

**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

Mengoptimalkan corak-corak Pemrosesan 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 pemrosesan 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 pemrosesan bagi corak-corak yang berganda.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	TITLE	i
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xii
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATION	xiv
	LIST OF APPENDICES	xv
1	INTRODUCTION	1
	1.1 Introduction	1
	1.2 Problem Background	3
	1.3 Problem Statement	4
	1.4 Project Aim	4
	1.5 Objectives	5
	1.6 Scope of the Study	5
	1.7 Significance of the Study	6

2	LITERITURE REVIEW	7
2.1	Introduction	7
	2.1.1 Event	7
	2.1.2 Event Requirements	9
	2.1.3 Virtual Events	10
	2.1.4 Simple and Raw Events	11
	2.1.5 Event Types	12
	2.1.6 Clock and Timing	13
	2.1.7 Event Processing	14
	2.1.8 Complex Event	14
	2.1.9 Derived Event	15
	2.1.10 Composite Event	16
	2.1.11 Relationships between Events	16
	2.1.12 Event Template and Event Pattern	17
	2.1.13 Event Cloud	18
	2.1.14 Event Infrastructure	19
	2.1.15 Complex Event Processing	21
	2.1.16 Complex Event Processing Software	22
2.2	Non-deterministic Finite Automaton (NFA)	24
	2.2.1 Formal Definition	25
	2.2.2 Properties of NFA	26
	2.2.3 Properties of NFA- ϵ	27
	2.2.4 Implementation	28
	2.2.5 Application of NFA- ϵ	30
2.3	Data Mining	31
	2.3.1 Association Rules	32
	2.3.2 A-priori Algorithm	35
2.4	Modeling Complex Events Using NFA	37
2.5	Rule Mining	38
2.6	Radio Frequency Identifier Devices	39
2.7	Related Studies on NFA, CEP, Association Rules, and RFID	40

2.8	Trends	45
2.9	Summary	46
3	METHODOLOGY	47
3.1	Introduction	47
3.2	Research Design	48
3.2.1	Investigation the feasibility of CEP patterns over NFA- Phase 1	49
3.2.2	Creating a Tool for Collecting Data from RFID Readers- Phase 2	49
3.2.3	Constructing a CEP Engine to Generate Event Patterns and Perform Pattern Matching- Phase 3	49
3.2.4	Developing a new Algorithm to Generate Optimized CEP Pattern- Phase 4	50
3.2.5	Proposing a new Technique for Pattern Matching using Location Buffer- Phase 5	50
3.2.6	A new Algorithm to Model Patterns using NFA- Phase 6	51
3.2.7	Developing a new Tool for the Case Study- Phase 7 and 8	52
3.3	A Case Study on Passive RFID Readers	53
3.3.1	Hardware Devices and Software Tools	53
3.3.2	Applying RFID Data	56
3.4	Collecting RFID Data	57
3.5	System Requirements for Proposed Study	58
3.6	Summary	59
4	EXPERIMENTAL RESULTS AND DISCUSSION	60
4.1	Introduction	60

4.2	Analysis of Proposed Models	61
4.3	Developing a new model to Construct Patterns	64
4.4	A new Tool for Receiving and Collecting Data from RFID Readers through Ports	65
4.5	Enhanced A-priori Algorithm (EAA) to Generate Optimized Patterns	67
4.6	Investigation on Feasibility of CEP Patterns over NFA Model	60
4.7	An Algorithm to Transform CEP Patterns to NFA	70
4.8	Applying Proposed Algorithm on the Case Study	72
4.9	Pattern Matching	77
4.9.1	Pattern Matching on Existing System and Proposed System	77
4.9.2	Location Buffer for Pattern Matching	78
4.10	Developing a Tool to Generate Patterns	82
4.11	Result of Running the Tool over the Case Study	83
4.12	Discussion	85
4.12.1	Comparing Results of Pattern Queries	85
4.12.2	One Pattern Query vs. Optimized Pattern	86
4.12.3	Best Case and Worst Case	86
4.12.4	Performance of the Proposed Optimized Techniques	87
4.13	Summary	89
5	CONCLUSION AND FUTURE WORK	90
5.1	Introduction	90

5.2	Discussion	91
5.3	Summary of Work	92
5.4	Contribution of the Study	93
5.5	Future Work	93
5.5.1	Optimizing Search Space by Purging Storage	94
5.5.2	Supporting Aggregation Queries in CEP	95
5.5.3	A Parallel System for Preventing False Negative	96
5.6	Summary	97
	REFERENCES	98

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Key elements for problem determination	10
2.2	Event infrastructure relationships and terminology	20
2.3	Example data base with 4 items and 5 transactions	33
2.4	Summary of Early Researches Over different concepts Using Non-deterministic Finite Automaton	41
2.5	Summary of Early Researches over different concepts Using Complex Event Processing (CEP)	42
2.6	Summary of Early Researches over different concepts Using A-priori algorithm	43
2.7	Summary of Early Researches over different concepts Using RFID technology	44
4.1	Some event tuples which stored in the Storage	73
4.2	Examples of pattern queries	74
4.3	Non-optimized pattern in a three-query pattern	75
4.4	Transition predicates on each state for each path	75
4.5	Optimized pattern considering location	76
4.6	Location buffer for optimized pattern	79
4.7	An event tuples in a location buffer	79
4.8	Results of running application	84

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	A model for making event pattern and pattern matching	3
2.1	Event types	16
2.2	Complex event type subset	16
2.3	Event infrastructure relationships and terminology	19
2.4	Apama Scenario Manager – A tool developed by Apama	23
2.5	State diagram for above transition function	29
2.6	A flowchart for A-priori algorithm	36
3.1	The Proposed Research design	48
3.2	Comparable steps of the proposed system and the existing system	52
3.3	Mercury 4 and Mercury 5 tag Readers	54
3.4	M4 and M5 tag reader ports	54
3.5	Thing Magic antennas	55
3.6	Dogbone tags	55
3.7	Ethernet connections	56
4.1	1 st proposed model for modeling system activities	61
4.2	2 nd proposed model for modeling system activities	62
4.3	A model for generating mined rules	63
4.4	A engine for CEP	65
4.5	A flowchart for generating optimized patterns	68

LIST OF ABBREVIATION

BAM	Business Activity Monitoring
BAP	Battery Assisted Passive
CEP	Complex Event Processing
CEDR	<i>Complex-Event</i> 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

APPENDIX NO.	TITLE	PAGE
A	Use Case Diagram and Notation for Existing and Proposed System	105
B	Entity Relation Diagram of the Proposed System	126
C	Shopping Mall RFID Data	127

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.

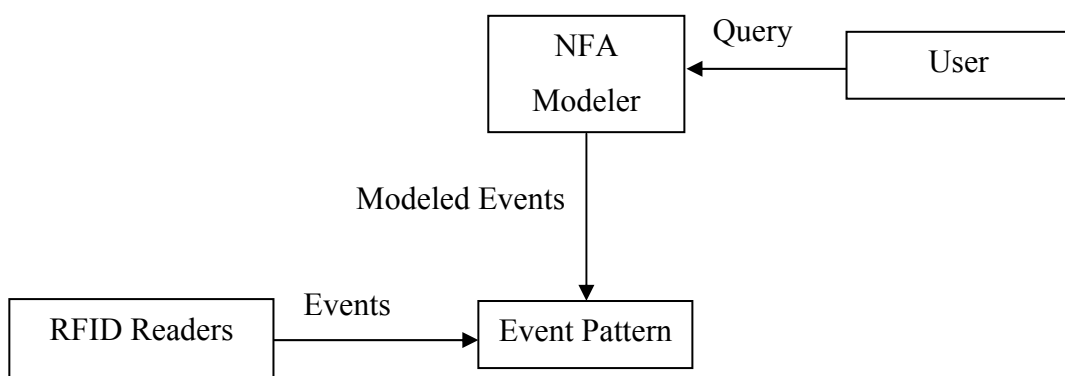


Figure 1.1: A model for making event pattern and pattern matching

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.