A RULE MODELING ENGINE FOR COMPLEX EVENT PROCESSING (A CASE STUDY ON PASSIVE RFID READERS FOR A VIRTUAL SHOPPING MALL)

BABAK BEHRAVESH

UNIVERSITI TEKNOLOGI MALAYSIA

A RULE MODELING ENGINE FOR COMPLEX EVENT PROCESSING (A CASE STUDY ON PASSIVE RFID FOR A VIRTUAL SHOPPING MALL)

BABAK BEHRAVESH

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

NOVEMBER 2009

Dedicated to my beloved family.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank ALLAH S.W.T. for all the achievements that I have gained today. Next, I wish to extend my grateful appreciation to all those who have contributed directly and indirectly to the preparation of this study. I would like to take this opportunity to thank my supervisor, Prof. Dr. Siti Mariyam Shamsuddin for attention, encouragement and guidance throughout the period of this study. Not forgetting my beloved wife, mother and father for all the supports and understandings that they have given to me. Not forgetting also, my examiners Assoc. Prof. Dr. Naomi Salim and Dr. Siti Zaiton binti Mohd Hashim for many helpful suggestions.

I am grateful to all my colleagues, friends, staff, and lecturers in Faculty of Computer Science and Information System, Universiti Teknologi Malaysia for their help and support at every step during this course of studies.

ABSTRACT

Optimizing Complex Event Processing (CEP) patterns become more interesting topic for researchers due to highly demanding in different areas including RFID based inventory management, Decision support systems, intrusion detection in networks, and many other systems dealing with pattern matching over time series data. Regular expression matching is a well-studied field. In order to achieve better results, one solution is to revise existing algorithms and techniques to make patterns shorter and reducing system overload. In this study, we proposed a complex event processing engine considering historical data in the process of generating more efficient pattern for incoming events. An algorithm is proposed to act on events based on the engine. We develop a pattern matching unit which is used to find match cases over arriving events. Experimental results have shown promising outcomes in reducing processing time with multiple patterns.

ABSTRAK

Mengoptimumkan corak-corak Pemprosesan Peristiwa Kompleks (CEP) kini menjadi lebih menarik disebabkan permintaan yang tinggi dari segenap lapangan termasuklah pengurusan inventori berlandaskan RFID, sistem sokongan keputusan, pengesanan pencerobohan dalam rangkaian, dan pelbagai lagi sistem berkaitan pemadanan corak ke atas data masa bersiri. Pemadanan pernyataan biasa merupakan suatu bidang yang telah matang. Bagi mencapai hasil yang lebih baik, salah satu penyelesaiannya ialah dengan menyemak algoritma-algoritma dan teknik-teknik supaya corak-corak menjadi ringkas justeru mengurangkan bebanan sistem. Untuk kajian ini, kami mencadangkan sebuah enjin pemprosesan peristiwa kompleks yang mengambil kira data sejarah di dalam proses penjanaan corak yang lebih efisien terhadap kemasukan jujukan peristiwa. Sebuah algoritma telah dicadangkan bagi bertindak ke atas peristiwa-peristiwa berdasarkan enjin tersebut. Kami membangunkan sebuah unit pemadanan corak yang digunakan untuk mencari kes-kes berpadanan terhadap semua ketibaan peristiwa. Keputusan eksperimen menunjukkan pengurangan masa pemprosesan bagi corak-corak yang berganda.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

BAM	Business Activity Monitoring
BAP	Battery Assisted Passive
CEP	Complex Event Processing
CEDR	*Complex-Event* Detection and Response
CQL	Continues Query Language
DFA	Deterministic Finite Automaton
ECA	Event Condition Action
EPL	Event Processing Language
ERD	Entity Relation Diagram
ESP	Event Stream Processing
KPI	Key Performance Indicator
NFA	Non-deterministic Finite Automaton
RAM	Random Access Memory
RF	Radio Frequency
RFID	Radio Frequency Identification
SASE	System-of-the-Art Stream Events
SQL	Structured Query Language
SQL_TS	Structured Query Language for Time Series

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview of Complex Event Processing

Complex Event Processing (CEP) is primarily an event processing concept that deals with the task of processing multiple events with the goal of identifying the meaningful events within the event cloud. CEP employs techniques such as detection of complex patterns of many events, event correlation and abstraction, event hierarchies, and relationships between events such as causality, membership, and timing, and event-driven processes. CEP ultimately creates complex events even if some or all of the source events are simple events [6, 7].

CEP needs some patterns for modeling, and modeling is required for implementation. In CEP, we deal with stream of events. For covering different CEP cases, we need a modeler that is able to express different cases (strict contiguity, skip till next, and skip till any match). Hence, Non-deterministic Finite Automaton (NFA) is used to illustrate the flow of arrival events, to act on certain and uncertain

complex events cases. Indeed, NFA can check on a predicate of moving from one state to the next state, and this can be mapped into CEP pattern.

Theoretically, NFA machine or nondeterministic finite automaton (NFA) is a finite state machine with each pair of state and input symbol may include several possible next states. This distinguishes it from the deterministic finite automaton (DFA), where the next possible state is uniquely determined. Although, DFA and NFA have distinct definitions, it is shown in formal theory that they are equivalent. As such, for any given NFA, one may construct an equivalent DFA, and vice-versa: this is the power set construction. Both types of automata recognize only regular languages. Non-deterministic finite state machines are sometimes studied by the name sub shifts of finite type. Non-deterministic finite state machines are generalized by probabilistic automata, which assign a probability to each state transition.

CEP can employ NFA as a model to construct CEP patterns; one such example is RFID devices. These devices are used to capture events from environment. In many complex event cases, we can find some relationship among captured events which can lead to some meaningful patterns. Hence, it is beneficial to define a scenario on RFID devices capture events. Subsequently, CEP goes through NFA to model its patterns for implementation.

1.2 Problem Background

Recently, CEP became a hot topic due to its broad usage in industry. High volume of events has been producing for the enterprise; hence it is plausible to optimize CEP patterns performance. There are several studies working on CEP optimization. Some of them are listed below:

a) State-of-the-Art Stream Event processor, SASE and SASE+ (UC Berkeley/ UMass Amherst)
b) Cayuga (Cornell University)
c) Aurora (Brandeis University, Brown University and MIT)

Figure 1.1 depicts the general idea of non-deterministic finite state automata (NFA), which is simulated, while new rules arriving. Ultimately, NFA model can use mined rules to generate Event Patterns appropriate for individual problem. Because the automata are non-deterministic, NFA can be represented in multiple states simultaneously, and it is determined based on selected strategy. Every time the automaton arrives at an acceptance state, a complex event is detected and constructed.

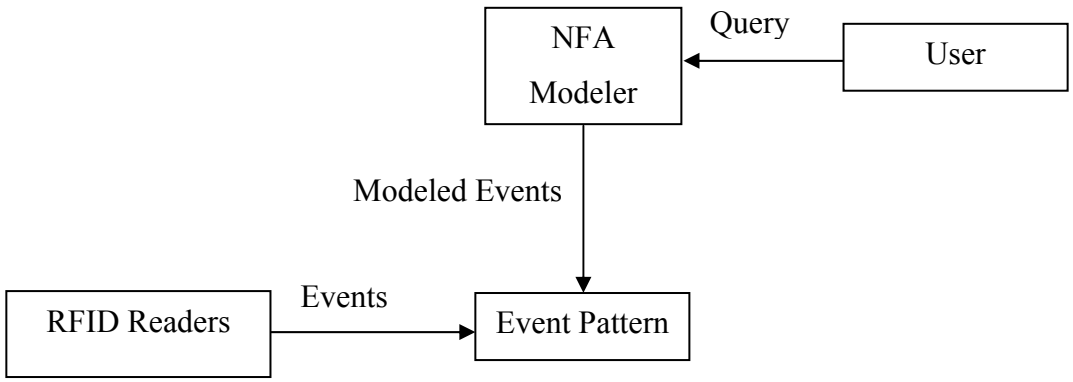

Figure 1.1: A model for making event pattern and pattern matching

Earlier researchers from Cornell and Massachusetts Amherst universities used NFA for modeling CEP patterns [1, 2, 16]. Although their work were on how to use NFA in making CEP patterns [2, 3, 4, 15, 18], but defining an efficient rule based technique to optimize CEP patterns over NFA is still remaining uncovered and challenging [1, 15, 16, 18].

1.3 Problem Statement

Due to limitations of the study on processing events in a certain time implementing engine for CEP environment, this study proposes optimized A-priori algorithm for generating patterns and location buffer to enhance pattern matching in CEP. The research questions of this study are stated as below:

a) *Provided set of queries belong to one case how we can make an optimized pattern?*
b) *Given a patterns extracted from set of queries and stored event-tuples how we can model patterns using NFA?*

1.4 Project Aim

The aim of the project is to propose a new engine for CEP patterns to process more events by optimizing CEP patterns.

1.5 Objectives

This study has few objectives as stated below:

a) To develop a new engine for Complex Event Processing which consists of:
   - A new algorithm to receive and collect event tuples from multiple RFID readers.
   - A new algorithm to generate event patterns, named Enhanced A-priory Algorithm (EAA).
   - A new algorithm for pattern matching.

b) To evaluate the proposed optimized pattern with existing patterns.
   - Investigation the feasibility of CEP patterns over NFA.
   - A new algorithm for generate NFA model given CEP patterns.

c) To compare the proposed system with existing system.
   - Developing two tools for proposed and existing system.
   - Implementing case study data on both systems.

1.6 Scope of the Study

This study is concentrating on using association rules to CEP patterns enhancement. The scopes are as follow:

a) A data set which has been retrieved from RFID devices and store in a database for applying proposed model on it.
b) Some models are investigated for CEP engine.

c) CEP patterns among stored event tuples are investigated for making more efficient patterns and applying for pattern matching.

1.7 Significance of Study

Nowadays one of the important issues in CEP pattern modeling is how can we find a better solution by considering high volume events and queries to process in a minimum time. Defining short and efficient patterns makes it possible for more events than before. Indeed such optimizations on patterns can reduce system overload while preserving the ability to cover different complex events types.