A Real-Time Event Stream Processing System for RFID Monitoring Applications

YU Ge

School of Information Science and Engineering

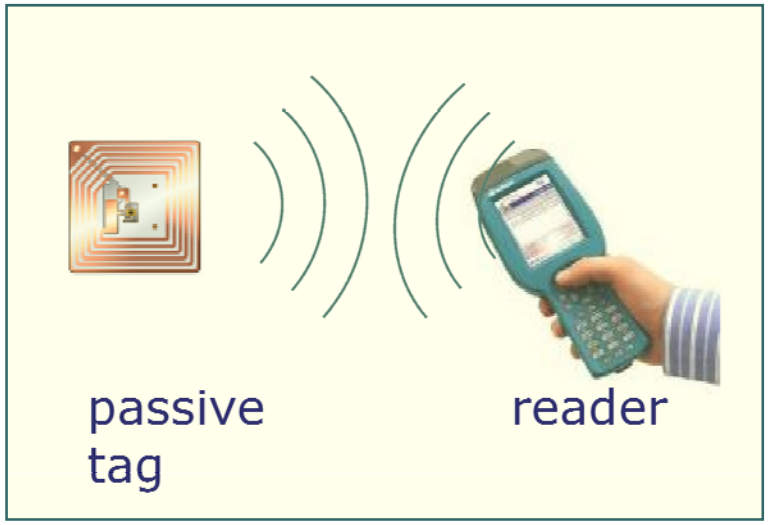
Northeastern University, China

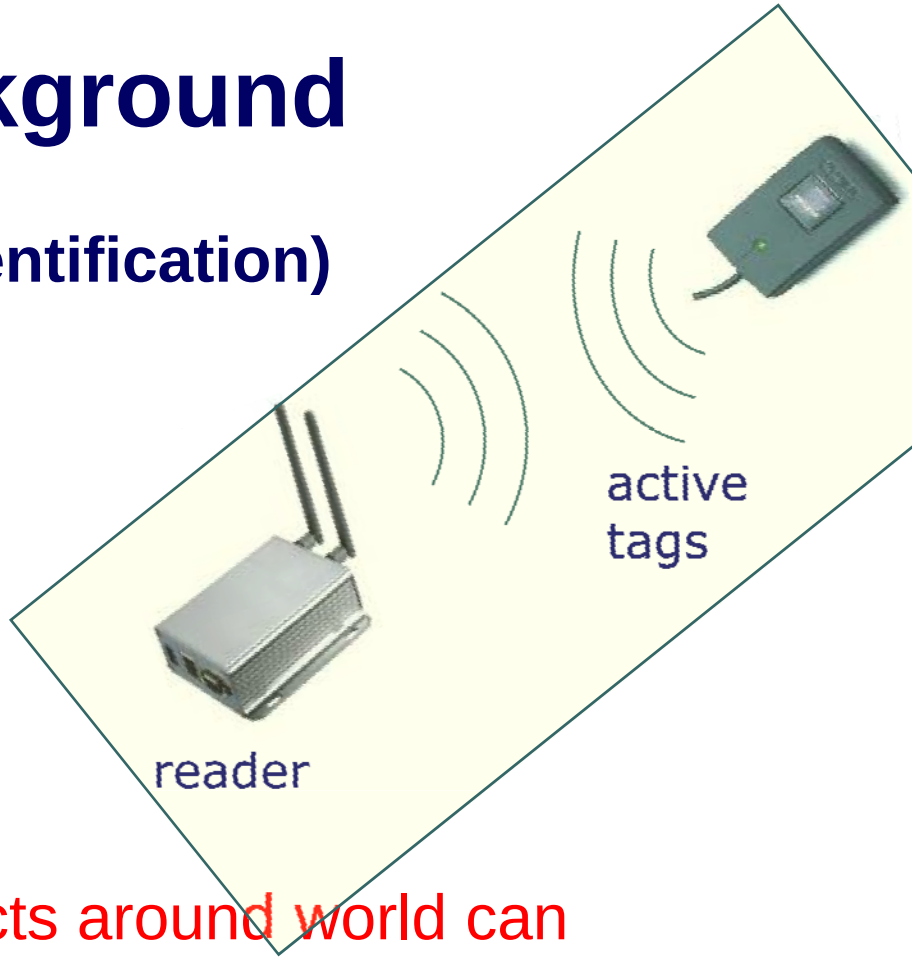
Outline

- Research Background
- Research Issues in Event-Centric Processing
- Overview of the REvent System
- Key Techniques
- Summary

Research Background

- **RFID (Radio Frequency IDentification)**





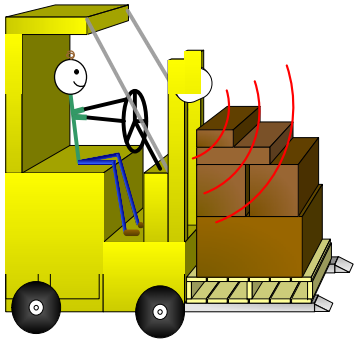
Through RFID, Billions of objects around world can report their identies, positions, histories, etc. => An internet of things![ITU Internet Reports 2005]

Research Background

- **Applications of RFID**
  - **being applied**
    - logistics management
    - asset tracking
    - access control and personal tracking
    - sensor network

    ……
  - **in future**
    - goods flow
    - people flow
    - equipment flow
    - money flow

    ……



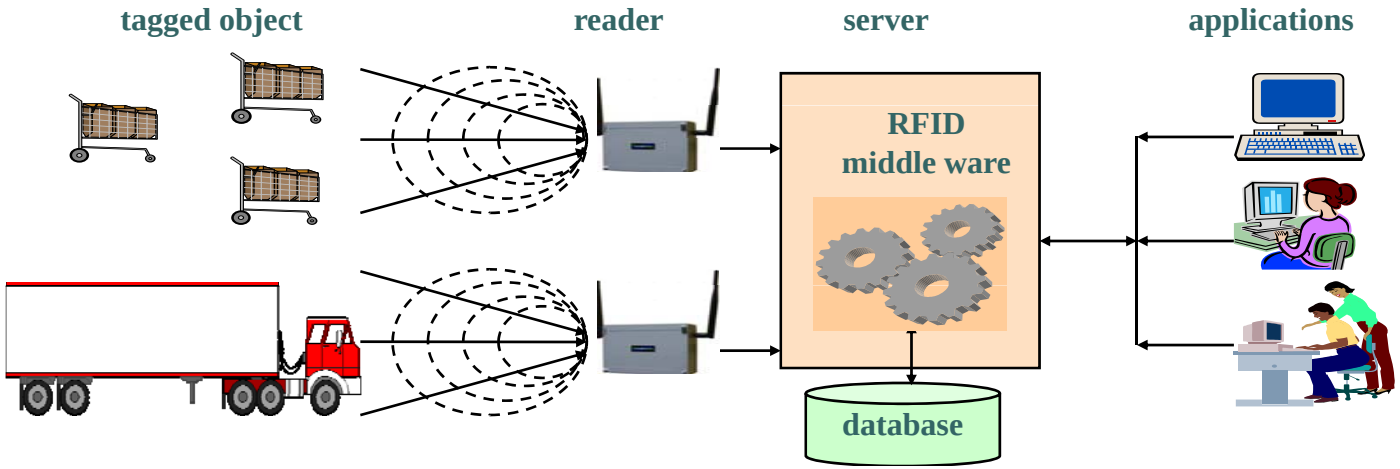
- A huge amount of RFID data could be generated continually
- need to be processed in real-time!

Research Background

- **Characteristics of RFID data**
  - **Spatio-temporal, dynamics and correlations**
    - Data contains time, location, and status
  - **Rich semantics.**
    - Objects carry a lot of information
    - related with its context status and background knowledge
  - **Uncertainty and heterogeneity**
    - missing readings, and repeat readings
    - dirty data
  - **Streaming, batching and massive volume**
    - automatically generated rapidly in form of streaming
    - Objects must be checked in a batch
    - needs to be processed in real time

Research Background

- **Current solution for RFID applications**
  - **data-centric systems**
    - Using traditional database technology, e.g. active databases, real-time databases.
  - **event-centric systems**
    - treat events as first class citizen
    - perform event processing directly on data source
    - typical systems: SASE system and Cayuga system



Research Issues

- **CEP (**Complex Event Processing)
- Four basic approaches of CEP
  1. finite automata based model
  2. Petri-net based model
  3. tree-based model
  4. directed graph based model

Research Issues

- Challenging
  - **Complex semantics, massive data volume, and inaccurate raw data**
    - ➢ to effectively manage RFID data, stream-style processing mechanisms is needed
  - **Complex constraints**
    - temporal order
    - numeric value interval
    - ➢ traditional ECA rules are not feasible
    - ➢ new event model and algebraic operations are needed

Research Issues

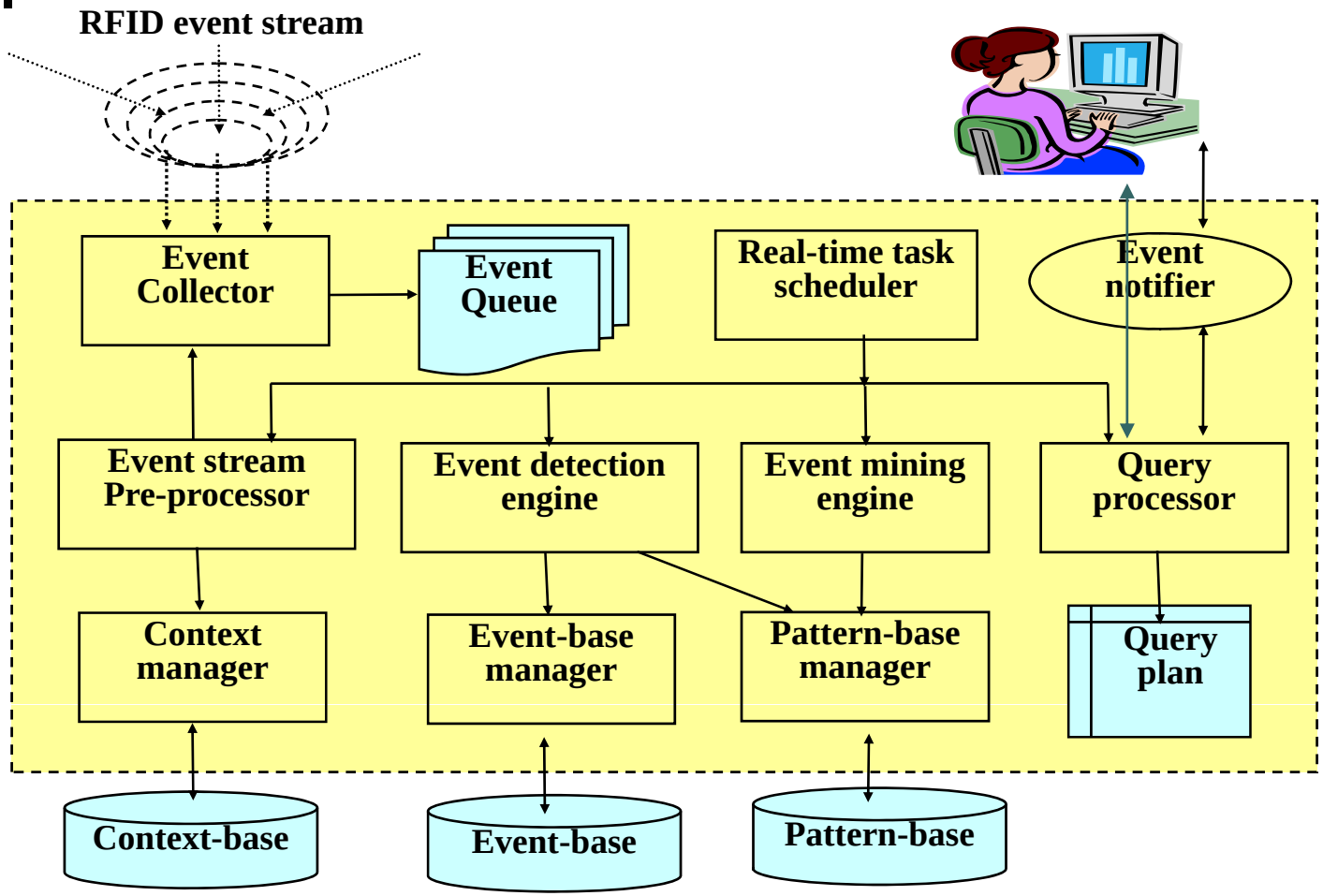
- Challenging
  - **Non-spontaneous composite events**
    - negation events
    - repetition events
    - ➢ the occurrence of the constituent event can not lead to the occurrence of the composite event
    - ➢ special query processing algorithms are needed
  - **Many queries registered over the same event stream**
    - ➢ new query optimization policy is required

Research Issues

- **Problems remain to be solved:**
  - how to process event happening orders of “and”, ”or” and ”same time”
  - how to compose composite events
  - how to make use of query sharing for optimization
  - how to schedule real-time tasks to meet deadline
  - how to deal with uncertain data with probability
  - how to perform distributed CEP

REvent : high-performance real-time event stream processing system

Overview of the REvent System



- The architecture of REvent system

Overview of the REvent System

- **Major modules of REvent system**
  - Real-time task scheduler
    - the proxy of the whole system
  - Query processor
    - registers queries for event, makes query execution plan
  - Event pre-processor
    - cleans raw data and generates basic events
  - Event detection engine
    - filtering and detection of composite events by matching rules
  - Event mining engine
    - summary analysis, temporal association mining, classification mining, and outlier detection

Overview of the REvent System

- **Other auxiliary modules:**
  - Context manager
    - provides the information about RFID context and background knowledge.
  - Event-base manager
    - provides temporary events and archived historic events
  - Rule-base manager
    - provides pre-defined rules for detection and stores mining posteriori rules.
  - Event collector
    - gets raw data from RFID readers

Overview of the REvent System

- **The basic data structures**
  - Event queue
    - a buffer containing basic events
  - Query plan
    - a buffer containing execution plans
  - Context-base
    - a database about context and background about the observed objects
  - Event-base
    - a database containing events
  - Rule-base
    - a database containing rules

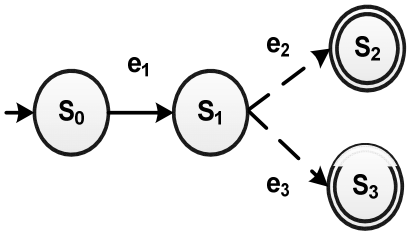
Overview of the REvent System

- Semantic event model
  - Basic event (EId, Etype, TS, Location, OID, Data)
- Event operation algebra
  - Sequence operations
  - Non-sequence operations: negation and repetition
  - Restriction operations: partition, within, continunity.
  - Complex operations : sliding-window, grouping and statistics.
- Query language
  - Express complex rules and application logics
  - XML based script language
  - SQL like event query language

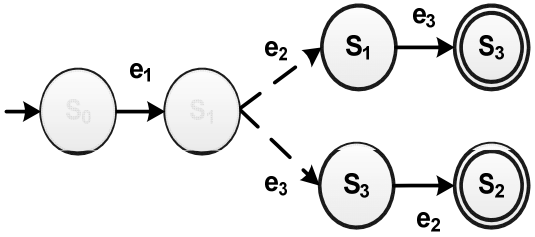
Overview of the REvent System

- **Examples of CEP operations**

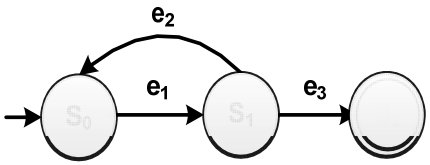
SEQ(E_1, ***DIS***(E_2,E_3))



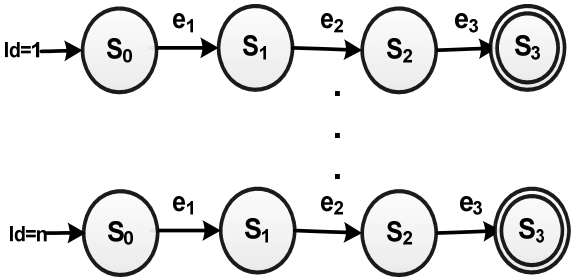
SEQ(E_1, ***CON***(E_2,E_3))



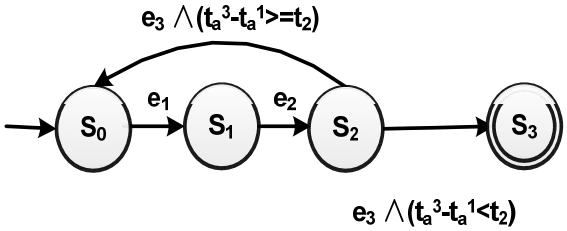
SEQ(E_1, ***NEG***(E_2),E_3)



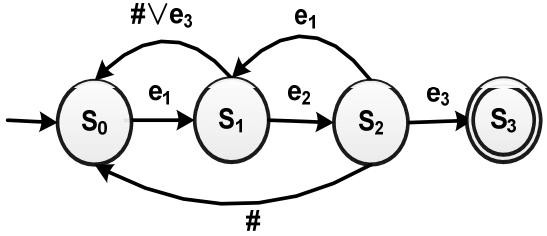
Partition(*SEQ*(E_1,E_2,E_3),id)



Within(*SEQ*(E_1,E_2,E_3), 0,t_2)

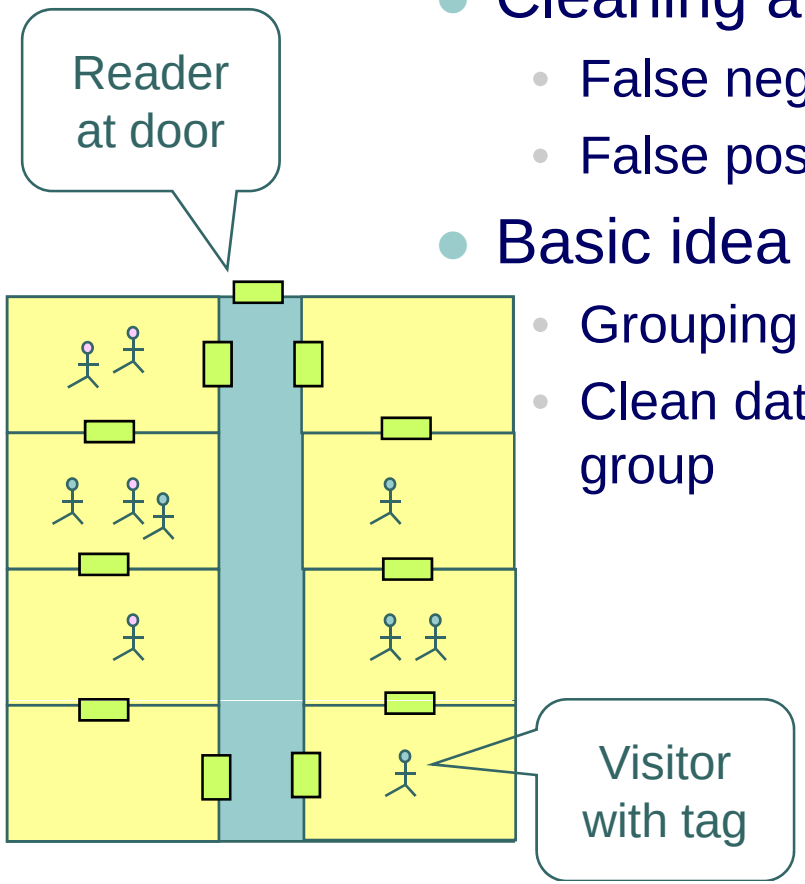


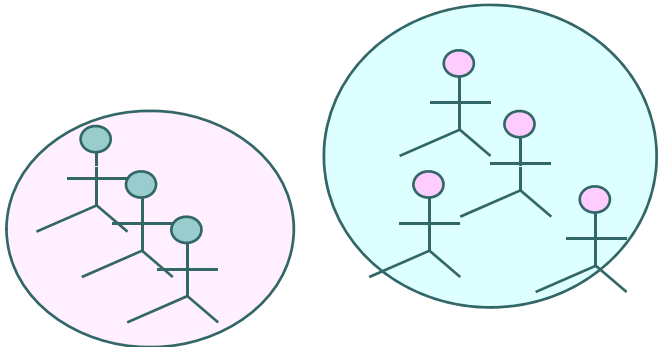
Continuity(*SEQ*(E_1,E_2,E_3), ALL)



Key Techniques

- Grouping based data cleaning(1)
  - Cleaning aim
    - False negative: missing in 30% probability
    - False positive: wrong readings, repetition
  - Basic idea
    - Grouping objects according to their association
    - Clean data according to the information within a group



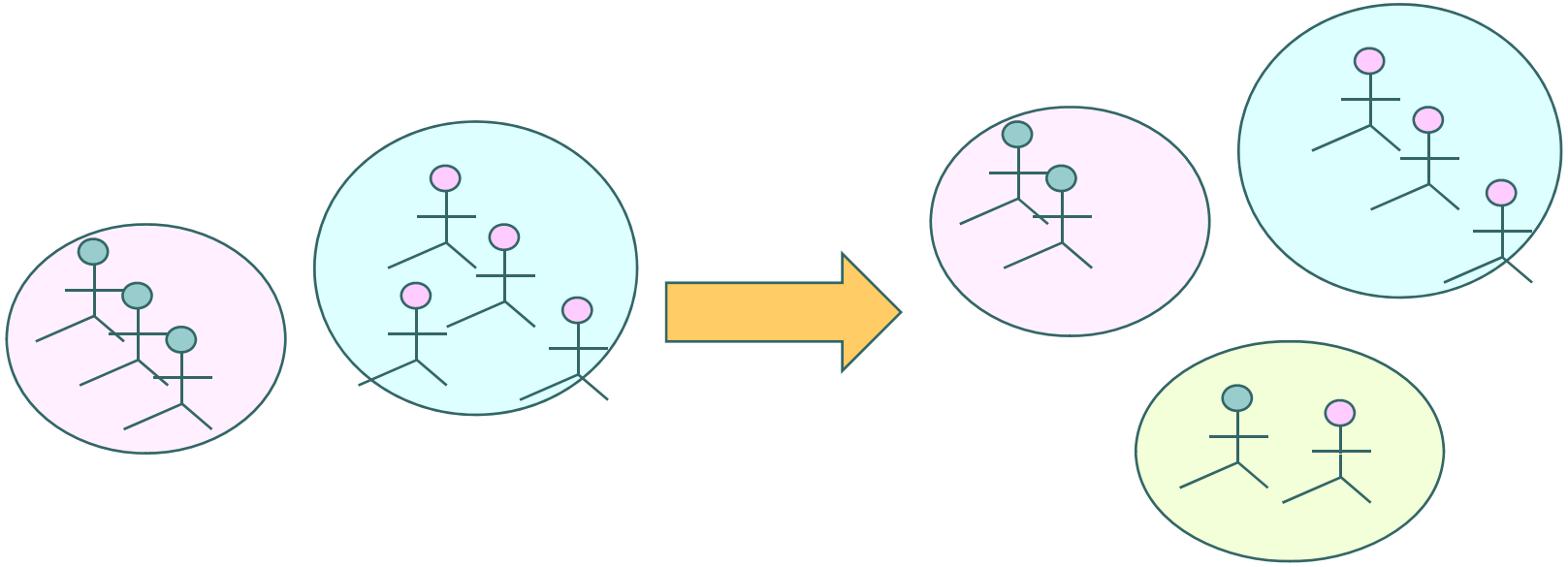


Key Techniques

- Pre-processing
  - Data cleaning
  - ✓Grouping based data cleaning strategy
- Real time query processing
  - Pattern matching
  - Real time scheduling
  - ✓Deadline-sensitive pattern matching

Key Techniques

- Grouping based data cleaning(2)
  - Group identification problems
    - Interferences of different groups
    - Changes of group memberships



Key Techniques

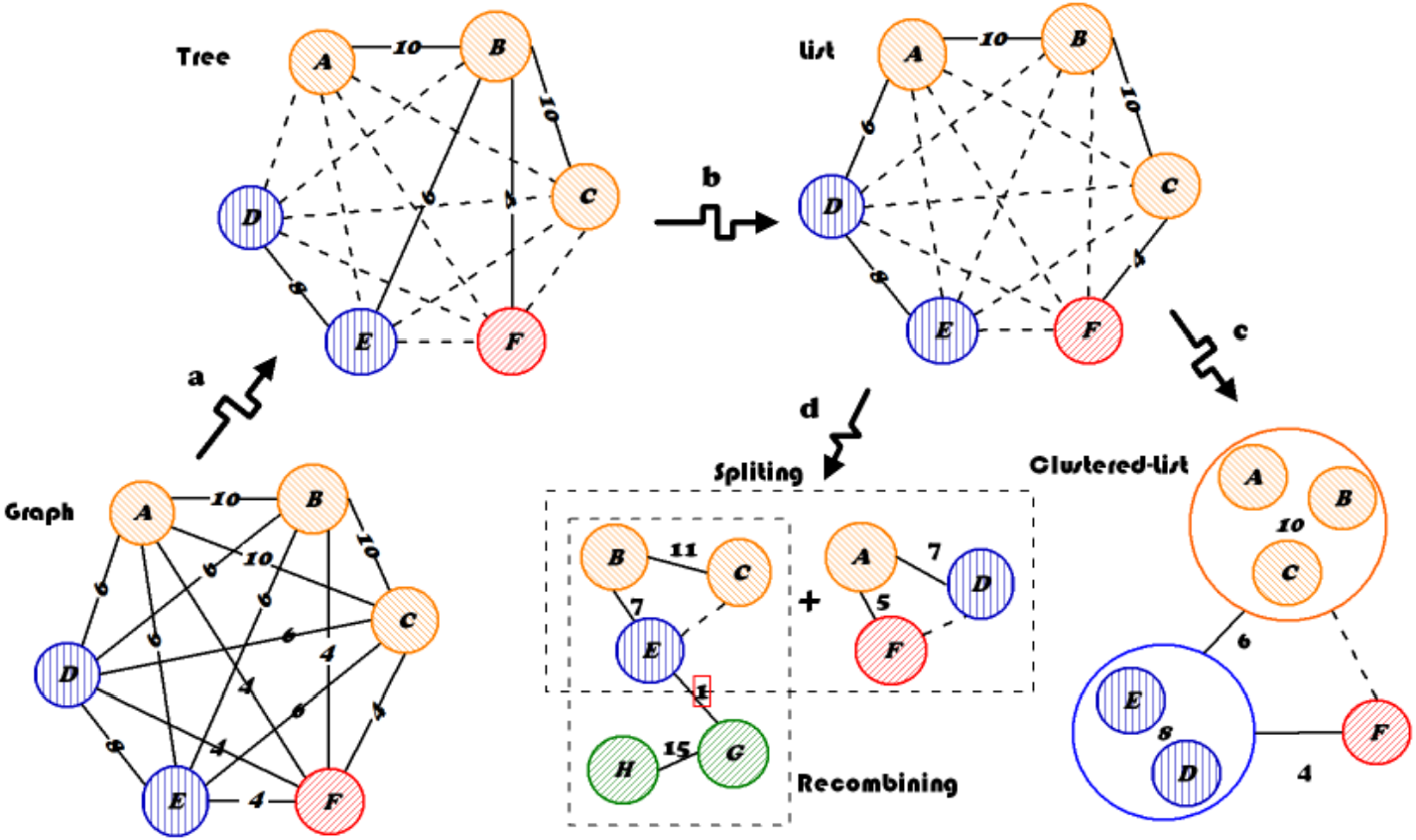
- Grouping based data cleaning(3)
  - Association degree

$$\delta^t(o_i,o_j)=\begin{cases}\delta^{t-1}(o_i,o_j)+\omega(t) & \text{iff } \exists k, o_i,o_j\in\Upsilon_k(t)\\ \delta^{t-1}(o_i,o_j) & \text{iff } o_i,o_j\in\tilde{\Upsilon}(t)\\ 0 & otherwise.\end{cases}$$

  - Graph based association model
  - Compressed tree model
  - List model
  - spliting and re-grouping algorithm

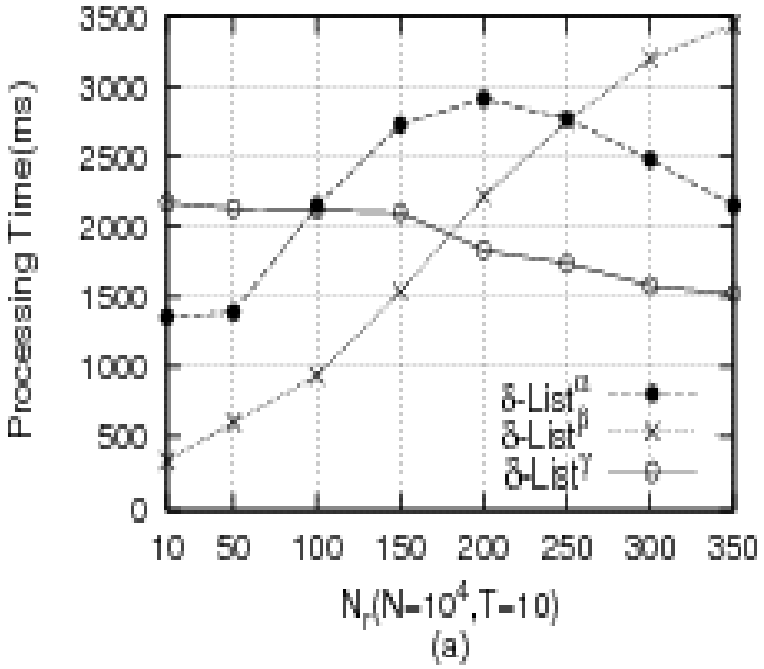
Key Techniques

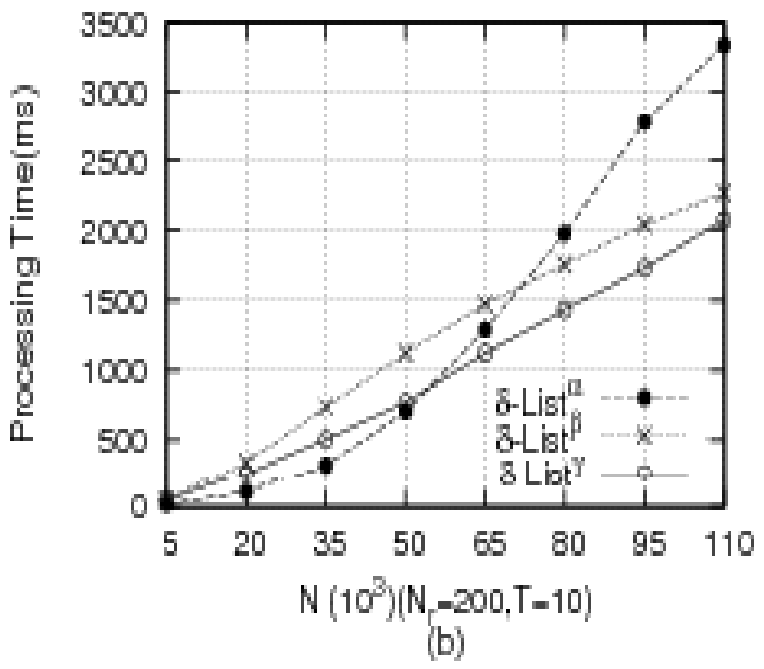
- Grouping based data cleaning(4)



Key Techniques

- Grouping based data cleaning(5)





Key Techniques

- **Deadline-sensitive pattern matching**(1)
  - Aim
    - more correct results that meet deadline
  - Basic idea
    - Statistics based approach
    - RFID data arrival model
    - Event processing model
    - Dead-line meeting model
    - Resource allocation model

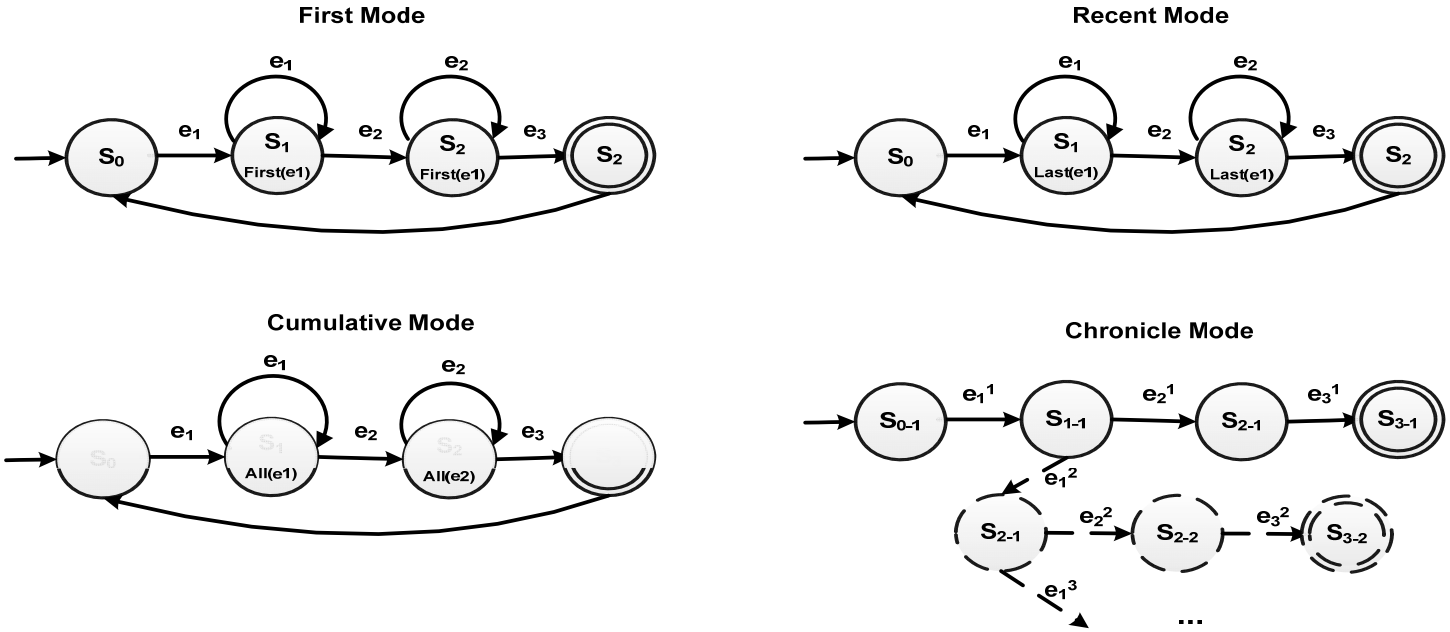
Key Techniques

Deadline-sensitive pattern matching(2)

- Possion distribution based arrival model

$$\pi(n,t) = P\{M(t) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

- NFA based processing model



Key Techniques

- **Deadline-sensitive pattern matching**(3)
  - Processing mode based cost model

Pattern	Query results of SEQ(E1,E2,E3) with $\langle e_1^1, e_1^2, e_2^1, e_2^2, e_3^1, e_3^2 \rangle$	Cost of other sub-events	Extra cost for terminating event
Unrestricted Mode	$(e_1^1, e_2^1, e_3^1), (e_1^1, e_2^2, e_3^1), (e_1^2, e_2^1, e_3^1), (e_1^2, e_2^2, e_3^1)$ $, (e_1^1, e_2^1, e_3^2), (e_1^1, e_2^2, e_3^2), (e_1^2, e_2^1, e_3^2), (e_1^2, e_2^2, e_3^2)$	$C_{insert} + C_{pointer} + \{C_{hash}\}$	$kC(n + \lfloor v_n w \rfloor - 1, n) \prod_{i=1}^{n-1} \frac{E(v_i)}{E(v_n)}$
First Mode	(e_1^1, e_2^1, e_3^1)	$C_{trans} + \{C_{hash}\}$	$\{C_{comp}\}$
Recent Mode	(e_1^2, e_2^2, e_3^1)	$C_{trans} + \{C_{hash}\}$	$\{C_{comp}\}$
Chronicle Mode	$(e_1^1, e_2^1, e_3^1) \quad (e_1^2, e_2^2, e_3^2)$	$C_{insert} + \{C_{hash}\}$	$C_{delete} + \{C_{comp}\}$
Cumulative Mode	$(\sum\{e_1^1, e_1^2\}, \sum\{e_2^1, e_2^2\}, e_3^1)$	$C_{trans} + C_{aggr} + \{C_{hash}\}$	$\{C_{comp}\}$
Continuous Mode	$(e_1^1, e_2^1, e_3^1), (e_1^1, e_2^2, e_3^1), (e_1^2, e_2^1, e_3^1), (e_1^2, e_2^2, e_3^1)$	$C_{insert} + \{C_{hash}\}$	$k \prod_{i=1}^{n-1} \frac{E(v_i)}{E(v_n)}$

Key Techniques

- **Deadline-sensitive pattern matching**(4)
  - Deadline meeting model

$$P\{t_w < t_d - t_p\} = \int_{-\infty}^{t_d - t_p} f(x)dx = 1 - e^{-2(\lambda - \lambda^2 t_p)(t_d - t_p))} \geq \sigma$$

$$\theta_\sigma(\mu) = 1 - e^{-2(\lambda - \lambda^2/\mu)(t_d - 1/\mu))}$$

  - Minimum resource model

$$\mu \geq \frac{2\lambda}{1 + t_d\lambda - \sqrt{(1 + t_d\lambda)^2 - 2(\lambda t_d \ln(1-\sigma) + 2\lambda t_d)}}$$

Key Techniques

- **Deadline-sensitive pattern matching**(5)
  - Disorder model

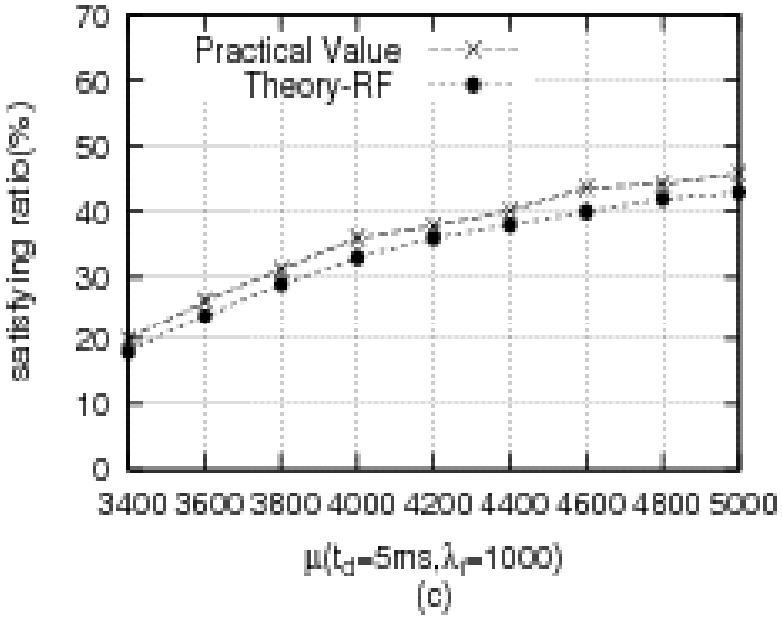
$$\omega_{avg}(E_1, E_2, \cdots, E_n) = 1 - \iiint_{\Omega} \omega(E_1, E_2, \cdots, E_n) \prod_{i=1}^{n-1} f_{i,i+1}(x) dv$$

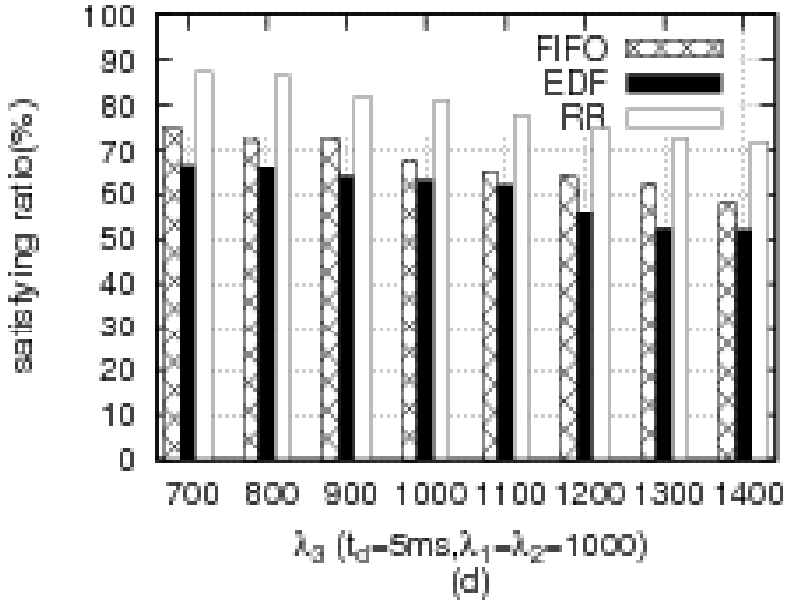
  - Least-disorder based scheduling strategy

$$\begin{cases} \frac{\partial}{\partial \mu_1} \omega_{avg}(\mu_1, \mu_2, \cdots, \mu_n) + t = 0 \\ \frac{\partial}{\partial \mu_2} \omega_{avg}(\mu_1, \mu_2, \cdots, \mu_n) + t = 0 \\ \cdots \\ \frac{\partial}{\partial \mu_n} \omega_{avg}(\mu_1, \mu_2, \cdots, \mu_n) + t = 0 \\ \mu_1 + \mu_2 + \ldots + \mu_n = \mu \\ \mu_n = \mu_{minn} \end{cases}$$

Key Techniques

- **Deadline-sensitive pattern matching**(5)
  - Simulation experimental results



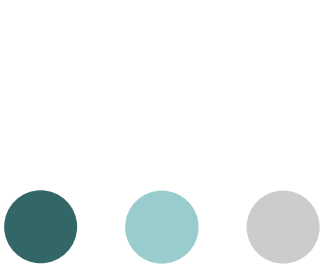


Summary

- Analyzing the characteristics of RFID event stream
- Exploring the challenging problems and key techniques
  - RFID event modeling, data cleaning, event detection, event mining, query optimization, real-time task scheduling, and system architecture
- Designing a high performance real-time event stream processing system

Summary

- Future works
  - Prototype implementation and performance evaluation
    - Scenario Simulator by Netlog
    - Real data set collected by experimental devices
  - Practical applications
    - Product location and tracking
    - People tracking in public area like museums, parks



The End

Thanks!