Complex Event Processing over Unreliable RFID Data Streams*

Yanming Nie, Zhanhuai Li, and Qun Chen

School of Computer, Northwestern Polytechnical University,
Xi'an 710072, China
yanming@mail.nwpu.edu.cn, {lizhh,chenbenben}@nwpu.edu.cn
http://www.nwpu.edu.cn/jsj

Abstract. Existing RFID complex event processing (CEP) techniques always assume that raw RIFD data has been first cleansed to filter out all unreliable readings upfront. But this may cause delayed triggering of matched complex events. Furthermore, since the cleansed event streams need to be temporarily buffered for CEP evaluation, it may generate a large number of intermediate results. To address these issues, we propose an approach to perform CEP directly over unreliable RFID event streams by incorporating cleansing requirements into complex event specifications, and then employ a non-deterministic finite automata (NFA) framework to evaluate the transformed complex events. Experimental results show that our approach is effective and efficient.

Keywords: Unreliable RFID data streams, CEP, NFA.

1 Introduction

CEP can correlate individual RFID readings and transform them into semantic-rich complex events, and therefore plays a key role in monitoring applications. For example, shoplifting in a retail store can be detected by processing a complex event in such scenario: an item has been picked up at a shelf and then taken out of the store without being first checked out in a specific time window.

Raw RFID data is inherently incomplete, noisy, and need to be cleansed. We classify the inaccuracy of RFID data into two categories: **unreliability** and **uncertainty**. Here, unreliability refers to the erroneous readings that can be corrected by deterministic cleansing rules. In a retail store, for instance, if an item has been picked up from a shelf and checked out later on, any other shelf readings for this item in-between should be false positives. Instead uncertainty refers to the inconsistent RFID readings which cannot be determinately eliminated by cleansing rules due to their ambiguities. An instance of uncertainty occurs when the two different shelf readers detect an item simultaneously. So either of the two readings can be false positive, but available RFID data cannot arbitrate. In

* A preliminary version of this work appeared as a 4-page paper at Application of Research of Computers, Vol 26 No 7, 2009 [15].

X. Du et al. (Eds.): APWeb 2011, LNCS 6612, pp. 278–289, 2011.

this paper, we focus only on RFID event stream containing unreliable readings, and leave the issues of CEP over uncertain RFID streams for future work.

The straightforward solution to CEP over unreliable RFID event stream is to filter out unreliable readings upfront, and then execute CEP. But this may cause CEP-enabled systems delaying to trigger corresponding response. Moreover, since cleansed events are needed to be temporarily recorded, it may generate large number of intermediate results. In this paper, we proposed an approach to incorporate cleansing requirements into CEP. Our approach is first to convert unreliability of RFID readings into its corresponding reliability constraints in complex event specifications, and then directly evaluate the transformed complex events over unreliable event streams. The main contributions are as following:

- We present a declarative **C**leansing **L**anguage for **U**nreliable RFID **E**vent **S**treams (**CLUES**), with which cleansing actions can be implicitly set without using a specific action clause.
- Based on CLUES, we propose an automated mechanism to enable evaluating complex event directly over unreliable RFID streams by implanting reliability check constraints into complex event specification.
- By extending the existing NFA implementation frameworks, we propose two approaches to evaluate complex events with reliability check constraints, i.e. a primitive approach and an advanced approach.

2 Related Work

Due to value-based constraints and sliding windows in RFID complex event specification, traditional evaluation frameworks [1][3] is no longer applicable. And their employed fixed data structures, such as tree [3], directed graph [1], finite automata [2], or Petri net [4], cannot adapt to necessary extensions required by RFID complex event specifications. To address these issues and to optimize CEP over huge-volume RFID data, Eugene et al [5] proposed a declarative specification language **SASE** and an NFA-based evaluation framework. Unfortunately, none of them can handle unreliable readings existed in RFID event streams.

RFID data is inherently incomplete, noisy, and need to be cleansed before being forwarded. **SMURF** [7] aims to capture the accurate time window of tag existence by viewing RFID stream as a statistical sample of the tags in physical world, and filters RFID data at the low level of edge device. In practice, however, RFID readings usually require to be analyzed and cleansed in a bigger context of business flow, so cleansing thus need to be executed probably within a large sliding window. A deferred approach was proposed in [8] for detecting RFID data anomalies by defining cleansing rules with **SQL-TS** and performing application-specific cleansing at query time. In [6], Wang et al provided with example rules for data filtering and cleansing. But these off-line and RDBMS-based solutions cannot be applied to RFID CEP. Other related work [11] [12] tried to perform interpretation and imputation over uncertain RFID data by fully exploiting the temporal and spatial relationships among the readings, but neither of them addressed the issues of RFID CEP.