Joint EMANICS/IRTF-NMRG Workshop on
Netflow/IPFIX Usage in Network Management
October 30, 2008, München, Germany

A Distributed Architecture for IP Traffic Analysis

Cristian Morariu

Department of Informatics IFI, Communication Systems Group CSG, University of Zürich

Overview
Proposed Approach
Application Scenarios

ifi

Overview

- One big problem with centralized traffic analysis: scalability
  - Link speeds increase every year by 50%-100%
  - DRAM speed increases 7-9% every year
  - → more data to be processed in a shorter time
- Solution 1: packet sampling and flow sampling
  - Problem: measurement accuracy
- Solution 2: dedicated hardware
  - Can partially solve the problem but at a high cost

Proposed Solution

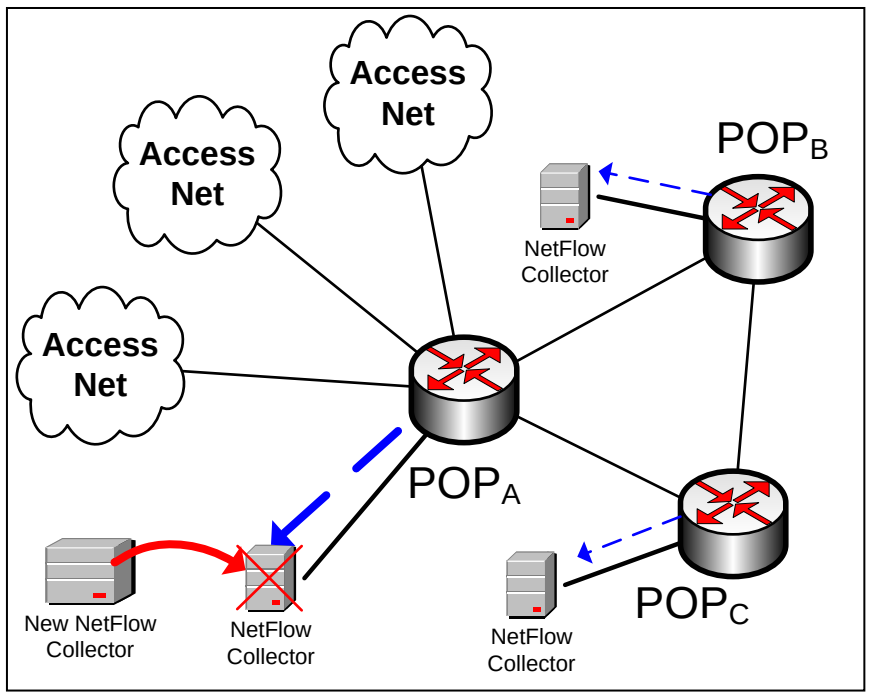
- Distribute traffic data to multiple traffic analyzers
  - Reduces the amount of data to be processed on a single node
  - Increases the time available to process each measurement data

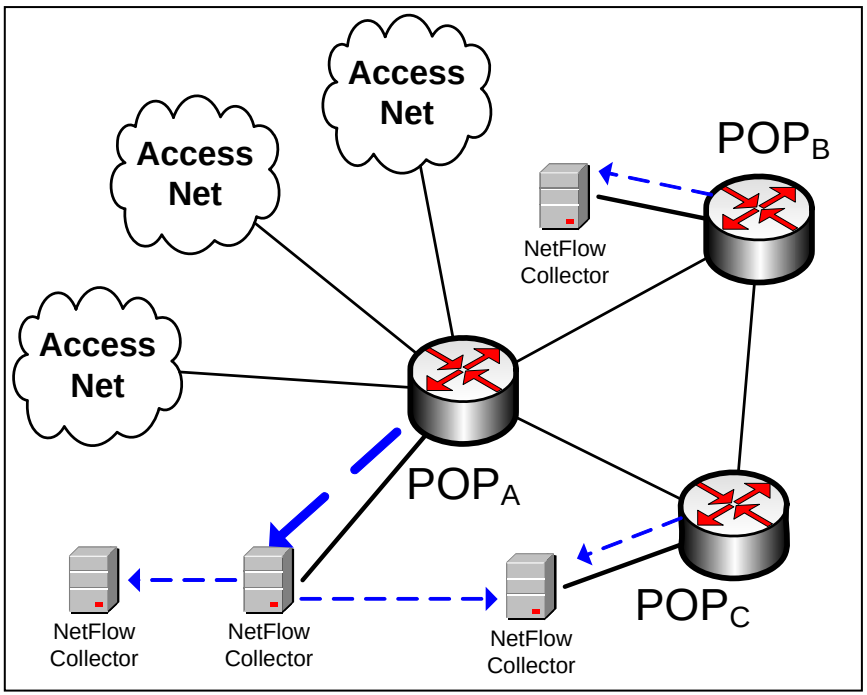
- Goals
  - Dynamic configuration
  - Avoid single point of failures
  - Provide load-balancing of workload
  - Use off-the shelve PCs (no dedicated hardware)
  - Increase the amount of data can be analyzed

- Applications:
  - Flow records storage
  - Multi-point delay measurement
  - Asymmetric routes detection

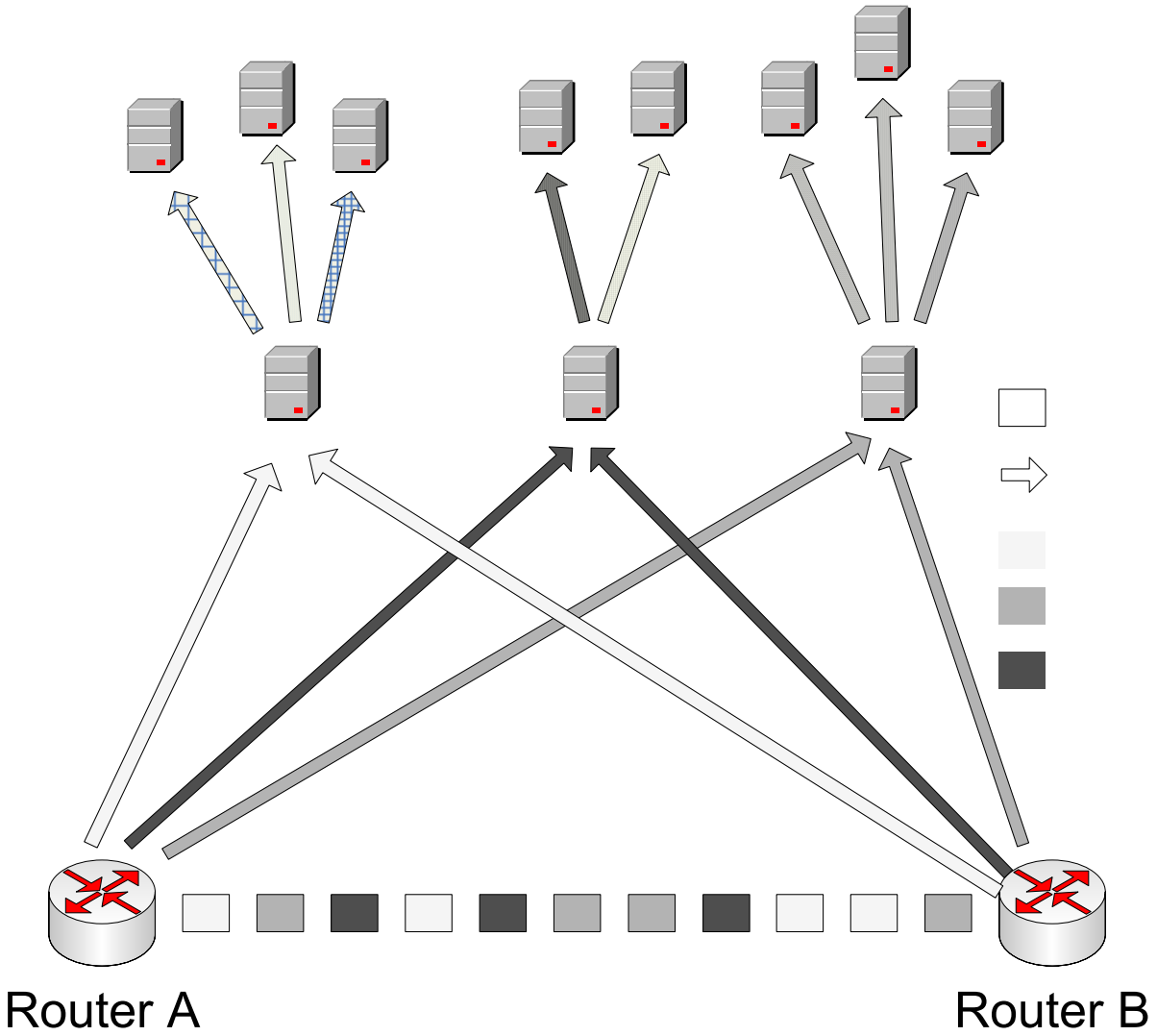
Proposed Approach

- Traditional approach:
  - Traffic increase → replace traffic analysis hardware
- Distributed approach:
  - Traffic increase → make use of available resources

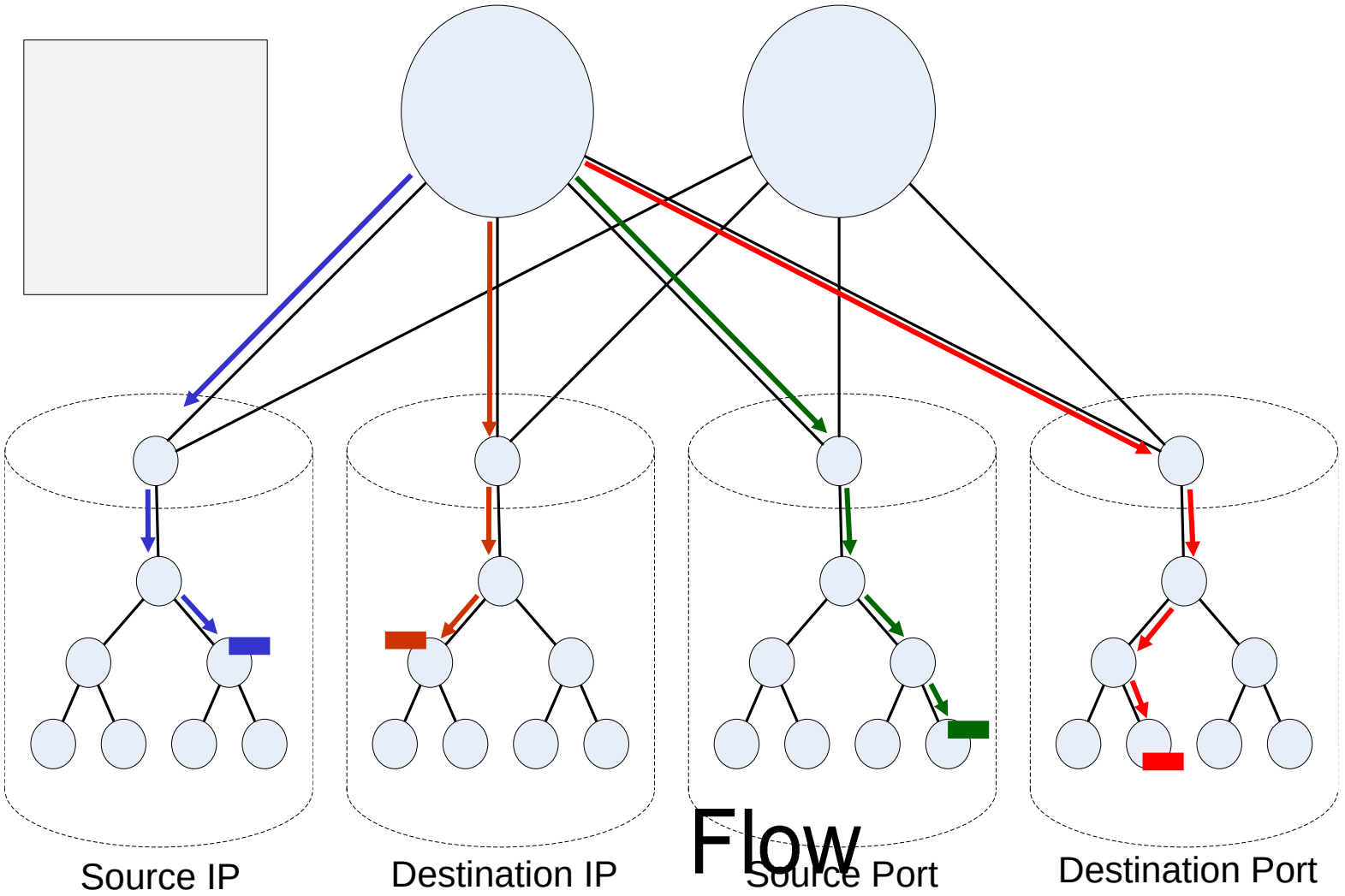




Distributed Architecture

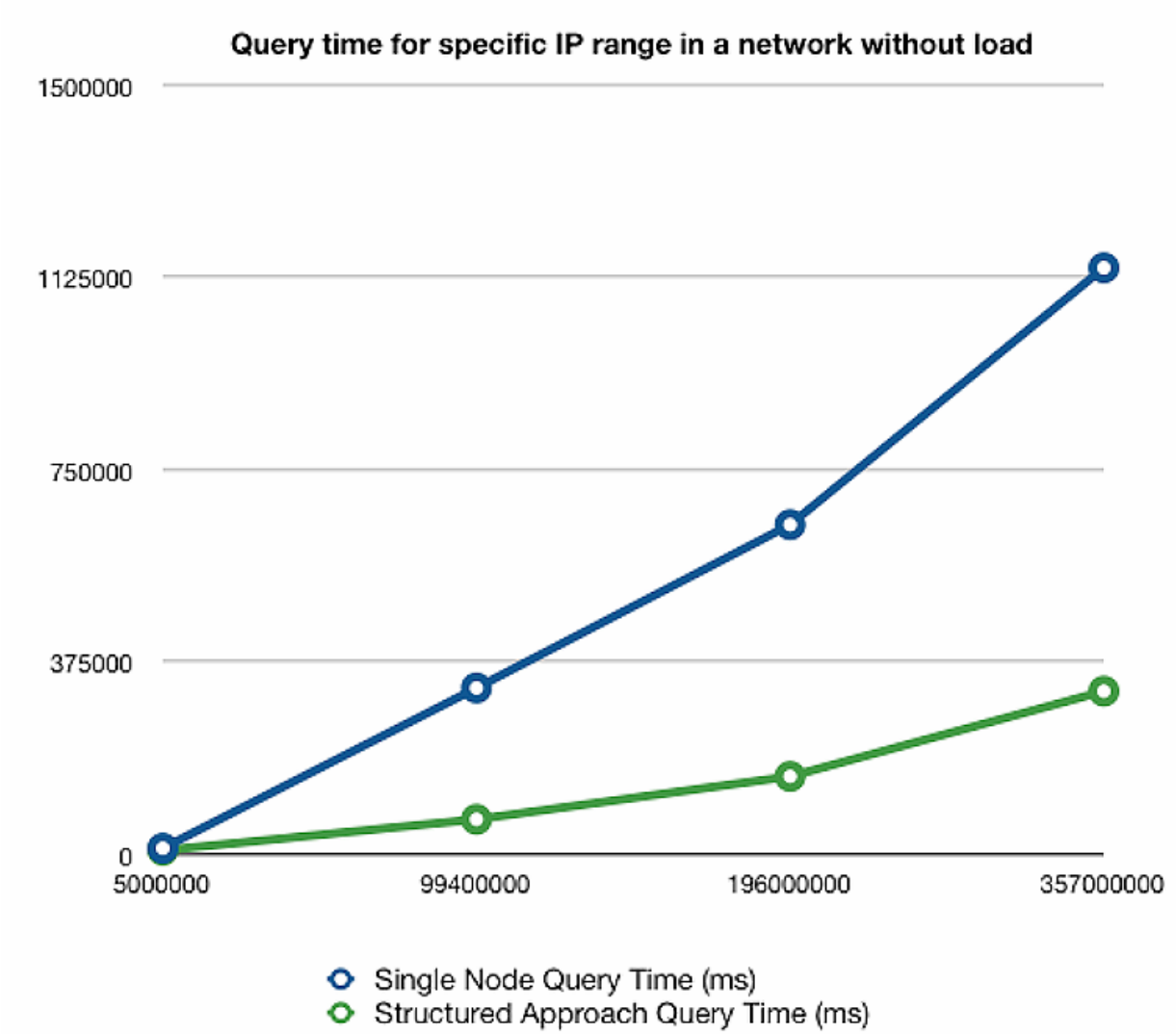


DIPStorage Architecture



C. Morariu, T. Kramis, B. Stiller: DIPStorage: Distributed Architecture for Storage of IP Flow Records. 16th Workshop on Local and Metropolitan Area Networks, September 2008.

DIPStorage Evaluation



Functional Evaluation

- distribution of IP flow records
- self organizing

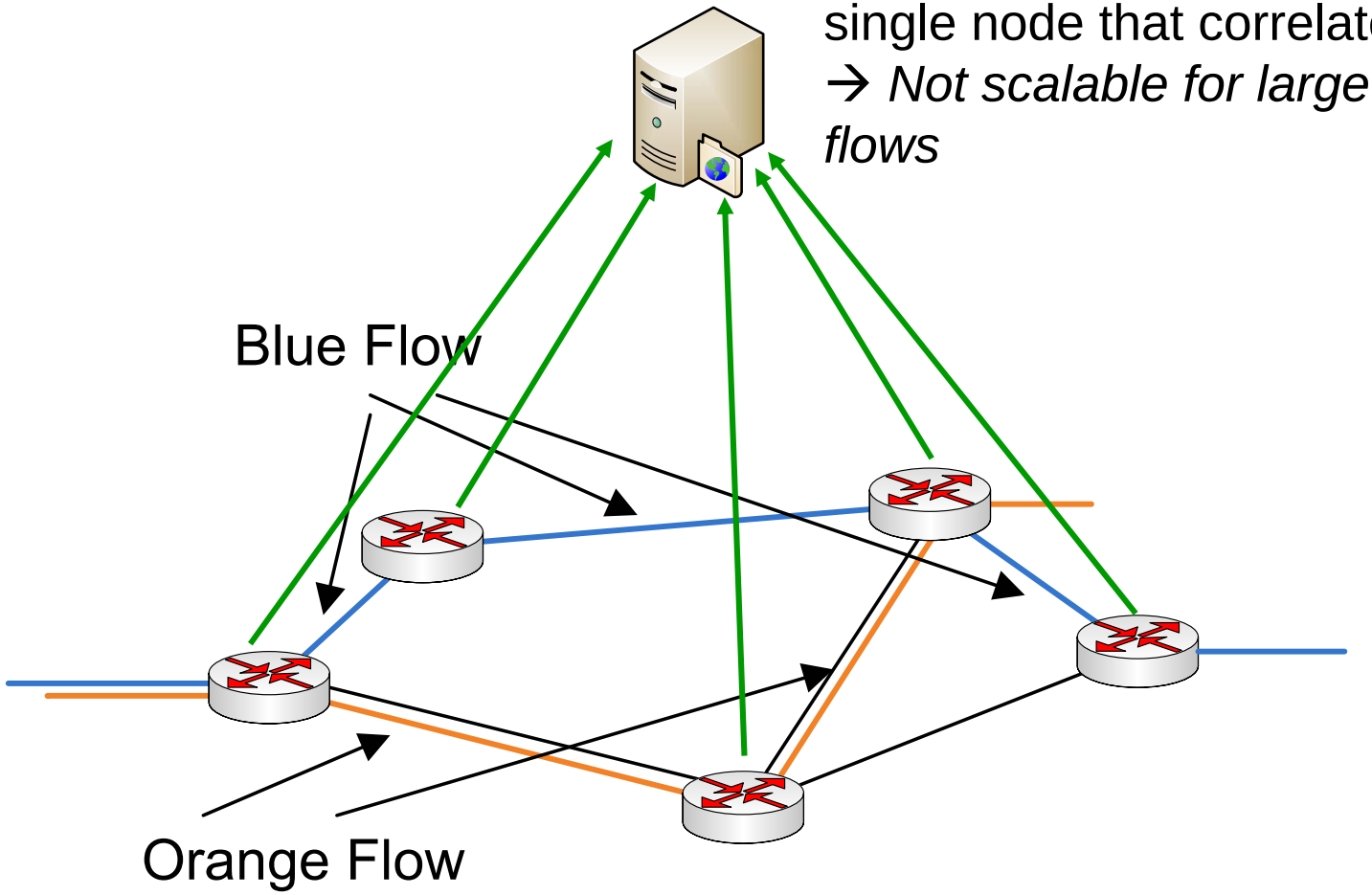
Performance evaluation

- 4 storage nodes used

- Speedup increase by a factor of 4 at high number of flow records

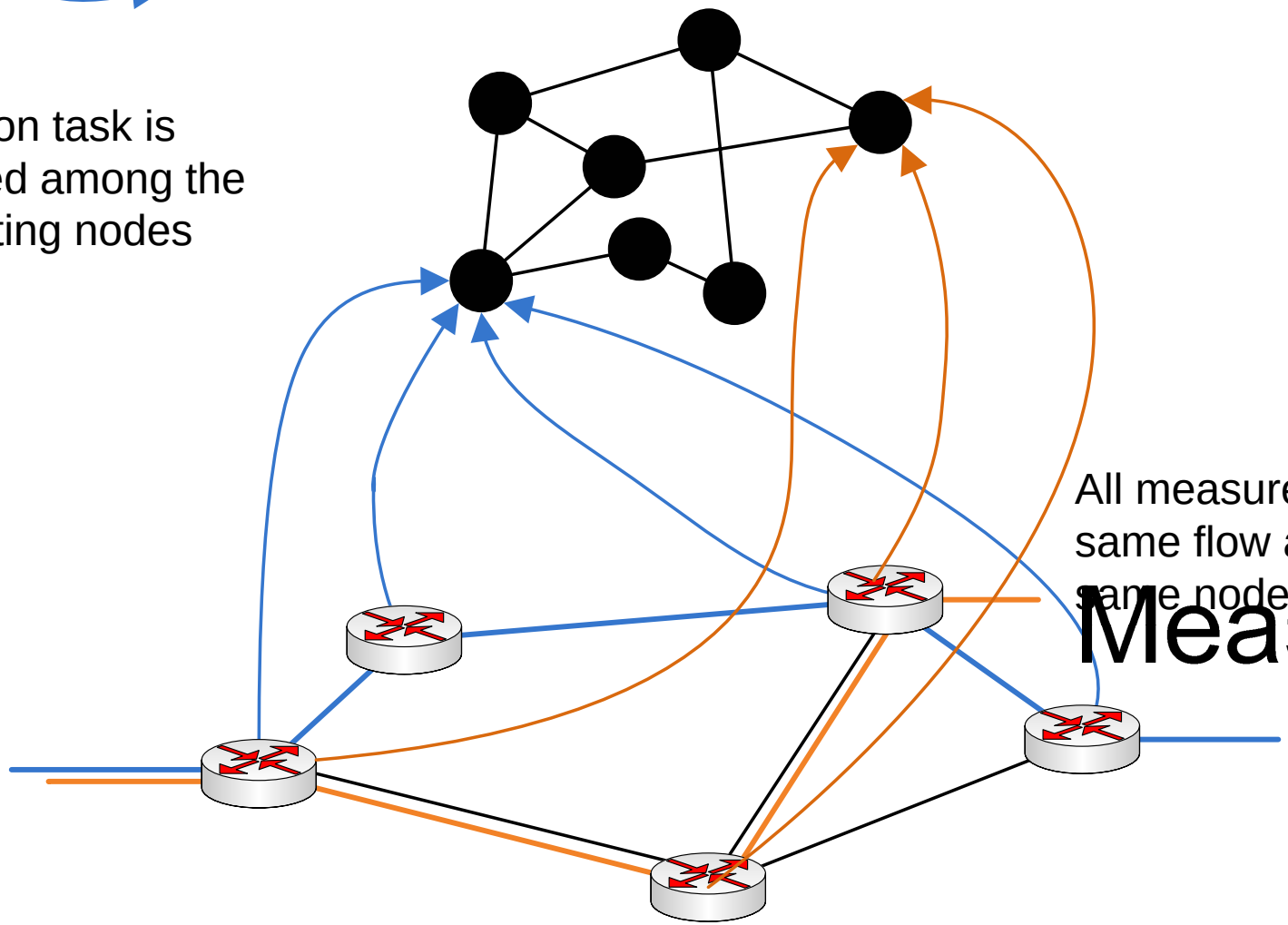
Applications: Multi-Point Delay Calculation

All measurement data goes to a single node that correlates it
→ *Not scalable for large number of flows*



Applications: Multi-Point Delay Calculation

Correlation task is distributed among the participating nodes

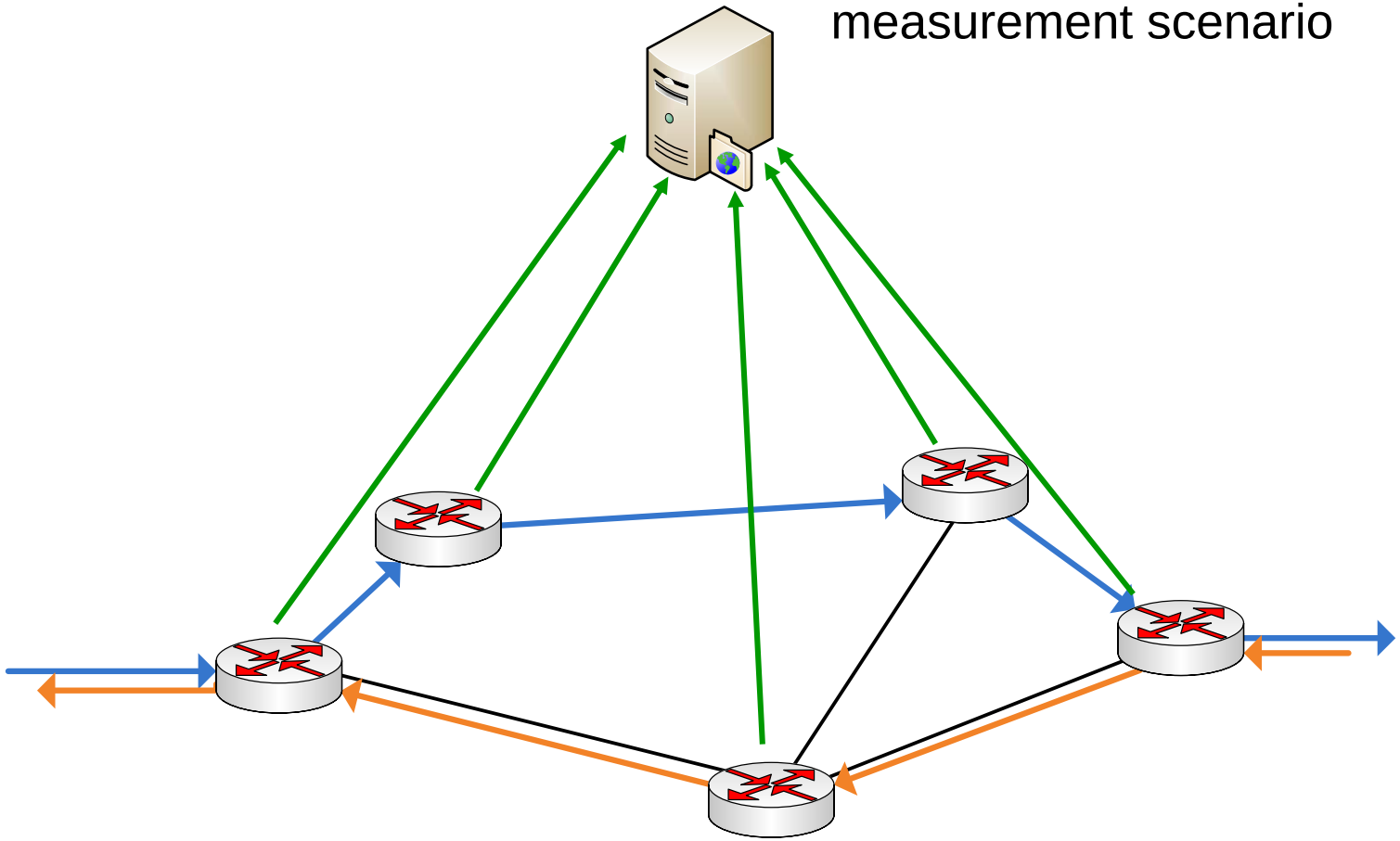


All measurements for the same flow are routed to the same node

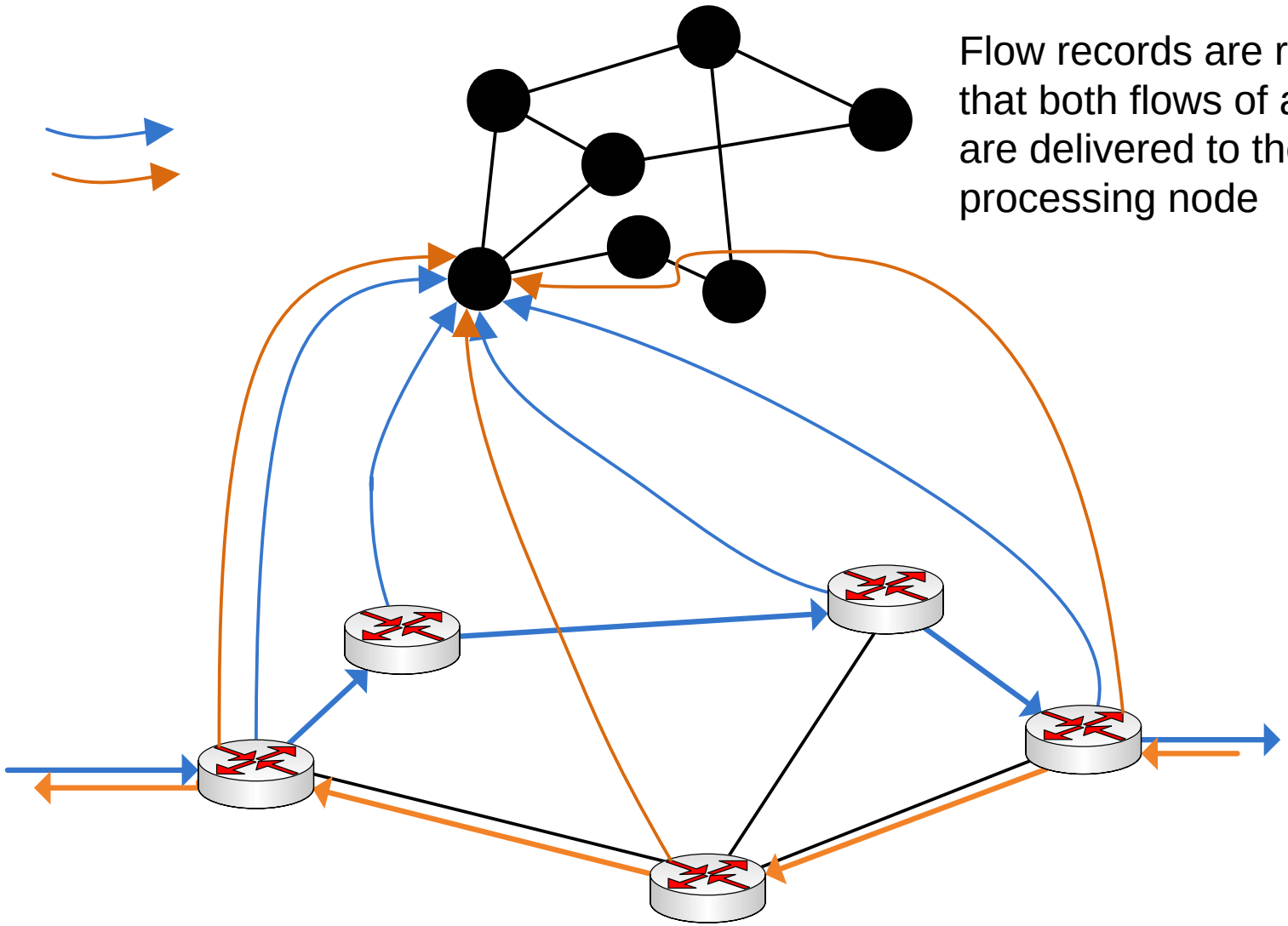
Measureme

Applications: Asymmetric Route Detection

Same problem as for the delay measurement scenario



Applications: Asymmetric Route Detection



Flow records are routed so that both flows of a bi-flow are delivered to the same processing node

Conclusion Remarks

- Existing distributed approaches to traffic analysis are based on static configurations
- A P2P based approach was not yet investigated
- Such an approach allows:
  - Scalability
  - More accurate results by processing more data
  - Increased storage space for flow records
  - Faster query response for IP flow records repositories
  - Support for different analysis applications