

Adaptive Stream Filters for Range Queries with Fraction-based Tolerance

Alan Kwan

Department of Computer Science
The University of Hong Kong

11-Jan-2007

Outline of the Talk

- Based on the paper
 - R. Cheng, B. Kao, S. Prabhaker, A. Kwan, and Y. Tu. **Adaptive Stream Filters for Entity-based Queries with Non-Value Tolerance**. In *Proc. of VLDB*, 2005.
- Introduction of Data Streams and related Applications
- Fraction-based Tolerance Protocol for Non-ranked Queries (FT-NRP)
- Enhancements on the Fraction-based Tolerance Protocol
 - Immediate Compensation
 - Incremental Deployment
- Summary

Data Streams

- Characteristics of **stream** and **applications**
 - Massive data volumes, fast data rate
 - Reactive systems - Real-time, continuous monitoring for special events, response time requirements are critical

Data Streams and Applications

- Emerging Data Stream Management System (DSMS)
 - **STREAM** [ABB03]
 - **AURORA** [ACC03]
 - **CACQ** [MSH02]
- Various stream applications
 - Network monitoring and traffic engineering
 - Telecom call records
 - Network security [BO03]
 - Habitat monitoring [MPS02]
 - Structural health monitoring

Data Streams and Applications

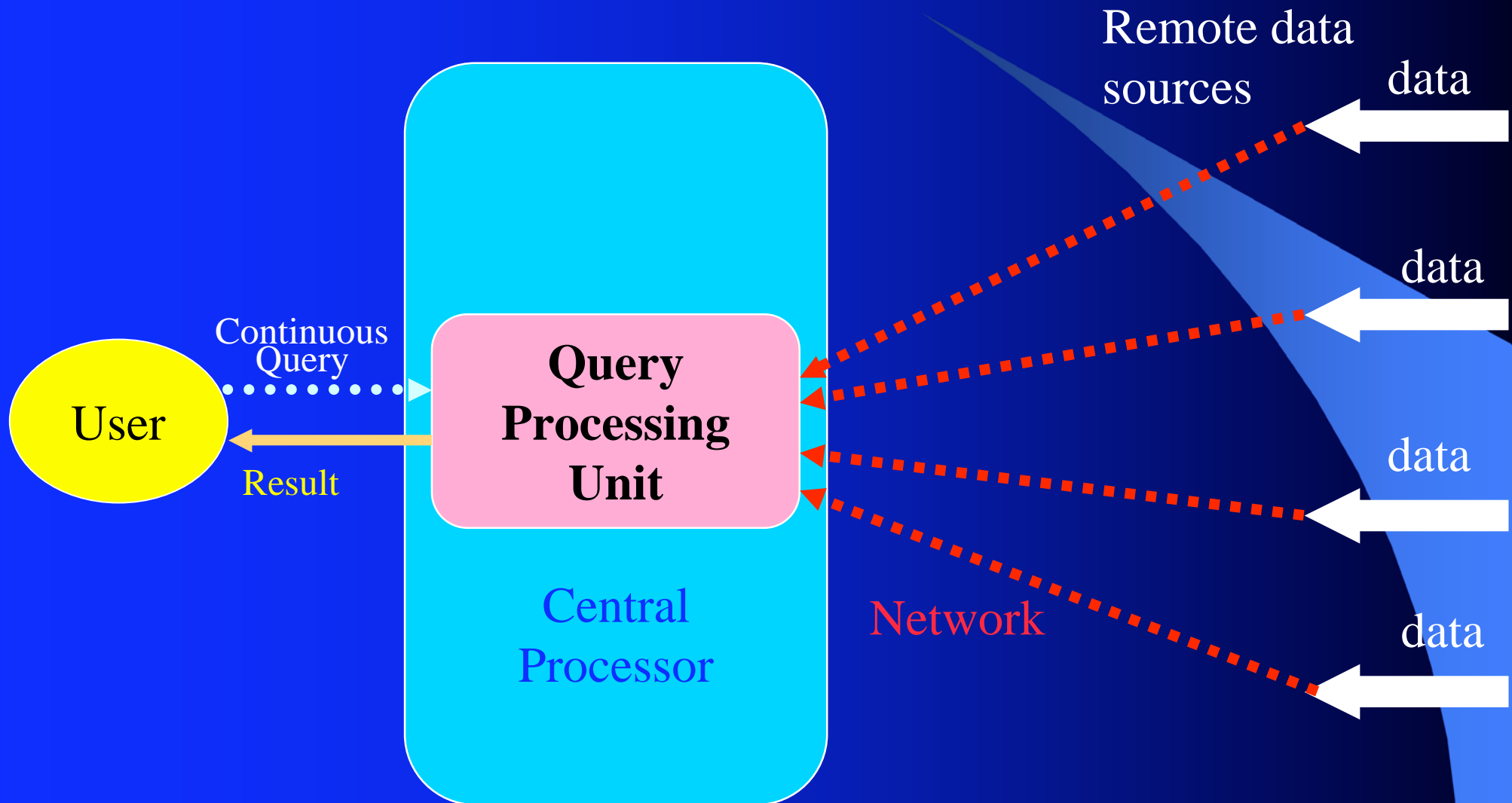
Future: Stonecutters Bridge in Hong Kong

- Stonecutters Bridge (2007)
- WASHMS: wind and structural health monitoring system



- Wind
- Temperature
- Accelerations
- Strains
- Traffic loading
- Humidity, rainfall
- Displacement (by GPS)
- Corrosion
- **TOTAL: 1114 channels !!!**

Model of DSMS for Stream Applications



The Motivation of Approximate Query Processing

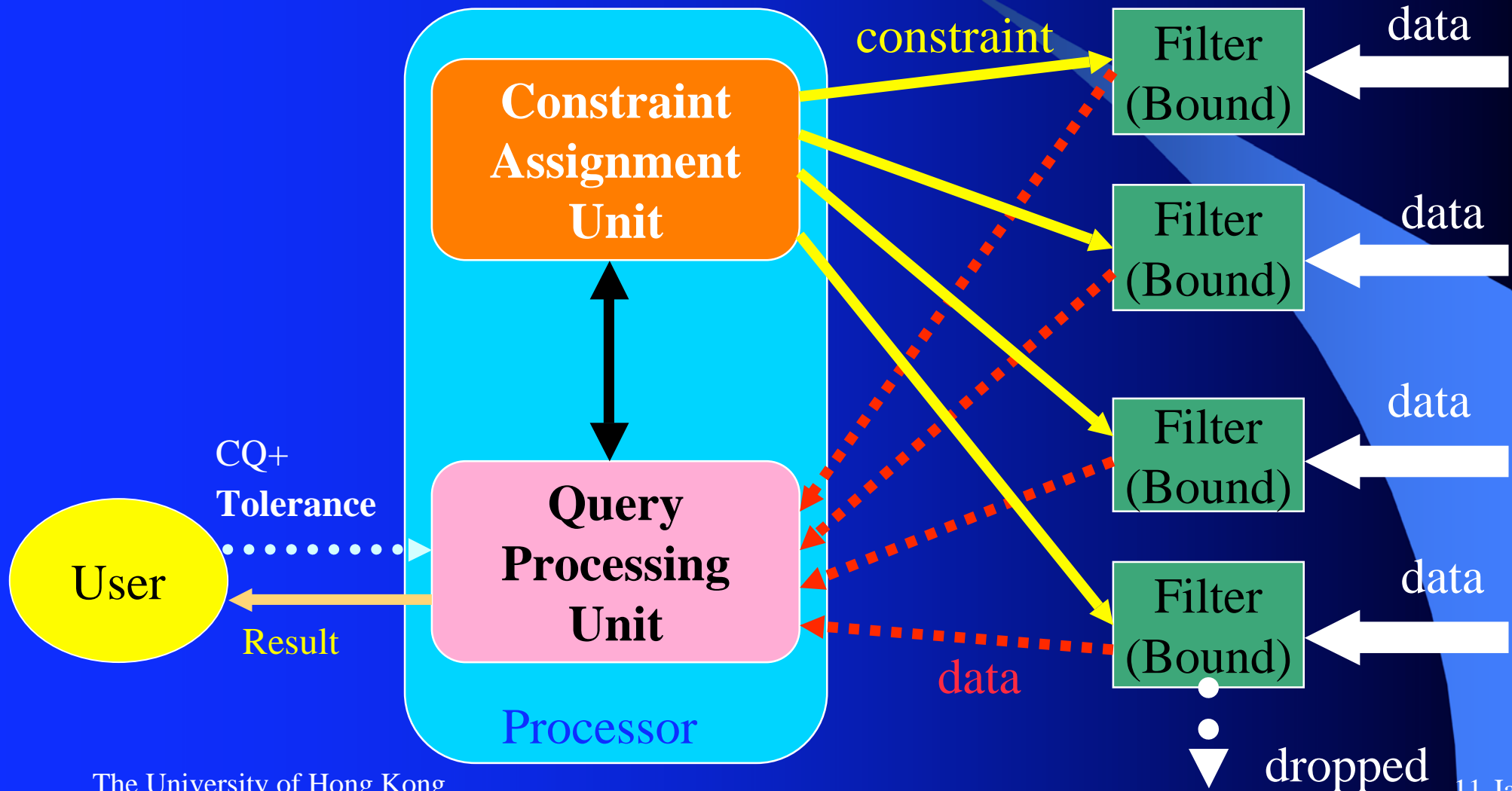
- **Reduce Limited Network Bandwidth**
- **Reduce Battery Energy Consumption of Sensors**

The Motivation of Approximate Query Processing

- Solutions

- Trade **query answer accuracy** for **communication overheads** [JCW04, KCG02, OJW03, OW00, OW02, SDR03]
 - Installing **Adaptive Filters** at data sources
 - Instructing the data sources not to transmit the unimportant updates

Adaptive Filter Model of DSMS with Error Tolerance



Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Query Model and Filters

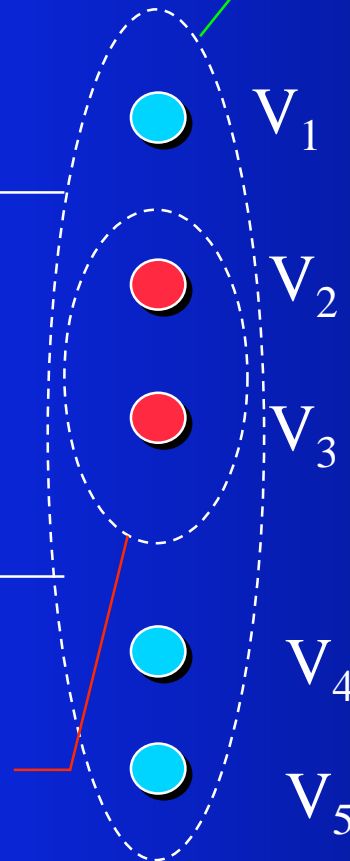
The background of the slide is a gradient of blue and black. The top half is a dark blue, which transitions into a lighter blue at the bottom. A curved, lighter blue shape is visible in the lower-left quadrant, and a black shape is in the lower-right quadrant.

Range Query

All Streams (S),
e.g. $S = \{s_1 \dots s_5\}$

Upper
Bound (u)

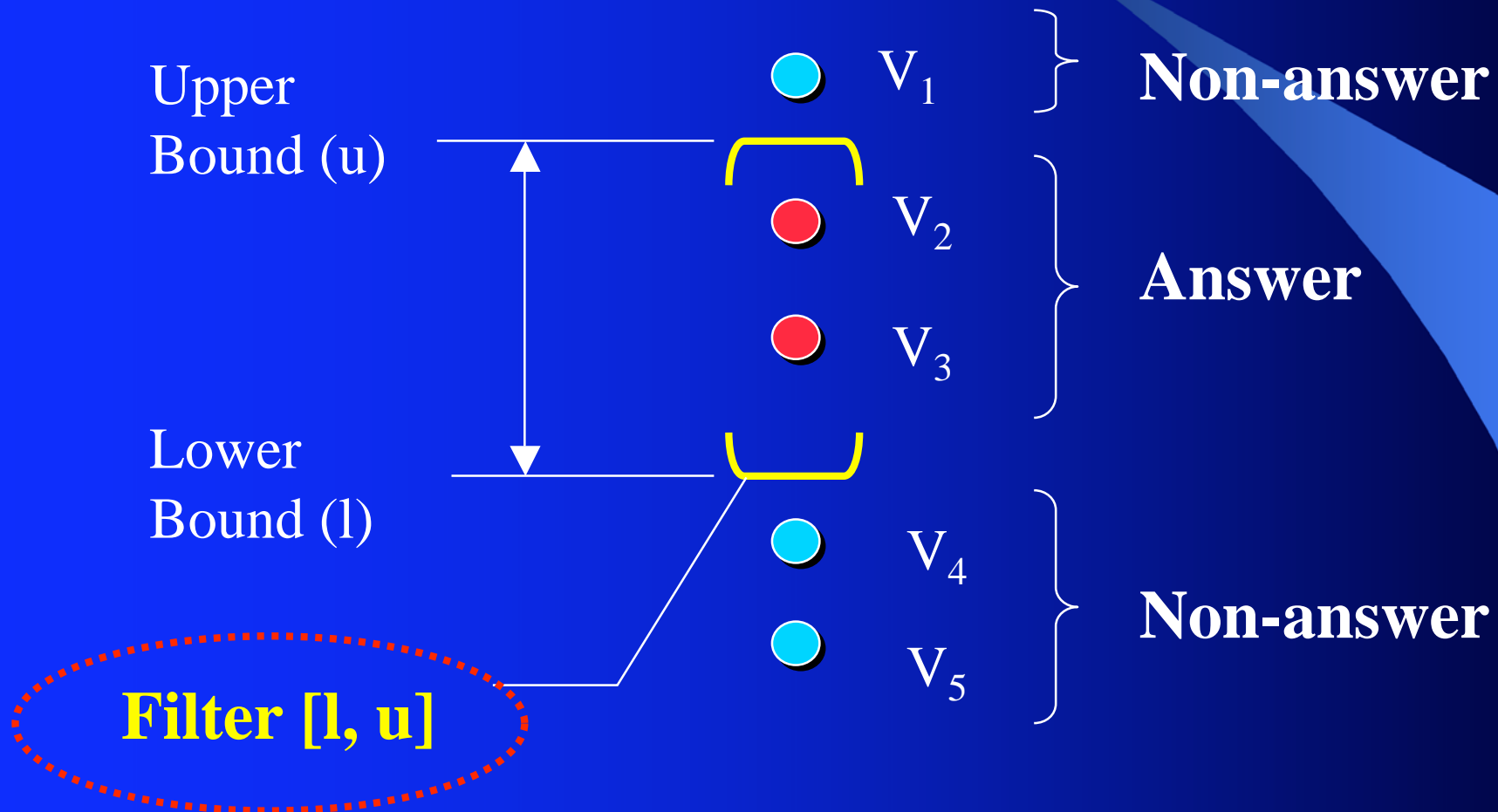
Lower
Bound (l)



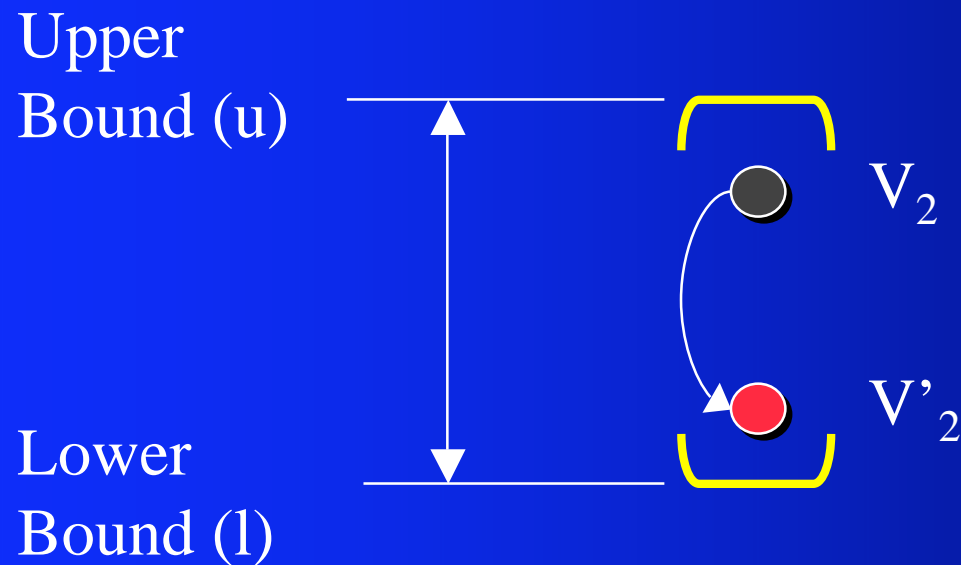
**SELECT S.id AS A
FROM AllStreams S
WHERE S.V
BETWEEN L AND U**

**Answer (A) is Set of
Identities, e.g. $A = \{s_2, s_3\}$**

Query and Filter Model



Update under Filter Model

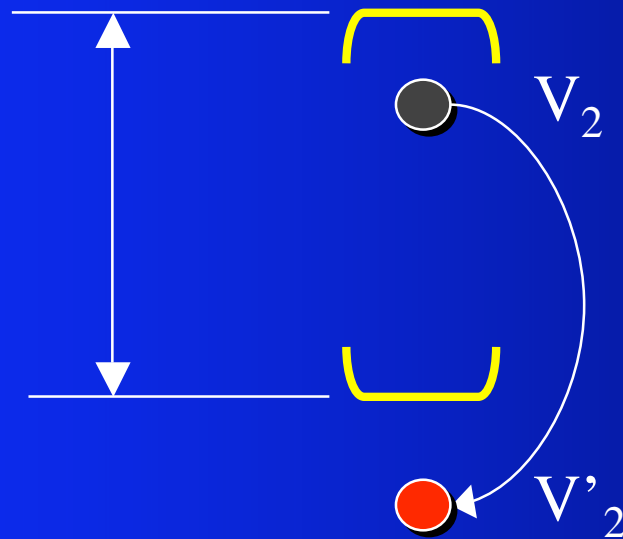


$V \in [l, u] \wedge V' \in [l, u]$
No Update!

Update under Filter Model

Upper
Bound (u)

Lower
Bound (l)



$V \in [l, u] \wedge$
 $V' \notin [l, u]$
Update!

Update under Filter Model

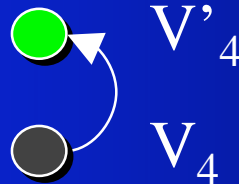
Upper
Bound (u)



Lower
Bound (l)



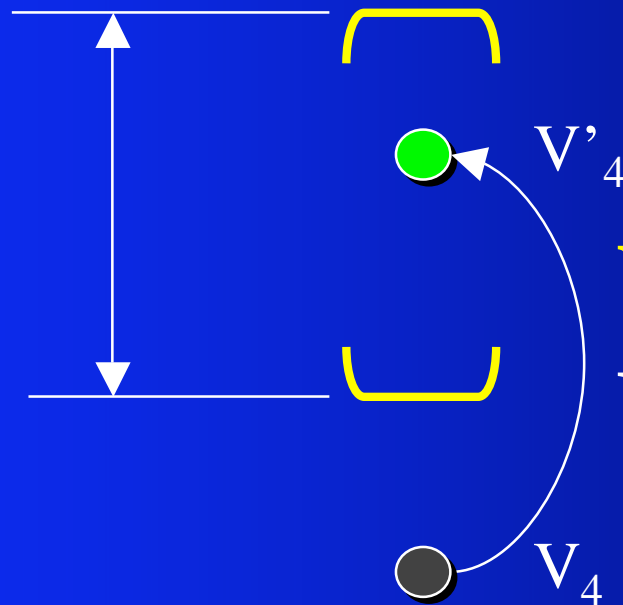
$V \notin [l, u] \wedge V' \notin [l, u]$
No Update!



Update under Filter Model

Upper
Bound (u)

Lower
Bound (l)



$V \notin [l, u] \wedge V' \in [l, u]$
Update!

Zero Tolerance (ZT)

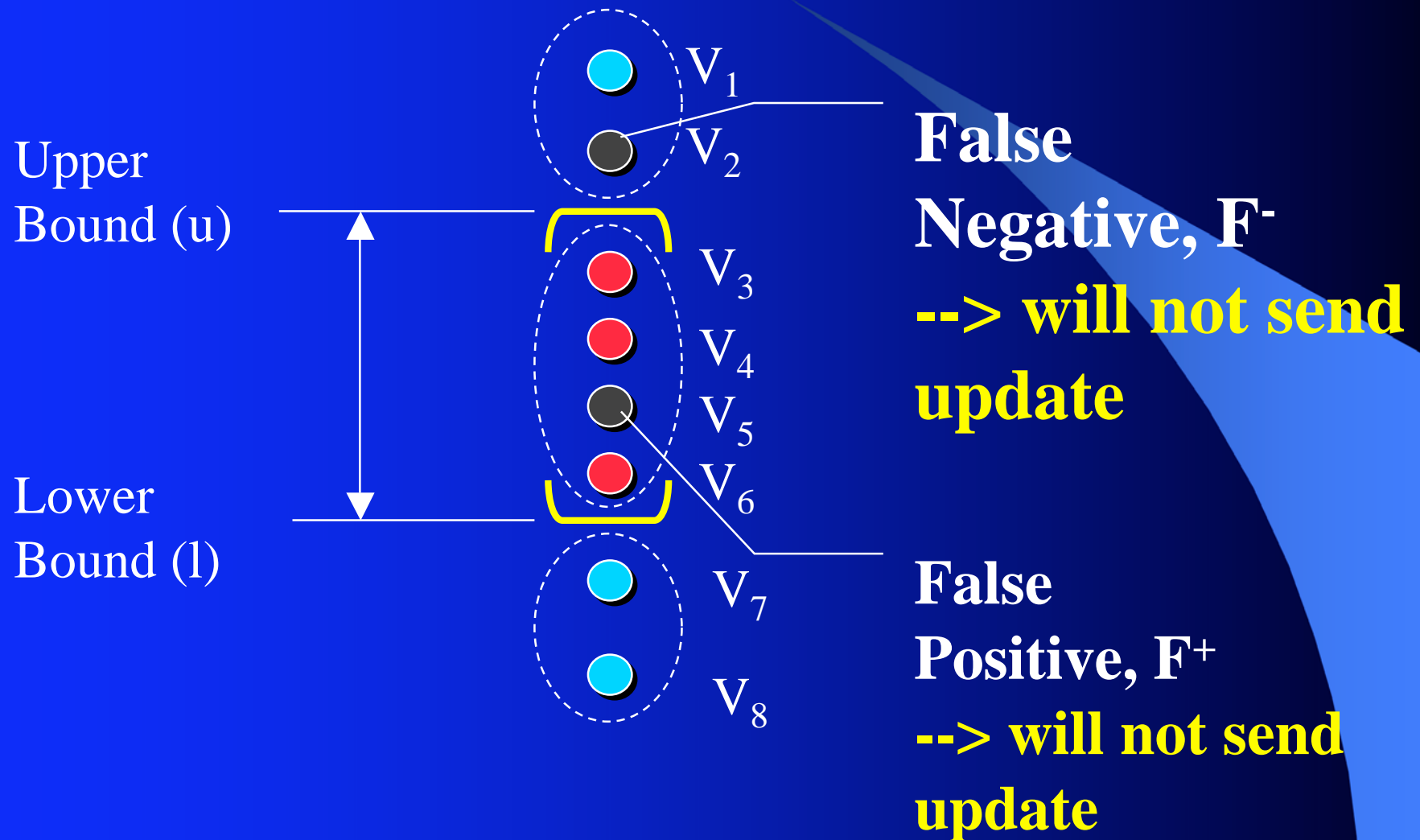
- ZT - Every cross-bound event is updated
- What about if non-zero tolerance is allowed?

Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

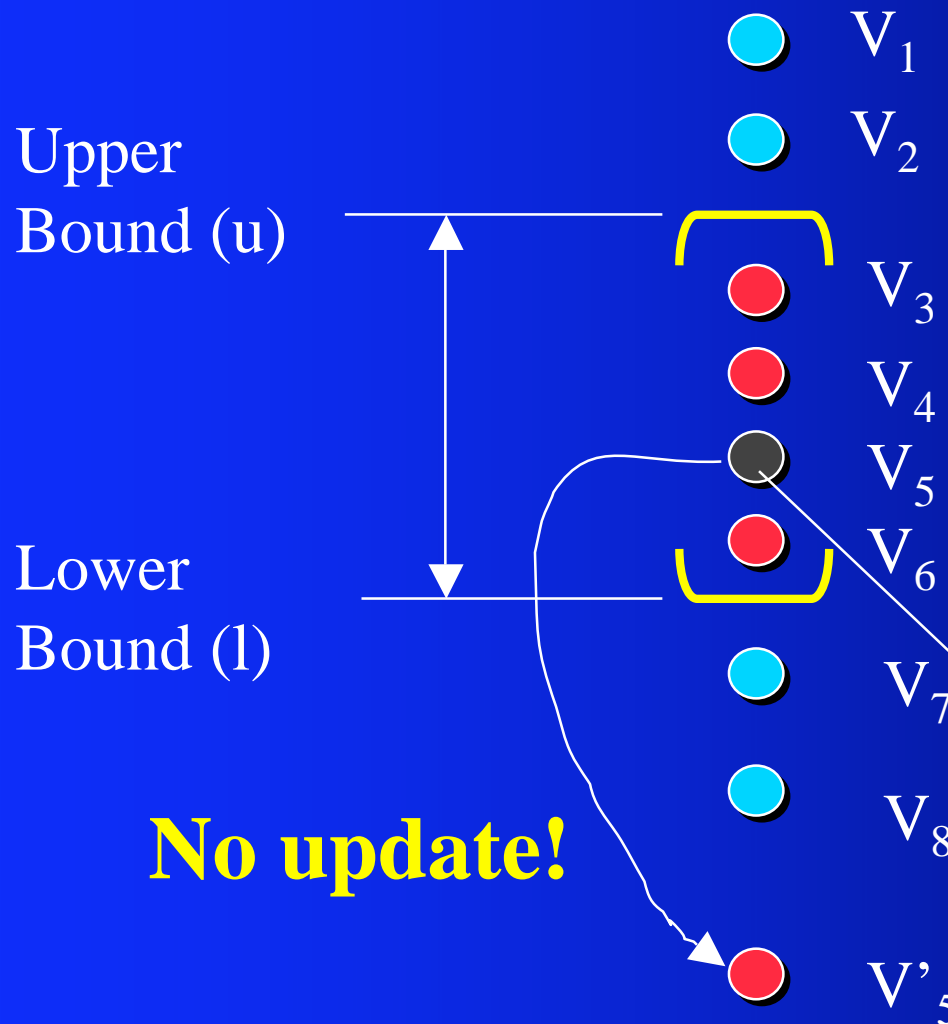
Fraction-based Tolerance

A decorative graphic element consisting of a blue gradient shape that starts as a thin wedge at the top and expands into a larger, curved shape at the bottom, set against a dark blue background.

Fraction-based Tolerance (FT)



False Positive Error Tolerance

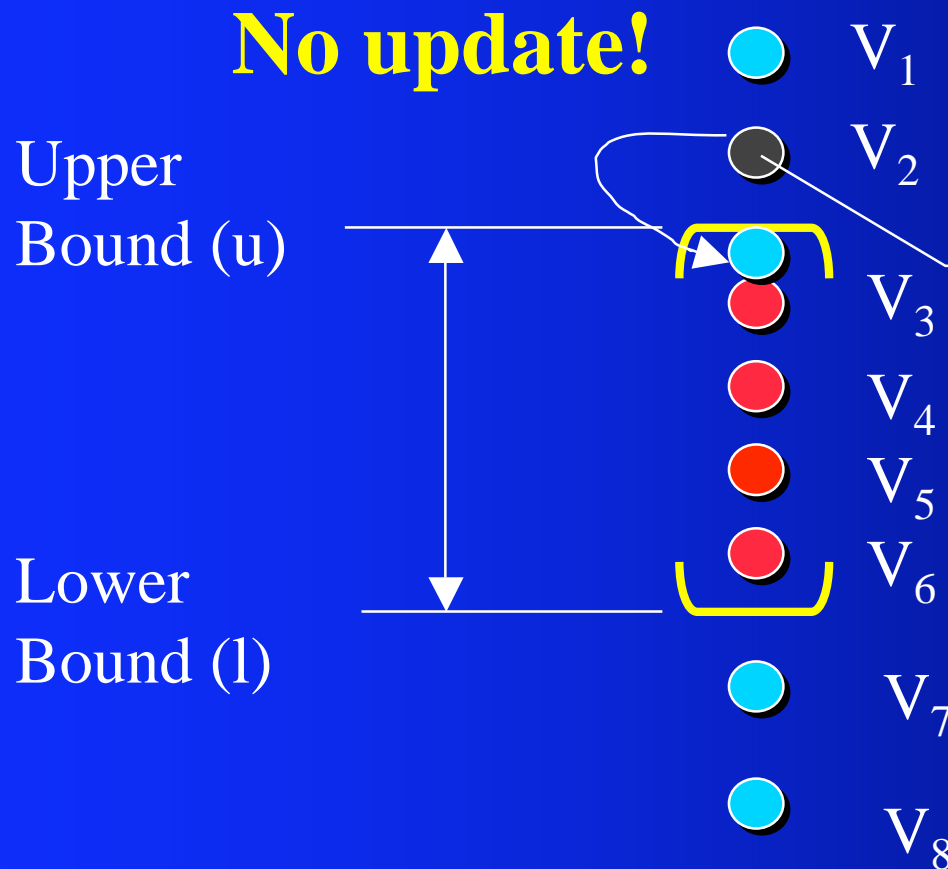


Correctness Criteria

$$\varepsilon^+ \geq F^+$$

$$F^+ = \frac{E^+}{|A|} = 25\%$$

False Negative Error Tolerance



$$F^- = \frac{E^-}{|A| - E^+ + E^-}$$

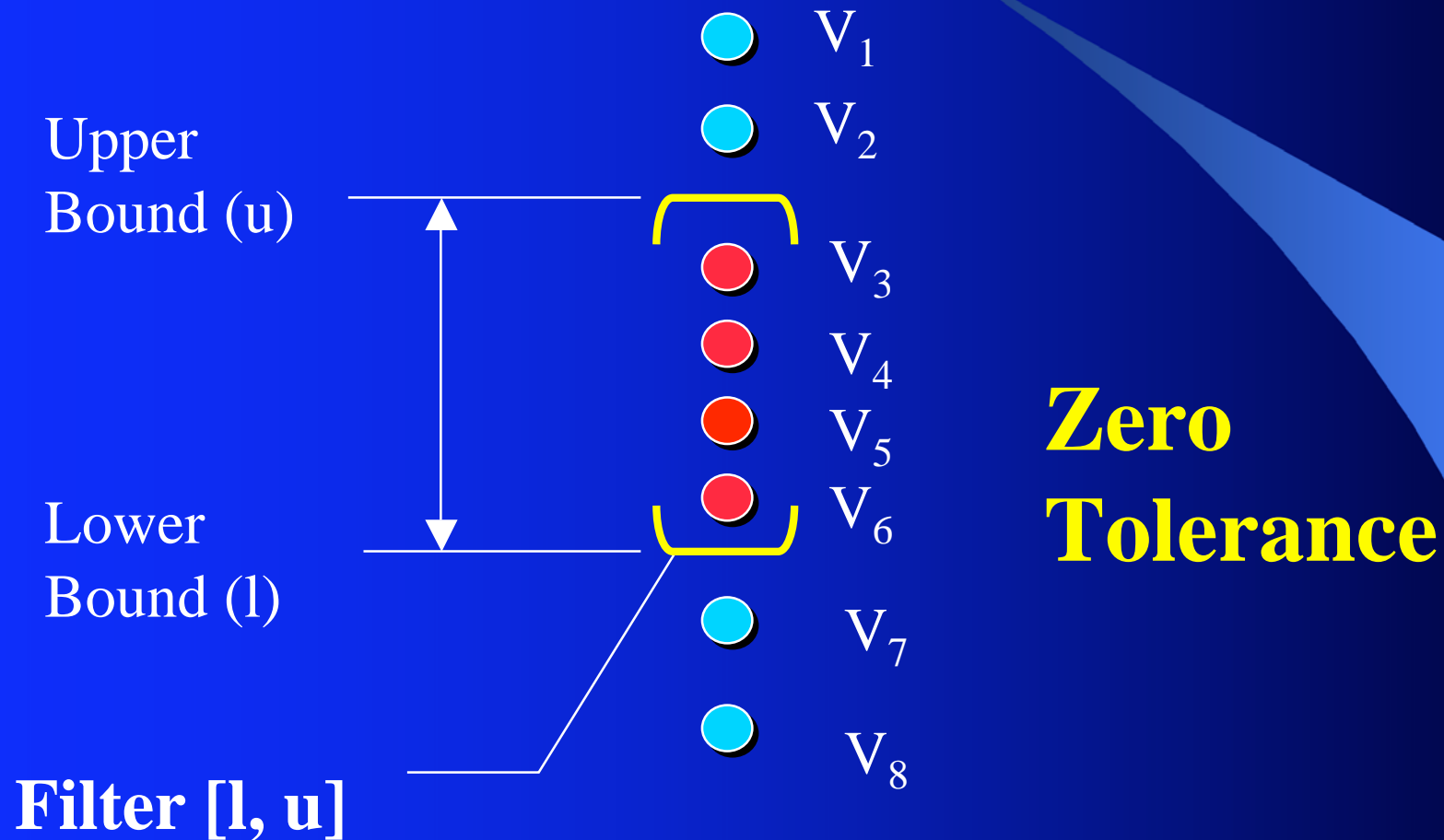
Correctness Criteria

$$\varepsilon^- \geq F^-$$

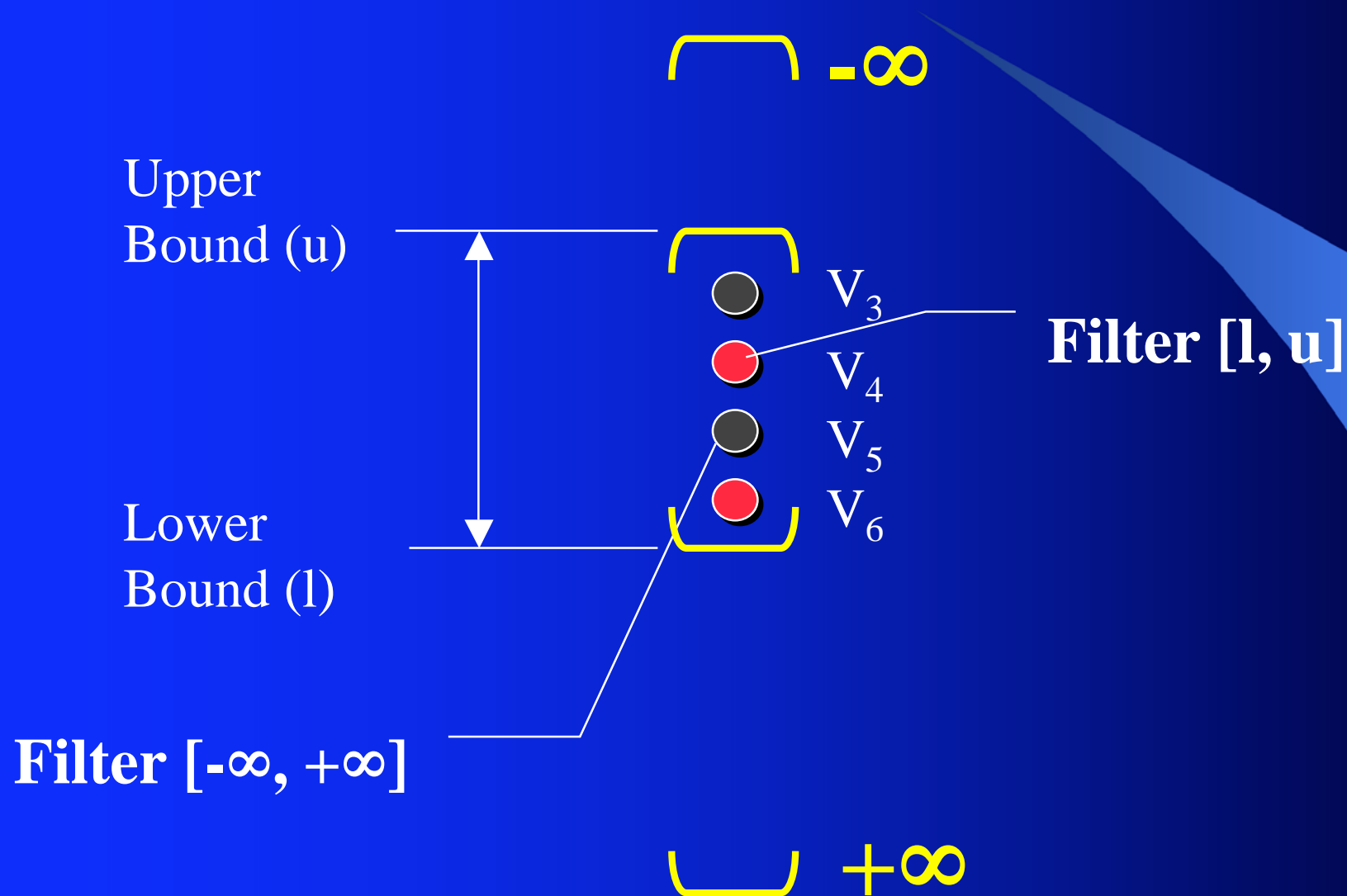
Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Filter Model for Fraction-based Tolerance

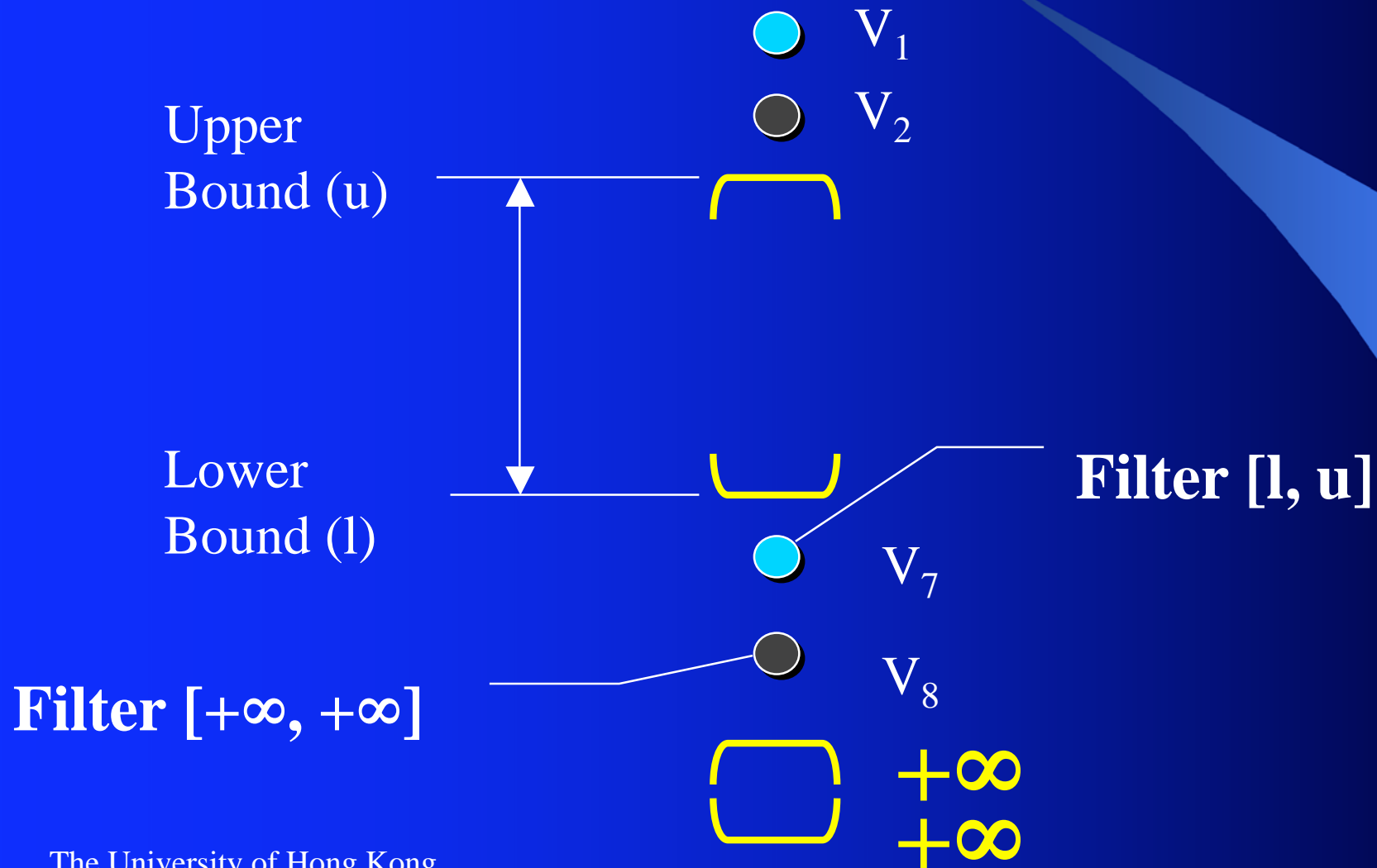
Filter Model for Error Tolerance



Filter Model for FP Streams



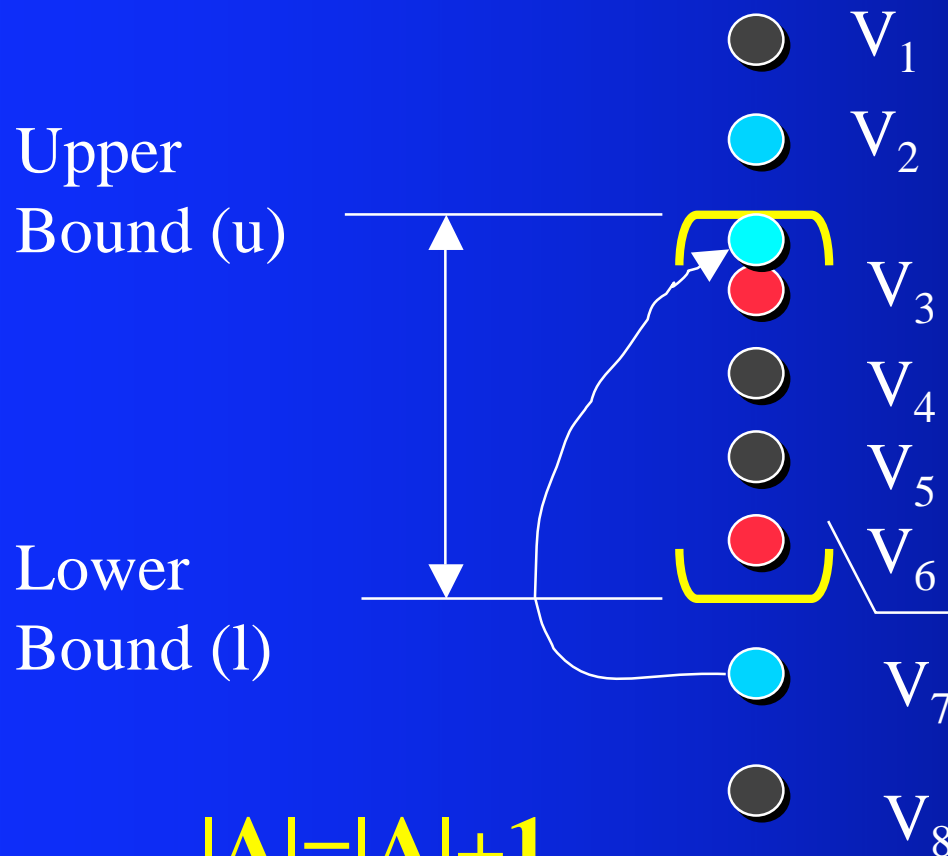
Filter Model for FN Streams



Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

The Fraction-based Tolerance Protocol for Non-ranked Queries (FT-NRP)

Fraction-based Tolerance Protocol for Non-Ranked Queries (FT-NRP)



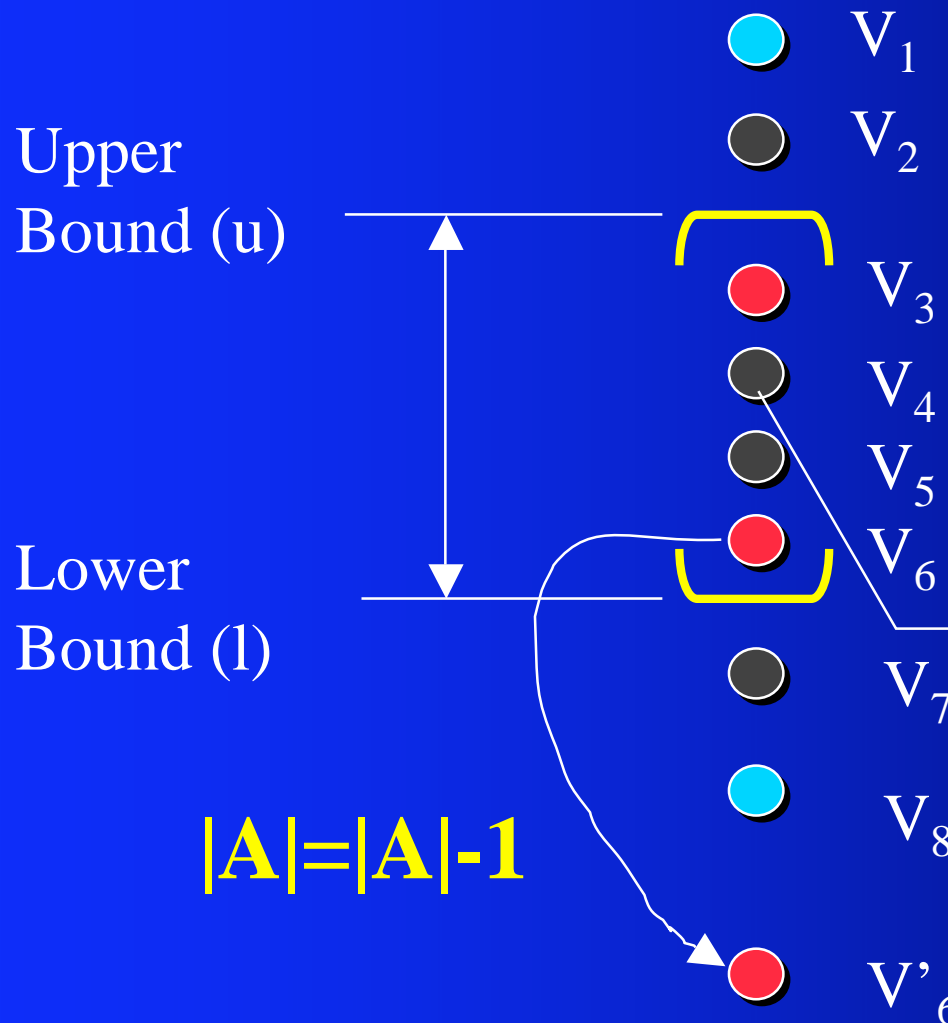
$$\epsilon^+ = 0.5$$

$$\epsilon^- = 0.5$$

$$F^+ = \frac{E^+}{|A|} = \frac{2}{5}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{5}$$

Fraction-based Tolerance Protocol for Non-Ranked Queries (FT-NRP)



$$\epsilon^+ = 0.5$$

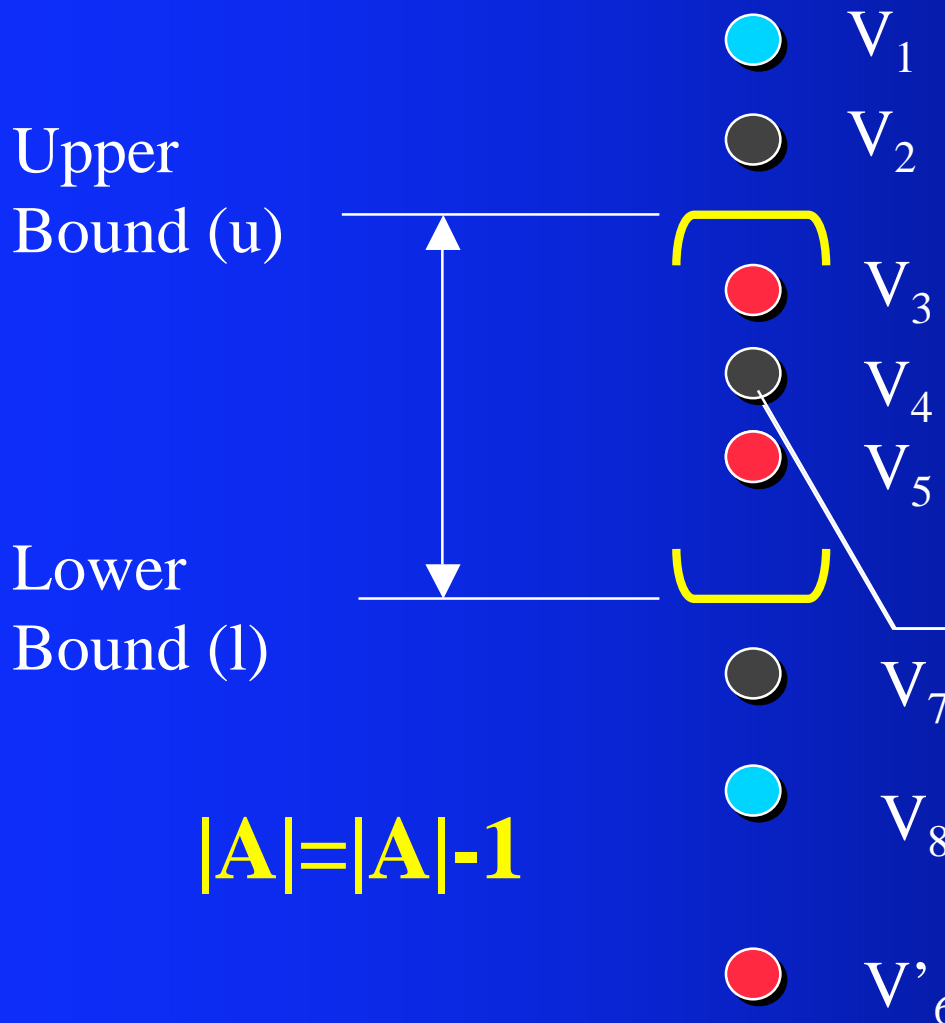
$$\epsilon^- = 0.5$$

$$F^+ = \frac{E^+}{|A|} = \frac{2}{3}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{3}$$

$$F^+ > \epsilon^+, F^- > \epsilon^- :- ($$

Fraction-based Tolerance Protocol for Non-Ranked Queries (FT-NRP)



$$|A| = |A| - 1$$

$$\epsilon^+ = 0.5$$

$$\epsilon^- = 0.5$$

$$F^+ = \frac{E^+}{|A|} = \frac{1}{3}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{4}$$

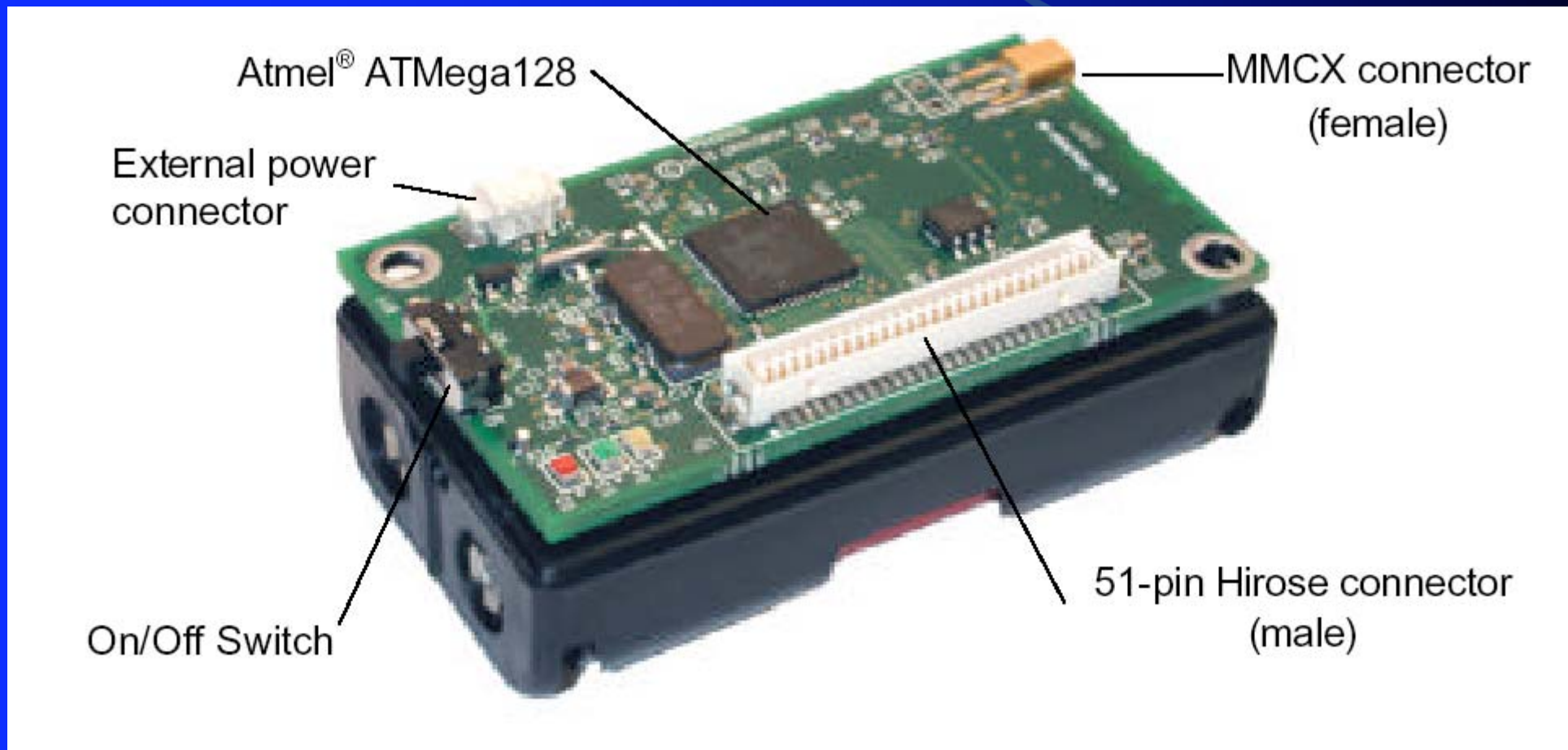
$$F^+ \leq \epsilon^+, F^- \leq \epsilon^- :-)$$

Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Energy Consumption Model of
Sensors

A decorative graphic element consisting of a blue gradient shape that starts as a thin wedge on the left and expands into a larger, curved shape on the right, positioned in the lower half of the slide.

Crossbow's Mote Sensor Board



Crossbow's Mote Sensor Board

- Battery Power: 3.6V
- Sensing Unit Current Consumption
 - Sleeping mode: 5 μ A
 - Sensing: 5mA
- Radio Unit
 - Sleeping mode: 2 μ A
 - Transmission: 12mA
 - Reception: 8mA
- Duty Cycle: 1%

Cost Model of Energy Consumptions

- d_s : Duration of sleeping mode
- d_w : Duration of waken mode
- C_u : Number of Uplink messages
- C_d : Number of Downlink messages
- **Energy for Sensing = $(0.19782d_w + 0.018d_s)mJ$**
- **Energy for Radio Transmission = $43.2 C_u mJ$**
- **Energy for Radio Reception = $28.8 C_d mJ$**

Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Experimental Results on the Basic
FT-NRP

Experimental Result 1 - FT-NRP Range Query with Real Data

- Experiment 1

- Evaluate #Messages vs. settings of epsilon +/-

- Data:

- Real Internet traffic data (TCP data)

- No. of streams = 1600

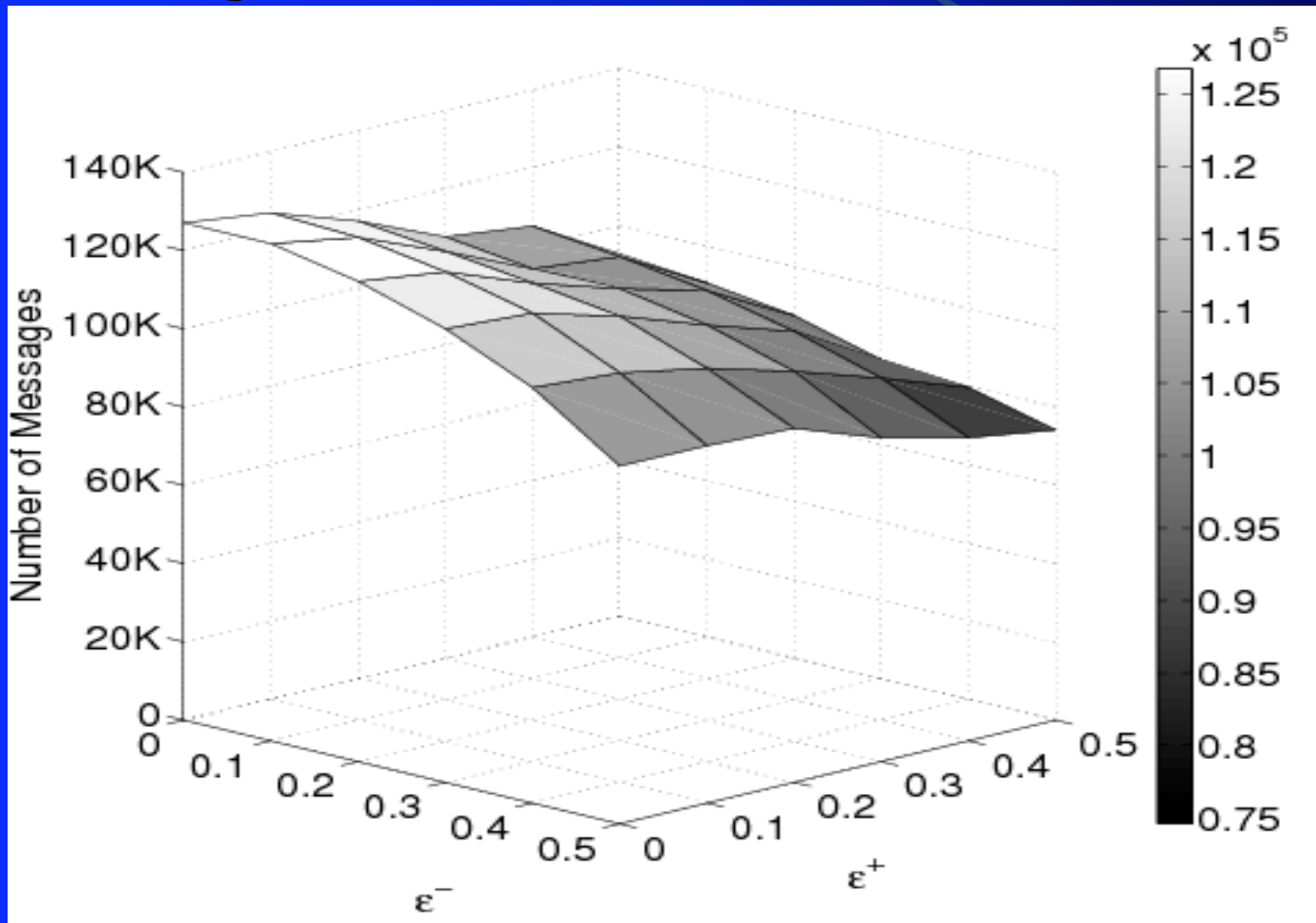
- Query:

- Range = [200, 350]

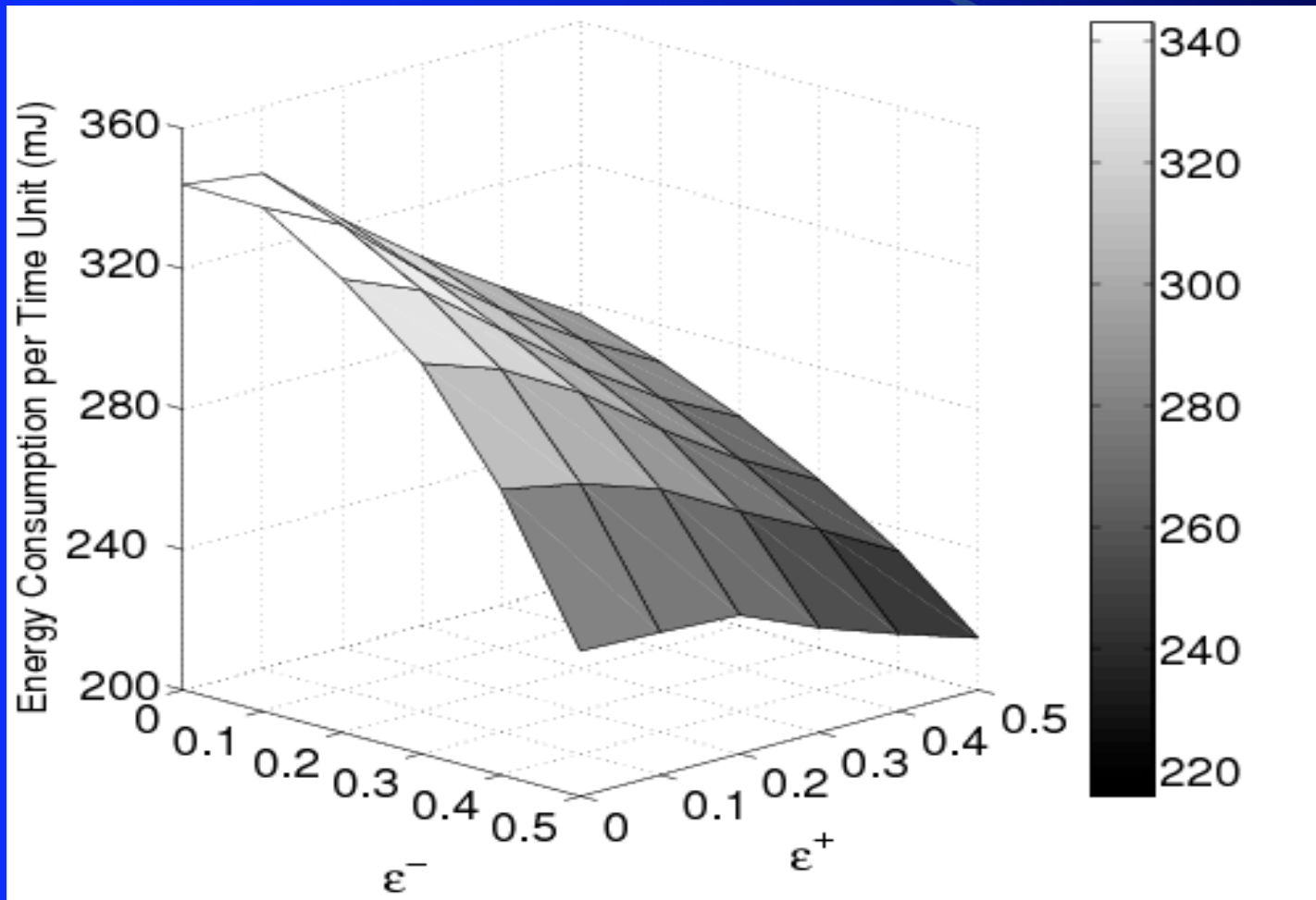
- Duration = 2,000,000 time units

Experimental Result 1 - FT-NRP

#Messages with Real Data



Experimental Result 1 - FT-NRP Energy Consumption with Real Data



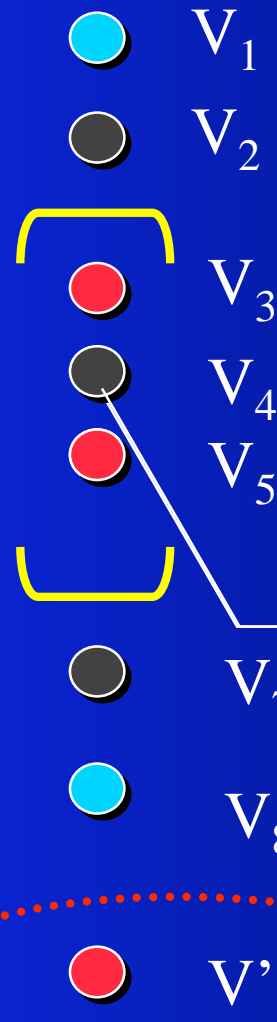
Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Enhancements 1:
Immediate Compensation

Problem of FT-NRP

**E^+ and E^- decrease as
Maintenance of
Correctness!!!**

**--> Slow down the lost
of infinite filters**



$$\varepsilon^+ = 0.5$$

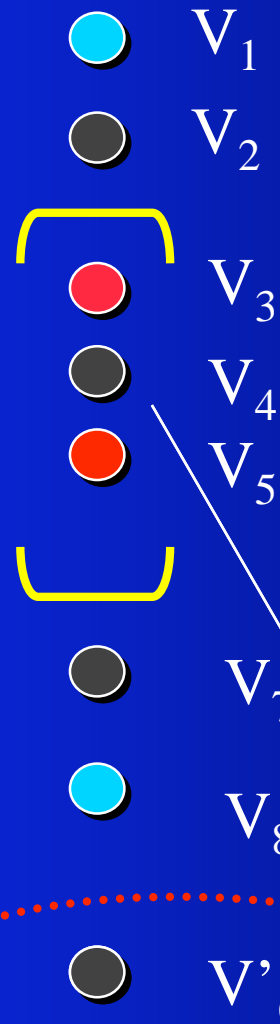
$$\varepsilon^- = 0.5$$

$$F^+ = \frac{E^+}{|A|} = \frac{1}{3}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{4}$$

Immediate Compensation for FT-NRP

Immediate set the s_6 as False Positive!



$$\varepsilon^+ = 0.5$$

$$\varepsilon^- = 0.5$$

$$F^+ = \frac{E^+}{|A|} = \frac{2}{4}$$

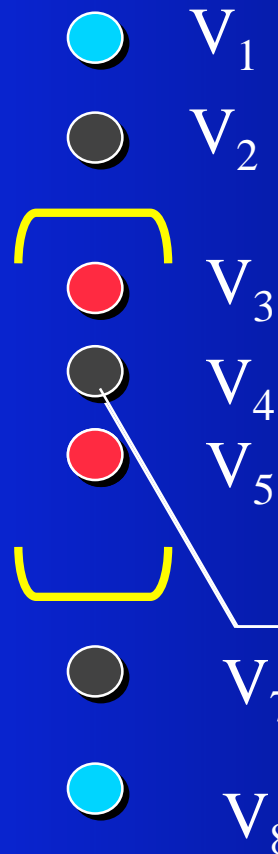
$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{4}$$

Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Enhancements 2:
Incremental Deployment

Problem of FT-NRP

**Filter must be consumed
for Maintenance of
Correctness!!!**



$$\varepsilon^+ = 0.5$$

$$\varepsilon^- = 0.5$$

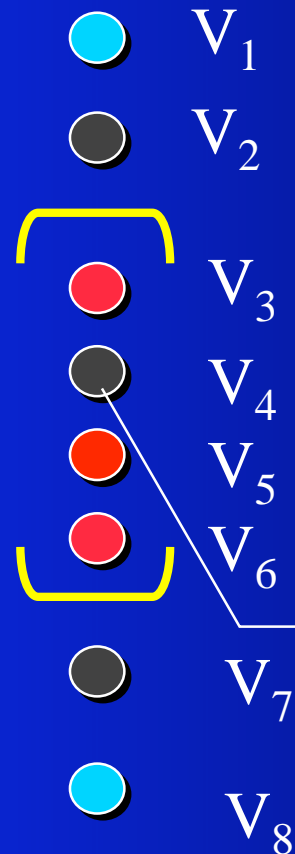
$$F^+ = \frac{E^+}{|A|} = \frac{1}{3}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{4}$$



Incremental Deployment for FT-NRP

Only assign
partial fraction of
filters -->
SLACK!



$$\varepsilon^+ = 0.5$$

$$\varepsilon^- = 0.5$$

$$F^+ = \frac{E^+}{|A|} = \frac{1}{4}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{4}$$

Incremental Deployment for FT-NRP

Assign the unused filter only when Update!

→ E^+ and E^- increase as Update happens
 → Message cost for maintenance is saved



$$\varepsilon^+ = 0.5$$

$$\varepsilon^- = 0.5$$

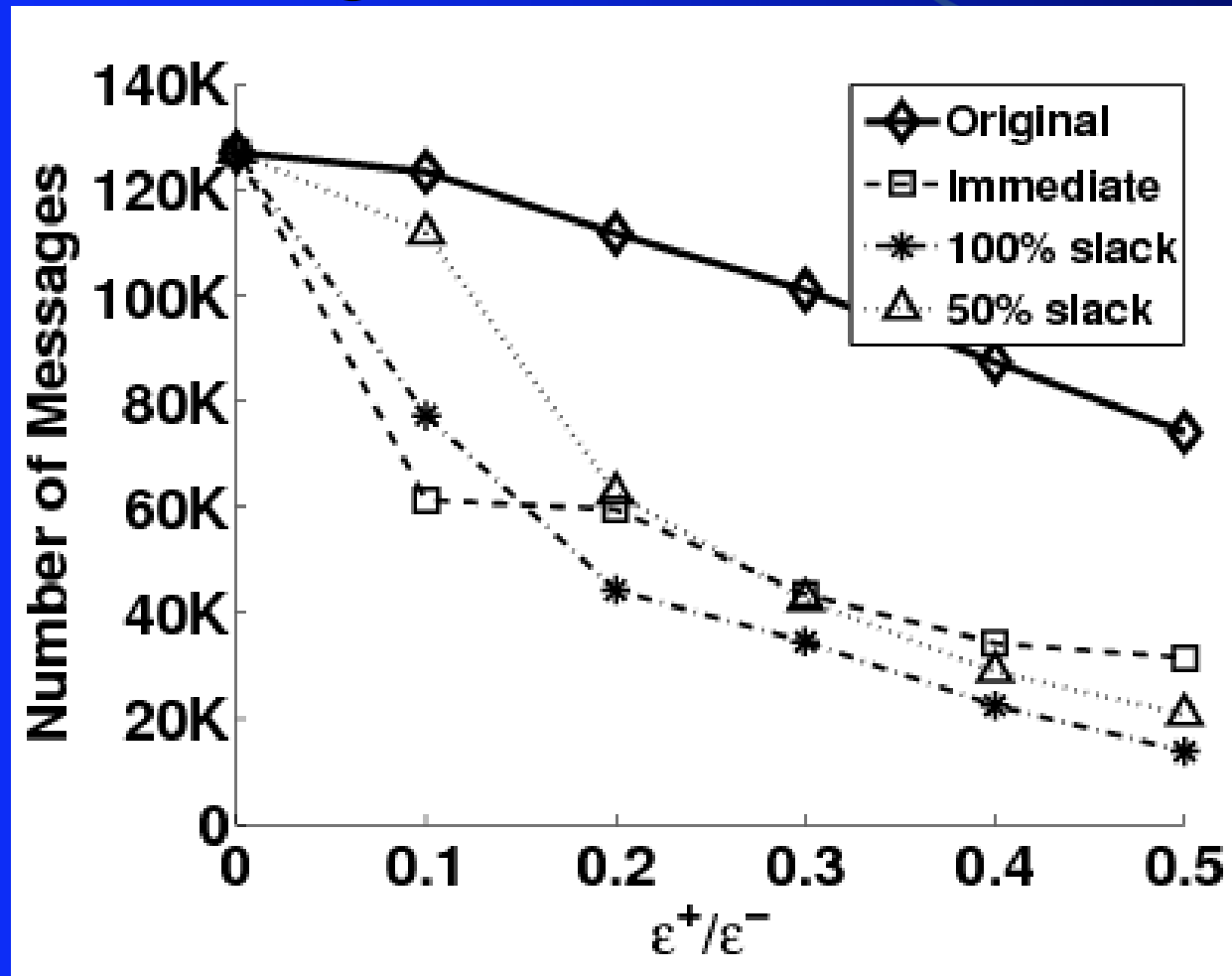
$$F^+ = \frac{E^+}{|A|} = \frac{2}{4}$$

$$F^- = \frac{E^-}{|A| - |E^+| + |E^-|} = \frac{2}{4}$$

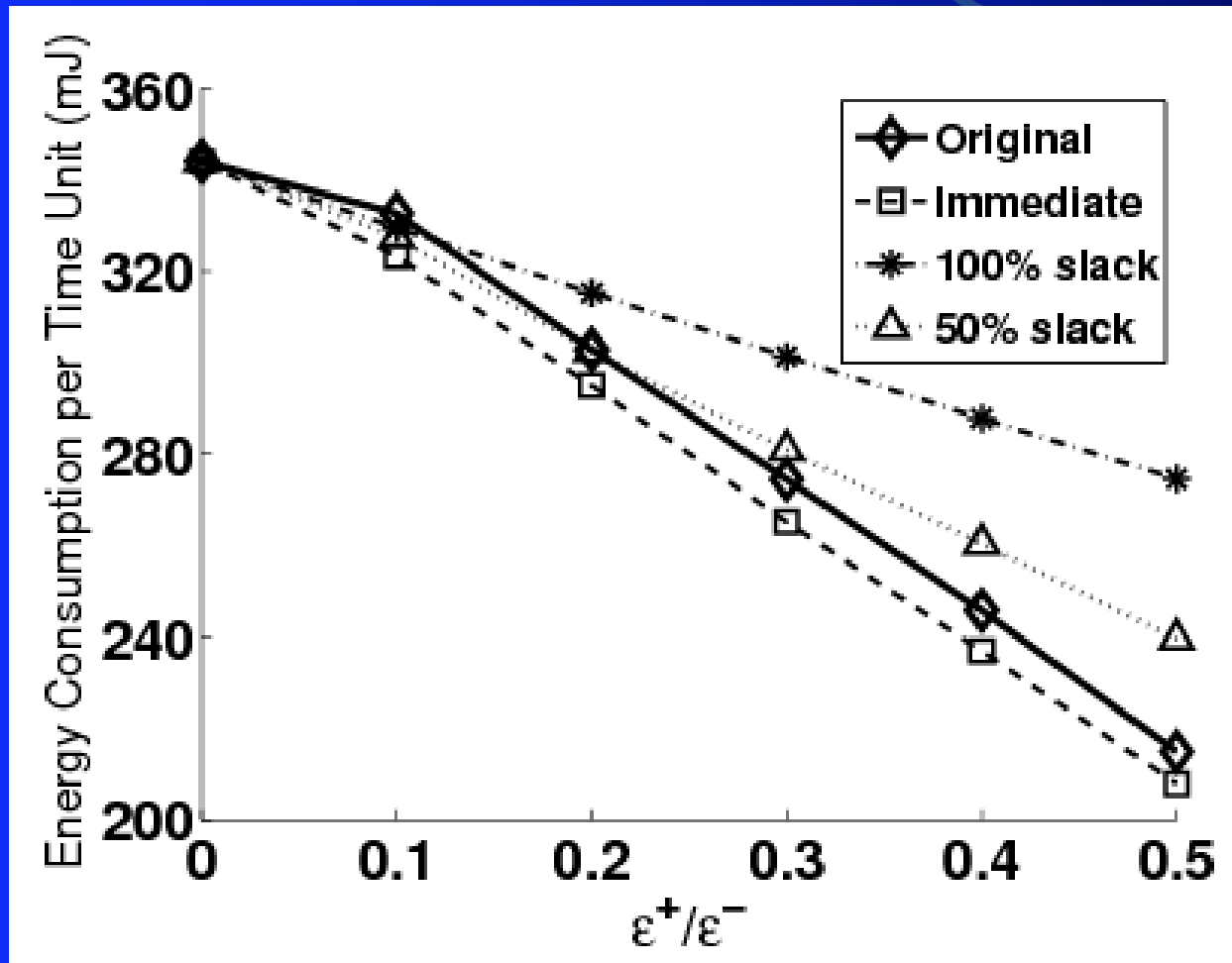
Adaptive Stream Filters for Entity-based Queries with Fraction-based Tolerance

Experimental Results of the Enhancements on FT-NRP

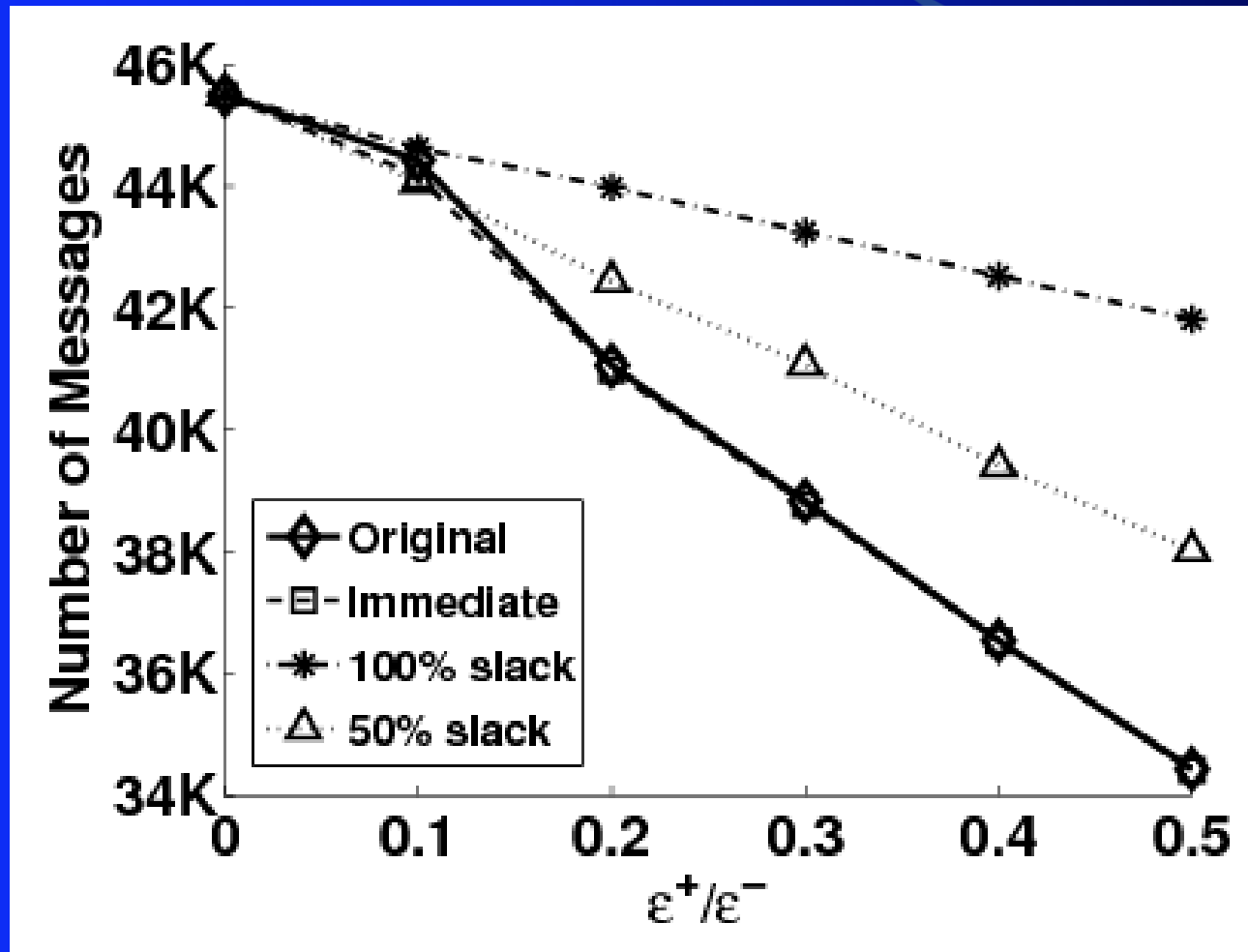
Experimental Result 2 - Performance on #Messages (Real Data)



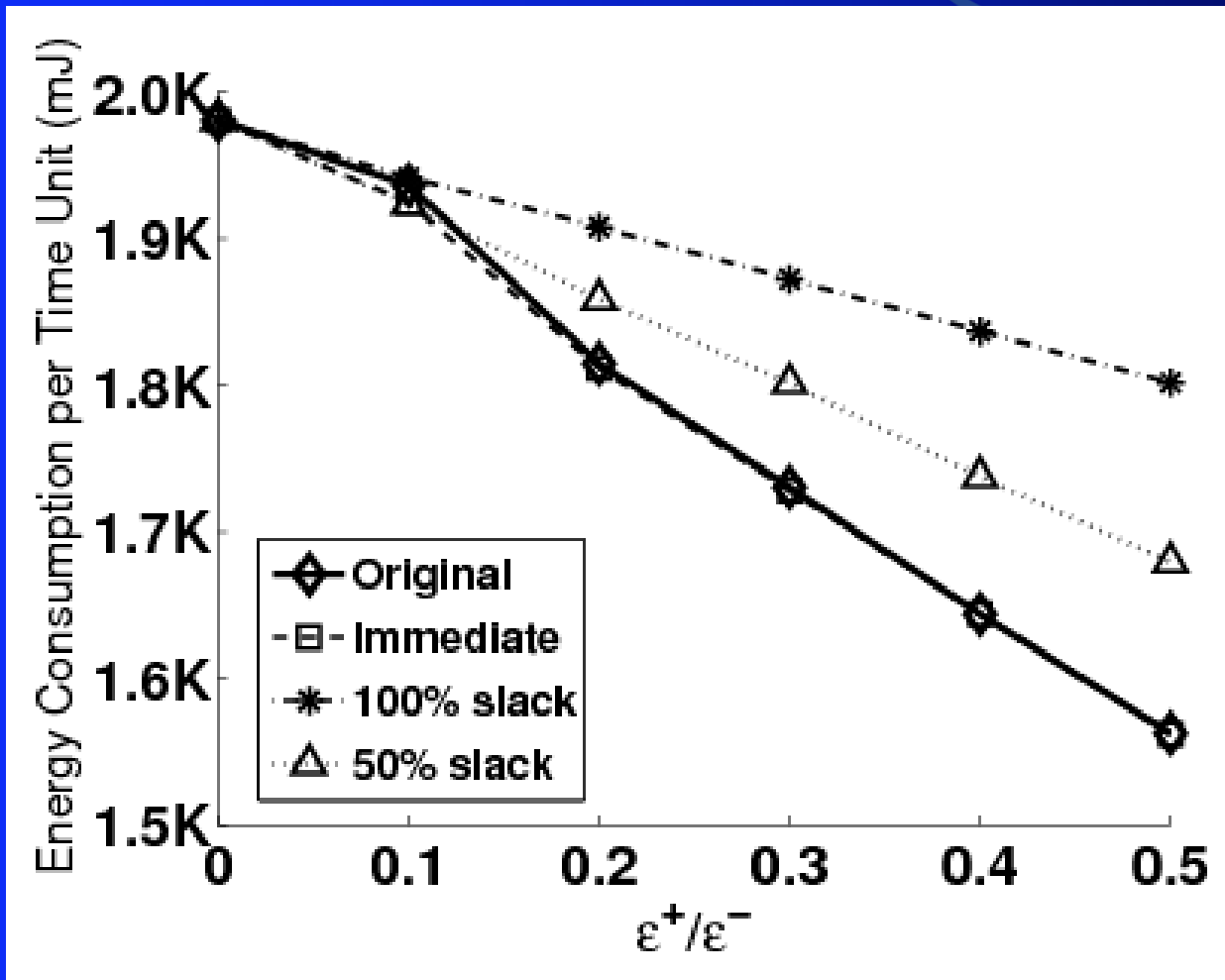
Experimental Result 2 - Performance on Energy (Real Data)



Experimental Result 2 - Performance on #Messages (Synthetic Data)



Experimental Result 2 - Performance on Energy (Synthetic Data)



Summary

- The motivation of using filters to reduce communication bandwidth and battery usage
- Fraction-based Tolerance Protocol for Range Queries (FT-NRP)
- The enhancement works on the FT-NRP
- The experimental results on performance evaluation of the protocols

References

- [ABB03] A. Arasu, B. Babcock, S. Babu, M. Datar, K. Ito, R. Motwani, I. Nishizawa, U. Srivastava, D. Thomas, R. Varma, and J. Widom. **STREAM: The Stanford Stream Data Manager**. *IEEE Data Engineering Bulletin*, 26 March 2003.
- [ACC03] D. Abadiand, D. Carneyand, U. Cetintemel, M. Cherniack, C. Convey, C. Erwin, E.Galvez, M.Hatoun, J.Hwang, A. Maskey, A. Rasin, A. Singer, M. Stonebraker, N. Tatbul, Y. Xing, R. Yan, and S. Zdonik. **Aurora: A data stream management system (demonstration)**. In *Proc. of 2003 ACM SIGMOD Intl. Conf. on Management of data*, 2003.

References

- [CKP05] R. Cheng, B. Kao, S. Prabhaker, A. Kwan, and Y. Tu. **Adaptive Stream Filters for Entity-based Queries with Non-Value Tolerance**. In *Proc. of VLDB*, 2005.
- [ITA00] Lawrence Berkeley National Laboratory. **The internet traffic archive, usa**. *http://ita.ee.lbl.gov*, 2000.
- [JCW04] A. Jain, E. Chang, and Y. Wang. **Adaptive stream resource management using kalman filters**. In *Proc. of ACM SIGMOD Intl. Conf. on Management of data*, 2004.
- [Crossbow] Crossbow. **MPR - Mote Processor Radio Board MIB - Mote Interface / Programming Board User's Manual**.

References

- [MSH02] S. Madden, M. Shah, J. M. Hellerstein, and V. Raman. **Continuously adaptive continuous queries over streams.** In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, June 2002.
- [NR03] J. Ni and C. V. Ravishankar. **Probabilistic spatial database operations.** In *Proc. of 8th Intl. Symposium on Spatial and Temporal Databases (SSTD)*, 2003.
- [OJW03] C. Olston, J. Jiang, and J. Widom. **Adaptive filters for continuous queries over distributed data streams.** In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, 2003.

References

- [OW00] C. Olston, and J. Widom. **Offering a precision-performance tradeoff for aggregation queries over replicated data.** In *Proc. of 26th VLDB*, September 2000.
- [OW02] C. Olston, and J. Widom. **Best-effort cache synchronization with source cooperation.** In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, June 2002.
- [SDR03] S. Shah, S. Dharmarajan, and K. Ramamritham. **An efficient and resilient approach to filtering and disseminating streaming data.** In *Proc. of the 29th VLDB Conf.*, 2003.

Thank You!