LARGE SCALE PRIVACY PRESERVING BLUETOOTH SENSING

Nelson Gonçalves

Advisers: Carlos Baquero
Rui José

LOCATION INFORMATION

- Type of Information that can be used by:
  - Service Providers: improve their service and obtain better understanding of users behavior to further improve future services and infrastructures.
  - Users: get personalized information
- **HOWEVER**, continuous monitoring, processing and storage of location data can create privacy problems.

PRIVACY IS IMPORTANT

- Location Information:
  - expose sensitive information
  - predict future whereabouts. E.g. PleaseRobMe(foursquare +Twitter)
- It might be a decisive factor in the popularity/sustainability of location based services

EXISTING APPROACHES & GOAL

- Regulatory Strategies - Government rules
- Privacy Policies - Trust-based mechanisms
- Anonymity: Disassociation between individual's personal information and actual identity
- Obfuscation: Degradation of the quality of data
- **GOAL**: Evaluate Stochastic summarizing techniques as an approach for large scale collaborative sensing scenarios

STOCHASTIC SUMMARIZING TECHNIQUES

- Probabilistic
  - Do not allow original item to be recreated from summary
- Space efficient
  - Size depends only on the number of input elements
- Trade-off between space and accuracy

PROOF OF CONCEPT

- 2 Bluetooth Collaborative Scanning Scenarios
  - Gate Counting
  - Movement Patterns

GATE COUNTING

- Count the number of unique devices (people) across nodes
  - Bloom Filters and Hash Sketches
- Criteria:
  - Accuracy
  - Size
  - Aggregation

GATE COUNTING RESULTS

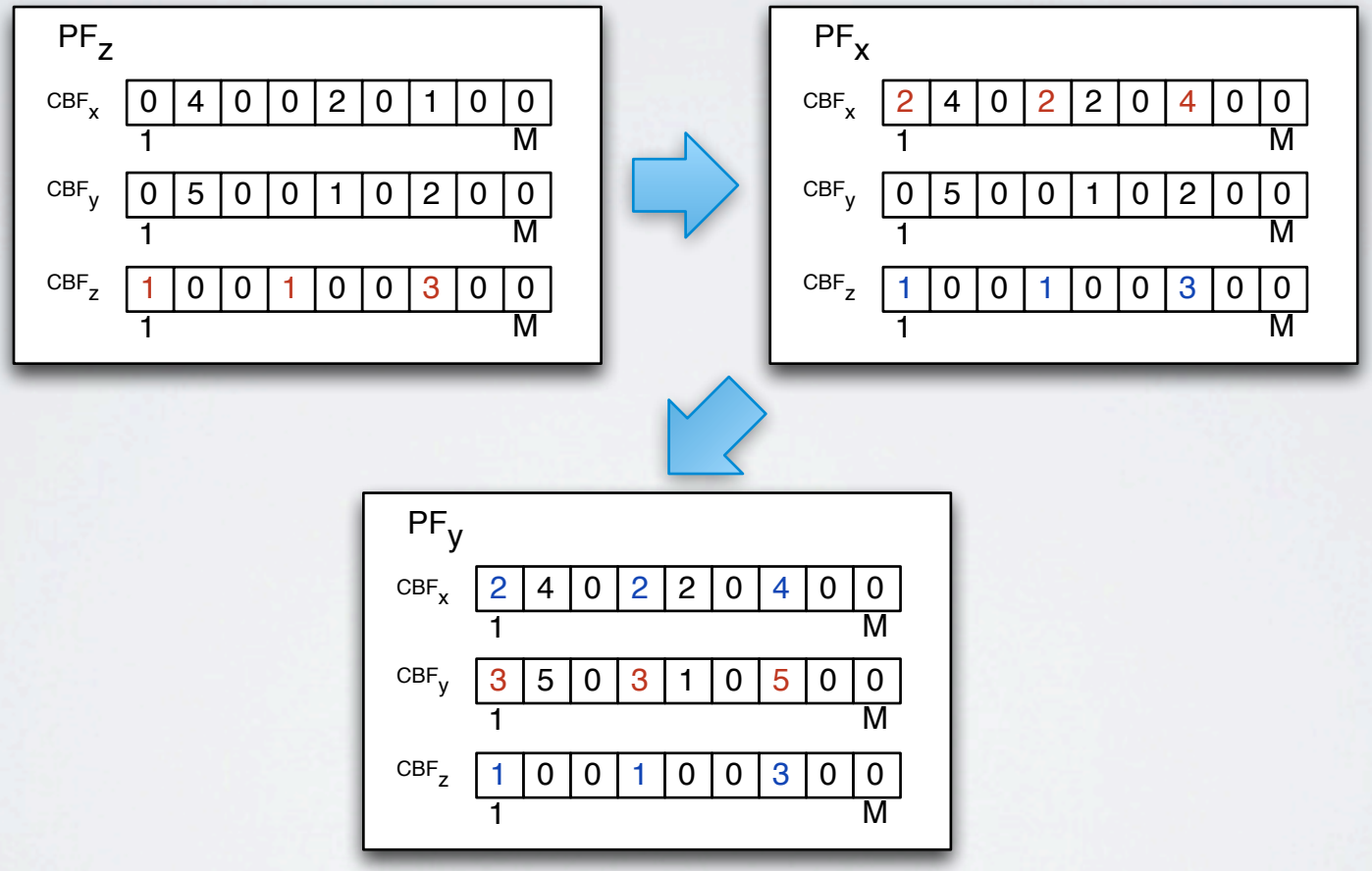
Technique Name	Accuracy	Size	Aggregation
Log Log Sketch	+*	++	+
HyperLogLog Sketch	+*	++	+
RIA DC Sketch	+	+	-
RIA LC Sketch	+	+	+
LC Sketch	+	+	+
Bloom Filter	+	-	+
Scalable Bloom Filter	+	-	-

* Bad Performance for small Cardinalities

MOVEMENT PATTERNS

- Obtain accurate macroscopic information
  - Allowing Plausible Deniability of individual information
- **New technique** - Precedence Filters
  - Counting Bloom Filters and Vector Clocks

PRECEDENCE FILTERS



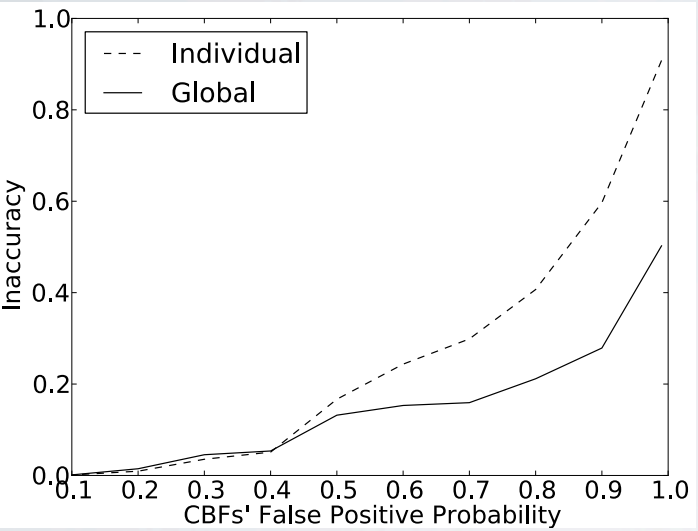
Trace: Z → X → Y

MOVEMENT PATTERNS

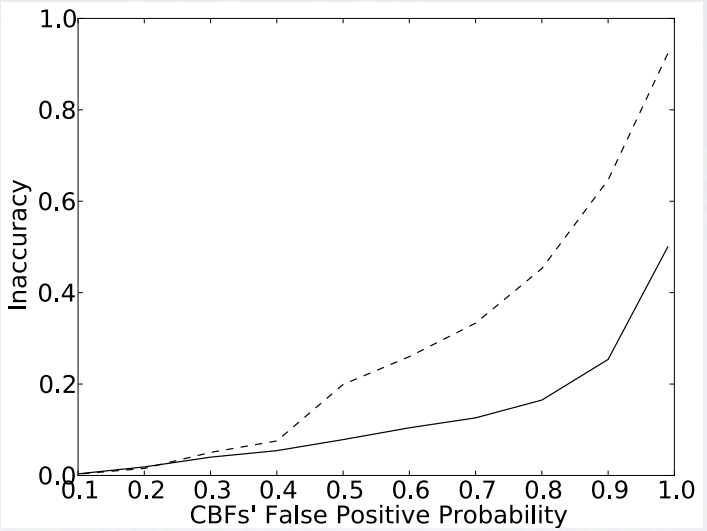
- Individual Metric
  - False probability "Individual X visited S1 before S2"
- Global Metric
  - Error "2% of the transitions are from Restaurant Y to Cafe Z"

MOVEMENT PATTERNS RESULTS

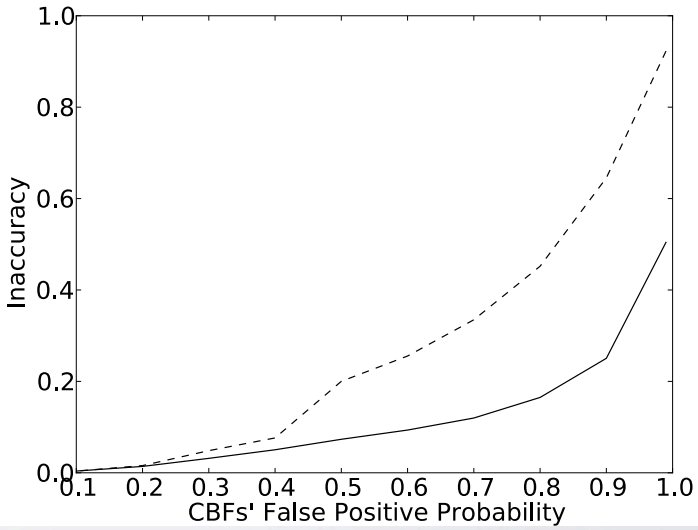
Number Devices : Maximum Trace Size : Average Trace Size



s2808 : 11 : 4



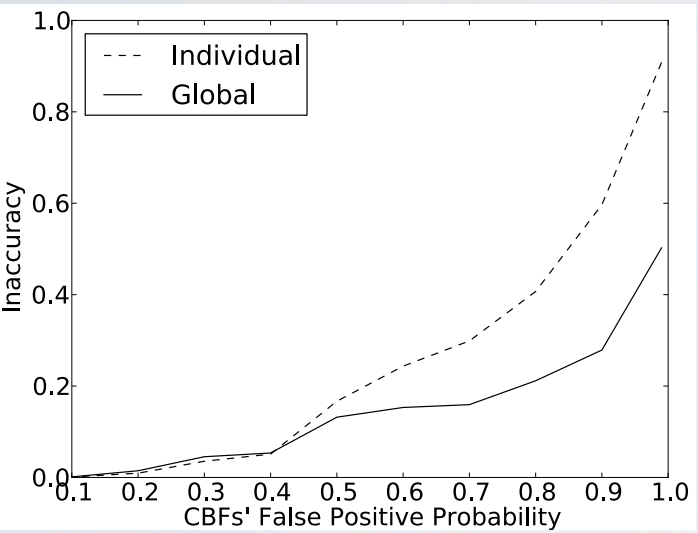
s10000 : 11 : 4



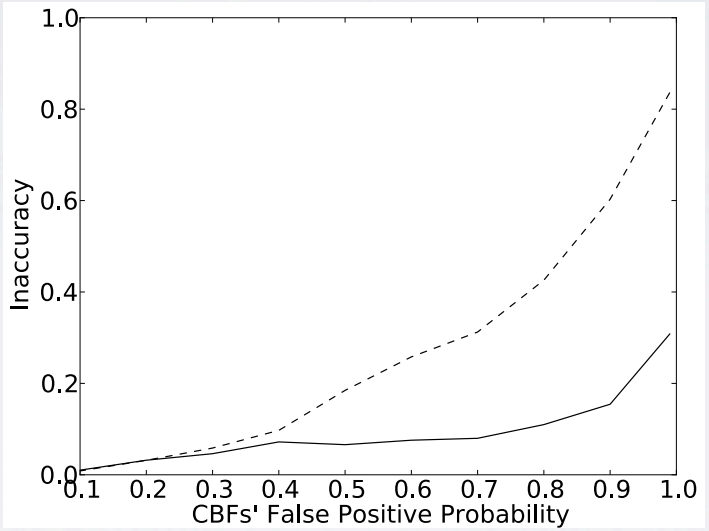
s100000 : 11 : 4

Increasing number of devices

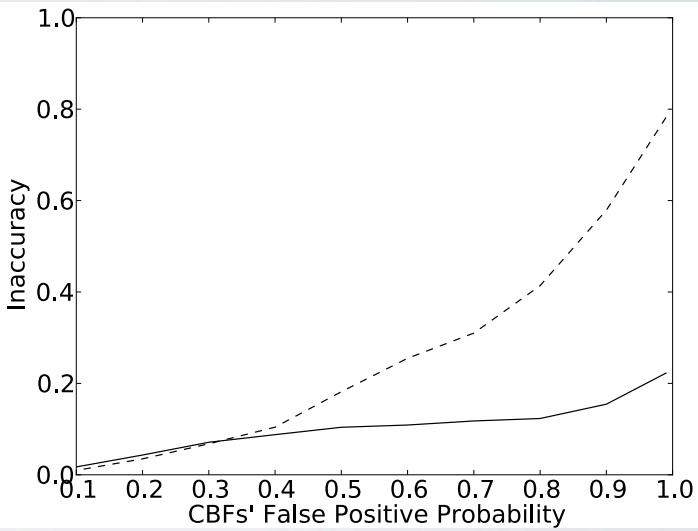
MOVEMENT PATTERNS RESULTS



s2805 : 11 : 4



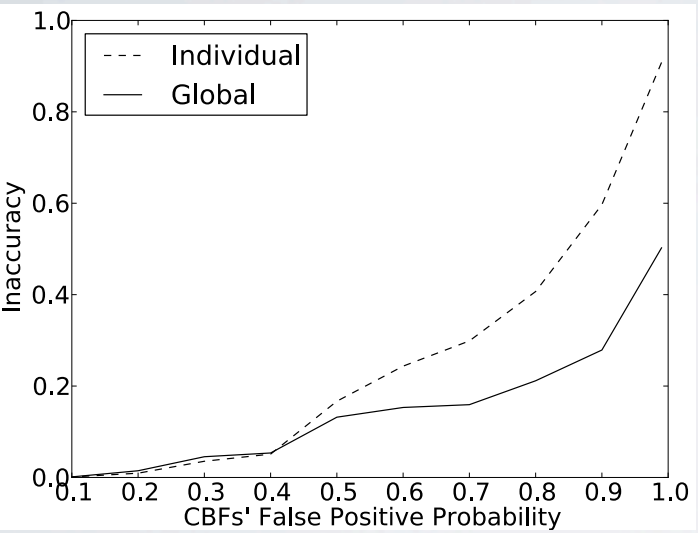
s2805 : 50 : 16



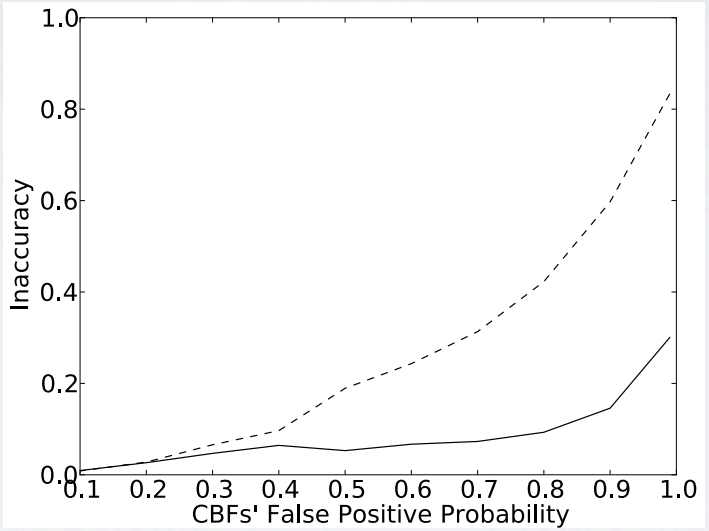
s2805 : 100 : 30

Increasing trace sizes

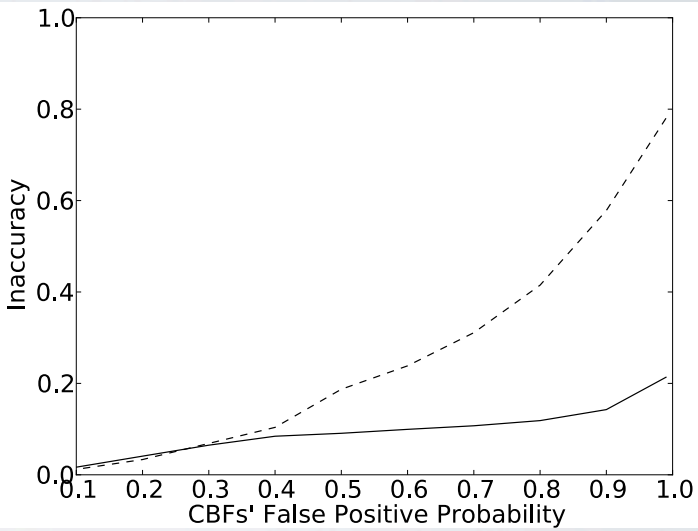
MOVEMENT PATTERNS RESULTS



s2805 : 11 : 4



s10000 : 50 : 16



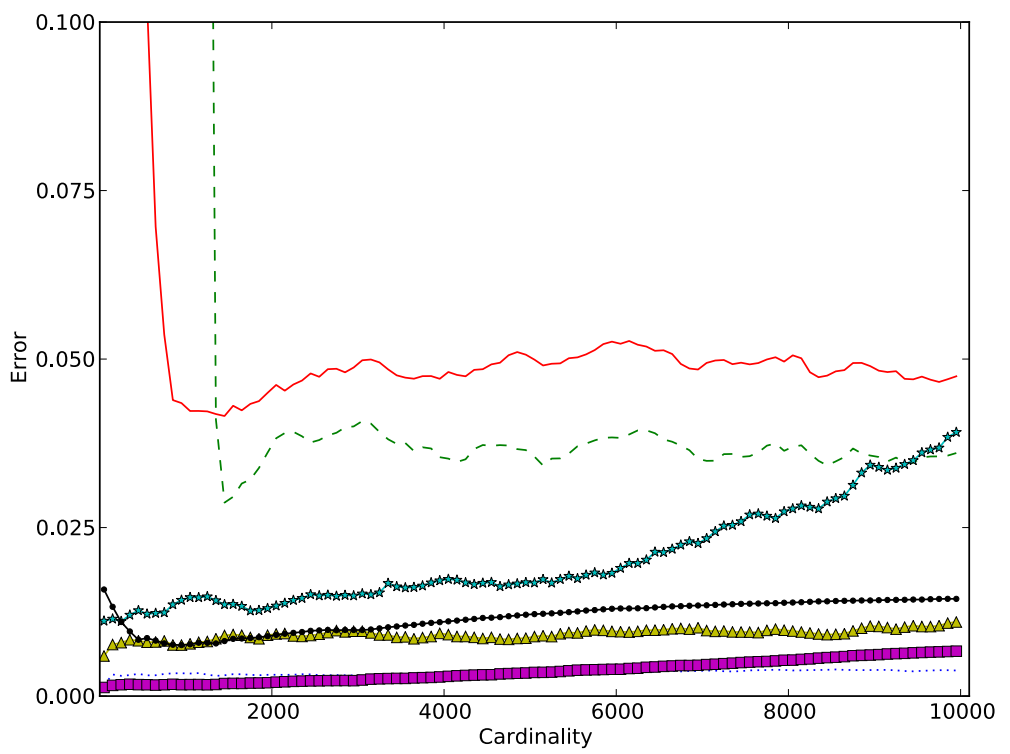
s100000 : 100 : 30

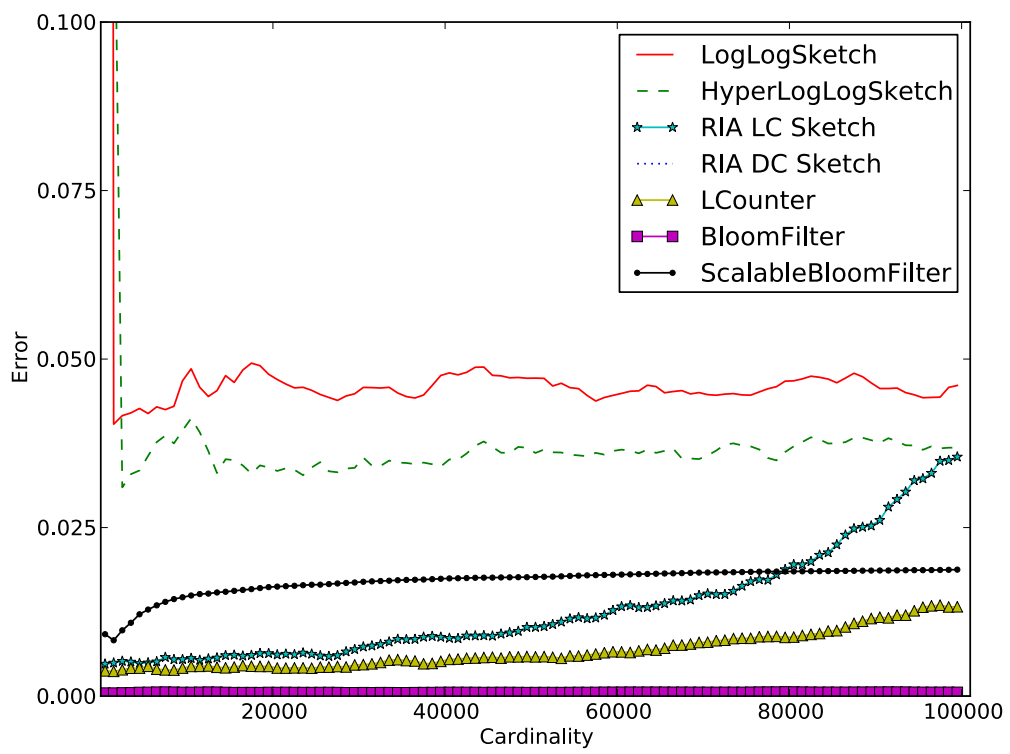
Increasing trace sizes and number of people

CONCLUSION

- Stochastic summarizing techniques are viable building tools
  - Analysis and use of existing techniques
  - Devised a new technique from existing ones
- Privacy has a cost

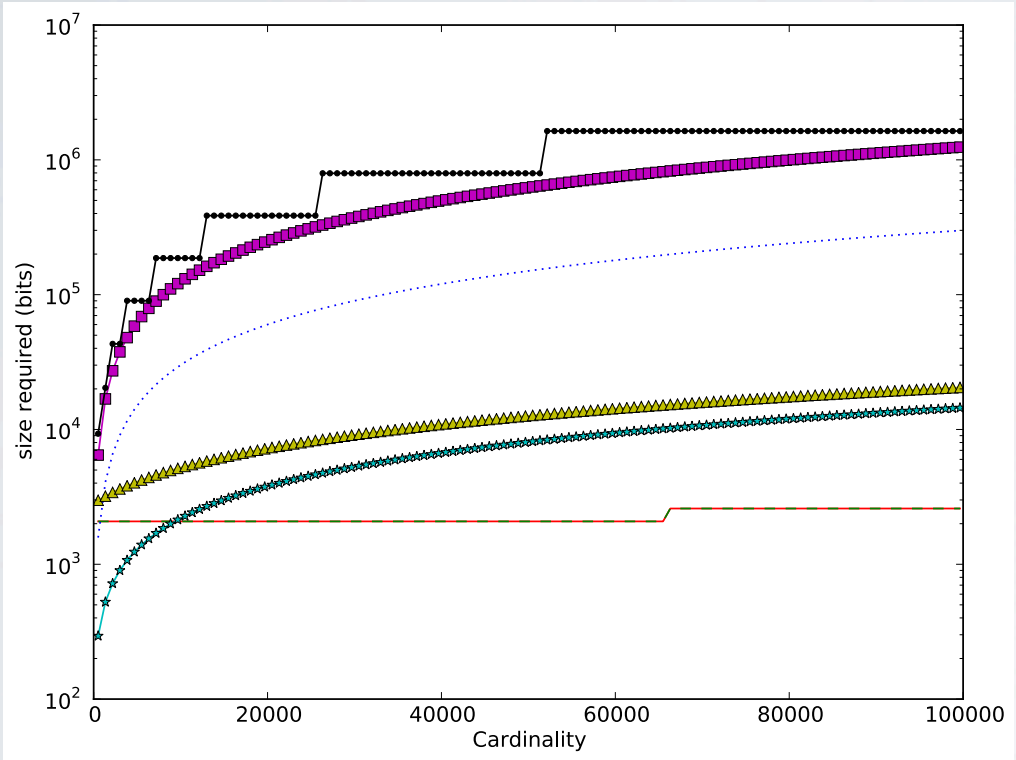
GATE COUNTING RESULTS



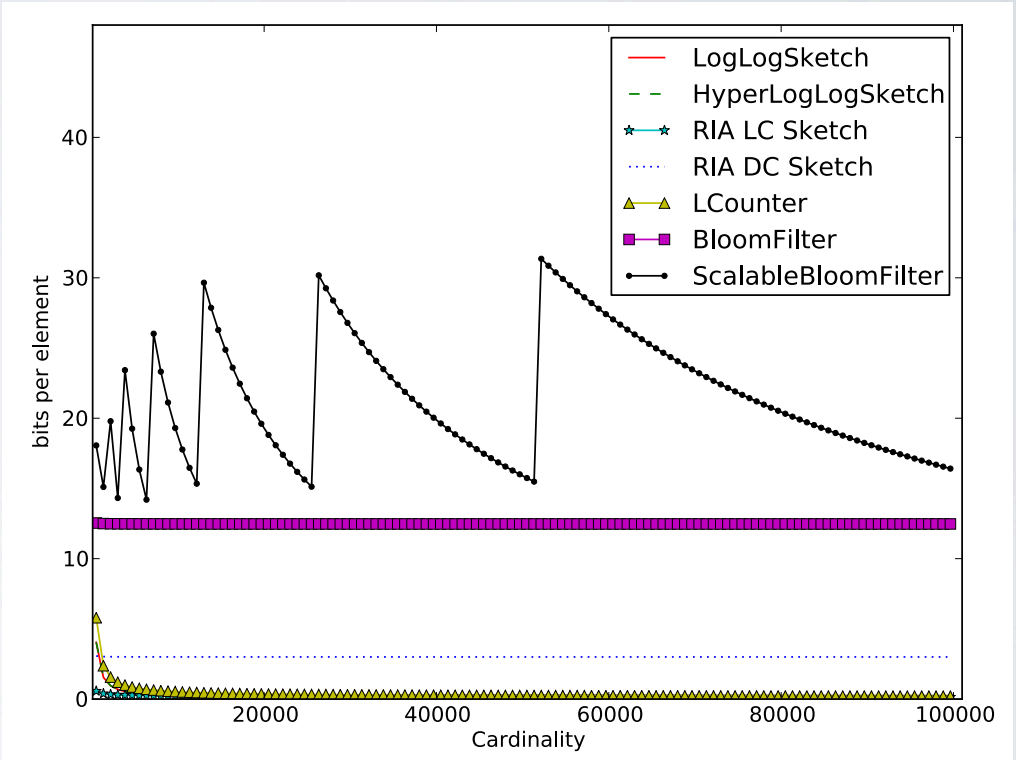


Accuracy

GATE COUNTING RESULTS

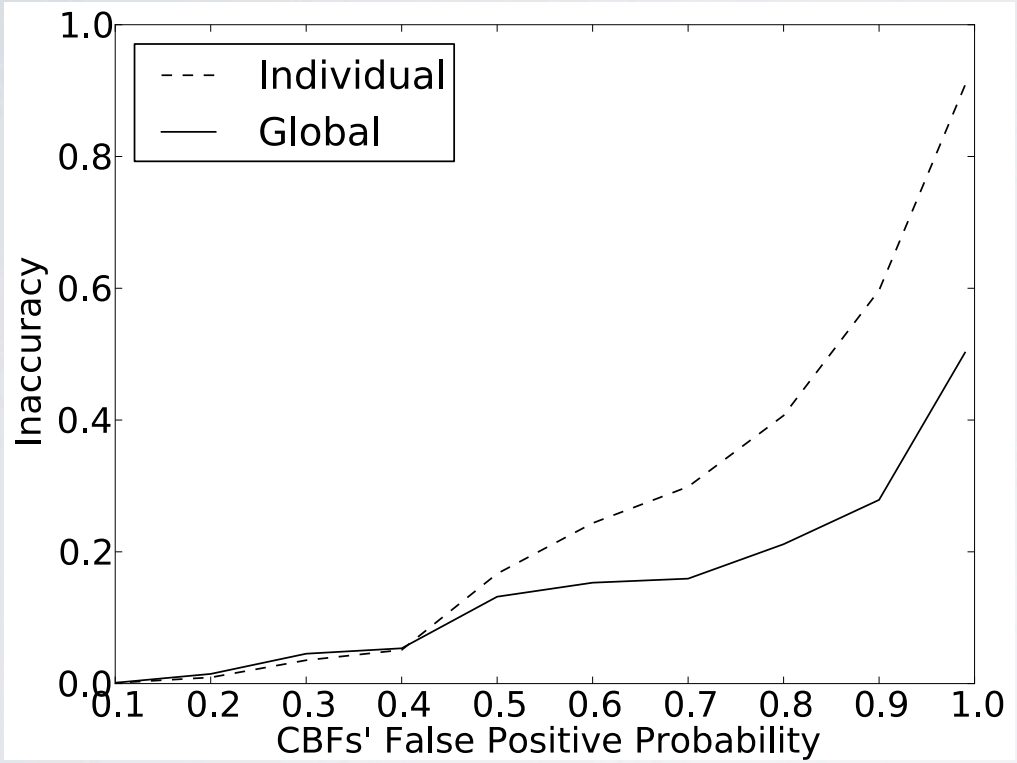


Total Size

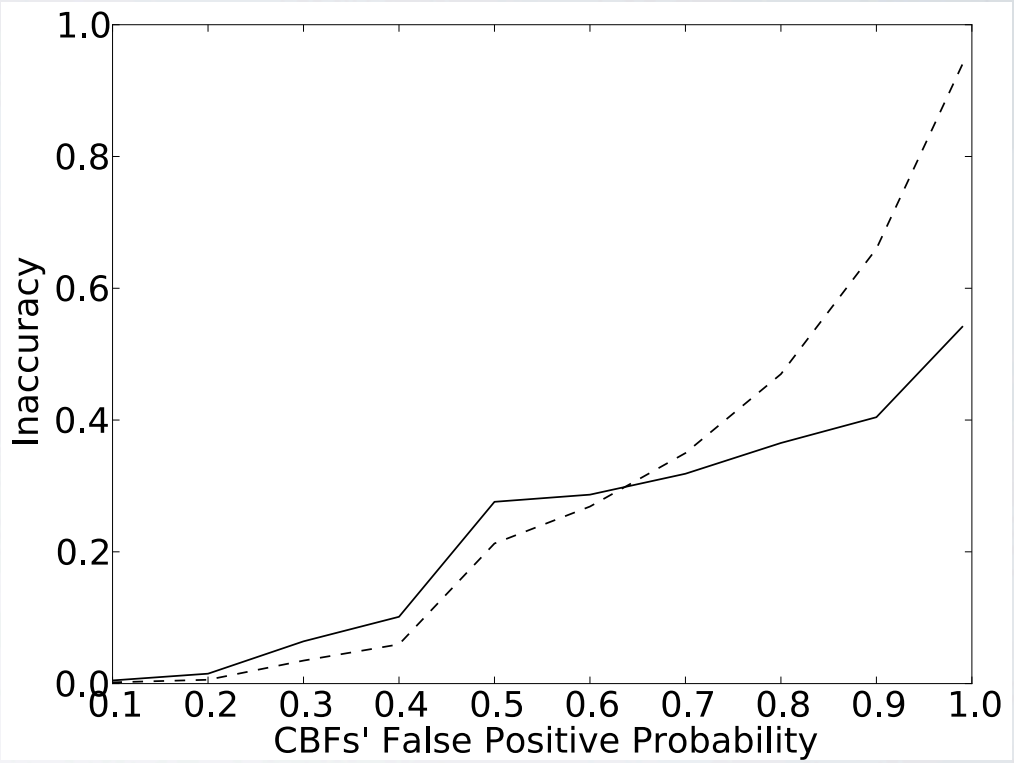


Bits per element

MOVEMENT PATTERNS RESULTS



s2805 : 11 : 4



r2805 : 11 : 4