การตรวจจับการบุกรุกด้วยเทคนิคการจำแนกในการทำเหมืองข้อมูล

ผดุง นันอำไพ

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
กุมภาพันธ์ 2561

การตรวจจับการบุกรุกด้วยเทคนิคการจำแนกในการทำเหมืองข้อมูล

ผดุง นันอำไพ

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
กุมภาพันธ์ 2561

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายผดุง นันอำไพ แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....	ประธานกรรมการ
(อาจารย์ ดร.ฉัตรตระกูล สมบัติธีระ)	(กรรมการบัณฑิตศึกษาประจำคณะ)
.....	กรรมการ
(ผศ.ดร.จารี ทองคำ)	(อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก)
.....	กรรมการ
(ผศ.ดร.แกมกาญจน์ สมประเสริฐศรี)	(อาจารย์บัณฑิตศึกษาประจำคณะ)
.....	กรรมการ
(รศ.ดร.สิทธิชัย บุษหมั่น)	(ผู้ทรงคุณวุฒิ)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

.....	.....
(ผศ.ดร.สุจิน บุตรดีสุวรรณ)	(ผศ.ดร.กริสน์ ชัยมูล)
คณบดีคณะวิทยาการสารสนเทศ	คณบดีบัณฑิตวิทยาลัย
	วันที่ 29 เดือน ก.พ. พ.ศ. 2561

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์จากบุคคลหลายท่าน ซึ่งท่านแรกที่ผู้วิจัยขอกราบขอบพระคุณคือ ผู้ช่วยศาสตราจารย์ ดร.จารี ทองคำ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำปรึกษา ตลอดจนปรับปรุงแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่อย่างดียิ่ง ผู้วิจัยตระหนักถึงความตั้งใจจริงและความทุ่มเทของอาจารย์ และขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอกราบขอบพระคุณ ประธานกรรมการสอบและกรรมการสอบในการสอบวิทยานิพนธ์ที่ช่วยชี้แนะ และนำเสนอแนวทางให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์

สุดท้ายนี้ ขอกราบขอบพระคุณ ครอบครัวนันอำไพ ที่คอยช่วยเหลือและเป็นกำลังใจตลอดจน ขอบคุณเพื่อนๆ ทุกคนที่ให้ความช่วยเหลือ ให้คำแนะนำ และเป็นกำลังใจเสมอมา จนการศึกษาในครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี

ผดุง นันอำไพ

ชื่อเรื่อง การตรวจจับการบุกรุกด้วยเทคนิคการจำแนกในการทำเหมืองข้อมูล
ผู้วิจัย นายผดุง นันอำไพ
ปริญญา วิทยาศาตรมหาบัณฑิต **สาขาวิชา** เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.จารี ทองคำ
มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2561

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองในการจำแนกรูปแบบการบุกรุกบนระบบเครือข่าย โดยใช้เทคนิคการจำแนกในการทำเหมืองข้อมูล 4 เทคนิคด้วยกันคือ เทคนิค Decision Table เทคนิค Naïve Bayes เทคนิค RIPPER และเทคนิค PART Decision list เพื่อหาแบบจำลองที่มีประสิทธิภาพและเหมาะสมสำหรับการตรวจจับการบุกรุกบนระบบเครือข่าย ในงานวิจัยนี้ใช้ชุดข้อมูลการบุกรุกระบบเครือข่ายจากฐานข้อมูลความรู้ KDD Cup'99 หลักการ 10-Fold Cross Validation ได้ถูกนำมาใช้ในการแบ่งชุดข้อมูลออกเป็นชุดเรียนรู้และชุดทดสอบ ผลการทดลองพบว่าแบบจำลองที่ใช้เทคนิค RIPPER มีค่าความถูกต้องเฉลี่ย (Accuracy) มากที่สุดคือ 99.97% และเทคนิค PART decision list มีค่าความถูกต้อง 99.96% ตามด้วยเทคนิค Decision Table มีค่าความถูกต้อง 99.76% และเทคนิคที่มีค่าความถูกต้องน้อยที่สุดคือเทคนิค Naïve Bayes มีค่าความถูกต้อง 92.78%

คำสำคัญ : การตรวจจับการบุกรุก; เทคนิคการจำแนกข้อมูล; การทำเหมืองข้อมูล; เทคนิคการจำแนกด้วยกฎ

TITLE Intrusion Detection Using Classification Techniques In Data Mining
AUTHOR Mr. Phadung Nanaumphai
DEGREE Master Degree of Science Program **MAJOR** Information Technology
ADVISORS Asst. Prof. Jaree Thongkum, Ph.D.
UNIVERSITY Mahasarakham University **YEAR** 2018

ABSTRACT

The objective of this paper is to compare the effectiveness of intrusion detection models using four classification techniques including Decision table, Naïve Bayes, RIPPER and PART decision list in data mining. In this thesis, the knowledge database "KDD Cup'99" is used. 10-fold cross validation is employed to divided data into training and testing sets. The experiment results showed that RIPPER has highest accuracy which is up to 99.97%. Then, PART Decision list has accuracy which is up to 99.96%. Follow by Decision Table has highest accuracy which is up to 99.76%. The lowest accuracy is Naïve Bayes 92.78%

Key Words : Intrusion Detection; Classification Techniques; Data Mining; Rule Based Techniques

สารบัญ

หน้า

กิตติกรรมประกาศ .....ก
บทคัดย่อภาษาไทย .....ข
บทคัดย่อภาษาอังกฤษ .....ค
สารบัญ .....ง
สารบัญตาราง .....ฉ
สารบัญรูป .....ช
บทที่ 1 บทนำ 1
1.1 หลักการและเหตุผล 1
1.2 วัตถุประสงค์ของการวิจัย 2
1.3 ความสำคัญของการวิจัย 2
1.4 ขอบเขตของการวิจัย 2
1.5 นิยามศัพท์เฉพาะ 3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4
2.1 ระบบเครือข่ายคอมพิวเตอร์ 4
2.2 ระบบตรวจจับการบุกรุก 8
2.3 รูปแบบการบุกรุกระบบเครือข่าย 10
2.4 การทำเหมืองข้อมูล 13
2.5 งานวิจัยที่เกี่ยวข้อง 18
บทที่ 3 วิธีดำเนินการวิจัย 20
3.1 การเตรียมข้อมูล 20
3.2 การสร้างแบบจำลอง 23
3.3 วัดประสิทธิภาพแบบจำลอง 23
บทที่ 4 ผลการวิจัย 25
4.1 ประสิทธิรูปของแบบจำลอง 25
4.2 การวิเคราะห์ประสิทธิรูปของแบบจำลอง 31
4.3 การแปลผลของแบบจำลอง .....

หน้า

บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 35
5.1 สรุปผลการวิจัย 35
5.2 อภิปรายผลการวิจัย 35

5.3 ข้อเสนอแนะการวิจัย.....36
เอกสารอ้างอิง.....37
ภาคผนวก.....41
ภาคผนวก ก ผลการรันแบบจำลองฉบับเต็ม..... 42
ประวัติย่อผู้วิจัย..... 87

สารบัญตาราง

หน้า

ตารางที่ 2.1 แสดงข้อมูลระดับบิตที่อยู่ใน TCP Header ซึ่งมีอยู่ทั้งหมด 6 บิต.....7
ตารางที่ 2.2 ตารางการตัดสินใจของ Decision Table 16
ตารางที่ 3.1 แอททริบิวของข้อมูล.....20
ตารางที่ 3.2 คลาส 22
ตารางที่ 2.2 ตารางการตัดสินใจของ Decision Table 33
ตารางที่ 4.1 ผลการจำแนกโดยใช้เทคนิค Decision Table.....25
ตารางที่ 4.2 ผลการจำแนกโดยใช้เทคนิค Naïve Bayes 27
ตารางที่ 4.3 ผลการจำแนกโดยใช้เทคนิค RIPPER.....28
ตารางที่ 4.4 ผลการจำแนกโดยใช้เทคนิค PART Decision list 29
ตารางที่ 4.5 ผลการวิเคราะห์ประสิทธิภาพของแบบจำลอง.....31

สารบัญรูป

หน้า

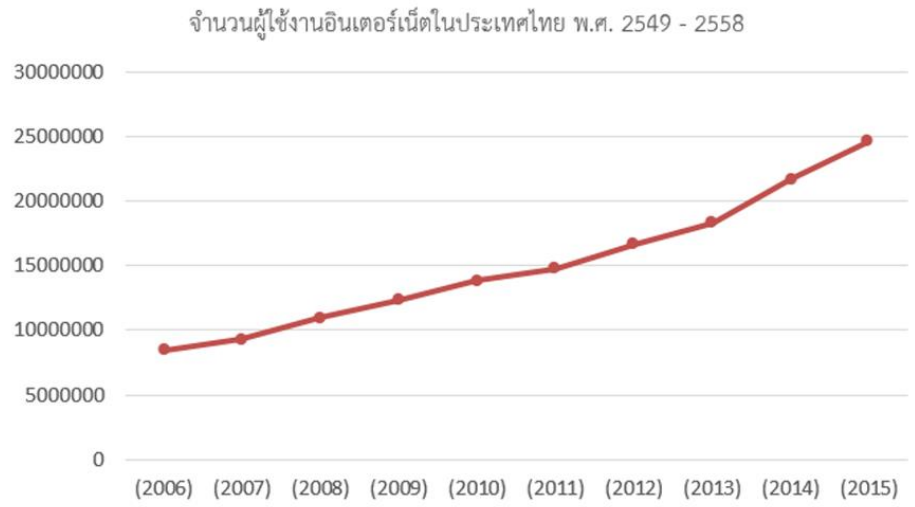
รูปที่ 1.1 แสดงจำนวนผู้ใช้งานอินเตอร์เน็ตในประเทศไทย พ.ศ. 2549 – 2558 1
รูปที่ 2.1 แสดงรูปแบบของ IP Header 5
รูปที่ 2.2 แสดงรูปแบบของ ICMP Header 6
รูปที่ 2.3 แสดงรูปแบบของ UDP Header 6
รูปที่ 2.4 แสดงรูปแบบของ TCP Header 7
รูปที่ 2.5 รูปแบบการทำงานของระบบตรวจจับการบุกรุก 8
รูปที่ 2.6 รูปแบบการทำงานของระบบตรวจจับการบุกรุกบนเครือข่าย 9
รูปที่ 2.7 แสดงการโจมตีแบบ Land Attacks 11
รูปที่ 2.8 แสดงขั้นตอนการทำเหมืองข้อมูล 14
รูปที่ 3.1 รูปแบบการการทดสอบตามวิธีการ 10 – fold cross validation 24

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

สังคมในปัจจุบันมีความต้องการในการรับและส่งข้อมูลสารสนเทศผ่านทางระบบเครือข่ายคอมพิวเตอร์เป็นจำนวนมาก อีกทั้งยังมีอุปกรณ์ต่างๆ ที่ใช้ในการรับและส่งข้อมูลที่แตกต่างกันไป ไม่ว่าจะเป็นเครื่องคอมพิวเตอร์ โทรศัพท์มือถือและอุปกรณ์อื่นๆ อีกมากมาย ซึ่งแนวโน้มในการใช้อินเตอร์เน็ตนั้นเพิ่มขึ้นทุกๆ ปี การสำรวจการมีการใช้เทคโนโลยีสารสนเทศและการสื่อสารในครัวเรือนตั้งแต่ปี พ.ศ. 2549 - 2558 โดยสำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร [1] พบว่าในปี พ.ศ. 2549 มีผู้ใช้งานอินเตอร์เน็ตประมาณ 8,465,823 คนและเพิ่มขึ้นเรื่อยๆ ทุกปีจนถึง 24,592,299 คน ในปี พ.ศ. 2558 ดังแสดงในรูปที่ 1.1



รูปที่ 1.1 แสดงจำนวนผู้ใช้งานอินเตอร์เน็ตในประเทศไทย พ.ศ. 2549 – 2558

ดังนั้นสิ่งที่มีความสำคัญและมองข้ามไม่ได้คือความปลอดภัยบนระบบเครือข่ายคอมพิวเตอร์ เนื่องจากปัจจุบันมีผู้ประสงค์ร้ายและต้องการหาผลประโยชน์ในระบบเครือข่ายคอมพิวเตอร์เป็นจำนวนมาก แม้จะมีอุปกรณ์และเครื่องมือต่างๆที่สามารถช่วยให้ระบบเครือข่ายคอมพิวเตอร์ปลอดภัยขึ้น แต่ผู้บุกรุกนั้นก็ยังมีเทคนิควิธีการและเครื่องมือต่างๆมากมายที่ใหม่ขึ้นและซับซ้อนขึ้นเพื่อหลีกเลี่ยงการตรวจจับ

การนำระบบตรวจจับการบุกรุก (Intrusion Detection System: IDS) เข้ามาใช้ทำให้ช่วยเพิ่มความปลอดภัยในระบบเครือข่ายคอมพิวเตอร์ได้มากขึ้น จึงมีงานวิจัยที่นำเสนอทฤษฎีและเทคนิคต่างๆ ที่นำมาใช้ในการวิเคราะห์รูปแบบการบุกรุก เช่นงานวิจัย [2] ได้นำเสนอโมเดลที่ใช้วิเคราะห์การบุกรุกโดยใช้เทคนิค TASVM แต่งานวิจัยดังกล่าวมีข้อบกพร่องในเรื่องของการประมวลผลที่ใช้เวลานานเกินไป

ในการเปรียบเทียบเทคนิคการตรวจจับการบุกรุกในงานวิจัยนี้จะเปรียบเทียบเพื่อวัดความแม่นยำและวัดความเร็ว โดยใช้เทคนิคการจำแนก 4 เทคนิคด้วยกันคือ เทคนิคตารางตัดสินใจ (Decision Table) เป็นวิธีที่นำมาทดสอบการทำงานร่วมกันของเงื่อนไขที่มีหลายเงื่อนไข มีลักษณะคล้ายกับต้นไม้ตัดสินใจ (Decision Tree) แต่จะอยู่ในรูปของตาราง เทคนิค Naïve Bayes [3] เป็นวิธีที่ได้รับความนิยมในการนำมาใช้จำแนกข้อมูล เนื่องจากมีแบบจำลองที่เข้าใจได้ง่ายและไม่ซับซ้อน ซึ่งเทคนิคนี้ใช้หลักการของความน่าจะเป็น เทคนิค RIPPER (Rule-Based Classification) เป็นอัลกอริทึมที่สามารถสร้างกฎเองได้ โดยการเรียนรู้จากข้อมูลที่เตรียมไว้ให้และเทคนิค PART decision list ซึ่งเป็นอัลกอริทึมที่สามารถจัดการกับข้อมูลที่หายไปและคุณลักษณะทางตัวเลขที่ต่างกันได้ดี โดยการวิจัยครั้งนี้มีจุดประสงค์เพื่อศึกษาข้อดีและข้อเสียในการจำแนกข้อมูลด้วยเทคนิคที่แตกต่างกัน เพื่อให้ได้เทคนิคการจำแนกข้อมูลบนระบบเครือข่ายคอมพิวเตอร์ที่มีประสิทธิภาพและลดเวลาในการประมวลผล

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการบุกรุกบนเครือข่าย
2. เพื่อพัฒนาแบบจำลองที่สารมารถจำแนกและตรวจจับการบุกรุกในระบบเครือข่าย

1.3 ความสำคัญของการวิจัย

1. ได้เทคนิคการจำแนกการตรวจจับการบุกรุกระบบเครือข่ายคอมพิวเตอร์ที่มีประสิทธิภาพ
2. ได้แบบจำลองที่มีประสิทธิภาพในการจำแนกการตรวจจับการบุกรุกระบบเครือข่าย

1.4 ขอบเขตของการวิจัย

1. ใช้ข้อมูลจากฐานข้อมูล KDD Cup'99 จำนวน 494,020 เรคคอร์ด
2. ปัจจัยที่ใช้ในการจำแนกคือ แอททริบิวจากฐานข้อมูลจำนวน 41 แอททริบิว
3. วิเคราะห์ข้อมูลโดยใช้หลักการ 10-fold cross validation เพื่อใช้ข้อมูลทุกชุดเป็นทั้งชุดการสอนและชุดการทดสอบ

4. ใช้เทคนิคการจำแนกข้อมูลโดยการเปรียบเทียบ 4 เทคนิคด้วยกัน คือ Decision Table, Naïve Bayes, RIPPER และ PART decision list

1.5 นิยามศัพท์เฉพาะ

1. การทำเหมืองข้อมูล (Data Mining) การหารูปแบบและความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ในระบบเครือข่ายคอมพิวเตอร์ เพื่อวิเคราะห์และตรวจจับการบุกรุก โดยใช้ขั้นตอนทางสถิติ

2. การจำแนกข้อมูล (Classification) เป็นกระบวนการวิเคราะห์ทางสถิติเพื่อจัดประเภทของข้อมูลในระบบเครือข่าย โดยวิเคราะห์ข้อมูลต่างๆที่อยู่ในแพ็คเกจ เพื่อหารูปแบบของกลุ่มข้อมูลใหม่

3. ระบบเครือข่าย หมายถึง คอมพิวเตอร์และอุปกรณ์ต่างๆที่เชื่อมต่อกันเพื่อให้สารมารถติดต่อสื่อสารและแลกเปลี่ยนข้อมูลกันได้ และรวมไปถึงการติดต่อสื่อสารผ่านอินเทอร์เน็ต

4. ผู้บุกรุก หมายถึง ผู้ที่ประสงค์ร้ายต่อระบบเครือข่ายคอมพิวเตอร์และผู้ที่แสวงหาผลประโยชน์ในทางที่ผิดบนระบบเครือข่ายคอมพิวเตอร์ทำให้เกิดความเสียหายแก่ผู้อื่น

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ผู้วิจัยได้ทำการศึกษาค้นคว้าทฤษฎีและงานวิจัยที่เกี่ยวข้องเรื่องต่างๆ ไม่ว่าจะเป็นรูปแบบการบุกรุกระบบเครือข่ายคอมพิวเตอร์ด้วยเทคนิควิธีต่างๆ องค์ประกอบของแพ็คเก็ตที่ใช้ส่งข้อมูลบนระบบเครือข่ายคอมพิวเตอร์ เทคนิคการจำแนกด้วยวิธีต่างๆในการทำเหมืองข้อมูลและทฤษฎีต่างๆที่เกี่ยวข้องกับงานวิจัย ดังนี้

2.1 ระบบเครือข่ายคอมพิวเตอร์

2.1.1 โปรโตคอล (Protocol)

โปรโตคอลเปรียบเสมือนภาษากลางที่คอมพิวเตอร์ให้สื่อสารกัน ซึ่งจำเป็นต้องมีมาตรฐานเดียวกัน เพื่อให้เครื่องคอมพิวเตอร์และอุปกรณ์ที่มีความแตกต่างกันและมีอยู่มากมายหลายรุ่นเข้าใจกันได้ เช่นเดียวกันกับภาษาที่มนุษย์ใช้สื่อสารกัน ซึ่งมนุษย์นั้นภาษากลางเปรียบเสมือนภาษาอังกฤษ ซึ่งการที่จะสื่อสารกันให้เข้าใจก็ต้องใช้ภาษาอังกฤษเหมือนกัน คอมพิวเตอร์ก็เช่นเดียวกัน ทั้งเครื่องคอมพิวเตอร์และอุปกรณ์บนเครือข่ายคอมพิวเตอร์จำเป็นต้องใช้โปรโตคอลชนิดเดียวกันเท่านั้นจึงจะสามารถติดต่อและส่งข้อมูลระหว่างกันได้ องค์ประกอบหลักของโปรโตคอลนั้นจะประกอบไปด้วย 3 ส่วนหลักๆ คือ

Syntax หมายถึง รูปแบบหรือโครงสร้างของข้อมูลที่แจ้งให้แอนติตี้ทราบว่าส่วนไหนคือที่อยู่ของผู้รับ ส่วนไหนคือที่อยู่ของผู้ส่ง และส่วนไหนคือข้อมูลข้อมูลจริงๆ
Semantics หมายถึง ความหมายของข้อมูลที่ได้รับมาแล้วทำการแปลง เพื่อให้รู้ว่าบิตแต่ละบิตทำอะไรได้บ้าง

Timing คือข้อกำหนดเวลาในการรับส่งข้อมูล เนื่องจากแอนติตี้แต่ละตัวมีเวลาในการรับส่งไม่เท่ากัน ซึ่งถ้าไม่มีข้อกำหนดเวลาในการรับส่งแล้ว แอนติตี้ที่ทำงานช้ากว่าจะไม่สามารถรับข้อมูลได้ทัน

โปรโตคอลนั้นเป็นองค์ประกอบที่สำคัญในการรับส่งข้อมูล ดังนั้นแล้วจึงจำเป็นต้องมีมาตรฐาน(Standard) เนื่องจากมีอุปกรณ์มามายหลายชนิดที่แตกต่างกันและใช้สื่อสารกันผ่านระบบเครือข่าย ดังนั้นจึงจำเป็นต้องมีกำกำหนดมาตรฐานเพื่อให้อุปกรณ์ที่แตกต่างกันสามารถสื่อสารกันได้

2.1.1.1 IP (Internet Protocol)

IP (Internet Protocol) เป็นโปรโตคอลในระดับ Network Layer ซึ่งทำหน้าที่ในการจัดการที่อยู่ผู้รับและผู้ส่งรวมทั้งควบคุมการส่งข้อมูลบางอย่างที่ใช้ในการค้นหาเส้นทาง กลไกการทำงานของ Internet Protocol จะมีความสามารถค้นหาเส้นทางที่ดีที่สุดได้ และสามารถเปลี่ยนแปลง

เส้นทางระหว่างการส่งข้อมูลได้ หามีปัญหาเกี่ยวกับเครือข่ายที่ทำให้ไม่สามารถส่งข้อมูลไปยังเส้นทางเดิมได้

การส่งข้อมูลของ Internet Protocol จะเป็นการเชื่อมต่อทุกครั้งในการส่งข้อมูล 1 Datagram โดยจะทำการแบ่งข้อมูลออกเป็นส่วนย่อยและทำไปประกอบเป็นข้อมูลเดิมเมื่อถึงปลายทาง โดยประมวลผลจาก IP Header ซึ่งลักษณะของ IP Header มีรูปแบบดังแสดงในรูปที่ 2.1

<table>
<tr><td>4-bit Version</td><td>Header Length</td><td>8-bit Type of Service</td><td colspan="2">16-bit Total Length in Byte</td></tr>
<tr><td colspan="3">16-bit Identification</td><td>3- bit Flag</td><td>16-bit Fragment Checksum</td></tr>
<tr><td colspan="2">8-bit Time to Live (TTL)</td><td>8-bit Protocol</td><td colspan="2">16-bit Header Checksum</td></tr>
<tr><td colspan="5">32-bit Source IP Address</td></tr>
<tr><td colspan="5">32-bit Destination IP Address</td></tr>
<tr><td colspan="5">Option</td></tr>
<tr><td colspan="5">Data</td></tr>
</table>

รูปที่ 2.1 แสดงรูปแบบของ IP Header

แต่ละส่วนของ IP Header จะมีความหมายดังนี้

Version หมายถึง เวอร์ชั่นของโปรโตคอลปัจจุบันที่ใช้งานอยู่คือ IPv4 และ IPv6

Header Length หมายถึง ความยาวของ Header

Type of Service (TOS) คือ ข้อมูลในการตัดสินใจเลือกเราเตอร์ แต่ปัจจุบันไม่มีการนำมาใช้แล้ว

Length คือ ความยาวของ Datagram ทั้งหมด คิดเป็นจำนวน Byte

Identification คือ หมายเลขของ Datagram เมื่อถูกแบ่งออกเป็นส่วน

Flag คือ ตัวกำหนดการทำงานเมื่อมีการแยกดาต้าแกรม

Fragment Offset คือ ตัวกำหนดตำแหน่งข้อมูลในดาต้าแกรมที่มีการแยกส่วน ทำให้ขั้นตอนการนำส่วนต่างๆมารวมกันนั้นสามารถทำได้ถูกต้อง

Time to Live (TTL) คือ ตัวกำหนดจำนวนครั้งในการส่ง โดยเริ่มนับจากมากไปหาน้อย จะถูกนับทุกครั้งเมื่อมีการส่ง 1 hop และเมื่อ TTL เป็น 0 แต่ข้อมูลยังไม่ถึงปลายทางข้อมูลนั้นจะถูกยกเลิกและเราเตอร์ตัวล่าสุดที่ข้อมูลไปถึงจะส่งข้อมูล ICMP แจ้งกลับไปยังปลายทางว่าเกิด Time out Protocol คือ ประเภทโปรโตคอลที่ใช้ในการส่ง เช่น TCP, UDP หรือ ICMP

Header Checksum คือ ส่วนที่ใช้ในการตรวจสอบความถูกต้องของข้อมูลในเฮดเดอร์

Source IP Address คือ หมายเลขไอพีหรือที่อยู่ผู้ส่ง

Destination IP Address คือ หมายเลขไอพีหรือที่อยู่ผู้รับ

Data คือข้อมูลจากโปรโตคอลระดับบน

2.1.1.2 ICMP (Internet Control Message Protocol)

โปรโตคอล ICMP คือโปรโตคอลที่ใช้สำหรับตรวจสอบและรายงานสถานภาพของ Datagram เมื่อเกิดปัญหากับ Datagram เช่น เมื่อไม่สารมารถส่ง Datagram ไปถึงยังปลายทางได้ โปรโตคอล ICMP จะทำการ รายงานไปยังต้นทางเพื่อแจ้งความผิดพลาดในการส่งที่เกิดขึ้น ซึ่งหากโปรโตคอล ICMP ไม่มีการรายงานกลับไปยังต้นทางก็แสดงว่า ข้อมูลที่ส่งไปยังปลายทางนั้นส่งได้สำเร็จ หรือการแจ้งความผิดพลาดมายังต้นทางนั้นเกิดปัญหาระหว่างทาง เนื่องจากไม่สามารถรับประกันได้ว่าเมื่อไม่มีการตอบกลับเพื่อแจ้งข้อผิดพลาดในการส่ง ข้อมูลที่ส่งไปยังปลายทางนั้นจะสำเร็จหรือไม่ โปรโตคอล ICMP จึงไม่มีความน่าเชื่อถือ ซึ่งจำเป็นต้องมีโปรโตคอลในระดับสูงกว่า Network Layer เพื่อใช้ในการจัดการการสื่อสารให้มีความน่าเชื่อถือ

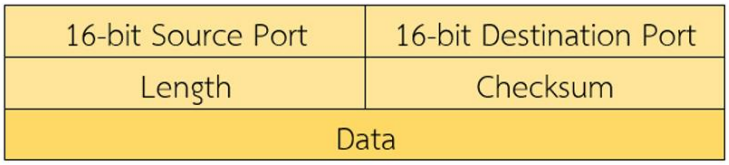
ในส่วนของ ICMP Message จะประกอบด้วย Type ขนาด 8 บิต Checksum ขนาด 16 บิต และส่วนของ Content ซึ่งจะมีขนาดแตกต่างกันไปตาม Type และ Code ดังแสดงในรูปที่ 2.2

รูปที่ 2.2 แสดงรูปแบบของ ICMP Header

2.1.1.3 UDP: (User Datagram Protocol)

โปรโตคอล UDP เป็นโปรโตคอลที่อยู่ในชั้น Transport Layer ของ OSI Model 7 Layer โปรโตคอล UDP นั้นจะทำการส่งข้อมูลครั้งละ 1 ชุด เรียกว่า UDP Datagram ซึ่งโปรโตคอลนี้จะไม่มีกลไกการตรวจสอบความสำเร็จในการรับหรือส่งข้อมูล

การตรวจสอบ Checksum ของโปรโตคอล UDP นั้นใช้เพื่อป้องกันการแก้ไขข้อมูลที่อาจเกิดขึ้นระหว่างการส่ง ซึ่งหากมีกรณีนี้เกิดขึ้นผู้ส่งจะไม่รู้ว่ามีข้อผิดพลาดดังกล่าวเกิดขึ้น แต่ข้อกำหนดของ UDP จะแจ้งให้ผู้รับทิ้งข้อมูลดังกล่าว หากการส่งข้อมูลมีความผิดพลาดระหว่างการส่งในระดับ IP เช่น ส่งไม่ถึงผู้รับหรือหมดเวลาในการส่ง ผู้ส่งจะได้รับข้อความแจ้งเตือนข้อผิดพลาด แต่หากเกิดความผิดพลาดในการส่งที่ระดับ UDP จะไม่มีการแจ้งให้ผู้ส่งทราบUDP Header มีรายละเอียดดังแสดงในรูปที่ 2.3



รูปที่ 2.3 แสดงรูปแบบของ UDP Header

Source Port Number คือ หมายเลขพอร์ตต้นทาง

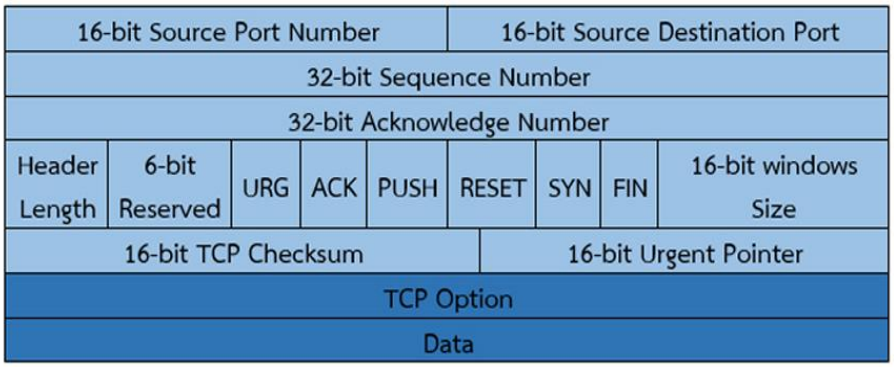
Destination Port Number คือ หมายเลขพอร์ตปลายทาง

UDP Length คือ ความยาวของ Datagram ทั้งส่วนของ Header และส่วนของข้อมูล

Checksum คือ ส่วนที่คอยตรวจสอบความถูกต้องของ UDP Datagram

2.1.1.4 TCP: (Transmission Control Protocol)

โปรโตคอล TCP เป็นโปรโตคอลที่อยู่ในชั้น Transport Layer ของ OSI Model 7 Layer ทำหน้าที่จัดการและควบคุมการรับส่งข้อมูลคล้ายกับโปรโตคอล UDP แต่จะมีรายละเอียดมากกว่ามีการสื่อสารอย่างเป็นกระบวนการและมีการตรวจสอบความถูกต้องของข้อมูลดังแสดงใน รูปที่ 2.4



รูปที่ 2.4 แสดงรูปแบบของ TCP Header

Source Port Number คือ หมายเลขพอร์ตต้นทาง

Destination Port Number คือ หมายเลขพอร์ตปลายทาง

Sequence Number คือ ส่วนที่ใช้ระบุหมายเลขลำดับในการส่งข้อมูล เพื่อใช้ในการจัดลำดับข้อมูลที่ถูกส่งมาได้ถูกต้อง

Acknowledgment Number คือ ส่วนที่ทำหน้าที่เช่นเดียวกับ Sequence Number แต่จะใช้ในการตอบรับ

Header Length คือ ความยาวของ Header

Flag คือตัวที่กำหนดการทำงานของ TCP Segment ซึ่ง Flag มีอยู่ทั้งหมด 6 บิต แบ่งได้ดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 แสดงข้อมูลระดับบิตที่อยู่ใน TCP Header ซึ่งมีอยู่ทั้งหมด 6 บิต

Type	Description
URG	ใช้บอกความหมายว่าเป็นข้อมูลด่วน และมีข้อมูลพิเศษมาด้วย
ACK	แจ้งว่าสามารถนำข้อมูลในฟิลด์ Acknowledge Number นำมาใช้งานได้
DSH	เป็นการแจ้งให้ผู้รับข้อมูลทราบว่าควรจะส่งข้อมูล Segment นี้ไปยัง

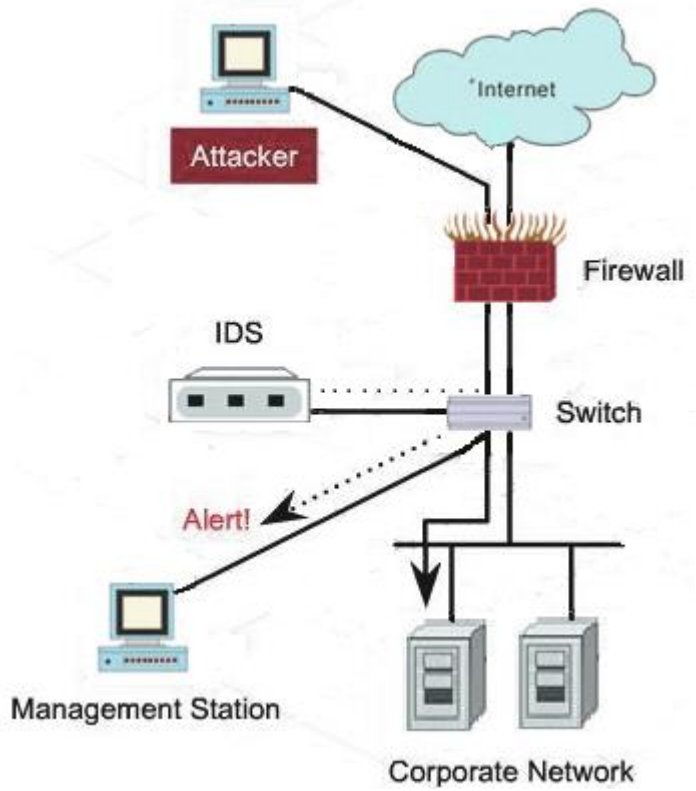
ตารางที่ 2.1 (ต่อ)

Type	Description
	Application ที่กำลังรออยู่โดยเร็ว
RST	ยกเลิกการเชื่อมอมต่อและให้เริ่มการสื่อสารใหม่เนื่องจากในกรณีที่เกิดการสับสน
SYN	ใช้ในการขอเชื่อต่อกับปลายทาง
FIN	ใช้ในการแจ้งปลายทางว่ายุติการติดต่อ

ใน TCP Header นั้นจำเป็นต้องมี Flag เพื่อการกำหนดการทำงานของ TCP Segment เพราะในการทำงานแต่ละอย่างจะมีฟิลด์ไม่เหมือนกัน ซึ่งถ้าไม่มี Flag เป็นตัวกำหนดการทำงานจะทำให้เกิดความผิดพลาดในการนำข้อมูลมาใช้งานได้

2.2 ระบบตรวจจับการบุกรุก

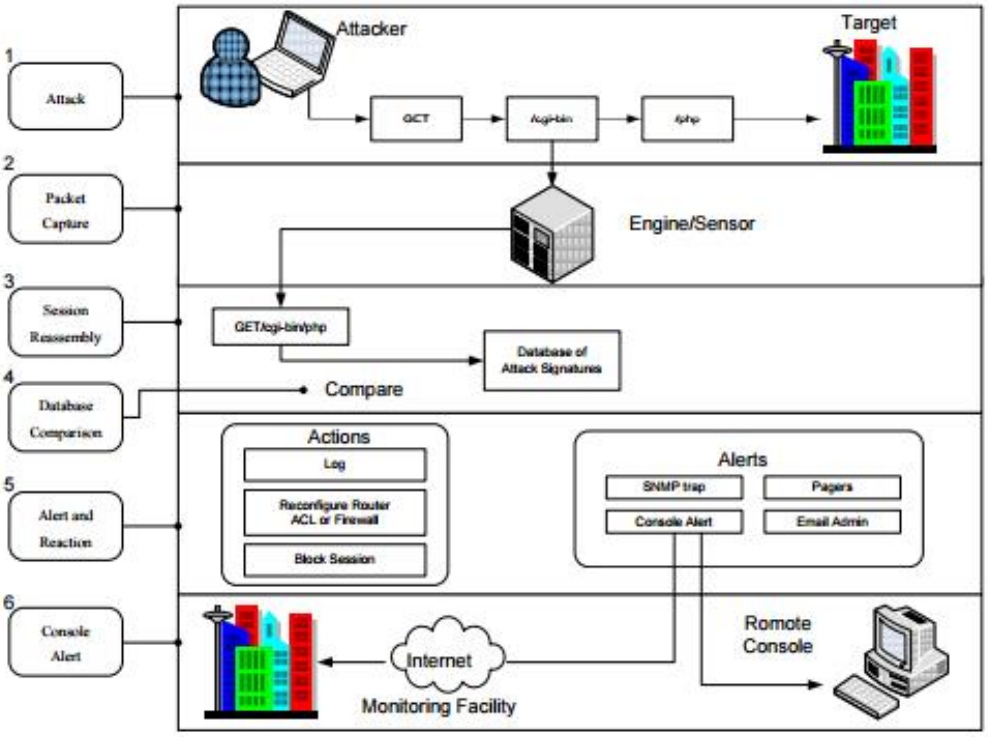
ระบบตรวจจับการบุกรุก (Intrusion Detection Systems) [4] คือซอฟต์แวร์หรือฮาร์ดแวร์ที่ถูกนำมาใช้ในการตรวจสอบข้อมูลที่ถูกรับและส่งภายในระบบเครือข่ายคอมพิวเตอร์ (Traffic) โดยระบบตรวจจับการบุกรุกจะทำหน้าที่ในการวิเคราะห์รูปแบบของแพ็คเก็ต (Packet) เพื่อหารูปแบบของแพ็คเก็ตที่มีความผิดปกติหรือมีพฤติกรรมที่เข้าข่ายบุกรุกระบบเครือข่าย และจะทำการแจ้งเตือนไปยังผู้ดูแลระบบให้ตรวจสอบเพื่อป้องกันและแก้ไขได้อย่างทันท่วงทีดังแสดงในรูปที่ 2.5



รูปที่ 2.5 รูปแบบการทำงานของระบบตรวจจับการบุกรุก

2.2.1 สถาปัตยกรรมระบบตรวจจับการบุกรุก

2.2.1.1 ระบบตรวจจับการบุกรุกบนเครือข่าย (Network-based IDSs) ทำงานบนระบบเครือข่ายโดยจะทำการตรวจสอบข้อมูลที่วิ่งอยู่บนระบบเครือข่ายเพื่อหาแพ็คเก็ตที่มีพฤติกรรมผิดปกติดังแสดงในรูปที่ 2.6



รูปที่ 2.6 รูปแบบการทำงานของระบบตรวจจับการบุกรุกบนเครือข่าย

2.2.1.2 ระบบตรวจจับการบุกรุกบนเครื่องคอมพิวเตอร์ (Host-based IDSs) ทำงานบนเครื่องคอมพิวเตอร์โดยจะทำการตรวจสอบข้อมูลในแพ็คเก็ตที่ถูกเก็บในเครื่องคอมพิวเตอร์ (Log File) และวิเคราะห์พฤติกรรมผิดปกติของการใช้งาน

2.2.2 หลักการทำงานของระบบตรวจจับการบุกรุก

2.2.2.1 การตรวจจับการใช้งานที่ผิดปกติ (Misuse Detection) การทำงานของระบบตรวจจับการบุกรุกนั้นจะทำการตรวจสอบและวิเคราะห์รูปของการบุกรุกที่เกิดขึ้นในระบบเครือขายโดยการเปรียบเทียบพฤติกรรมของแพ็คเก็ตที่เข้ามาในระบบเครือข่ายกับข้อมูลการบุกรุกที่มีอยู่แล้ว ซึ่งถ้าหากพฤติกรรมของแพ็คเก็ตที่เข้ามาในระบบเครือข่ายตรงกับข้อมูลการบุกรุกที่ถูกเก็บอยู่ระบบจะทำการแจ้งให้ผู้ดูแลระบบทราบ

2.2.2.2 การตรวจจับการเหตุการณ์ที่ผิดปกติ (Anomaly-based Detection) คือการตรวจสอบและวิเคราะห์พฤติกรรมของการบุกรุกที่เกิดขึ้นในระบบเครือข่ายเพื่อเปรียบเทียบรูปแบบของแพ็คเก็ตที่เข้ามาในระบบเครือข่ายกับข้อมูลการใช้งานปกติที่มีอยู่แล้ว ซึ่งถ้าการเปรียบเทียบแล้วไม่ตรงกันแสดงว่าแพ็คเก็ตที่เข้ามามีพฤติกรรมที่จะบุกรุกระบบเครือขาย ระบบจะทำการแจ้งให้ผู้ดูแลระบบทราบ

2.3 รูปแบบการบุกรุกระบบเครือข่าย

2.3.1 Backdoor Attacks

Backdoor จัดอยู่ในประเภทของมัลแวร์ (Malware) โจมตีโดยการเปิดช่องโหว่เพื่อให้ผู้ประสงค์ร้ายเข้าถึงเครื่องคอมพิวเตอร์ได้ทุกเวลาที่มีการเปิดเครื่อง ส่วนใหญ่แฮกเกอร์จะทำการฝังโค๊ดไว้ในโปรแกรมที่ผู้ใช้เป็นคนติดตั้งเองและเมื่อมีการติดตั้งโปรแกรมจะแอบเปิดช่องโหว่ไว้เบื้องหลังเพื่อให้แฮกเกอร์สามารถเข้ามาได้อีกครั้ง โดยไม่จำเป็นต้องแฮกอะไรอีก

2.3.2 Teardrop Attacks

คือการส่ง Fragment IP ที่มีขนาดเหลื่อมล้ำหรือมีขนาดใหญ่เกินไป ทำให้ TCP/IP ของหลายระบบปฏิบัติการไม่สามารถจัดการกับ Fragment ได้อย่างถูกต้องทำให้เครื่องเป้าหมายเกิดความสับสนและหยุดทำงานในที่สุด

2.3.3 Load module

เป็นการพยายามเข้าถึงระบบของเป้าหมายโดยไม่ได้รับอนุญาต มีจุดประสงค์เพื่อเข้าถึงข้อมูลส่วนตัวหรือทำลายระบบงานของเป้าหมาย เป็นลักษณะที่ผู้โจมตีพยายามใช้งานในส่วนที่ไม่ได้รับอนุญาตในการเข้าถึงเช่น การเข้าใช้งานส่วนของ Super user (root)

2.3.4 Rootkit

เป็นรูปแบบการโจมตีที่สามารถซ่อนตัวในโปรแกรมหลัก (Root) ของเครื่องเป้าหมายได้ และเนื่องจาก Rootkit ไม่ใช่ไวรัส (Virus) หรือหนอน (Worm) จึงทำให้โปรแกรมแอนตี้ไวรัสไม่สามารถตรวจพบได้

2.3.5 Neptune Attacks

เป็นการโจมตีแบบ Dos (Denial of Service) เพื่อให้เครื่องของเหยื่อนั้นใช้ทรัพยากรหน่วยความจำมากเกินไป โดยการส่งแพ็คเก็ต TCP เพื่อขอเชื่อมต่อกับเครื่องเป้าหมายเป็นจำนวนมากด้วยทีอยู่ปลอม จึงทำให้เครื่องเป้าหมายไม่สามารถเชื่อมต่อกับเครื่องปลายทางได้ แต่พื้นที่หน่วยความจำยังถูกจัดสรรให้การเชื่อมต่อที่ถูกขอเข้ามา ทำให้หน่วยความจำของเครื่องเป้าหมายเต็มและไม่สารมารถให้บริการต่อได้

2.3.6 Phf Attacks

Phf Attacks จัดอยู่ในประเภทการโจมตีแบบ DoS (Denial of Service) เป็นการโจมตีเพื่อรบกวนและขัดขวางการทำงานของเครื่องคอมพิวเตอร์และอุปกรณ์เครือข่ายของเป้าหมาย

2.3.7 Satan

เป็นการตรวจสอบเป้าหมายโดยการส่งแพ็คเก็ตจำนวนมากไปยัง IP Address ต่างๆ เพื่อให้เครื่องเป้าหมายตอบกลับและรวบรวมข้อมูลต่างๆเพื่อหาช่องโหว่ในการโจมตี เช่น หมายเลข

พอร์ตที่กำลังเปิดอยู่ Service ที่กำลังทำงานอยู่ และสามารถสั่งเปิด Service ให้ทำงานผ่านพอร์ตที่ต้องการได้

2.3.8 Buffer Overflow Attacks

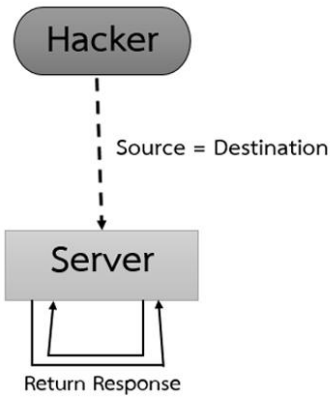
ลักษณะการโจมตีแบบ Buffer Overflow จะเป็นการเขียนข้อมูลเกินขอบเขตที่กำหนดทำให้ข้อมูลที่เขียนนั้นไปทับกับข้อมูลอื่นที่อยู่ในระบบ ทำให้การประมวลผลข้อมูลผิดพลาด ส่งผลให้เกิดการเปลี่ยนแปลงการทำงานของระบบให้เป็นไปตามที่ผู้โจมตีต้องการ

2.3.9 FTP Write Attacks

เป็นการโจมตีเว็บไซต์จากระยะไกลโดยผู้โจมตีจะเขียนไฟล์ลงไปใน Directory ของเครื่องเป้าหมายผ่านการเชื่อมต่อทาง โปรโตคอล FTP (File Transfer Protocol) ซึ่งเป็นโปรโตคอลที่ใช้ในการถ่ายโอนข้อมูล หากเครื่องเป้าหมายไม่ได้ตั้งค่าการป้องกันไว้ผู้โจมตีจะสามารถเขียนไฟล์ประเภทใดก็ได้ที่ผู้โจมตีต้องการ

2.3.10 Land Attacks

การโจมตีแบบ Land Attacks ผู้โจมตีจะทำการปลอม IP Address ของเครื่องต้นทางให้เหมือนกับเครื่องเป้าหมาย และทำการส่งคำขอเพื่อเชื่อมต่อไปยังเครื่องเป้าหมาย แต่เนื่องจากเครื่องต้นทางและเครื่องเป้าหมายมี IP Address เหมือนกันทำให้โปรโตคอลของเครื่องเป้าหมายไม่สามารถแยกแยะได้จึงทำการตอบกลับด้วย SYN ACK ออกไป ผลก็คือ SYN ACK จะย้อนกลับเข้าหาเครื่องของเป้าหมาย และเมื่อมีการส่งคำขออย่างต่อเนื่องจะทำให้เครื่องเป้าหมายเกิดปัญหาในการจัดสรรหน่วยความจำดังแสดงในรูปที่ 2.7



รูปที่ 2.7 แสดงการโจมตีแบบ Land Attacks

2.3.11 Spyware

คือซอฟต์แวร์ที่ถูกสร้างขึ้นเพื่อตรวจสอบการทำงานของคอมพิวเตอร์ โดยการทำงานของ Spyware จะส่งตัวเองไปทางช่องโหว่ของเว็บไซต์ต่างๆ เช่น popup โฆษณาโดยมีจุดประสงค์เพื่อให้เกิดความเสียหายต่อผู้ใช้งาน

2.3.12 IP Scanning

ผู้โจมตีนั้นสามารถค้นหาเป้าหมายที่ต้องการจะโจมตีได้ผ่านการทำ IP Scanning เปรียบเทียบได้กับการที่โจรนั้นเลือกบ้านที่ต้องการปล้นก่อนที่จะทำการปล้นจริง โดยผู้ที่โจมตีนั้นจะค้นหาว่ามีเครื่องใดในระบบที่กำลังใช้งานอยู่และมีช่องทางใดที่สามารถโจมตีได้บ้าง

2.3.13 Multi-hop Attacks

เป็นการโจมที่ผ่านการสื่อสารของเครือข่ายไร้สาย ซึ่งโดยทั่วไปการสื่อสารของเครือข่ายไร้สายจะเป็นแบบ Multi-hop การโจมตีแบบ Multi-hop Attacks จะเป็นการรบกวนช่องทางการสื่อสารและจะทำให้การส่งข้อมูลจากต้นทางไม่สามารถไปถึงปลายทางได้ ทำให้เกิดผลเสียต่อระบบเครือข่าย

2.3.14 Smurf Flooding Attacks

การโจมตีแบบ Smurf Flooding Attacks เป็นการส่ง ICMP Echo Request ไปยัง Broadcast Address ในเครือข่ายที่เป็นตัวกลางโดยปลอม Source IP Address เป็น IP Address ของระบบที่ต้องการโจมตี ซึ่งจะทำให้เครือข่ายที่เป็นตัวกลางส่ง ICMP Echo Reply กลับไปยัง IP Address ของเป้าหมายทันที ทำให้มีการใช้งานแบนวิดธ์อย่างเต็มที่และไม่สามารถให้บริการต่อได้

2.3.15 PoD Attacks

การโจมตีแบบ PoD ซึ่งย่อมาจาก Ping of Death นั้นเป็นการโจมตีเครื่องคอมพิวเตอร์ของเป้าหมายโดยการส่ง Ping ที่มีรูปแบบผิดปกติไปยังเครื่องของเป้าหมาย ทำให้เครื่องคอมพิวเตอร์ของเป้าหมายไม่สามารถจัดการกับ Ping ที่มีขนาดใหญ่กว่าขนาดแพ็คเก็ต IP สูงสุดและทำให้การเชื่อมต่อของเครื่องเป้าหมายล้มเหลว

2.3.16 Perl Attacks

Perl Attacks จัดอยู่ในการโจมตีแบบ DoS (Denial of Service) จุดประสงค์เพื่อทำให้เครื่องของเป้าหมายไม่สามารถใช้งานระบบเครือข่ายได้ โดยการส่งแพ็คเก็ตจำนวนมากไปยังเครื่องเป้าหมาย

2.3.17 IMAP Attacks

การโจมตีแบบ IMAP Attacks จะโจมตีผ่านทางโปรโตคอล IMAP (Internet Message Access Protocol) ซึ่งเป็นโปรโตคอลที่ใช้ในการส่งอีเมล์ มีความสามารถในการเข้าถึงข้อมูลทั้งในแบบออฟไลน์และออนไลน์

2.3.18 Warez master Attacks

เป็นการโจมตีโดยใช้ประโยชน์จากข้อผิดพลาดของระบบที่เกี่ยวข้องกับ FTP (File Transfer Protocol) โดยปกติผู้ใช้งานทั่วไปจะไม่ได้รับสิทธิ์ในการเขียนข้อมูลบนเซิร์ฟเวอร์ แต่จะมีการอนุญาตให้ดาวน์โหลดข้อมูลได้ การโจมตีลักษณะนี้เกิดขึ้นเมื่อมีการขอสิทธิ์ในการดาวน์โหลดไฟล์แต่ FTP Server ให้สิทธิ์ในการเขียนไฟล์ด้วยโดยไม่ตั้งใจ จึงทำให้ผู้โจมตีสามารถเขียนไฟล์ที่ต้องการไปยัง Server ได้

2.3.19 Warez client Attacks

เป็นการโจมตีที่เกิดขึ้นหลังจากสามารถโจมตีแบบWarez master Attacks สำเร็จ และสามารถวางไฟล์ที่ต้องการลงในไดเร็กทอรี่ของเครื่องเป้าหมายได้ เมื่อมีการเข้าถึงและดาวน์โหลดไฟล์ดังกล่าวโดยเครื่อง Client จะทำให้ผู้โจมตีนั้นสามารถโจมตีเครื่อง Client ได้ด้วย

2.3.20 Nmap

เป็นเครื่องมือที่มีวัตถุประสงค์ที่ใช้ในแสกนระบบเครือข่าย ซึ่งถูกประเมินว่าเป็นเครื่องมือที่สมบูรณ์แบบที่สุดที่ใช้ในการแสกน โดยหลักๆนั้นผู้โจมตีจะแสกนเพื่อที่จะหาข้อมูลเกี่ยวกับเครื่องนั้นๆ เช่น ดูว่าเปิดพอร์ตไหนไว้บ้าง หรือดูว่า ใช้ระบบปฏิบัติการแบบไหน มีการเปิด firewall หรือไม่เป็น

2.3.21 Port Scanning

เป็นการที่ผู้โจมตีทำการค้นหาช่องโหว่เพื่อเข้าโจมตีระบบ โดยการทำ Port Scanning นั้น ผู้โจมตีสามารถได้ข้อมูลจำนวนมากจากเป้าหมายที่โจมตีได้ เช่น Service ต่างๆที่กำลังรันอยู่ในเครื่องของเป้าหมาย การสนับสนุนการเข้าสู่ระบบด้วย Anonymous หรือไม่ และ Service ด้านเครือข่ายอื่นๆ การทำ Port Scanning นั้นจะเป็นการส่งแพ็คเก็ตไปยังแต่ละพอร์ตและรอผลตอบกลับว่าพอร์ตดังกล่าวกำลังถูกใช้อยู่หรือไม่

2.3.22 Guess Password

เป็นการพยายามเข้าสู่ระบบของผู้ที่ไม่ใช่เจ้าของบัญชีตัวจริง โดยการเดารหัสผ่านด้วยวิธีการต่างๆ เช่นการทำ Brute Force Attack หากบัญชีใดที่ถูกโจมตีตั้งรหัสผ่านที่สามารถเดาได้ง่ายนั้น การโจมตีด้วยวิธีนี้ผู้โจมตีจะใช้เวลาไม่นานก็สามารถเข้าสู่ระบบได้

2.4 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล [5] เป็นกระบวนการหาประโยชน์หรือความรู้ต่างๆที่ไม่เคยรู้มาก่อนจากฐานข้อมูลขนาดใหญ่ เช่น รูปแบบ ความสัมพันธ์และกฎ ที่ซ่อนอยู่ภายในฐานข้อมูล เพื่อนำความรู้ที่ได้ไปใช้ประโยชน์และประกอบการตัดสินใจได้ ยกตัวอย่างเช่นการเก็บข้อมูลลูกค้าของห้างสรรพสินค้าโดยการให้ลูกค้านั้นสมัครสมาชิก และนำข้อมูลเหล่านั้นไปกลั่นกรองโดยผ่านเทคนิคการทำเหมืองข้อมูลเพื่อที่จะสามารถจัดโปรโมชั่นต่างๆให้ลูกค้า และสามารถจัดวางสินค้าที่ลูกค้าชอบซื้อควบคู่กันไว้ใกล้ๆกัน การจำแนกข้อมูล (Classification) เป็นหนึ่งในเทคนิคการทำเหมืองข้อมูล ซึ่งการจำแนกข้อมูลนั้นเป็นกระบวนการทางสถิติเพื่อจัดประเภท และวิเคราะห์หารูปแบบของกลุ่มข้อมูลใหม่ โดยทำการแบ่งข้อมูลออกเป็น 2 กลุ่มหลักๆ คือ ข้อมูลที่ใช้ในการสอน (Training Data) และข้อมูลที่ใช้ในการทดสอบ (Test Set) เพื่อนำผลลัพธ์ที่ได้มาเป็นแบบจำลองหรือโมเดล

2.4.1 งานของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลนั้นสามารถนำไปใช้ประโยชน์กับงานได้หลายด้าน แต่ในทางปฏิบัติจริงนั้นไม่มีเทคนิคหรือเครื่องมือใดที่สามารถนำไปใช้งานกับงานทุกชนิดได้อย่างเหมาะสม แต่ละเทคนิคนั้นจะเหมาะสมกับชนิดของงานแตกต่างกันไป ซึ่งจะมีเทคนิคในการทำเหมืองข้อมูลที่แตกต่างกันไปขึ้นอยู่กับชนิดของงานดังนี้

2.4.1.2 การจัดหมวดหมู่ (Classification) การจัดหมวดหมู่ถือเป็นงานหลักๆของการทำเหมืองข้อมูล โดยการจัดหมวดหมู่นั้นจะทำการสำรวจจุดเด่นของข้อมูล เพื่อที่จะใช้เป็นตัวแบ่งหมวดหมู่ ยกตัวอย่างการจัดหมวดหมู่สัตว์ ซึ่งจะสำรวจลักษณะเด่นว่าควรจัดให้อยู่ในสัตว์กลุ่มไหน เช่น สัตว์ปีก สัตว์เลื้อยคลาน สัตว์เลี้ยงลูกด้วยนม เป็นต้น

2.4.1.3 การประเมินค่า (Estimation) การประเมินค่าข้อมูลอย่างต่อเนื่องจะก่อให้เกิดประโยชน์ต่อธุรกิจได้ การนำข้อมูลต่างๆที่มีอยู่มาวิเคราะห์เพื่อประเมินตัวแปรที่เราไม่รู้ค่าแน่นอนเช่น การประเมินรายได้รวมของครอบครัว

2.4.1.4 การทำนายล่วงหน้า (Prediction) เป็นการทำงานที่มีลักษณะคล้ายกันกับการจัดหมวดหมู่ การทำนายล่วงหน้าจะใช้สถิติย้อนหลังเพื่อประเมินพฤติกรรมและโอกาสที่จะเกิดขึ้นในอนาคต ตัวอย่างการทำนายเช่น การทำนายปริมาณน้ำฝนในแต่ละปี การทำนายจำนวนลูกค้าในห้างสรรพสินค้าในอีก 3 เดือนข้างหน้า เป็นต้น

2.4.1.5 การจัดกลุ่มโดยอาศัยความใกล้ชิด (Affinity Group) คือการตัดสินใจรวมสิ่งใดสิ่งหนึ่งให้เข้าไปอยู่ด้วยกัน โดยอาศัยลักษณะที่มีความใกล้ชิดกันเช่นการจัดวางสินค้าในห้างสรรพสินค้า

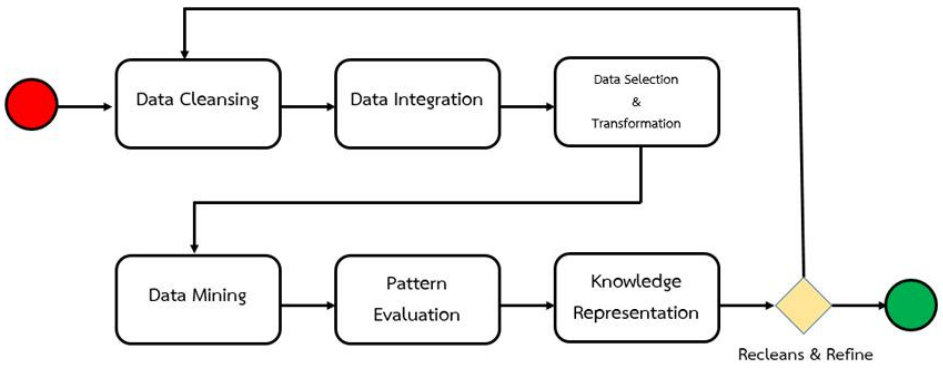
2.4.1.6 การรวมตัว (Clustering) คือการรวมส่วนต่างๆในแต่ละส่วนให้อยู่รวมกันเป็นกลุ่มย่อยหรือที่เรียกว่าคลัสเตอร์ (Clusters) ซึ่งแต่ละส่วนย่อยอาจจะประกอบด้วยส่วนที่ต่างชนิดกัน ซึ่งความแตกต่างของการรวมตัวจากการจัดหมวดหมู่คือ การรวมตัวจะไม่พึ่งพาอาศัยการกำหนดหมวดหมู่ล่วงหน้า

2.4.1.7 การบรรยาย (Description)

วัตถุประสงค์ของการทำเหมืองข้อมูลด้วยวิธีการนี้ คือต้องการที่จะอธิบายความสับสนของฐานข้อมูลเพื่อให้เพิ่มความเข้าใจกระบวนการให้มากขึ้น การทำเหมืองข้อมูลส่วนใหญ่นั้นต้องการข้อมูลจำนวนมากเพื่อที่จะสร้างกฎที่ใช้ในการจัดกลุ่ม การจำแนก การทำนายล่วงหน้า ดังนั้นถ้าข้อมูลมีขนาดใหญ่เท่าไหร่ก็ยิ่งจะนำไปสู่ความน่าเชื่อถือของผลสรุป

2.4.2 ขั้นตอนการทำเหมืองข้อมูล

ขั้นตอนการทำเหมืองข้อมูล [6] ที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้นั้นจะประกอบไปด้วยขั้นตอนแยกย่อยที่แสดงรูปที่ 2.8



รูปที่ 2.8 แสดงขั้นตอนการทำเหมืองข้อมูล

2.4.2.1 การทำความสะอาดข้อมูล (Data cleansing) เป็นขั้นตอนการลบข้อมูลต่างๆที่ไม่ต้องการ รวมถึงการปรับปรุงแก้ไขข้อมูลให้มีประสิทธิภาพ

2.4.2.2 การคัดเลือกข้อมูลและการเปลี่ยนรูปแบบข้อมูล (Data Selection & Data Transformation) เป็นขั้นตอนเลือกการนำข้อมูลที่จะใช้ในการวิเคราะห์จากแหล่งข้อมูลและทำการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมในการทำเหมืองข้อมูล เพราะข้อมูลที่นำมาจากแต่ละแหล่งนั้นจะมีความแตกต่างกัน ขั้นตอนนี้จึงจำเป็นต้องเลือกข้อมูลและแปลงรูปแบบข้อมูลให้อยู่ในรูปแบบที่โปรแกรมสามารถคำนวณได้

2.4.2.3 การทำเหมืองข้อมูล (Data Mining) เป็นขั้นตอนการหารูปแบบที่มีประโยชน์จากข้อมูลที่ได้นำมาวิเคราะห์

2.4.2.4 การประเมินผลรูปแบบ (Pattern Evaluation) เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูลเพื่อนำไปเป็นตัวแทนของความรู้

2.4.2.5 การนำเสนอความรู้ (Knowledge Representation) เป็นขั้นตอนการนำเสนอความรู้ที่ได้จากการค้นหาความรู้จากการทำเหมืองข้อมูล โดยใช้เทคนิคในการนำเสนอเพื่อให้สามารถเข้าใจได้ง่าย

2.4.3 การวัดประสิทธิภาพ

การวัดประสิทธิภาพในงานวิจัยนี้ได้เปรียบเทียบค่าความแม่นยำ ค่าระลึก ค่า F-Measure และค่าความถูกต้อง [3]

2.4.3.1 ค่าความแม่นยำ Precision คำนวณจากจำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้นหารด้วยจำนวนของข้อมูลที่ถูกจำแนกว่าเป็นกลุ่มนั้นทั้งหมด โดยค่า Precision นั้นจะเป็นการวัดความสามารถของแบบจำลองโดยการขจัดข้อมูลที่ไม่เกี่ยวข้องออกไป ผลลัพธ์จะบ่งบอกว่าสามารถขจัดการจำแนกประเภทที่ผิดพลาดได้มากน้อยเพียงใด ดังสมการที่ 1

$$\text{precision} = \frac{tp}{tp+fp} \qquad (2.1)$$

2.4.3.2 ค่าระลึก Recall คำนวณจากจำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้นหารด้วยจำนวนของข้อมูลที่มีอยู่จริงในกลุ่มนั้น โดยค่า Recall นั้นจะเป็นการวัดความสามารถของแบบจำลองว่าการจำแนกข้อมูลที่เกี่ยวข้องออกมานั้นมีประสิทธิภาพมากน้อยเพียงใด ดังสมการที่ 2

$$\text{recll} = \frac{tp}{2tp+fn} \qquad (2.2)$$

2.4.3.3 ค่า F-Measure เป็นการวัดค่าความแม่ยำ Precision และค่าระลึก Recall พร้อมกันของแบบจำลอง ดังสมการที่ 3

$$\text{F - Measure} = \frac{2 \times \text{precision} \times \text{Recll}}{\text{precision} + \text{Recll}} \qquad (2.3)$$

2.4.3.4 ค่า Accuracy เป็นการวัดค่าความถูกต้องในการจำแนก ดังสมการที่ 4

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (2.4)$$

TP (True Positive) คือ จำนวนข้อมูลที่จำแนกถูกต้องในคลาสนั้นๆ

TN (True Negative) คือ จำนวนข้อมูลที่จำแนกถูกต้องทีไม่อยู่ในคลาสนั้นๆ

FP (False Positive) คือ จำนวนข้อมูลที่จำแนกไม่ถูกต้องที่อยู่ในคลาสนั้นๆ

FN (False Negative) คือ จำนวนข้อมูลที่จำแนกไม่ถูกต้องที่ไม่อยู่ในคลาสนั้นๆ

2.4.4 เทคนิค Decision Table

เทคนิคตารางตัดสินใจ (Decision Table) [7] เป็นวิธีที่นำมาทดสอบการทำงานร่วมกันของเงือนไขที่มีหลายเงื่อนไข มีลักษณะคล้ายกับต้นไม้ตัดสินใจ (Decision Tree) แต่จะอยู่ในรูปของตาราง ซึ่งตารางจะประกอบไปด้วยเงื่อนไข (Conditions) และการกระทำ (Actions) ดังแสดงในตารางที่ 2.2

ตารางที่ 2.2 ตารางการตัดสินใจของ Decision Table

เงื่อนไข (Condition)	กฎการตัดสินใจ / การกระทำ (Action)
ระบุเงื่อนไขสำหรับการพิจารณาการทำงาน	กฎที่เป็นไปได้ภายใต้เงื่อนไขที่ระบุ
การกระทำที่เป็นไปได้	ระบุการเลือกการกระทำภายใต้กฎ

2.4.5 เทคนิค Naïve Bayes

เทคนิคการจำแนกข้อมูลด้วย Naïve Bayes [3] เป็นวิธีที่ได้รับความนิยมในการนำมาใช้จำแนกข้อมูล เนื่องจากมีแบบจำลองที่เข้าใจได้ง่ายและไม่ซับซ้อน ซึ่งเทคนิคนี้ใช้หลักการของความน่าจะเป็นโดยแสดงดังสมการที่ 6

$$P(B/A) = \frac{P(A \sqcap B)}{P(B)} \quad (2.5)$$

จากสมการสามารถอธิบายได้ดังนี้

P(A/B) หมายถึง ความน่าจะเป็นที่มีเหตุการณ์ B เกิดขึ้นก่อนและมีเหตุการณ์ A เกิดขึ้นตาม

P(A ∏ B) หมายถึง ความน่าจะเป็นที่มีเหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน

P(B) หมายถึง ความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้นในลักษณะเดียวกัน หากเหตุการณ์ A เกิดขึ้นก่อนจะเขียนสมการได้ดังนี้

$$P(B/A) = \frac{P(A \prod B)}{P(A)} \quad (2.6)$$

จากทั้ง 2 แบบจะมีค่าที่เหมือนกันอยู่คือ P(A∏B) ดังนั้นสามารถเขียนเป็นสมการได้ดังนี้

$$P(A \prod B) = P(A/B) \times P(B) = P(B/A) \times P(A) \quad (2.7)$$

$$P(B/A) = \frac{P(A/B) \times P(B)}{P(A)} \quad (2.8)$$

สมการดังกล่าวคือ Bayes theorem หรือทฤษฎีของเบย์ ในการนำไปใช้จำแนกข้อมูลนั้น เพื่อให้เข้าใจง่าย B จะถูกเปลี่ยนเป็น C ซึ่งหมายถึง ค่าของคลาส (Class) และ A คือ แอททริบิว (Attribute) ดังสมการต่อไปนี้

$$P(C/A) = \frac{P(A/C) \times P(C)}{P(A)} \quad (2.9)$$

จากสมการสามารถอธิบายได้ดังนี้

P(C/A) หมายถึง ค่าความน่าจะเป็นของข้อมูลที่มีแอททริบิว A จะมีคลาส C

P(A/C) หมายถึง ค่าความน่าจะเป็นของข้อมูลที่เป็น Training data จะมีคลาส C และมีแอททริบิว A

P(C) หมายถึง คือความน่าจะเป็นของคลาส C

2.4.6 เทคนิค RIPPER

เทคนิค RIPPER (Rule-Based Classification) [8] เป็นเทคนิคที่พัฒนาจากเทคนิค IRIP สามารถสร้างกฎเองได้ โดยการเรียนรู้จากข้อมูลที่เตรียมไว้ให้ซึ่งกฎที่สร้างขึ้นจะอยู่ในรูปของ if then else โดยมีขั้นตอนหลักอยู่ 3 ขั้นตอน ขั้นตอนแรกคือการสร้างกฎเริ่มต้น (Building) ซึ่งจะแบ่งเป็น 2 กระบวนการคือกระบวนการเจริญเติบโต (Growth) โดยกระบวนการนี้จะทำการเพิ่มจำนวนกฎให้เหมาะสมกับข้อมูลจากนั้นจะตัดกฎที่ไม่ไม่จำเป็นหรือกฎที่ลดประสิทธิภาพในการเรียนรู้ออก (Pruning) ขั้นตอนที่ 2 คือขั้นตอนการเพิ่มประสิทธิภาพ (Optimization) โดยจะมีการเพิ่มคุณลักษณะให้แต่ละกฎ ขั้นตอนที่ 3 คือขั้นตอนการลบกฎออกจาก Rule set และเลือกเฉพาะกฎที่ดีที่สุดเก็บไว้

2.4.7 เทคนิค PART decision list

เทคนิค PART decision list [9] เป็นเทคนิคที่พัฒนาจาก C4.5 และ RIPPER โดยรวมทั้ง 2 เทคนิคเข้าด้วยกัน มีจุดเด่นคือสามารถเรียนรู้กฎได้เองเหมือนเทคนิค RIPPER และสามารถจัดการกับข้อมูลที่หายไปและคุณลักษณะทางตัวเลขที่ต่างกันได้ดี และได้มีการทดสอบแล้วว่ามีความแม่นยำมากกว่าเทคนิค RIPPER

2.5 งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ผู้วิจัยได้ทำการศึกษาเทคนิควิธีการตรวจจับการบุกรุกโดยการใช้เทคนิคเหมืองข้อมูล ซึ่งในปัจจุบันมีผู้นำเสนอเทคนิคต่างๆ อย่างต่อเนื่อง ดังนี้

อรนุช พันโท และมนต์ชัย เทียนทอง [10] ได้เสนอการเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบข้อมูลการเรียนรู้ VARK ด้วยเทคนิคเหมืองข้อมูล โดยใช้วิธีการ 10–fold validation แล ใช้เทคนิคการจำแนกข้อมูล 3 เทคนิคคือ Bayes, Decision Tree, และ Rules-Based ผลปรากฏว่าการ จำแนกข้อมูลด้วยเทคนิค Decision Tree มีประสิทธิภาพสูงที่สุดคือ 82.78%

ลลินี อินทราราม [11] ได้เสนอการจำแนกข้อมูลรูปแบบการบุกรุกบนระบบเครือข่าย โดยใช้แบบจำลอง แรนดอมฟอเรส ซัพพอร์ทเวกเตอร์แมชชีนและการผสมผสานขั้นตอนเข้าด้วยกัน ซึ่งเลือกใช้ชุดข้อมูลจากฐานข้อมูลความรู้ KDD Cup’99 ผลปรากฏว่าแบบจำลองที่ให้ผลดีที่สุดคือ แรนดอมฟอเรสโดยมีค่าความถูกต้อง 99.95%

เยาวภา ภารสำเร็จ [12] ได้เปรียบเทียบอัลกอริทึมที่ใช้ในการวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา โดยเปรียบเทียบอัลกอริทึมเหมืองข้อมูล 3 อัลกอริทึมคือ C4.5, Naïve Bayes และ k-Nearest Neighbor โดยใช้ข้อมูลของนักศึกษาจำนวน 4,591 ชุด ผลการเปรียบเทียบปรากฏว่า C4.5 ได้ผลที่ดีที่สุดโดยมีค่าความถูกต้อง 73.55%

ปรีชา สมหวัง และศิริวัฒน์ โทศิริกุล [13] ได้นำเสนอแนวคิดในการตรวจจับการใช้งานคอมพิวเตอร์ในทางที่ผิด โดยใช้การวิธีการวิเคราะห์องค์ประกอบหลักเพื่อคัดแยกข้อมูลและใช้ฟัซซี ซีมีนเพื่อจัดกลุ่ม ผลการทดลองมีความแม่นยำ 81.48% และมีความผิดพลาดในการตรวจจับ 18.52%

ปวีณา ชัยวนารมย์ [14] ได้เสนอการพยากรณ์การเกิดความเครียดในหลายระดับด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้กลุ่มตัวอย่าง 300 คน และใช้อัลกอริทึมในการทำเหมืองข้อมูลจำนวน 6 อัลลกอริทึมในการสร้างแบบจำลอง คือ Bayesian Network, Naïve Bayesian, Decision Tree, Decision Table, Partial Rules (PART) และ Multilayer Perceptron (MLP) จากการทดสอบพบว่าแบบจำลองที่เหมาะสมในการพยากรณ์ความเครียดคือแบบจำลองของอัลกอริทึม Multilayer Perceptron (MLP) ซึ่งมีค่าความถูกต้องเท่ากับ 81% ค่าความแม่นยำเท่า 0.81 ค่าความระลึกเท่ากับ 0.81 และค่าความเหวี่ยงเท่ากับ 0.81

กรมวุฒิ นงนุช และธนพล เจนสุทธิเวชกุล [15] ได้เสนอเทคนิคการตรวจจับการบุกรุกเครือข่ายโดยใช้โครงข่ายประสาทเทียมกับกระบวนการทางพันธุศาสตร์ ซึ่งเลือกใช้ชุดข้อมูลจากฐานข้อมูลความรู้ KDD Cup’99 ผลปรากฏว่าการตรวจจับการบุกรุกมีความถูกต้องอยู่ที่ 93.35%

ณัฐวุฒิ ปั้นรูป และอัฐพร กิ่งบู [16] ได้เสนอวิธีการจำแนกข้อมูลให้ดีขึ้นโดยมีการรวมเทคนิคจัดกลุ่มเคมีนและเทคนิคจำแนกข้อมูลเอสวีเอ็มแบบหลายกลุ่มทำงานร่วมกัน ซึ่งเลือกใช้ชุด

ข้อมูลจากฐานข้อมูลความรู้ KDD Cup’99 ผลปรากฏว่าเทคนิคที่ได้ผลดีที่สุดคือเทคนิค KMMSVM มีค่าความถูกต้องอยู่ที่ 99.32% ตามด้วยเทคนิคการจัดกลุ่มเคมีนซึ่งมีค่าความถูกต้องอยู่ที่ 99.19%

ธนกร มีหินกอง [17] ได้เสนอเทคนิคการตรวจหาการบุกรุกเป็นระบบที่ใช้ในการตรวจหาผู้ที่บุกรุกเข้ามาในเครือข่ายคอมพิวเตอร์เพื่อมุ่งทำลายระบบหรือขโมยข้อมูล ด้วยเทคนิคเมืองข้อมูลวิเคราะห์ด้วยกฎความสัมพันธ์จากโครงข่ายประสาทเทียม ผลจากการทดลองพบว่าระบบการตรวจหาการบุกรุกที่ได้พัฒนาขึ้นนี้สามารถรายงานผลได้อย่างรวดเร็วโดยมีค่าความเที่ยงที่ 97.4 %

นิเวศ จิระวิชิตชัย [18] ได้เปรียบเทียบอัลกอริทึมเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติทดสอบประสิทธิภาพในการจำแนก สำหรับข้อมูลทางการแพทย์ Decision tree, Support Vector Machine ผลการเปรียบเทียบปรากฏว่า Decision tree ให้ประสิทธิภาพที่ดีที่สุด 99.57%

ไพชยนต์ คงไชย [19] ได้เสนอการพัฒนาขั้นตอนวิธีเพื่อจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์แบบคลุมเครือที่กะทัดรัดเพื่อเพิ่มประสิทธิภาพการจำแนกประเภทข้อมูล โดยเปรียบเทียบอัลกอริทึม CCFAR กับอีก 9 อัลกอริทึม คือ C4.5, RIPPER, OneR, CBA, GARC, OAC, FURIA, CFAR และ CFARC ผลปรากฏว่าอัลกอริทึม CCFAR มีความเหมาะสมของกฎอยู่ในอันดับที่ 1 คือ 0.9104 ซึ่งมีค่าความเหมาะสมของกฎแตกต่างจากอัลกอริทึม CFARC ที่เป็นอันดับ 2 ไม่มาก

บทที่ 3

วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยในงานวิจัยนี้ ได้ใช้ข้อมูลจากฐานข้อมูลความรู้ KDD Cup’99 มาใช้ในการเปรียบเทียบหาเทคนิคที่มีประสิทธิภาพในการจำแนก ซึ่งมีขั้นตอนคือ การเตรียมข้อมูล การสร้างแบบจำลอง การวัดประสิทธิภาพแบบจำลอง

3.1 การเตรียมข้อมูล

งานวิจัยนี้ผู้วิจัยใช้ข้อมูลจากฐานข้อมูลความรู้ KDD Cup'99 เป็นชุดข้อมูลที่ในการทดสอบระบบความปลอดภัยของระบบเครือข่าย ใช้ในการแข่งขัน The Third International Knowledge Discovery and Data Mining ซึ่งข้อมูลที่สร้างขึ้นจากการเก็บข้อมูลการโจมตีที่ Lincoln Laboratory ของสถาบัน MIT (Massachusetts Institute of Technology) ใช้ใชระยะเวลาในการเก็บข้อมูล 9 สัปดาห์ และผู้วิจัยได้คัดเลือกข้อมูลจำนวน 494,020 เรคคอร์ด 41 แอททริบิว คลาสผลลัพธ์จำนวน 23 คลาส ดังตารางที่ 3.1

ตารางที่ 3.1 แอททริบิวของข้อมูล

ลำดับ	ชื่อแอททริบิว	สัญลักษณ์แอททริบิว	ชนิดแอททริบิว
1	Duration	duration	Numeric
2	Protocol type	protocol_type	Nominal
3	Service	service	Nominal
4	Flag	flag	Nominal
5	Source bytes	src_bytes	Numeric
6	Destination bytes	dst_bytes	Numeric
7	land	land	Nominal
8	Wrong fragment	wrong_fragment	Numeric
9	Urgent	urgent	Numeric
10	Hot	hot	Numeric
11	Number failed logins	num_failed_logins	Numeric
12	Logged in	logged_in	Nominal

ตารางที่ 3.1 (ต่อ)

ลำดับ	ชื่อแอททริบิว	สัญลักษณ์แอททริบิว	ชนิดแอททริบิว
13	Logged number compromised	lnum_compromised	Numeric
14	Logged root shell	lroot_shell	Numeric
15	ISU attempted	isu_attempted	Numeric
16	Logged number root	lnum_root	Numeric
17	Logged number file creations	lnum_file_creations	Numeric
18	Logged number shells	lnum_shells	Numeric
19	logged number access files	lnum_access_files	Numeric
20	logged number outbound cmds	lnum_outbound_cmds	Numeric
21	Is host login	is_host_login	Nominal
22	Is guest login	is_guest_login	Nominal
23	count	count	Numeric
24	Service count	srv_count	Numeric
25	S error rate	serror_rate	Numeric
26	Service SYN error rate	srv_serror_rate	Numeric
27	Error rate	rerror_rate	Numeric
28	Service REJ error rate	srv_rerror_rate	Numeric
29	Same service rate	same_srv_rate	Numeric
30	Diff service rate	diff_srv_rate	Numeric
31	Service diff host rate	srv_diff_host_rate	Numeric
32	Destination host count	dst_host_count	Numeric
33	Destination host service count	dst_host_srv_count	Numeric
34	Destination host same service rate	dst_host_same_srv_rate	Numeric
35	Destination host diff service rate	dst_host_diff_srv_rate	Numeric

ตารางที่ 3.1 (ต่อ)

ลำดับ	ชื่อแอททริบิว	สัญลักษณ์แอททริบิว	ชนิดแอททริบิว
36	Destination host same source port rate	dst_host_same_src_port_rate	Numeric
37	Destination host service diff host rate	dst_host_srv_diff_host_rate	Numeric
38	Destination host SYN error rate	dst_host_serror_rate	Numeric
39	Destination host Service SYN error rate	dst_host_srv_serror_rate	Numeric
40	Destination host REJ error rate	dst_host_rerror_rate	Numeric
41	Destination host Service REJ error rate	dst_host_srv_rerror_rate	Numeric
42	Class	label	No

ตารางที่ 3.2 คลาส

ลำดับ	ชื่อคลาส
1	Back
2	Teardrop
3	Load module
4	Neptune
5	Root kit
6	PHF
7	Satan
8	Buffer overflow
9	FTP write
10	Land
11	Spy
12	IP sweep
13	Multi hop

ตารางที่ 3.2 (ต่อ)

ลำดับ	ชื่อคลาส
14	Smurf
15	Pod
16	Perl
17	Warez client
18	Nmap
19	Imap
20	Warez master
21	Port sweep
22	Normal
23	Guess password

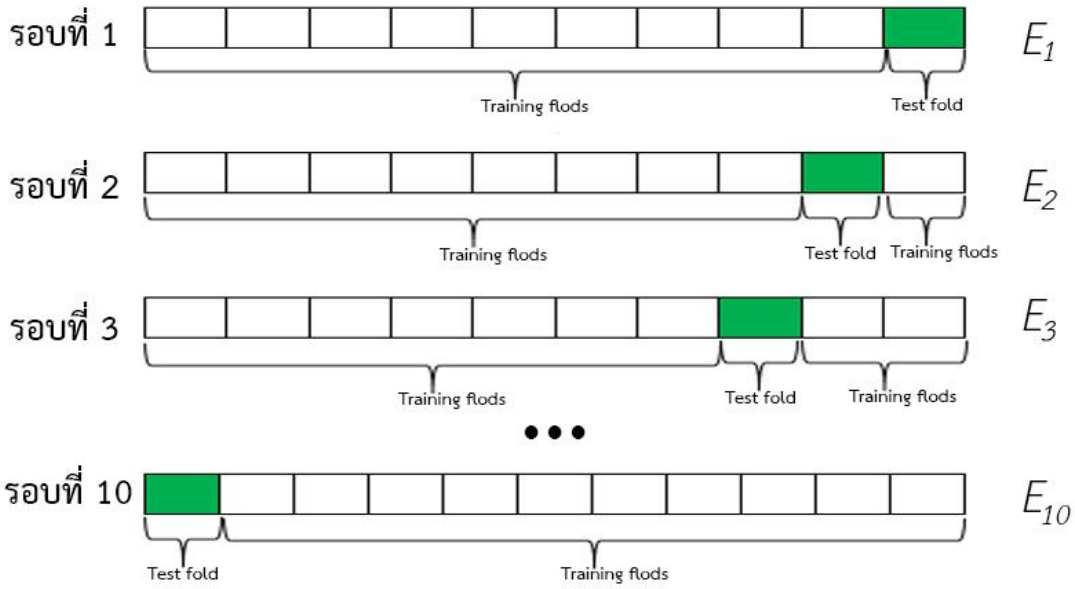
3.2 การสร้างแบบจำลอง

การสร้างแบบจำลองนั้นผู้วิจัยได้เลือกเทคนิคการจำแนกที่มีประสิทธิภาพ 4 เทคนิค ดังนี้

1. เทคนิค Decision Table
2. เทคนิค Naïve Bayes
3. เทคนิค RIPPER
4. เทคนิค PART Decision list

3.3 วัดประสิทธิภาพแบบจำลอง

การวัดประสิทธิภาพแบบจำลองในงานวิจัยนี้ผู้วิจัยได้นำหลักการ 10-fold cross validation มาใช้ เพื่อให้ข้อมูลทุกชุดเป็นทั้งชุดการสอนและชุดการทดสอบ โดยใช้ข้อมูล 9 ชุดเพื่อเป็นข้อมูลในการสอน (Training data) และใช้ข้อมูลอีก 1 ชุด เป็นข้อมูลในการทดสอบ (Testing Data) หลังจากนั้นทำการสลับข้อมูลและวนทำซ้ำจนครบ 10 รอบ ดังแสดงในรูปที่ 3.1



รูปที่ 3.1 รูปแบบการการทดสอบตามวิธีการ 10 – fold cross validation

ซึ่งวิธีการ 10-fold cross validation มีสมการเป็นดังนี้

$$E = \frac{1}{10}\sum_{\varepsilon=1}^{10} E_i \tag{3.1}$$

การวัดประสิทธิภาพของแบบจำลอง ค่าที่ใช้ในการวัดประสิทธิภาพ ได้แก่

1. ค่าความแม่นยำ (Precision)
2. ค่าเรียกคืน (Recall)
3. ค่า F-Measure

บทที่ 4

ผลการวิจัย

ผลการวิจัยการจำแนกข้อมูลสำหรับตรวจจับการบุกรุกโดยใช้เทคนิคการจำแนก Decision Table, Naïve Bayes, RIPPER และ PART Decision list ทดลองสร้างโมเดลจากฐานข้อมูล KDD Cup'99 ซึ่งใช้ข้อมูลจำนวน 494,020 เรคคอร์ด แอททริบิวจำนวน 41 แอททริบิวและคลาสผลลัพธ์ 23 คลาส ทดลองด้วยวิธีแบบ 10-fold cross validation สามารถแสดงประสิทธิภาพของแบบจำลองและแบบจำลองดังต่อไปนี้

4.1 ประสิทธิภาพของแบบจำลอง

4.1.1 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Decision Table จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Decision Table มีกฎจำนวน 1,105 กฎ สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.1

ตารางที่ 4.1 ผลการจำแนกโดยใช้เทคนิค Decision Table

Class	Precision (%)	Recall (%)	F-Measure (%)
back	0	99.8	95.3
teardrop	99.7	99.6	99.6
loadmodule	0	0	0
neptune	99.9	100	99.9
rootkit	0	0	0
phf	100	50	66.7
satan	99.3	96.1	97.7
buffer_overflow	100	50	66.7
ftp_write	0	0	0
land	94.7	85.7	90
spy	0	0	0
ipsweep	91.7	92.2	92
multihop	0	0	0

ตารางที่ 4.1 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
back	0	99.8	95.3
teardrop	99.7	99.6	99.6
loadmodule	0	0	0
neptune	99.9	100	99.9
rootkit	0	0	0
phf	100	50	66.7
satan	99.3	96.1	97.7
buffer_overflow	100	50	66.7
ftp_write	0	0	0
land	94.7	85.7	90
spy	0	0	0
ipsweep	91.7	92.2	92
multihop	0	0	0
smurf	99.9	100	99.9
pod	98.5	99.2	98.9
perl	0	0	0
warezclient	83.7	93.5	88.3
nmap	91.1	48.9	63.7
imap	100	8.3	15.4
warezmaster	50	15	23.1
portsweep	90.3	91.6	91
normal	99.9	99.4	99.6
guess_passwd	100	90.6	95
Average	**65.16**	**57.39**	**60.12**
Standard Deviation	**44.29**	**42.71**	**41.94**

จากตารางที่ 4.1 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Decision Table พบว่าค่า Precision มี Average อยู่ที่ 65.16% และ Standard Deviation อยู่ที่ 44.29% ค่า

Recall มี Average อยู่ที่ 57.39% และ Standard Deviation อยู่ที่ 42.71% และค่า F-measure มี Average อยู่ที่ 60.12% และ Standard Deviation อยู่ที่ 41.94

4.1.2 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Naïve Bayes

จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Naïve Bayes สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.2

ตารางที่ 4.2 ผลการจำแนกโดยใช้เทคนิค Naïve Bayes

Class	Precision (%)	Recall (%)	F-Measure (%)
back	10	99.7	99.8
teardrop	9.24	99.9	96
loadmodule	0	0	0
neptune	10	99.9	10
rootkit	0.33	40	6
phf	10	50	66.7
satan	9.82	91.3	94.6
buffer_overflow	2.88	76.7	41.8
ftp_write	0.87	25	12.9
land	1.89	100	31.8
spy	0	0	0
ipsweep	7.29	92.9	81.7
multihop	0	0	0
smurf	10	100	100
pod	7.79	98.9	87.1
perl	0	0	0
warezclient	8.15	99.6	89.6
nmap	3.41	47.2	39.6
imap	2.86	83.3	42.6
warezmaster	2.21	75	34.1
portsweep	9.05	97.7	93.9

ตารางที่ 4.2 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
normal	9.99	98.9	99.4
guess_passwd	8.5	96.2	90.3
Average	**5.40**	**68.36**	**52.95**
Standard Deviation	**4.07**	**37.92**	**38.92**

จากตารางที่ 4.2 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Naïve Bayes พบว่าค่า Precision มี Average อยู่ที่ 5.40% และ Standard Deviation อยู่ที่ 4.07% ค่า Recall มี Average อยู่ที่ 68.36% และ Standard Deviation อยู่ที่ 37.92% และค่า F-measure มี Average อยู่ที่ 52.95% และ Standard Deviation อยู่ที่ 38.92

4.1.3 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค RIPPER

จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค RIPPER มีกฎจำนวน 61 กฎ สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.3

ตารางที่ 4.3 ผลการจำแนกโดยใช้เทคนิค RIPPER

Class	Precision (%)	Recall (%)	F-Measure (%)
back	100	100	100
teardrop	100	100	100
loadmodule	20	11.1	14.3
neptune	100	100	100
rootkit	12.5	10	11.1
phf	66.7	100	80
satan	99.5	99.4	99.5
buffer_overflow	73.3	73.3	73.3
ftp_write	0	0	0
land	95.5	100	97.7
spy	0	0	0
ipsweep	99.7	99.3	99.5

ตารางที่ 4.3 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
ipsweep	99.7	99.3	99.5
multihop	28.6	28.6	28.6
smurf	100	100	100
pod	99.2	100	99.6
perl	50	33.3	40
warezclient	98.8	99.4	99.1
nmap	98.3	97.8	98
imap	100	100	100
warezmaster	73.7	70	71.8
portsweep	99.7	89.8	99.3
normal	99.9	100	99.9
guess_passwd	100	94.3	97.1
Average	**74.58**	**74.19**	**74.30**
Standard Deviation	**35.71**	**37.29**	**36.52**

จากตารางที่ 4.3 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค RIPPER พบว่าค่า Precision มี Average อยู่ที่ 74.58% และ Standard Deviation อยู่ที่ 35.71% ค่า Recall มี Average อยู่ที่ 74.19% และ Standard Deviation อยู่ที่ 37.29% และค่า F-measure มี Average อยู่ที่ 74.30% และ Standard Deviation อยู่ที่ 36.52

4.1.4 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค PART Decision list

จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค PART Decision list มีกฎจำนวน 98 กฎ สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการจำแนกโดยใช้เทคนิค PART Decision list

Class	Precision (%)	Recall (%)	F-Measure (%)
back	100	99.9	99.9
teardrop	100	100	100

ตารางที่ 4.4 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
loadmodule	28.6	20	25
neptune	100	100	100
rootkit	28.6	20	23.5
phf	100	100	100
satan	98.9	99.4	99.1
buffer_overflow	80	93.3	86.2
ftp_write	60	37.5	46.2
land	90.5	90.5	90.5
spy	0	0	0
ipsweep	99.4	99.9	99.7
multihop	14.3	14.3	14.3
smurf	100	100	100
pod	99.6	98.9	99.2
perl	100	100	100
warezclient	99	98.5	98.8
nmap	99.1	98.3	98.7
imap	88.9	66.7	76.2
warezmaster	71.4	75	73.2
portsweep	99.5	98.7	99.1
normal	99.9	99.9	99.9
guess_passwd	96.2	95.2	95.2
Average	**80.60**	**78.52**	**79.33**
Standard Deviation	**30.88**	**33.19**	**31.96**

จากตารางที่ 4.4 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค PART Decision list พบว่าค่า Precision มี Average อยู่ที่ 80.60% และ Standard Deviation อยู่ที่ 30.88% ค่า Recall มี Average อยู่ที่ 78.52% และ Standard Deviation อยู่ที่ 33.19% และค่า F-measure มี Average อยู่ที่ 79.33% และ Standard Deviation อยู่ที่ 31.96

4.2 การวิเคราะห์ประสิทธิภาพของแบบจำลอง

ในงานวิจัยนี้ผู้วิจัยได้วัดประสิทธิภาพโดยใช้ค่า Precision, Recall, F-Measure และ Accuracy

ตารางที่ 4.5 ผลการวิเคราะห์ประสิทธิภาพของแบบจำลอง

แบบจำลอง	Precision	Recall	F-Measure	Accuracy
Decision Table	65.16%	57.39%	60.12%	99.7569%
Naïve Bayes	5.40%	68.36%	52.95%	92.7794%
RIPPER	74.58%	74.19%	74.30%	99.9698%
PART	80.60%	78.52%	79.33%	99.9694%

จากตารางที่ 4.5 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลองการตรวจจับการบุกรุกทั้ง 4 แบบจำลองผลปรากฏว่าแบบจำลอง RIPPER นั้นมีประสิทธิภาพสูงที่สุดคือ มีค่า Accuracy อยู่ที่ 99.9698% มีค่า F-Measure อยู่ที่ 74.30% มีค่า Recall อยู่ที่ 74.19% และมีค่า Precision อยู่ที่ 74.58%

4.3 การแปลผลของแบบจำลอง

จากการทดลองสร้างแบบจำลองแบบจำลองที่สร้างจาก PART Decision list สามารถจำแนกการตรวจจับการบุกรุกได้มีประสิทธิภาพมากที่สุด ดังตัวอย่างของ 10 กฎแรก ดังนี้

กฎที่ 1

srv_count > 322 AND protocol_type = icmp: smurf (280461.0/1.0)

กฎที่ 2

same_srv_rate <= 0.32 AND dst_host_diff_srv_rate <= 0.14 AND
src_bytes <= 0 AND dst_host_same_src_port_rate <= 0.02 AND
diff_srv_rate <= 0.58: neptune (106158.0)

กฎที่ 3

wrong_fragment <= 0 AND lnum_compromised > 0 AND
src_bytes > 10073: back (2129.0)

กฎที่ 4

wrong_fragment <= 0 AND srv_serror_rate <= 0.51 AND
dst_host_srv_diff_host_rate <= 0.48 AND same_srv_rate > 0.11 AND
dst_host_srv_rerror_rate <= 0.99 AND hot <= 0 AND
dst_host_same_src_port_rate <= 0.99 AND src_bytes > 6 AND
dst_host_srv_serror_rate <= 0.02 AND lnum_file_creations <= 0 AND
count <= 46 AND lnum_compromised <= 0 AND flag = SF AND
dst_bytes > 0 AND src_bytes <= 30683 AND dst_host_srv_count > 4:
normal (74653.0/1.0)

กฎที่ 5

wrong_fragment > 0 AND protocol_type = udp: teardrop (979.0)

กฎที่ 6

dst_host_srv_serror_rate > 0.82 AND flag = SH AND
srv_count <= 80: nmap (103.0)

กฎที่ 7

srv_serror_rate > 0.51 AND dst_host_diff_srv_rate > 0.7 AND
same_srv_rate <= 0.25: satan (171.0)

กฎที่ 8

srv_serror_rate > 0.51 AND src_bytes <= 0 AND land = 0 AND
dst_host_serror_rate > 0.68 AND flag = S0 AND
dst_host_same_src_port_rate <= 0.17: neptune (1019.0)

กฎที่ 9

count > 327 AND diff_srv_rate > 0.73: satan (1144.0)

กฎที่ 10

dst_host_srv_diff_host_rate <= 0.48 AND
flag = S1 AND
hot <= 0: normal (53.0/1.0)

เป็นการยกตัวอย่างการสร้างกฎของแบบจำลอง PART Decision list จำนวน 10 กฎ ซึ่งมีความหมายดังนี้

กฎที่ 1 หมายถึง ถ้า Service count มากกว่า 322 และ Protocol type เป็นประเภทโปรโตคอล ICMP จะเป็นการบุกรุกแบบ Smurf Flooding Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 280,461 กรณี และมีกรณีที่ไม่ตรงกับเงื่อนไขเพียง 1 กรณีเท่านั้น

กฎที่ 2 หมายถึง ถ้า Same service rate น้อยกว่าหรือเท่ากับ 0.32 และ Destination host diff service rate น้อยกว่าหรือเท่ากับ 0.14 และ Source bytes น้อยกว่าหรือเท่ากับ 0 และ Destination host same source port rate น้อยกว่าหรือเท่ากับ 0.02 และ Diff Service rate น้อยกว่าหรือเท่ากับ 0.58 จะเป็นการบุกรุกแบบ Neptune Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 106,158 และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 3 หมายถึง ถ้า Wrong fragment น้อยกว่าหรือเท่ากับ 0 และ Logged number compromised มากกว่า 0 และ Source bytes มากกว่า 10,073 จะเป็นการบุกรุกแบบ Backdoor Attacks ซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 2,129 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 4 หมายถึง ถ้า Wrong fragment น้อยกว่าหรือเท่ากับ 0 และ Service SYN error rate น้อยกว่าหรือเท่ากับ 0.51 และ Destination host service diff host rate น้อยกว่าหรือเท่ากับ 0.48 และ Same service rate มากกว่า 0.11 และ Destination host service REJ error rate น้อยกว่าหรือเท่ากับ 0.99 และ Hot น้อยกว่าหรือเท่ากับ 0 และ Destination host same source port rate น้อยกว่าหรือเท่ากับ 0.99 และ Source bytes มากกว่า 6 และ Destination host service SYN error rate น้อยกว่าหรือเท่ากับ 0.02 และ Logged number file creations น้อยกว่าหรือเท่ากับ 0 และ Count น้อยกว่าหรือเท่ากับ 46 และ Logged number compromised น้อยกว่าหรือเท่ากับ 0 และ Destination bytes มากกว่า 0 และ Source bytes น้อยกว่าหรือเท่ากับ 30,683 และ Destination host service count มากกว่า 4 จะเป็นการเข้าใช้งานแบบปกติ (Normal) ไม่มีการบุกรุกซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 74,653 กรณี และมีกรณีที่ไม่ตรงกับเงื่อนไขเพียง 1 กรณีเท่านั้น

กฎที่ 5 หมายถึง ถ้า Wrong fragment มากกว่า 0 และ Protocol type เป็นประเภทโปรโตคอล UDP จะเป็นการบุกรุกแบบ Teardrop Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไข 979 กรณีและไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 6 หมายถึง ถ้า Destination host service SYN error rate มากกว่า 0.82 และ Flag เท่ากับ SH และ Service count น้อยกว่าหรือเท่ากับ 80 จะเป็นการบุกรุกโดยใช้ ซอฟต์แวร์ Nmap ซึ่งมีกรณีที่ตรงกับเงื่อนไข 103 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 7 หมายถึง ถ้า Service SYN error rate มากกว่า 0.51 และ Destination host diff service rate มากกว่า 0.7 และ Same service rate น้อยกว่าหรือเท่ากับ 0.25 จะเป็นการบุกรุกแบบ Satan ซึ่งมีกรณีที่ตรงกับเงื่อนไข 171 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 8 หมายถึง ถ้า Service SYN error rate มากกว่า 0.51 และ Source byte มากกว่าหรือเท่ากับ 0 และ Land เท่ากับ 0 และ Destination host SYN error rate มากว่า 0.68

และ Flag เท่ากับ 0 และ Destination host same source port rate น้อยกว่าหรือเท่ากับ 0.17 จะเป็นการบุกรุกแบบ Neptune Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไข 1,019 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 9 หมายถึง ถ้า Count มากกว่า 327 และ Diff service rate มากกว่า 0.73 จะเป็นการบุกรุกแบบ Satan ซึ่งมีกรณีที่ตรงกับเงื่อนไข 1,144 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 10 หมายถึง ถ้า Destination host service diff host rate น้อยกว่าหรือเท่ากับ 0.48 และ Flag เท่ากับ 1 และ Hot น้อยกว่าหรือเท่ากับ 0 จะเป็นการเข้าใช้งานแบบปกติ (Normal) ไม่มีการบุกรุกซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 74,653 กรณี และมีกรณีที่ไม่ตรงกับเงื่อนไขเพียง 1 กรณีเท่านั้น

บทที่ 4

ผลการวิจัย

ผลการวิจัยการจำแนกข้อมูลสำหรับตรวจจับการบุกรุกโดยใช้เทคนิคการจำแนก Decision Table, Naïve Bayes, RIPPER และ PART Decision list ทดลองสร้างโมเดลจากฐานข้อมูล KDD Cup'99 ซึ่งใช้ข้อมูลจำนวน 494,020 เรคคอร์ด แอททริบิวจำนวน 41 แอททริบิวและคลาสผลลัพธ์ 23 คลาส ทดลองด้วยวิธีแบบ 10-fold cross validation สามารถแสดงประสิทธิภาพของแบบจำลองและแบบจำลองดังต่อไปนี้

4.1 ประสิทธิภาพของแบบจำลอง

4.1.1 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Decision Table จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Decision Table มีกฎจำนวน 1,105 กฎ สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.1

ตารางที่ 4.1 ผลการจำแนกโดยใช้เทคนิค Decision Table

Class	Precision (%)	Recall (%)	F-Measure (%)
back	0	99.8	95.3
teardrop	99.7	99.6	99.6
loadmodule	0	0	0
neptune	99.9	100	99.9
rootkit	0	0	0
phf	100	50	66.7
satan	99.3	96.1	97.7
buffer_overflow	100	50	66.7
ftp_write	0	0	0
land	94.7	85.7	90
spy	0	0	0
ipsweep	91.7	92.2	92
multihop	0	0	0

ตารางที่ 4.1 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
back	0	99.8	95.3
teardrop	99.7	99.6	99.6
loadmodule	0	0	0
neptune	99.9	100	99.9
rootkit	0	0	0
phf	100	50	66.7
satan	99.3	96.1	97.7
buffer_overflow	100	50	66.7
ftp_write	0	0	0
land	94.7	85.7	90
spy	0	0	0
ipsweep	91.7	92.2	92
multihop	0	0	0
smurf	99.9	100	99.9
pod	98.5	99.2	98.9
perl	0	0	0
warezclient	83.7	93.5	88.3
nmap	91.1	48.9	63.7
imap	100	8.3	15.4
warezmaster	50	15	23.1
portsweep	90.3	91.6	91
normal	99.9	99.4	99.6
guess_passwd	100	90.6	95
Average	**65.16**	**57.39**	**60.12**
Standard Deviation	**44.29**	**42.71**	**41.94**

จากตารางที่ 4.1 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Decision Table พบว่าค่า Precision มี Average อยู่ที่ 65.16% และ Standard Deviation อยู่ที่ 44.29% ค่า

Recall มี Average อยู่ที่ 57.39% และ Standard Deviation อยู่ที่ 42.71% และค่า F-measure มี Average อยู่ที่ 60.12% และ Standard Deviation อยู่ที่ 41.94

4.1.2 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Naïve Bayes

จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Naïve Bayes สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.2

ตารางที่ 4.2 ผลการจำแนกโดยใช้เทคนิค Naïve Bayes

Class	Precision (%)	Recall (%)	F-Measure (%)
back	10	99.7	99.8
teardrop	9.24	99.9	96
loadmodule	0	0	0
neptune	10	99.9	10
rootkit	0.33	40	6
phf	10	50	66.7
satan	9.82	91.3	94.6
buffer_overflow	2.88	76.7	41.8
ftp_write	0.87	25	12.9
land	1.89	100	31.8
spy	0	0	0
ipsweep	7.29	92.9	81.7
multihop	0	0	0
smurf	10	100	100
pod	7.79	98.9	87.1
perl	0	0	0
warezclient	8.15	99.6	89.6
nmap	3.41	47.2	39.6
imap	2.86	83.3	42.6
warezmaster	2.21	75	34.1
portsweep	9.05	97.7	93.9

ตารางที่ 4.2 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
normal	9.99	98.9	99.4
guess_passwd	8.5	96.2	90.3
Average	**5.40**	**68.36**	**52.95**
Standard Deviation	**4.07**	**37.92**	**38.92**

จากตารางที่ 4.2 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค Naïve Bayes พบว่าค่า Precision มี Average อยู่ที่ 5.40% และ Standard Deviation อยู่ที่ 4.07% ค่า Recall มี Average อยู่ที่ 68.36% และ Standard Deviation อยู่ที่ 37.92% และค่า F-measure มี Average อยู่ที่ 52.95% และ Standard Deviation อยู่ที่ 38.92

4.1.3 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค RIPPER

จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค RIPPER มีกฎจำนวน 61 กฎ สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.3

ตารางที่ 4.3 ผลการจำแนกโดยใช้เทคนิค RIPPER

Class	Precision (%)	Recall (%)	F-Measure (%)
back	100	100	100
teardrop	100	100	100
loadmodule	20	11.1	14.3
neptune	100	100	100
rootkit	12.5	10	11.1
phf	66.7	100	80
satan	99.5	99.4	99.5
buffer_overflow	73.3	73.3	73.3
ftp_write	0	0	0
land	95.5	100	97.7
spy	0	0	0
ipsweep	99.7	99.3	99.5

ตารางที่ 4.3 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
ipsweep	99.7	99.3	99.5
multihop	28.6	28.6	28.6
smurf	100	100	100
pod	99.2	100	99.6
perl	50	33.3	40
warezclient	98.8	99.4	99.1
nmap	98.3	97.8	98
imap	100	100	100
warezmaster	73.7	70	71.8
portsweep	99.7	89.8	99.3
normal	99.9	100	99.9
guess_passwd	100	94.3	97.1
Average	**74.58**	**74.19**	**74.30**
Standard Deviation	**35.71**	**37.29**	**36.52**

จากตารางที่ 4.3 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค RIPPER พบว่าค่า Precision มี Average อยู่ที่ 74.58% และ Standard Deviation อยู่ที่ 35.71% ค่า Recall มี Average อยู่ที่ 74.19% และ Standard Deviation อยู่ที่ 37.29% และค่า F-measure มี Average อยู่ที่ 74.30% และ Standard Deviation อยู่ที่ 36.52

4.1.4 ผลการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค PART Decision list

จากการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค PART Decision list มีกฎจำนวน 98 กฎ สามารถแสดงค่า Precision Recall และ F-measure ได้ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการจำแนกโดยใช้เทคนิค PART Decision list

Class	Precision (%)	Recall (%)	F-Measure (%)
back	100	99.9	99.9
teardrop	100	100	100

ตารางที่ 4.4 (ต่อ)

Class	Precision (%)	Recall (%)	F-Measure (%)
loadmodule	28.6	20	25
neptune	100	100	100
rootkit	28.6	20	23.5
phf	100	100	100
satan	98.9	99.4	99.1
buffer_overflow	80	93.3	86.2
ftp_write	60	37.5	46.2
land	90.5	90.5	90.5
spy	0	0	0
ipsweep	99.4	99.9	99.7
multihop	14.3	14.3	14.3
smurf	100	100	100
pod	99.6	98.9	99.2
perl	100	100	100
warezclient	99	98.5	98.8
nmap	99.1	98.3	98.7
imap	88.9	66.7	76.2
warezmaster	71.4	75	73.2
portsweep	99.5	98.7	99.1
normal	99.9	99.9	99.9
guess_passwd	96.2	95.2	95.2
Average	**80.60**	**78.52**	**79.33**
Standard Deviation	**30.88**	**33.19**	**31.96**

จากตารางที่ 4.4 แสดงผลการทดลองการจำแนกการตรวจจับการบุกรุกด้วยเทคนิค PART Decision list พบว่าค่า Precision มี Average อยู่ที่ 80.60% และ Standard Deviation อยู่ที่ 30.88% ค่า Recall มี Average อยู่ที่ 78.52% และ Standard Deviation อยู่ที่ 33.19% และค่า F-measure มี Average อยู่ที่ 79.33% และ Standard Deviation อยู่ที่ 31.96

4.2 การวิเคราะห์ประสิทธิภาพของแบบจำลอง

ในงานวิจัยนี้ผู้วิจัยได้วัดประสิทธิภาพโดยใช้ค่า Precision, Recall, F-Measure และ Accuracy

ตารางที่ 4.5 ผลการวิเคราะห์ประสิทธิภาพของแบบจำลอง

แบบจำลอง	Precision	Recall	F-Measure	Accuracy
Decision Table	65.16%	57.39%	60.12%	99.7569%
Naïve Bayes	5.40%	68.36%	52.95%	92.7794%
RIPPER	74.58%	74.19%	74.30%	99.9698%
PART	80.60%	78.52%	79.33%	99.9694%

จากตารางที่ 4.5 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลองการตรวจจับการบุกรุกทั้ง 4 แบบจำลองผลปรากฏว่าแบบจำลอง RIPPER นั้นมีประสิทธิภาพสูงที่สุดคือ มีค่า Accuracy อยู่ที่ 99.9698% มีค่า F-Measure อยู่ที่ 74.30% มีค่า Recall อยู่ที่ 74.19% และมีค่า Precision อยู่ที่ 74.58%

4.3 การแปลผลของแบบจำลอง

จากการทดลองสร้างแบบจำลองแบบจำลองที่สร้างจาก PART Decision list สามารถจำแนกการตรวจจับการบุกรุกได้มีประสิทธิภาพมากที่สุด ดังตัวอย่างของ 10 กฎแรก ดังนี้

กฎที่ 1

srv_count > 322 AND protocol_type = icmp: smurf (280461.0/1.0)

กฎที่ 2

same_srv_rate <= 0.32 AND dst_host_diff_srv_rate <= 0.14 AND
src_bytes <= 0 AND dst_host_same_src_port_rate <= 0.02 AND
diff_srv_rate <= 0.58: neptune (106158.0)

กฎที่ 3

wrong_fragment <= 0 AND lnum_compromised > 0 AND
src_bytes > 10073: back (2129.0)

กฎที่ 4

wrong_fragment <= 0 AND srv_serror_rate <= 0.51 AND
dst_host_srv_diff_host_rate <= 0.48 AND same_srv_rate > 0.11 AND
dst_host_srv_rerror_rate <= 0.99 AND hot <= 0 AND
dst_host_same_src_port_rate <= 0.99 AND src_bytes > 6 AND
dst_host_srv_serror_rate <= 0.02 AND lnum_file_creations <= 0 AND
count <= 46 AND lnum_compromised <= 0 AND flag = SF AND
dst_bytes > 0 AND src_bytes <= 30683 AND dst_host_srv_count > 4:
normal (74653.0/1.0)

กฎที่ 5

wrong_fragment > 0 AND protocol_type = udp: teardrop (979.0)

กฎที่ 6

dst_host_srv_serror_rate > 0.82 AND flag = SH AND
srv_count <= 80: nmap (103.0)

กฎที่ 7

srv_serror_rate > 0.51 AND dst_host_diff_srv_rate > 0.7 AND
same_srv_rate <= 0.25: satan (171.0)

กฎที่ 8

srv_serror_rate > 0.51 AND src_bytes <= 0 AND land = 0 AND
dst_host_serror_rate > 0.68 AND flag = S0 AND
dst_host_same_src_port_rate <= 0.17: neptune (1019.0)

กฎที่ 9

count > 327 AND diff_srv_rate > 0.73: satan (1144.0)

กฎที่ 10

dst_host_srv_diff_host_rate <= 0.48 AND
flag = S1 AND
hot <= 0: normal (53.0/1.0)

เป็นการยกตัวอย่างการสร้างกฎของแบบจำลอง PART Decision list จำนวน 10 กฎ ซึ่งมีความหมายดังนี้

กฎที่ 1 หมายถึง ถ้า Service count มากกว่า 322 และ Protocol type เป็นประเภทโปรโตคอล ICMP จะเป็นการบุกรุกแบบ Smurf Flooding Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 280,461 กรณี และมีกรณีที่ไม่ตรงกับเงื่อนไขเพียง 1 กรณีเท่านั้น

กฎที่ 2 หมายถึง ถ้า Same service rate น้อยกว่าหรือเท่ากับ 0.32 และ Destination host diff service rate น้อยกว่าหรือเท่ากับ 0.14 และ Source bytes น้อยกว่าหรือเท่ากับ 0 และ Destination host same source port rate น้อยกว่าหรือเท่ากับ 0.02 และ Diff Service rate น้อยกว่าหรือเท่ากับ 0.58 จะเป็นการบุกรุกแบบ Neptune Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 106,158 และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 3 หมายถึง ถ้า Wrong fragment น้อยกว่าหรือเท่ากับ 0 และ Logged number compromised มากกว่า 0 และ Source bytes มากกว่า 10,073 จะเป็นการบุกรุกแบบ Backdoor Attacks ซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 2,129 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 4 หมายถึง ถ้า Wrong fragment น้อยกว่าหรือเท่ากับ 0 และ Service SYN error rate น้อยกว่าหรือเท่ากับ 0.51 และ Destination host service diff host rate น้อยกว่าหรือเท่ากับ 0.48 และ Same service rate มากกว่า 0.11 และ Destination host service REJ error rate น้อยกว่าหรือเท่ากับ 0.99 และ Hot น้อยกว่าหรือเท่ากับ 0 และ Destination host same source port rate น้อยกว่าหรือเท่ากับ 0.99 และ Source bytes มากกว่า 6 และ Destination host service SYN error rate น้อยกว่าหรือเท่ากับ 0.02 และ Logged number file creations น้อยกว่าหรือเท่ากับ 0 และ Count น้อยกว่าหรือเท่ากับ 46 และ Logged number compromised น้อยกว่าหรือเท่ากับ 0 และ Destination bytes มากกว่า 0 และ Source bytes น้อยกว่าหรือเท่ากับ 30,683 และ Destination host service count มากกว่า 4 จะเป็นการเข้าใช้งานแบบปกติ (Normal) ไม่มีการบุกรุกซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 74,653 กรณี และมีกรณีที่ไม่ตรงกับเงื่อนไขเพียง 1 กรณีเท่านั้น

กฎที่ 5 หมายถึง ถ้า Wrong fragment มากกว่า 0 และ Protocol type เป็นประเภทโปรโตคอล UDP จะเป็นการบุกรุกแบบ Teardrop Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไข 979 กรณีและไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 6 หมายถึง ถ้า Destination host service SYN error rate มากกว่า 0.82 และ Flag เท่ากับ SH และ Service count น้อยกว่าหรือเท่ากับ 80 จะเป็นการบุกรุกโดยใช้ ซอฟต์แวร์ Nmap ซึ่งมีกรณีที่ตรงกับเงื่อนไข 103 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 7 หมายถึง ถ้า Service SYN error rate มากกว่า 0.51 และ Destination host diff service rate มากกว่า 0.7 และ Same service rate น้อยกว่าหรือเท่ากับ 0.25 จะเป็นการบุกรุกแบบ Satan ซึ่งมีกรณีที่ตรงกับเงื่อนไข 171 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 8 หมายถึง ถ้า Service SYN error rate มากกว่า 0.51 และ Source byte มากกว่าหรือเท่ากับ 0 และ Land เท่ากับ 0 และ Destination host SYN error rate มากว่า 0.68

และ Flag เท่ากับ 0 และ Destination host same source port rate น้อยกว่าหรือเท่ากับ 0.17 จะเป็นการบุกรุกแบบ Neptune Attack ซึ่งมีกรณีที่ตรงกับเงื่อนไข 1,019 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 9 หมายถึง ถ้า Count มากกว่า 327 และ Diff service rate มากกว่า 0.73 จะเป็นการบุกรุกแบบ Satan ซึ่งมีกรณีที่ตรงกับเงื่อนไข 1,144 กรณี และไม่มีกรณีที่ไม่ตรงกับเงื่อนไขเลย

กฎที่ 10 หมายถึง ถ้า Destination host service diff host rate น้อยกว่าหรือเท่ากับ 0.48 และ Flag เท่ากับ 1 และ Hot น้อยกว่าหรือเท่ากับ 0 จะเป็นการเข้าใช้งานแบบปกติ (Normal) ไม่มีการบุกรุกซึ่งมีกรณีที่ตรงกับเงื่อนไขถึง 74,653 กรณี และมีกรณีที่ไม่ตรงกับเงื่อนไขเพียง 1 กรณีเท่านั้น

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การตรวจจับการบุกรุกด้วยเทคนิคการจำแนกในการทำเหมืองข้อมูล โดยการเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการบุกรุกบนระบบเครือโดยใช้เทคนิค 4 เทคนิคด้วยกันคือ Decision Table, Naïve Bayes, RIPPER และ PART Decision list พบว่าแบบจำลองที่ใช้เทคนิค RIPPER นั้นมีความถูกต้องมากที่สุดตามด้วยเทคนิค PART Decision list และ Decision Table กับ Naïve Bayes ตามลำดับแต่อย่างไรก็ตามแบบจำลองของเทคนิค RIPPER และแบบจำลองของเทคนิค PART Decision list นั้นก็มีความแตกต่างกันเพียงเล็กน้อย จึงทำให้แบบจำลองที่ใช้เทคนิค RIPPER และ PART Decision list นั้นเหมาะที่จะนำไปใช้ในการตรวจจับการบุกรุกบนระบบเครือข่าย เพราะสามารถเพิ่มความปลอดภัยและประสิทธิภาพให้กับระบบเครือข่ายได้ ส่วนแบบจำลองที่ใช้เทคนิค Naïve Bayes นั้นมีความเหมาะสมน้อยที่สุดที่จะนำไปใช้ในการตรวจจับการบุกรุกบนระบบเครือข่าย

5.2 อภิปรายผลการวิจัย

จากการทดลองโดยเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการบุกรุกบนระบบเครือข่าย ซึ่งใช้ข้อมูลจากฐานข้อมูลความรู้ KDD Cup'99Dataset ที่มีข้อมูลจำนวน 494,020 เรคคอร์ด มีแอททรบิวจำนวน 41 แอททริบิว และมีคลาสผลลัพธ์จำนวน 23 คลาสเป็นชุดข้อมูลที่ใช้ในทดลองและใช้วิธีการ 10-fold validation ในการสร้างแบบจำลองโดยใช้อัลกอริทึม 4 อัลกอริทึม คือ Decision Table, Naïve Bayes, RIPPER และ PART decision list จากผลการทดลองนั้นแบบจำลองที่สร้างจากอัลกอริทึม Decision Table นั้นมีค่าความถูกต้องเท่ากับ 99.7569% มีค่าความแม่นยำ (Precision) เท่ากับ 0.998 มีค่าระลึก (Recall) เท่ากับ 0.997 และค่า F-Measure เท่ากับ 0.997 แบบจำลองที่สร้างจากอัลกอริทึม Naïve Bayes นั้นมีค่าความถูกต้องเท่ากับ 92.7794% มีค่าความแม่นยำ (Precision) เท่ากับ 0.998 มีค่าระลึก (Recall) เท่ากับ 0.997 และค่า F-Measure เท่ากับ 0.997 แบบจำลองที่สร้างจากอัลกอริทึม RIPPER นั้นมีค่าความถูกต้องเท่ากับ 99.9698% มีค่าความแม่นยำ (Precision) เท่ากับ 1 มีค่าระลึก (Recall) เท่ากับ 1 และค่า F-Measure เท่ากับ 1และแบบจำลองที่สร้างจากอัลกอริทึม PART Decision list นั้นมีค่าความถูกต้องเท่ากับ 99.9694% มีค่าความแม่นยำ (Precision) เท่ากับ 1 มีค่าระลึก (Recall) เท่ากับ 1 และค่า F-Measure เท่ากับ 1 ผลจากค่าทางสถิติ

เอกสารอ้างอิง

เอกสารอ้างอิง

[1] สำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร. [ออนไลน์]. 2559. [สืบค้นเมื่อ วันที่ 30 พฤศจิกายน 2559] ; ได้จาก: http://service.nso.go.th/nso/web/statseries/statseries22.html.

[2] P. Tang, R. Jiang, and M. Zhao. "Feature Selection and Design of Intrusion System Based on K-Means and Triangle Area Support Vector Machine". *International Conference on Future Network* 2010; 16-18 November 2010; Brisbane, Australia: pp. 144-148.

[3] เอกสิทธิ์ พัชรวงศ์ศักดา. *An Introduction to Data Minig Techniques* vol. 2, 2557.

[4] ธนกร มีหินกอง, ประสงค์ ประณีตพลกรัง. "ระบบตรวจหาการบุกรุก". *Information Science Institute of Sripatum University,* 2555.

[5] เอกสิทธิ์ พัชรวงศ์ศักดา. "คู่มือการใช้งาน *Weka Explorer* เบื้องต้น". กรุงเทพฯ: หสม.ดาต้า คิวบ์; 2556.

[6] บุญเสริม กิจศิริกุล. "อัลกอริทึมการทำเหมืองข้อมูล" [ออนไลน์]. 2546. [สืบค้นเมื่อ วันที่ 30 พฤศจิกายน 2559]; ได้จาก: https://www.cp.eng.chula.ac.th/~boonserm/publication/AlgoDataMining.pdf.

[7] ธนาวุฒิ เอื้อชัยกุล. "การสร้างชุดคำสั่งบีเพลสำหรับกฎธุรกิจจากตารางตัดสินใจ". กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย; 2551.

[8] Anil Rajput. "J48 and JRIP Rules for E-Governance Data". *International Journal of Computer Science and Security,* vol.5 2011; 5(2): 201-207.

[9] Eibe Frank, Ian H., Witten. "Generating Accurate Rule Sets Without Global Optimization". [Online]. 1998. [cited 23 August 2016]; Available from: https://pdfs.semanticscholar.org/3998/12d46345ad7d93f5510b1bbda30948e7a65c.pdf

[10] อรนุช พันโท, มนต์ชัย เทียนทอง. "การเปรียบเทียบประสิทธิภาพการจำแนกรูปแบบการเรียนรู้ VARK ด้วยเทคนิคเหมืองข้อมูล". วารสารเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี (Journal of Industrial Technology, UBRU) 2557; 4(1) : 1-11

[11] ลลินี อินทราราม. "การวิเคราะห์รูปแบบการบุกรุกข้อมูลบนเครือข่าย โดยใช้การจำแนกข้อมูลของแบบจำลองแรนดอมฟอเรส ซัพพอร์ทเวกเตอร์แมชชีน แบบผสมผสานกัน". [วิทยานิพนธ์ปริญญาวิทยาศาสตร์มหาบัณฑิต]. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ; 2554.

[12] เยาวภา ภารสำเร็จ. "การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา". วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม 2556; ฉบับพิเศษ: 281-289.

[13] ปรีชา สมหวัง, ศิริวัฒน์ โทศิริกุล. "ระบบตรวจจับการใช้งานคอมพิวเตอร์ในทางที่ผิด". ใน: National Conference on Information Technology 2010. 28-29 October 2010; Thailand. pp. 409-414.

[14] ปวีณา ชัยวนารมณ์. "การพัฒนาแบบจำลองเพื่อพยากรณ์การเกิดความเครียดในหลายระดับด้วยเทคนิคการทำเหมืองข้อมูล". ใน: การประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ครั้งที่ 1; 2559. 22 มิถุนายน 2559; มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จังหวัดพระนครศรีอยุธยาม, ประเทศไทย: หน้า 406-416.

[15] กรมวุฒิ นงนุช, ธนพล เจนสุทธิเวชกุล. "ระบบตรวจสอบการบุกรุกเครือข่ายโดยใช้โครงข่ายประสาทเทียมกับกระบวนการทางพันธุศาสตร์". ใน: National Conference on Computer and Information Technology, 2010. 28-29 October 2010; Thailand: pp. 415-421.

[16] ณัฐวุฒิ ปั้นรูป, อัฐพร กิ่งบู. "Data Classification by K-Means and Multi-Class SVM for Intrusion Detection System". ใน: National Conference on Information Technology, 2015. 29-30 October 2015; Chiang Mai, Thailand: pp. 40-45.

[17] ธนกร มีหินกอง. "สถาปัตยกรรมความรู้ด้านความมั่งคงปลอดภัยไซเบอร์เพื่อสนับสนุนระบบตรวจหาการบุกรุกแบบปรับตัวด้วยเทคนิคกฎความสัมพันธ์". The Journal of KMUTNB 2558; 25(20): 277-288.

[18] นิเวศ จิระวิชิตชัย. "การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ" มหาวิทยาลัยราชภัฏสวนสุนันทา, [ออนไลน์]. 2553. [สืบค้นเมื่อ วันที่ 30 พฤศจิกายน 2559]; ได้จาก:http://www.ssruir.ssru.ac.th/bitstream/ssruir/354/1/080-53.pdf.

[19] ไพชยนต์ คงไชย. "การพัฒนาขั้นตอนวิธีเพื่อจำแนกประเภทข้อมูลด้วยกฎความสัมพันธ์แบบคลุมเครือที่กะทัดรัด". [วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต]. นครราชสีมา: มหาวิทยาลัยเทคโนโลยีสุรนารี; 2557.

[20] "KDD Cup 1999 Data". [Online]. 1999. [cited 23 August 2016]; Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[21] Roberto Di Pietro, Luigi V. Mancini. "Intrusion Detection Systems". United States: Springer International Publishing; 2008.

[22] Guan Lihe, Hu Feng and Han Fengqing. "A rule induction algorithm in incomplete decision table based on attribute order". Journal of Intelligent & Fuzzy Systems 2016; 30(2): pp. 845-859.

[23] Jens Hühn, Eyke Hüllermeier. "FURIA: An Algorithm For Unordered Fuzzy Rule Induction". Journal of Data Mining and Knowledge Discovery 2009; 19(3): pp. 293-319.

[24] ปองหทัย กาญจนภาชน์, มณเฑียร รัตนศิริวงศ์วุฒิ. "ระบบจัดการเหตุเสียเครือข่ายไอพีด้วยการจำแนกกลุ่มข้อคามเหตุเสียแบบอัตโนมัติ". ใน: The Eleventh National Conference on Computing and Information Technology 2015. 18 August 2015; Chiang Rai, Thailand: pp. 40-45.

[25] พยุง มีสัจ. "ระบบฟัซซีและโครงข่ายประสาทเทียม". วารสารวิจัยพลังงาน 2557; 11(2): หน้า 67-78.

[26] Paul Mather, Brandt Tso. "Classification Methods for Remotely Sensed Data". Taylor & Francis group; 2009.

[27] เกรียงไกร พิพิธหิรัญการ. "การเปรียบเทียบประสิทธิภาพในการทำนายพฤติกรรมผู้บริโภคโดยใช้เทคนิคการทำเหมืองข้อมูลระหว่างกฎความสัมพันธ์สำหรับจำแนกและต้นไม้ตัดสินใจ". [วิทยานิพนธ์ปริญญาวิทยาศาสตร์มหาบัณฑิต]. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ; 2550.

[28] มัณฑนา ธรรมรักษา. "โปรโตคอลค้นหาเส้นทางเชิงกรีดีสำหรับการสื่อสารระหว่างยานพาหนะ". [วิทยานิพนธ์ปริญญาวิทยาศาสตร์มหาบัณฑิต]. กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์; 2553.

[29] อรรถพล ป้อมสถิตย์. "Enhanced Efficiency of Intrusion Detection Systems with Honey Pot in Cyber Security". KKU Science Journal 2559; 44(2) : 384-397.

[30] วิชุดา เขตอุดมคีรี. "การสร้างกรณีทดสอบเอ็ชทีเอ็มแอลเอ็กซ์เอ็มแอลสคีมาสำหรับโปรแกรมประยุกต์บนเว็บโดยใช้ตารางตัดสินใจ". [วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต]. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย; 2549.

[31] มารุด คำภักดี. "Intrusion Detection System Base on Snort". Faculty of Science, Khon Kaen University; 2012.

[32] Kittikhun Thongkanchorn. "Evaluation of Three Intrusion Detection Systems Under Various Attacks". Faculty of Graduate Studies, Mahidol University; 2012.

ต่างๆจะเห็นได้ว่าแบบจำลองที่สร้างจากอัลกอริทึมทั้ง 4 อัลกอริทึม มีประสิทธิภาพค่อนข้างสูงแต่แบบจำลองที่มีประสิทธิภาพมากที่สุดคือแบบจำลองที่จากอัลกอริทึม RIPPER ตามด้วย PART Decision list, Decision Table และ Naïve Bayes ตามลำดับ

5.3 ข้อเสนอแนะการวิจัย

จากการทดลองการตรวจจับการบุกรุกด้วยเทคนิคการจำแนกทั้ง 4 เทคนิค ถึงแม้แบบจำลองนั้นมีความแม่นยำค่อนข้างสูงแต่ในปัจจุบันนั้นมีรูปแบบการบุกรุกระบบเครือข่ายหลากและซับซ้อนขึ้น การนำระบบตรวจจับการบุกรุกไปใช้ในระบบเครือข่ายอาจช่วยทำให้ระบบมีความปลอดภัยมากขึ้นแต่ก็ยังไม่สามารถตรวจจับรูปแบบการบุกรุกทั้งหมดได้ อีกทั้งฐานข้อมูลความรู้ KDD Cup'99 ที่ได้ใช้ในงานวิจัยครั้งนี้เป็นข้อมูลที่ถูกสร้างขึ้นนานหลายปีแล้ว ปัจจุบันนั้นอาจมีรูปแบบการโจมตีและการบุกรุกระบบเครือข่ายรูปแบบใหม่ๆเกิดขึ้น อาจทำให้แบบจำลองของระบบตรวจจับการบุกรุกในงานวิจัยนี้ไม่สามารถตรวจจับการบุกรุกดังกล่าวได้ แต่อย่างไรก็ตามถึงแม้ระบบเครือข่ายจะมีซอฟต์แวร์หรือฮาร์ดแวร์ด้านความปลอดภัยเข้ามาช่วยเพิ่มความปลอดภัยมากเพียงใด ก็ยังมีช่องโหว่ให้ผู้ที่ไม่หวังดีเข้าโจมตีหรือบุกรุกเครือข่ายได้เสมอ จึงเป็นหน้าที่ของผู้ดูแลระบบที่ต้องเรียนรู้เทคโนโลยีและคอยสังเกตความผิดปกติของระบบเครือข่าย และผู้ใช้เองก็ต้องระมัดระวังข้อมูลส่วนตัวเมื่อต้องใช้ระบบเครือข่ายที่รู้สึกว่าไม่มีความปลอดภัย

ภาคผนวก

ภาคผนวก ก
ผลการรันแบบจำลองฉบับเต็มของแบบจำลอง Naïve bayes

ผลการรับแบบจำลองฉบับเต็มของแบบจำลอง Naïve bayes

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: kdd_cup_1999

Instances: 494020

Attributes: 42

duration

protocol_type

service

flag

src_bytes

dst_bytes

land

wrong_fragment

urgent

hot

num_failed_logins

logged_in

lnum_compromised

lroot_shell

lsu_attempted

lnum_root

lnum_file_creations

lnum_shells

lnum_access_files

lnum_outbound_cmds

is_host_login

is_guest_login

count

srv_count
serror_rate
srv_serror_rate
rerror_rate
srv_rerror_rate
same_srv_rate
diff_srv_rate
srv_diff_host_rate
dst_host_count
dst_host_srv_count
dst_host_same_srv_rate
dst_host_diff_srv_rate
dst_host_same_src_port_rate
dst_host_srv_diff_host_rate
dst_host_serror_rate
dst_host_srv_serror_rate
dst_host_rerror_rate
dst_host_srv_rerror_rate
label

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Class

Attribute back teardrop loadmodule neptune
rootkit phf satan buffer_overflow ftp_write land
spy ipsweep multihop smurf pod perl

warezclient nmap imap warezmaster portsweep normal guess_passwd

(0) (0) (0) (0.22) (0) (0) (0) (0) (0) (0) (0) (0) (0) (0) (0) (0.57) (0) (0) (0) (0) (0) (0) (0) (0) (0.2) (0)

==========

==========

==========

==========

==========

==========

==========

duration

mean 0.0849 0 33.7823 0 102.906 5.8469 0 91.2121 32.1581 0 315.7344 0 183.7607 0 0 38.9796 615.0515 0 5.8469 8.1857 1915.1853 216.5055 3.0889

std. dev. 3.898 3.898 38.3684 3.898 202.5977 10.1272 3.898 96.8734 46.684 3.898 11.6939 3.898 237.336 3.898 3.898 11.0251 2206.2586 3.898 13.9185 35.6807 7281.6117 1359.1476 13.6582

weight sum 2203 979 9 107201 10 4 1589 30 8 21 2 1247 7 280790 264 3 1020 231 12 20 1040 97277 53

precision 23.3877 23.3877 23.3877 23.3877 23.3877 23.3877 23.3877 23.3877 23.3877 23.3877

23.3877 23.3877 23.3877 23.3877 23.3877 23.3877
23.3877 23.3877 23.3877 23.3877 23.3877 23.3877
23.3877

protocol_type
tcp 2204.0 1.0 10.0 107202.0
8.0 5.0 1417.0 31.0 9.0 22.0 3.0
95.0 8.0 1.0 1.0 4.0 1021.0 104.0
13.0 21.0 1040.0 76813.0 54.0
udp 1.0 980.0 1.0 1.0 4.0
1.0 171.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 26.0 1.0
1.0 1.0 19178.0 1.0
icmp 1.0 1.0 1.0 1.0 1.0
1.0 4.0 1.0 1.0 1.0 1.0 1154.0
1.0 280791.0 265.0 1.0 1.0 104.0 1.0
1.0 2.0 1289.0 1.0
[total] 2206.0 982.0 12.0 107204.0
13.0 7.0 1592.0 33.0 11.0 24.0 5.0
1250.0 10.0 280793.0 267.0 6.0 1023.0
234.0 15.0 23.0 1043.0 97280.0 56.0

service
vmnet 1.0 1.0 1.0 102.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 5.0 1.0 1.0
smtp 1.0 1.0 1.0 121.0
1.0 1.0 3.0 1.0 1.0 1.0 1.0

2.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 3.0 9599.0 1.0
ntp_u 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 381.0 1.0
shell 1.0 1.0 1.0 112.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0
kshell 1.0 1.0 1.0 99.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
aol 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
imap4 1.0 1.0 1.0 106.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
13.0 1.0 1.0 1.0 1.0
urh_i 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 15.0 1.0
netbios_ssn 1.0 1.0 1.0 107.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0

tftp_u 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0
mtp 1.0 1.0 1.0 106.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 3.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
uucp 1.0 1.0 1.0 105.0 1.0
1.0 2.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0
nnsp 1.0 1.0 1.0 106.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
echo 1.0 1.0 1.0 112.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0
tim_i 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 6.0 1.0 1.0 1.0 1.0
1.0 1.0 3.0 1.0
ssh 1.0 1.0 1.0 103.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 2.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 2.0 2.0 1.0
iso_tsap 1.0 1.0 1.0 116.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0

1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0
time 1.0 1.0 1.0 104.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 3.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 53.0 1.0
netbios_ns 1.0 1.0 1.0 102.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
systat 1.0 1.0 1.0 114.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 3.0 1.0 1.0
hostnames 1.0 1.0 1.0 103.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
login 1.0 1.0 1.0 103.0 1.0
1.0 1.0 1.0 3.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
efs 1.0 1.0 1.0 103.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0
supdup 1.0 1.0 1.0 102.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 5.0 1.0 1.0

http_8001			1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0		
courier			1.0	1.0	1.0	108.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	2.0	1.0	1.0			
ctf			1.0	1.0	1.0	97.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	2.0	1.0
1.0	1.0	1.0	1.0			
finger			1.0	1.0	1.0	178.0	1.0
1.0	4.0	1.0	1.0	21.0	1.0	2.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	2.0	469.0	1.0			
nntp			1.0	1.0	1.0	107.0	1.0
1.0	2.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	2.0	1.0
1.0	1.0	1.0	1.0			
ftp_data			1.0	1.0	4.0	171.0	
2.0	1.0	4.0	9.0	5.0	1.0	1.0
4.0	4.0	1.0	1.0	1.0	709.0	1.0
1.0	19.0	3.0	3799.0	1.0		
red_i			1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	2.0	1.0			
ldap			1.0	1.0	1.0	102.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0

1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
http 2204.0 1.0 1.0 193.0
1.0 5.0 3.0 1.0 1.0 1.0 1.0
4.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 4.0 61886.0 1.0
ftp 1.0 1.0 2.0 105.0 2.0
1.0 2.0 2.0 3.0 1.0 1.0 2.0
3.0 1.0 1.0 1.0 308.0 1.0 1.0
3.0 4.0 374.0 1.0
pm_dump 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0
exec 1.0 1.0 1.0 100.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
klogin 1.0 1.0 1.0 107.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
auth 1.0 1.0 1.0 109.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 221.0 1.0
netbios_dgm 1.0 1.0 1.0 100.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0

other 1.0 1.0 1.0 92.0 4.0
1.0 1247.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 6.0 1.0 1.0
1.0 261.0 5633.0 1.0
link 1.0 1.0 1.0 100.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 3.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0
X11 1.0 1.0 1.0 1.0 1.0
1.0 3.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 10.0 1.0
discard 1.0 1.0 1.0 116.0 1.0
1.0 2.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
private 1.0 980.0 1.0 101318.0
1.0 1.0 315.0 1.0 1.0 1.0 1.0
69.0 1.0 1.0 1.0 1.0 1.0 125.0
1.0 1.0 726.0 7367.0 1.0
remote_job 1.0 1.0 1.0 119.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
2.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
IRC 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 43.0 1.0
daytime 1.0 1.0 1.0 103.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0

1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
pop_3 1.0 1.0 1.0 119.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 2.0
1.0 1.0 4.0 80.0 1.0
pop_2 1.0 1.0 1.0 102.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0
gopher 1.0 1.0 1.0 113.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
4.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
sunrpc 1.0 1.0 1.0 105.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 3.0 1.0 1.0
name 1.0 1.0 1.0 98.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
2.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0
rje 1.0 1.0 1.0 109.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 3.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0
domain 1.0 1.0 1.0 113.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
2.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 4.0 1.0

uucp_path 1.0 1.0 1.0 106.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
http_2784 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0
Z39_50 1.0 1.0 1.0 92.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
domain_u 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 5863.0 1.0
csnet_ns 1.0 1.0 1.0 124.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 4.0 1.0 1.0
whois 1.0 1.0 1.0 108.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 2.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 3.0 1.0 1.0
eco_i 1.0 1.0 1.0 1.0 1.0
1.0 3.0 1.0 1.0 1.0 1.0 1150.0
1.0 1.0 1.0 1.0 1.0 103.0 1.0
1.0 1.0 390.0 1.0
bgp 1.0 1.0 1.0 107.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0

1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
sql_net 1.0 1.0 1.0 110.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0
printer 1.0 1.0 1.0 108.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 3.0 1.0 1.0
telnet 1.0 1.0 6.0 198.0 6.0
1.0 2.0 22.0 1.0 2.0 3.0 2.0
3.0 1.0 1.0 4.0 1.0 2.0 1.0
1.0 3.0 220.0 54.0
ecr_i 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 5.0
1.0 280791.0 260.0 1.0 1.0 2.0 1.0
1.0 2.0 346.0 1.0
urp_i 1.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 538.0 1.0
netstat 1.0 1.0 1.0 93.0 1.0
1.0 2.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 3.0 1.0 1.0
http_443 1.0 1.0 1.0 99.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0 1.0

harvest 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
[total] 2273.0 1049.0 79.0 107271.0
80.0 74.0 1659.0 100.0 78.0 91.0 72.0
1317.0 77.0 280860.0 334.0 73.0 1090.0
301.0 82.0 90.0 1110.0 97347.0 123.0

flag
RSTR 92.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 2.0 1.0 1.0
1.0 777.0 32.0 5.0
S3 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 2.0 1.0 1.0
1.0 1.0 8.0 3.0
SF 2106.0 980.0 10.0 1.0
11.0 5.0 188.0 30.0 9.0 1.0 3.0
1162.0 8.0 280791.0 265.0 4.0 1017.0 129.0
7.0 21.0 2.0 91709.0 3.0
RSTO 1.0 1.0 1.0 456.0
1.0 1.0 1.0 2.0 1.0 1.0 1.0
4.0 1.0 1.0 1.0 1.0 2.0 1.0
1.0 1.0 8.0 68.0 46.0
SH 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 104.0 5.0
1.0 1.0 1.0 1.0

OTH 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 8.0 2.0 1.0
S2 6.0 1.0 1.0 1.0 1.0
1.0 2.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 2.0 1.0 1.0
1.0 1.0 18.0 1.0
RSTOS0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 12.0 1.0 1.0
S1 3.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 2.0
1.0 1.0 55.0 1.0
S0 1.0 1.0 1.0 86745.0 1.0
1.0 173.0 1.0 1.0 22.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 2.0
1.0 19.0 52.0 1.0
REJ 1.0 1.0 1.0 20003.0 1.0
1.0 1230.0 1.0 1.0 1.0 1.0 84.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 221.0 5342.0 1.0
[total] 2214.0 990.0 20.0 107212.0
21.0 15.0 1600.0 41.0 19.0 32.0 13.0
1258.0 18.0 280801.0 275.0 14.0 1031.0
242.0 23.0 31.0 1051.0 97288.0 64.0

src_bytes

mean 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 294248.5286 0 0 0
666707.3462 589.8461 0
std. dev. 35029.5867 35029.5867 35029.5867
35029.5867 35029.5867 35029.5867 35029.5867 35029.5867
35029.5867 35029.5867 35029.5867 35029.5867 35029.5867
35029.5867 35029.5867 35029.5867 35029.5867 35029.5867
35029.5867 35029.5867 35029.5867 1179596.3794 35029.5867
35029.5867 35029.5867 21490326.5249 35029.5867 35029.5867
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 210177.5205 210177.5205 210177.5205
210177.5205 210177.5205 210177.5205 210177.5205 210177.5205
210177.5205 210177.5205 210177.5205 210177.5205 210177.5205
210177.5205 210177.5205 210177.5205 210177.5205 210177.5205
210177.5205 210177.5205 210177.5205 210177.5205 210177.5205

dst_bytes
mean 8093.3851 0 2937.8625 0
4230.522 8172.5994 0.3025 6329.7584 5408.3378 0
1201.8529 0 213037.0029 0 0 2403.7057
710.2715 0 54924.6754 3922054.4907 0 3392.9916
45.3529
std. dev. 603.4524 80.1235 2666.5074 80.1235
7168.6871 80.1235 80.1235 12260.9937 12887.8756 80.1235
240.3706 80.1235 354125.1099 80.1235 80.1235 80.1235
1089.8824 80.1235 179142.1649 2141865.1996 80.1235
37578.0193 268.4649

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 480.7411 480.7411 480.7411 480.7411
480.7411 480.7411 480.7411 480.7411 480.7411 480.7411
480.7411 480.7411 480.7411 480.7411 480.7411 480.7411
480.7411 480.7411 480.7411 480.7411 480.7411 480.7411
480.7411

land
1 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 22.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 2.0 1.0
0 2204.0 980.0 10.0 107202.0
11.0 5.0 1590.0 31.0 9.0 1.0 3.0
1248.0 8.0 280791.0 265.0 4.0 1021.0 232.0
13.0 21.0 1041.0 97277.0 54.0
[total] 2205.0 981.0 11.0 107203.0
12.0 6.0 1591.0 32.0 10.0 23.0 4.0
1249.0 9.0 280792.0 266.0 5.0 1022.0 233.0
14.0 22.0 1042.0 97279.0 55.0

wrong_fragment
mean 0 2.9862 0 0 0
0 0 0 0 0 0 0 0
0 1.4716 0 0 0 0 0
0 0 0

std. dev. 0.25 0.25 0.25 0.25
0.25 0.25 0.25 0.25 0.25 0.25 0.25
0.25 0.25 0.25 0.25 0.25 0.25 0.25
0.25 0.25 0.25 0.25 0.25

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 1.5 1.5 1.5 1.5 1.5
1.5 1.5 1.5 1.5 1.5 1.5 1.5
1.5 1.5 1.5 1.5 1.5 1.5 1.5
1.5 1.5 1.5 1.5

urgent

mean 0 0 0 0 0.1
0 0 0 0.375 0 0 0
0 0 0 0 0 0 0 0
0 0 0

std. dev. 0.1667 0.1667 0.1667 0.1667
0.3 0.1667 0.1667 0.1667 0.696 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1

hot

mean 1.424 0 0.9524 0
0.2857 1.4286 0.0018 2.0952 0.3571 0 0
0 2.8571 0 0 0 8.2087 0
0.2381 0.9286 0.0027 0.045 1.4016

std. dev. 0.2381 0.2381 1.3469 0.2381
0.5714 0.2381 0.2381 1.5089 0.6186 0.2381
0.2381 0.2381 4.7687 0.2381 0.2381 0.2381
12.5853 0.2381 0.5324 4.0475 0.2381 0.8581
0.2381

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 1.4286 1.4286 1.4286 1.4286
1.4286 1.4286 1.4286 1.4286 1.4286 1.4286
1.4286 1.4286 1.4286 1.4286 1.4286 1.4286
1.4286 1.4286 1.4286 1.4286 1.4286 1.4286
1.4286

num_failed_logins

mean 0 0 0 0 0.1
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0.0002 1.0566

std. dev. 0.1667 0.1667 0.1667 0.1667
0.3 0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.5635

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1

logged_in
1 2204.0 1.0 9.0 1.0 6.0
5.0 7.0 31.0 7.0 1.0 2.0 5.0
5.0 1.0 1.0 4.0 1021.0 1.0 2.0
3.0 1.0 69939.0 2.0
0 1.0 980.0 2.0 107202.0
6.0 1.0 1584.0 1.0 3.0 22.0 2.0
1244.0 4.0 280791.0 265.0 1.0 1.0 232.0
12.0 19.0 1041.0 27340.0 53.0
[total] 2205.0 981.0 11.0 107203.0
12.0 6.0 1591.0 32.0 10.0 23.0 4.0
1249.0 9.0 280792.0 266.0 5.0 1022.0 233.0
14.0 22.0 1042.0 97279.0 55.0

lnum_compromised
mean 0 0 0 0 0
0 0 0 0 0 0 0
11.4805 0 0 0 0 0 0
0 0 0.0268 0
std. dev. 6.697 6.697 6.697 6.697
6.697 6.697 6.697 6.697 6.697 6.697 6.697

6.697 18.1523 6.697 6.697 6.697 6.697 6.697
6.697 6.697 6.697 6.697 6.697
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 40.1818 40.1818 40.1818 40.1818
40.1818 40.1818 40.1818 40.1818 40.1818 40.1818
40.1818 40.1818 40.1818 40.1818 40.1818 40.1818
40.1818 40.1818 40.1818 40.1818 40.1818 40.1818
40.1818

lroot_shell
mean 0 0 0.3333 0
0.2 1 0 0.6 0 0 0
0 0.2857 0 0 1 0 0
0 0 0 0.0002 0
std. dev. 0.1667 0.1667 0.4714 0.1667
0.4 0.1667 0.1667 0.4899 0.1667 0.1667 0.1667
0.1667 0.4518 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1

lsu_attempted

mean 0 0 0 0 0
0 0 0 0 0 0.5 0 0
0 0 0 0 0 0 0 0
0.0002 0

std. dev. 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
0.5 0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1

lnum_root

mean 0 0 0 0 0
0 0 0 0 0 0 0
14.9323 0 0 0 0 0 0
0 0 0.0285 0

std. dev. 8.7105 8.7105 8.7105 8.7105
8.7105 8.7105 8.7105 8.7105 8.7105 8.7105
8.7105 8.7105 23.6101 8.7105 8.7105 8.7105
8.7105 8.7105 8.7105 8.7105 8.7105 8.7105
8.7105

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2

1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 52.2632 52.2632 52.2632 52.2632
52.2632 52.2632 52.2632 52.2632 52.2632 52.2632
52.2632 52.2632 52.2632 52.2632 52.2632 52.2632
52.2632 52.2632 52.2632 52.2632 52.2632 52.2632
52.2632

lnum_file_creations
mean 0 0 1.281 0
0.4941 0 0 0.8235 0.4118 0 0.8235
0 1.4118 0 0 1.6471 0 0
0 1.1529 0 0.0058 0
std. dev. 0.2745 0.2745 1.5091 0.2745
0.7548 0.2745 0.2745 1.0198 0.7132 0.2745
0.8235 0.2745 1.372 0.2745 0.2745 0.2745
0.2745 0.2745 0.2745 4.6615 0.2745 0.2745
0.2745
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1.6471 1.6471 1.6471 1.6471
1.6471 1.6471 1.6471 1.6471 1.6471 1.6471
1.6471 1.6471 1.6471 1.6471 1.6471 1.6471
1.6471 1.6471 1.6471 1.6471 1.6471 1.6471
1.6471

lnum_shells

mean 0 0 0.4444 0 0
0 0 0 0 0 0.5 0
0.4286 0 0 1 0 0 0
0 0 0.0004 0

std. dev. 0.1667 0.1667 0.8315 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667
0.5 0.1667 0.7284 0.1667 0.1667 0.1667 0.1667
0.1667 0.1667 0.1667 0.1667 0.1667 0.1667

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1

lnum_access_files

mean 0 0 0.1481 0 0
1.3333 0 0 0.5 0 0.6667 0
0.381 0 0 0 0 0 0
0 0 0.0066 0

std. dev. 0.2222 0.2222 0.419 0.2222
0.2222 0.2222 0.2222 0.2222 0.6455 0.2222
0.6667 0.2222 0.9331 0.2222 0.2222 0.2222
0.2222 0.2222 0.2222 0.2222 0.2222 0.2222
0.2222

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2

1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1.3333 1.3333 1.3333 1.3333
1.3333 1.3333 1.3333 1.3333 1.3333 1.3333
1.3333 1.3333 1.3333 1.3333 1.3333 1.3333
1.3333 1.3333 1.3333 1.3333 1.3333 1.3333
1.3333

lnum_outbound_cmds
mean 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0
std. dev. 0.0017 0.0017 0.0017 0.0017
0.0017 0.0017 0.0017 0.0017 0.0017 0.0017
0.0017 0.0017 0.0017 0.0017 0.0017 0.0017
0.0017 0.0017 0.0017 0.0017 0.0017 0.0017
0.0017
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01

is_host_login
1 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0

1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0
0 2204.0 980.0 10.0 107202.0
11.0 5.0 1590.0 31.0 9.0 22.0 3.0
1248.0 8.0 280791.0 265.0 4.0 1021.0 232.0
13.0 21.0 1041.0 97278.0 54.0
[total] 2205.0 981.0 11.0 107203.0
12.0 6.0 1591.0 32.0 10.0 23.0 4.0
1249.0 9.0 280792.0 266.0 5.0 1022.0 233.0
14.0 22.0 1042.0 97279.0 55.0

is_guest_login
1 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 3.0 1.0 1.0 1.0
3.0 1.0 1.0 1.0 308.0 1.0 1.0
3.0 1.0 372.0 2.0
0 2204.0 980.0 10.0 107202.0
11.0 5.0 1590.0 31.0 7.0 22.0 3.0
1248.0 6.0 280791.0 265.0 4.0 714.0 232.0
13.0 19.0 1041.0 96907.0 53.0
[total] 2205.0 981.0 11.0 107203.0
12.0 6.0 1591.0 32.0 10.0 23.0 4.0
1249.0 9.0 280792.0 266.0 5.0 1022.0 233.0
14.0 22.0 1042.0 97279.0 55.0

count
mean 3.5292 59.9252 1.7416 188.5847
1.1495 1.045 431.4262 9.1262 1.045 4.8269
1.045 1.045 1.045 507.0242 4.271 1.045

1.3646 1.3662 2.09 1.2017 18.3426 8.238
1.5971
std. dev. 1.6967 41.7948 1.1015 69.0255
0.3135 0.1742 168.0869 30.0198 0.1742 16.4517
0.1742 0.1742 0.1742 18.2177 5.1496 0.1742
0.4837 2.9646 1.2067 0.4984 63.8827 17.6869
0.5598
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1.045 1.045 1.045 1.045
1.045 1.045 1.045 1.045 1.045 1.045 1.045
1.045 1.045 1.045 1.045 1.045 1.045 1.045
1.045 1.045 1.045 1.045 1.045

srv_count
mean 3.923 49.5552 1.4527 11.0384
1.1985 1.3619 2.7606 1.4527 1.2257 1.9716
1.0896 17.9379 1.2452 507.019 5.6417 1.0896
1.4527 12.768 95.5174 1.253 1.2477 11.0258
1.6652
std. dev. 1.8589 28.3938 0.7264 6.1494
0.3269 0.4718 11.2234 0.7618 0.3603 0.4278
0.1816 17.2262 0.3813 18.2132 4.5926 0.1816
0.7618 16.4148 68.5687 0.5197 0.4013 21.7718
0.5836
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2

1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1.0896 1.0896 1.0896 1.0896
1.0896 1.0896 1.0896 1.0896 1.0896 1.0896
1.0896 1.0896 1.0896 1.0896 1.0896 1.0896
1.0896 1.0896 1.0896 1.0896 1.0896 1.0896
1.0896

serror_rate
mean 0.003 0.0713 0 0.809
0 0 0.106 0.0659 0 0.9681 0
0 0 0 0 0 0.0025 0.4459
0.696 0 0.0188 0.0016 0.0379
std. dev. 0.0338 0.2135 0.0018 0.3929
0.0018 0.0018 0.0868 0.2467 0.0018 0.1428
0.0018 0.0018 0.0018 0.0018 0.0018 0.0018
0.047 0.4971 0.3648 0.0018 0.1164 0.0279
0.1645
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.011 0.011 0.011 0.011
0.011 0.011 0.011 0.011 0.011 0.011 0.011
0.011 0.011 0.011 0.011 0.011 0.011 0.011
0.011 0.011 0.011 0.011 0.011

srv_serror_rate
mean 0.0031 0 0 0.8091
0 0 0.1089 0 0 1 0

0 0 0 0 0 0.0026 0.4459
0.6133 0 0.0233 0.0018 0.0377
std. dev. 0.0341 0.0033 0.0033 0.393
0.0033 0.0033 0.3115 0.0033 0.0033 0.0033
0.0033 0.0033 0.0033 0.0033 0.0033 0.0033
0.0471 0.4971 0.3085 0.0033 0.1422 0.0265
0.1639
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02

rerror_rate
mean 0.0404 0.0018 0 0.1908
0 0 0.775 0.0171 0 0.032 0
0.069 0 0 0 0 0.0025 0
0 0 0.9721 0.0559 0.9245
std. dev. 0.1274 0.0067 0.0022 0.3929
0.0022 0.0022 0.2865 0.0897 0.0022 0.1429
0.0022 0.2534 0.0022 0.0022 0.0022 0.0022
0.0469 0.0022 0.0022 0.0022 0.1484 0.229
0.2456
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 0.0132 0.0132 0.0132 0.0132
0.0132 0.0132 0.0132 0.0132 0.0132 0.0132
0.0132 0.0132 0.0132 0.0132 0.0132 0.0132
0.0132 0.0132 0.0132 0.0132 0.0132 0.0132
0.0132

srv_rerror_rate
mean 0.0942 0 0 0.1908
0 0.125 0.7734 0.0333 0 0 0
0.069 0 0 0 0 0.002 0
0.0267 0 0.9681 0.0562 0.9245
std. dev. 0.1696 0.0033 0.0033 0.3929
0.0033 0.2165 0.4186 0.1795 0.0033 0.0033
0.0033 0.2534 0.0033 0.0033 0.0033 0.0033
0.0442 0.0033 0.0189 0.0033 0.1685 0.2278
0.2456
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02 0.02 0.02
0.02 0.02 0.02 0.02 0.02

same_srv_rate
mean 0.9989 0.927 0.881 0.0704
1 1 0.0511 0.9173 1 0.9208 1
1 1 1 1 1 0.9967 0.9936
1 1 0.8233 0.9855 1

std. dev. 0.0174 0.2131 0.1792 0.0675
0.0017 0.0017 0.184 0.2585 0.0017 0.249
0.0017 0.0017 0.0017 0.0017 0.0017 0.0017
0.0396 0.0721 0.0017 0.0017 0.319 0.0922
0.0017

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 0.0102 0.0102 0.0102 0.0102
0.0102 0.0102 0.0102 0.0102 0.0102 0.0102
0.0102 0.0102 0.0102 0.0102 0.0102 0.0102
0.0102 0.0102 0.0102 0.0102 0.0102 0.0102
0.0102

diff_srv_rate

mean 0.0023 0.0096 0.241 0.0657
0 0 0.9058 0.0381 0 0.0513 0
0 0 0 0 0 0.0065 0.0045
0 0 0.2143 0.0182 0

std. dev. 0.0352 0.0193 0.3616 0.0337
0.0022 0.0022 0.2673 0.1795 0.0022 0.2128
0.0022 0.0022 0.0022 0.0022 0.0022 0.0022
0.0793 0.0657 0.0022 0.0022 0.3811 0.1167
0.0022

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 0.013 0.013 0.013 0.013
0.013 0.013 0.013 0.013 0.013 0.013 0.013
0.013 0.013 0.013 0.013 0.013 0.013 0.013
0.013 0.013 0.013 0.013 0.013

srv_diff_host_rate
mean 0.1122 0 0 0.0001
0 0.25 0.0003 0 0.125 0.8095 0
0.652 0.1429 0 0.3559 0 0.0113 0.4286
0.2169 0 0.0006 0.134 0
std. dev. 0.2365 0.0026 0.0026 0.0085
0.0026 0.433 0.0127 0.0026 0.3307 0.3927
0.0026 0.4763 0.3499 0.0026 0.2967 0.0026
0.101 0.4949 0.1568 0.0026 0.0207 0.2783
0.0026
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.0159 0.0159 0.0159 0.0159
0.0159 0.0159 0.0159 0.0159 0.0159 0.0159
0.0159 0.0159 0.0159 0.0159 0.0159 0.0159
0.0159 0.0159 0.0159 0.0159 0.0159 0.0159
0.0159

dst_host_count
mean 206.9637 239.3166 2.6667 254.73
178.9 255 253.3505 3.0667 1.5 2.9048 255
3.9471 74 254.9815 66.6174 193 84.4667
114.9264 26.75 32.2 249.9529 148.5137 26.0189

std. dev. 81.0181 41.1853 1.8257 6.7764
116.245 0.1667 15.7052 2.5157 0.5 4.7097
0.1667 10.8952 114.4764 1.6301 104.9934 87.6812
111.5407 115.8941 68.8865 68.5278 24.0107 103.3954
15.2656
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1
1 1

dst_host_srv_count
mean 206.9637 58.2278 3.5556 11.0308
6 247.75 4.4877 13.9 22.75 4.2381 47.5
156.0922 1.5714 254.9081 55.7803 1.6667 40.2078
63.632 5.5833 8.65 1.5356 202.0662 26.0377
std. dev. 81.0181 38.5542 2.6713 6.7205
11.7218 5.0683 20.3004 27.0177 35.6818 2.467
0.5 103.7832 0.7284 3.9439 63.1184 0.4714
33.223 77.0113 3.3281 5.5432 5.8039 86.9122
15.2352
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 1 1 1 1 1
1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1
1 1

dst_host_same_srv_rate

mean 1 0.2468 0.8356 0.0433
0.306 0.9725 0.0147 1 0.875 0.87 0.185
0.9303 0.7157 0.9997 0.6597 0.0133 0.7354
0.5265 0.9167 0.9 0.0041 0.845 1

std. dev. 0.0017 0.1602 0.2556 0.0288
0.4544 0.0192 0.0805 0.0017 0.2165 0.3185
0.005 0.248 0.4495 0.0141 0.4205 0.0125
0.3588 0.4845 0.2764 0.3 0.038 0.3052
0.0017

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01

dst_host_diff_srv_rate

mean 0 0.2028 0.2056 0.0661
0.014 0.0075 0.9033 0 0.25 0.0267 0.02
0.0738 0.0029 0 0.1107 0.03 0.0191
0.4392 0.0008 0.004 0.7647 0.0565 0

std. dev. 0.0017 0.2988 0.2967 0.0108
0.0092 0.0043 0.2459 0.0017 0.433 0.0655
0.0017 0.2614 0.0045 0.0021 0.2528 0.0216

0.0517 0.4834 0.0028 0.0124 0.2744 0.18
0.0017
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01

dst_host_same_src_port_rate
mean 0.0103 0.2468 0.7356 0.0001
0.25 0 0.0735 0.6807 0.875 0.87 0
0.9303 0.7143 0.9997 0.6597 0.0033 0.6613
0.9598 0.625 0.9 0.8873 0.134 0.1053
std. dev. 0.0536 0.1602 0.3745 0.0038
0.4031 0.0017 0.2384 0.3604 0.2165 0.3185
0.0017 0.248 0.4518 0.0141 0.4205 0.0047
0.4595 0.1288 0.3769 0.3 0.2447 0.2808
0.1962
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01

dst_host_srv_diff_host_rate

mean 0 0 0.2083 0 0.0094 0 0.0002 0.0745 0.1172 0.5446 0 0.6031 0 0 0.2086 0 0.0995 0.112 0 0 0 0.0246 0.0189

std. dev. 0.0026 0.0026 0.2503 0.0026 0.0281 0.0026 0.0075 0.2263 0.1955 0.4172 0.0026 0.2536 0.0026 0.0026 0.2367 0.0026 0.1135 0.1272 0.0026 0.0026 0.0026 0.0496 0.1361

weight sum 2203 979 9 107201 10 4 1589 30 8 21 2 1247 7 280790 264 3 1020 231 12 20 1040 97277 53

precision 0.0156

dst_host_serror_rate

mean 0.0022 0.0208 0 0.8091 0 0 0.1113 0 0 0.8937 0.2222 0 0 0 0.0664 0 0.0111 0.4346 0.5699 0.0202 0.0117 0.0021 0.1022

std. dev. 0.0048 0.0347 0.0017 0.3929 0.0017 0.0017 0.1002 0.0017 0.0017 0.2742 0.0017 0.0017 0.0017 0.0017 0.1688 0.0017 0.0506 0.4867 0.2502 0.0858 0.0631 0.0294 0.1473

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.0101 0.0101 0.0101 0.0101
0.0101 0.0101 0.0101 0.0101 0.0101 0.0101
0.0101 0.0101 0.0101 0.0101 0.0101 0.0101
0.0101 0.0101 0.0101 0.0101 0.0101 0.0101
0.0101

dst_host_srv_serror_rate
mean 0.0028 0 0 0.809
0.0254 0 0.1083 0 0 0.6479 0.3169
0 0 0 0 0 0.0033 0.4459
0.5739 0 0.0233 0.0012 0.1031
std. dev. 0.0058 0.0023 0.0023 0.393
0.0761 0.0023 0.3107 0.0023 0.0023 0.3748
0.007 0.0023 0.0023 0.0023 0.0023 0.0023
0.0323 0.4971 0.2502 0.0023 0.1423 0.0158
0.1473
weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53
precision 0.0141 0.0141 0.0141 0.0141
0.0141 0.0141 0.0141 0.0141 0.0141 0.0141
0.0141 0.0141 0.0141 0.0141 0.0141 0.0141
0.0141 0.0141 0.0141 0.0141 0.0141 0.0141
0.0141

dst_host_rerror_rate

mean 0.0504 0.2201 0 0.1908
0.073 0 0.7803 0.0213 0 0.0062 0
0.0691 0.0114 0 0.0008 0.23 0.004 0
0.0025 0.006 0.8902 0.0577 0.8792

std. dev. 0.0742 0.2915 0.0017 0.393
0.219 0.0017 0.254 0.0488 0.0017 0.0191
0.0017 0.2425 0.0181 0.0032 0.0026 0.3253
0.0191 0.0017 0.0083 0.0183 0.2441 0.225
0.1907

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01

dst_host_srv_rerror_rate

mean 0.0504 0 0.0111 0.1908
0.025 0 0.7734 0.0213 0 0 0
0.0662 0 0 0 0 0.0005 0
0 0 0.9653 0.0558 0.8792

std. dev. 0.0742 0.0017 0.0314 0.393
0.075 0.0017 0.4186 0.0488 0.0017 0.0017
0.0017 0.2473 0.0017 0.0017 0.0017 0.0017
0.0067 0.0017 0.0017 0.0017 0.1759 0.2189
0.1907

weight sum 2203 979 9 107201
10 4 1589 30 8 21 2
1247 7 280790 264 3 1020 231
12 20 1040 97277 53

precision 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01 0.01 0.01
0.01 0.01 0.01 0.01 0.01

Time taken to build model: 2.53 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	458349	92.7794 %
Incorrectly Classified Instances	35671	7.2206 %
Kappa statistic	0.8806	
Mean absolute error	0.0063	
Root mean squared error	0.0772	
Relative absolute error	12.1894 %	
Root relative squared error	48.1426 %	
Total Number of Instances	494020	

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.975 0.005 0.485 0.975 0.648 0.686 0.999 0.957 back

0.995 0.001 0.638 0.995 0.777 0.796 0.999 0.636 teardrop

0.556 0.000 0.022 0.556 0.041 0.109 0.999 0.026 loadmodule

0.996 0.000 1.000 0.996 0.998 0.997 1.000 1.000 neptune

0.500 0.003 0.003 0.500 0.006 0.038 0.977 0.002 rootkit

0.750 0.000 0.071 0.750 0.130 0.231 0.995 0.750 phf

0.954 0.002 0.580 0.954 0.722 0.743 0.996 0.604 satan

0.133 0.000 0.018 0.133 0.032 0.049 0.999 0.080 buffer_overflow

0.750 0.003 0.004 0.750 0.009 0.057 0.997 0.004 ftp_write

0.952 0.000 0.323 0.952 0.482 0.554 1.000 0.357 land

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 spy

0.966 0.009 0.205 0.966 0.339 0.443 0.994 0.181 ipsweep

0.429 0.000 0.068 0.429 0.118 0.171 1.000 0.072 multihop

0.999 0.000 1.000 0.999 0.999 0.998 1.000 1.000 smurf

0.985 0.037 0.014 0.985 0.028 0.115 0.998 0.372 pod

0.333 0.000 0.333 0.333 0.333 0.333 1.000 0.386 perl

0.478 0.006 0.143 0.478 0.220 0.259 0.988 0.300 warezclient

0.446 0.001 0.167 0.446 0.243 0.272 0.995 0.112 nmap

0.917 0.000 0.141 0.917 0.244 0.360 0.942 0.317 imap

0.900 0.001 0.055 0.900 0.103 0.222 0.994 0.276 warezmaster

0.907 0.001 0.650 0.907 0.757 0.767 0.998 0.506 portsweep

0.652 0.001 0.997 0.652 0.788 0.774 0.999 0.994 normal

0.943 0.001 0.070 0.943 0.130 0.256 0.989 0.477 guess_passwd

Weighted Avg. 0.928 0.000 0.989 0.928 0.950 0.947 1.000 0.991

=== Confusion Matrix ===

a b c d e f g h i j k l m n o p q r s t u v w <-- classified as

2148 0 0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 48 3 | a = back

0 974 0 0 3 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 | b = teardrop

0 0 5 0 0 0 0 2 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 | c = loadmodule

0 0 0 106757 0 0 23 0 0 41 0 0 0 0 0 0 0 367 5 0 5 3 0 | d = neptune

0 0 0 0 5 0 0 0 2 0 0 0 1 0 0 0 0 0 1 0 1 0 | e = rootkit

1 0 0 0 0 3 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 | f = phf
0 18 0 0 3 0 1516 0 0 0 0 0 0 0 11
0 0 3 0 1 6 31 0 | g = satan
0 0 6 0 4 0 0 4 3 0 0 0 9 0 0
0 0 0 0 2 0 2 0 | h = buffer_overflow
0 0 0 0 0 0 0 0 6 0 0 1 0 0 0
0 0 0 0 0 0 1 0 | i = ftp_write
0 0 0 0 0 0 0 0 0 20 0 0 0 0 0
0 0 0 0 0 1 0 0 | j = land
0 0 0 0 0 0 0 0 0 0 2 0 0 0 0
0 0 0 0 0 0 0 0 | k = spy
0 0 0 0 0 0 0 0 31 0 0 1205 0 0 0
0 0 0 0 2 9 0 0 | l = ipsweep
0 0 0 0 1 0 0 0 1 0 0 0 3 0 0
0 0 0 1 1 0 0 0 | m = multihop
0 142 0 0 0 0 90 0 0 0 0 0 0 280381
61 0 0 0 7 0 0 109 0 | n = smurf
0 0 0 0 2 0 0 0 2 0 0 0 0 0 260
0 0 0 0 0 0 0 0 | o = pod
0 0 0 0 2 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 | p = perl
0 0 30 0 60 0 2 129 230 0 0 13 0 0
13 0 488 3 0 47 1 4 0 | q = warezclient
0 0 1 0 1 0 0 0 2 0 0 99 0 0 22
0 0 103 0 2 0 1 0 | r = nmap
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 11 0 0 0 0 | s = imap
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 2 0 0 18 0 0 0 | t = warezmaster

0 0 0 0 0 0 85 0 1 0 0 0 0 0 0
0 0 2 0 0 943 9 0 | u = portsweep
2281 393 190 6 1587 39 897 85 1101 1 0 4553 30
24 18122 2 2926 139 53 253 485 63446 664 | v = normal
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 0 1 50 | w = guess_passwd

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ นามสกุล	นายผดุง นันอำไพ
วัน เดือน ปีเกิด	วันที่ 3 กุมภาพันธ์ พ.ศ. 2535
จังหวัด และประเทศที่เกิด	จังหวัดกาฬสินธุ์ ประเทศไทย
ประวัติการศึกษา	พ.ศ. 2551 ปริญญาบริหารธุรกิจบัณฑิต (บธ.บ.) มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน พ.ศ. 2561 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยมหาสารคาม
ที่อยู่ที่สามารถติดต่อได้	บ้านเลขที่ 122 หมู่ 6 ตำบลบึงวิชัย อำเภอเมือง จังหวัดกาฬสินธุ์ 46000