Foxtrot:
Phase space data representation for correlation-aware aggregation

Tom Parker <T.E.V.Parker@tudelft.nl>

TU Delft
Delft University of Technology

Current Aggregation Techniques

- Variants on the theme of averaging
  - Remove the interesting anomalies
  - Lose spatial information
- Aggregation without data awareness loses information
- Need to look again at application scenarios
  - Specifically, periodic data measurement
  - (Aggregation is less useful for event-based apps)

TU Delft
Delft University of Technology

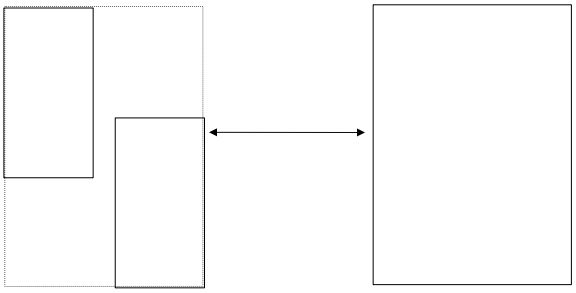
Foxtrot

- Data aggregation
- Minimal information loss
- Only aggregate "similar" data
- Aggregate less when there's more to say

Sensors

- A set of sensor readings is a point in phase space
  - Both sensors as we normally see them
  - And location data
- Several sets (or one) define a region
- Regions can be aggregated

Aggregation example



- Axes = sensor readings
  - (Can be physical locations)

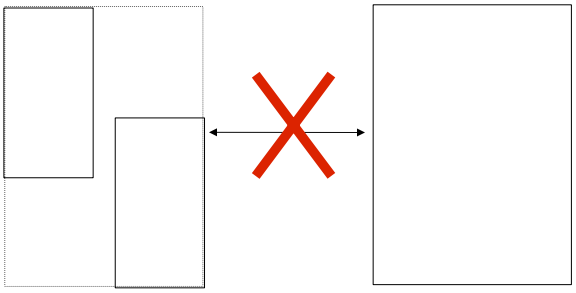
TU Delft
Delft University of Technology

Similarity

- Answering "can we aggregate these regions?"
- Simple ranges
  - Threshold limits on region size
  - (Limits on expansion rates)
- Complex ranges
  - User-defined function
  - Usually: Multi-sensor thresholding

TU Delft
Delft University of Technology

Limits on Merging



- Sometimes, we can't merge some points
  - Try others!

TU Delft
Delft University of Technology

Hints to Routing

- Early aggregation is better
- Foxtrot can give "hints" to routing algorithms
  - Data-aware stateless clusters
  - Move towards closest-to-sink node
  - Then to the wider world

TU Delft
Delft University of Technology

Auxiliary uses

- In some cases, no message transfer required
  - Can determine currently unneeded sensors
  - But, no node counts
- Battery level can be a sensor
  - Less battery => more aggregation ?
- Normal aggregates are difficult
  - Limited additional information can be derived

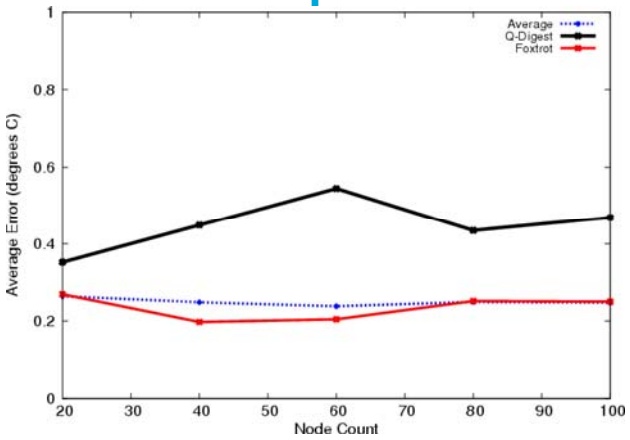
TU Delft
Delft University of Technology

Simulations

- 30m x 30m region
- 14m radio range
- Perfect links
- 20,40,60,80,100 nodes
- Metric: difference between true values and data reported to sink
- Tested with different data scenarios
- Compare with averaging and Q-Digest
- Values: 20-21 (Spread); 20-21/30-31 (Division)

TU Delft
Delft University of Technology

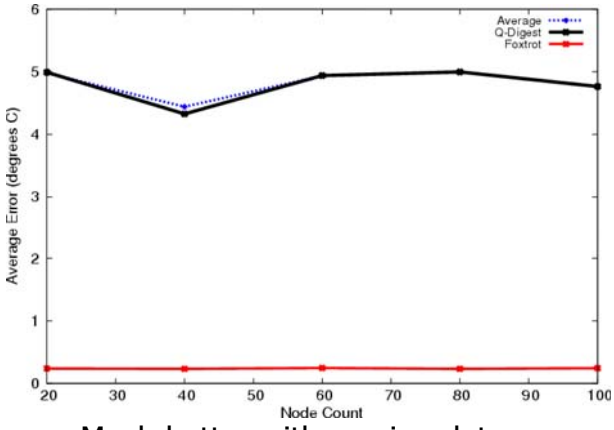
Results - “Spread” scenario



- Similar error in trivial scenarios

TU Delft
Delft University of Technology

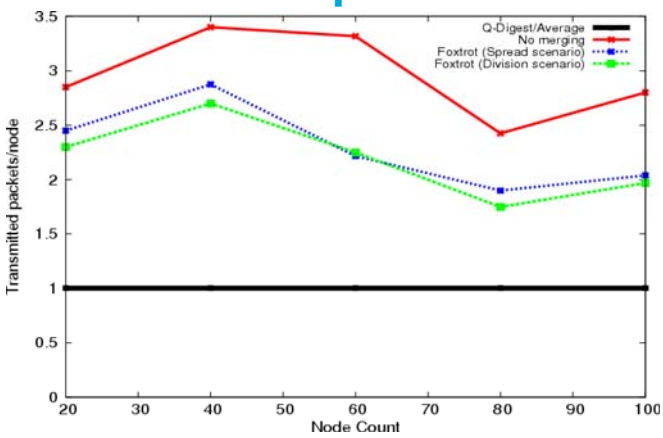
Results - “Division” scenario



- Much better with varying data

TU Delft
Delft University of Technology

Results – packet rates



- Disadvantage: Increased data rate

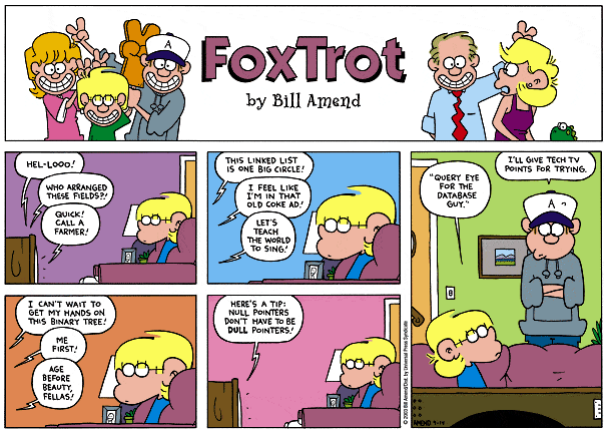
TU Delft
Delft University of Technology

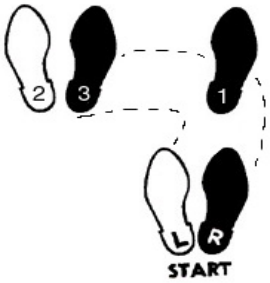
Summary

- Automatic multi-dimensional aggregation
- Information loss < user thresholds
- "Similarity" metric can also help routing
  - Not vital, but does improve aggregation
- Sometimes we can't aggregate
  - When there's lots of information
- Results have a greater level of trust

TU Delft
Delft University of Technology

Foxtrot is not...





TU Delft
Delft University of Technology