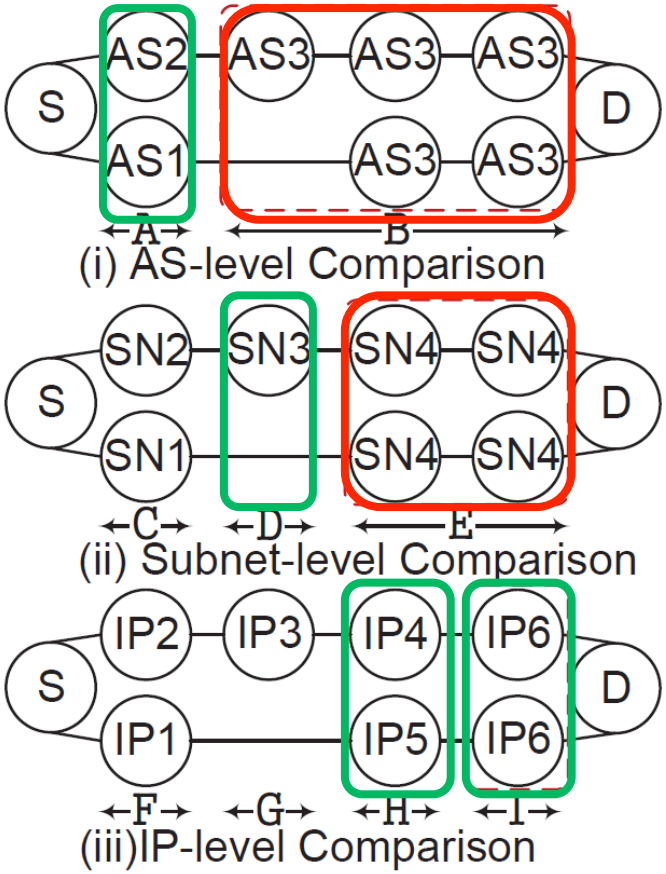
Root cause: fixed resolutions

- No resolution fits the whole paths' comparison.



Solution: dynamic resolutions

- Adjust the resolutions on different path segments.

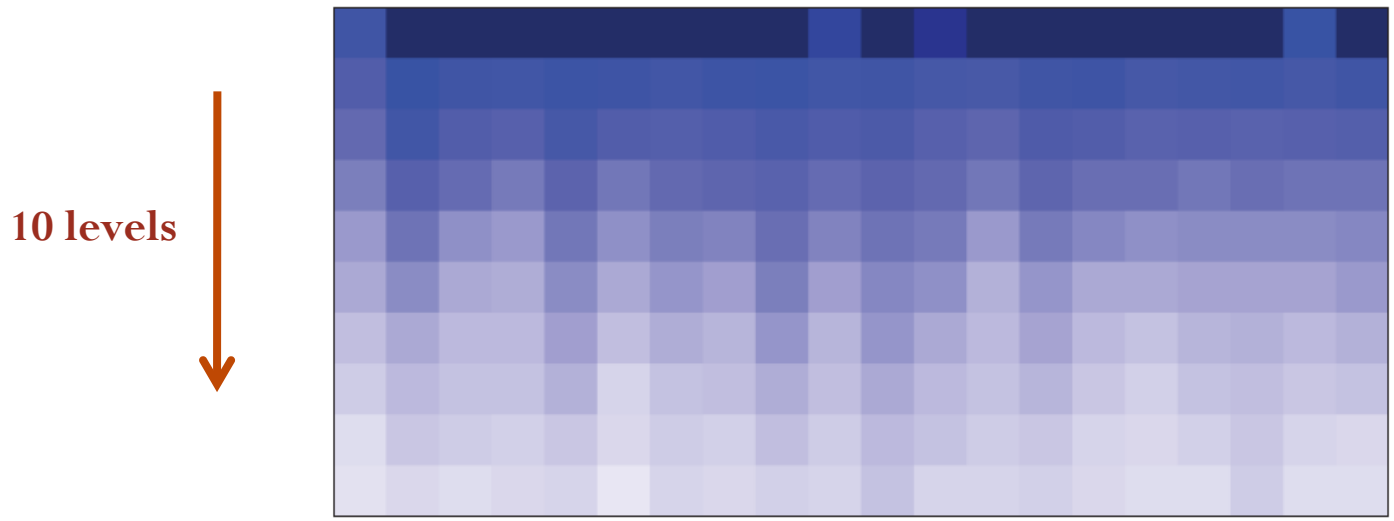


Efficiency

- For n-level route analytics:
  - Existing approaches: n times the computational load
  - New approach: constant order of load under an ideal case
- Results on 87 million routes &52 billion comparisons:
  - Ark: 86% node comparisons eliminated
  - iPlane: 85% node comparisons eliminated
  - FastMapping: 83% node comparisons eliminated

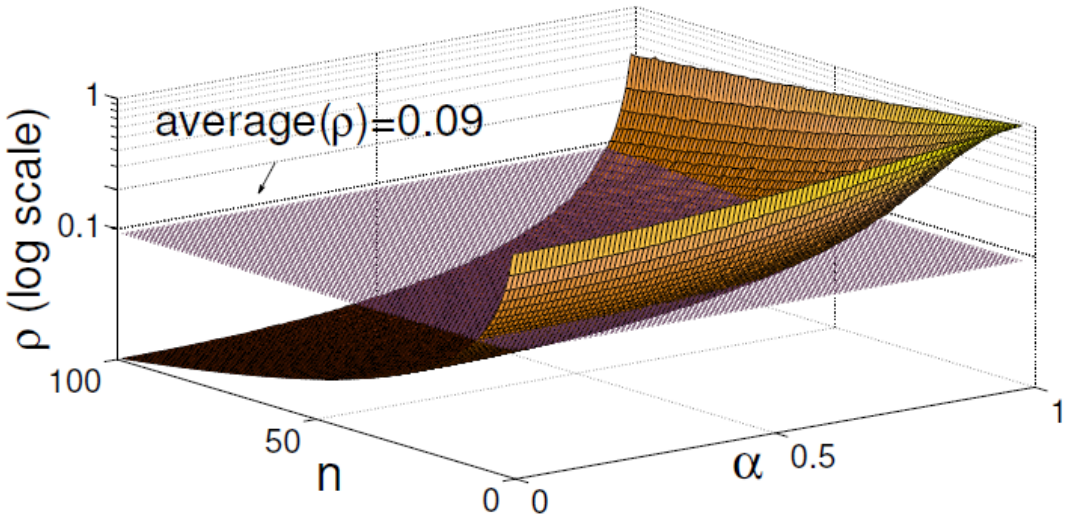
Efficiency (Cont'd)

- Simulation on 10 levels: a heatmap
- Darker regions contain more comparisons

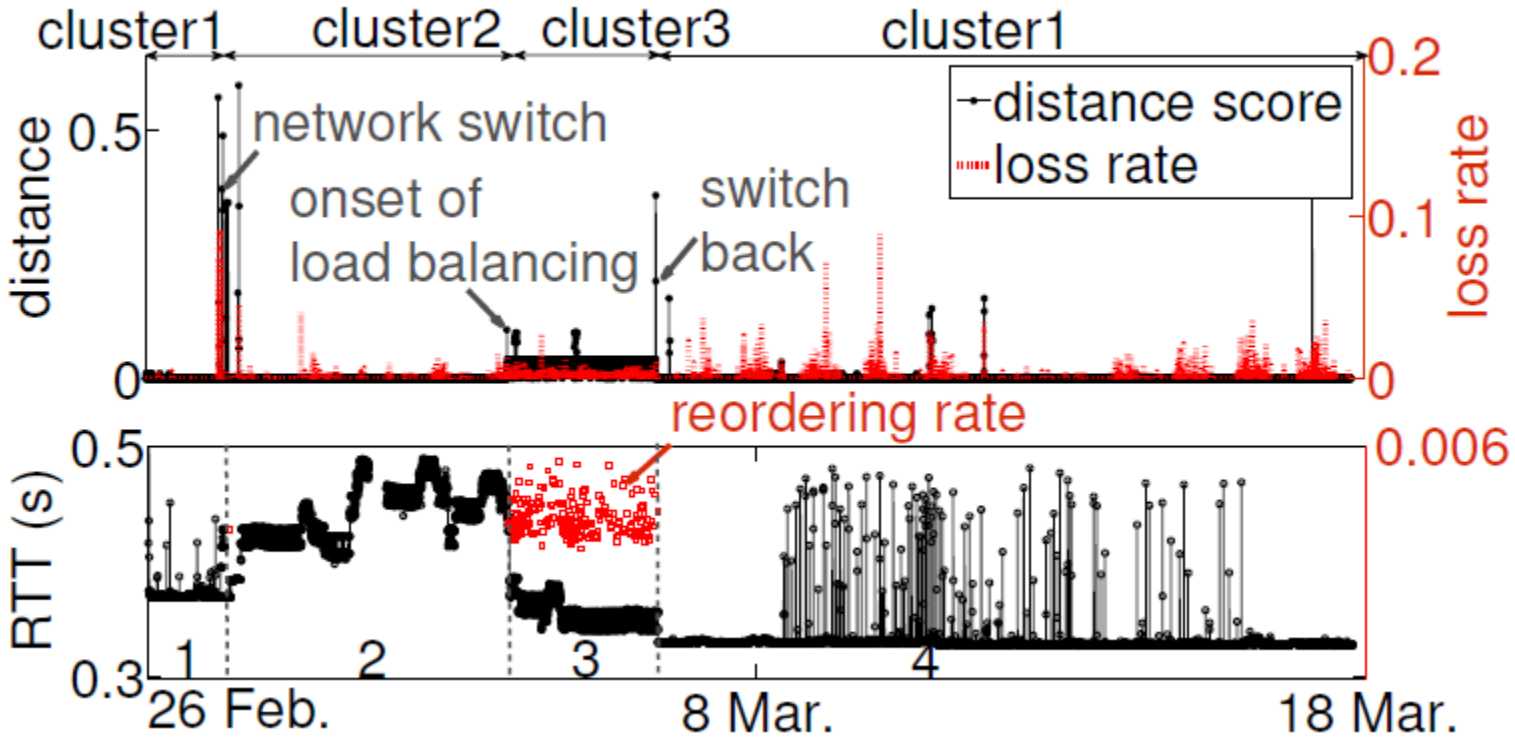


Efficiency (Cont'd)

- Simulation on 100 levels:
- 90% node comparisons eliminated on average



Showcase: an ISP transition monitoring



Conclusions

- Multi-level route analytics underpins many applications
- Existing approaches are inefficient
- Propose a new, faster approach

Acknowledgement

- Edmond Chan
- Xiapu Luo
- Waiting Fok
- Rocky Chang

- Work done at the Hong Kong Polytechnic University
- Project website: www.oneprobe.org

Thanks!