

**WEB BASED SYSTEM FOR MICROSOFT ACTIVE
DIRECTORY REPORTING AND EVENT CORRELATION
USING DATA MINING**

M.S.P. Perera

This dissertation was submitted in requirements for the Master of Engineering degree Master of Science in computer science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

December 2008

93368

Abstract

Microsoft Active Directory is very popular in large and medium scale organizations as a system for centralized management of users, desktops, servers, printers and mail boxes etc. This provides a centralized console for managing and viewing the objects very easily. The Active Directory data repository could be used to generate many management reports that would be useful for taking future management decisions and analyzing the health of the organization's security. Many events are generated as a result of user activities and status changes of the objects. These events are reflected on the active directories and event logs. The correlation and outlier analysis of the events is important to filter out thousands of non critical events and be pro-active on important critical events.

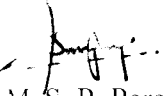
This thesis discusses generating management reports, by querying the Active Directory database and providing real time alerts to system administrators on critical events, with the use of data mining techniques such as event correlation and outlier analysis.

The scope of the event analysis is limited to data generated in the Microsoft Active Directory.

Keywords: Microsoft Active Directory, MS AD, Event Correlation, AD Reports, Outlier Analysis, Event Log Clustering.

Declaration

The work included in this report was done by me, and only by me, and the work has not been submitted for any other academic qualification at any institution.



M. S. P. Perera
22nd December 2008

I certify that the declaration above by the candidate is true to the best of my knowledge.

UOM Verified Signature

Dr. Amal Shehan Perera
22nd December 2008



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table of Content

List of Figures	vii
List of Tables.....	viii
List of Tables.....	viii
Abstract.....	1
Acknowledgement.....	2
1. Introduction	3
2. Literature Review	5
2.1 Web Application Development Tools.....	6
2.2 Querying LDAP Databases	8
2.3 Querying Microsoft Active Directory	9
2.3.1 Active Directory Query Syntax.....	10
2.4 Types of Management Reports.....	11
2.5 Syslog Server for Collecting Microsoft Event Logs.....	11
2.5.1 Introduction to Syslog-ng	13
2.6 Forwarding Active Directory Events to Syslog Server.....	13
2.7 Event Correlation and Outlier Analysis	14
2.7.1 Event Correlation Approaches	15
2.7.2 Introduction to SEC.....	16
2.7.3 Data Mining for Frequent Item Set	17
2.7.4 Data Clustering.....	18
2.7.5 Outlier Analysis	19
3. Methodology	24
3.1 System Development	24
3.1.1 Waterfall Model	24
3.1.2 Spiral Model / Iterative Model.....	25
3.1.3 Top-Down Model.....	25
3.1.4 Bottom-Up Model	26
3.1.5 Hybrid Model	26
3.1.6 Rapid Prototyping	26
3.1.7 Object-Oriented Model.....	27
3.1.8 Selected SDLC Methodology	27
3.2 System Architecture	28
3.2.1 Management Report Generation	28
3.2.2 Event Correlation	29
3.2.3 Outlier Analysis	29
3.3 Software Architecture	31
3.4 Active Directory Reports.....	32
3.4.1 The time parameters in MS Active Directory.....	33
3.4.2 MS Active Directory Reports and Corresponding ldap Queries.....	33
3.5 Event Gathering	34
3.6 Event Correlation.....	35
3.7 Outlier Analysis	35
4. Results.....	37

4.1	AD Reports.....	37
4.2	Event Correlation.....	38
4.3	Event Clustering and Outlier Analysis.....	40
4.4	Event Clusters Observed.....	41
4.5	Outliers Observed.....	41
5.	Analysis and Discussion of Results.....	43
6.	Analysis and Discussion of Results.....	45
7.	Conclusions.....	47
8.	Future Work.....	48
	List of Symbols, Notations, Abbreviations and Acronyms.....	49
	References.....	50
	Annex I - Sample SEC rules.....	53
	Annex II – Querying Active Directory – Java Code.....	55
	Annex III – Windows Epoch.....	57
	Annex IV – Syslog-ng and MySQL Configuration.....	58
	Annex V – Sample SEC Rules.....	60
	Annex VI – Product Web Pages.....	61



University of Moratuwa, Sri Lanka.
 Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

Figure 1 - Spiral Model.....	25
Figure 2 – Management Report Generation.....	28
Figure 3 – Event Correlation.....	29
Figure 4 – Outlier Analysis.....	30
Figure 5- Software Architecture.....	31
Figure 6 – AD Reporting Tool Login Page	37
Figure 7 – Dash Board.....	37
Figure 8 – User Reports.....	38
Figure 9 – Event Correlation Output.....	39
Figure 10 – Event Correlation Rules.....	40
Figure 11 – No of Event Clusters / Outliers when initial threshold is 10 and step value is 1	44
Figure 12 – No of Event Clusters / Outliers when initial threshold is 10 and step value is 2	45



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Tables

Table 1 - Syslog Log Levels	12
Table 2 - Syslog Facility	12
Table 3 - Event Patterns	21
Table 4 - Ldap Queries	34
Table 5 - Event Correlation Output & Rules	39
Table 6 - Event Clusters Observed	41
Table 7 - Outliers Observed	42
Table 8 - No of Event Clusters and Outliers at Threshold Value 10, Step 1 and Iterations from 10 to 5	43
Table 9 - No of Event Clusters and Outliers at Threshold Value 10, Step 2 and Iterations from 5 to 2	44
Table 9 - MySQL Table Structure	59



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Acknowledgement

First of all, I would like to thank my supervisor Dr. Amal Shehan Perera for being the supervisor for my M.Sc. project. I am also thankful to Dr. Shehan for his patience, friendliness, advice and guidance.

I would like to extend my gratitude to Dr. Gihan Dias for encouraging me to complete my M.Sc. project within the allocated period. Without his encouragement and support I wouldn't have completed the M.Sc. Project.

I would like to thank our project coordinator Dr. Sanath Jayasena too, for the proper guidance and effort taken in encouraging all students to complete the final thesis in time.

Finally I thank Sri Lanka Telecom, my employer for providing me the infrastructure and valuable time allocation for my research and study.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk