การตรวจหาค่าผิดปรกติของปริมาณการใช้งานโปรแกรมประยุกต์เว็บ โดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค

นายไพศาล ดีวรรณกิจ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549

ANOMALY DETECTION IN WEB APPLICATION USAGE USING SAX REPRESENTATION FOR TIME SERIES

Mr. Paisal Deewannakij

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2006

หัวข้อวิทยานิพนธ์	การตรวจหาค่าผิดปรกติของปริมาณการใช้งานโปรแกรมประยุกต์เว็บโดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค
โดย	นายไพศาล ดีวรรณกิจ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.โชติรัตน์ รัตนามหัทธนะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.พรศิริ หมื่นไชยศรี)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.โชติรัตน์ รัตนามหัทธนะ)

..... กรรมการ
(อาจารย์ ดร.อติวงศ์ สุชาโต)

..... กรรมการ
(อาจารย์ ธงชัย โรจน์กังสดาล)

ไพศาล ดีวรรณกิจ : การตรวจหาค่าผิดปรกติของปริมาณการใช้งานโปรแกรมประยุกต์เว็บโดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค. (ANOMALY DETECTION IN WEB APPLICATION USAGE USING SAX REPRESENTATION FOR TIME SERIES) อาจารย์ที่ปรึกษา : อ.ดร.โชติรัตน์ รัตนามหัทธนะ, 117 หน้า.

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์ในการพัฒนาวิธีการตรวจหาค่าผิดปรกติที่ดัดแปลงจากข้อมูลปริมาณการใช้งานโปรแกรมประยุกต์เว็บ เพื่อเพิ่มประสิทธิภาพในการตรวจหาค่าผิดปรกติ จึงนำเสนอแนวทางในการตรวจหาค่าผิดปรกติโดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค ทั้งยังออกแบบและพัฒนาเครื่องมือในการสร้างข้อมูลที่มีรูปแบบต่างๆ ไว้ใช้สำหรับทดสอบประสิทธิภาพของวิธีการตรวจหาค่าผิดปรกติรูปแบบอื่นๆ อีกด้วย

งานวิจัยนี้ได้มีการวิเคราะห์รูปแบบของปริมาณการใช้งานเว็บไซต์ต่างๆ และกำหนดค่าพารามิเตอร์ที่สำคัญ เพื่อให้การตรวจหาค่าผิดปรกติเป็นไปอย่างมีประสิทธิภาพ นอกจากนี้ยังได้เปรียบเทียบผลจากวิธีในการตรวจหาค่าผิดปรกติที่ดัดแปลง และวิธีในการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ ซึ่งได้ผลสรุปจากการทดลองว่า สามารถพิจารณาค่าผิดปรกติเป็นไปอย่างมีประสิทธิภาพ โดยเครื่องมือในการสร้างข้อมูลที่มีรูปต่างๆ นั้น ได้มีการพิจารณารูปแบบของข้อมูลที่ทำการสร้าง และพบว่าสามารถสร้างข้อมูลที่มีรูปแบบต่างๆ ได้ตามต้องการเช่นกัน

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต ไพศาล ดีวรรณกิจ
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา
ปีการศึกษา 2549

4871434221 : MAJOR COMPUTER SCIENCE

KEY WORD: SAX / WEB APPLICATION SECURITY / TIME SERIES / ANOMALY DETECTION

PAISAL DEEWANNAKIJ : ANOMALY DETECTION IN WEB APPLICATION USAGE USING SAX REPRESENTATION FOR TIME SERIES. THESIS ADVISOR : CHOTIRAT RATANAMAHATANA, Ph.D., 117 pp.

The objectives of this research are to present and develop an anomaly detection algorithm that improves the detection performance, using SAX Time Series representation. Furthermore, this research designs and develops a web application usage data generator. Any arbitrary usage pattern could be conveniently generated and used in various anomaly detection methods.

Experimenting with massive data from real web application usages, this research also analyzes and determines optimal parameters. The results from both brute-force and the proposed anomaly detection methods are compared. This research work has demonstrated its effectiveness in decreasing loop counts and discovering anomalies. And web application usage generator can also be used to create various usage patterns.

Department.....Computer Engineering.....Student's signature.....

Field of study.....Computer Science.....Advisor's signature.....

Academic year ....2006.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี เนื่องมาจากความช่วยเหลืออย่างดียิ่งของท่าน อ.ดร.โชติรัตน์ รัตนามหัทธนะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้สละเวลาให้คำปรึกษาแนะนำแนวทางเกี่ยวกับงานวิจัยอย่างดีตลอดมาจนเสร็จสมบูรณ์ และผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่านที่ได้ให้คำแนะนำ ข้อคิดเห็น ข้อเสนอแนะ และแนวทางในการพัฒนางานวิจัยนี้

ขอขอบคุณ สำนักบริการสารเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย และบริษัท Ethnic Earth จำกัด ที่เอื้อเฟื้อข้อมูลที่นำมาใช้ในการวิจัย

ขอขอบคุณ พี่ธุรการภาคฯ ทุกๆ คนที่ช่วยอำนวยความสะดวกในการทำงาน และคอยแนะนำสิ่งดีๆ เสมอมา

สุดท้ายนี้ ขอกราบขอบพระคุณคุณพ่อคุณแม่ที่ให้โอกาสเราได้เกิด ได้เติบโต ได้เลี้ยงดูเป็นอย่างดี และคอยสนับสนุนในด้านการศึกษาเป็นอย่างดีเสมอมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฎ
สารบัญภาพ ..... ฑ
บทที่
1 บทนำ 1
1.1 ความเป็นมาและความสำคัญของปัญหา 1
1.2 วัตถุประสงค์การวิจัย 3
1.3 ขอบเขตการวิจัย 3
1.4 ประโยชน์ที่คาดว่าจะได้รับ 3
1.5 ขั้นตอนและวิธีดำเนินงานวิจัย 4
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5
2.1 อนุกรมเวลา 5
2.2 การทำข้อมูลให้เป็นบรรทัดฐาน 5
2.2.1 Z-Score 5
2.2.2 Piecewise Aggregation Approximation 5
2.3 SAX 6
2.4 การวัดระยะห่าง 8
2.4.1 การหาระยะห่างแบบยุคลิด 8
2.4.2 การหาระยะห่างแบบแช็ค 8
2.5 การตรวจหาค่าผิดปรกติ 10
2.6 การตรวจหาค่าผิดปรกติในอนุกรมเวลา 12
2.7 การตรวจหาค่าผิดปรกติในอนุกรมเวลาจากทุกความเป็นไปได้ 13
2.8 งานวิจัยที่เกี่ยวข้อง 13
2.8.1 A Signal Analysis of Network Traffic Anomalies 13
2.8.2 Visually Mining and Monitoring Massive Time Series 14

บทที่ หน้า

2.8.3 HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications 14
2.8.4 Recent Advances in Network Intrusion Detection System Tuning 19
3 การออกแบบการตรวจหาค่าผิดปรกติและสร้างข้อมูลบันทึกการใช้งานบริการเว็บ 21
3.1 งานที่เกี่ยวกับการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บ 21
3.1.1 กระบวนการในการตรวจหาค่าผิดปรกติ 22
3.1.1.1 การแปลงข้อมูลให้อยู่ในรูปของอนุกรมเวลา 22
3.1.1.2 การลดมิติของข้อมูลด้วยวิธี PAA 24
3.1.1.3 การกำหนดขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูล 24
3.1.1.4 การกำหนดจำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติ 24
3.1.1.5 ขั้นตอนวิธีที่ใช้ในการตรวจหาค่าผิดปรกติ 25
3.1.1.6 โครงสร้างข้อมูลที่นำมาใช้กับขั้นตอนวิธี 26
3.1.1.7 อันดับของค่าผิดปรกติที่เกิดขึ้น 28
3.1.1.8 เกณฑ์ในการเลือกค่าผิดปรกติ 28
3.1.1.9 การจำลองข้อมูลแบบต่อเนื่อง 28
3.1.1.10 แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ 29
3.1.2 กระบวนการในการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบไม่ต่อเนื่อง 31
3.1.3 กระบวนการในการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบต่อเนื่อง 32
3.2 โปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ 33
3.2.1 การวิเคราะห์การสร้างข้อมูล 33
3.2.1.1 การสร้างข้อมูลที่มีรูปแบบต่างๆ 33
3.2.1.2 การสุ่มค่าตัวเลข 33
3.2.1.3 การสร้างแผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานบริการเว็บ 33
3.2.1.4 แฟ้มโครงแบบที่ใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บ 34
3.2.2 กระบวนการในการสร้างข้อมูล 37
3.3 ข้อมูลที่ใช้ในการวิจัย 38
3.3.1 ข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติ 38

บทที่ หน้า

3.3.2 ข้อมูลบันทึกการใช้งานโปรแกรมเว็บจากเว็บไซต์จากจุฬาลงกรณ์มหาวิทยาลัย.... 39
3.3.3 ข้อมูลที่สร้างขึ้นมาจากโปรแกรม 41
3.3.4 ข้อมูลบันทึกการใช้งานโปรแกรมเว็บจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ..... 43
3.3.5 ข้อมูลบันทึกการใช้งานโปรแกรมเว็บจากเว็บไซต์วิทยุออนไลน์
ของหนังสือพิมพ์ผู้จัดการ..... 45
4 ผลการตรวจหาค่าผิดปรกติและสร้างข้อมูลบันทึกการใช้งานบริการเว็บ 48
4.1 ข้อมูลที่ใช้ในการวิจัย..... 48
4.1.1 แฟ้มโครงแบบที่ใช้ในการทดลองสร้างข้อมูลบันทึกการใช้งาน
โปรแกรมประยุกต์เว็บ 48
4.1.1.1 แฟ้มโครงแบบ A..... 48
4.1.1.2 แฟ้มโครงแบบ B..... 49
4.1.1.3 แฟ้มโครงแบบ C..... 50
4.1.2 ข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติ 52
4.1.2.1 ข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่มีการใช้งานจริง 52
4.1.2.2 ข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่ถูกสร้างขึ้น 52
4.2 การดำเนินงานวิจัย 53
4.2.1 การสร้างข้อมูลบันทึกการใช้งานบริการเว็บ..... 53
4.2.2 การสร้างแผนภูมิกราฟในการใช้งานบริการเว็บ และการแปลงข้อมูล
ให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์เท่านั้น..... 53
4.2.3 การตรวจหาค่าผิดปรกติ..... 54
4.3 ผลการวิจัย..... 55
4.3.1 ผลการสร้างข้อมูลบันทึกการเข้าใช้งานของเว็บไซต์ 55
4.3.2 ผลการตรวจหาค่าผิดปรกติ 57
4.3.2.1 ผลการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่อง
ของข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่มีการใช้งานจริง 57
4.3.2.2 ผลการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่อง
ของข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่ถูกสร้างขึ้น 71

บทที่ หน้า

4.3.2.3 ผลการตรวจหาค่าผิดปรกติจากข้อมูลแบบต่อเนื่อง 80
4.4 เวลาที่ใช้ในการตรวจหาค่าผิดปรกติ 87
4.4.1 คุณลักษณะของคอมพิวเตอร์ 87
4.4.2 รุ่นของโปรแกรม 87
4.4.3 เวลาที่ใช้ในการตรวจหาค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง 88
5 การทดสอบประสิทธิภาพของการตรวจหาค่าผิดปรกติ 90
5.1 การเปรียบเทียบผลการตรวจหาค่าผิดปรกติที่ได้กับวิธีทุกความเป็นไปได้ 90
5.2 การวิเคราะห์ผลการเปรียบเทียบกับวิธีทุกความเป็นไปได้ 91
5.3 การวิเคราะห์ผลการสร้างข้อมูลบันทึกการเข้าใช้งานของเว็บไซต์ 92
6 สรุปผลการวิจัยและข้อเสนอแนะ 93
6.1 สรุปผลการวิจัย 93
6.1.1 ผลการทดสอบโปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึก
การเข้าใช้งานบริการเว็บ 93
6.1.2 ผลการทดสอบโปรแกรมที่ใช้ในการตรวจหาค่าผิดปรกติของเครื่องบริการ
โปรแกรมประยุกต์เว็บ 94
6.2 ปัญหาที่พบจากการวิจัย 94
6.3 ข้อเสนอแนะ 94
รายการอ้างอิง 95
ภาคผนวก 98
ภาคผนวก ก แฟ้มโครงแบบเอกซ์เอ็มแอลของการสร้างข้อมูลการบันทึกการใช้งานของเว็บ . 99
ภาคผนวก ข แฟ้มโครงแบบของการสร้างแผนภูมิกราฟและแปลงข้อมูล
ให้มีเพียงข้อมูลที่นำมาวิเคราะห์เท่านั้น 112
ภาคผนวก ค แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ 116
ประวัติผู้เขียนวิทยานิพนธ์ 117

สารบัญตาราง

ตาราง หน้า

2.1 ตัวอย่างจุดแบ่งพื้นที่ใต้กราฟของการแจกแจงเกาส์เซียน 7

2.2 ค่าระยะห่างสำหรับข้อมูลที่แทนค่าสัญลักษณ์แล้วโดยใช้สัญลักษณ์ทั้งหมด 4 ตัว 9

2.3 ตัวอย่างการหาค่าระยะห่างที่ค่า *r* และ *c* มีความแตกต่างกันมากกว่า 1 10

2.4 ข้อดีและข้อเสียของการตรวจหาค่าผิดปรกติโดยการใช้ลักษณะบ่งชี้ 11

2.5 ข้อดีและข้อเสียของการตรวจหาค่าผิดปรกติจากความผิดปรกติ 12

2.6 รหัสเทียมของการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ 15

2.7 รหัสเทียมของการตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก 16

3.1 ตัวอย่างของข้อมูลการร้องขอบริการจากเว็บไซต์ที่ผ่านการแปลงให้มีเพียงข้อมูลที่ใช้ในการวิเคราะห์ 22

3.2 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ที่ผ่านการแปลงให้มีเพียงข้อมูลที่ใช้ในการวิเคราะห์ 22

3.3 ตัวอย่างแฟ้มโครงแบบของการแปลงข้อมูลให้มีเพียงข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติ . 23

3.4 จำนวนตัวอักษรที่ถูกเลือกในการตรวจหาค่าผิดปรกติที่ขึ้นกับช่วงเวลาที่ทำ PAA 25

3.5 รหัสเทียมของการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ที่มีการจัดเก็บข้อมูลระยะห่างของแต่ละลำดับย่อยในแถวลำดับ 26

3.6 ตัวอย่างแฟ้มโครงแบบของการตรวจหาค่าผิดปรกติ 29

3.7 แฟ้มโครงแบบเอกซ์เอ็มแอลที่ใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บ 34

3.8 ส่วนเพิ่มเติมในแฟ้มโครงแบบเอกซ์เอ็มแอลที่ใช้ในการสร้างลักษณะข้อมูลบันทึกการใช้งานบริการเว็บที่ผิดปรกติ 36

3.9 ตัวอย่างของข้อมูลการร้องขอบริการจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย 39

3.10 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย 39

3.11 ตัวอย่างของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นจากโปรแกรม 41

3.12 ความหมายของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นจากโปรแกรม 41

3.13 ตัวอย่างของข้อมูลการร้องขอบริการจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ 43

3.14 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ 43

3.15 ตัวอย่างของข้อมูลการร้องขอบริการจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ 45

3.16 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ 46

ตาราง หน้า

4.1 ลักษณะตำแหน่งและชนิดของข้อมูลที่แตกต่างกัน
ของข้อมูลบันทึกการใช้งานบริการเว็บไซต์ 54
4.2 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล A 58
4.3 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล A 59
4.4 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล B 61
4.5 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล B 62
4.6 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล C
ที่ผ่านการทำ PAA 24 ชั่วโมง 63
4.7 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล C
ที่ผ่านการทำ PAA 8 ชั่วโมง 65
4.8 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล C
ที่ผ่านการทำ PAA 8 ชั่วโมง 65
4.9 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล D 68
4.10 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล D 70
4.11 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล E 72
4.12 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล E 73
4.13 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล F 74
4.14 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล F 76
4.15 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล G 78
4.16 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล G 79
4.17 ผลการตรวจหาค่าผิดปรกติของข้อมูล A ที่จำลองเป็นข้อมูลแบบต่อเนื่อง
จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 81
4.18 ผลการตรวจหาค่าผิดปรกติของข้อมูล B ที่จำลองเป็นข้อมูลแบบต่อเนื่อง
จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 83
4.19 ผลการตรวจหาค่าผิดปรกติของข้อมูล D ที่จำลองเป็นข้อมูลแบบต่อเนื่อง
จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 84
4.20 ผลการตรวจหาค่าผิดปรกติของข้อมูล F ที่จำลองเป็นข้อมูลแบบต่อเนื่อง
จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 85

ตาราง หน้า

4.21 ผลการตรวจหาค่าผิดปรกติของข้อมูล G ที่จำลองเป็นข้อมูลแบบต่อเนื่อง จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 86
4.22 คุณลักษณะของคอมพิวเตอร์ 87
ก.1 แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับสร้างข้อมูลแบบที่ 1 99
ก.2 แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับสร้างข้อมูลแบบที่ 2 101
ก.3 แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับสร้างข้อมูลแบบที่ 3 103
ข.1 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัย 112
ข.2 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์หนังสือพิมพ์ผู้จัดการ 113
ข.3 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ 114
ข.4 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์ที่ถูกสร้างขึ้น 114
ค.1 แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ 116

สารบัญภาพ

ภาพประกอบ หน้า

2.1 แสดงเส้นอนุกรมเวลา c ที่ถูกแปลงให้เป็น PAA $\bar{c}$ เทียบกับเส้นอนุกรมเวลาแบบเดิม 6

2.2 เส้นอนุกรมเวลาที่ถูกแปลงให้เป็น PAA แล้วทำการแทนค่าข้อมูลให้อยู่ในรูปสัญลักษณ์..... 8

2.3 ระยะห่างของข้อมูลที่เป็นสัญลักษณ์จำนวน 2 สายอักขระ 9

2.4 ค่าผิดปรกติในอนุกรมเวลา 12

2.5 โครงสร้างข้อมูลสายอักขระ..... 17

2.6 โครงสร้างข้อมูลแถวลำดับ (ซ้าย) และโครงสร้างต้นไม้แบบแต่งเติม (ขวา) 18

3.1 โครงสร้างต้นไม้แบบแต่งเติมรวมกับโครงสร้างข้อมูลแถวลำดับ 28

3.2 กระบวนการในการตรวจหาค่าผิดปรกติจากปริมาณการใช้งานเว็บ
โดยใช้ข้อมูลแบบไม่ต่อเนื่อง 31

3.3 กระบวนการในการตรวจหาค่าผิดปรกติจากปริมาณการใช้งานเว็บ
โดยใช้ข้อมูลแบบต่อเนื่อง 32

3.4 ส่วนประกอบต่างๆในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ 37

3.5 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย 40

3.6 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ที่ถูกสร้างขึ้นจากโปรแกรม..... 43

3.7 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ..... 45

3.8 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ..... 47

4.1 แผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานที่ถูกสร้างขึ้นจากโปรแกรมตาม
แฟ้มโครงแบบ A 55

4.2 แผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานที่ถูกสร้างขึ้นจากโปรแกรมตาม
แฟ้มโครงแบบ B 56

4.3 แผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานที่ถูกสร้างขึ้นจากโปรแกรมตาม
แฟ้มโครงแบบ C 56

4.4 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล A จากการตรวจหาค่าผิดปรกติทั้งสองวิธี..... 57

4.5 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล B จากการตรวจหาค่าผิดปรกติทั้งสองวิธี..... 60

4.6 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล C จากการตรวจหาค่าผิดปรกติ
จากทุกความเป็นไปได้ที่ผ่านการทำ PAA 24 ชั่วโมง 63

4.7 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล C จากการตรวจหาค่าผิดปรกติทั้งสองวิธี
ที่ผ่านการทำ PAA 8 ชั่วโมง 64

ภาพประกอบ หน้า

4.8 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล D จากการตรวจหาค่าผิดปรกติ วิธีทุกความเป็นไปได้ 67
4.9 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล D จากการตรวจหาค่าผิดปรกติ วิธีที่ผู้เขียนดัดแปลง 69
4.10 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล E จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 71
4.11 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล F จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 74
4.12 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล F จากการตรวจหาค่าผิดปรกติทั้งสองวิธี 77

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัญหาการตรวจหาค่าผิดปรกติของข้อมูลที่รับส่งกันในเครือข่ายคอมพิวเตอร์ ทั้งที่เกิดขึ้นจากการใช้งานในปริมาณที่เปลี่ยนแปลงไปจากเดิม หรือเกิดจากการบุกรุกเข้ามายังเครือข่ายคอมพิวเตอร์นั้น เป็นปัญหาที่ทุกคนต้องให้ความใส่ใจและมีความจำเป็นต่อองค์กร ซึ่งปริมาณการใช้งานที่ผิดปรกติ หรือที่เกิดจากการบุกรุกที่เกิดขึ้น ไม่เพียงเกิดขึ้นในเครือข่ายเท่านั้น แต่ยังสามารถโจมตีไปยังเครื่องคอมพิวเตอร์ต่างๆ ที่อยู่ในเครือข่ายอีกด้วย อาทิเช่น เครื่องบริการเว็บ (Web Server) เครื่องบริการเว็บเซอร์วิส (Web Service Server) เครื่องบริการชื่อโดเมน (Domain Name Server) เครื่องบริการแฟ้ม (File Server) เครื่องบริการอีเมล (Email Server) เครื่องบริการโปรแกรมประยุกต์เว็บ (Web Application Server) และเครื่องคอมพิวเตอร์ของผู้ใช้ เป็นต้น

การตรวจหาค่าผิดปรกติของข้อมูลมีหลายวิธี หนึ่งในวิธีที่นิยมใช้กันอย่างแพร่หลายนั้นคือการจับยึดข้อมูลบนเครือข่าย [1][2][3][4][5][6] ข้อเสียของวิธีนี้ คือ หากข้อมูลที่รับส่งระหว่างกันเป็นไปตามกฎที่กำหนดไว้ และมีการเข้ารหัสข้อมูลที่รับส่งระหว่างกันก็จะสามารถตรวจหาค่าผิดปรกติได้จากปริมาณการใช้งานเท่านั้น ซึ่งเครื่องบริการเว็บ และเครื่องบริการเว็บเซอร์วิส หรือที่มักเรียกรวมกันว่า เครื่องบริการโปรแกรมประยุกต์เว็บ ที่มีการเข้ารหัสข้อมูล ก็จะอยู่ในข่ายของปัญหานี้ด้วย และเนื่องจากเครื่องที่ให้บริการดังกล่าวเป็นที่นิยมมากขึ้นในปัจจุบัน ทำให้มีการโจมตีแบบต่างๆ เกิดขึ้นมากมาย [7] เช่น การทำให้เครื่องบริการเว็บปฏิเสธการให้บริการ หนอนและไวรัสบนเว็บ [8] การเข้าถึงข้อมูลสำคัญโดยใส่ข้อมูลเฉพาะแบบ และ Cross Site scripting เป็นต้น ทำให้มีงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์การโจมตีขึ้น [9] งานวิจัยที่ทำการจำแนกประเภทของการโจมตีของเครื่องบริการเว็บตามสาเหตุที่แท้จริง และตำแหน่งของจุดเชื่อมต่อของเว็บ [10] ก็เป็นหนึ่งในงานวิจัยที่ได้มีการทำวิจัยไว้ รวมไปถึงการจัดตั้งกลุ่มต่างๆที่จัดตั้งขึ้นเพื่อความปลอดภัยของเว็บ ไม่ว่าจะเป็นโครงการความปลอดภัยของโปรแกรมประยุกต์เว็บ (OWASP) [11] องค์กรสำหรับความปลอดภัยของโปรแกรมประยุกต์เว็บ (Web Application Security Consortium) [7][12][13] ฯลฯ เพื่อเป็นสื่อกลางในการพัฒนา และเปิดโอกาสให้ร่วมกันพัฒนาความปลอดภัยของโปรแกรมประยุกต์เว็บ ทั้งยังมีคู่มือในการป้องกันการโจมตี การพัฒนามอดูลที่รองรับความปลอดภัยของเว็บและวิธีการใช้งาน [14][15][16]

อย่างไรก็ตาม การทดสอบวิธีที่ใช้ในการตรวจหาค่าผิดปรกติว่ามีประสิทธิภาพนั้น ส่วนใหญ่เป็นการเก็บข้อมูลจากเครือข่ายที่ใช้งานจริงหรือเครื่องที่ให้บริการจริง และมีการโจมตีเกิดขึ้น

จริง [1][2][4][5][6] ซึ่งทำให้บางครั้งผู้ที่ต้องการเรียนรู้และพัฒนาต่อเพียงบางส่วน จำเป็นต้องศึกษาการทำงานทั้งหมด หรือหากต้องการตรวจหาค่าผิดปรกติจากข้อมูลทดสอบก็สามารถทำได้โดยใช้ชุดข้อมูลของ DARPA [17] ที่มีข้อมูลการโจมตีของเครือข่ายที่ไม่ทันสมัย ซึ่งหากผู้ที่ทำงานวิจัยต้องการข้อมูลที่มีความสมบูรณ์มากที่สุด ผู้ที่ทำงานวิจัยจำเป็นต้องจัดเตรียมอุปกรณ์เพื่อทำการสร้างข้อมูลจำลอง ที่การติดตั้งอุปกรณ์จำเป็นต้องอาศัยความรู้ ประสบการณ์ และความชำนาญ รวมไปถึงรูปแบบในการโจมตีจำนวนมากทั้งแบบเก่าและแบบใหม่ เพื่อใช้ในการทดสอบ [1][2][4][5][6] ซึ่งข้อมูลการโจมตีที่จัดเตรียมต้องเป็นความลับ เนื่องจากอาจทำให้เกิดความเสียหายได้หากข้อมูลดังกล่าวเปิดเผยสู่ผู้ที่ไม่ได้ทำงานวิจัย โดยข้อมูลที่นำมาตรวจหาค่าผิดปรกติสำหรับเครื่องบริการโปรแกรมประยุกต์เว็บนั้นก็ประสบปัญหาเช่นเดียวกัน ทั้งนี้ชุดข้อมูลของ DARPA [17] ยังไม่มีข้อมูลของการโจมตีและปริมาณการใช้งานของเครื่องบริการเว็บโดยเฉพาะ

นอกจากนี้ ได้มีงานวิจัยเป็นจำนวนมากที่นำข้อมูลปริมาณการใช้งานของเครือข่ายมาวิเคราะห์โดยวิธีต่างๆ เช่น การใช้เทคนิค Deviation Score [1] การวิเคราะห์ทางสถิติ [2] และการวิเคราะห์แบบอนุกรมเวลา [3] เป็นต้น ซึ่งพบว่าสามารถตรวจหาค่าผิดปรกติที่เกิดขึ้นในเครือข่ายได้เป็นอย่างดี ซึ่งปริมาณการใช้งานที่อยู่ในเครื่องบริการเว็บนั้นก็เป็นข้อมูลที่อยู่ในรูปแบบอนุกรมเวลาแบบหนึ่ง โดยมีงานวิจัยที่เกี่ยวกับการตรวจหาค่าผิดปรกติของอนุกรมเวลาโดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค (SAX - Symbolic Aggregate approXimation) [18] เพื่อใช้ในการตรวจหาค่าผิดปรกติ ซึ่งข้อดีของวิธีแซ็ค [19] นั้น คือความสามารถในการลดมิติและความซับซ้อนของข้อมูล โดยยังสามารถตรวจหาค่าผิดปรกติได้อย่างรวดเร็วและมีประสิทธิภาพ

จากปัญหาและเทคนิคดังกล่าว ทำให้ผู้เขียนเกิดแนวคิดในการออกแบบและพัฒนาเครื่องมือในการตรวจหาค่าผิดปรกติ ทั้งยังได้ดัดแปลงขั้นตอนวิธีและโครงสร้างข้อมูลจากการตรวจหาค่าผิดปรกติแบบเดิม จากความสามารถในการตรวจหาค่าผิดปรกติมากสุดค่าเดียว เป็นการตรวจหาค่าผิดปรกติเพื่อช่วยในการพิจารณาไว้ใช้กับการตรวจหาค่าผิดปรกติของเครื่องบริการ ทั้งยังออกแบบและพัฒนาเครื่องมือที่ใช้สร้างข้อมูลบันทึกการใช้งานบริการเว็บที่มีรูปแบบปรกติและผิดปรกติที่หลากหลาย เพื่อใช้ในการวัดประสิทธิภาพของขั้นตอนวิธีต่างๆในการตรวจหาค่าผิดปรกติ โดยไม่จำเป็นต้องจัดเตรียมอุปกรณ์ที่ต้องอาศัยความชำนาญ รวมไปถึงรูปแบบการโจมตีที่ใช้ในการจัดเก็บข้อมูล

1.2 วัตถุประสงค์การวิจัย

1. การตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บ จากข้อมูลการลงบันทึกการเข้าใช้บริการเว็บ เพื่อเป็นอีกหนึ่งองค์ประกอบ ในการตรวจหาการค่าผิดปรกติของโปรแกรมประยุกต์เว็บ เพื่อเพิ่มความถูกต้องและความรวดเร็วในการตรวจหาค่าผิดปรกติ

2. เพื่อสร้างเครื่องมือสำหรับการสร้างข้อมูลการลงบันทึกการเข้าใช้บริการเว็บ (Web Access Log) สำหรับการตรวจหาค่าผิดปรกติจากปริมาณการใช้งาน โดยจะเป็นการสร้างข้อมูลปริมาณการใช้งานเท่านั้น ไม่รวมถึงเนื้อข้อมูลที่ใช้รับส่งกัน

3. ศึกษาวิธีการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บจากข้อมูลแบบต่อเนื่อง ด้วยขั้นตอนวิธีของงานวิจัยนี้ ว่าสามารถตรวจหาค่าผิดปรกติได้ดีเพียงใด

1.3 ขอบเขตการวิจัย

1. เครื่องมือในการสร้างข้อมูล ที่จะใช้ในการตรวจหาค่าผิดของโปรแกรมประยุกต์เว็บนั้น จะเป็นการสร้างข้อมูลการลงบันทึกการเข้าใช้บริการเว็บ ของเครื่องบริการเว็บที่มีการจัดเก็บข้อมูลที่เป็นไปตามมาตรฐาน

2. นำแนวคิดในการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค มาประยุกต์ใช้กับการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลงกับข้อมูลการลงบันทึกการเข้าใช้บริการเว็บ โดยใช้ข้อมูลแบบไม่ต่อเนื่อง (Non-streaming Data) ในการหาแต่ละครั้ง

3. ศึกษาความเป็นไปได้ในการนำข้อมูลแบบต่อเนื่อง (Streaming Data) โดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค มาประยุกต์ใช้กับการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลงกับข้อมูลการลงบันทึกการเข้าใช้บริการเว็บ

4. ตรวจสอบความถูกต้องจากการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบต่อเนื่อง (Streaming data) มาเปรียบเทียบกับการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบไม่ต่อเนื่อง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้วิธีที่มีประสิทธิภาพไว้เป็นอีกหนึ่งองค์ประกอบ เพื่อเพิ่มความแม่นยำในการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บ

2. ได้โปรแกรมสำหรับสร้างข้อมูลสำหรับการทดสอบ การตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บ

3. สามารถใช้การแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค มาประยุกต์ใช้กับงานอนุกรมเวลาทางด้านเครือข่าย

4. สามารถใช้ขั้นตอนวิธีในการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลง มาประยุกต์ใช้กับข้อมูลอนุกรมเวลาในรูปแบบอื่นๆ

1.5 ขั้นตอนและวิธีดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บทั่วไป
2. ศึกษางานวิจัยอื่นที่เกี่ยวข้องกับการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บโดยการใช้การแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค
3. ออกแบบ พัฒนา โปรแกรมที่ใช้สร้างข้อมูลเพื่อการทดสอบ และโปรแกรมที่ใช้ตรวจหาค่าผิดปรกติ โดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค และขั้นตอนวิธีในการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลง
4. ทดสอบการตรวจหาค่าผิดปรกติ จากข้อมูลที่สร้างขึ้น และข้อมูลจริง
5. จัดเก็บผลลัพธ์ที่ได้และวิเคราะห์ระบบเบื้องต้นเพื่อปรับปรุงแก้ไข
6. สรุปผลและจัดทำเอกสารวิทยานิพนธ์

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 อนุกรมเวลา (Time Series) [20]

คือ เซตของข้อมูลเชิงปริมาณที่เกิดขึ้นตามลำดับเวลาอย่างต่อเนื่องกันโดยข้อมูลไม่จำเป็นที่จะต้องมีขนาดเท่ากัน อาจเป็นปี เดือน สัปดาห์ วัน ชั่วโมง วินาที หรือส่วนของวินาทีก็ได้ ตัวอย่างเช่น ดัชนีตลาดหลักทรัพย์ในแต่ละวันเมื่อปิดทำการซื้อขายในแต่ละวัน ปริมาณผู้เข้าใช้งานเว็บในแต่ละวัน เป็นต้น

2.2 การทำข้อมูลให้เป็นบรรทัดฐาน

2.2.1 Z-Score Normalization

คือ การทำข้อมูลให้เป็นบรรทัดฐาน โดยค่าเฉลี่ยมีค่าเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 1 โดยคำนวณจากสมการต่อไปนี้

$$z = \frac{x - \bar{x}}{std}$$

x คือ ค่าของข้อมูลในอนุกรมเวลา

$\bar{x}$ คือ ค่าเฉลี่ยซึ่งหาได้จากสมการ $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

std คือ ส่วนเบี่ยงเบนมาตรฐาน ซึ่งหาได้จากสมการ $std = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

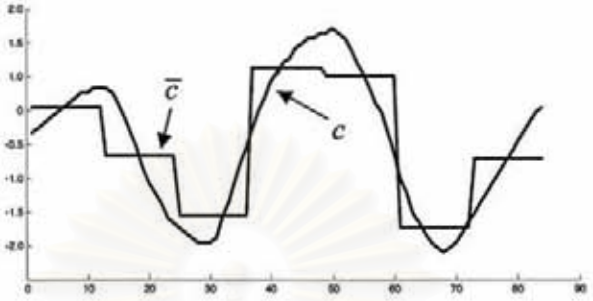
2.2.2 Piecewise Aggregation Approximation (PAA) [18][19]

คือ วิธีการลดมิติของข้อมูลโดยการแบ่งข้อมูลออกเป็นส่วนๆ ส่วนละเท่าๆ กัน และมีการคำนวณค่าเฉลี่ยของแต่ละส่วนเพื่อเป็นตัวแทนของข้อมูลแต่ละส่วน ดังรูปที่ 2.1 โดยกำหนดให้ความยาวของอนุกรมเวลา c มีจำนวน n จุดข้อมูล (Data Points) และทำการแบ่งข้อมูลออกเป็น w ส่วน โดยที่แต่ละส่วนจะมีจำนวนจุดข้อมูลที่เท่ากัน ซึ่งจะต้องทำการหาค่าเฉลี่ย $\bar{c}$ ของจุดข้อมูลแต่ละส่วนที่ถูกทำการแบ่ง โดยค่าเฉลี่ยของค่า $\bar{c}$ ตัวที่ i สามารถหาได้จากสมการต่อไปนี้

$$\overline{c_i} = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j$$

c_i คือ ค่าเฉลี่ยของจุดข้อมูลแต่ละส่วนที่ถูกทำการแบ่ง

c_j คือ ค่าของข้อมูลในแต่ละจุด



รูปที่ 2.1 แสดงเส้นอนุกรมเวลา c ที่ถูกแปลงให้เป็น PAA $\overline{c}$ เทียบกับเส้นอนุกรมเวลาแบบเดิม

2.3 SAX [18][19]

SAX (Symbolic Aggregate approXimation) เป็นการแทนค่าสัญลักษณ์ให้กับอนุกรมเวลาที่มีประสิทธิภาพรูปแบบหนึ่ง โดยมีคุณสมบัติพิเศษที่ทำการลดความซับซ้อนและมิติของข้อมูล โดยยังสามารถคงลักษณะต่างๆที่สำคัญ ของอนุกรมเวลาได้อย่างถูกต้อง

ข้อมูลก่อนที่จะแปลงเป็นค่าสัญลักษณ์นั้น จะต้องทำการลดมิติของข้อมูลด้วยวิธี PAA และทำข้อมูลให้เป็นบรรทัดฐานด้วยวิธี Z-Score โดยข้อมูลที่มีอยู่ ณ ขณะนี้จะมีการแจกแจงเกาส์เซียน ทำให้สามารถกำหนดจุดแบ่ง (Breakpoint) β ที่จะสร้างพื้นที่ใต้กราฟของการแจกแจงเกาส์เซียนที่มีขนาดเท่ากันได้ โดยจุดแบ่งเหล่านี้สามารถหาได้จากตารางสถิติ ดังตัวอย่างในตารางที่ 2.1 ที่แสดงจำนวนจุดแบ่งขนาดต่างๆ β_i ที่แบ่งพื้นที่ใต้กราฟของการแจกแจงเกาส์เซียนให้มีขนาดเท่ากันตั้งแต่ 3 ส่วนถึง 10 ส่วน ซึ่งก็คือจำนวนตัวอักษร (a) ที่ใช้ในการแทนค่าสัญลักษณ์

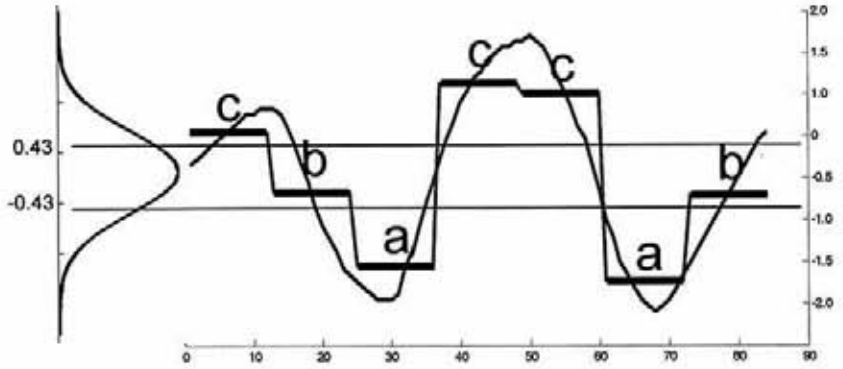
ตารางที่ 2.1 ตัวอย่างจุดแบ่งพื้นที่ใต้กราฟของการแจกแจงเกาส์เซียน

a / β_i	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

หลังจากทำการลดมิติของข้อมูลด้วยวิธี PAA และทำข้อมูลให้เป็นบรรทัดฐานด้วยวิธี Z-Score สามารถทำการแทนค่าสัญลักษณ์ดังแสดงในรูปที่ 2.2 ได้โดย

- ตัวอย่างเช่น หากขนาดของสัญลักษณ์ที่ต้องการใช้ (a) มีค่าเท่ากับ 3 ให้ดูจุดแบ่งพื้นที่ใต้กราฟที่มีขนาดเท่ากับ 3 จากตารางที่ 2.1 จะได้ค่าที่น้อยที่สุดของจุดแบ่ง β_1 มีค่าเท่ากับ -0.43 และ ค่ามากที่สุดของจุดแบ่ง β_{a-1} ซึ่งเท่ากับ β_2 มีค่าเท่ากับ 0.43
- การเปรียบเทียบข้อมูลที่ผ่านการลดมิติกับจุดแบ่งพื้นที่ใต้กราฟ มีเงื่อนไขดังนี้

1. หากจุดของกราฟที่มีค่าของข้อมูลน้อยกว่าค่าที่น้อยที่สุดของจุดแบ่ง ($x < \beta_1$) แล้ว ค่าของข้อมูลตัวนั้นก็จะถูกแทนด้วยสัญลักษณ์ "a"
2. หากจุดของกราฟที่มีค่าของข้อมูลมากกว่าหรือเท่ากับจุดแบ่ง แต่น้อยกว่าจุดแบ่งถัดไป ($\beta_1 \le x < \beta_{a-1}$) แล้ว ค่าของข้อมูลตัวนั้นก็จะถูกแทนด้วยสัญลักษณ์ "b"
3. หากจุดของกราฟที่มีค่าของข้อมูลมากกว่าค่าที่มากที่สุดของจุดแบ่ง ($x \ge \beta_{a-1}$) แล้ว ค่าของข้อมูลตัวนั้นก็จะถูกแทนด้วยสัญลักษณ์ "c"
4. หลักการนี้ใช้กับขนาดของสัญลักษณ์ทุกขนาด



รูปที่ 2.2 เส้นอนุกรมเวลาที่ถูกแปลงให้เป็น PAA แล้วทำการแทนค่าข้อมูลให้อยู่ในรูปสัญลักษณ์

จากรูปที่ 2.2 แสดงอนุกรมเวลาที่มีขนาดข้อมูล 84 จุดข้อมูล ถูกแบ่งออกเป็น 7 ส่วนที่มีขนาดเท่าๆ กัน และมีจำนวนของสัญลักษณ์เท่ากับ 3 (n = 84, w = 7 และ a = 3) และพบว่าอนุกรมเวลาสามารถแปลงได้เป็นอักขระ cbaccab

2.4 การวัดระยะห่าง

2.4.1 การหาระยะห่างแบบยุคลิด (Euclidean Distance)

เป็นการหาระยะห่างที่ง่ายและใช้งานกันอย่างแพร่หลายที่สุด โดยมีหลักการดังนี้ หากเรามีอนุกรมเวลาอยู่สองตัว ได้แก่ อนุกรมเวลา Q และอนุกรมเวลา C ที่มีขนาดเท่ากันแล้ว การหาระยะห่างของอนุกรมเวลาทั้งสองสามารถหาได้จากสมการ

$$D(Q,C) \equiv \sqrt{\sum_{i=1}^{n}(q_i - c_i)^2}$$

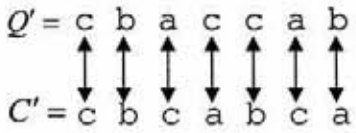
q_i คือ จุดที่ i บนอนุกรมเวลา Q

c_i คือ จุดที่ i บนอนุกรมเวลา C

n คือ ความยาวของอนุกรมเวลาทั้งสอง

2.4.2 การหาระยะห่างแบบแซ็ค (SAX Distance) [19]

การหาระยะห่างของอนุกรมเวลาที่เป็นค่าสัญลักษณ์ สามารถทำได้โดยนำค่าระยะห่างของตัวอักขระแต่ละคู่ในตำแหน่งเดียวกันมารวมกัน ดังแสดงในรูปที่ 2.3



รูปที่ 2.3 ระยะห่างของข้อมูลที่เป็นสัญลักษณ์จำนวน 2 สายอักขระ

การหาค่าระยะห่างระหว่างอักขระแต่ละคู่ในตำแหน่งเดียวกันของ 2 สายอักขระ ที่มีการกำหนดจำนวนสัญลักษณ์ที่ใช้เป็น 4 ให้ทำการสร้างตารางที่มีขนาด 4x4 ซึ่งแต่ละคู่ตัวอักษรที่ต้องการหาค่าระยะห่างให้กำหนดสัญลักษณ์ตัวใดตัวหนึ่งเป็นแถว (Row) และให้สัญลักษณ์ที่เหลือเป็นสดมภ์ (Column) เช่น หากต้องการหาระยะห่างของสัญลักษณ์ *a* และสัญลักษณ์ *b* โดยกำหนดสัญลักษณ์ *a* ให้เป็นแถว (*r*) ซึ่งมีค่าเท่ากับ 1 และให้สัญลักษณ์อีกตัวหนึ่งเป็นสดมภ์ (*c*) ก็คือ *b* ซึ่งมีค่าเท่ากับ 2 แล้วให้ดูค่าที่อยู่ในตารางที่ 2.2 ที่เป็นตำแหน่งของแถวและสดมภ์นั้น ซึ่งในกรณีนี้จะได้ระยะห่างที่มีค่าเท่ากับ 0

ตารางที่ 2.2 ค่าระยะห่างสำหรับข้อมูลที่แทนค่าสัญลักษณ์แล้ว
โดยใช้สัญลักษณ์ทั้งหมด 4 ตัว

	a	b	c	d
a	0	0	0.67	1.34
b	0	0	0	0.67
c	0.67	0	0	0
d	1.34	0.67	0	0

โดยสามารถคำนวณหาค่าระยะห่างที่อยู่ในแต่ละตำแหน่งของตารางที่ 2.2 ได้จากสมการดังต่อไปนี้

$$cell_{r,c} = \begin{cases} 0 & ,if\ |r-c| \le 1 \\ \beta_{\max(r,c)-1} - \beta_{\min(r,c)}, & otherwise \end{cases}$$

r คือ ตำแหน่งของแถวที่ทำการหาค่าความแตกต่าง

c คือ ตำแหน่งของสดมภ์ที่ทำการหาค่าความแตกต่าง

$\beta_{\max(r,c)-1}$ คือ ค่าจุดแบ่งจากตารางที่ 2.1 ซึ่งหาได้จากการหาค่าที่มากที่สุดระหว่าง r และ c แล้วลบด้วย 1 เช่น หากกำหนดให้ r มีค่าเท่ากับ 1 และc มีค่าเท่ากับ 2 แล้วค่าจุดแบ่งที่ได้จะมีค่าเท่ากับ $\beta_{\max(1,2)-1} = \beta_{2-1} = \beta_1$ เป็นต้น

$\beta_{\min(r,c)}$ คือ ค่าจุดแบ่งจากตารางที่ 2.1 ซึ่งหาได้จากการหาค่าที่น้อยที่สุดระหว่าง r และ c เช่น หากกำหนดให้ r มีค่าเท่ากับ 1 และ c มีค่าเท่ากับ 2 แล้ว ค่าจุดแบ่งที่ได้จะมีค่าเท่ากับ $\beta_{\min(1,2)} = \beta_1$ เป็นต้น

หากกำหนดจำนวนของสัญลักษณ์ที่ใช้เท่ากับ 4 จะได้ค่าจุดแบ่งจากตารางที่ 2.1 ที่ทำการแบ่งพื้นที่ใต้กราฟของการแจกแจงเกาส์เซียนเป็น -0.67 0 และ 0.67 ตามลำดับ

การหาระยะห่างระหว่างสัญลักษณ์ a และ b สามารถดูได้จาก

- โดยนำค่า r และ c มาพิจารณาว่าต่างกันน้อยกว่าหรือเท่ากับ 1 หรือไม่ ถ้าใช่ระยะห่างระหว่างตัวอักษรสัญลักษณ์นั้นก็จะมีค่าเท่ากับ 0 เช่น dist(a,b) มีค่า r และ c มีค่าเท่ากับ 1 และ 2 ตามลำดับ จึงทำให้ระยะห่างเท่ากับ 0 และ dist (b,c) มีค่า r และ c มีค่าเท่ากับ 2 และ 3 ตามลำดับ จึงทำให้ระยะห่างเท่ากับ 0 เช่นกัน
- หากค่า r และ c มีความต่างกันมากกว่า 1 ให้คำนวณตามสมการ ดังตัวอย่างในตารางที่ 2.3

ตารางที่ 2.3 ตัวอย่างการหาค่าระยะห่างที่ค่า r และ c มีความแตกต่างกันมากกว่า 1

สัญลักษณ์	r	C	$\beta_{\max(r,c)-1}$	$\beta_{\min(r,c)}$	ระยะห่าง ($\beta_{\max(r,c)-1} - \beta_{\min(r,c)}$)
a,c	1	3	0	-0.67	D(a,c) = 0.67
a,d	1	4	0.67	-0.67	D(a,d) = 1.34

2.5 การตรวจหาค่าผิดปรกติ (Anomaly Detection) [21]

คือ การตรวจหาลักษณะข้อมูลที่มีรูปแบบการโจมตีหรือมีการใช้งานแตกต่างไปจากการใช้งานปกติ โดยการตรวจหาค่าผิดปรกติมีประโยชน์มากมาย ตัวอย่างเช่น ทำให้ทราบว่ามีการโจมตีหรือไม่ ช่วยกำหนดนโยบายในการใช้งานเครือข่าย ทำให้มีข้อมูลในการตรวจสอบร่องรอยของการโจมตีในภายหลัง ช่วยให้ง่ายต่อการจัดการทรัพยากรในการใช้งานเครือข่ายให้เป็นไปอย่างมีประสิทธิภาพ เป็นต้น โดยทั่วไปแล้วการตรวจหาค่าผิดปรกติในเครือข่ายสามารถแบ่งได้เป็น 2 กลุ่มใหญ่ๆ ดังต่อไปนี้

- การตรวจหาค่าผิดปรกติโดยการใช้ลักษณะบ่งชี้ (Signature Based)

เป็นการตรวจหาค่าผิดปรกติโดยการเปรียบเทียบความเหมือนของข้อมูลเครือข่าย กับลักษณะบ่งชี้ของการโจมตี ซึ่งถ้าเหมือนกันแสดงว่ามีการโจมตีเกิดขึ้น แต่ถ้าไม่เหมือนกันแสดงว่าไม่มีการโจมตีเกิดขึ้น โดยมีข้อดีและข้อเสียดังแสดงในตารางที่ 2.4

ตารางที่ 2.4 ข้อดีและข้อเสียของการตรวจหาค่าผิดปรกติโดยการใช้ลักษณะบ่งชี้

ข้อดี
- การพัฒนาลักษณะบ่งชี้ในการตรวจสอบสามารถทำได้ง่าย - สามารถตรวจหาค่าผิดปรกติได้ง่าย หากพฤติกรรมการใช้งานเครือข่ายมีรูปแบบที่คงที่ - สามารถยกเลิกลักษณะบ่งชี้ที่ไม่ต้องการตรวจสอบได้ หากต้องการตรวจสอบเพียงบางโปรโตคอล
ข้อเสีย
- ข้อมูลโอเวอร์โหลด หากมีปริมาณการใช้งานมากเนื่องจากข้อมูลที่จะนำไปใช้ในการวิเคราะห์มีจำนวนมาก - ไม่สามารถตรวจหาค่าผิดปรกติจากการโจมตีแบบใหม่ได้ - ไม่สามารถตรวจหาค่าผิดปรกติจากการเปลี่ยนแปลงพฤติกรรมการใช้งานได้ - ไม่สามารถตรวจหาค่าผิดปรกติจากข้อมูลที่เข้ารหัสได้ - ไม่สามารถตรวจหาค่าผิดปรกติจากประเภทข้อมูลที่รับส่งผ่านช่องทางต่างๆได้ (Covert Channel - ช่องทางแอบแฝง) - มีการแจ้งเตือนว่าเป็นค่าผิดปรกติทั้งที่ความเป็นจริงแล้วเป็นการใช้งานปรกติ (False Positive) เป็นจำนวนมาก - การตรวจหาค่าผิดปรกติทำได้ช้า เนื่องจากจำเป็นต้องพัฒนาลักษณะบ่งชี้ตลอดเวลา ทำให้ลักษณะบ่งชี้ที่ต้องเทียบกับข้อมูลเครือข่ายมีจำนวนมาก

- การตรวจหาค่าผิดปรกติจากความผิดปรกติ (Anomaly Based)

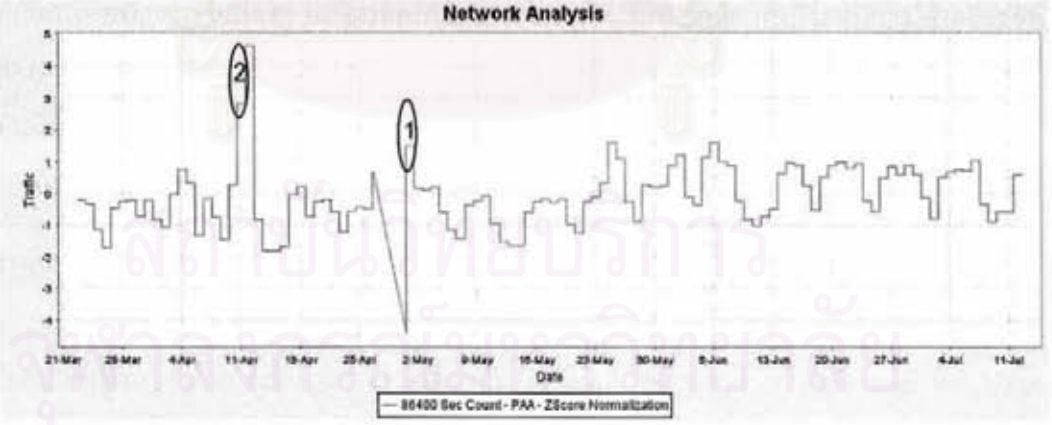
เป็นการตรวจหาค่าผิดปรกติที่ทำงานแตกต่างกับการใช้ลักษณะบ่งชี้อย่างสิ้นเชิง โดยจะเป็นการเรียนรู้ถึงลักษณะที่ปรกติก่อน หากพบว่ามีข้อมูลที่ไม่เหมือนกับลักษณะปรกติที่เรียนรู้ไว้ก็จะทำการแจ้งเตือนว่าเป็นการโจมตี โดยมีข้อดีและข้อเสียดังแสดงในตารางที่ 2.5

ตารางที่ 2.5 ข้อดีและข้อเสียของการตรวจหาค่าผิดปรกติจากความผิดปรกติ

ข้อดี
- สามารถตรวจหาค่าผิดปรกติจากข้อมูลที่เข้ารหัสได้ - สามารถทำงานได้รวดเร็ว หากมีการปรับแต่งข้อมูลในการตรวจหาค่าผิดปรกติแล้ว เพราะไม่จำเป็นที่จะต้องสร้างลักษณะบ่งชี้ใหม่ - สามารถตรวจหาค่าผิดปรกติได้ หากการโจมตีทำให้การใช้งานไม่อยู่ในแบบรูปที่เป็นปรกติ - สามารถตรวจสอบข้อมูลที่มีการรับส่งผ่านช่องทางแอบแฝงได้
ข้อเสีย
- ไม่สามารถตรวจหาค่าผิดปรกติแบบรูปใหม่หรือแบบรูปที่แตกต่างจากพฤติกรรมที่กำหนดว่าเป็นพฤติกรรมปรกติได้ - ไม่สามารถตรวจสอบพฤติกรรมผิดปรกติที่เกิดขึ้นอย่างไม่ชัดเจนได้ - การกำหนดกฎในการตรวจหาค่าผิดปรกติเป็นไปได้ยาก - ต้องเรียนรู้รายละเอียดของพฤติกรรมในการใช้งานที่เป็นปรกติ

2.6 **การตรวจหาค่าผิดปรกติในอนุกรมเวลา** (Anomaly Detection in Time Series) [18]

เป็นการตรวจหาค่าผิดปรกติจากความผิดปรกติในอนุกรมเวลา โดยการหาลำดับย่อยในอนุกรมเวลาที่มีลักษณะคล้ายกับลำดับย่อยอื่นๆในอนุกรมเวลาน้อยที่สุด ซึ่งลำดับย่อยดังกล่าวเรียกว่าลำดับย่อยที่มีค่าผิดปรกติ ดังตัวอย่างแสดงในรูปที่ 2.4



รูปที่ 2.4 ค่าผิดปรกติในอนุกรมเวลา

จากรูปที่ 2.4 แสดงปริมาณการร้องขอการบริการเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย ระหว่างวันที่ 22 มีนาคม ค.ศ. 2006 เวลา 14 นาฬิกา 52 นาที 32 วินาที ถึงวันที่ 12 กรกฎาคม

ค.ศ. 2006 เวลา 12 นาฬิกา 5 นาที 3 วินาที ที่ผ่านการตรวจหาค่าผิดปรกติแล้วพบข้อมูลที่มีลักษณะที่ผิดปรกติอยู่ 2 ตำแหน่งด้วยกันคือ ตำแหน่งที่ 1 คือช่วงวันที่ 27 เมษายน ถึงวันที่ 2 พฤษภาคม มีลักษณะของข้อมูลที่ขาดช่วงอันเนื่องมาจากปัญหาในการเขียนบันทึกการใช้งานบริการเว็บของโปรแกรมประยุกต์เว็บ และตำแหน่งที่ 2 คือวันที่ 11 เมษายน มีลักษณะข้อมูลที่มีปริมาณการส่งออกมากกว่าวันอื่น โดยพบว่าวันนั้นมีการดาวน์โหลด (Download) ไฟล์พีดีเอฟ (PDF) ซึ่งในหากสังเกตในรูปกราฟแล้ว จะพบว่ามีลักษณะที่แตกต่างจากข้อมูลในตำแหน่งอื่นๆ อย่างชัดเจน ทำให้พบว่าหากตรวจหาค่าผิดปรกติได้เพียง 1 ตำแหน่งจะทำให้ไม่ทราบลักษณะผิดปรกติอื่นๆ ที่ส่งผลต่อโปรแกรมประยุกต์เว็บที่มีอันดับรองลงมา

2.7 การตรวจหาค่าผิดปรกติในอนุกรมเวลาจากทุกความเป็นไปได้ (Brute Force Anomaly Detection in Time Series) [18]

เป็นการตรวจหาค่าผิดปรกติจากความผิดปรกติในอนุกรมเวลา โดยจะเป็นการสร้างลำดับย่อยที่เป็นไปได้ทั้งหมดในการตรวจหาค่าผิดปรกติ และทำการเปรียบเทียบลำดับย่อยที่สร้างขึ้นมาทั้งหมดในอนุกรมเวลาเพื่อหาว่าลำดับย่อยใดมีค่าผิดปรกติ ซึ่งมีการกล่าวถึงรายละเอียดในงานวิจัยที่เกี่ยวข้องหัวข้อ 2.8.3 หัวข้อย่อยการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ (Brute Force)

2.8 งานวิจัยที่เกี่ยวข้อง

จากการค้นคว้างานวิจัยที่เกี่ยวข้องกับการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บ ผู้เขียนได้พบงานวิจัยต่างๆ ที่เกี่ยวข้องกับการตรวจหาค่าผิดปรกติจากปริมาณการใช้เครือข่าย การตรวจหาค่าผิดปรกติโดยใช้วิธีแซ็ค และยังพบว่างานวิจัยการวิเคราะห์อนุกรมเวลาด้วยวิธีแซ็คนั้นมีประสิทธิภาพในการตรวจหาค่าผิดปรกติกับข้อมูลหลายประเภท ซึ่งมีรายละเอียดดังนี้

2.8.1 Barford, P., Kline, J., Plonka, D. และ Ron, A. "A Signal Analysis of Network Traffic Anomalies" [1]

เป็นงานวิจัยที่นำข้อมูลที่รับส่งกันในเครือข่ายคอมพิวเตอร์มาวิเคราะห์หาค่าผิดปรกติ โดยมีการนำมาจากแหล่งข้อมูล 2 แหล่งด้วยกัน อันได้แก่ ข้อมูลจาก IP Flow และ ข้อมูลจาก SNMP มาวิเคราะห์ร่วมกัน โดยผู้วิจัยใช้เทคนิค Wavelet เพื่อแยกลักษณะของสัญญาณที่เกิดขึ้นในข้อมูล เพื่อนำข้อมูลที่ได้ผ่านการกรองนั้น มาวิเคราะห์หาค่าผิดปรกติด้วยวิธีที่ผู้วิจัยได้คิดค้นขึ้นคือเทคนิค Deviation Score โดยได้นำไปเปรียบเทียบกับเทคนิค Holt-Winters Forecasting ซึ่งผลที่ได้รับก็คือ เทคนิค Deviation Score มีผลการตรวจหาค่าผิดปรกติที่ดีกว่า แต่ผู้วิจัยไม่ได้ทำ

การทดสอบว่าเทคนิคใดมีการแจ้งเตือนว่ามีค่าผิดปรกติที่ไม่ถูกต้องมากกว่า นอกจากนี้ข้อมูลที่นำมาวิเคราะห์นั้นก็เป็นเพียงข้อมูลของมหาวิทยาลัยแห่งเดียวโดยมีระยะเวลาของข้อมูลที่ใช้เก็บเพียง 6 เดือน และเป็นข้อมูลที่เก็บจากปริมาณการใช้งานในเครือข่ายทุกๆ 5 นาที ไม่ใช่ปริมาณการใช้งานเครือข่ายทั้งหมด

2.8.2 Lin, J., Keogh, E., Lonardi, S., Lankford, J. P. and Nystrom, D. M. "Visually Mining and Monitoring Massive Time Series" [22]

เป็นงานวิจัยที่นำเสนอเครื่องมือสร้างภาพนามธรรมเพื่อช่วยในการวิเคราะห์ข้อมูลอนุกรมเวลา โดยใช้เทคนิคในการแปลงอนุกรมเวลาเป็นตัวแทนสัญลักษณ์ด้วยวิธีแซ็ค และเข้ารหัสข้อมูลตัวแทนสัญลักษณ์ให้อยู่ในรูปต้นไม้แบบต่อท้าย (Suffix Tree) ซึ่งแบบรูป (Pattern) ความถี่ถูกแปลงเป็นสีต่างๆ และแบบรูปคุณสมบัติอื่นๆ ถูกแปลงเป็นคุณสมบัติของภาพต่างๆ โดยมีการประเมินเปรียบเทียบขั้นตอนวิธีของงานวิจัยนี้กับขั้นตอนวิธีแบบกลุ่มที่ทันสมัยหลายแบบ ด้วยชุดข้อมูลจริงและชุดข้อมูลสังเคราะห์หลายชุด และพบว่างานวิจัยนี้ สามารถหาสิ่งที่ต้องการได้ถูกต้อง ทั้งวิธีที่ต้องการความรู้ก่อนหน้า (Prior Knowledge) วิธีที่ไม่ต้องการความรู้ก่อนหน้า โดยงานวิจัยนี้สามารถทำงานได้รวดเร็วกว่าขั้นตอนวิธีอื่น โดยทางผู้วิจัยกล่าวว่าจะขยายขอบเขตของชุดข้อมูลจากเดิมที่เป็นชุดข้อมูลจำกัดให้เป็นข้อมูลที่ได้รับมาแบบต่อเนื่อง

2.8.3 Keogh, E., Lin, J. and Fu, A. "HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications" [18]

เป็นงานวิจัยที่ทำการสร้างขั้นตอนวิธีที่นำมาตรวจหาค่าผิดปรกติ (Discord) ของอนุกรมเวลา โดยทางผู้วิจัยพบว่าการตรวจหาค่าผิดปรกติที่เคยปฏิบัติกันมามีการใช้พารามิเตอร์จำนวนหลายตัวด้วยกัน แต่การตรวจหาค่าผิดปรกติของงานวิจัยนี้เป็นการใช้พารามิเตอร์เพียงตัวเดียวเท่านั้นคือขนาดของอนุกรมเวลาย่อย โดยมีขั้นตอนวิธี (Algorithm) ที่ใช้ในการตรวจหาค่าผิดปรกติ ดังต่อไปนี้

- การตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้

มีหลักการดังนี้ คือมีการแบ่งอนุกรมเวลา T ความยาว $|T|$ ออกเป็นลำดับย่อย (Subsequence) ที่แต่ละลำดับย่อยมีความยาว n จะได้จำนวนลำดับย่อยทั้งหมดเท่ากับ $|T| - n + 1$ ซึ่งการตรวจหาค่าผิดปรกติสามารถตรวจได้จากระยะห่างจากลำดับย่อยทั้งหมดที่เกิดขึ้นในอนุกรมเวลา โดยแบ่งการวนรอบการค้นหาออกเป็น 2 ส่วน ได้แก่การวนรอบภายนอก $|T| - n + 1$ ครั้ง และการวนรอบภายใน $|T| - n + 1$ ครั้ง โดยการวนรอบภายนอกในครั้งแรกนั้น จะหาระยะห่างระหว่างลำดับย่อยแรกกับการวนรอบภายในซึ่งก็คือลำดับย่อยที่เกิดขึ้นทั้งหมดในอนุกรมเวลา

แล้วเลือกระยะห่างที่น้อยที่สุด ซึ่งระยะห่างที่น้อยที่สุดดังกล่าวจะต้องไม่ใช่ลำดับย่อยที่เป็นส่วนหนึ่งของลำดับย่อยที่เริ่มทำการค้นหา (Non-self Match) เมื่อหาระยะห่างของลำดับย่อยแรกกับลำดับย่อยที่เหลือทั้งหมดเสร็จสิ้นแล้ว ก็จะทำการหาระยะห่างจากลำดับย่อยถัดไป เพื่อหาระยะห่างที่มีเงื่อนไขเช่นเดียวกับการวนรอบภายนอกครั้งแรก ทำเช่นนี้ไปเรื่อยๆ จนหมดการวนรอบของลำดับย่อยที่มีอยู่ทั้งหมด จากนั้นให้เลือกระยะห่างของลำดับย่อยที่มีค่ามากที่สุดจากระยะห่างที่น้อยที่สุดของแต่ละลำดับย่อยที่ได้จากการค้นหาดังกล่าวและระบุว่าลำดับย่อยที่มีระยะห่างดังกล่าว เป็นลำดับย่อยที่มีค่าผิดปรกติ ดังแสดงรหัสเทียม (Pseudo Code) ในตารางที่ 2.6

ซึ่งผู้วิจัยพบว่าการตรวจหาค่าผิดปรกติด้วยวิธีตรวจหาคำตอบในทุกความเป็นไปได้ทำงานได้ช้า ทางผู้วิจัยจึงได้เสนอขั้นตอนวิธีที่ทำการแก้ไขวิธีการตรวจหาคำตอบในทุกความเป็นไปได้ โดยเพิ่มเงื่อนไขในการวนซ้ำ และทำการเรียงลำดับข้อมูลก่อนที่จะทำการตรวจหาค่าผิดปรกติ พบว่าจำนวนการวนรอบในการตรวจหาค่าผิดปรกตินั้นน้อยลง ทำให้ทำงานได้เร็วขึ้น ซึ่งก็คือการตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก (Heuristic)

ตารางที่ 2.6 รหัสเทียมของการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้
(ที่มา:Jessica Lin et al.) [18]

```
1  Function [dist,loc] = Brute_Force(T,n)
2    best_so_far_dist = 0
3    best_so_far_loc = NaN
4
5    For p = 1 to |T| - n + 1                          // Begin Outer Loop
6      Nearest_neighbor_dist = infinity
7      For q = 1 to |T| - n + 1                        // Begin Inner Loop
8        IF |p - q| ≥ n                                // non-self match?
9          IF Dist(tp,…,tp+n-1, tq,…,tq+n-1) < nearest_neighbor_dist
10           Nearest_neighbor_dist = Dist(tp,…,tp+n-1, tq,…,tq+n-1)
11         End
12       End                                           // End non-self match test
13     End                                             // End Inner Loop
14     IF nearest_neighbor_dist > best_so_far_dist
15       Best_so_far_dist = nearest_neighbor_dist
16       Best_so_far_loc = p
17     End
18 End                                                 // End Outer Loop
19 Return [best_so_far_dist, best_so_far_loc]
```

- การตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก

หลักการจะเหมือนกับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ แต่ทางผู้วิจัยพบว่าการหาค่าระยะห่างตามเงื่อนไขของการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ ของแต่ละลำดับย่อยนั้น ไม่มีความจำเป็นต้องหาจากลำดับย่อยทั้งหมด ซึ่งการหาค่าระยะห่างดังกล่าวต้องการหาเพียงค่าที่มีค่าน้อยกว่าระยะห่างที่ดีที่สุดปัจจุบันของแต่ละลำดับย่อย ดังนั้นงานวิจัยนี้จึงมีการแก้รหัสเทียม เพื่อให้รองรับกับแนวคิดนี้ โดยการเพิ่มแนวคิดที่จำเป็นต้องมีการเรียงลำดับข้อมูลที่อยู่ในโครงสร้างข้อมูลก่อนที่จะนำมาวนซ้ำ และการตรวจสอบให้หยุดการวนรอบภายในเมื่อระยะห่างที่เกิดขึ้นในการค้นหา ณ ตำแหน่งที่มีค่าระยะห่างน้อยกว่าค่าระยะห่างที่น้อยที่สุดที่เคยค้นพบจากการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ ดังแสดงในบรรทัดที่ 9 - 11 ของตารางที่ 2.7

ตารางที่ 2.7 รหัสเทียมของการตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก

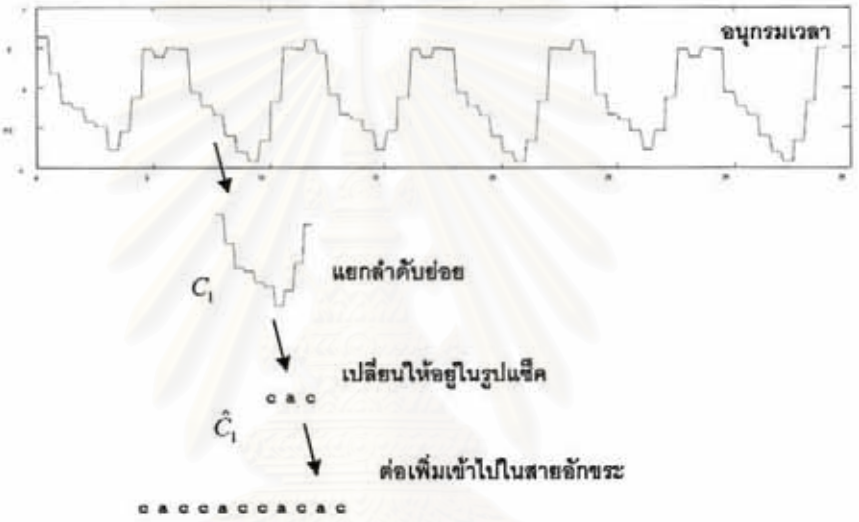
(ที่มา:Jessica Lin et al.) [18]

```
1  Function [dist,loc] = Brute_Force(T,n)
2    best_so_far_dist = 0
3    best_so_far_loc = NaN
4
5    For p = 1 to |T| - n + 1                          // Begin Outer Loop
6      nearest_neighbor_dist = infinity
7      For q = 1 to |T| - n + 1                        // Begin Inner Loop
8        IF |p - q| ≥ n                                // non-self match?
9          IF Dist(tp,…,tp+n-1, tq,…,tq+n-1) < best_so_far_dist
10           Break                                     // Break out of Inner Loop
11         End
12         IF Dist(tp,…,tp+n-1, tq,…,tq+n-1) < nearest_neighbor_dist
13           Nearest_neighbor_dist = Dist(tp,…,tp+n-1, tq,…,tq+n-1)
14         End
15       End                                           // End non-self match test
16     End                                             // End Inner Loop
17     IF nearest_neighbor_dist > best_so_far_dist
18       best_so_far_dist = nearest_neighbor_dist
19       best_so_far_loc = p
20     End
21   End                                               // End Outer Loop
22   Return [best_so_far_dist, best_so_far_loc]
```

หลังจากนั้นทางผู้วิจัยได้ทำการพัฒนาต่อ โดยนำข้อมูลที่จะทำการตรวจหาค่าผิดปรกติมาแปลงให้อยู่ในรูปสัญลักษณ์ด้วยวิธีแซ็คก่อน แล้วจึงสร้างโครงสร้างข้อมูลรูปแบบต่างๆ ที่นำมาใช้กับขั้นตอนวิธีที่ได้ทำการวิจัยไว้ ดังนี้

- โครงสร้างข้อมูลสายอักขระ

เป็นโครงสร้างข้อมูลที่มีการแปลงอนุกรมเวลาให้อยู่ในรูปของสายอักขระเรียงต่อกัน โดยใช้วิธีแซ็ค ดังแสดงในรูปที่ 2.5 ซึ่งการตรวจหาค่าผิดปรกตินั้น จะใช้การวนซ้ำตอนนอกและตอนในกับข้อมูลสายอักขระชุดเดียวกัน โดยโครงสร้างชนิดนี้สามารถใช้ได้กับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ หากกำหนดให้สายอักขระมีความยาว m และขนาดของลำดับย่อยเป็น n แล้ว จำนวนของลำดับย่อยที่สามารถเกิดขึ้นได้มีค่าเท่ากับ $(m - n) + 1$



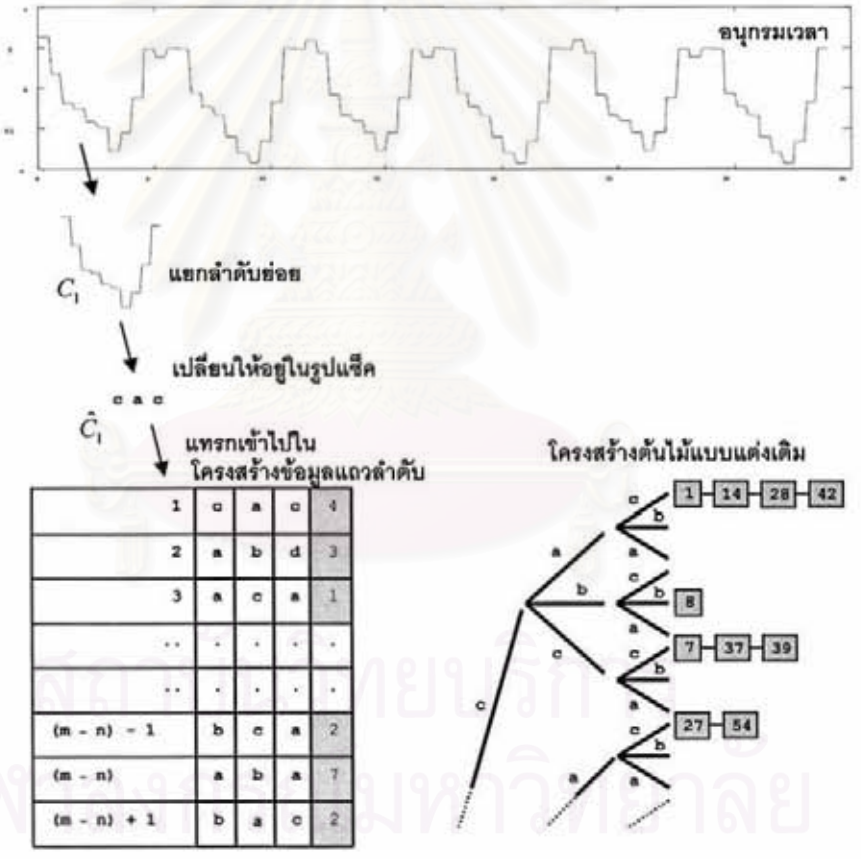
รูปที่ 2.5 โครงสร้างข้อมูลสายอักขระ (ที่มา:Jessica Lin et al.) [18]

- โครงสร้างข้อมูลแถวลำดับ (Array) และโครงสร้างต้นไม้แบบแต่งเติม (Augmented Tree)

เป็นโครงสร้างข้อมูลที่มาจากการแปลงอนุกรมเวลาให้อยู่ในรูปสายอักขระเรียงต่อกัน และมีการปรับแต่งให้มีโครงสร้างข้อมูลสองแบบคือ โครงสร้างข้อมูลแถวลำดับ เป็นโครงสร้างที่ใช้สำหรับการวนซ้ำตอนนอก และ โครงสร้างต้นไม้แบบแต่งเติม เป็นโครงสร้างที่ใช้สำหรับการวนซ้ำตอนใน ดังแสดงในรูปที่ 5 ซึ่งโครงสร้างชนิดนี้ สามารถใช้ได้กับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ การตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก โดยโครงสร้างข้อมูลทั้งสองแบบ มีรายละเอียดดังต่อไปนี้

โครงสร้างข้อมูลแถวลำดับ จะทำการเก็บสายอักขระที่เกิดขึ้นในแต่ละลำดับย่อยลงในแถวลำดับแต่ละแถว และจำนวนที่สายอักขระนั้นเกิดขึ้นในอนุกรมเวลาทั้งหมด จากรูปที่ 2.6 (ซ้าย) ข้อมูลตัวแรกของแถวลำดับที่ 1 จะมีสายอักขระที่เกิดขึ้นคือ c a c และมีจำนวนของสายอักขระที่มีค่าเดียวกับสายอักขระที่เกิดที่ตำแหน่งที่ 1 นี้ทั้งหมด 4 ตัว ถ้ากำหนดความยาวของสายอักขระให้เท่ากับ m ที่ลำดับย่อยมีขนาดเท่ากับ n จำนวนของลำดับย่อยที่สามารถเกิดขึ้นได้จะมีค่าเท่ากับ $(m - n) + 1$

โครงสร้างต้นไม้แบบแต่งเติม จะทำการสร้างโครงสร้างต้นไม้จากข้อมูลของสายอักขระที่เกิดขึ้นในแต่ละลำดับย่อย และข้อมูลในโครงสร้างต้นไม้ก็จะทำการเก็บตำแหน่งที่สายอักขระนั้นเกิดขึ้น หากมีตำแหน่งที่เกิดสายอักขระที่ซ้ำกันก็จะถูกเก็บใส่ไว้ในข้อมูลโครงสร้างต้นไม้ต่อกันไป จากรูปที่ 2.6 (ขวา) ข้อมูลของโครงสร้างต้นไม้ c a c แสดงให้เห็นว่ามีตำแหน่งที่เกิดสายอักขระเดียวกันทั้งหมดนี้ 4 ตำแหน่ง คือ ตำแหน่งที่ 1 14 28 และ 42



รูปที่ 2.6 โครงสร้างข้อมูลแถวลำดับ (ซ้าย)
และโครงสร้างต้นไม้แบบแต่งเติม (ขวา)
(ที่มา:Jessica Lin et al.) [18]

โดยการตรวจหาค่าผิดปรกติจากโครงสร้างข้อมูลชนิดนี้ จะต้องมีการเรียงลำดับข้อมูลที่อยู่ในโครงสร้างข้อมูลแถวลำดับ โดยจะเลือกเฉพาะแถวลำดับที่มีจำนวนที่สายอักขระเกิดขึ้นน้อยที่สุด ให้อยู่ในลำดับแรกๆในการวนซ้ำตอนนอก หลังจากนั้นข้อมูลที่เหลือจะไม่มีการเรียงลำดับแต่อย่างใดก่อนที่จะทำการวนซ้ำตอนนอก ส่วนการวนซ้ำตอนในก็จะทำการวนหาข้อมูลที่อยู่ในสายอักขระเดียวกันจากโครงสร้างต้นไม้แบบแต่งเติมก่อน แล้วจึงวนหาข้อมูลจากข้อมูลที่เหลือจากข้อมูลที่อยู่ในโครงสร้างต้นไม้ต่อไป

หลังจากนั้นได้ทำการประเมินผล โดยการนำขั้นตอนวิธีการตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก และโครงสร้างข้อมูลแถวลำดับ และโครงสร้างต้นไม้แบบแต่งเติมดังกล่าว ไปใช้ในการตรวจหาการค่าผิดปรกติจากข้อมูลแหล่งต่างๆ ทั้งทางด้านอากาศยาน การแพทย์ ทางด้านสาธารณูปโภค แล้วนำมาเปรียบเทียบกับขั้นตอนวิธีอื่นๆที่ใช้ในการตรวจหาค่าผิดปรกติ พบว่าสามารถทำการตรวจหาค่าผิดปรกติ ได้เป็นอย่างดี และมีจำนวนการวนรอบที่ใช้ในการตรวจหาค่าผิดปรกติที่น้อยลงกว่าวิธีการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้มาก

อย่างไรก็ตามขั้นตอนวิธีที่ผู้วิจัยเสนอมาทั้งสองวิธีนั้น หากข้อมูลมีลักษณะเหมือนหรือคล้ายกับลำดับย่อยที่เริ่มทำการค้นหา (ระยะห่างระหว่างลำดับย่อยเท่ากับศูนย์) แม้เพียงครั้งเดียวก็จะไม่สามารถค้นพบความผิดปรกตินั้นได้ และความผิดปรกติที่เกิดขึ้นรองลงมา กล่าวคือการตรวจหาค่าผิดปรกติที่มีค่าระยะห่างระหว่างลำดับย่อยรองลงมาก็ไม่สามารถหาได้เช่นกัน

2.8.4 Sommers, J., Yegneswaran, V. and Barford, P. "Recent Advances in Network Intrusion Detection System Tuning" [5]

เป็นงานวิจัยที่ทำการสร้างข้อมูลที่รับส่งกันในเครือข่ายคอมพิวเตอร์ โดยมุ่งเน้นไปที่การสร้างลักษณะของข้อมูล เพื่อใช้ในการแก้ไขปัญหาของระบบการตรวจหาการบุกรุกผ่านทางเครือข่าย โดยมีการสร้าง Trident Framework เพื่อทำการสร้างข้อมูลที่รับส่งกันในเครือข่ายคอมพิวเตอร์ ที่มีความสมจริงมากยิ่งขึ้น ทั้งการสร้างข้อมูลเครือข่ายที่เหมือนการใช้งานตามปกติ โดยมีการสร้างออโตมาตากำหนดลำดับในการรับส่งข้อมูลระหว่างกันโดยเป็นมอดูลที่ขยายต่อจาก Harphoon Framework และการสร้างการรับส่งข้อมูลที่มีการโจมตีระหว่างกัน โดยมีการเรียกใช้งานของ Mace Framework และ การสร้างการโจมตีจากข้อมูล DARPA [17] ซึ่งทำการสร้างข้อมูลแล้วไปทดสอบกับ ระบบการตรวจหาการบุกรุกผ่านทางเครือข่ายที่มีลักษณะต่างกัน 3 ตัวด้วยกัน ได้แก่ โบร (Bro) [23] สน้อร์ต (Snort) [24] และ บลีดสน้อร์ต (Bleed Snort) [25] ซึ่งผลที่ได้รับคือ โบร เหมาะกับการตรวจหาการบุกรุกการเชื่อมต่อแบบมีการวิเคราะห์สถานะของ Packet (Stateful) แต่สน้อร์ตเหมาะกับการการตรวจหาการบุกรุกแบบระบุเงื่อนไข โดย

บลีดสน้อร์ต คือ สน้อร์ต ที่มีจำนวนหลักเกณฑ์ในการตรวจหาการบุกรุกมากกว่า สน้อร์ต ธรรมดา มีการสร้างการเตือนภัยมากกว่า ซึ่งทำให้โอกาสที่จะเกิดการแจ้งว่ามีค่าผิดปรกติ ทั้งๆ ที่จริงแล้วไม่มีค่าผิดปรกติ (False-Positive หรือ False-Alarm) มากกว่า ทำให้ประสิทธิภาพในการตรวจจับลดลง และยังพบสิ่งที่น่าสนใจอีกว่า การจำลองการโจมตีวิธีเฉพาะวิธีหนึ่ง ทำให้ข้อมูลที่จะทำการตรวจหาการบุกรุกของทั้งโบร และบลีดสน้อร์ด สูญหายไปเป็นจำนวนมาก อย่างไรก็ตามการที่จะทดสอบความสามารถของระบบการตรวจหาการบุกรุกนี้ต้องอาศัยอุปกรณ์เครือข่ายที่พร้อม และความรู้ความชำนาญทางด้านเครือข่ายเป็นอย่างดี

บทที่ 3
การออกแบบการตรวจหาค่าผิดปรกติ และสร้างข้อมูลบันทึกการใช้งานบริการเว็บ

งานวิจัยนี้เป็นการตรวจหาค่าผิดปรกติของบันทึกการใช้งานบริการเว็บโดยการแทนค่าสัญลักษณ์ในอนุกรมเวลาด้วยวิธีแซ็ค เพื่อให้ทราบถึงลักษณะผิดปรกติต่างๆ ที่เกิดขึ้นกับโปรแกรมประยุกต์เว็บที่ใช้งานอยู่ เพื่อป้องกันและตรวจสอบความเสียหายอันเนื่องจากมาจากค่าผิดปรกติดังกล่าว ซึ่งข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกตินั้น พบปัญหาที่เกี่ยวกับความยุ่งยากในการจัดเก็บข้อมูล ปริมาณข้อมูล ความหลากหลายของการโจมตี ความล้าสมัยของข้อมูลเพื่อนำมาวิเคราะห์ จึงได้มีการสร้างโปรแกรมเพื่อใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บอีกด้วย

ขั้นตอนในการออกแบบและดำเนินการตรวจหาค่าผิดปรกติจะเริ่มจากการแปลงข้อมูลบันทึกการใช้งานบริการเว็บให้อยู่ในรูปข้อมูลอนุกรมเวลา และทำการแทนค่าอนุกรมเวลาดังกล่าวให้เป็นสัญลักษณ์ด้วยวิธีแซ็ค แล้วจึงทำการตรวจหาค่าผิดปรกติจากข้อมูลสัญลักษณ์ โดยการสร้างข้อมูลบันทึกการใช้งานบริการเว็บนั้น ก็จะทำการอ่านลักษณะข้อมูลที่ต้องการสร้างจากแฟ้มโครงแบบเอกซ์เอ็มแอล (XML Configuration) โดยมีรายละเอียดของแนวคิดและขั้นตอนดังนี้

3.1 งานที่เกี่ยวกับการตรวจหาค่าผิดปรกติของเครื่องบริการโปรแกรมประยุกต์เว็บ

การตรวจหาค่าผิดปรกตินี้ จะเป็นการตรวจหาค่าผิดปรกติจากปริมาณการส่งข้อมูลออกจากเครื่องบริการโปรแกรมประยุกต์เว็บ เนื่องจากข้อมูลการบันทึกการใช้บริการเว็บจะมีเพียงข้อมูลปริมาณการส่งข้อมูลออกเท่านั้น (ในกรณีที่เป็นโปรแกรมประยุกต์เว็บทั่วๆไป ซึ่งเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัยก็เป็นหนึ่งในกรณีนี้ด้วย) แต่ในความเป็นจริงแล้วหากมีโปรแกรมประยุกต์เว็บใดๆ ที่สามารถเก็บข้อมูลได้ทั้งปริมาณการส่งและการรับข้อมูลของเครื่องบริการโปรแกรมประยุกต์เว็บ [12] การตรวจหาค่าผิดปรกติยังสามารถทำได้ดีขึ้นอีกด้วย

โปรแกรมที่พัฒนาขึ้นเพื่อตรวจหาค่าผิดปรกตินี้ได้พัฒนาด้วยภาษาจาวา (Java) เพื่อให้สามารถทำงานได้บนทุกแพลตฟอร์ม ตามคุณสมบัติของภาษาจาวา โดยการตรวจหาค่าผิดปรกติมีรายละเอียดที่เกี่ยวข้องดังต่อไปนี้

3.1.1 กระบวนการในการตรวจหาค่าผิดปรกติ

3.1.1.1 การแปลงข้อมูลให้อยู่ในรูปของอนุกรมเวลา

ในการตรวจหาค่าผิดปรกตินั้นมีการใช้วันเวลาที่ทำการร้องขอบริการเว็บ และปริมาณข้อมูลที่ส่งจากผู้ให้บริการเว็บกลับไปยังผู้ร้องขอบริการเว็บ แต่เนื่องจากข้อมูลบันทึกการใช้งานจากเว็บไซต์แต่ละแห่งมีรูปแบบของข้อมูลที่แตกต่างกัน โดยมีข้อมูลที่ใช้ในการวิเคราะห์เหมือนกัน เพื่อลดความยุ่งยากและซับซ้อนในการตรวจหาค่าผิดปรกติ ข้อมูลดังกล่าวจะถูกรวบรวมและแปลงข้อมูลให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์เท่านั้น คือ วันที่ เวลาที่มีรายละเอียดจนถึงวินาที ที่มีการร้องขอใช้บริการ และขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บ โดยหากมีการร้องขอบริการเว็บในช่วงเวลาที่เป็นวินาทีเดียวกันเป็นจำนวนหลายครั้ง การแปลงจะทำการรวมขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บให้ ดังตัวอย่างข้อมูลแสดงในตารางที่ 3.1 และ ความหมายของข้อมูลที่ผ่านการแปลงแสดงในตารางที่ 3.2 โดยโปรแกรมที่แปลงข้อมูลให้มีเพียงข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติสามารถปรับแต่งได้จากแฟ้มโครงแบบ (Configuration File) ที่มีรายละเอียดดังแสดงในตารางที่ 3.3

ตารางที่ 3.1 ตัวอย่างของข้อมูลการร้องขอบริการจากเว็บไซต์ที่ผ่านการแปลงให้มีเพียงข้อมูลที่ใช้ในการวิเคราะห์

```
10/Jan/2007:09:22:41 +0700,18750.0
10/Jan/2007:09:22:42 +0700,182686.0
10/Jan/2007:09:22:43 +0700,94355.0
10/Jan/2007:09:22:44 +0700,290600.0
10/Jan/2007:09:22:45 +0700,52571.0
10/Jan/2007:09:22:46 +0700,211010.0
10/Jan/2007:09:22:47 +0700,283687.0
10/Jan/2007:09:22:48 +0700,195909.0
10/Jan/2007:09:22:49 +0700,490487.0
```

ตารางที่ 3.2 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ที่ผ่านการแปลงให้มีเพียงข้อมูลที่ใช้ในการวิเคราะห์

10/Jan/2007:09:22:41 +0700,18750.0		
	ค่าที่ปรากฏ	**คำอธิบาย**
1	10/Jan/2007 :09:22:41 +0700	วันที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึง วันที่สิบเดือนมกราคม ปีค.ศ. สองพันเจ็ด เวลาเก้านาที ยี่สิบสองนาที สี่สิบเอ็ดนาที ณ เวลาที่การร้องขอบริการเว็บนั้น เครื่องผู้ให้บริการอยู่ในเขตเวลาที่ต้องบวกเพิ่มเจ็ดชั่วโมงจากเวลาปานกลางกรีนิช
2	18750.0	ขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บ มีหน่วยเป็นไบต์

ตารางที่ 3.3 ตัวอย่างแฟ้มโครงแบบของการแปลงข้อมูล
ให้มีเพียงข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติ

```
1  # for cu web aug - dec
2  src_dir=e:/workspace/Thesis/out-cu-aug-dec/
3  files=data-series-sec-paa.txt
4  date_pos=0
5  size_pos=1
6  src_ip_pos=1
7  dst_ip_pos=1
8  log_type=general
9  date_type=DATE_TYPE_STRING
10 delim=,
11 out_dir=./out-cu-aug-dec/
```

โดยในแต่ละบรรทัดของแฟ้มโครงแบบที่ใช้ในการแปลงให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์ มีรายละเอียดดังต่อไปนี้

- บรรทัดที่ 1 แสดงหมายเหตุของแฟ้มโครงแบบนี้ สำหรับข้อมูลใดๆ ที่ขึ้นต้นด้วยเครื่องหมายสี่เหลี่ยม (#)

- บรรทัดที่ 2 - 3 แสดงที่อยู่ของไฟล์ และชื่อของไฟล์ที่เก็บข้อมูลปริมาณการใช้งานเว็บตามลำดับ

- บรรทัดที่ 4 - 7 แสดงตำแหน่งของข้อมูลที่อยู่ในไฟล์ เรียงลำดับดังต่อไปนี้ ตำแหน่งของวันที่เวลาการร้องขอบริการเว็บ ตำแหน่งของขนาดข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บ เลขที่อยู่ไอพีของผู้ให้บริการเว็บ และเลขที่อยู่ไอพีของผู้ร้องขอบริการเว็บ

- บรรทัดที่ 8 แสดงประเภทของไฟล์ร้องขอบริการเว็บจากโปรแกรมบริการเว็บแบบต่างๆ ณ ขณะนี้รองรับอยู่ 3 โปรแกรมด้วยกันคือ จากโปรแกรมอะแพชชี (Apache) โปรแกรมไอไอเอส (IIS) และโปรแกรมที่แปลงให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์

- บรรทัดที่ 9 แสดงประเภทของวันที่ร้องขอบริการเว็บในรูปแบบต่างๆ ณ ขณะนี้รองรับอยู่ 3 ประเภทด้วยกันคือ วันเวลาที่มีรายละเอียดจนถึงวินาทีที่เป็นตัวเลข วันเวลาที่มีรายละเอียดจนถึงมิลลิวินาทีที่เป็นตัวเลข และวันเวลาที่มีรายละเอียดจนถึงวินาทีที่เป็นตัวอักษร

- บรรทัดที่ 10 แสดงอักขระคั่นของข้อมูลแต่ละบรรทัดของไฟล์ร้องขอบริการเว็บ

- บรรทัดที่ 11 แสดงโฟลเดอร์ที่เก็บข้อมูลหลังจากสร้างข้อมูลใหม่ หรือตรวจหาค่าผิดปรกติ

3.1.1.2 การลดมิติข้อมูลด้วยวิธี PAA

จากการวิเคราะห์ข้อมูลบันทึกการใช้งานบริการเว็บ พบว่าการลดมิติด้วยวิธีทำ PAA ของปริมาณข้อมูลนั้น ไม่สามารถใช้จำนวนข้อมูลในการลดมิติได้เพราะความหนาแน่นของข้อมูลในแต่ละช่วงเวลาแตกต่างกันมาก และการระบุเวลาและตำแหน่งที่เกิดค่าผิดปรกติมีโอกาสที่จะเกิดความคลาดเคลื่อนสูง จึงได้ลดมิติของข้อมูลด้วยวิธี PAA ตามช่วงเวลาที่ข้อมูลปรากฏดังต่อไปนี้ 1 ชั่วโมง 2 ชั่วโมง 4 ชั่วโมง 8 ชั่วโมง และ 24 ชั่วโมง ตามลำดับ

3.1.1.3 การกำหนดขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูล

จากการวิเคราะห์ข้อมูลแล้วพบว่า ขนาดของสัญลักษณ์ที่ใช้ควรมีขนาดตั้งแต่ 3 ถึง 10 ตัวอักษร เพื่อให้สามารถตรวจหาค่าผิดปรกติได้

3.1.1.4 การกำหนดจำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติ

จากการวิเคราะห์ข้อมูลปริมาณการส่งข้อมูลจากเครื่องให้บริการเว็บไปยังผู้ร้องขอบริการเว็บจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย เว็บไซต์ของหนังสือพิมพ์ผู้จัดการ และเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการนั้น พบว่ามีรูปแบบของข้อมูลที่คล้ายคลึงกัน ทั้งปริมาณข้อมูลระหว่างกลางวันกับกลางคืนแตกต่างกัน และปริมาณข้อมูลระหว่างวันธรรมดากับวันหยุดแตกต่างกัน

ดังนั้นเพื่อให้การตรวจหาค่าผิดปรกติเป็นไปอย่างครบถ้วนถูกต้องที่สุด การตรวจหาค่าผิดปรกติจะทำตามช่วงเวลาที่ทำ PAA และความยาวของตัวอักษรที่ถูกเลือกในแต่ละช่วงเวลาให้เป็นไปตามรูปแบบจากการวิเคราะห์เป็น 1 วัน และ 1 สัปดาห์ เช่น หากมีการทำ PAA ของช่วงเวลา 1 ชั่วโมงแล้ว จำนวนของตัวอักษรที่ถูกเลือกในการตรวจหาค่าผิดปรกติคือ 24 ตัวอักษร เท่ากับ 24 ชั่วโมง หรือ 1 วัน และ 168 ตัวอักษร เท่ากับ 168 ชั่วโมง หรือ 1 สัปดาห์ หรือหากมีการทำ PAA ของช่วงเวลา 8 ชั่วโมงแล้ว จำนวนของตัวอักษรที่ถูกเลือกในการตรวจหาค่าผิดปรกติคือ 3 ตัวอักษร เท่ากับ 24 ชั่วโมง หรือ 1 วัน และ 21 ตัวอักษร เท่ากับ 168 ชั่วโมง หรือ 1 สัปดาห์ เป็นต้น ดังแสดงในตารางที่ 3.4 หลังจากนั้นโปรแกรมจะเลือกแสดงค่าผิดปรกติที่ใกล้เคียง หรือตรงตามเกณฑ์ที่กำหนด ในช่วงเวลาที่ทำ PAA แตกต่างกัน

ตารางที่ 3.4 จำนวนตัวอักษรที่ถูกเลือกในการตรวจหาค่าผิดปรกติ ที่ขึ้นกับช่วงเวลาที่ทำ PAA

ช่วงเวลา (ชั่วโมง)	จำนวนตัวอักษร	
	รูปแบบ 1 วัน	รูปแบบ 7 วัน
1	24	168
2	12	84
4	6	42
8	3	21
24	3	7

จากตารางที่ 3.4 จะพบว่าในช่วงเวลา 24 ชั่วโมงหรือ 1 วันนั้น แทนที่จะมีการเลือกจำนวนตัวอักษรที่ใช้ในการตรวจหาค่าผิดปรกติให้เป็น 1 แต่กลับมีการเลือกให้มีจำนวนตัวอักษร 3 ตัวอักษรนั้น อันเนื่องมาจากว่าหากเลือกเพียง 1 ตัวอักษร จะทำให้ไม่มีข้อมูลข้างเคียงมาช่วยในการตรวจหาค่าผิดปรกติ ซึ่งจะทำให้การตรวจหาค่าผิดปรกติทำได้ไม่ดี ดังนั้นในกรณีนี้จึงเลือกจำนวนตัวอักษรเป็น 3 ตัวอักษร สำหรับกรณีที่ช่วงเวลาที่ผ่านการทำ PAA เป็น 24 ชั่วโมง หากพบค่าผิดปรกติที่ตรงตามเกณฑ์ในการเลือกค่าผิดปรกติหลายช่วงเวลา ช่วงเวลาที่ลดมิติของข้อมูลที่มีค่ามากที่สุดที่พบจะถูกเลือกว่าเป็นค่าที่ผิดปรกติ

3.1.1.5 ขั้นตอนวิธีที่ใช้ในการตรวจหาค่าผิดปรกติ

- การตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้โดยมีการเก็บข้อมูลระยะห่างของแต่ละลำดับย่อยไว้ในแถวลำดับ (Array)

จากรหัสเทียมที่ใช้ในการตรวจหาค่าผิดปรกติทั้งสองแบบที่กล่าวมา [18] ผู้เขียนพบว่าหากความผิดปรกติที่มีลักษณะที่เหมือนหรือคล้ายกันและมีการเกิดซ้ำมากกว่า 1 ครั้ง กล่าวคือในกรณีที่ลำดับย่อยที่ทำการค้นหานั้น ถึงแม้ว่าจะไม่ได้เป็นส่วนหนึ่งของลำดับย่อยเริ่มต้นที่เริ่มทำการค้นหา แต่มีลักษณะข้อมูลเหมือนลำดับย่อยที่เริ่มทำการค้นหา (ระยะห่างระหว่างลำดับย่อยเท่ากับศูนย์) แม้เพียงครั้งเดียว ก็จะไม่สามารถค้นพบความผิดปรกตินั้นได้ และความผิดปรกติที่เกิดขึ้นรองลงมา กล่าวคือการตรวจหาค่าผิดปรกติที่มีค่าระยะห่างระหว่างลำดับย่อยรองลงมาก็ไม่สามารถหาได้ ทางผู้เขียนจึงได้เสนอขั้นตอนวิธีที่มีความสามารถหาค่าผิดปรกติ แบบเดียวกับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ [18] ทั้งยังสามารถตรวจหาค่าผิดปกติลักษณะที่กล่าวมาดังกล่าว

ได้อีกด้วย ดังนั้นจึงมีการแก้รหัสเทียมเพื่อให้รองรับกับแนวคิดนี้ โดยการเพิ่มการจัดเก็บข้อมูลระยะห่างทั้งหมดที่เกิดขึ้นจากการวนรอบภายนอกไว้ในตัวแปรแถวลำดับ จากการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ ดังแสดงในบรรทัดที่ 16 – 18 และเพื่อทำการวัดประสิทธิภาพในการวนรอบการตรวจหาค่าผิดปรกติจึงได้เพิ่มตัวแปรเพื่อทำการนับจำนวนครั้งในการวนรอบ ดังแสดงในบรรทัดที่ 10 ของตารางที่ 3.5

ตารางที่ 3.5 รหัสเทียมของการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ที่มีการจัดเก็บข้อมูลระยะห่างของแต่ละลำดับย่อยในแถวลำดับ

```
1  Function [dist,loc] = Brute_Force(T,n)
2    Best_so_far_dist = 0
3    Best_so_far_loc = NaN
4    Distance_array = new Array();
5
6    For p = 1 to |T| - n + 1                          // Begin Outer Loop
7      nearest_neighbor_dist = infinity
8      For q = 1 to |T| - n + 1                        // Begin Inner Loop
9        IF |p - q| ≥ n                                // non-self match?
10         Loopcount += 1                              //  Count Loop
11         IF Dist(tp,…,tp+n-1, tq,…,tq+n-1) < nearest_neighbor_dist
12           Nearest_neighbor_dist = Dist(tp,…,tp+n-1, tq,…,tq+n-1)
13         End
14       End                                           // End non-self match test
15     End                                             // End Inner Loop
16     IF nearest_neighbor_dist != infinity
17       distance_array.add(nearest_neighbor_dist)
18     End
19     IF nearest_neighbor_dist > best_so_far_dist
20       Best_so_far_dist = nearest_neighbor_dist
21       Best_so_far_loc = p
22     End
23 End                                                 // End Outer Loop
24 Return [best_so_far_dist, best_so_far_loc, distance_array]
```

3.1.1.6 โครงสร้างข้อมูล (Data Structure) ที่นำมาใช้กับขั้นตอนวิธี

ในการตรวจหาค่าผิดปรกติของงานวิจัยนี้มีการใช้โครงสร้างข้อมูลในการตรวจหาค่าผิดปรกติ 2 แบบด้วยกัน คือ

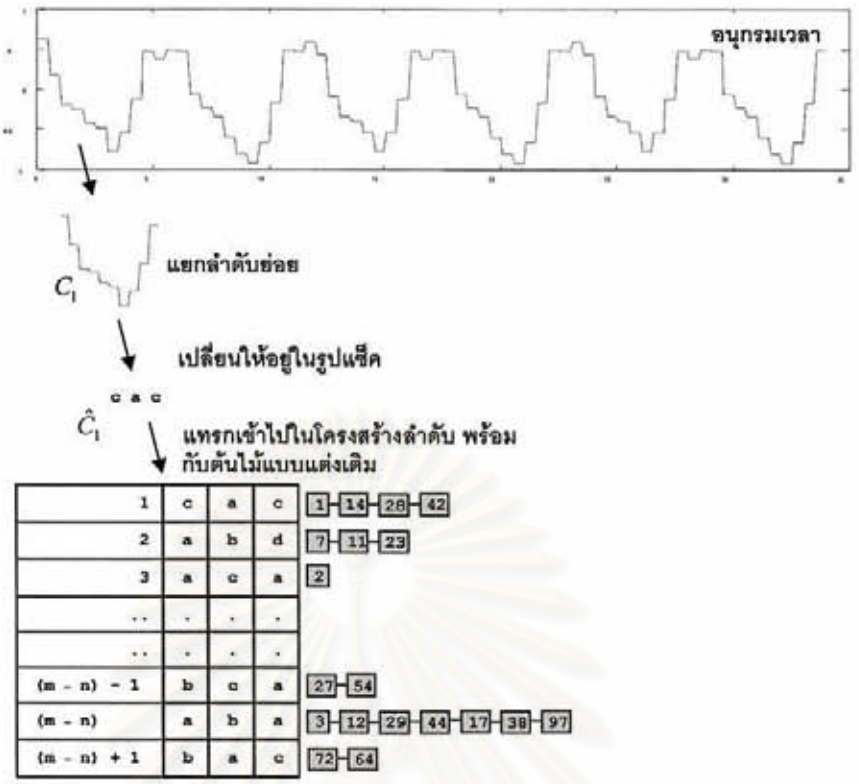
- โครงสร้างข้อมูลสายอักขระ [18]

เป็นโครงสร้างข้อมูลที่มีการแปลงอนุกรมเวลาให้อยู่ในรูปของสายอักขระเรียงต่อกัน โดยใช้วิธีแซ็ค

- โครงสร้างต้นไม้แบบแต่งเติมรวมกับโครงสร้างข้อมูลแถวลำดับ (Array-based Augmented Tree)

เป็นโครงสร้างข้อมูลที่เสนอโดยผู้เขียน โดยมีรายละเอียดดังต่อไปนี้ โครงสร้างนี้มาจากการแปลงอนุกรมเวลาให้อยู่ในรูปสายอักขระทั้งหมด และมีการปรับแต่งให้มีโครงสร้างข้อมูลแบบเดียวให้เป็นโครงสร้างข้อมูลแถวลำดับ โดยข้อมูลในแถวลำดับแต่ละตัวจะทำการเก็บสายอักขระที่เกิดขึ้นในแต่ละลำดับย่อย และตำแหน่งที่สายอักขระนั้นเกิดขึ้นในอนุกรมเวลาทั้งหมด ซึ่งโครงสร้างชนิดนี้ สามารถใช้ได้กับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ [18] การตรวจหาค่าผิดปรกติจากวิธีศึกษาสำนึก [18] การตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ที่เพิ่มการเก็บข้อมูลระยะห่างของแต่ละลำดับย่อยไว้ในแถวลำดับ ดังแสดงในรูปที่ 3.1 ข้อมูลตัวแรกของแถวลำดับที่ 1 จะมีสายอักขระที่เกิดขึ้นคือ c a c และมีตำแหน่งของสายอักขระที่มีค่าเดียวกับสายอักขระที่เกิดที่ตำแหน่งที่ 1 นี้ ทั้งหมด 4 ตำแหน่ง คือ ตำแหน่งที่ 1 14 28 และ 42 ถ้ากำหนดความยาวของสายอักขระให้เท่ากับ m ที่ลำดับย่อยมีขนาดเท่ากับ n แล้ว จำนวนของลำดับย่อยที่สามารถเกิดขึ้นได้จะมีค่าน้อยกว่า หรือเท่ากับ $(m - n) + 1$

โดยการตรวจหาค่าผิดปรกติจากโครงสร้างข้อมูลชนิดนี้ จะใช้การวนซ้ำตอนนอกและตอนในกับโครงสร้างข้อมูลแถวลำดับชุดเดียวกัน แต่จะใช้จำนวนการวนรอบในการตรวจหาค่าผิดปรกติน้อยกว่าหรือเท่ากับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ อันเนื่องมาจากข้อมูลสายอักขระที่แถวลำดับใดๆ มีตำแหน่งที่เกิดข้อมูลสายอักขระมากกว่า 1 ตำแหน่ง ก็จะทำให้จำนวนข้อมูลในการวนรอบลดลงไปด้วย ซึ่งเมื่อนำโครงสร้างข้อมูลชนิดนี้มาใช้ร่วมกับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ โดยมีการเก็บข้อมูลระยะห่างของแต่ละลำดับย่อยไว้ในแถวลำดับแล้ว นอกจากสามารถตรวจหาค่าผิดปรกติที่มีลักษณะข้อมูลที่ไม่ซ้ำได้แล้ว ยังสามารถตรวจหาค่าผิดปรกติที่มีลักษณะข้อมูลเหมือนลำดับย่อยที่เริ่มทำการค้นหา (ระยะห่างระหว่างลำดับย่อยเท่ากับศูนย์) ที่มีการเกิดมากกว่าหนึ่งครั้ง และความผิดปรกติที่เกิดขึ้นรองลงมาได้อีกด้วย



รูปที่ 3.1 โครงสร้างต้นไม้แบบแต่งเติมรวมกับโครงสร้างข้อมูลแถวลำดับ

3.1.1.7 อันดับของค่าผิดปรกติที่เกิดขึ้น

ในการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้เกิดจากการคำนวณระยะห่างในแต่ละลำดับย่อย โดยระยะห่างที่มีค่ามากที่สุดจะถือว่าเป็นค่าผิดปรกติอันดับที่ 1 ระยะห่างที่มีค่ารองลงมาจะถือว่าเป็นค่าผิดปรกติอันดับที่ 2 เป็นเช่นนี้ไปเรื่อย

3.1.1.8 เกณฑ์ในการเลือกค่าผิดปรกติ

เนื่องจากในการตรวจหาค่าผิดปรกตินั้นพบค่าผิดปรกติหลายอันดับ และในแต่ละอันดับก็จะมีค่าผิดปรกติหลายตำแหน่งเช่นกัน ดังนั้นจึงมีเกณฑ์ในการเลือกค่าผิดปรกติที่เหมาะสมที่สุด โดยให้จำนวนค่าผิดปรกติที่เกิดขึ้นในแต่ละอันดับ และจำนวนอันดับของค่าผิดปรกติที่เกิดขึ้นไม่เกินค่าที่กำหนด ซึ่งจากการทดลองควรมีค่าทั้งสองเป็น 5

3.1.1.9 การจำลองข้อมูลแบบต่อเนื่อง

เป็นการจำลองข้อมูลแบบต่อเนื่องจากข้อมูลแบบไม่ต่อเนื่อง โดยเป็นการเลือกข้อมูลที่เป็นข้อมูลหลักตามช่วงเวลาที่กำหนดทุกช่วงเวลามาตรวจหาค่าผิดปรกติ ซึ่งข้อมูลหลักที่เลือกจะต้องเป็นข้อมูลที่มีระยะเวลามากสุดที่มีรูปแบบชัดเจน และจากการวิเคราะห์รูปแบบของข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัย เว็บไซต์ของ

หนังสือพิมพ์ผู้จัดการ และเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ พบว่าช่วงเวลา 7 วัน หรือ 1 สัปดาห์ เป็นช่วงเวลาที่มีรูปแบบของข้อมูลที่มีระยะเวลาที่มากที่สุดที่มีรูปแบบที่ชัดเจน และมีการกำหนดระยะเวลาของข้อมูลหลักที่ใช้ในการตรวจหาค่าผิดปรกติเพิ่มเติมเป็น 15 วัน และ 30 วันตามลำดับ โดยในครั้งแรกจะเลือกช่วงเวลาที่นำมาตรวจหาค่าผิดปรกติก่อน จากนั้นจึงเลือกข้อมูลจากอนุกรมเวลาตามลำดับของข้อมูล และตามช่วงเวลาที่เลือก โดยข้อมูลช่วงเวลาที่ถูกเลือกเป็นข้อมูลหลัก จากนั้นเลือกข้อมูลวันถัดไปจากช่วงเวลาที่เลือกหนึ่งวันมาเรียงต่อกับข้อมูลที่เลือกดังกล่าว แปลงให้อยู่ในรูปสัญลักษณ์และตรวจหาค่าผิดปรกติ โดยการตรวจหาค่าผิดปรกติครั้งต่อไป จะคัดข้อมูลที่เป็นข้อมูลเก่าออกจากข้อมูลหลักหนึ่งวัน และเลือกข้อมูลวันถัดไปจากช่วงเวลาที่เลือกหนึ่งวันมาเรียงต่อกับข้อมูลหลักใหม่มาตรวจหาค่าผิดปรกติ ทำเช่นนี้ไปเรื่อยๆ จนข้อมูลหมด จากนั้นให้เลือกช่วงเวลาถัดไปตามที่กำหนด โดยใช้ข้อมูลอนุกรมเวลาและการตรวจหาค่าผิดปรกติเช่นเดียวกับช่วงเวลาที่ถูกเลือกครั้งแรก ทำเช่นนี้ไปเรื่อยๆ จนช่วงเวลาที่ถูกเลือกหมด เมื่อได้ผลการตรวจหาค่าผิดปรกติทั้งหมดจึงนำไปเปรียบเทียบกับการตรวจหาค่าผิดปรกติแบบไม่ต่อเนื่อง ซึ่งหลักเกณฑ์ต่างๆในการแปลงให้อยู่ในรูปสัญลักษณ์ และการตรวจหาค่าผิดปรกติเป็นเช่นเดียวกับข้อมูลแบบไม่ต่อเนื่อง

3.1.1.10 แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ

เนื่องจากข้อมูลบันทึกการใช้งานจากเว็บไซต์แต่ละแห่งมีรูปแบบของข้อมูลที่แตกต่างกัน จึงมีการสร้างแฟ้มโครงแบบเพื่อให้สามารถปรับแต่งค่าพารามิเตอร์ต่างๆที่ใช้ในการตรวจหาค่าผิดปรกติได้ ดังแสดงรายละเอียดในตารางที่ 3.6

ตารางที่ 3.6 ตัวอย่างแฟ้มโครงแบบของการตรวจหาค่าผิดปรกติ

```
1  # for cu web aug - dec
2  src_dir=e:/workspace/Thesis/out-cu-aug-dec/
3  files=data-series-sec-paa.txt
4  Date_pos=0
5  Size_pos=1
6  src_ip_pos=1
7  dst_ip_pos=1
8  log_type=general
9  date_type=DATE_TYPE_STRING
10 delim=,
11 out_dir=./out-cu-aug-dec/
12 sec_counts=3600,7200,14400,28800,86400
13 sax_counts=3-10
14 #sax_window_length is in source code
15 sax_window_length=3,4,5,6,7,8,9,10
```

ตารางที่ 3.6 (ต่อ) ตัวอย่างแฟ้มโครงแบบของการตรวจหาค่าผิดปรกติ

```
16  Streaming_day=7,15,30
17  sec_streaming_counts=3600,7200,14400,28800,86400
18  sax_streaming_counts=3-10
19  #sax_streaming_window_length is in source code
20  sax_streaming_window_length=24,12,6,3
21  knn=1000
22  Streaming_flag=yes
23  non_streaming_flag=yes
24  discord_best_count=3
25  discord_best_order=5
```

โดยในแต่ละบรรทัดของแฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ มีรายละเอียดดังต่อไปนี้

- บรรทัดที่ 1 แสดงหมายเหตุของแฟ้มโครงแบบนี้ สำหรับข้อมูลใดๆ ที่ขึ้นต้นด้วยเครื่องหมายสี่เหลี่ยม (#)

- บรรทัดที่ 2 - 3 แสดงที่อยู่ของไฟล์ และชื่อของไฟล์ที่เก็บข้อมูลปริมาณการใช้งานเว็บตามลำดับ

- บรรทัดที่ 4 - 7 แสดงตำแหน่งของข้อมูลที่อยู่ในไฟล์ เรียงลำดับดังต่อไปนี้ ตำแหน่งของวันที่เวลาการร้องขอบริการเว็บ ตำแหน่งของขนาดข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บ เลขที่อยู่ไอพีของผู้ให้บริการเว็บ และเลขที่อยู่ไอพีของผู้ร้องขอบริการเว็บ

- บรรทัดที่ 8 แสดงประเภทของไฟล์ร้องขอบริการเว็บจากโปรแกรมบริการเว็บแบบต่างๆ ณ ขณะนี้รองรับอยู่ 3 โปรแกรมด้วยกันคือ จากโปรแกรมอะแพชชี โปรแกรมไอไอเอส และโปรแกรมที่แปลงให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์

- บรรทัดที่ 9 แสดงประเภทของวันที่ร้องขอบริการเว็บในรูปแบบต่างๆ ณ ขณะนี้รองรับอยู่ 3 ประเภทด้วยกันคือ วันเวลาที่มีรายละเอียดจนถึงวินาทีที่เป็นตัวเลข วันเวลาที่มีรายละเอียดจนถึงมิลลิวินาทีที่เป็นตัวเลข และวันเวลาที่มีรายละเอียดจนถึงวินาทีที่เป็นตัวอักษร

- บรรทัดที่ 10 แสดงอักขระคั่นของข้อมูลแต่ละบรรทัดของไฟล์ร้องขอบริการเว็บ

- บรรทัดที่ 11 แสดงโฟลเดอร์ที่เก็บข้อมูลหลังจากสร้างข้อมูลใหม่ หรือตรวจหาค่าผิดปรกติ

- บรรทัดที่ 12 และ 17 แสดงช่วงเวลาที่ใช้ทำ PAA มีหน่วยเป็นวินาที สำหรับข้อมูลแบบต่อเนื่อง และข้อมูลแบบไม่ต่อเนื่องตามลำดับ โดยตัวเลข 3600 หมายถึง 3600 วินาที คือ 1 ชั่วโมง เป็นต้น

- บรรทัดที่ 13 และ 18 แสดงจำนวนตัวอักษรที่ใช้ในการแทนค่าสัญลักษณ์ของข้อมูลแบบไม่ต่อเนื่อง และข้อมูลแบบต่อเนื่องตามลำดับ

- บรรทัดที่ 15 และ 20 แสดงจำนวนตัวอักษรที่ใช้ในการตรวจหาค่าผิดปรกติของข้อมูลแบบต่อเนื่อง และข้อมูลแบบไม่ต่อเนื่องตามลำดับ

- บรรทัดที่ 16 แสดงจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง

- บรรทัดที่ 21 แสดงค่าที่มากที่สุดของจำนวนลำดับค่าผิดปรกติที่ต้องการค้นหา

- บรรทัดที่ 22 - 23 แสดงการแจ้งให้โปรแกรมทราบว่าให้ตรวจหาค่าผิดปรกติจากข้อมูลแบบต่อเนื่องและไม่ต่อเนื่องตามลำดับ

- บรรทัดที่ 24 แสดงค่าผิดปรกติที่มีจำนวนข้อมูลในแต่ละอันดับไม่เกินค่าที่กำหนด

- บรรทัดที่ 25 แสดงค่าผิดปรกติที่มีอันดับไม่เกินค่าที่กำหนด

3.1.2 กระบวนการในการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบไม่ต่อเนื่อง

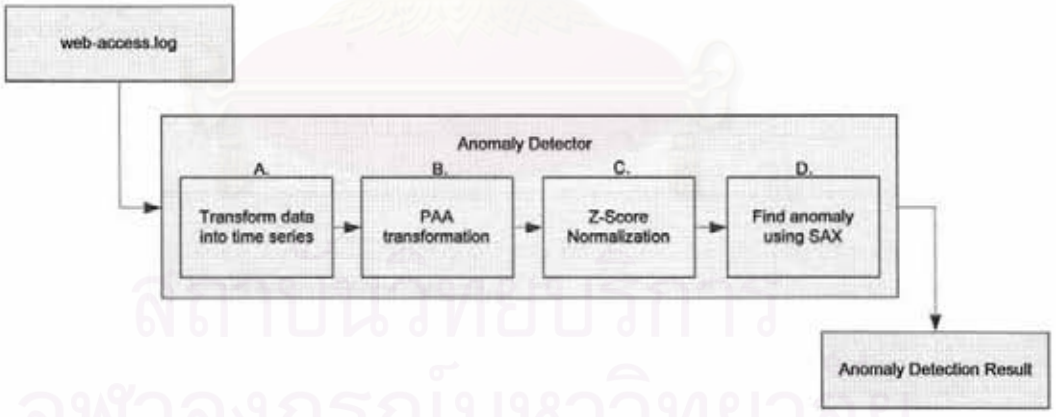
ในการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่อง จะเป็นการนำขั้นตอนวิธีที่ใช้ในการตรวจหาค่าผิดปรกติวิธีที่ผู้เขียนดัดแปลงในวิธีทุกความเป็นไปได้ มาใช้ร่วมกับโครงสร้างข้อมูลที่ผู้เขียนดัดแปลง และโครงสร้างข้อมูลสายอักขระ โดยในการตรวจหาค่าผิดปรกติจะมีกระบวนการดังต่อไปนี้ หรือดังแสดงในรูปที่ 3.2

A. แปลงข้อมูลให้อยู่ในรูปของข้อมูลอนุกรมเวลา

B. ลดมิติของข้อมูลด้วยวิธี PAA

C. ทำข้อมูลให้เป็นบรรทัดฐานด้วยวิธี Z-Score

D. ตรวจหาค่าผิดปรกติจากข้อมูลที่แปลงให้อยู่ในรูปของสัญลักษณ์โดยใช้วิธีแซ็ค



รูปที่ 3.2 กระบวนการในการตรวจหาค่าผิดปรกติจากปริมาณการใช้งานเว็บโดยใช้ข้อมูลแบบไม่ต่อเนื่อง

3.1.3 กระบวนการในการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบต่อเนื่อง

ในความเป็นจริงข้อมูลมีการรับส่งกันอย่างต่อเนื่องตลอดเวลา ในการตรวจหาค่าผิดปรกติหากต้องตรวจหาค่าจากข้อมูลเก่าทั้งหมดรวมกับข้อมูลใหม่นั้น เป็นไปได้ยาก หรือเป็นไปไม่ได้เลย ทั้งนี้เนื่องมาจากการที่ทรัพยากรมีอยู่อย่างจำกัดและไม่เพียงพอ เช่น หน่วยความจำ ความเร็วในการประมวลผล หรือขนาดความจุของฮาร์ดดิสค์ เป็นต้น ซึ่งจะมีการเลียนแบบข้อมูลแบบต่อเนื่องให้เป็นข้อมูลแบบไม่ต่อเนื่อง โดยทำการเลือกข้อมูลหนึ่งสัปดาห์เป็นข้อมูลหลักทำการแปลงให้อยู่ในรูปสัญลักษณ์ จากนั้นทำการเลือกข้อมูลวันใหม่ที่จากข้อมูลแบบต่อเนื่องในแต่ละวันมาทำการแปลงให้อยู่ในรูปสัญลักษณ์ แล้วนำข้อมูลสัญลักษณ์ทั้งสองชุดมาเรียงต่อกัน หลังจากนั้นจึงทำการตรวจหาค่าผิดปรกติ แล้วนำมาเทียบกับการตรวจหาค่าผิดปรกติโดยใช้ข้อมูลแบบไม่ต่อเนื่องว่าสามารถตรวจหาค่าผิดปรกติได้ในช่วงวันเดียวกันหรือไม่ โดยในการตรวจหาค่าผิดปรกติจะมีขั้นตอนดังต่อไปนี้ หรือดังแสดงในรูปที่ 3.3

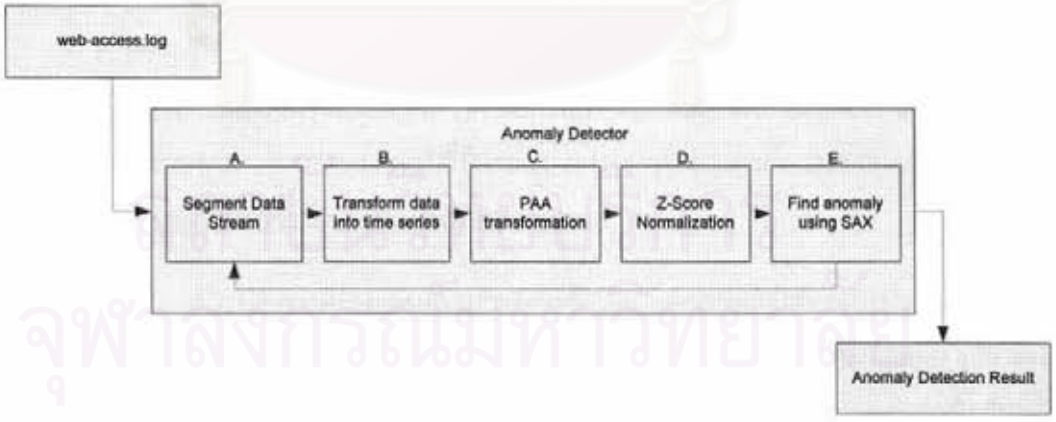
A. เลือกข้อมูลตามช่วงเวลาที่กำหนด เช่น หนึ่งสัปดาห์ หรือสิบห้าวัน หรือสามสิบวัน ที่เรียงต่อกัน แล้วเลือกข้อมูลวันถัดไปหลังช่วงเวลาที่เลือก มาเรียงต่อกับข้อมูลในช่วงเวลาที่ถูกเลือกขึ้นมา

B. แปลงข้อมูลดิบให้อยู่ในรูปของข้อมูลอนุกรมเวลา

C. ลดมิติของข้อมูลด้วยวิธี PAA

D. ทำข้อมูลให้เป็นบรรทัดฐานด้วยวิธี Z-Score

E. ตรวจหาค่าผิดปรกติจากข้อมูลที่แปลงให้อยู่ในรูปของสัญลักษณ์โดยใช้วิธีแซ็ค แล้วทำการเลือกข้อมูลตามช่วงเวลาในข้อ A ที่เป็นช่วงเวลาถัดไปอีกหนึ่งวัน แล้วเริ่มทำขั้นตอนต่างๆ ตามลำดับในการตรวจหาค่าผิดปรกติ ทำเช่นนี้ไปเรื่อยๆ จนกว่าข้อมูลจะหมด



รูปที่ 3.3 กระบวนการในการตรวจหาค่าผิดปรกติจากปริมาณการใช้งานเว็บโดยใช้ข้อมูลแบบต่อเนื่อง

3.2 โปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ

3.2.1 การวิเคราะห์การสร้างข้อมูล

3.2.1.1 การสร้างข้อมูลที่มีรูปแบบต่างๆ

โปรแกรมนี้มีจุดประสงค์เพื่อทำให้มีข้อมูลที่ใช้ในงานวิจัยได้ง่ายขึ้น จากเดิมที่ต้องเก็บข้อมูลจากเครื่องบริการเว็บจริง หรือต้องมีการจำลองการร้องขอบริการมายังเว็บมาเป็นข้อมูลดิบที่พร้อมจะใช้งานทันที โดยโปรแกรมนี้สามารถจำลองลักษณะข้อมูลได้หลายรูปแบบ รวมไปถึงรูปแบบที่เป็นฤดูกาล ตามการเลือกค่าที่อยู่ในแฟ้มโครงแบบที่อยู่ในรูปแบบเอ็กซ์เอ็มแอล สามารถสร้างค่าผิดปรกติอันเนื่องมาจากการใช้งานมากเกินไปหรือการใช้งานน้อยเกินไปได้ ตัวอย่างเช่น ได้มีการนำข้อมูลจริงซึ่งเป็นเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัยมาแสดงให้อยู่ในรูปกราฟแล้วพบว่าการเรียกใช้งานบริการโปรแกรมประยุกต์เว็บมีลักษณะข้อมูลเป็นฤดูกาล ดังนั้นการเลียนแบบการบันทึกการใช้งานของเว็บก็จะเป็นการสร้างข้อมูลเลียนแบบการใช้งานซึ่งมีลักษณะข้อมูลเป็นฤดูกาลด้วย เช่น ช่วงกลางวันจะมีการเรียกใช้บริการเว็บไซต์เป็นจำนวนมาก แต่ช่วงกลางคืนจะมีการเรียกใช้บริการเว็บไซต์น้อยลง ซึ่งข้อมูลจริงซึ่งเป็นเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัยนั้น เป็นข้อมูลที่เป็นรูปแบบประเภทหนึ่งเท่านั้น หากต้องการข้อมูลที่มีรูปแบบอื่นๆ จะไม่สามารถทำได้ หากไม่มีโปรแกรมนี้ เป็นต้น

3.2.1.2 การสุ่มค่าตัวเลข

ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บจะมีการสร้างข้อมูลที่ต้องการค่าสุ่ม อันได้แก่ วันที่และเวลาที่มีการร้องขอบริการเว็บ ปริมาณข้อมูลที่ผู้ให้บริการส่งกลับมาให้ผู้ร้องขอบริการ จำนวนครั้งในการร้องขอบริการ หมายเลขพอร์ต (Port) ที่ทำการติดต่อระหว่างผู้ให้บริการกับผู้ร้องขอบริการเว็บ เลขที่อยู่ไอพี (IP Address) ของทั้งผู้ให้บริการและผู้ร้องขอบริการเว็บ โดยใช้ตัวสร้างเลขสุ่ม (Random Number Generator) ที่เป็นฟังก์ชันของภาษาจาวา (Java) ที่มีการแจกแจงเอกรูป (Uniform Distribution) กล่าวคือตัวแปรสุ่มทุกตัวจะมีความน่าจะเป็นเท่ากันหมด

3.2.1.3 การสร้างแผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานบริการเว็บ

หลังจากที่ทำการสร้างข้อมูลด้วยโปรแกรม Web Access Generator แล้ว จะมีการสร้างแผนภูมิกราฟปริมาณการใช้งานจากโปรแกรม Chart Generator เพื่อทำการตรวจสอบว่าโปรแกรมที่ใช้สร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บนั้นทำงานได้อย่างถูกต้อง ซึ่งโปรแกรม Chart Generator นี้ไม่ได้ใช้สำหรับการสร้างแผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานที่สร้างโดยโปรแกรม Web Access Generator เท่านั้น แต่ยัง

ใช้สำหรับสร้างแผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งานจากเว็บไซต์ต่างๆ อีกด้วย และเพื่ออำนวยความสะดวกในการปรับแต่งรูปแบบของข้อมูล ซึ่งการปรับเปลี่ยนรูปแบบของข้อมูลกำหนดได้ตามแฟ้มโครงแบบเอกซ์เอ็มแอล และมีการใช้แฟ้มโครงแบบเอกซ์เอ็มแอลเดียวกับการตรวจหาค่าผิดปรกติ ดังแสดงในตารางที่ 3.7 ซึ่งจะทำการสร้างแผนภูมิกราฟจากข้อมูลที่ผ่านการลดมิติด้วยวิธี PAA และทำให้เป็นบรรทัดฐานด้วยวิธี Z-Score แล้ว ตามค่าที่กำหนดไว้ในบรรทัดที่ 12 หรือ sec_counts

3.2.1.4 แฟ้มโครงแบบที่ใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บ

เนื่องจากมีการสร้างข้อมูลบันทึกการใช้งานที่มีรูปแบบของข้อมูลที่หลากหลาย จึงมีการสร้างแฟ้มโครงแบบเพื่อให้สามารถปรับแต่งค่าพารามิเตอร์ต่างๆที่ใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บได้ ดังแสดงรายละเอียดในตารางที่ 3.7

ตารางที่ 3.7 แฟ้มโครงแบบเอกซ์เอ็มแอลที่ใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บ

```
1  <?xml version="1.0"?>
2  <data_generator>
3      <template>
4          <web>
5              <folder value="d:/workspace/cputil/log/web-access-gen.log"/>
6          </web>
7      </template>
8      <data>
9          <cp_web value="web">
10             <year value="2006">
11                 <month value="3-4">
12                     <week value="1-6">
13                         <day value="1-5">
14                             <hour value="0-6">
15                                 <traffic value="8500-9000">
16                                     <size value="1400-1500"/>
17                                 </traffic>
18                                 <dst_ip_range value="202.3.145.3-202.3.145.3">
19                                     <port_range value="80"/>
20                                 </dst_ip_range>
21                                 <ip_range value="192.168.1.0-192.168.1.254">
22                                     <port_range value="1024-5000"/>
23                                 </ip_range>
24                             </hour>
25                             <hour value="6-0">
26                                 <traffic value="8500-9500">
27                                     <size value="1500-1700"/>
28                                 </traffic>
```

ตารางที่ 3.7 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลที่ใช้ในการสร้างข้อมูลบันทึกการใช้งานบริการเว็บ

```
                                <dst_ip_range value="202.3.145.3-202.3.145.3">
                                    <port_range value="80"/>
                                </dst_ip_range>
                                <ip_range value="192.168.1.0-192.168.1.254">
                                    <port_range value="1024-5000"/>
                                </ip_range>
                            </hour>
                        </day>
                        <day value="6-7">
                            <hour value="0-6">
                                <traffic value="9500-10000">
                                    <size value="1400-1500"/>
                                </traffic>
                                <dst_ip_range value="202.3.145.3-202.3.145.3">
                                    <port_range value="80"/>
                                </dst_ip_range>
                                <ip_range value="192.168.1.0-192.168.1.254">
                                    <port_range value="1024-5000"/>
                                </ip_range>
                            </hour>
                            <hour value="6-0">
                                <traffic value="10000-11000">
                                    <size value="1500-1700"/>
                                </traffic>
                                <dst_ip_range value="202.3.145.3-202.3.145.3">
                                    <port_range value="80"/>
                                </dst_ip_range>
                                <ip_range value="192.168.1.0-192.168.1.254">
                                    <port_range value="1024-5000"/>
                                </ip_range>
                            </hour>
                        </day>
                    </week>
                </month>
            </year>
        </cp_web>
    </data>
</data_generator>
```

โดยในแต่ละบรรทัดของแฟ้มโครงแบบเอกซ์เอ็มแอลมีรายละเอียดดังต่อไปนี้

- บรรทัดที่ 5 แสดงชื่อและที่อยู่ของไฟล์ที่จะทำการเก็บข้อมูลบันทึกการใช้งานเว็บที่ถูกสร้างขึ้น

- บรรทัดที่ 10 - 14 แสดงช่วงเวลาในการสร้างข้อมูลที่มีการร้องขอการบริการเว็บ โดยมีช่วงเวลาที่เป็นหน่วยของปี ช่วงเวลาที่เป็นหน่วยของเดือน ช่วงเวลาที่เป็นหน่วยของสัปดาห์ ช่วงเวลาที่เป็นหน่วยของวัน และช่วงเวลาที่เป็นหน่วยของชั่วโมงตามลำดับ

- บรรทัดที่ 15 - 17 แสดงการสุ่มจำนวนครั้งของการร้องขอการบริการเว็บตามตัวเลขที่กำหนดไว้ (บรรทัดที่ 15) โดยในแต่ละครั้งของการร้องขอบริการให้มีการสุ่มปริมาณการส่งข้อมูลตามตัวเลขที่กำหนดไว้ (บรรทัดที่ 16)

- บรรทัดที่ 18 - 20 แสดงการสุ่มเลขที่อยู่ไอพีของการเครื่องที่ให้บริการเว็บตามตัวเลขที่อยู่ไอพีที่กำหนดไว้ (บรรทัดที่ 18) โดยในแต่ละครั้งของการร้องขอบริการให้มีการสุ่มหมายเลขของพอร์ตที่ทำการรับข้อมูลที่ร้องขอตามตัวเลขที่กำหนดไว้ (บรรทัดที่ 19)

- บรรทัดที่ 21 - 23 แสดงการสุ่มเลขที่อยู่ไอพีของการเครื่องที่ร้องขอบริการเว็บตามตัวเลขที่อยู่ไอพีที่กำหนดไว้ (บรรทัดที่ 18) โดยในแต่ละครั้งของการร้องขอบริการให้มีการสุ่มหมายเลขของพอร์ตที่ทำการร้องขอบริการเว็บตามตัวเลขที่กำหนดไว้ (บรรทัดที่ 22)

- บรรทัดที่ 25 - 35 ถ้าหากในช่วงเวลาของวันดังกล่าวมีลักษณะของข้อมูลช่วงเวลาที่แตกต่างกัน สามารถใส่รายละเอียดของข้อมูลในช่วงเวลาดังกล่าวได้เหมือนกับข้อมูลที่แสดงในบรรทัดที่ 15 - 23

- บรรทัดที่ 37 - 60 ถ้าหากในช่วงเวลาสัปดาห์มีลักษณะของข้อมูลในแต่ละวันที่แตกต่างกัน สามารถทำได้โดยใส่ช่วงของวัน ดังแสดงในบรรทัดที่ 37 - 60 และสามารถใส่รายละเอียดของข้อมูลในช่วงของวันดังกล่าวได้เหมือนกับข้อมูลที่แสดงในบรรทัดที่ 13 - 36

หากต้องการให้มีการสร้างลักษณะข้อมูลบันทึกการใช้งานบริการเว็บที่ผิดปรกติสามารถทำได้โดยการสร้างส่วนย่อยใหม่ที่อยู่ภายใต้ระดับของส่วนย่อย <data> ซึ่งการสร้างส่วนย่อยดังกล่าว หากมีข้อมูลที่มีลักษณะที่ได้ทำการสร้างไปแล้วจะถูกเพิ่มเติมด้วยส่วนย่อยที่ถูกสร้างใหม่ ดังแสดงในตารางที่ 3.8

ตารางที่ 3.8 ส่วนเพิ่มเติมในแฟ้มโครงแบบเอกซ์เอ็มแอล
ที่ใช้ในการสร้างลักษณะข้อมูลบันทึกการใช้งานบริการเว็บที่ผิดปรกติ

```
1  <data>
2    ....
3    <anomaly_web1 value="web">
4        <year value="2006">
5            <month value="3">
6                <week value="3">
7                    <day value="5-7">
8                        <hour value="0-0">
9                            <traffic value="10000-22500">
10                               <size value="1000-2000"/>
```

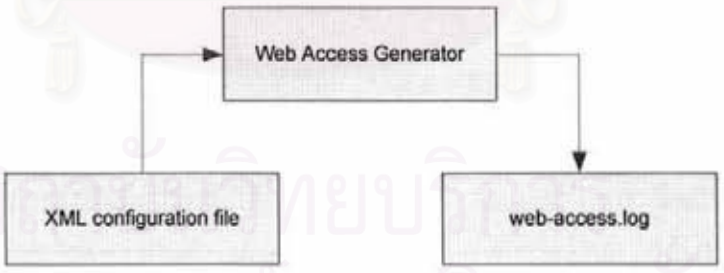
ตารางที่ 3.8 (ต่อ) ส่วนเพิ่มเติมในแฟ้มโครงแบบเอกซ์เอ็มแอล
ที่ใช้ในการสร้างลักษณะข้อมูลบันทึกการใช้งานบริการเว็บที่ผิดปรกติ

```
                        </traffic>
                        <dst_ip_range value="202.3.145.3-202.3.145.3">
                            <port_range valuse="80"/>
                        </dst_ip_range>
                        <ip_range value="192.168.1.0-192.168.1.254">
                            <port_range value="1024-5000"/>
                        </ip_range>
                    </hour>
                </day>
            </week>
        </month>
    </year>
</anomaly_web1>
```

3.2.2 กระบวนการในการสร้างข้อมูล

โปรแกรมนี้พัฒนาขึ้นด้วยภาษาจาวา (Java) เพื่อให้สามารถทำงานได้บนทุกแพลตฟอร์ม ตามคุณสมบัติของภาษาจาวา โดยตัวโปรแกรมนี้จะแบ่งการทำงานออกเป็น 3 ส่วน ดังแสดงในรูปที่ 3.4 คือ

- ส่วนจัดการการอ่านค่าจากแฟ้มโครงแบบที่อยู่ในรูปแบบเอกซ์เอ็มแอล (XML Configuration File)
- ส่วนจัดการการสร้างข้อมูลการบันทึกการใช้งานของเว็บ (Web Access Generator)
- ส่วนการสร้างบันทึกการใช้งานของเว็บ (web-access.log)



รูปที่ 3.4 ส่วนประกอบต่างๆในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ

3.3 ข้อมูลที่ใช้ในการวิจัย

3.3.1 ข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติ

ข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกตินั้น จะใช้ข้อมูลบันทึกการใช้งานเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย เว็บไซต์ของหนังสือพิมพ์ผู้จัดการ เว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ และข้อมูลที่สร้างขึ้น เพื่อทำการตรวจหาค่าผิดปรกติ และเปรียบเทียบผลลัพธ์ที่ได้กับวิธีทุกความเป็นไปได้ ซึ่งลักษณะของข้อมูลของโปรแกรมประยุกต์เว็บที่จะนำมาใช้วิเคราะห์ มีอยู่ 2 ลักษณะ ด้วยกัน คือ

- ข้อมูลแบบไม่ต่อเนื่อง

คือ ข้อมูลทั้งหมดที่พร้อมจะใช้ประมวลผล โดยอาจมาจากข้อมูลแบบต่อเนื่องก็ได้ ตัวอย่างเช่น ข้อมูลกระจายเสียงวิทยุช่วงเวลาใดช่วงเวลาหนึ่ง ข้อมูลการใช้งานโปรแกรมประยุกต์เว็บช่วงเวลาใดช่วงเวลาหนึ่งเป็นต้น โดยในงานวิจัยนี้จะเป็นการอ่านข้อมูลที่มีอยู่ทั้งหมดของโปรแกรมประยุกต์เว็บที่ต้องการตรวจหาค่าผิดปรกติ

- ข้อมูลแบบต่อเนื่อง [29]

คือ ข้อมูลที่ใช้เทคนิคในการส่งข้อมูลอย่างต่อเนื่องตลอดเวลา ซึ่งในขณะที่ทำการประมวลผลข้อมูลนั้นไม่จำเป็นที่จะต้องให้การส่งข้อมูลทั้งหมดเสร็จสิ้นก่อน เนื่องจากในความเป็นจริงแล้วข้อมูลบางประเภทนั้น ไม่มีจำนวนข้อมูลทั้งหมดที่แน่นอนเนื่องจากมีข้อมูลเพิ่มขึ้นตลอดเวลา ตัวอย่างเช่น ข้อมูลการกระจายเสียงวิทยุ ข้อมูลการใช้งานโปรแกรมประยุกต์เว็บ เป็นต้น โดยในงานวิจัยนี้จะเป็นการแบ่งข้อมูลที่มีขนาดใหญ่ที่มีอยู่ ออกเป็นช่วงเวลาหลายๆ ช่วงเวลา แล้วทำการตรวจหาค่าผิดปรกติของข้อมูลในแต่ละช่วงเวลา ที่มีค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง แต่เนื่องจากข้อมูลแบบต่อเนื่องยังพบปัญหาต่างๆ อีกเป็นจำนวนมาก ดังนั้นในงานวิจัยนี้จึงเป็นการศึกษาความเป็นไปได้เท่านั้น

ในการตรวจหาค่าผิดปรกติของโปรแกรมประยุกต์เว็บจากปริมาณการใช้งานนั้น ทางผู้เขียนใช้ข้อมูลการลงบันทึกการเข้าใช้บริการเว็บ (Web Access Log) ซึ่งเป็นข้อมูลที่ทำการจัดเก็บเพื่อใช้ในการวิเคราะห์ในด้านต่างๆ ทั้งทางด้านความสามารถในการให้บริการ สถานะของผู้ให้บริการเว็บ และยังรวมไปถึงการตรวจสอบปัญหาที่เกิดขึ้นในตัวผู้ให้บริการเว็บอีกด้วย ซึ่งเวิลด์ไวด์เว็บคอนซอร์เทียม (W3C) ได้มีการกำหนดมาตรฐานการลงบันทึกการให้บริการเว็บ (Common Log Format) [26] สำหรับโปรแกรมที่ให้บริการเว็บ

การอ้างถึงข้อมูลที่จัดเก็บในการลงบันทึกการเข้าใช้บริการเว็บนั้น เราจะใช้การอ้างอิงจากโปรแกรมที่มีมาตรฐานการลงบันทึกการให้บริการเว็บ โดยงานวิจัยนี้จะมีการใช้ข้อมูลในการวิเคราะห์หาค่าผิดปรกติจากแหล่งข้อมูล 4 แหล่งด้วยกัน คือ

3.3.2 ข้อมูลการบันทึกการใช้งานโปรแกรมเว็บจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย เป็นการนำข้อมูลที่มีเว็บไซต์ที่เปิดให้บริการตลอดทุกวัน 24 ชั่วโมง นำมาเพื่อใช้ในการวิเคราะห์หาค่าผิดปรกติ โดยตัวอย่างของข้อมูลการร้องขอบริการแสดงในตารางที่ 3.9 ความหมายของข้อมูลการร้องขอบริการแสดงในตารางที่ 3.10 และในรูปที่ 3.5 แสดงให้เห็นแผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัยตามลำดับ

ตารางที่ 3.9 ตัวอย่างข้อมูลการร้องขอบริการจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย

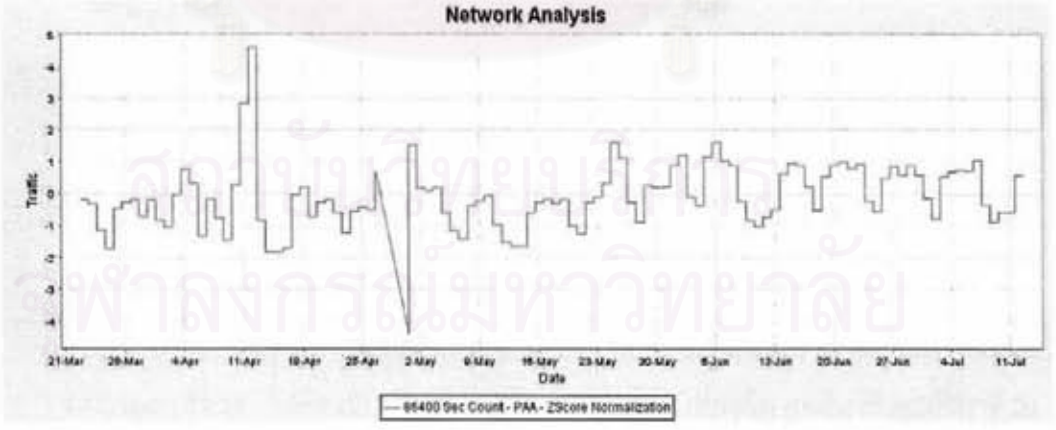
```
203.146.201.34 - - [22/Mar/2006:14:52:32 +0700] "GET
/chula/resources/images/news/im223344.jpg HTTP/1.0" 500 691
"http://www.chula.ac.th/chula/th/main.html" "Mozilla/4.0 (compatible; MSIE 6.0;
Windows NT 5.1; SIMBAR Enabled)"
203.151.33.220 - - [22/Mar/2006:14:52:32 +0700] "GET /chula/th/home.html HTTP/1.1" 500
691 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)"
203.146.201.34 - - [22/Mar/2006:14:52:32 +0700] "GET
/chula/resources/images/news/im060249.jpg HTTP/1.0" 500 691
"http://www.chula.ac.th/chula/th/main.html" "Mozilla/4.0 (compatible; MSIE 6.0;
Windows NT 5.1; SIMBAR Enabled)"
192.169.41.37 - - [22/Mar/2006:14:52:32 +0700] "GET / HTTP/1.0" 500 691
"http://search.yahoo.com/search?p=Faculty+of+Veterinary+Science%2C+Chulalongkorn+Unive
rsity&prssweb=Search&ei=UTF-8&fr=slv1-iy&x=wrt" "Mozilla/4.0 (compatible; MSIE 6.0;
Windows NT 5.0)"
203.146.201.34 - - [22/Mar/2006:14:52:32 +0700] "GET
/system/modules/org.chula.webcontent.main.th/resources/images/go.gif HTTP/1.0" 500 691
"http://www.chula.ac.th/chula/th/main.html" "Mozilla/4.0 (compatible; MSIE 6.0;
Windows NT 5.1; SIMBAR Enabled)"
```

ตารางที่ 3.10 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย

161.200.71.24 - - [22/Mar/2006:14:52:35 +0700] "GET / HTTP/1.1" 200 691 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; InfoPath.1; .NET CLR 2.0.50727)"		
	ค่าที่ปรากฏ	**คำอธิบาย**
1	161.200.71.24	เลขที่อยู่ไอพี (IP Address) ของเครื่องที่ทำการร้องขอบริการเว็บ
2	-	ค่าที่บ่งบอกตัวตนของผู้ร้องขอบริการจากโปรโตคอล identd [27] ซึ่งเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัยไม่มีข้อมูลนี้
3	-	userid ที่บอกถึงตัวตนของผู้ที่ทำการร้องขอบริการเว็บ ซึ่งเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัยไม่มีข้อมูลนี้

ตารางที่ 3.10 (ต่อ) ความหมายของข้อมูลการร้องขอบริการ
จากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย

4	[22/Mar/2006 :14:52:35 +0700]	วันที่เวลาที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึง วันที่ยี่สิบสอง เดือนมีนาคม ปีค.ศ. สองพันหก เวลาสิบสี่นาฬิกา ห้าสิบสองนาที สามสิบห้าวินาที ณ เวลาที่การร้องขอบริการเว็บนั้นเครื่องผู้ให้บริการ อยู่ในเขตเวลาที่ต้องบวกเพิ่มเจ็ดชั่วโมงจากเวลาปานกลางกรีนิช (Greenwich Mean Time – GMT)
5	"GET / HTTP/1.1"	ข้อมูลในการร้องขอบริการเว็บ ซึ่งในกรณีนี้ จะหมายถึงการร้องขอบริการผ่านวิธี GET โดยทำการร้องขอไฟล์ที่พาธ (path) ที่ root (/) ซึ่งผู้ให้บริการเว็บจะทำการส่งข้อมูลหน้าเว็บเพจที่เป็นค่าดีฟอลต์ (default) ไปให้ โดยผู้ร้องขอบริการใช้โปรโตคอล HTTP รุ่น 1.1 ซึ่งข้อมูลนี้เป็นข้อมูลที่ให้รายละเอียดในการร้องขอบริการเว็บ
6	200	รหัสของสถานภาพที่ผู้ให้บริการตอบกลับไปยังผู้ร้องขอบริการเว็บ โดยข้อมูลนี้เป็นข้อมูลที่มีความสำคัญมากเนื่องจากว่า เป็นตัวที่ระบุถึงการให้บริการครั้งนั้นๆ ว่าสำเร็จ หรือ ไม่สำเร็จ หรือกรณีอื่นๆ [28] ซึ่งในกรณีนี้คือสำเร็จ
7	691	ขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บโดยไม่รวมถึงเฮดเดอร์ (Header) ในการ ตอบกลับ ซึ่งถ้าไม่มีข้อมูลตอบกลับจะมีค่าเป็น - หรือ 0 มีหน่วยเป็นไบต์ (Byte)



รูปที่ 3.5 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย

จากรูปที่ 3.5 แสดงให้เห็นถึงปริมาณการส่งข้อมูลจากเครื่องให้บริการเว็บของเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยไปยังผู้ร้องขอบริการเว็บ ระหว่างวันที่ 21 มีนาคม ค.ศ. 2006 จนถึงวันที่ 11 กรกฎาคม ค.ศ. 2006 ที่ได้ผ่านการลดมิติของข้อมูลด้วยวิธี PAA ที่มีช่วงเวลา 1 วัน และทำให้เป็นบรรทัดฐานด้วยวิธี Z-Score

3.3.3 ข้อมูลที่สร้างขึ้นมาจากโปรแกรม โดยกำหนดให้ข้อมูลมีรูปแบบที่เป็นไปตามมาตรฐานการลงบันทึกการให้บริการเว็บ ซึ่งข้อมูลที่ถูกสร้างสามารถเปลี่ยนแปลงลักษณะปริมาณการใช้งานได้ตามต้องการ ไม่ว่าจะเป็นการจำลองสภาพการใช้งานสูง รูปแบบการใช้งานในลักษณะแปลกๆ เป็นต้น สามารถนำมาสร้างข้อมูลเพื่อใช้ในการวิเคราะห์หาค่าผิดปรกติได้ โดยตัวอย่างของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นแสดงในตารางที่ 3.11 ความหมายของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นแสดงในตารางที่ 3.12 และในรูปที่ 3.6 แสดงให้เห็นแผนภูมิกราฟข้อมูลการร้องขอบริการเว็บที่สร้างขึ้นมาจากโปรแกรม โดยตัวอย่างแฟ้มโครงแบบเอกซ์เอ็มแอลที่ทำการสร้างข้อมูลการบันทึกการใช้งานของเว็บ แสดงในตารางที่ ก.1 ของภาคผนวก ก ดังนี้

ตารางที่ 3.11 ตัวอย่างของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นจากโปรแกรม

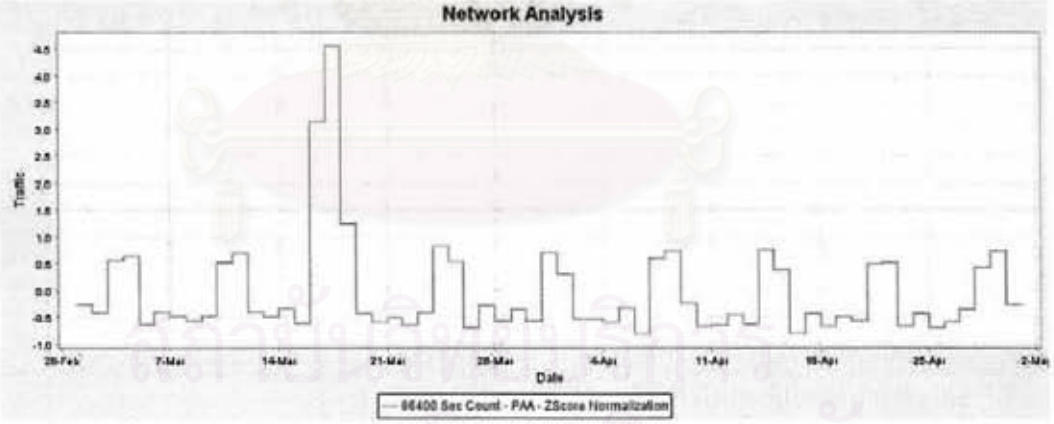
```
192.168.1.219 - - [01/Mar/2006:00:00:00 +0700] "GET /aaa.txt HTTP/1.1" 200 1987 4893
202.3.145.3 80
192.168.1.39 - - [01/Mar/2006:00:00:00 +0700] "GET /aaa.txt HTTP/1.1" 200 1549 1410
202.3.145.3 80
192.168.1.214 - - [01/Mar/2006:00:00:00 +0700] "GET /aaa.txt HTTP/1.1" 200 1644 2165
202.3.145.3 80
192.168.1.188 - - [01/Mar/2006:00:00:00 +0700] "GET /aaa.txt HTTP/1.1" 200 1908 4264
202.3.145.3 80
```

ตารางที่ 3.12 ความหมายของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นจากโปรแกรม

192.168.1.219 - - [01/Mar/2006:00:00:00 +0700] "GET /aaa.txt HTTP/1.1" 200 1987 4893 202.3.145.3 80		
	ค่าที่ปรากฏ	**คำอธิบาย**
1	192.168.1.219	เลขที่อยู่ไอพีของเครื่องที่ทำการร้องขอบริการเว็บ
2	-	ค่าที่บ่งบอกตัวตนของผู้ร้องขอบริการจากโปรโตคอล identd [27] ซึ่งไม่มีการสร้างข้อมูลนี้
3	-	userid ที่บอกถึงตัวตนของผู้ที่ทำการร้องขอบริการเว็บ ซึ่งไม่มีการสร้างข้อมูลนี้
4	[01/Mar/2006:00:00:00 +0700]	วันที่เวลาที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้ จะหมายถึงวันที่หนึ่งเดือนมีนาคม ปีค.ศ. สองพันหก เวลาเที่ยงคืน ศูนย์นาที ศูนย์วินาที ณ เวลาที่การร้องขอบริการเว็บนั้น เครื่องผู้ให้บริการอยู่ในเขตเวลาที่ต้องบวกเพิ่มเจ็ดชั่วโมงจากเวลาปานกลางกรีนิช

ตารางที่ 3.12 (ต่อ) ความหมายของข้อมูลการร้องขอบริการที่ถูกสร้างขึ้นจากโปรแกรม

5	"GET /aaa.txt HTTP/1.1"	ข้อมูลในการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึงการร้องขอบริการ ผ่านวิธี GET โดยทำการร้องขอไฟล์ aaa.txt พาธ / ซึ่งผู้ให้บริการเว็บจะทำการส่งข้อมูลหน้าเว็บเพจ aaa.txt ไปให้ โดยผู้ร้องขอบริการใช้โปรโตคอล HTTP รุ่น 1.1 ซึ่งข้อมูลนี้เป็นข้อมูลที่ให้รายละเอียดในการร้องขอบริการเว็บ
6	200	รหัสของสถานภาพที่ผู้ให้บริการตอบกลับไปยังผู้ร้องขอบริการเว็บ โดยข้อมูลนี้เป็นข้อมูลที่มีความสำคัญมากเนื่องจากว่า เป็นตัวที่ระบุถึงการให้บริการครั้งนั้นๆว่าสำเร็จ หรือ ไม่สำเร็จ หรือกรณีอื่นๆ [28] ซึ่งในกรณีนี้คือสำเร็จ
7	1987	ขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บโดยไม่รวมถึงเฮดเดอร์ในการตอบกลับ
8	4893	ค่าของพอร์ตที่ผู้ร้องขอบริการส่งข้อมูลมายังผู้ให้บริการ ในกรณีนี้คือ พอร์ตหมายเลข 4893
9	202.3.145.3	เลขที่อยู่ไอพีของเครื่องให้บริการเว็บ
10	80	ค่าของพอร์ตที่ผู้ให้บริการเว็บรับข้อมูลจากผู้ร้องขอบริการ ในกรณีนี้คือ พอร์ตหมายเลข 80



รูปที่ 3.6 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์
ที่ถูกสร้างขึ้นจากโปรแกรม

จากรูปที่ 3.6 แสดงให้เห็นถึงปริมาณการส่งข้อมูลจากเครื่องให้บริการเว็บที่ถูกสร้างขึ้นไปยังผู้ร้องขอบริการเว็บ ระหว่างวันที่ 1 มีนาคม ค.ศ. 2006 จนถึงวันที่ 30 เมษายน ค.ศ. 2006 ที่ได้

ผ่านการลดมิติของข้อมูลด้วยวิธี PAA ที่มีช่วงเวลา 1 วัน และทำให้เป็นบรรทัดฐานด้วยวิธี Z-Score

3.3.4 ข้อมูลการบันทึกการใช้งานโปรแกรมเว็บจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ เป็นการนำข้อมูลที่มีเว็บไซต์ที่เปิดให้บริการตลอดทุกวัน 24 ชั่วโมง ที่มีปริมาณการร้องขอเป็นจำนวนมาก เพื่อใช้ในการวิเคราะห์หาค่าผิดปรกติที่เกิดขึ้นในแวดวงที่หลากหลาย ซึ่งตัวอย่างของข้อมูลการร้องขอบริการแสดงในตารางที่ 3.13 โดยความหมายของข้อมูลการร้องขอบริการแสดงในตารางที่ 3.14 และในรูปที่ 3.7 แสดงให้เห็นแผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ

ตารางที่ 3.13 ตัวอย่างของข้อมูลการร้องขอบริการจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ

```
2007-02-06 23:59:59 202.57.155.216 GET /JavaScript/Cookies.js - 80 - 124.120.116.94
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727;+
InfoPath.1;+.NET+CLR+1.1.4322) 200 0 64 0 579
2007-02-06 23:59:59 202.57.155.216 GET /Home/images/photo_7.gif - 80 - 134.173.232.13
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+1.1.4322;
+.NET+CLR+2.0.50727;+InfoPath.1) 200 0 0 301 441
2007-02-06 23:59:59 202.57.155.216 GET /images/icon_mcard.gif - 80 - 64.198.232.69
Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.1.4322;+InfoPat
h.1;+.NET+CLR+2.0.50727) 200 0 0 2595 484
2007-02-06 23:59:59 202.57.155.216 GET /images/icon_18.gif - 80 - 58.8.120.117 M
ozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.1.4322;+InfoPath.1)
 200 0 0 956 555
2007-02-06 23:59:59 202.57.155.216 GET /images/bg_adrate.gif - 80 - 64.198.232.69
Mozilla/4.0+(compatible;+MSIE+7.0;+Windows+NT+5.1;+.NET+CLR+1.1.4322;+InfoPath
.1;+.NET+CLR+2.0.50727) 200 0 0 406 485
2007-02-06 23:59:59 202.57.155.216 GET /images/SmallCss.gif - 80 - 203.121.160.61
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+InfoPath.1;+.NET+CLR+1
.1.4322) 200 0 0 475 456
```

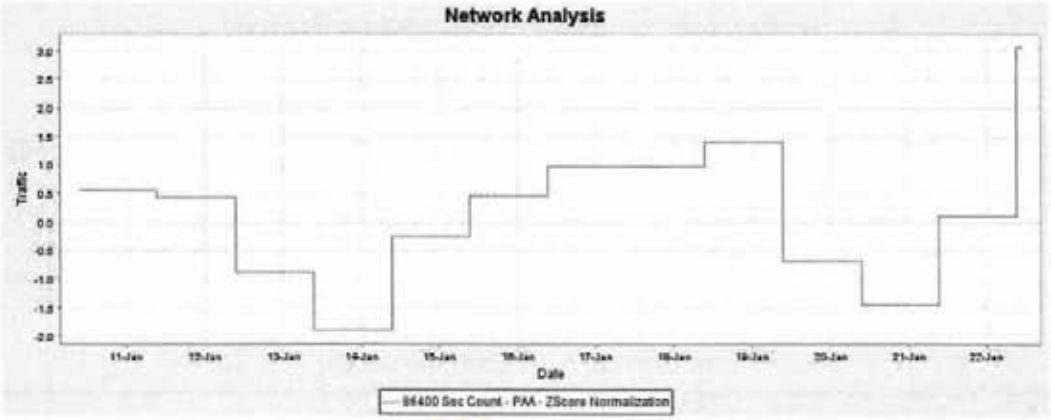
ตารางที่ 3.14 ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ

2007-02-06 23:59:59 202.57.155.216 GET /JavaScript/Cookies.js - 80 - 124.120.116.94 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1 ;+SV1;+.NET+CLR+2.0.50727;+InfoPath.1;+.NET+CLR+1.1.4322) 200 0 64 0 579

	ค่าที่ปรากฏ	คำอธิบาย
1	2007-02-06	วันที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึง วันที่หกเดือนกุมภาพันธ์ ปีค.ศ. สองพันเจ็ด
2	23:59:59	เวลาที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึง เวลายี่สิบสามนาฬิกา ห้าสิบเก้านาที ห้าสิบเก้าวินาที
3	202.57.155.216	เลขที่อยู่ไอพีของเครื่องผู้ให้บริการเว็บ

ตารางที่ 3.14 (ต่อ) ความหมายของข้อมูลการร้องขอบริการจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ

4	GET /JavaScript/Cookies.js	ข้อมูลในการร้องขอบริการเว็บ ซึ่งในกรณีนี้ จะหมายถึงการร้องขอบริการผ่านวิธี GET โดยทำการร้องขอไฟล์ที่พาธ JavaScript ร้องขอไฟล์ที่มีชื่อว่า Cookie.js ซึ่งข้อมูลนี้เป็นข้อมูลที่ให้รายละเอียดในการร้องขอบริการเว็บ
5	-	ค่าของพารามิเตอร์ที่ส่งมาพร้อมกับการร้องขอบริการเว็บ ซึ่งในกรณีนี้ไม่มีพารามิเตอร์ที่ส่งมา
6	80	ค่าของพอร์ตที่ผู้ให้บริการเว็บรับข้อมูลจากผู้ร้องขอบริการ ในกรณีนี้คือ พอร์ตหมายเลข 80
7	-	userid ที่บอกถึงตัวตนของผู้ที่ทำการร้องขอบริการเว็บ ซึ่งเว็บไซต์ของหนังสือพิมพ์ผู้จัดการไม่มีข้อมูลนี้
8	124.120.116.94	เลขที่อยู่ไอพีของเครื่องที่ทำการร้องขอบริการเว็บ
9	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727;+InfoPath.1;+.NET+CLR+1.1.4322)	รายละเอียดของโปรแกรมเบราเซอร์ที่ทำการร้องขอบริการเว็บ
10	200	รหัสของสถานภาพที่ผู้ให้บริการตอบกลับไปยังผู้ร้องขอบริการเว็บ โดยข้อมูลนี้เป็นข้อมูลที่มีความสำคัญมากเนื่องจากว่า เป็นตัวที่ระบุถึงการให้บริการครั้งนั้นๆว่าสำเร็จ หรือ ไม่สำเร็จ หรือกรณีอื่นๆ [28] ซึ่งในกรณีนี้คือสำเร็จ
11	-	รหัสย่อยของสถานภาพที่แสดงถึงรายละเอียดของความผิดพลาด ซึ่งในกรณีนี้ไม่มีข้อผิดพลาดเกิดขึ้น
12	64	รหัสของสถานภาพของโปรแกรมวินโดวส์ซึ่งในกรณีนี้คือ ไม่ได้รับการตอบรับจากผู้ร้องขอบริการว่าได้รับข้อมูลเสร็จสิ้นแล้ว ซึ่งมีค่าเท่ากับ 0 แสดงว่าทำงานสำเร็จ
13	0	ขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บ ซึ่งถ้าไม่มีข้อมูลตอบกลับจะมีค่าเป็น 0 มีหน่วยเป็นไบต์
14	579	ขนาดของข้อมูลที่ทำการร้องขอไปยังผู้ให้บริการเว็บ ซึ่งถ้าไม่มีข้อมูลที่ร้องขอจะมีค่าเป็น 0 มีหน่วยเป็นไบต์



รูปที่ 3.7 แผนภูมิกราฟข้อมูลการร้องขอบริการเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ

จากรูปที่ 3.7 แสดงให้เห็นถึงปริมาณการส่งข้อมูลจากเครื่องให้บริการเว็บของเว็บไซต์หนังสือพิมพ์ผู้จัดการไปยังผู้ร้องขอบริการเว็บ ระหว่างวันที่ 11 มกราคม ค.ศ. 2007 จนถึงวันที่ 22 มกราคม ค.ศ. 2006 ที่ได้ผ่านการลดมิติของข้อมูลด้วยวิธี PAA ที่มีช่วงเวลา 1 วัน และทำให้เป็นบรรทัดฐานด้วยวิธี Z-Score

3.3.5 ข้อมูลการบันทึกการใช้งานโปรแกรมเว็บจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ เป็นการนำข้อมูลที่มีเว็บไซต์ที่เปิดให้บริการตลอดทุกวัน 24 ชั่วโมง ที่มีปริมาณการร้องขอเป็นจำนวนไม่มากนัก เพื่อใช้ในการวิเคราะห์หาค่าผิดปรกติที่เกิดขึ้นในแวดวงที่หลากหลาย ซึ่งตัวอย่างของข้อมูลการร้องขอบริการแสดงในตารางที่ 3.15 โดยความหมายของข้อมูลการร้องขอบริการแสดงในตารางที่ 3.16 และในรูปที่ 3.8 แสดงให้เห็นแผนภูมิกราฟข้อมูลการร้องขอบริการจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ

ตารางที่ 3.15 ตัวอย่างของข้อมูลการร้องขอบริการจาก
เว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ

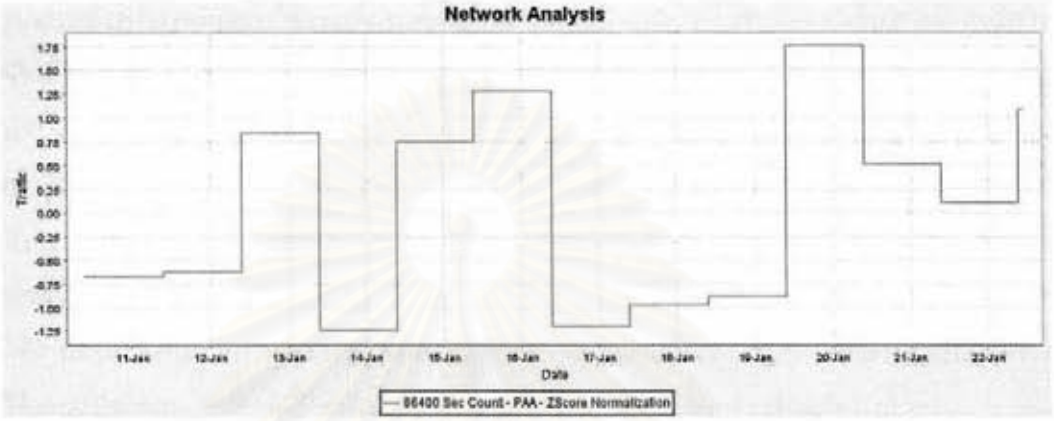
```
2007-02-07 00:00:00 203.150.28.104 - 202.57.155.222 80 GET /radio/images/post.gif -
304 165 590 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR
+2.0.50727;+InfoPath.1)
2007-02-07 00:00:00 203.188.14.72 - 202.57.155.222 80 GET /radio/images/pic01.gif -
200 5028 484 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+NetCapt
or+7.5.4;+.NET+CLR+2.0.50727;+.NET+CLR+1.1.4322)
2007-02-07 00:00:00 203.188.14.72 - 202.57.155.222 80 GET /radio/images/pic02.gif -
200 627 484 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+NetCapto
r+7.5.4;+.NET+CLR+2.0.50727;+.NET+CLR+1.1.4322)
2007-02-07 00:00:00 202.57.155.222 - 202.57.155.222 80 GET /radio/Default.asp - 200 0
117 Wget/1.8.2
2007-02-07 00:00:00 203.188.14.72 - 202.57.155.222 80 GET /radio/images/bullet.gif -
200 300 485 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+NetCapt
or+7.5.4;+.NET+CLR+2.0.50727;+.NET+CLR+1.1.4322)
2007-02-07 00:00:00 203.188.14.72 - 202.57.155.222 80 GET /radio/images/spacer.gif -
200 292 485 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+NetCapt
or+7.5.4;+.NET+CLR+2.0.50727;+.NET+CLR+1.1.4322)
```

ตารางที่ 3.16 ความหมายของข้อมูลการร้องขอบริการ
จากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ

2007-02-07 00:00:00 203.150.28.104 - 202.57.155.222 80 GET /radio/images/post.gif - 304 165 590 Mozilla/4.0+(compatible; +MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727;+InfoPath.1)		
	ค่าที่ปรากฏ	คำอธิบาย
1	2007-02-07	วันที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึง วันที่เจ็ด เดือนกุมภาพันธ์ ปีค.ศ. สองพันเจ็ด
2	00:00:00	เวลาที่ทำการร้องขอบริการเว็บ ซึ่งในกรณีนี้จะหมายถึง เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที
3	203.150.28.104	เลขที่อยู่ไอพีของเครื่องที่ทำการร้องขอบริการเว็บ
4	-	ค่าของพอร์ตที่ผู้ร้องขอบริการในกรณีนี้คือไม่ได้เก็บข้อมูลนี้ไว้
5	202.57.155.222	เลขที่อยู่ไอพีของเครื่องผู้ให้บริการเว็บ
6	80	ค่าของพอร์ตที่ผู้ให้บริการเว็บรับข้อมูลจากผู้ร้องขอบริการ ในกรณีนี้คือ พอร์ตหมายเลข 80
7	GET /radio/images/ post.gif	ข้อมูลในการร้องขอบริการเว็บ ซึ่งในกรณีนี้ จะหมายถึงการร้องขอบริการผ่านวิธี GET โดยทำการร้องขอไฟล์ที่พาธ radio/images ร้องขอไฟล์ที่มีชื่อว่า post.gif ซึ่งข้อมูลนี้เป็นข้อมูลที่ให้รายละเอียดในการร้องขอบริการเว็บ
8	-	ค่าของพารามิเตอร์ที่ส่งมาพร้อมกับการร้องขอบริการเว็บ ซึ่งในกรณีนี้ไม่มีพารามิเตอร์ที่ส่งมา
9	304	รหัสของสถานภาพที่ผู้ให้บริการตอบกลับไปยังผู้ร้องขอบริการเว็บ โดยข้อมูลนี้เป็นข้อมูลที่มีความสำคัญมากเนื่องจากว่า เป็นตัวที่ระบุถึงการให้บริการครั้งนั้นๆว่าสำเร็จ หรือ ไม่สำเร็จ หรือกรณีอื่นๆ [28] ซึ่งในกรณีนี้คือพบว่าไฟล์ที่ทำการร้องขอไม่ได้มีการเปลี่ยนแปลงจากการร้องขอครั้งก่อน
10	165	ขนาดของข้อมูลที่ส่งกลับไปยังผู้ร้องขอบริการเว็บ ซึ่งถ้าไม่มีข้อมูลตอบกลับจะมีค่าเป็น 0 มีหน่วยเป็นไบต์
11	590	ขนาดของข้อมูลที่ทำการร้องขอไปยังผู้ให้บริการเว็บ ซึ่งถ้าไม่มีข้อมูลที่ร้องขอจะมีค่าเป็น 0 มีหน่วยเป็นไบต์

ตารางที่ 3.16 (ต่อ) ความหมายของข้อมูลการร้องขอบริการ
จากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ

12	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+2.0.50727;+InfoPath.1)	รายละเอียดของโปรแกรมเบราเซอร์ที่ทำการร้องขอบริการเว็บ



รูปที่ 3.8 แผนภูมิกราฟข้อมูลการร้องขอบริการจาก
เว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ

จากรูปที่ 3.8 แสดงให้เห็นถึงปริมาณการส่งข้อมูลจากเครื่องให้บริการเว็บของเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการไปยังผู้ร้องขอบริการเว็บ ระหว่างวันที่ 11 มกราคม ค.ศ. 2007 จนถึงวันที่ 22 มกราคม ค.ศ. 2007 ที่ได้ผ่านการลดมิติของข้อมูลด้วยวิธี PAA ที่มีช่วงเวลา 1 วัน และทำให้เป็นบรรทัดฐานด้วยวิธี Z-Score

บทที่ 4
ผลการตรวจหาค่าผิดปรกติและสร้างข้อมูลบันทึกการใช้งานบริการเว็บ

4.1 ข้อมูลที่ใช้ในการวิจัย

4.1.1 แฟ้มโครงแบบที่ใช้ในการทดลองสร้างข้อมูลบันทึกการใช้งานโปรแกรมประยุกต์เว็บ

ในการสร้างข้อมูลบันทึกการใช้งานโปรแกรมประยุกต์เว็บนั้น สามารถสร้างข้อมูลได้หลายรูปแบบทั้งรูปแบบที่มีการใช้งานที่ปรกติ คือ เหมือนกับลักษณะการใช้งานจากเว็บไซต์ที่ให้บริการจริงที่มีลักษณะเป็นฤดูกาล ตัวอย่างเช่น ข้อมูลบันทึกการใช้งานจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย ข้อมูลปริมาณการใช้งานจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ และข้อมูลปริมาณการใช้งานจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ เป็นต้น ทั้งยังสามารถสร้างลักษณะที่ปริมาณการใช้งานที่ผิดปรกติได้หลายตำแหน่งเช่นกัน ซึ่งในการสร้างข้อมูลในงานวิจัยนี้จะเป็นการสร้างข้อมูลที่มีลักษณะปริมาณการใช้งานที่มีลักษณะที่ปรกติ ข้อมูลที่มีลักษณะปริมาณการใช้งานที่มีลักษณะที่ปรกติที่มีการเพิ่มค่าผิดปรกติ 1 ตำแหน่ง และข้อมูลที่มีลักษณะปริมาณการใช้งานที่มีลักษณะที่ปรกติที่มีการเพิ่มค่าผิดปรกติ 3 ตำแหน่ง ดังรายละเอียดต่อไปนี้

4.1.1.1 แฟ้มโครงแบบ A เป็นแฟ้มโครงแบบที่ทำการสร้างข้อมูลบันทึกการใช้งานโปรแกรมประยุกต์เว็บที่มีลักษณะการใช้งานปรกติให้คล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ โดยเป็นการสร้างข้อมูลระหว่างวันที่ 1 มีนาคม ค.ศ. 2006 เวลา 0 นาฬิกา 0 นาที 0 วินาที ถึงวันที่ถึงวันที่ 30 เมษายน ค.ศ. 2006 เวลา 23 นาฬิกา 59 นาที 59 วินาที ดังแสดงในตารางที่ ก.1 ในภาคผนวก ก. ซึ่งมีรายละเอียดข้อมูลดังต่อไปนี้

- ช่วงเวลาในการร้องขอบริการเว็บคือทุกวัน ในช่วงเดือนมีนาคมถึงเดือนเมษายน ปีค.ศ. 2006
- ปริมาณการใช้งานระหว่างวันธรรมดา อันได้แก่วันจันทร์ถึงวันศุกร์มีความแตกต่างกับปริมาณการใช้งานในวันหยุด อันได้แก่วันเสาร์และวันอาทิตย์
- ปริมาณการใช้งานของวันธรรมดาในช่วงเวลาเริ่มวันใหม่จนถึงหกนาฬิกา ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างแปดพันห้าร้อยถึงเก้าพันครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันสี่ร้อยถึงหนึ่งพันห้าร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของวันธรรมดาในช่วงเวลาหกนาฬิกาจนถึงเที่ยงคืน ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างแปดพันห้าร้อยถึงเก้าพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันห้าร้อยถึงหนึ่งพันเจ็ดร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของวันหยุดในช่วงเวลาเริ่มวันใหม่จนถึงหกนาฬิกา ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างเก้าพันห้าร้อยถึงหนึ่งหมื่นครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันสี่ร้อยถึงหนึ่งพันห้าร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของวันหยุด ในช่วงเวลาหกนาฬิกาจนถึงเที่ยงคืน ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างหนึ่งหมื่นถึงหนึ่งหมื่นหนึ่งพันครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันห้าร้อยถึงหนึ่งพันเจ็ดร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

4.1.1.2 แฟ้มโครงแบบ B คือแฟ้มโครงแบบที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A แต่ได้มีการเพิ่มลักษณะของรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 ทั้งวัน ให้มีปริมาณการใช้งานบริการเว็บมากกว่าปรกติ ดังแสดงในตารางที่ ก.2 ในภาคผนวก ก. โดยมีรายละเอียดข้อมูลดังต่อไปนี้

- ช่วงเวลาในการร้องขอบริการเว็บคือทุกวัน ในช่วงเดือนมีนาคมถึงเดือนเมษายน ปีค.ศ. 2006

- ปริมาณการใช้งานของค่าผิดปรกติวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 รวมทั้งวัน มีการสุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างหนึ่งหมื่นถึงสองหมื่นสองพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่ง

ข้อมูลที่มีขนาดระหว่างหนึ่งพันถึงสองพันไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

4.1.1.3 แฟ้มโครงแบบ C คือแฟ้มโครงแบบที่มีลักษณะการใช้งานปรกติให้คล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ โดยเพิ่มตำแหน่งข้อมูลที่ผิดปรกติไว้ 3 ตำแหน่ง ได้แก่ ตำแหน่งที่ 1 คือวันที่ 19 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ตำแหน่งที่ 2 คือวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บมากกว่าปรกติ และตำแหน่งที่ 3 คือวันที่ 16 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ดังแสดงในตารางที่ ก.3 ในภาคผนวก ก. โดยมีรายละเอียดข้อมูลดังต่อไปนี้

- ช่วงเวลาในการร้องขอบริการเว็บคือทุกวัน ในช่วงเดือนมีนาคมถึงเดือนเมษายน ปีค.ศ. 2006

- ปริมาณการใช้งานระหว่างวันธรรมดา อันได้แก่วันจันทร์ถึงวันศุกร์มีความแตกต่างกับปริมาณการใช้งานในวันหยุด อันได้แก่วันเสาร์และวันอาทิตย์

- ปริมาณการใช้งานของวันธรรมดาในช่วงเวลาเริ่มวันใหม่จนถึงหกนาฬิกา ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างแปดพันห้าร้อยถึงเก้าพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันถึงหนึ่งพันห้าร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของวันธรรมดาในช่วงเวลาหกนาฬิกาจนถึงเที่ยงคืน ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างแปดพันห้าร้อยถึงเก้าพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันถึงหนึ่งพันเจ็ดร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของวันหยุดในช่วงเวลาเริ่มวันใหม่จนถึงหกนาฬิกา ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างเก้าพันห้าร้อยถึงหนึ่งหมื่นห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันสี่ร้อยถึงหนึ่งพันห้าร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของวันหยุด ในช่วงเวลาหกนาฬิกาจนถึงเที่ยงคืน ให้สุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างหนึ่งหมื่นห้าร้อยถึงหนึ่งหมื่นหนึ่งพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันห้าร้อยถึงหนึ่งพันเจ็ดร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของค่าผิดปรกติตำแหน่งที่ 1 ในวันที่ 19 มีนาคม ค.ศ. 2006 ตำแหน่งที่ 3 ในวันที่ 16 เมษายน ค.ศ. 2006 รวมทั้งวัน มีการสุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างสี่พันถึงสี่พันพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันถึงหนึ่งพันหนึ่งร้อยไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

- ปริมาณการใช้งานของค่าผิดปรกติตำแหน่งที่ 2 ในวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 รวมทั้งวัน มีการสุ่มจำนวนการร้องขอบริการในแต่ละวันที่มีจำนวนระหว่างหนึ่งหมื่นห้าพันถึงสองหมื่นสองพันห้าร้อยครั้ง โดยการร้องขอบริการแต่ละครั้งให้มีการสุ่มปริมาณการส่งข้อมูลที่มีขนาดระหว่างหนึ่งพันถึงสองพันไบต์ เลขที่อยู่ไอพีของผู้ให้บริการเว็บคือ 202.3.145.3 ที่พอร์ตหมายเลข 80 เลขที่อยู่ไอพีของผู้ที่ร้องขอบริการเว็บให้มีการสุ่มเลขที่อยู่ไอพีให้อยู่ในช่วง 192.168.1.0 ถึง 192.168.1.254 และให้มีการสุ่มพอร์ตของผู้ที่ร้องขอบริการเว็บให้อยู่ระหว่างหมายเลข 1024 และ 5000

4.1.2 ข้อมูลที่ใช้ในการตรวจหาค่าผิดปรกติ

4.1.2.1 ข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่มีการใช้งานจริง

- ข้อมูล A เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัย ระหว่างวันที่ 22 มีนาคม ค.ศ. 2006 เวลา 14 นาฬิกา 52 นาที 32 วินาที ถึงวันที่ 12 กรกฎาคม ค.ศ. 2006 เวลา 12 นาฬิกา 5 นาที 3 วินาที ที่มีลักษณะปริมาณการใช้งานกลางวันมากกว่ากลางคืน และมีปริมาณการใช้งานวันธรรมดามากกว่าวันหยุด

- ข้อมูล B เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัย ระหว่างวันที่ 10 สิงหาคม ค.ศ. 2006 เวลา 16 นาฬิกา 51 นาที 49 วินาที ถึงวันที่ 21 ธันวาคม ค.ศ. 2006 เวลา 13 นาฬิกา 15 นาที 32 วินาที ที่มีลักษณะปริมาณการใช้งานกลางวันมากกว่ากลางคืน และมีปริมาณการใช้งานวันธรรมดามากกว่าวันหยุด

- ข้อมูล C เป็นข้อมูลจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ ระหว่างวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลา 3 นาฬิกา 53 นาที 1 วินาที ถึงวันที่ 28 กุมภาพันธ์ ค.ศ. 2007 เวลา 23 นาฬิกา 59 นาที 59 วินาที ที่มีลักษณะปริมาณการใช้งานกลางวันมากกว่ากลางคืน และมีปริมาณการใช้งานวันธรรมดามากกว่าวันหยุด

- ข้อมูล D เป็นข้อมูลจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ ระหว่างวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลา 3 นาฬิกา 57 นาที 53 วินาที ถึงวันที่ 28 กุมภาพันธ์ ค.ศ. 2007 เวลา 23 นาฬิกา 59 นาที 59 วินาที ที่มีลักษณะปริมาณการใช้งานกลางวันมากกว่ากลางคืน และมีปริมาณการใช้งานวันธรรมดามากกว่าวันหยุด

4.1.2.2 ข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่ถูกสร้างขึ้น

- ข้อมูล E เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ A ที่เป็นลักษณะข้อมูลที่มีลักษณะการใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์

- ข้อมูล F เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ B ที่เป็นลักษณะข้อมูลที่มีลักษณะการใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ และมีการเพิ่มรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 ทั้งวัน

- ข้อมูล G เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ C ที่เป็นลักษณะข้อมูลที่มีลักษณะการใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ และมีการเพิ่มรูปแบบที่ผิดปรกติไว้ 3 ตำแหน่ง ได้แก่ ตำแหน่งที่ 1 คือวันที่

19 มีนาคม ค.ศ. 2006 ทั้งวัน ตำแหน่งที่ 2 คือวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน และตำแหน่งที่ 3 คือวันที่ 16 เมษายน ค.ศ. 2006 ทั้งวัน

4.2 การดำเนินงานวิจัย

4.2.1 การสร้างข้อมูลบันทึกการใช้งานบริการเว็บ เป็นการสร้างข้อมูลตามแฟ้มโครงแบบที่กำหนด ใช้คำสั่งในการดำเนินงานดังนี้

java –jar WebAccessGenerator.jar แฟ้มโครงแบบเอกซ์เอ็มแอล

โดยคำสั่งดำเนินงานดังกล่าวใช้แฟ้มโครงแบบเอกซ์เอ็มแอล A B และ C ในตารางที่ ก.1 ก.2 และ ก.3 ในภาคผนวก ก. ตามลำดับ

4.2.2 การสร้างแผนภูมิกราฟในการใช้งานบริการเว็บ และการแปลงข้อมูลให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์เท่านั้น เป็นการสร้างแผนภูมิกราฟ และทำการแปลงข้อมูลให้มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์เท่านั้น ต้องใช้คำสั่งดำเนินงานดังนี้

java –jar ChartGenerator.jar แฟ้มโครงแบบ

โดยแฟ้มโครงแบบดังกล่าวจะมีลักษณะที่แตกต่างกันที่ตำแหน่งที่เก็บปริมาณข้อมูลที่ส่งกลับจากผู้ให้บริการเว็บไปยังผู้ร้องขอบริการเว็บ ตำแหน่งของวันที่เวลาที่มีการร้องขอบริการเว็บ รูปแบบของวันที่เวลาที่มีการร้องขอบริการเว็บ ตำแหน่งของเลขที่อยู่ไอพีของผู้ให้บริการเว็บ ตำแหน่งเลขที่อยู่ไอพีของผู้ร้องขอบริการเว็บ และชนิดของแฟ้มบันทึกการใช้งาน โดยให้ข้อมูลบันทึกการใช้งานเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยแทนด้วยข้อมูล ก. แสดงในตารางที่ ข.1 ในภาคผนวก ข. ข้อมูลบันทึกการใช้งานเว็บไซต์ผู้จัดการแทนด้วยข้อมูล ข. แสดงในตารางที่ ข.2 ในภาคผนวก ข. ข้อมูลบันทึกการใช้งานเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการแทนด้วยข้อมูล ค. แสดงในตารางที่ ข.3 ในภาคผนวก ข. และ ข้อมูลบันทึกการใช้งานเว็บไซต์ที่ถูกสร้างขึ้นแทนด้วยข้อมูล ง. ดังแสดงในตารางที่ข.4 ในภาคผนวก ข. โดยสิ่งที่แตกต่างกันในแต่ละแฟ้มโครงแบบ มีรายละเอียดดังแสดงในตารางที่ 4.1

จากตารางที่ 4.1 แสดงให้เห็นถึงลักษณะข้อมูลที่แตกต่างกันสำหรับข้อมูลต่างๆ ดังนี้

ข้อมูลการบันทึกการใช้งานบริการเว็บไซต์ข้อมูล ก. หรือข้อมูลบันทึกการใช้งานบริการเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย และข้อมูล ง. หรือข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่ถูกสร้างขึ้น มีตำแหน่งปริมาณข้อมูลอยู่ตำแหน่งที่ 10 ตำแหน่งวันที่เวลาที่มีการร้องขอบริการเว็บ

อยู่ตำแหน่งที่ 4 ตำแหน่งเลขที่อยู่ไอพีผู้ให้บริการอยู่ตำแหน่งที่ 9 ตำแหน่งเลขที่อยู่ไอพีผู้ร้องขอบริการอยู่ตำแหน่งที่ 3 ชนิดของแฟ้มข้อมูลเป็นแฟ้มบันทึกการใช้งานมาจากโปรแกรมอะแพชชี

ตารางที่ 4.1 ลักษณะตำแหน่งและชนิดของข้อมูลที่แตกต่างกันของข้อมูลบันทึกการใช้งานบริการเว็บไซต์

ตัวแปร / ข้อมูล	ตำแหน่งปริมาณข้อมูล	ตำแหน่งวันที่เวลา	ตำแหน่งเลขที่อยู่ไอพีผู้ให้บริการ	ตำแหน่งเลขที่อยู่ไอพีผู้ร้องขอบริการ	ชนิดของแฟ้ม
ก.	10	4	9	3	apache
ข.	14	1	9	3	iis
ค.	11	1	3	5	iis
ง.	10	4	9	3	apache

- ข้อมูลการบันทึกการใช้งานบริการเว็บไซต์ข้อมูล ข. หรือข้อมูลบันทึกการใช้งานบริการเว็บไซต์ของหนังสือพิมพ์ผู้จัดการ มีตำแหน่งปริมาณข้อมูลอยู่ตำแหน่งที่ 14 ตำแหน่งวันที่เวลาที่มีการร้องขอบริการเว็บอยู่ตำแหน่งที่ 1 ตำแหน่งเลขที่อยู่ไอพีผู้ให้บริการอยู่ตำแหน่งที่ 9 ตำแหน่งเลขที่อยู่ไอพีผู้ให้บริการอยู่ตำแหน่งที่ 3 ชนิดของแฟ้มข้อมูลเป็นแฟ้มบันทึกการใช้งานมาจากโปรแกรมประยุกต์เว็บไอไอเอส

- ข้อมูลการบันทึกการใช้งานบริการเว็บไซต์ข้อมูล ค. หรือข้อมูลบันทึกการใช้งานบริการเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ มีตำแหน่งปริมาณข้อมูลอยู่ตำแหน่งที่ 11 ตำแหน่งวันที่เวลาที่มีการร้องขอบริการเว็บอยู่ตำแหน่งที่ 1 ตำแหน่งเลขที่อยู่ไอพีผู้ให้บริการอยู่ตำแหน่งที่ 3 ตำแหน่งเลขที่อยู่ไอพีผู้ให้บริการอยู่ตำแหน่งที่ 5 ชนิดของแฟ้มข้อมูลเป็นแฟ้มบันทึกการใช้งานมาจากโปรแกรมประยุกต์เว็บไอไอเอส

4.2.3 การตรวจหาค่าผิดปรกติ เป็นการตรวจหาค่าผิดปรกติของข้อมูลบันทึกการใช้งานบริการเว็บไซต์ตามแฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ

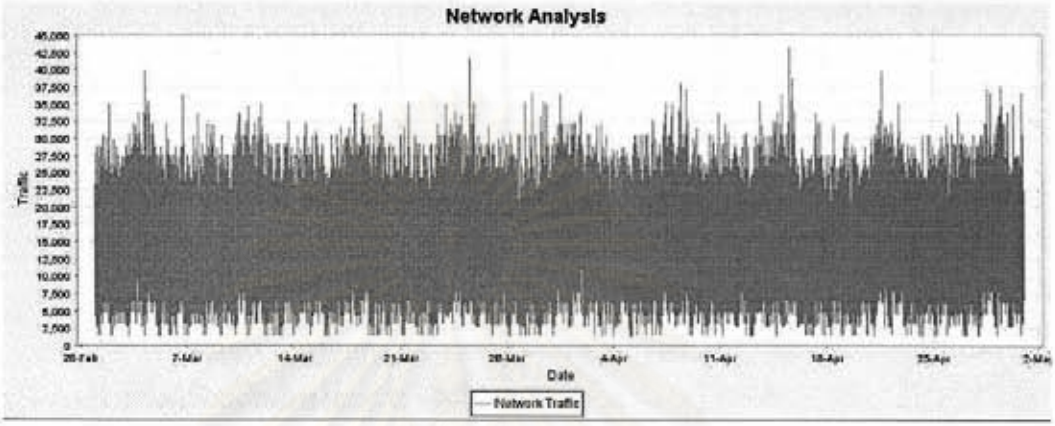
java –jar Discorder.jar แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ

โดยแฟ้มโครงแบบดังกล่าวจะมีลักษณะที่เหมือนกันทั้งหมด เนื่องมาจากว่าข้อมูลดังกล่าวถูกแปลงให้อยู่ในรูปที่มีเพียงข้อมูลที่นำมาใช้ในการวิเคราะห์เท่านั้น ดังแสดงในตารางที่ ค.1 ในภาคผนวก ค.

4.3 ผลการวิจัย

4.3.1 ผลการสร้างข้อมูลบันทึกการเข้าใช้งานของเว็บไซต์

- เมื่อสร้างข้อมูลบันทึกการใช้งานของเว็บไซต์ตามแฟ้มโครงแบบ A แล้ว จึงสร้างแผนภูมิกราฟดังแสดงในรูปที่ 4.1 เพื่อตรวจสอบว่าข้อมูลที่สร้างขึ้นนั้นเป็นไปตามที่กำหนดไว้ในแฟ้มโครงแบบ A หรือไม่

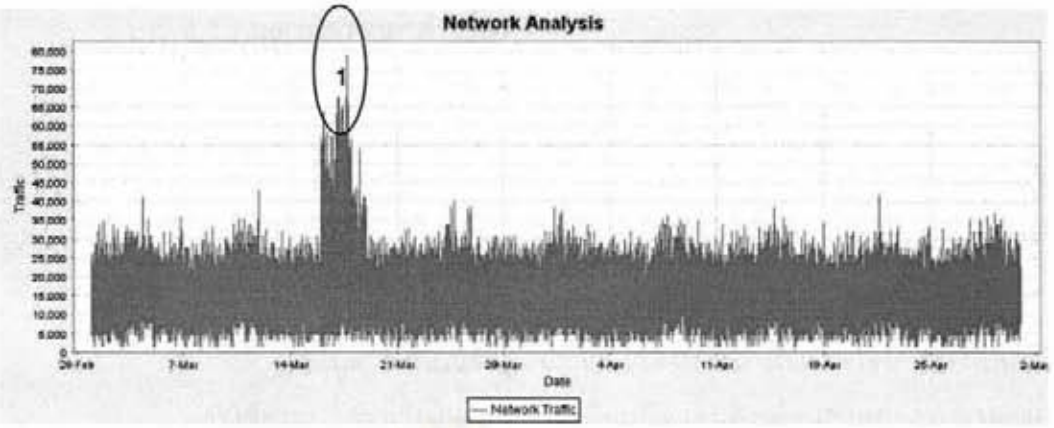


รูปที่ 4.1 แผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งาน
ที่ถูกสร้างขึ้นจากโปรแกรมตามแฟ้มโครงแบบ A

จากรูปที่ 4.1 แสดงให้เห็นถึงแผนภูมิกราฟที่มีเพียงการสุ่มปริมาณการใช้งานตามแฟ้มโครงแบบ A ซึ่งไม่มีการใส่ค่าผิดปรกติลงไปในข้อมูลดังกล่าว

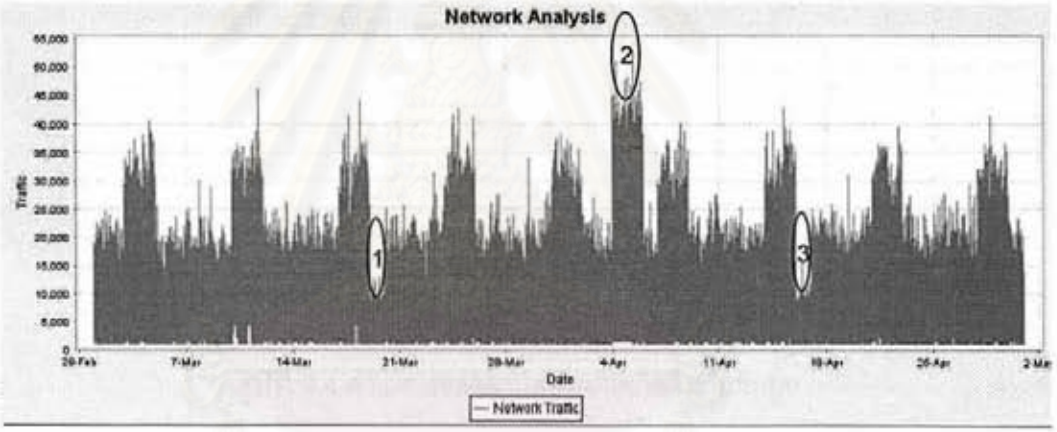
- เมื่อสร้างข้อมูลบันทึกการใช้งานของเว็บไซต์ตามแฟ้มโครงแบบ B แล้ว จึงสร้างแผนภูมิกราฟดังแสดงในรูปที่ 4.2 เพื่อตรวจสอบว่าข้อมูลที่สร้างขึ้นนั้นเป็นไปตามที่กำหนดไว้ในแฟ้มโครงแบบ B หรือไม่

จากรูปที่ 4.2 แสดงให้เห็นถึงแผนภูมิกราฟที่มีเพียงการสุ่มปริมาณการใช้งานตามแฟ้มโครงแบบ B ซึ่งมีการใส่ค่าผิดปรกติลงไปในข้อมูลดังกล่าว 1 ตำแหน่ง



รูปที่ 4.2 แผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งาน
ที่ถูกสร้างขึ้นจากโปรแกรมตามแฟ้มโครงแบบ B

- เมื่อสร้างข้อมูลบันทึกการใช้งานของเว็บไซต์ตามแฟ้มโครงแบบ C แล้ว จึงสร้างแผนภูมิกราฟดังแสดงในรูปที่ 4.3 เพื่อตรวจสอบว่าข้อมูลที่สร้างขึ้นนั้นเป็นไปตามที่กำหนดไว้ในแฟ้มโครงแบบ C หรือไม่



รูปที่ 4.3 แผนภูมิกราฟจากข้อมูลบันทึกการเข้าใช้งาน
ที่ถูกสร้างขึ้นจากโปรแกรมตามแฟ้มโครงแบบ C

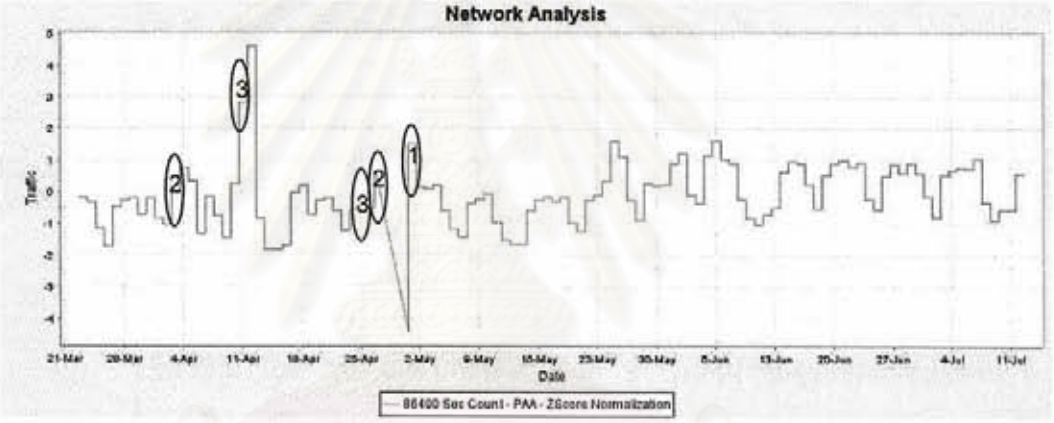
จากรูปที่ 4.3 แสดงให้เห็นถึงแผนภูมิกราฟที่มีเพียงการสุ่มปริมาณการใช้งานตามแฟ้มโครงแบบ C ซึ่งมีการใส่ค่าผิดปรกติลงไปในข้อมูลดังกล่าว 3 ตำแหน่ง

4.3.2 ผลการตรวจหาค่าผิดปรกติ

ในการตรวจหาค่าผิดปรกตินั้น มีความจำเป็นที่จะต้องใช้ค่าพารามิเตอร์ต่างๆ ที่ผ่านการวิเคราะห์มาแล้วจากการดำเนินงานวิจัย โดยได้ผลดังต่อไปนี้

4.3.2.1 ผลการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องของข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่มีการใช้งานจริง

- ข้อมูล A ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนมีนาคมถึงเดือนกรกฎาคม สามารถตรวจหาค่าผิดปรกติได้ที่การลดมิติของข้อมูลด้วยวิธี PAA 24 ชั่วโมง ทั้งจากวิธีทุกความเป็นไปได้ และวิธีที่ผู้เขียนดัดแปลง ซึ่งเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.4



รูปที่ 4.4 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล A จากการตรวจหาค่าผิดปรกติทั้งสองวิธี

จากรูปที่ 4.4 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 1 วัน พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล A ซึ่งมีทั้งหมด 3 อันดับ 5 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.2

ตารางที่ 4.2 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล A

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	11,556	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
1	มีหนึ่งช่วงเวลา คือวันที่ 30 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายโมง ห้าสิบเจ็ดนาที ยี่สิบวินาที ถึงวันที่ 1 พฤษภาคม ค.ศ. 2006 ช่วงเวลาบ่ายโมง ห้าสิบเจ็ดนาที ยี่สิบวินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่หนึ่ง คือวันที่ 3 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 4 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที และช่วงเวลาที่สอง คือ วันที่ 26 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 27 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที		
3	มีสองช่วงเวลา โดยช่วงเวลาที่หนึ่ง คือวันที่ 11 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 12 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที และช่วงเวลาที่สอง คือ วันที่ 25 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 26 เมษายน ค.ศ. 2006 ช่วงเวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที		

จากตารางที่ 4.2 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 24 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 3 พบอันดับของค่าผิดปรกติทั้งหมด 3 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 11,556 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 9 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

- ค่าผิดปรกติอันดับที่ 1 พบว่าเกิดจากการที่โปรแกรมประยุกต์เว็บของจุฬาลงกรณ์มหาวิทยาลัย ไม่ได้ทำการเขียนบันทึกการใช้งานบริการเว็บ

- ค่าผิดปรกติอันดับที่ 2 ที่พบช่วงวันที่ 3 ถึง 4 เมษายน ค.ศ. 2006 นั้นไม่พบความผิดปรกติใดๆ

- ค่าผิดปรกติอันดับที่ 2 ที่พบช่วงวันที่ 26 ถึง 27 เมษายน ค.ศ. 2006 พบว่าเกิดจากการที่โปรแกรมประยุกต์เว็บของจุฬาลงกรณ์มหาวิทยาลัย ไม่ได้ทำการเขียนบันทึกการใช้งานบริการเว็บ

- ค่าผิดปรกติอันดับที่ 3 ที่พบช่วงวันที่ 11 ถึง 12 เมษายน ค.ศ. 2006 พบว่าเกิดจากการที่ช่วงเวลานั้นมีคนดาวน์โหลดไฟล์พีดีเอฟที่มีขนาดใหญ่

- ค่าผิดปรกติอันดับที่ 3 ที่พบช่วงวันที่ 25 ถึง 26 เมษายน ค.ศ. 2006 พบว่าเกิดจากการที่โปรแกรมประยุกต์เว็บของจุฬาลงกรณ์มหาวิทยาลัย ไม่ได้ทำการเขียนบันทึกการใช้งานบริการเว็บ

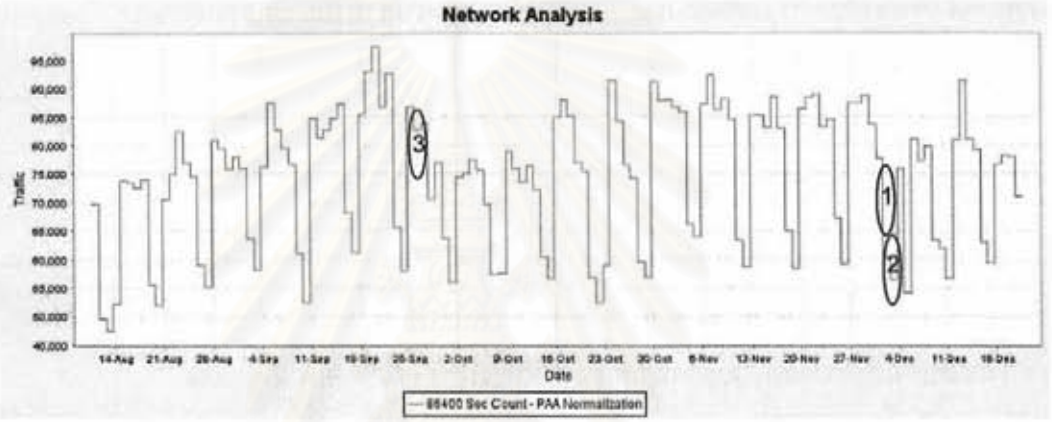
โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ดัดแปลงโดยผู้เขียนสามารถตรวจหาค่าผิดปรกติได้ตำแหน่งเดียวกับวิธีทุกความเป็นไปได้ ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล A

วิธี	ผู้เขียนดัดแปลง	ช่วงเวลา (วัน)	3
จำนวนครั้ง	6,742	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
1	มีหนึ่งช่วงเวลา คือวันที่ 30 เมษายน ค.ศ. 2006 เวลาบ่ายโมง ห้าสิบเจ็ดนาที ยี่สิบวินาที ถึงวันที่ 1 พฤษภาคม ค.ศ. 2006 เวลาบ่ายโมง ห้าสิบเจ็ดนาที ยี่สิบวินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่หนึ่ง คือวันที่ 3 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 4 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที และช่วงเวลาที่สอง คือ วันที่ 26 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 27 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที		
3	มีสองช่วงเวลา โดยช่วงเวลาที่หนึ่ง คือวันที่ 11 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 12 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที และช่วงเวลาที่สอง คือ วันที่ 25 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที ถึงวันที่ 26 เมษายน ค.ศ. 2006 เวลาบ่ายสองโมง ห้าสิบสองนาที สามสิบสองวินาที		

จากตารางที่ 4.3 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลเหมือนกับวิธีทุกความเป็นไปได้ทุกประการ แต่มีจำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าน้อยกว่า

- ข้อมูล B ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนสิงหาคมถึงเดือนธันวาคม สามารถตรวจหาค่าผิดปรกติได้ที่การลดมิติของข้อมูลด้วยวิธี PAA 24 ชั่วโมง ทั้งจากวิธีทุกความเป็นไปได้ และวิธีที่ผู้เขียนดัดแปลง ซึ่งเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.4



รูปที่ 4.5 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล B จากการตรวจหาค่าผิดปรกติทั้งสองวิธี

จากรูปที่ 4.5 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 1 วัน พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล B ซึ่งมีทั้งหมด 3 อันดับ 3 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.4

จากตารางที่ 4.4 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 24 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 3 พบอันดับของค่าผิดปรกติทั้งหมด 3 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 7,364 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 9 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

ตารางที่ 4.4 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล B

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	16,512	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
1	มีหนึ่งช่วงเวลา คือวันที่ 2 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที ถึงวันที่ 3 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที		
2	มีหนึ่งช่วงเวลา คือวันที่ 3 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที ถึงวันที่ 4 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที		
3	มีหนึ่งช่วงเวลา คือวันที่ 26 กันยายน ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที ถึงวันที่ 27 กันยายน ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที		

- ค่าผิดปรกติอันดับที่ 1 และอันดับที่ 2 พบว่าวันที่ 4 และวันที่ 5 ธันวาคม ค.ศ. 2006 มีปริมาณผู้ร้องขอบริการเว็บน้อยกว่าที่เคยเป็นมา อันเนื่องมาจากวันที่ 4 และ 5 ธันวาคม ค.ศ. 2006 เป็นวันหยุดครั้งแรกของเดือนธันวาคม จึงทำให้การตรวจหาค่าผิดปรกติพบว่าค่าผิดปรกติอันดับที่ 1 คือวันที่ 2 ธันวาคม ค.ศ. 2006 และค่าผิดปรกติอันดับที่ 2 คือวันที่ 3 ธันวาคม 2006 ตามลำดับ

- ค่าผิดปรกติอันดับที่ 3 ที่พบช่วงวันที่ 26 ถึง 27 กันยายน ค.ศ. 2006 นั้น ไม่พบความผิดปรกติใดๆ

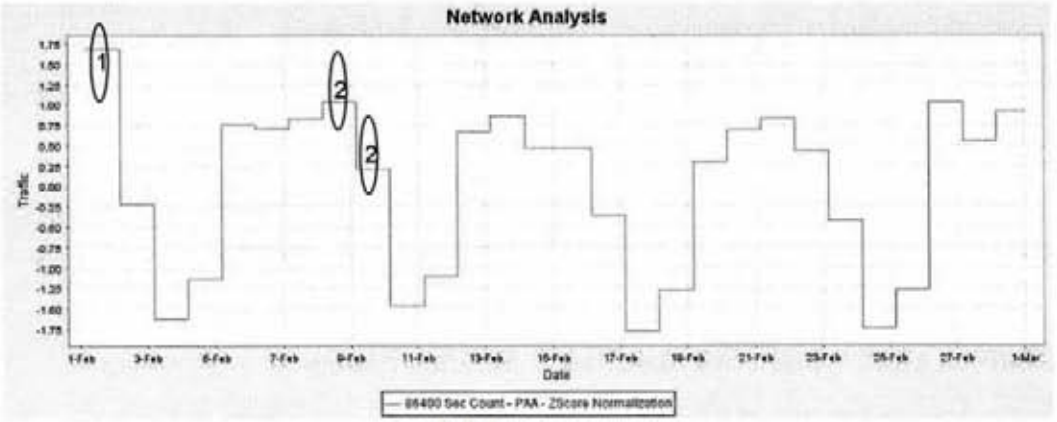
โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ดัดแปลงโดยผู้เขียนสามารถตรวจหาค่าผิดปรกติได้ตำแหน่งเดียวกับวิธีทุกความเป็นไปได้ ดังแสดงในตารางที่ 4.5

ตารางที่ 4.5 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล B

วิธี	ผู้เขียนดัดแปลง	ช่วงเวลา (วัน)	3
จำนวนครั้ง	7,364	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
1	มีหนึ่งช่วงเวลา คือวันที่ 2 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที ถึงวันที่ 3 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที		
2	มีหนึ่งช่วงเวลา คือวันที่ 3 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที ถึงวันที่ 4 ธันวาคม ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที		
3	มีหนึ่งช่วงเวลา คือวันที่ 26 กันยายน ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที ถึงวันที่ 27 กันยายน ค.ศ. 2006 เวลาบ่ายสี่โมง ห้าสิบเอ็ดนาที สี่สิบเก้าวินาที		

จากตารางที่ 4.5 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลเหมือนกับวิธีทุกความเป็นไปได้ทุกประการ แต่มีจำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าน้อยกว่า

- ข้อมูล C เป็นข้อมูลจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการที่มีบันทึกการใช้งานบริการเว็บไซต์เดือนกุมภาพันธ์ สามารถตรวจหาค่าผิดปรกติได้ที่การลดมิติของข้อมูลด้วยวิธี PAA 24 ชั่วโมง จากวิธีทุกความเป็นไปได้เท่านั้น ส่วนการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลงพบค่าผิดปรกติของข้อมูลที่มีการลดมิติของข้อมูลด้วยวิธี PAA 8 ชั่วโมง ซึ่งเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.6



รูปที่ 4.6 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล C
จากการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ที่ผ่านการทำ PAA 24 ชั่วโมง

จากรูปที่ 4.6 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 24 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล C ซึ่งมีทั้งหมด 2 อันดับ 3 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.6

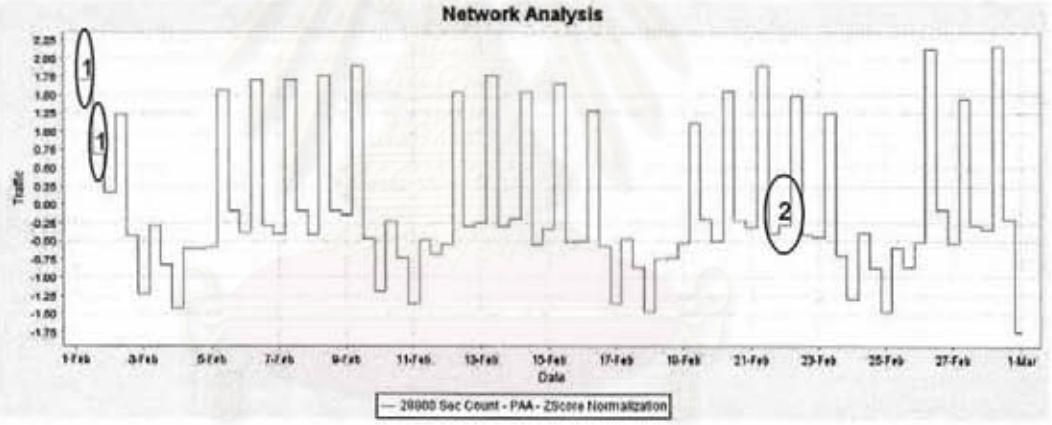
ตารางที่ 4.6 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล C
ที่ผ่านการทำ PAA 24 ชั่วโมง

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	552	อันดับที่พบค่าผิดปรกติ	2
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	7
อันดับที่			
1	มีหนึ่งช่วงเวลา คือวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 2 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 8 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 9 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที และช่วงเวลาที่ 2 คือ วันที่ 9 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 10 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที		

จากตารางที่ 4.6 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 24 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 3 พบอันดับของค่าผิดปรกติทั้งหมด 2 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 552 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 7 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

. ค่าผิดปรกติอันดับที่ 1 และอันดับที่ 2 พบว่าวันที่ 1 วันที่ 8 และวันที่ 9 กุมภาพันธ์ ค.ศ. 2007 ไม่มีค่าผิดปรกติใดๆ แต่เนื่องจากการตรวจหาค่าผิดปรกติเป็นการคำนวณทางคณิตศาสตร์จึงสามารถตรวจหาค่าผิดปรกติได้

โดยวิธีที่ผู้เขียนดัดแปลงพบค่าผิดปรกติพบค่าผิดปรกติที่ได้มาจากการลดมิติของข้อมูลด้วยวิธี PAA 8 ชั่วโมง ซึ่งวิธีทุกความเป็นไปได้ก็พบค่าผิดปรกติตำแหน่งเดียวกันเช่นกัน เพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.7



รูปที่ 4.7 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล C
จากการตรวจหาค่าผิดปรกติทั้งสองวิธีที่ผ่านการทำ PAA 8 ชั่วโมง

จากรูปที่ 4.7 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 8 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล C ซึ่งมีทั้งหมด 2 อันดับ 3 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.7

ตารางที่ 4.7 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล C ที่ผ่านการทำ PAA 8 ชั่วโมง

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	1
จำนวนครั้ง	1,892	อันดับที่พบค่าผิดปรกติ	2
ขนาด PAA	8 ชั่วโมง	ขนาดของสัญลักษณ์	10
อันดับที่			
1	มีสองช่วงเวลา โดยช่วงเวลาที่หนึ่ง คือวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเอ็ดนาฬิกา ห้าสิบสามนาที หนึ่งวินาที และช่วงเวลาที่สอง คือวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเอ็ดนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเก้านาฬิกา ห้าสิบสามนาที หนึ่งวินาที		
2	มีหนึ่งช่วงเวลา คือวันที่ 22 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 22 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเอ็ดนาฬิกา ห้าสิบสามนาที หนึ่งวินาที		

จากตารางที่ 4.7 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 8 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 1 พบอันดับของค่าผิดปรกติทั้งหมด 2 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 1,892 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 10 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

- ค่าผิดปรกติอันดับที่ 1 และอันดับที่ 2 พบว่าวันที่ 1 และวันที่ 22 กุมภาพันธ์ ค.ศ. 2007 ไม่ได้มีค่าผิดปรกติใดๆ แต่เนื่องจากการตรวจหาค่าผิดปรกติเป็นการคำนวณทางคณิตศาสตร์จึงสามารถตรวจหาค่าผิดปรกติได้

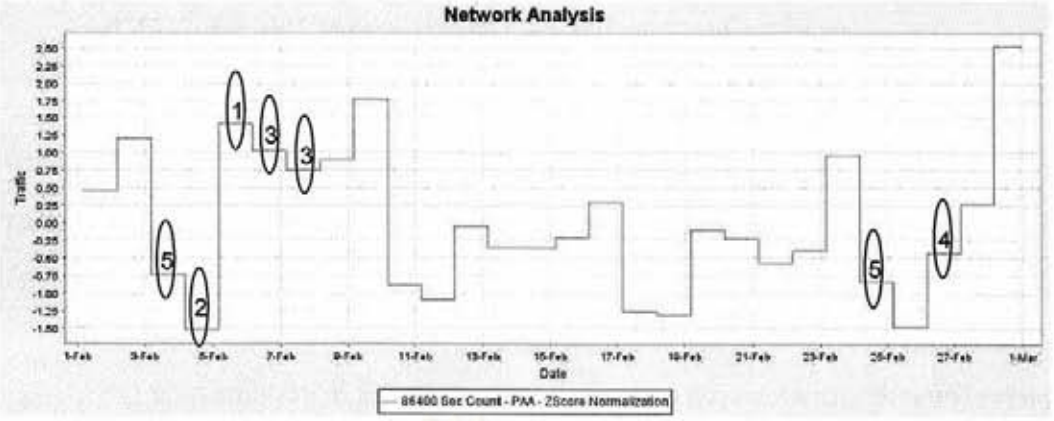
โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ดัดแปลงโดยผู้เขียนสามารถตรวจหาค่าผิดปรกติได้ตำแหน่งเดียวกับวิธีทุกความเป็นไปได้ ดังแสดงในตารางที่ 4.8

ตารางที่ 4.8 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล C ที่ผ่านการทำ PAA 8 ชั่วโมง

วิธี	ผู้เขียนดัดแปลง	ช่วงเวลา (วัน)	1
จำนวนครั้ง	1,892	อันดับที่พบค่าผิดปรกติ	2
ขนาด PAA	8 ชั่วโมง	ขนาดของสัญลักษณ์	10
อันดับที่			
1	มีสองช่วงเวลา โดยช่วงเวลาที่หนึ่ง คือวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเอ็ดนาฬิกา ห้าสิบสามนาที หนึ่งวินาที และช่วงเวลาที่สอง คือวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเอ็ดนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเก้านาฬิกา ห้าสิบสามนาที หนึ่งวินาที		
2	มีหนึ่งช่วงเวลา คือวันที่ 22 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบสามนาที หนึ่งวินาที ถึงวันที่ 22 กุมภาพันธ์ ค.ศ. 2007 เวลาสิบเอ็ดนาฬิกา ห้าสิบสามนาที หนึ่งวินาที		

จากตารางที่ 4.8 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลเหมือนกับวิธีทุกความเป็นไปได้ทุกประการ โดยในครั้งนี้จำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าเท่ากับวิธีทุกความเป็นไปได้

- ข้อมูล D เป็นข้อมูลจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการที่มีบันทึกการใช้งานบริการเว็บไซต์เดือนกุมภาพันธ์ สามารถตรวจหาค่าผิดปรกติได้ที่การลดมิติของข้อมูลด้วยวิธี PAA 24 ชั่วโมง จากวิธีทุกความเป็นไปได้และวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ตำแหน่งของค่าผิดปรกติที่แตกต่างกัน โดยเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.8



รูปที่ 4.8 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล D จากการตรวจหาค่าผิดปรกติวิธีทุกความเป็นไปได้

จากรูปที่ 4.8 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 24 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล D ซึ่งมีทั้งหมด 5 อันดับ 7 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.9

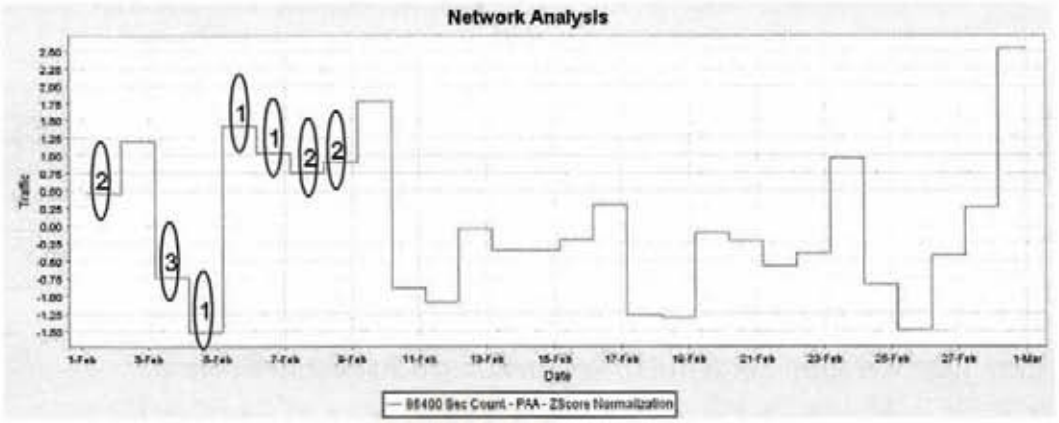
จากตารางที่ 4.9 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 24 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 3 วัน พบอันดับของค่าผิดปรกติทั้งหมด 5 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 552 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 7 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

- ค่าผิดปรกติทุกอันดับพบว่า ณ ช่วงเวลาดังกล่าว มีปริมาณการใช้งานช่วงวันที่ 5 ถึงวันที่ 10 กุมภาพันธ์ ค.ศ. 2006 ที่มีปริมาณการใช้งานที่แตกต่างกว่าช่วงเวลาอื่น โดยมีปริมาณการใช้งานที่ใกล้เคียงกันมาก และวันที่ 4 กุมภาพันธ์ ค.ศ. 2006 มีการเปลี่ยนแปลงปริมาณการใช้งานที่แตกต่างจากวันอื่นๆ มากที่สุด โดยจากการตรวจสอบพบว่าไม่มีความผิดปรกติใดๆ เกิดขึ้น แต่เนื่องจากการตรวจหาค่าผิดปรกติเป็นการคำนวณทางคณิตศาสตร์จึงสามารถตรวจหาค่าผิดปรกติได้

ตารางที่ 4.9 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล D

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	552	อันดับที่พบค่าผิดปรกติ	5
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	7
อันดับที่			
1	มีหนึ่งช่วงเวลา คือวันที่ 5 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 6 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		
2	มีหนึ่งช่วงเวลา คือวันที่ 4 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 5 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		
3	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 6 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 7 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที และช่วงเวลาที่ 2 คือวันที่ 7 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 8 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		
4	มีหนึ่งช่วงเวลา คือวันที่ 26 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 27 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		
5	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 3 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 4 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที และช่วงเวลาที่ 2 คือวันที่ 25 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 26 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		

โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ผู้เขียนดัดแปลงสามารถตรวจหาค่าผิดปรกติได้ตำแหน่งที่ใกล้เคียงกับวิธีทุกความเป็นไปได้ เนื่องจากมีการเก็บลำดับย่อยที่มีอักขระซ้ำกันไว้ในโครงสร้างแถวลำดับ จึงทำให้ผลที่ได้แตกต่างกัน และเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีที่ผู้เขียนดัดแปลงในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.9



รูปที่ 4.9 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล D
จากการตรวจหาค่าผิดปรกติวิธีที่ผู้เขียนดัดแปลง

จากรูปที่ 4.9 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 24 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล D ซึ่งมีทั้งหมด 3 อันดับ 7 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.10

จากตารางที่ 4.10 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลใกล้เคียงกับวิธีทุกความเป็นไปได้ โดยในครั้งนี้จำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าน้อยกว่าวิธีทุกความเป็นไปได้ ซึ่งผลจากการวิเคราะห์ค่าผิดปรกติพบว่า สามารถหาช่วงเวลาที่ข้อมูลมีปริมาณการใช้งานที่แตกต่างกว่าช่วงเวลาอื่น และมีปริมาณการใช้งานที่ใกล้เคียงกันมากได้เช่นกัน โดยมีรายละเอียดของช่วงเวลาที่แตกต่างกันดังต่อไปนี้

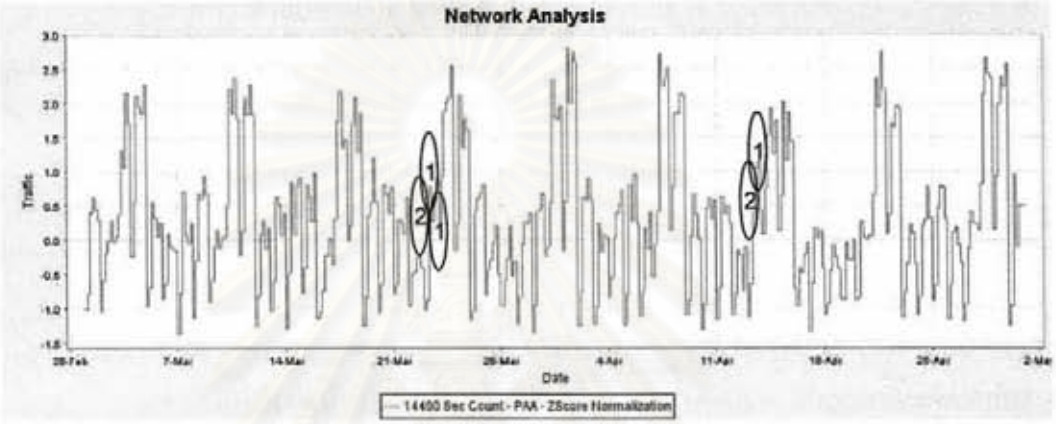
- วิธีที่ผู้เขียนดัดแปลงไม่พบค่าผิดปรกติที่เกิดจากวิธีทุกความเป็นไปได้อันดับที่ 4 คือ วันที่ 27 กุมภาพันธ์ ค.ศ. 2007 และอันดับที่ 5 คือ วันที่ 24 กุมภาพันธ์ ค.ศ. 2007
- วิธีที่ผู้เขียนดัดแปลงพบค่าผิดปรกติอันดับที่ 2 คือ วันที่ 1 กุมภาพันธ์ ค.ศ. 2007 ในขณะที่วิธีทุกความเป็นไปได้ไม่พบค่าผิดปรกติที่ตำแหน่งนี้

ตารางที่ 4.10 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล D

วิธี	ผู้เขียนดัดแปลง	ช่วงเวลา (วัน)	3
จำนวนครั้ง	330	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	6
อันดับที่			
1	มีสามช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 4 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 5 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ช่วงเวลาที่ 2 คือ วันที่ 5 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 6 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที และช่วงเวลาที่ 3 คือ วันที่ 6 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 7 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		
2	มีสามช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 1 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 2 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ช่วงเวลาที่ 2 คือ วันที่ 7 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 8 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที และช่วงเวลาที่ 3 คือ วันที่ 8 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 9 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		
3	มีหนึ่งช่วงเวลา คือวันที่ 3 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที ถึงวันที่ 4 กุมภาพันธ์ ค.ศ. 2007 เวลาสามนาฬิกา ห้าสิบเจ็ดนาที สามสิบสามวินาที		

4.3.2.2 ผลการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องของข้อมูลบันทึกการใช้งานบริการเว็บไซต์ที่ถูกสร้างขึ้น

- ข้อมูล E เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ A ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ สามารถตรวจหาค่าผิดปรกติได้ ซึ่งเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.10



รูปที่ 4.10 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล E จากการตรวจหาค่าผิดปรกติทั้งสองวิธี

จากรูปที่ 4.10 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 4 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล E ซึ่งมีทั้งหมด 2 อันดับ 5 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.11

โดยในตารางที่ 4.11 เป็นผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 4 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 1 วันพบอันดับของค่าผิดปรกติทั้งหมด 2 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 126,380 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 6 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

ตารางที่ 4.11 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล E

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	1
จำนวนครั้ง	126,380	อันดับที่พบค่าผิดปรกติ	2
ขนาด PAA	4 ชั่วโมง	ขนาดของสัญลักษณ์	6
อันดับที่			
1	มีสามช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 23 มีนาคม ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบเอ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ช่วงเวลาที่ 2 คือวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบเอ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบห้านาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที และช่วงเวลาที่ 3 คือวันที่ 13 เมษายน ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบเอ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 23 มีนาคม ค.ศ. 2006 เวลาสามนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที สามสิบสองวินาที และช่วงเวลาที่ 2 คือ วันที่ 13 เมษายน ค.ศ. 2006 เวลาสามนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที		

- ค่าผิดปรกติอันดับที่ 1 และอันดับที่ 2 พบว่าวันที่ 23 มีนาคม ค.ศ. 2006 และ 13 เมษายน ค.ศ. 2006 ณ ช่วงเวลาดังกล่าวไม่ได้มีการใส่ค่าผิดปรกติใดๆ ลงไปในข้อมูลที่ถูกสร้างขึ้นมา และเนื่องจากว่าเป็นการสุ่มข้อมูลจึงทำให้ข้อมูลดังกล่าวมีลักษณะที่ผิดปรกติได้หากมีการคำนวณอย่างละเอียด โดยการตรวจหาค่าผิดปรกติเป็นการคำนวณทางคณิตศาสตร์จึงสามารถตรวจหาค่าผิดปรกติได้

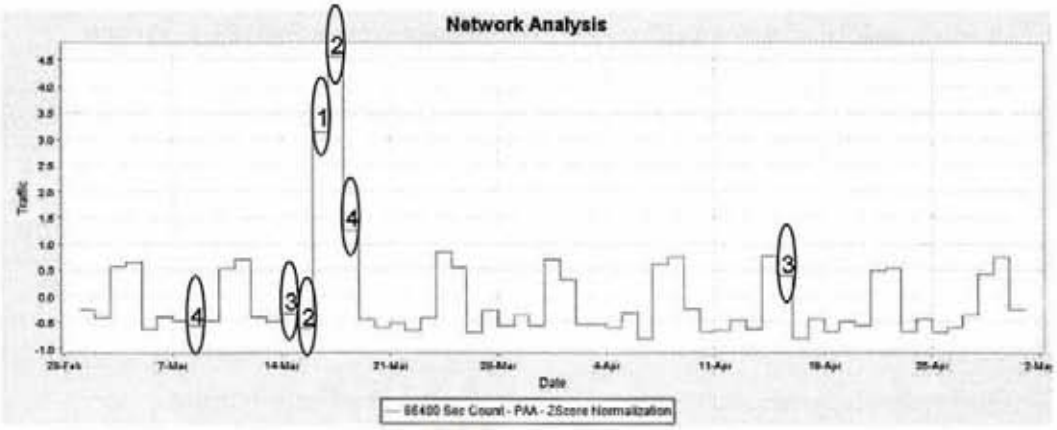
โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ดัดแปลงโดยผู้เขียนสามารถตรวจหาค่าผิดปรกติได้ตำแหน่งเดียวกับวิธีทุกความเป็นไปได้ ดังแสดงในตารางที่ 4.12

ตารางที่ 4.12 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล E

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	1
จำนวนครั้ง	86,324	อันดับที่พบค่าผิดปรกติ	2
ขนาด PAA	4 ชั่วโมง	ขนาดของสัญลักษณ์	6
อันดับที่			
1	มีสามช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 23 มีนาคม ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบเอ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ช่วงเวลาที่ 2 คือวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบเอ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบห้านาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที และช่วงเวลาที่ 3 คือวันที่ 13 เมษายน ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาสิบเอ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 23 มีนาคม ค.ศ. 2006 เวลาสามนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที สามสิบสองวินาที และช่วงเวลาที่ 2 คือ วันที่ 13 เมษายน ค.ศ. 2006 เวลาสามนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที ถึงวันที่ 23 มีนาคม ค.ศ. 2006 เวลาเจ็ดนาฬิกา ห้าสิบเก้านาที ห้าสิบสามวินาที		

จากตารางที่ 4.12 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลเหมือนกับวิธีทุกความเป็นไปได้ทุกประการ โดยจำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าน้อยกว่า

- ข้อมูล F เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ B ที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A คือ การใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มลักษณะของรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานมากกว่าปรกติ สามารถตรวจหาค่าผิดปรกติได้ ซึ่งเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.11



รูปที่ 4.11 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล F
จากการตรวจหาค่าผิดปรกติทั้งสองวิธี

จากรูปที่ 4.11 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 24 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล F ซึ่งมีทั้งหมด 4 อันดับ 7 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.13

ตารางที่ 4.13 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล F

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	3,192	อันดับที่พบค่าผิดปรกติ	4
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
1	มีหนึ่งช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 16 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 17 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 15 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 16 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที และช่วงเวลาที่ 2 คือ วันที่ 17 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 18 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		

ตารางที่ 4.13 (ต่อ) ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล F

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	3,192	อันดับที่พบค่าผิดปรกติ	4
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
3	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 14 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 15 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ช่วงเวลาที่ 2 คือ วันที่ 15 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 16 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
4	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 8 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 9 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที และช่วงเวลาที่ 2 คือ วันที่ 18 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 19 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		

จากตารางที่ 4.13 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 24 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 3 วัน พบอันดับของค่าผิดปรกติทั้งหมด 4 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 3,192 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 9 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

- ค่าผิดปรกติอันดับที่ 1 ที่ช่วงเวลาที่ 1 คือวันที่ 16 มีนาคม ค.ศ. 2006 อันดับที่ 2 ที่ช่วงเวลาที่ 2 คือ วันที่ 15 มีนาคม ค.ศ. 2006 และ วันที่ 17 มีนาคม ค.ศ. 2006 อันดับที่ 3 คือ วันที่ 14 มีนาคม ค.ศ. 2006 และอันดับที่ 4 คือวันที่ 18 มีนาคม 2006 พบว่าเกิดจากการเพิ่มลักษณะของรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006

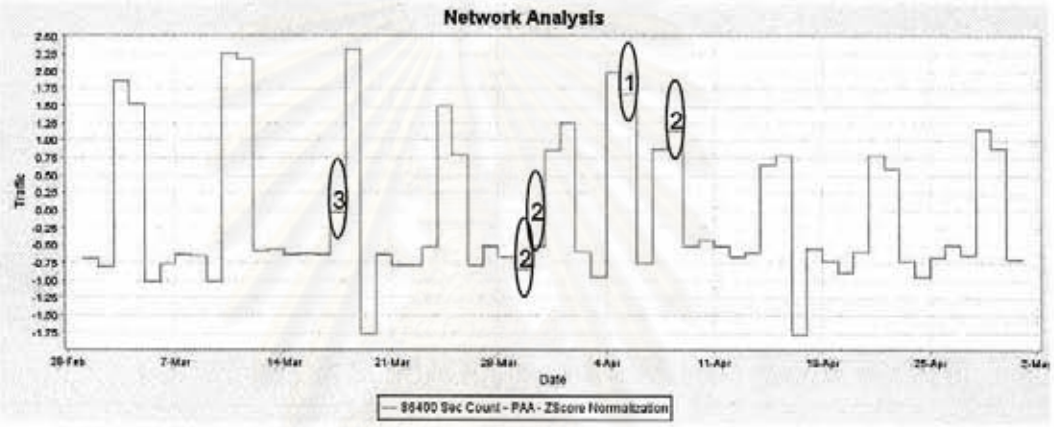
โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ดัดแปลงโดยผู้เขียน สามารถตรวจหาค่าผิดปรกติได้ตำแหน่งเดียวกับวิธีทุกความเป็นไปได้ ดังแสดงในตารางที่ 4.14

ตารางที่ 4.14 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล F

วิธี	ผู้เขียนดัดแปลง	ช่วงเวลา (วัน)	3
จำนวนครั้ง	2,220	อันดับที่พบค่าผิดปรกติ	4
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	9
อันดับที่			
1	มีหนึ่งช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 16 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 17 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
2	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 15 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 16 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที และช่วงเวลาที่ 2 คือ วันที่ 17 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 18 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
3	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 14 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 15 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ช่วงเวลาที่ 2 คือ วันที่ 15 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 16 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
4	มีสองช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 8 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 9 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที และช่วงเวลาที่ 2 คือ วันที่ 18 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 19 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		

จากตารางที่ 4.14 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลเหมือนกับวิธีทุกความเป็นไปได้ทุกประการ โดยจำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าน้อยกว่า

- ข้อมูล G เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ C ที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A คือ การใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มตำแหน่งข้อมูลที่ผิดปรกติไว้ 3 ตำแหน่ง ได้แก่ ตำแหน่งที่ 1 คือวันที่ 19 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ตำแหน่งที่ 2 คือวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บมากกว่าปรกติ และตำแหน่งที่ 3 คือวันที่ 16 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ สามารถตรวจหาค่าผิดปรกติได้ ซึ่งเพื่อให้ง่ายต่อการพิจารณาค่าผิดปรกติจึงได้มีการสร้างแผนภูมิกราฟ พร้อมบอกตำแหน่งของค่าผิดปรกติที่ค้นพบจากวิธีทุกความเป็นไปได้ในแผนภูมิกราฟ ดังแสดงในรูปที่ 4.12



รูปที่ 4.12 ตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล F จากการตรวจหาค่าผิดปรกติทั้งสองวิธี

จากรูปที่ 4.12 แสดงให้เห็นการตรวจหาค่าผิดปรกติที่ข้อมูลอนุกรมเวลาผ่านการแปลง PAA ที่ระดับ 24 ชั่วโมง พร้อมทั้งตำแหน่งของค่าผิดปรกติที่เกิดขึ้นกับข้อมูล F ซึ่งมีทั้งหมด 3 อันดับ 5 ตำแหน่ง ซึ่งรายละเอียดของผลการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้แสดงในตารางที่ 4.15

ตารางที่ 4.15 ผลการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ของข้อมูล G

วิธี	ทุกความเป็นไปได้	ช่วงเวลา (วัน)	3
จำนวนครั้ง	3,192	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	10
อันดับที่			
1	มีหนึ่งช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 5 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 6 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
2	มีสามช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 29 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 30 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ช่วงเวลาที่ 2 คือ วันที่ 30 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 31 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที และช่วงเวลาที่ 3 คือ วันที่ 8 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 9 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
3	มีหนึ่งช่วงเวลา คือ วันที่ 17 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 18 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		

จากตารางที่ 4.15 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีทุกความเป็นไปได้ ที่มีช่วงเวลาที่ลดมิติของข้อมูลด้วยการแปลง PAA ที่ระดับ 24 ชั่วโมง จำนวนของสัญลักษณ์ที่ใช้ในการตรวจหาค่าผิดปรกติที่มีค่าเท่ากับ 3 วัน พบอันดับของค่าผิดปรกติทั้งหมด 3 อันดับด้วยกัน โดยใช้จำนวนครั้งในการวนรอบในการตรวจหาค่าผิดปรกติ 3,192 ครั้ง ซึ่งขนาดของสัญลักษณ์ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาด้วยวิธีแซ็คเป็น 10 และมีการวิเคราะห์ค่าผิดปรกติที่เกิดขึ้นดังต่อไปนี้

- ค่าผิดปรกติอันดับที่ 1 คือวันที่ 5 เมษายน ค.ศ. 2006 พบว่าเกิดจากการเพิ่มปริมาณการใช้งานเว็บมากกว่าปรกติในวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน และอันดับที่ 3 คือ วันที่ 17 มีนาคม ค.ศ. 2006 พบว่าเกิดจากการปริมาณการใช้งานเว็บน้อยกว่าปรกติในวันที่ 19 มีนาคม ค.ศ. 2006 ทั้งวัน ทั้งยังพบว่าค่าผิดปรกติอันดับที่ 2 นั้นไม่ได้เกิดจากการสร้างข้อมูลที่มีค่าผิดปรกติแต่อย่างใด ซึ่งเกิดจากการคำนวณทางคณิตศาสตร์ที่มีระยะห่างมากเป็นอันดับ 2

และปริมาณการใช้งานเว็บที่น้อยกว่าปรกติในวันที่ 16 เมษายน ค.ศ. 2006 ก็ไม่สามารถตรวจพบค่าผิดปรกติได้เช่นกัน ซึ่งเกิดจากการคำนวณทางคณิตศาสตร์แล้วพบว่าปริมาณการใช้งานจากมากไปน้อยที่มีลักษณะเหมือนข้อมูลวันที่ 15 มีนาคม ค.ศ. 2006 ถึงวันที่ 16 มีนาคม 2006 มีเป็นจำนวนมาก

โดยพบว่าผลของการตรวจหาค่าผิดปรกติด้วยวิธีที่ดัดแปลงโดยผู้เขียนสามารถตรวจหาค่าผิดปรกติได้ตำแหน่งเดียวกับวิธีทุกความเป็นไปได้ ดังแสดงในตารางที่ 4.16

ตารางที่ 4.16 ผลการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลงของข้อมูล G

วิธี	ผู้เขียนดัดแปลง	ช่วงเวลา (วัน)	3
จำนวนครั้ง	1,672	อันดับที่พบค่าผิดปรกติ	3
ขนาด PAA	24 ชั่วโมง	ขนาดของสัญลักษณ์	10
อันดับที่			
1	มีหนึ่งช่วงเวลา โดยช่วงเวลาที่ 1 คือวันที่ 5 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 6 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
2	มีสามช่วงเวลา โดยช่วงเวลาที่ 1 คือ วันที่ 29 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 30 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ช่วงเวลาที่ 2 คือ วันที่ 30 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 31 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที และช่วงเวลาที่ 3 คือ วันที่ 8 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 9 เมษายน ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		
3	มีหนึ่งช่วงเวลา คือ วันที่ 17 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที ถึงวันที่ 18 มีนาคม ค.ศ. 2006 เวลาศูนย์นาฬิกา ศูนย์นาที ศูนย์วินาที		

จากตารางที่ 4.16 แสดงผลจากการตรวจหาค่าผิดปรกติจากวิธีที่ผู้เขียนดัดแปลง ซึ่งได้ผลเหมือนกับวิธีทุกความเป็นไปได้ทุกประการ โดยจำนวนครั้งที่ใช้ในการตรวจหาค่าผิดปรกติมีค่าน้อยกว่า

4.3.2.3 ผลการตรวจหาค่าผิดปรกติจากข้อมูลแบบต่อเนื่อง

ในการตรวจหาค่าผิดปรกติจากข้อมูลแบบต่อเนื่องนั้น ได้มีการกำหนดจำนวนวันที่เก็บข้อมูลเบื้องต้นของข้อมูลแบบไม่ต่อเนื่องไว้ 3 จำนวนด้วยกันคือ 7 วัน 15 วัน และ 30 วัน ซึ่งได้มีการเปรียบเทียบระหว่างผลของการตรวจหาค่าผิดปรกติของช่วงเวลาดังกล่าว กับผลของการตรวจหาค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่องจากทุกความเป็นไปได้ โดยการใช้ตำแหน่งค่าผิดปรกติที่เกิดขึ้นจากข้อมูลแบบไม่ต่อเนื่อง กับตำแหน่งค่าผิดปรกติที่ตรวจพบจากจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่องต่างๆ โดยกำหนดวิธีการตรวจหาค่าผิดปรกติ BF คือ วิธีทุกความเป็นไปได้ และ MB คือ วิธีที่ผู้เขียนดัดแปลง ซึ่งข้อมูลที่จะนำมาตรวจหาค่าผิดปรกติแบบไม่ต่อเนื่องนั้นได้ผ่านการวิเคราะห์ว่าต้องมีความผิดปรกติที่เกิดขึ้นจริงจากปริมาณใช้งาน ซึ่งได้แก่

- ข้อมูล A ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนมีนาคมถึงเดือนกรกฎาคม
- ข้อมูล B ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนสิงหาคมถึงเดือนธันวาคม
- ข้อมูล D เป็นข้อมูลจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการที่มีบันทึกการใช้งานบริการเว็บไซต์เดือนกุมภาพันธ์
- ข้อมูล F เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ B ที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A คือ การใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มลักษณะของรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานมากกว่าปรกติ
- ข้อมูล G เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ C ที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A คือ การใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มตำแหน่งข้อมูลที่ผิดปรกติไว้ 3 ตำแหน่ง ได้แก่ ตำแหน่งที่ 1 คือวันที่ 19 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ตำแหน่งที่ 2 คือวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บมากกว่าปรกติ และตำแหน่งที่ 3 คือวันที่ 16 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ

เมื่อดำเนินการตรวจหาค่าผิดปรกติแล้ว ได้ผลดังต่อไปนี้

- ข้อมูล A ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนมีนาคมถึงเดือนกรกฎาคม พบว่ามีตำแหน่งค่าผิดปรกติที่ควรพิจารณา 3 อันดับ 3 ตำแหน่งด้วยกัน ได้แก่ ตำแหน่งที่ 1 คือวันที่ 30 เม.ย. ค.ศ. 2006 ที่ข้อมูลบันทึกการใช้งานบริการเว็บหายไป ตำแหน่งที่ 2 คือวันที่ 26 เม.ย. ค.ศ. 2006 ที่ข้อมูลบันทึกการใช้งานบริการเว็บหายไป และตำแหน่งที่ 3 คือวันที่ 11 เม.ย. ค.ศ. 2006 ที่มีการดาวน์โหลดไฟล์พีดีเอฟขนาดใหญ่ มีช่วงเวลาของการลดมิติของข้อมูลด้วยวิธี PAA ที่เลือกเป็นเกณฑ์ในการตรวจหาค่าผิดปรกติเป็นช่วงเวลาเดียวกับค่าผิดปรกติที่พบจากข้อมูลแบบไม่ต่อเนื่อง คือ 24 ชั่วโมง และพบว่าการตรวจหาค่าผิดปรกติของจำนวนวันที่ถูกแบ่งออกมา เพื่อทำให้เป็นข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาค่าผิดปรกติได้ในบางช่วงเวลา ดังแสดงในตารางที่ 4.17

ตารางที่ 4.17 ผลการตรวจหาค่าผิดปรกติของข้อมูล A ที่จำลองเป็นข้อมูลแบบต่อเนื่องจากการตรวจหาค่าผิดปรกติทั้งสองวิธี

ค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง		ข้อมูลเบื้องต้น		วิธี	อันดับที่พบ	จำนวนครั้ง
อันดับที่	วันที่	จำนวนวัน	วันที่เริ่ม			
1	30 เม.ย. 2006	7	26 เม.ย. 2006	BF	3	12
			26 เม.ย. 2006	MB	3	12
		15	21 เม.ย. 2006	BF	2	12
			21 เม.ย. 2006	MB	2	12
		30	2 เม.ย. 2006	BF	2	702
			2 เม.ย. 2006	MB	2	596
2	26 เม.ย. 2006	7	24 เม.ย. 2006	BF	4	12
			24 เม.ย. 2006	MB	4	12
		15	18 เม.ย. 2006	BF	2	12
			17 เม.ย. 2006	MB	2	12
		30	12 เม.ย. 2006	BF	5	702
			12 เม.ย. 2006	MB	5	646

ตารางที่ 4.17 (ต่อ) ผลการตรวจหาค่าผิดปรกติของข้อมูล A ที่จำลองเป็นข้อมูลแบบต่อเนื่องจากการตรวจหาค่าผิดปรกติทั้งสองวิธี

ค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง		ข้อมูลเบื้องต้น		วิธี	อันดับที่พบ	จำนวนครั้ง
อันดับที่	วันที่	จำนวนวัน	วันที่เริ่ม			
3	11 เม.ย. 2006	7	8 เม.ย. 2006	BF	2	12
			8 เม.ย. 2006	MB	2	12
		15	-	BF	-	-
			-	MB	-	-
		30	24 มี.ค. 2006	BF	5	702
			24 มี.ค. 2006	MB	5	702

จากตารางที่ 4.17 พบว่าตำแหน่งค่าผิดปรกติที่พบอันดับที่ 1 และ 3 ที่เกิดจากการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาได้จากจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาจากข้อมูลแบบไม่ต่อเนื่อง ส่วนในกรณีค่าผิดปรกติอันดับที่ 2 นั้น จำนวนวันที่เก็บข้อมูลเบื้องต้นเป็น 15 วัน ไม่สามารถตรวจพบว่าเป็นค่าผิดปรกติอันดับแรกได้ ทั้งจากวิธีทุกความเป็นไปได้และวิธีที่ผู้เขียนดัดแปลง ทั้งยังพบว่าค่าผิดปรกติที่พบทุกครั้งนั้นสามารถหาได้จากทั้งสองวิธีและจำนวนครั้งของวิธีที่ผู้เขียนดัดแปลงมีค่าน้อยกว่าหรือเท่ากับวิธีทุกความเป็นไปได้อีกด้วย

- จากข้อมูล B ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนสิงหาคมถึงเดือนธันวาคม พบว่ามีตำแหน่งค่าผิดปรกติที่ควรพิจารณา 2 อันดับ 2 ตำแหน่งด้วยกัน ได้แก่ ตำแหน่งที่ 1 คือวันที่ 2 ธันวาคม ค.ศ. 2006 และตำแหน่งที่ 2 คือวันที่ 3 ธันวาคม ค.ศ. 2006 ที่ค้นพบจากปริมาณผู้ร้องขอบริการเว็บน้อยกว่าที่เคยเป็นมา อันเนื่องมาจากวันที่ 4 และ 5 ธันวาคม ค.ศ. 2006 เป็นวันหยุดครั้งแรกของเดือนธันวาคม และช่วงเวลาของการลดมิติของข้อมูลด้วยวิธี PAA ที่เลือกเป็นเกณฑ์ในการตรวจหาค่าผิดปรกติเป็นช่วงเวลาเดียวกับค่าผิดปรกติที่พบจากข้อมูลแบบไม่ต่อเนื่อง คือ 24 ชั่วโมง มีการตรวจหาค่าผิดปรกติของจำนวนวันที่ถูกแบ่งออกมา

เพื่อทำให้เป็นข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาค่าผิดปรกติได้ในบางช่วงเวลา ดังแสดงในตารางที่ 4.18

ตารางที่ 4.18 ผลการตรวจหาค่าผิดปรกติของข้อมูล B ที่จำลองเป็นข้อมูลแบบต่อเนื่องจากการตรวจหาค่าผิดปรกติทั้งสองวิธี

ค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง		ข้อมูลเบื้องต้น		วิธี	อันดับที่พบ	จำนวนครั้ง
อันดับที่	วันที่	จำนวนวัน	วันที่เริ่ม			
1	2 ธ.ค. 2006	7	30 พ.ย. 2006	BF	2	12
			30 พ.ย. 2006	MB	2	12
		15	30 พ.ย. 2006	BF	5	132
			30 พ.ย. 2006	MB	5	132
		30	4 พ.ย. 2006	BF	3	702
			5 พ.ย. 2006	MB	3	592
2	3 ธ.ค. 2006	7	-	BF	-	-
			-	MB	-	-
		15	1 ธ.ค. 2006	BF	3	132
			1 ธ.ค. 2006	MB	3	132
		30	10 พ.ย. 2006	BF	4	702
			12 พ.ย. 2006	MB	2	418

จากตารางที่ 4.18 พบว่าตำแหน่งค่าผิดปรกติอันดับที่ 1 ที่เกิดจากการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาได้จากจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาจากข้อมูลแบบไม่ต่อเนื่อง ส่วนในกรณีค่าผิดปรกติอันดับที่ 2 นั้น จำนวนวันที่เก็บข้อมูลเบื้องต้นเป็น 7 วัน ไม่สามารถตรวจพบว่าเป็นค่าผิดปรกติอันดับแรกได้ ทั้งจากวิธีทุกความเป็นไปได้และวิธีที่ผู้เขียนดัดแปลง ทั้งยังพบว่าค่าผิดปรกติที่พบทุกครั้งนั้น สามารถหาได้จากทั้งสองวิธีและจำนวนครั้งของวิธีที่ผู้เขียนดัดแปลงมีค่าน้อยกว่าหรือเท่ากับวิธีทุกความเป็นไปได้อีกด้วย

- จากข้อมูล D ที่เป็นข้อมูลจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการที่มีบันทึกการใช้งานบริการเว็บไซต์เดือนกุมภาพันธ์ พบว่ามีตำแหน่งค่าผิดปรกติที่ควรพิจารณา 1 อันดับ 2 ตำแหน่งด้วยกัน ได้แก่ ตำแหน่งที่ 1 คือ วันที่ 4 กุมภาพันธ์ ค.ศ. 2006 ที่วันถัดไปมีการเปลี่ยนแปลงปริมาณการใช้งานที่แตกต่างจากวันอื่นๆ มากที่สุด และตำแหน่งที่ 2 คือวันที่ 5 กุมภาพันธ์ ค.ศ. 2006 ที่มีปริมาณการใช้งานช่วงวันที่ 5 ถึงวันที่ 10 กุมภาพันธ์ ค.ศ. 2006 แตกต่างกว่าช่วงเวลาอื่น โดยมีปริมาณการใช้งานที่ใกล้เคียงกันมาก โดยช่วงเวลาของการลดมิติของข้อมูลด้วยวิธี PAA ที่เลือกเป็นเกณฑ์ในการตรวจหาค่าผิดปรกติเป็นช่วงเวลาเดียวกับค่าผิดปรกติที่พบจากข้อมูลแบบไม่ต่อเนื่อง คือ 24 ชั่วโมง มีการตรวจหาค่าผิดปรกติของจำนวนวันที่ถูกแบ่งออกมา เพื่อทำให้เป็นข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาค่าผิดปรกติได้ในบางช่วงเวลา ดังแสดงในตารางที่ 4.19

ตารางที่ 4.19 ผลการตรวจหาค่าผิดปรกติของข้อมูล D ที่จำลองเป็นข้อมูลแบบต่อเนื่องจากการตรวจหาค่าผิดปรกติทั้งสองวิธี

ค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง		ข้อมูลเบื้องต้น		วิธี	อันดับที่พบ	จำนวนครั้ง
อันดับที่	วันที่	จำนวนวัน	วันที่เริ่ม			
1	5 ก.พ. 2006	7	5 ก.พ. 2006	BF	2	12
			5 ก.พ. 2006	MB	2	12
		15	5 ก.พ. 2006	BF	2	132
			3 ก.พ. 2006	MB	2	92
	6 ก.พ. 2006	7	3 ก.พ. 2006	BF	2	12
			3 ก.พ. 2006	MB	2	12
		15	6 ก.พ. 2006	BF	2	132
			3 ก.พ. 2006	MB	2	92

จากตารางที่ 4.19 พบว่าตำแหน่งค่าผิดปรกติที่เกิดจากการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาได้จากจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาจากข้อมูลแบบไม่ต่อเนื่อง ทั้งยังพบว่าค่า

ผิดปรกติที่พบทุกครั้งนั้น สามารถหาได้จากทั้งสองวิธีและจำนวนครั้งของวิธีที่ผู้เขียนดัดแปลงมีค่าเท่ากับวิธีทุกความเป็นไปได้อีกด้วย

- จากข้อมูล F ที่เป็นข้อมูลที่ถูกสร้างขึ้น โดยมีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มลักษณะของรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานมากกว่าปรกติ พบว่ามีตำแหน่งค่าผิดปรกติที่ควรพิจารณา 1 อันดับ 1 ตำแหน่งด้วยกัน คือวันที่ 16 มีนาคม ค.ศ. 2006 ที่ค้นพบจากปริมาณผู้ร้องขอบริการเว็บที่มีปริมาณการใช้งานแตกต่างจากช่วงเวลาอื่นมาก และช่วงเวลาของการลดมิติของข้อมูลด้วยวิธี PAA ที่เลือกเป็นเกณฑ์ในการตรวจหาค่าผิดปรกติเป็นช่วงเวลาเดียวกับค่าผิดปรกติที่พบจากข้อมูลแบบไม่ต่อเนื่อง คือ 24 ชั่วโมง มีการตรวจหาค่าผิดปรกติของจำนวนวันที่ถูกแบ่งออกมา เพื่อทำให้เป็นข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาค่าผิดปรกติได้ในบางช่วงเวลา ดังแสดงในตารางที่ 4.20

ตารางที่ 4.20 ผลการตรวจหาค่าผิดปรกติของข้อมูล F ที่จำลองเป็นข้อมูลแบบต่อเนื่องจากการตรวจหาค่าผิดปรกติทั้งสองวิธี

ค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง		ข้อมูลเบื้องต้น		วิธี	อันดับที่พบ	จำนวนครั้ง
อันดับที่	วันที่	จำนวนวัน	วันที่เริ่ม			
1	16 มี.ค. 2006	7	12 มี.ค. 2006	BF	4	12
			12 มี.ค. 2006	MB	4	12
		15	6 มี.ค. 2006	BF	5	132
			6 มี.ค. 2006	MB	5	132
		30	1 มี.ค. 2006	BF	5	702
			12 มี.ค. 2006	MB	2	360

จากตารางที่ 4.20 พบว่าตำแหน่งค่าผิดปรกติอันดับที่ 1 ที่เกิดจากการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาได้จากจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาจากข้อมูลแบบไม่ต่อเนื่อง ทั้งยังพบว่าค่าผิดปรกติที่พบทุกครั้งนั้น สามารถหาได้จากทั้งสองวิธีและจำนวนครั้งของวิธีที่ผู้เขียนดัดแปลงมีค่าน้อยกว่าหรือเท่ากับวิธีทุกความเป็นไปได้อีกด้วย

- จากข้อมูล G ที่เป็นข้อมูลที่ถูกสร้างขึ้น โดยมีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มตำแหน่งข้อมูลที่ผิดปรกติไว้ 3 ตำแหน่ง ได้แก่ ตำแหน่งที่ 1 คือวันที่ 19 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ตำแหน่งที่ 2 คือวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บมากกว่าปรกติ และตำแหน่งที่ 3 คือวันที่ 16 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ พบว่ามีตำแหน่งค่าผิดปรกติที่ควรพิจารณา 2 อันดับ 2 ตำแหน่งด้วยกัน โดยอันดับที่ 1 คือวันที่ 5 เมษายน ค.ศ. 2006 และอันดับที่ 2 คือวันที่ 17 มีนาคม ค.ศ. 2006 ที่มีปริมาณการใช้งานเปลี่ยนแปลงมาก โดยช่วงเวลาของการลดมิติของข้อมูลด้วยวิธี PAA ที่เลือกเป็นเกณฑ์ในการตรวจหาค่าผิดปรกติเป็นช่วงเวลาเดียวกับค่าผิดปรกติที่พบจากข้อมูลแบบไม่ต่อเนื่อง คือ 24 ชั่วโมง มีการตรวจหาค่าผิดปรกติของจำนวนวันที่ถูกแบ่งออกมา เพื่อทำให้เป็นข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาค่าผิดปรกติได้ในบางช่วงเวลา ดังแสดงในตารางที่ 4.21

ตารางที่ 4.21 ผลการตรวจหาค่าผิดปรกติของข้อมูล G ที่จำลองเป็นข้อมูลแบบต่อเนื่องจากการตรวจหาค่าผิดปรกติทั้งสองวิธี

ค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง		ข้อมูลเบื้องต้น		วิธี	อันดับที่พบ	จำนวนครั้ง
อันดับที่	วันที่	จำนวนวัน	วันที่เริ่ม			
1	5 เม.ย. 2006	7	2 เม.ย. 2006	BF	2	12
			2 เม.ย. 2006	MB	2	12
		15	23 มี.ค. 2006	BF	2	132
			23 มี.ค. 2006	MB	2	112
		30	8 มี.ค. 2006	BF	3	702
			8 มี.ค. 2006	MB	3	652
2	17 มี.ค. 2006	7	12 มี.ค. 2006	BF	2	12
			12 มี.ค. 2006	MB	2	12
		15	-	BF	-	-
			-	MB	-	-
		30	-	BF	-	-
			-	MB	-	-

จากตารางที่ 4.21 พบว่าตำแหน่งค่าผิดปรกติอันดับที่ 1 ที่เกิดจากการตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องนั้น สามารถตรวจหาได้จากจำนวนวันที่เก็บข้อมูลเบื้องต้นสำหรับการตรวจหาจากข้อมูลแบบไม่ต่อเนื่อง ส่วนในกรณีค่าผิดปรกติอันดับที่ 2 นั้น จำนวนวันที่เก็บข้อมูลเบื้องต้นเป็น 15 วัน และ 30 วัน ไม่สามารถตรวจพบว่าเป็นค่าผิดปรกติอันดับแรกได้ ทั้งจากวิธีทุกความเป็นไปได้และวิธีที่ผู้เขียนดัดแปลง ทั้งยังพบว่าค่าผิดปรกติที่พบทุกครั้งนั้นสามารถหาได้จากทั้งสองวิธีและจำนวนครั้งของวิธีที่ผู้เขียนดัดแปลงมีค่าน้อยกว่าหรือเท่ากับวิธีทุกความเป็นไปได้อีกด้วย

4.4 เวลาที่ใช้ในการตรวจหาค่าผิดปรกติ

4.4.1 คุณลักษณะของคอมพิวเตอร์ (Computer Specification)

ในการดำเนินการตรวจหาค่าผิดปรกติของงานวิจัยนี้ใช้คอมพิวเตอร์ในการตรวจหาค่าผิดปรกติ โดยรายละเอียดคุณลักษณะของคอมพิวเตอร์แสดงในตารางที่ 4.22

ตารางที่ 4.22 คุณลักษณะของคอมพิวเตอร์

ชนิดของคอมพิวเตอร์ (Computer)	โน้ตบุ๊กคอมพิวเตอร์ (Notebook Computer)
รุ่น	*acer* Aspire 5112NWLMi
ชนิดของซีพียู (CPU)	AMD® Turion™ 64x2 Mobile Technology TL50
รายละเอียดของซีพียู	มีการประมวลผลแบบ 2 แกน (Dual Core) แต่ละแกนมีความเร็วในการประมวลผล 1.6 กิกะเฮิร์ตซ์ (Gigahertz)
แรม (RAM)	1.536 กิกะไบต์ (Gigabytes)
ความจุฮาร์ดดิสก์ (Hard Disk Capacity)	100 กิกะไบต์

4.4.2 รุ่น (Version) ของโปรแกรม

- ระบบปฏิบัติการ (Operating System) ใช้โปรแกรมไมโครซอฟต์วินโดวส์เอกซ์พี รุ่น 5.1 เซอร์วิสแพ็ค 2 (Microsoft® Windows XP Version 5.1 Service Pack 2)
- โปรแกรมจาวารุ่น 1.6.0-b105 (Java™ version 1.6.0-b105)

4.4.3 เวลาที่ใช้ในการตรวจหาค่าผิดปรกติของข้อมูลแบบไม่ต่อเนื่อง

ในการตรวจหาค่าผิดปรกติของงานวิจัยนี้ได้มีการดำเนินงานการตรวจหาค่าผิดปรกติด้วยวิธีทุกความเป็นไปได้ และวิธีที่ผู้เขียนดัดแปลงไปพร้อมกัน ดังนั้นเวลาที่ใช้ในการตรวจหาค่าผิดปรกติจะเป็นเวลาที่มีการตรวจหาค่าผิดปรกติจากทั้งสองวิธีเช่นกัน ซึ่งพบว่าเวลาที่ใช้ในการตรวจหาค่าผิดปรกติจะมีค่ามากหรือน้อยนั้น ขึ้นอยู่กับจำนวนวันที่ต้องการตรวจหาค่าผิดปรกติและความถี่ของข้อมูลที่มีในแต่ละวัน หากค่าจำนวนวันและความถี่มีค่ามากเท่าไร เวลาที่ใช้ก็จะมีค่ามากยิ่งขึ้น ดังแสดงรายละเอียดดังต่อไปนี้

4.4.3.1 ข้อมูล A ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนมีนาคมถึงเดือนกรกฎาคม ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 1 ชั่วโมง 10 นาที

4.4.3.2 ข้อมูล B ที่เป็นข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยที่มีบันทึกการใช้งานบริการเว็บไซต์ระหว่างเดือนสิงหาคมถึงเดือนธันวาคม ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 2 ชั่วโมง 28 นาที

4.4.3.3 ข้อมูล C เป็นข้อมูลจากเว็บไซต์ของหนังสือพิมพ์ผู้จัดการที่มีบันทึกการใช้งานบริการเว็บไซต์เดือนกุมภาพันธ์ ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 7 นาที

4.4.3.4 ข้อมูล D เป็นข้อมูลจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการที่มีบันทึกการใช้งานบริการเว็บไซต์เดือนกุมภาพันธ์ ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 3 นาที

4.4.3.5 ข้อมูล E เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ A ที่เป็นลักษณะข้อมูลที่มีลักษณะการใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 13 นาที

4.4.3.6 ข้อมูล F เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ B ที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A คือ การใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มลักษณะของรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง คือวันที่ 16 ถึงวันที่ 18 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานมากกว่าปรกติ ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 10 นาที

4.4.3.7 ข้อมูล G เป็นข้อมูลที่สร้างขึ้นจากแฟ้มโครงแบบ C ที่มีลักษณะข้อมูลเบื้องต้นเหมือนกับแฟ้มโครงแบบ A คือ การใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ แต่ได้มีการเพิ่มตำแหน่งข้อมูลที่ผิดปรกติไว้ 3 ตำแหน่ง ได้แก่ ตำแหน่งที่ 1 คือวันที่ 19 มีนาคม ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ตำแหน่งที่ 2 คือวันที่ 4 ถึงวันที่ 5 เมษายน ค.ศ. 2006 ทั้งวัน มี

ปริมาณการใช้งานเว็บมากกว่าปรกติ และตำแหน่งที่ 3 คือวันที่ 16 เมษายน ค.ศ. 2006 ทั้งวัน มีปริมาณการใช้งานเว็บน้อยกว่าปรกติ ใช้เวลาในการตรวจหาค่าผิดปรกติประมาณ 10 นาที

บทที่ 5
การทดสอบประสิทธิภาพของการตรวจหาค่าผิดปรกติ

การทดสอบประสิทธิภาพของงานวิจัยนี้ มีการนำผลจากการตรวจหาค่าผิดปรกติด้วยวิธีทุกความเป็นไปได้เปรียบเทียบกับการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลง

5.1 การเปรียบเทียบผลการตรวจหาค่าผิดปรกติที่ได้กับวิธีทุกความเป็นไปได้

1. ข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยระหว่างเดือนมีนาคมถึงเดือนกรกฎาคม พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลเหมือนกัน แต่จำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงคิดเป็น 58.43 เปอร์เซ็นต์เมื่อเทียบกับจำนวนครั้งของวิธีทุกความเป็นไปได้ โดยมีจำนวน 6,742 ครั้ง ซึ่งน้อยกว่าวิธีทุกความเป็นไปได้ที่มีจำนวน 11,556 ครั้ง หรือ

2. ข้อมูลจากเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัยระหว่างเดือนสิงหาคมถึงเดือนธันวาคม พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลเหมือนกัน แต่จำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงคิดเป็น 44.60 เปอร์เซ็นต์เมื่อเทียบกับจำนวนครั้งของวิธีทุกความเป็นไปได้ โดยมีจำนวน 7,364 ครั้ง ซึ่งน้อยกว่าวิธีทุกความเป็นไปได้ที่มีจำนวน 16,512 ครั้ง

3. ข้อมูลจากเว็บไซต์หนังสือพิมพ์ผู้จัดการเดือนกุมภาพันธ์ พบว่าไม่สามารถตรวจหาค่าผิดปรกติที่มีการลดมิติของข้อมูลด้วยวิธี PAA ที่ช่วงเวลาที่มากที่สุดเหมือนกันได้ แต่พบว่าที่ช่วงเวลาที่ลดมิติของข้อมูลด้วยวิธี PAA ที่มีค่าน้อยกว่าถัดไปของทั้งสองวิธี พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลเหมือนกัน และจำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงมีค่า 1,892 ครั้งเท่ากับวิธีทุกความเป็นไปได้

4. ข้อมูลจากเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการเดือนกุมภาพันธ์ พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลใกล้เคียงกัน โดยวิธีที่ผู้เขียนดัดแปลงไม่พบค่าผิดปรกติอันดับที่ 4 คือ วันที่ 27 กุมภาพันธ์ ค.ศ. 2007 และอันดับที่ 5 คือ วันที่ 24 กุมภาพันธ์ ค.ศ. 2007 และวิธีที่ผู้เขียนดัดแปลงพบค่าผิดปรกติอันดับที่ 2 คือ วันที่ 1 กุมภาพันธ์ ค.ศ. 2007 ในขณะที่วิธีทุกความเป็นไปได้ไม่พบค่าผิดปรกติที่ตำแหน่งนี้ โดยจำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงคิดเป็น 59.78 เปอร์เซ็นต์เมื่อเทียบกับจำนวนครั้งของวิธีทุกความเป็นไปได้ โดยมีค่า 330 ครั้ง ซึ่งน้อยกว่าวิธีทุกความเป็นไปได้ที่มีจำนวน 552 ครั้ง

5. ข้อมูลจากเว็บไซต์ที่ถูกสร้างข้อมูลบันทึกการใช้งานโปรแกรมประยุกต์เว็บที่มีลักษณะการใช้งานปรกติให้คล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลเหมือนกัน แต่จำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงคิด

เป็น 68.31 เปอร์เซ็นต์เมื่อเทียบกับจำนวนครั้งของวิธีทุกความเป็นไปได้ โดยมีค่า 86,324 ครั้งน้อยกว่าวิธีทุกความเป็นไปได้ที่มีจำนวนครั้ง 126,380 ครั้ง

6. ข้อมูลจากเว็บไซต์ที่ถูกสร้างข้อมูลบันทึกการใช้งานโปรแกรมประยุกต์เว็บที่มีลักษณะการใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ และมีการเพิ่มรูปแบบที่ผิดปรกติไว้ 1 ตำแหน่ง พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลเหมือนกัน แต่จำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงคิดเป็น 69.55 เปอร์เซ็นต์เมื่อเทียบกับจำนวนครั้งของวิธีทุกความเป็นไปได้ โดยมีค่า 2,220 ครั้งน้อยกว่าวิธีทุกความเป็นไปได้ที่มีจำนวนครั้ง 3,192 ครั้ง

7. ข้อมูลจากเว็บไซต์ที่ถูกสร้างข้อมูลบันทึกการใช้งานโปรแกรมประยุกต์เว็บที่มีลักษณะการใช้งานเว็บไซต์ที่มีลักษณะการใช้งานปรกติคล้ายคลึงกับลักษณะการใช้งานจากเว็บไซต์ และมีการเพิ่มรูปแบบที่ผิดปรกติไว้ 3 ตำแหน่ง พบว่าสามารถตรวจหาค่าผิดปรกติได้ผลเหมือนกัน แต่จำนวนครั้งในการตรวจหาค่าผิดปรกติของวิธีที่ผู้เขียนดัดแปลงคิดเป็น 52.38 เปอร์เซ็นต์เมื่อเทียบกับจำนวนครั้งของวิธีทุกความเป็นไปได้ โดยมีค่า 1,672 ครั้งน้อยกว่าวิธีทุกความเป็นไปได้ที่มีจำนวนครั้ง 3,192 ครั้ง

5.2 การวิเคราะห์ผลการเปรียบเทียบกับวิธีทุกความเป็นไปได้

จากการเปรียบเทียบผลการตรวจหาค่าผิดปรกติทั้งจากวิธีทุกความเป็นไปได้และจากวิธีที่ผู้เขียนดัดแปลงพบว่า การตรวจหาค่าผิดปรกติด้วยวิธีที่ผู้เขียนดัดแปลง สามารถตรวจหาค่าผิดปรกติได้เช่นเดียวกับการตรวจหาค่าผิดปรกติด้วยวิธีทุกความเป็นไปได้ โดยมีรายละเอียดดังต่อไปนี้

- ข้อมูลมีลักษณะที่ผิดปรกติอย่างชัดเจน

พบว่าจำนวนครั้งในการตรวจหาค่าผิดปรกติด้วยวิธีที่ผู้เขียนดัดแปลงมีค่าน้อยกว่าหรือเท่ากับวิธีทุกความเป็นไปได้ สำหรับข้อมูลทั้งแบบไม่ต่อเนื่องและต่อเนื่อง

- ข้อมูลมีลักษณะที่ผิดปรกติอย่างไม่ชัดเจน

พบว่าค่าผิดปรกติที่ตรวจพบจากวิธีที่ผู้เขียนดัดแปลงนั้นพบค่าผิดปรกติที่มีการลดมิติของข้อมูลด้วยวิธี PAA ที่ช่วงเวลาน้อยกว่าวิธีทุกความเป็นไปได้ ซึ่งเมื่อเทียบผลค่าผิดปรกติดังกล่าวกับการลดมิติของข้อมูลด้วยวิธี PAA ที่ความละเอียดเดียวกับวิธีทุกความเป็นไปได้แล้ว พบว่ามีค่าผิดปรกติเหมือนกัน

สาเหตุที่วิธีที่ผู้เขียนดัดแปลงนั้นไม่สามารถพบค่าผิดปรกติที่มีการลดมิติของข้อมูลด้วยวิธี PAA ที่ช่วงเวลาเดียวกับวิธีทุกความเป็นไปได้ มีผลมาจากโครงสร้างข้อมูลที่สร้างขึ้นมานั้นมีการเก็บชุดลำดับย่อยที่ซ้ำกันไว้เป็นโครงสร้างแถวลำดับทำให้เกิดผลดังต่อไปนี้

1. เมื่อทำการคำนวณค่าระยะห่างของลำดับย่อยในแต่ละการวนรอบจะทำให้ไม่พบลำดับย่อยที่มีอักขระชุดเดียวกับลำดับย่อยที่คำนวณค่าระยะห่าง ทำให้ค่าระยะห่างที่ได้มีค่าไม่เท่ากันและอาจมีผลให้ค่าผิดปรกติที่ได้ไม่เหมือนกัน

2. ข้อมูลลำดับย่อยที่ซ้ำกันที่เป็นค่าผิดปรกติมีจำนวนมากกว่าเกณฑ์ในการเลือกค่าผิดปรกติ

3. ข้อมูลลำดับย่อยที่ซ้ำกันที่เป็นค่าผิดปรกติรวมกับข้อมูลลำดับย่อยอื่นแล้วมีจำนวนมากกว่าเกณฑ์ในการเลือกค่าผิดปรกติ

สำหรับผลของการทดลองตรวจหาค่าผิดปรกติจากข้อมูลแบบไม่ต่อเนื่องนั้น พบว่าค่าผิดปรกติที่ตรวจหาได้จากข้อมูลแบบไม่ต่อเนื่องนั้น อาจตรวจหาไม่พบ อันเนื่องมาจากลักษณะข้อมูลข้างเคียงนั้นมีลักษณะต่างกัน และยิ่งมีจำนวนข้อมูลหลักมากเท่าไร ยิ่งทำให้การตรวจหาค่าผิดปรกติดีขึ้นเท่านั้น แต่ก็จะทำให้เวลาที่ใช้ในการตรวจหาค่าผิดปรกติมากขึ้นตามไปด้วย

5.3 การวิเคราะห์การสร้างข้อมูลบันทึกการเข้าใช้งานของเว็บไซต์

จากผลการทดลองสรุปได้ว่า การสร้างข้อมูลบันทึกการเข้าใช้งานเว็บไซต์สามารถสร้างข้อมูลที่มีรูปแบบต่างๆ ตามต้องการได้ โดยการสร้างข้อมูลที่มีรูปแบบใดๆ นั้น จำเป็นต้องใช้การวิเคราะห์ก่อนที่จะสร้าง เพื่อให้ข้อมูลที่ถูกสร้างนั้นสามารถเป็นตัวแทนของข้อมูลจากแหล่งอื่นๆ ได้

บทที่ 6
สรุปผลการวิจัยและข้อเสนอแนะ

6.1 สรุปผลการวิจัย

งานวิจัยนี้แบ่งออกเป็น 2 ส่วนด้วยกันคือ

- การพัฒนาโปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ

ในการพัฒนาโปรแกรมที่ใช้ในการสร้างข้อมูลบันทึกการเข้าใช้งานบริการเว็บนั้น เริ่มจากการพัฒนาให้โปรแกรมสามารถสร้างข้อมูลได้อย่างยืดหยุ่น ไม่ว่ารูปแบบของข้อมูลจะเป็นรูปแบบใด โดยมีการสร้างแฟ้มโครงแบบเอกซ์เอ็มแอลเพื่อใช้ในการสร้างข้อมูลที่มีรูปแบบที่หลากหลาย เพื่อลดความจำเป็นในการแก้ไขโปรแกรม หลังจากนั้นโปรแกรมจะอ่านรูปแบบของข้อมูลที่มีในแฟ้มโครงแบบเอกซ์เอ็มแอล เพื่อสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บตามต้องการ โดยรูปแบบที่ทำการสร้างแฟ้มโครงแบบเอกซ์เอ็มแอล จะผ่านการวิเคราะห์จากข้อมูลการบันทึกการเข้าใช้งานบริการเว็บที่มีอยู่และเกิดขึ้นจริง

- การพัฒนาโปรแกรมที่ใช้ในการตรวจหาค่าผิดปรกติของเครื่องบริการโปรแกรมประยุกต์เว็บ

ขั้นตอนการออกแบบและพัฒนาโปรแกรมที่ใช้ในการตรวจหาค่าผิดปรกติของเครื่องบริการโปรแกรมประยุกต์เว็บ เริ่มจากการแปลงข้อมูลบันทึกการเข้าใช้งานบริการเว็บให้เป็นข้อมูลอนุกรมเวลา จากนั้นลดมิติของข้อมูลด้วยวิธี PAA ปรับข้อมูลให้เป็นบรรทัดฐานด้วยวิธี Z-Score แล้วแปลงข้อมูลให้อยู่ในรูปของสัญลักษณ์โดยใช้วิธีแซ็ค ซึ่งข้อมูลที่เป็นสัญลักษณ์นั้น จะถูกนำมาตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้และจากวิธีที่ดัดแปลงโดยผู้เขียนมาเปรียบเทียบกัน โดยค่าพารามิเตอร์ต่างๆ ที่ใช้ในทุกขั้นตอนนั้นจะผ่านการวิเคราะห์ก่อนที่จะนำมาใช้

โดยงานวิจัยนี้ได้ทดสอบโปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บโดยการสร้างข้อมูลที่มีรูปแบบต่างๆ และทดสอบโปรแกรมที่ใช้ในการตรวจหาค่าผิดปรกติของเครื่องบริการโปรแกรมประยุกต์เว็บ ซึ่งสรุปผลการทดสอบได้ดังนี้

6.1.1 ผลการทดสอบโปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ

ข้อมูลที่ได้จากการสร้างผ่านโปรแกรม เมื่อนำมาทำให้อยู่ในรูปของแผนภูมิกราฟพบว่าสามารถสร้างข้อมูลบันทึกการใช้งานได้ตามต้องการ

6.1.2 ผลการทดสอบโปรแกรมที่ใช้ในการตรวจหาค่าผิดปรกติ ของเครื่องบริการโปรแกรมประยุกต์เว็บ

ได้มีการทดสอบการตรวจหาค่าผิดปรกติจากข้อมูลของเว็บไซต์ต่างๆ ทั้งวิธีการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้ และวิธีที่ผู้เขียนดัดแปลง สามารถตรวจหาค่าผิดปรกติได้เหมือนกัน หากข้อมูลมีลักษณะที่แตกต่างกันอย่างชัดเจน โดยจำนวนครั้งในการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลงมีจำนวนครั้งที่น้อยกว่าหรือเท่ากับการตรวจหาค่าผิดปรกติจากทุกความเป็นไปได้มาก

จากการทดสอบโปรแกรมได้ผลดังนี้

- โปรแกรมที่ใช้ในการสร้างข้อมูลการบันทึกการเข้าใช้งานบริการเว็บ สามารถนำไปใช้เพื่อช่วยในการสร้างข้อมูลที่มีรูปแบบที่หลากหลาย เพื่อใช้ในการเปรียบเทียบความสามารถในการตรวจหาค่าผิดปรกติได้
- โปรแกรมที่ใช้ในการตรวจหาค่าผิดปรกติของเครื่องบริการโปรแกรมประยุกต์เว็บ สามารถนำไปใช้ในการตรวจหาค่าผิดปรกติที่เกิดขึ้นในโปรแกรมประยุกต์เว็บได้
- ขั้นตอนวิธีที่ใช้ในการตรวจหาค่าผิดปรกติของเครื่องบริการโปรแกรมประยุกต์เว็บ สามารถนำไปตรวจหาค่าผิดปรกติกับข้อมูลที่อยู่ในรูปของอนุกรมเวลาแบบอื่นๆ ได้

6.2 ปัญหาที่พบจากการวิจัย

ปัญหาที่พบจากการทำการวิจัยบางประการ ที่น่าจะเป็นประโยชน์และสามารถนำไปเป็นแนวทางในการแก้ปัญหาในงานวิจัยที่ใกล้เคียงอื่นๆ ต่อไป ดังนี้

1. ปัญหาของเวลาที่ใช้ในการตรวจหาค่าผิดปรกติ โดยพบว่าหากข้อมูลมีจำนวนมาก เวลาที่ใช้ในการตรวจหาค่าผิดปรกติยิ่งต้องใช้เวลามากเช่นกัน
2. ปัญหาของจำนวนข้อมูลหลักที่นำมาใช้ในการตรวจหาค่าผิดปรกติจากข้อมูลแบบต่อเนื่อง พบว่ายิ่งมีจำนวนข้อมูลหลักมากเท่าไร ยิ่งจะทำให้การตรวจหาค่าผิดปรกติทำได้ดีเท่านั้น แต่เวลาที่ใช้ในการตรวจหาค่าผิดปรกติก็จะมากขึ้นเช่นกัน

6.3 ข้อเสนอแนะ

1. สามารถสร้างข้อมูลบันทึกการใช้งานบริการเว็บที่หลากหลายเพื่อใช้ในการเปรียบเทียบความสามารถของขั้นตอนวิธีในการตรวจหาค่าผิดปรกติอื่นๆ
2. สามารถนำขั้นตอนวิธีในการตรวจหาค่าผิดปรกติที่ผู้เขียนดัดแปลง ไปใช้กับการตรวจหาค่าผิดปรกติกับข้อมูลที่อยู่ในรูปแบบของอนุกรมเวลาในโดเมนอื่นๆ

รายการอ้างอิง

[1] Barford, P.; Kline, J.; Plonka, D.; and Ron, A. A Signal Analysis of Network Traffic Anomalies. In Proceedings of 2nd ACM SIGCOMM Workshop on Internet Measurement, pp. 71-82, Marseille, France, 2002.

[2] Barford, P.; and Plonka, D. Characteristics of Network Traffic Flow Anomalies. In Proceedings of 1st ACM SIGCOMM Workshop on Internet Measurement, pp. 69-73, California, USA, 2001.

[3] Brutlag, J. Aberrant behavior detection in time series for network monitoring. In Proceedings of USENIX Fourteenth System Administration Conference LISA XIV, New Orleans, LA, December 2000.

[4] Sommers, J.; Yegneswaran, V.; and Barford, P. A Framework for Malicious Workload Generation. In Proceedings of ACM SIGCOMM/USENIX Internet Measurement Conference. pp. 82-87, Taormina, Italy, October 2004.

[5] Sommers, J.; Yegneswaran, V.; and Barford, P. Recent Advances in Network Intrusion Detection System Tuning. In Proceedings of 40th IEEE Conference on Information Sciences and Systems, March 2006.

[6] Thottan, M.; and Chuanyi, J. Anomaly Detection in IP Networks. The Transactions on Signal Processing 51, 8 (2003) : 2191-2204.

[7] Top Ten Most Critical Web Application Security Vulnerabilities 2004 Update [Online]. Available from: http://www.owasp.org[2006, August 1]

[8] Hoffman, B. Analysis of Web Application Worms and Viruses. Black Hat Federal Europe 2006. Netherlands, 2006.

[9] Auger, R.; et al. Web Application Firewall Evaluation Criteria. Web Application Security Consortium. 2006.

[10] Seo, J.; Kim, H. S.; Cho, S.; and Cha, S. Web Server Attack Categorization based on Root Causes and Their Locations. In Proceedings of Information Technology: Coding and Computing (ITTC'04). Volume 1, pp. 90-96, 2004.

[11] OWASP, About The Open Web Application Security Project [Online]. 2006. Avaiable from: http://www.owasp.org/index.php/About_The_Open_Web_Application_Security_Project[2006, August 1]

[12] Breach Security. ModSecurity [Online]. 2004.Available from: http://www.modsecurity.org[2006,August 1]

[13] Grossman, J.; and Auger, R. Web Application Security Consortium: Charter [Online]. 2004. Available from: http://www.webappsec.org/aboutus.shtml[2006, August 1]

[14] Barnett, R. C. Preventing Web Attacks with Apache. Addison Wesley Professional. 2006.

[15] Moyer, S. Defending Black Box Web Application. Black Hat USA. Las Vegas, 2006.

[16] Thinking Stone. ModSecurity for Apache User Guide [Online]. 2006. Available from: http://www.modsecurity.org[2006, August 1]

[17] Zissman, Marc. Data Sets Overview [Online]. 2001. Available from: http://www.ll.mit.edu/IST/ideval/data/data_index.html[2006, August 1]

[18] Keogh, E.; Lin, J.; and Fu, A. HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications. In Proceedings of 5th IEEE International Conference on Data Mining (ICDM 2005), pp. 8, Nov 27-30, 2005.

[19] Keogh, E.; Lin, J.; Lonardi, S.; and Chiu, B. We Have Seen the Future, and It Is Symbolic. In Proceedings of 2nd workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation. Volume 32, page 83, New Zealand, 2004.

[20] เศรษฐศาสตร์น่ารู้ [Online]. Available from: http://www.fpo.go.th/fseg/Source/ECO/ECO24.htm[2006, December 1]

[21] Scott, C.; Wolfe, P.; and Hayes, B. Snort FOR DUMMIES. Wiley Publishing, 2004.

[22] Lin, J.; Keogh, E.; Lonardi, S.; Lankford, J. P.; and Nystrom, D. M. Visually Mining and Monitoring Massive Time Series. In Proceedings of 10th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 460-469, Seattle, Washington, USA, 2004.

[23] Lawrence Berkeley National Laboratory, Bro Intrusion Detection System [Online]. 2003. Available from: http://www.bro-ids.org[2006 August 1]

[24] Sourcefire. Snort [Online]. Available from: http://www.snort.org[2006, August 1]

[25] Bleeding Edge Threat [Online]. Available from: http://www.bleedingsnort.com [2006, August 1]

[26] Logging Control in W3C httpd [Online]. Available from: http://www.w3.org/Daemon/User/Config/Logging.html#common-logfile-format[2006, August 1]

[27] Basak, R. Identd for Windows 98/Me [Online]. Available from: http://identd.sourceforge.net/[2006, August 1]

[28] 10 Status Code Definitions [Online]. Available from: http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html[2006, August 1]

[29] Glossary [Online]. Available from: http://dtp.epsb.net/glossary.htm[2006, December 1]

ภาคผนวก

ภาคผนวก ก
แฟ้มโครงแบบเอกซ์เอ็มแอลของการสร้างข้อมูลการบันทึกการใช้งานของเว็บ

ตารางที่ ก.1 แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับสร้างข้อมูลแบบที่ 1

```
<?xml version="1.0"?>
<data_generator>
  <template>
      <web>
          <folder value="d:/workspace/cputil/log/web-access-gen.log"/>
      </web>
  </template>
  <data>
      <cp_web value="web">
          <year value="2006">
              <month value="3-4">
                  <week value="1-6">
                      <day value="1-5">
                          <hour value="0-6">
                              <traffic value="8500-9000">
                                  <size value="1400-1500"/>
                              </traffic>
                              <dst_ip_range value="202.3.145.3-202.3.145.3">
                                  <port_range value="80"/>
                              </dst_ip_range>
                              <ip_range value="192.168.1.0-192.168.1.254">
                                  <port_range value="1024-5000"/>
                              </ip_range>
                          </hour>
                          <hour value="6-0">
                              <traffic value="8500-9500">
                                  <size value="1500-1700"/>
                              </traffic>
                              <dst_ip_range value="202.3.145.3-202.3.145.3">
                                  <port_range value="80"/>
                              </dst_ip_range>
                              <ip_range value="192.168.1.0-192.168.1.254">
                                  <port_range value="1024-5000"/>
                              </ip_range>
                          </hour>
                      </day>
                      <day value="6-7">
                          <hour value="0-6">
                              <traffic value="9500-10000">
                                  <size value="1400-1500"/>
                              </traffic>
```

ตารางที่ ก.1 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 1

```
                        <dst_ip_range value="202.3.145.3-202.3.145.3">
                            <port_range value="80"/>
                        </dst_ip_range>
                        <ip_range value="192.168.1.0-192.168.1.254">
                            <port_range value="1024-5000"/>
                        </ip_range>
                    </hour>
                    <hour value="6-0">
                        <traffic value="10000-11000">
                            <size value="1500-1700"/>
                        </traffic>
                        <dst_ip_range value="202.3.145.3-202.3.145.3">
                            <port_range value="80"/>
                        </dst_ip_range>
                        <ip_range value="192.168.1.0-192.168.1.254">
                            <port_range value="1024-5000"/>
                        </ip_range>
                    </hour>
                </day>
            </week>
        </month>
    </year>
  </cp_web>
 </data>
</data_generator>
```

ตารางที่ ก.2 แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับสร้างข้อมูลแบบที่ 2

```
<?xml version="1.0"?>
<data_generator>
  <template>
      <web>
          <folder value="d:/workspace/cputil/log/web-access-gen.log"/>
      </web>
  </template>
  <data>
      <cp_web value="web">
          <year value="2006">
              <month value="3-4">
                  <week value="1-6">
                      <day value="1-5">
                          <hour value="0-6">
                              <traffic value="8500-9000">
                                  <size value="1400-1500"/>
                              </traffic>
                              <dst_ip_range value="202.3.145.3-202.3.145.3">
                                  <port_range value="80"/>
                              </dst_ip_range>
                              <ip_range value="192.168.1.0-192.168.1.254">
                                  <port_range value="1024-5000"/>
                              </ip_range>
                          </hour>
                          <hour value="6-0">
                              <traffic value="8500-9500">
                                  <size value="1500-1700"/>
                              </traffic>
                              <dst_ip_range value="202.3.145.3-202.3.145.3">
                                  <port_range value="80"/>
                              </dst_ip_range>
                              <ip_range value="192.168.1.0-192.168.1.254">
                                  <port_range value="1024-5000"/>
                              </ip_range>
                          </hour>
                      </day>
                      <day value="6-7">
                          <hour value="0-6">
                              <traffic value="9500-10000">
                                  <size value="1400-1500"/>
                              </traffic>
                              <dst_ip_range value="202.3.145.3-202.3.145.3">
                                  <port_range value="80"/>
                              </dst_ip_range>
```

ตารางที่ ก.2 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 2

```
                        <ip_range value="192.168.1.0-192.168.1.254">
                            <port_range value="1024-5000"/>
                        </ip_range>
                    </hour>
                    <hour value="6-0">
                        <traffic value="10000-11000">
                            <size value="1500-1700"/>
                        </traffic>
                        <dst_ip_range value="202.3.145.3-202.3.145.3">
                            <port_range value="80"/>
                        </dst_ip_range>
                        <ip_range value="192.168.1.0-192.168.1.254">
                            <port_range value="1024-5000"/>
                        </ip_range>
                    </hour>
                </day>
            </week>
        </month>
    </year>
  </cp_web>
 </data>
 <anomaly_web1 value="web">
    <year value="2006">
        <month value="3">
            <week value="3">
                <day value="5-7">
                    <hour value="0-0">
                        <traffic value="10000-22500">
                            <size value="1000-2000"/>
                        </traffic>
                        <dst_ip_range value="202.3.145.3-202.3.145.3">
                            <port_range value="80"/>
                        </dst_ip_range>
                        <ip_range value="192.168.1.0-192.168.1.254">
                            <port_range value="1024-5000"/>
                        </ip_range>
                    </hour>
                </day>
            </week>
        </month>
    </year>
  </anomaly_web1>
</data_generator>
```

ตารางที่ ก.3 แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
<?xml version="1.0"?>
<data_generator>
  <template>
    <web>
      <folder value="e:/workspace/thesis/log/web-access-gen.log"/>
    </web>
  </template>
  <data>
    <min_web value="web">
      <year value="2006">
        <month value="3-4">
          <week value="1-6">
            <day value="1-7">
              <hour value="0-0">
                <traffic value="4000-4500">
                  <size value="1000-1100"/>
                </traffic>
                <dst_ip_range value="202.3.145.3-202.3.145.3">
                  <port_range value="80"/>
                </dst_ip_range>
                <ip_range value="192.168.1.0-192.168.1.254">
                  <port_range value="1024-5000"/>
                </ip_range>
              </hour>
                  </day>
                </week>
            </month>
          </year>
    </min_web>
    <cp_web1 value="web">
      <year value="2006">
        <month value="3-3">
          <week value="1-3">
            <day value="1-7">
              <hour value="0-6">
                <traffic value="4500-5000">
                  <size value="1400-1500"/>
                </traffic>
                <dst_ip_range value="202.3.145.3-202.3.145.3">
                  <port_range value="80"/>
                </dst_ip_range>
                <ip_range value="192.168.1.0-192.168.1.254">
                  <port_range value="1024-5000"/>
                </ip_range>
              </hour>
              <hour value="6-0">
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
            <traffic value="4500-5000">
              <size value="1500-1700"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
        </day>
        <day value="6-7">
          <hour value="0-6">
            <traffic value="5500-6000">
              <size value="1400-1500"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
          <hour value="6-0">
            <traffic value="6000-7000">
              <size value="1500-1700"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
        </day>
      </week>
    </month>
  </year>
</cp_web1>
<cp_web2 value="web">
  <year value="2006">
    <month value="3-3">
      <week value="4-4">
        <day value="2-7">
          <hour value="0-6">
            <traffic value="4500-5000">
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
            <size value="1400-1500"/>
          </traffic>
          <dst_ip_range value="202.3.145.3-202.3.145.3">
            <port_range value="80"/>
          </dst_ip_range>
          <ip_range value="192.168.1.0-192.168.1.254">
            <port_range value="1024-5000"/>
          </ip_range>
        </hour>
        <hour value="6-0">
          <traffic value="4500-5500">
            <size value="1500-1700"/>
          </traffic>
          <dst_ip_range value="202.3.145.3-202.3.145.3">
            <port_range value="80"/>
          </dst_ip_range>
          <ip_range value="192.168.1.0-192.168.1.254">
            <port_range value="1024-5000"/>
          </ip_range>
        </hour>
      </day>
      <day value="6-7">
        <hour value="0-6">
          <traffic value="5500-6000">
            <size value="1400-1500"/>
          </traffic>
          <dst_ip_range value="202.3.145.3-202.3.145.3">
            <port_range value="80"/>
          </dst_ip_range>
          <ip_range value="192.168.1.0-192.168.1.254">
            <port_range value="1024-5000"/>
          </ip_range>
        </hour>
        <hour value="6-0">
          <traffic value="6000-7000">
            <size value="1500-1700"/>
          </traffic>
          <dst_ip_range value="202.3.145.3-202.3.145.3">
            <port_range value="80"/>
          </dst_ip_range>
          <ip_range value="192.168.1.0-192.168.1.254">
            <port_range value="1024-5000"/>
          </ip_range>
        </hour>
      </day>
    </week>
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
      </month>
    </year>
  </cp_web2>
  <cp_web3 value="web">
    <year value="2006">
      <month value="3-3">
        <week value="5-6">
          <day value="1-7">
            <hour value="0-6">
              <traffic value="4500-5000">
                <size value="1400-1500"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
            <hour value="6-0">
              <traffic value="4500-5500">
                <size value="1500-1700"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
          </day>
          <day value="6-7">
            <hour value="0-6">
              <traffic value="5500-6000">
                <size value="1400-1500"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
            <hour value="6-0">
              <traffic value="6000-7000">
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
            <size value="1500-1700"/>
          </traffic>
          <dst_ip_range value="202.3.145.3-202.3.145.3">
            <port_range value="80"/>
          </dst_ip_range>
          <ip_range value="192.168.1.0-192.168.1.254">
            <port_range value="1024-5000"/>
          </ip_range>
        </hour>
      </day>
    </week>
  </month>
 </year>
</cp_web3>
<cp_web4 value="web">
 <year value="2006">
  <month value="4-4">
   <week value="1-3">
    <day value="1-7">
     <hour value="0-6">
      <traffic value="4500-5000">
       <size value="1400-1500"/>
      </traffic>
      <dst_ip_range value="202.3.145.3-202.3.145.3">
       <port_range value="80"/>
      </dst_ip_range>
      <ip_range value="192.168.1.0-192.168.1.254">
       <port_range value="1024-5000"/>
      </ip_range>
     </hour>
     <hour value="6-0">
      <traffic value="4500-5500">
       <size value="1500-1700"/>
      </traffic>
      <dst_ip_range value="202.3.145.3-202.3.145.3">
       <port_range value="80"/>
      </dst_ip_range>
      <ip_range value="192.168.1.0-192.168.1.254">
       <port_range value="1024-5000"/>
      </ip_range>
     </hour>
    </day>
    <day value="6-7">
     <hour value="0-6">
      <traffic value="5500-6000">
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
                <size value="1400-1500"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
            <hour value="6-0">
              <traffic value="6000-7000">
                <size value="1500-1700"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
          </day>
        </week>
      </month>
    </year>
  </cp_web4>
  <cp_web5 value="web">
    <year value="2006">
      <month value="4-4">
        <week value="4-4">
          <day value="2-7">
            <hour value="0-6">
              <traffic value="4500-5000">
                <size value="1400-1500"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
            <hour value="6-0">
              <traffic value="4500-5500">
                <size value="1500-1700"/>
              </traffic>
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
        </day>
        <day value="6-7">
          <hour value="0-6">
            <traffic value="5500-6000">
              <size value="1400-1500"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
          <hour value="6-0">
            <traffic value="6000-7000">
              <size value="1500-1700"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
        </day>
      </week>
    </month>
  </year>
</cp_web5>
<cp_web6 value="web">
  <year value="2006">
    <month value="4-4">
      <week value="5-6">
        <day value="1-7">
          <hour value="0-6">
            <traffic value="4500-5000">
              <size value="1400-1500"/>
            </traffic>
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
          <hour value="6-0">
            <traffic value="4500-5500">
              <size value="1500-1700"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
        </day>
        <day value="6-7">
          <hour value="0-6">
            <traffic value="5500-6000">
              <size value="1400-1500"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
          <hour value="6-0">
            <traffic value="6000-7000">
              <size value="1500-1700"/>
            </traffic>
            <dst_ip_range value="202.3.145.3-202.3.145.3">
              <port_range value="80"/>
            </dst_ip_range>
            <ip_range value="192.168.1.0-192.168.1.254">
              <port_range value="1024-5000"/>
            </ip_range>
          </hour>
        </day>
      </week>
    </month>
```

ตารางที่ ก.3 (ต่อ) แฟ้มโครงแบบเอกซ์เอ็มแอลสำหรับการสร้างข้อมูลแบบที่ 3

```
    </year>
  </cp_web6>
  <anomaly_web1 value="web">
    <year value="2006">
      <month value="4">
        <week value="2">
          <day value="3-4">
            <hour value="0-0">
              <traffic value="11000-18500">
                <size value="1000-2000"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
          </day>
        </week>
      </month>
    </year>
  </anomaly_web1>
  <anomaly_web1 value="web">
    <year value="2006">
      <month value="4">
        <week value="2">
          <day value="3-4">
            <hour value="0-0">
              <traffic value="11000-18500">
                <size value="1000-2000"/>
              </traffic>
              <dst_ip_range value="202.3.145.3-202.3.145.3">
                <port_range value="80"/>
              </dst_ip_range>
              <ip_range value="192.168.1.0-192.168.1.254">
                <port_range value="1024-5000"/>
              </ip_range>
            </hour>
          </day>
        </week>
      </month>
    </year>
  </anomaly_web1>
```

ภาคผนวก ข
แฟ้มโครงแบบของการสร้างแผนภูมิกราฟ
และแปลงข้อมูลให้มีเพียงข้อมูลที่นำมาวิเคราะห์เท่านั้น

ตารางที่ ข.1 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์จุฬาลงกรณ์มหาวิทยาลัย

```
# for cu web
src_dir=e:/workspace/Thesis/log/
files=cu-web.log
date_pos=3
size_pos=9
src_ip_pos=8
dst_ip_pos=2
log_type=apache
date_type=DATE_TYPE_STRING
# space is unicode character \u0020
delim=\u0020
out_dir=./out-cu/
stream_counts=14400,28800,86400,172800,259200,345600,432000,604800
packet_counts=50,100,200,400,800,1600,3200
sec_counts=3600,7200,14400,28800,86400
sax_counts=3-10
sax_window_length=3,4,5,6,7,8,10
sec7_counts=3600,7200,14400,28800,86400
sax7_counts=3-10
sax7_window_length=3,6,12,24
knn=1000
streaming_flag=yes
non_streaming_flag=yes
```

ตารางที่ ข.2 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์หนังสือพิมพ์ผู้จัดการ

```
# for manager
src_dir=d:/workspace/Thesis/log/
files=manager.log
date_pos=0
size_pos=13
src_ip_pos=8
dst_ip_pos=2
log_type=iis
date_type=DATE_TYPE_STRING
#space is Unicode character \u0020
delim=\u0020
out_dir=./out-manager/
stream_counts=14400,28800,86400,172800,259200,345600,432000,604800
packet_counts=50,100,200,400,800,1600,3200
sec_counts=3600,7200,14400,28800,86400
sax_counts=3-10
sax_window_length=3,4,5,6,7,8,10
sec7_counts=3600,7200,14400,28800,86400
sax7_counts=3-10
sax7_window_length=3,6,12,24
knn=1000
streaming_flag=yes
non_streaming_flag=yes
```

ตารางที่ ข.3 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์วิทยุออนไลน์ของหนังสือพิมพ์ผู้จัดการ

```
# for manager radio
src_dir=d:/workspace/Thesis/log/
files=radio.log
date_pos=0
size_pos=10
src_ip_pos=2
dst_ip_pos=4
log_type=iis
date_type=DATE_TYPE_STRING
#space is unicode character \u0020
delim=\u0020
out_dir=./out-radio/
stream_counts=14400,28800,86400,172800,259200,345600,432000,604800
packet_counts=50,100,200,400,800,1600,3200
sec_counts=3600,7200,14400,28800,86400
sax_counts=3-10
sax_window_length=3,4,5,6,7,8,10
sec7_counts=3600,7200,14400,28800,86400
sax7_counts=3-10
sax7_window_length=3,6,12,24
knn=1000
streaming_flag=yes
non_streaming_flag=yes
```

ตารางที่ ข.4 แฟ้มโครงแบบแปลงข้อมูลของเว็บไซต์ที่ถูกสร้างขึ้น

```
# for cu web
src_dir=e:/workspace/Thesis/log/
files=web-access-gen.log
date_pos=3
size_pos=9
src_ip_pos=8
dst_ip_pos=2
log_type=apache
date_type=DATE_TYPE_STRING
# space is unicode character \u0020
delim=\u0020
out_dir=e:/workspace/thesis/out/
stream_counts=14400,28800,86400,172800,259200,345600,432000,604800
packet_counts=50,100,200,400,800,1600,3200
sec_counts=3600,7200,14400,28800,86400
sax_counts=3-10
#sax_window_length is in source code
sax_window_length=3,4,5,6,7,8,10
streaming_day=7,15,30
sec_streaming_counts=3600,7200,14400,28800,86400
sax_streaming_counts=3-10
#sax_streaming_window_length is in source code
sax_streaming_window_length=24,12,6,3
```

ภาคผนวก ค
แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ

ตารางที่ ค.1 แฟ้มโครงแบบที่ใช้ในการตรวจหาค่าผิดปรกติ

```
# for all web
src_dir=e:/workspace/Thesis/out/
files=web-data.txt
date_pos=0
size_pos=1
src_ip_pos=1
dst_ip_pos=1
log_type=general
date_type=DATE_TYPE_STRING
delim=,
out_dir=./out /
sec_counts=3600,7200,14400,28800,86400
sax_counts=3-10
#sax_window_length is in source code
sax_window_length=3,4,5,6,7,8,9,10
streaming_day=7,15,30
sec_streaming_counts=3600,7200,14400,28800,86400
sax_streaming_counts=3-10
#sax_streaming_window_length is in source code
sax_streaming_window_length=24,12,6,3
knn=1000
streaming_flag=yes
non_streaming_flag=yes
```

ประวัติผู้เขียนวิทยานิพนธ์

นายไพศาล ดีวรรณกิจ เกิดเมื่อวันที่ 8 สิงหาคม พ.ศ. 2520 ที่จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาวิทยาศาสตรบัณฑิต จากภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2542 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2548