การพัฒนาชุดซอฟต์แวร์ตรวจจับผู้บุกรุกโดยใช้การสุ่มกระแสข้อมูล

นายกิติศักดิ์ จิรวรรณกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

DEVELOPMENT OF INTRUSION DETECTION SOFTWARE SUITE USING TRAFFIC SAMPLING

Mr. Kitisak Jirawannakool

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2007

หัวข้อวิทยานิพนธ์	การพัฒนาชุดซอฟต์แวร์ตรวจจับผู้บุกรุกโดยใช้การสุ่มกระแสข้อมูล
โดย	นายกิติศักดิ์ จิรวรรณกูล
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.ยรรยง เต็งอำนวย

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(อาจารย์ จารุมาตร ปิ่นทอง)

..... อาจารย์ที่ปรึกษา

(อาจารย์ ดร.ยรรยง เต็งอำนวย)

..... กรรมการ

(อาจารย์ ธงชัย โรจน์กังสดาล)

..... กรรมการ

(ดร.โกเมน พิบูลย์โรจน์)

กิติศักดิ์ จิรวรรณกูล : การพัฒนาชุดซอฟต์แวร์ตรวจจับผู้บุกรุกโดยใช้การสุ่มกระแสข้อมูล. (DEVELOPMENT OF INTRUSION DETECTION SOFTWARE SUITE USING TRAFFIC SAMPLING) อาจารย์ที่ปรึกษา : อ.ดร.ยรรยง เต็งอำนวย, 46 หน้า.

วิทยานิพนธ์นี้เป็นการพัฒนาชุดซอฟต์แวร์ตรวจจับผู้บุกรุกโดยการสุ่มกระแสข้อมูล ซึ่งใช้การสุ่มแบบ Stratified random ประยุกต์ร่วมกับการตรวจจับผู้บุกรุกของโปรแกรมสนอร์ต เพื่อลดปริมาณข้อมูลของระบบเครือข่าย สำหรับการตรวจจับการบุกรุกในระบบเครือข่ายขนาดใหญ่ ซึ่งเหมาะสำหรับเครื่องคอมพิวเตอร์ขนาดเล็ก

โปรแกรมสนอร์ต เป็นโปรแกรมตรวจจับการบุกรุกผ่านทางระบบเครือข่ายที่ได้รับความนิยมสูงในปัจจุบัน แต่การตรวจจับการบุกรุกนั้นต้องรวบรวมข้อมูลทั้งหมดในระบบเครือข่าย ด้วยเทคนิคการตรวจจับการบุกรุกสองวิธีคือ 1) วิธีการวิเคราะห์พฤติกรรมของข้อมูล และ 2) การวิเคราะห์ลักษณะของการโจมตีเปรียบเทียบกับกฎของโปรแกรม ในระบบเครือข่ายขนาดใหญ่ การรวบรวมข้อมูลจะมีปริมาณมากและมีความต้องการประสิทธิภาพของเครื่องคอมพิวเตอร์สูงสำหรับใช้ในการตรวจจับการบุกรุกของโปรแกรมสนอร์ต

ดังนั้นในงานวิจัยนี้เสนอวิธีการลดปริมาณข้อมูลโดยใช้การสุ่มด้วยวิธี Stratified random และถูกนำไปใช้ในโปรแกรมสนอร์ต งานวิจัยนี้ได้วัดประสิทธิภาพของการตรวจจับผู้บุกรุกประเภทสแกนพอร์ตด้วยโปรแกรมสนอร์ตร่วมกับการสุ่มแบบ Stratified random ซึ่งผลการทดลองพบว่า การสุ่มช่วยลดปริมาณข้อมูลลงอย่างมาก โดยไม่มีผลกระทบต่อประสิทธิภาพการตรวจจับด้วยวิธีการวิเคราะห์ลักษณะการโจมตีเปรียบเทียบกับกฎของโปรแกรมสนอร์ต อย่างไรก็ตามระบบนี้มีผลกระทบต่อประสิทธิภาพการตรวจจับด้วยการวิเคราะห์พฤติกรรม

ภาควิชา....วิศวกรรมคอมพิวเตอร์.... ลายมือชื่อนิสิต.....

สาขาวิชา....วิทยาศาสตร์คอมพิวเตอร์.... ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา.....2550.....

4871401521 : MAJOR COMPUTER SCIENCE

KEY WORD: IDS/ SAMPLING/ SNORT/ sFlow/ TRAFFIC SAMPLING

KITISAK JIRAWANNAKOOL: DEVELOPMENT OF INTRUSION DETECTION SOFTWARE SUITE USING TRAFFIC SAMPLING. THESIS ADVISOR: YUNYONG TENG-AMNUAY, Ph.D., 46 pp.

The objective of this research is to propose the intrusion detection software suite based on the open source intrusion detection system (Snort) with Stratified random sampling. By applying Stratified random sampling, it significantly reduces the amount of collected data needed for the intrusion detection of the Snort. The proposed solution is suitable to the computer with low computation power to be able to handle the large amount data of the large network for intrusion detection.

Snort is the famous network IDS software. Data is collected from the network to analyze the traffic pattern and determine if there is any attack to the network. Snort executes two intrusion detection techniques 1) Behavior-based intrusion detection technique which analyzes the intrusion traffic pattern and 2) Rules-based intrusion detection technique which analyzes intrusion signature within the traffic. Snort requires high performance computer to handle the large amount of data of the large network.

Therefore, this research proposes the software suite that includes Snort with Stratified random sampling. The research investigates the trade-off of applying Stratified random sampling to the Snort in several aspects such as the decreasing amount of collected data, the accuracy of the intrusion detection. By experimented with the real generated data of the port scan attack, the Snort with Stratified random sampling reduces the amount of collected data but still provides the detection accuracy of the Snort in Rule-based detection. However the Snort with Stratified random sampling reduces the accuracy of Behavior-based intrusion detection.

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

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

Academic year .....2007.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์ และความช่วยเหลืออย่างยิ่งจาก อ.ดร.ยรรยง เต็งอำนวย อาจารย์ที่ปรึกษา ซึ่งให้ข้อคิด แนวทาง และคำปรึกษา ตลอดจนเป็นผู้ตรวจทานแก้ไข ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วง ขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ อ.จารุมาตร ปิ่นทอง และอ.ธงชัย โรจน์กังสดาล ประธานกรรมการและกรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำในการแก้ไขวิทยานิพนธ์ให้มีคุณภาพยิ่งขึ้น ขอขอบพระคุณคณาจารย์ในภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่านที่ประสิทธิ์ประสาทความรู้อันมีค่ายิ่งแก่ผู้วิจัย ตลอดทั้งเจ้าหน้าที่ธุรการประจำภาควิชาวิศวกรรมคอมพิวเตอร์ที่ให้ความช่วยเหลือและอำนวยความสะดวกในด้านต่างๆ

ขอขอบพระคุณ ดร.โกเมน พิบูลย์โรจน์ ดร.ศิวรักษ์ ศิวโมกษธรรม และดร.กิตติ วงศ์ถาวราวัฒน์ ที่ให้คำแนะนำเกี่ยวกับการทำวิจัย ซึ่งเป็นพื้นฐานของการเขียนวิทยานิพนธ์นี้ คุณณัฏฐ์ ปิยะปราโมทย์ คุณไตรรัตน์ พุทธรักษา และคุณจิรพัฒน์ สุมานนท์ ที่ให้ความช่วยเหลือเรื่องการเขียนโปรแกรมสำหรับการทดสอบทั้งหมด คุณฌานิน เหลืองอิงคะสุต และคุณทรงฤทธิ์ ศรีลาศักดิ์ ที่ให้ความช่วยเหลือเรื่องการเก็บข้อมูลสำหรับการทดสอบ อ.วชิรา พรหมสาขา ณ สกลนคร และ ผศ.นริศ เจริญพร ที่ให้ความช่วยเหลือเรื่องข้อมูลและหนังสือเกี่ยวกับสถิติศาสตร์ คุณชวลิต ทินกรสูติบุตร คุณเลอศักดิ์ ลิ้มวิวัฒน์กุล และคุณศิวพงษ์ นิยมพานิช ที่ให้ความช่วยเหลือเกี่ยวกับการใช้งานระบบปฏิบัติการลินุกซ์และเทคนิคการสแกนพอร์ต คุณณัฐพงษ์ แสงเลิศศิลปชัย และคุณเลิศพงษ์ เลิศไพศาลวงศ์ ที่ให้ความช่วยเหลือเรื่องงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์ฉบับนี้ นอกจากนี้ขอขอบคุณ คุณพนิตา เมนะเนตร ที่ช่วยตรวจรายงานและช่วยแนะนำการเตรียมการนำเสนอหัวข้อและวิทยานิพนธ์

ที่สำคัญที่สุดขอบพระคุณ คุณพ่อ คุณแม่ คุณยาย พี่ๆ และเพื่อนๆ ที่เป็นแรงผลักดัน เป็นกำลังใจ ที่สำคัญให้ตลอดการศึกษาครั้งนี้

สารบัญ

หน้า

หน้า

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัญหาการบุกรุกโจมตีระบบเครือข่ายคอมพิวเตอร์ในปัจจุบันนั้นทวีความรุนแรงและเพิ่มจำนวนขึ้นอย่างรวดเร็ว โดยเป้าหมายของผู้ที่ประสงค์ร้ายเหล่านี้จะมุ่งไปที่ระบบเครือข่ายระดับกลางถึงระดับใหญ่ อาจจะเพียงเพื่อชื่อเสียง หรือทรัพย์สินที่เป็นข้อมูลซึ่งมีค่าบนระบบเครือข่าย

ในระบบเครือข่ายระดับกลางถึงระดับใหญ่ มีปริมาณการใช้งานระบบเครือข่ายที่เข้าออกระบบนั้นมีปริมาณมหาศาล ถ้าหากนำข้อมูลทั้งหมดมาผ่านเครื่องแม่ข่ายที่ให้บริการตรวจจับผู้บุกรุก (Intrusion Detection System) ก็อาจจะทำได้ยาก หรือจะทำให้เครื่องแม่ข่ายดังกล่าวหยุดให้บริการก็เป็นได้ แต่ถ้าหากทำได้ก็อาจจำเป็นต้องใช้เครื่องแม่ข่ายที่มีประสิทธิภาพสูง ซึ่งสิ่งที่ตามมาคือค่าใช้จ่ายสำหรับอุปกรณ์ต่างๆ เหล่านี้

ดังนั้นจากปัญหาทั้งหมดที่ได้กล่าวมาในข้างต้นนั้น การวิจัยนี้จึงประยุกต์เทคโนโลยีการสุ่มกระแสข้อมูลหรือเอสโฟล์ว (sFlow) [1] ที่ใช้วิธีการเลือกดักข้อมูล (packet) และสุ่มออกมาเพื่อลดปริมาณข้อมูลให้เพียงพอต่อการใช้งานกับเครื่องแม่ข่ายที่มีขนาดเล็ก จากนั้นจึงนำข้อมูลที่ได้มาเข้ากระบวนการของระบบตรวจจับผู้บุกรุก เพื่อหาความผิดปกติที่เกิดขึ้นกับระบบเครือข่ายคอมพิวเตอร์ได้ ทำให้เจ้าหน้าที่ดูแลระบบเครือข่ายทำงานได้สะดวกขึ้น

1.2 วัตถุประสงค์ของการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อพัฒนาระบบซอฟต์แวร์ที่อาศัยการสุ่มแพ็กเก็ตที่ใช้ในการวิเคราะห์หาผู้บุกรุก เพื่อลดปริมาณข้อมูล

1.3 ขอบเขตวิทยานิพนธ์

1. ใช้หลักการทำงานของเอสโฟล์ว ตามแนวทางของ RFC 3176 [2]
2. ใช้เครื่องแม่ข่ายที่เป็น Intel based และใช้ระบบปฏิบัติการลินุกซ์
3. สามารถปรับค่าการสุ่มได้อย่างอัตโนมัติได้ เพื่อให้เครื่องสามารถทำงานได้ทันกับกระแสข้อมูลที่เข้ามา
4. ใช้ข้อมูลเพื่อทดสอบจากหลากหลายแหล่งข้อมูล
5. มีส่วนของการแสดงผลแบบเว็บเพื่อช่วยในการแจ้งเตือน

1.4 ขั้นตอนการดำเนินงาน

การจัดทำโครงงานมีขั้นตอนการปฏิบัติดังนี้

1. ศึกษาทฤษฎีที่เกี่ยวข้องกับการสุ่มดักแพ็กเก็ต โดยมุ่งประเด็นไปที่เทคโนโลยีของเอสโฟล์ว ว่ามีวิธีการที่ใช้ในการสุ่มอย่างไร และรูปแบบของข้อมูลที่เก็บนั้นจะสามารถส่งต่อให้โปรแกรมตรวจจับผู้บุกรุกได้หรือไม่

2. ศึกษารูปแบบและวิธีใช้งานโปรแกรมตรวจจับผู้บุกรุกสนอร์ต (Snort)

3. ศึกษาวิธีการพัฒนาโปรแกรมบนระบบปฏิบัติการลินุกซ์ รวมทั้งพัฒนาซอฟต์แวร์

4. ทดสอบการติดตั้งและการใช้งาน

5. ทดสอบหาค่าอัตราการสุ่มที่เหมาะสม ซึ่งสามารถทำให้เครื่องที่นำมาใช้นั้นสามารถทำงานได้อย่างต่อเนื่อง กล่าวคือระบบสามารถทำงานได้ทันต่อการรับกระแสข้อมูล

6. กำหนดค่าการสุ่มที่แนะนำสำหรับองค์กรหรือหน่วยงานต่างๆ ตามขนาดของระบบเครือข่าย

7. เขียนสรุปเป็นรายงานรูปเล่ม

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อลดการใช้งานทรัพยากรของเครื่องแม่ข่ายที่จะนำมาให้บริการ หาอัตราการสุ่มที่เหมาะสมกับเครื่องและค่าที่แนะนำสำหรับองค์กรตามขนาดของระบบเครือข่าย

2. เป็นซอฟต์แวร์ที่ใช้งานง่ายและช่วยลดภาระของผู้ดูแลระบบเครือข่ายในองค์กรต่างๆ

3. มีการพัฒนาระบบเพื่อความปลอดภัยบนระบบเครือข่ายของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งจะเป็นผลให้ผู้ที่สนใจ สามารถนำความรู้ในโครงงาน นำไปประยุกต์หรือต่อยอดพัฒนาต่อไปได้

4. อำนวยความสะดวกกับหน่วยงานหรือองค์กรที่ไม่มีบุคคลากรที่จะเฝ้าระวังระบบเครือข่ายตลอด 24 ชั่วโมง

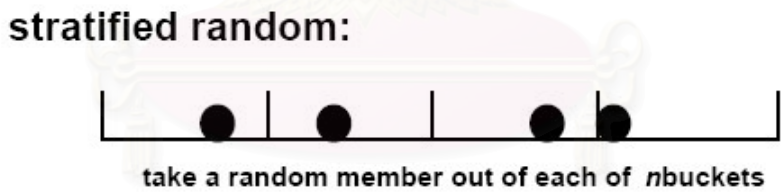
5. มีค่าใช้จ่ายต่ำเพราะเป็นซอฟต์แวร์ประเภท open source และสามารถใช้งานกับเครื่องคอมพิวเตอร์ส่วนบุคคล (PC) ได้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้กล่าวถึงการนำทฤษฎีต่างๆ มาใช้ เช่น เอสโฟล์ว ซึ่งเป็นแนวความคิดหลักและกล่าวถึงวิธีการสุ่มที่นำมาใช้ในงานวิจัย กระบวนการทำงานของระบบตรวจจับผู้บุกรุกชื่อสนอร์ต และตัวอย่างเทคนิคการสแกนพอร์ต ที่จะแสดงคำอธิบายและลักษณะข้อมูลที่เกิดจากการสแกนพอร์ตด้วย รวมทั้งผลงานวิจัยที่เกี่ยวข้องกับการสุ่มและระบบตรวจจับผู้บุกรุกด้วย

2.1 เอสโฟล์ว

เอสโฟล์ว (sFlow) [1] เป็นเทคโนโลยีหนึ่งที่ใช้ในการจับกระแสข้อมูลที่วิ่งผ่านไปในระบบเครือข่าย โดยเอสโฟล์วย่อมาจาก Sampling Flow ซึ่งก็คือโฟล์ว (flow) หรือกระแสข้อมูลที่มาผ่านการสุ่ม ด้วยการเลือกสุ่มเป็นบางตัวโดยใช้หลักสถิติที่ว่าข้อมูลตัวแทนจากการสุ่มสามารถแทนข้อมูลทั้งหมดได้ ซึ่งวิธีการสุ่มนั้นก็มีหลากหลายวิธี แต่วิธีที่ใช้ในการวิจัยนี้คือวิธีการสุ่มแบบ stratified random [3] เนื่องจากวิธีการสุ่มนี้ส่งผลกระทบต่อความเสียหายของข้อมูลในเชิงเวลาน้อยสุดเมื่อเปรียบเทียบกับการสุ่มแบบอื่นๆ และมีหลายบทความวิจัยใช้การสุ่มวิธีนี้ [3][8] ซึ่งวิธีการสุ่มแบบ stratified random นี้เป็นการสุ่มโดยเลือกสมาชิกออกจากทุกๆ กลุ่มที่แบ่ง ดังรูปที่ 2.1



รูปที่ 2.1 รูปแบบวิธีการสุ่มแบบ stratified random

2.2 ระบบตรวจจับผู้บุกรุก

ระบบตรวจจับผู้บุกรุก (Intrusion Detection System : IDS) เป็นระบบที่ช่วยตรวจหาสิ่งผิดปกติที่แปลกปลอมเข้ามายังเครื่องหรือระบบเครือข่ายได้ ซึ่งสามารถตรวจจับได้ทั้งสิ่งผิดปกติบนเครื่อง (host-based IDS) และ ระบบที่ใช้ตรวจจับผู้บุกรุกบนระบบเครือข่ายคอมพิวเตอร์ (network-based IDS) ซึ่งในงานวิจัยนี้จะกล่าวถึงโปรแกรมที่ชื่อสนอร์ต (Snort) [10] เป็นระบบที่ใช้ตรวจจับผู้บุกรุกบนระบบเครือข่ายที่ได้รับความนิยมเป็นอย่างมาก และยังเป็นโปรแกรมประเภทโอเพ่นซอร์ส (open source) ด้วย ซึ่งมีการทำงานอยู่สองลักษณะคือ

2.2.1 **การตรวจจับด้วยพฤติกรรมของข้อมูล** (Preprocessor)

เป็นโมดูลหนึ่งในโปรแกรมสนอร์ตที่ใช้ในการวิเคราะห์พฤติกรรมของปริมาณการใช้งานระบบเครือข่าย ซึ่งผู้ดูแลระบบสามารถปรับแต่งค่าขีดแบ่งขั้นต่ำ (threshold) เพื่อให้โปรแกรมมีความแม่นยำในการวิเคราะห์มากยิ่งขึ้น ซึ่งการตรวจจับพฤติกรรมของโปรแกรมสนอร์ตที่เกี่ยวกับการสแกนพอร์ตนั้นประกอบด้วย

2.2.1.1 Open Port

2.2.1.2 TCP/UDP Portscan

2.2.1.3 TCP/UDP Portsweep

2.2.1.4 TCP/UDP Decoy Portscan

2.2.1.5 TCP/UDP Distributed Portscan

2.2.2 **การตรวจจับด้วยกฎ**

ในโปรแกรมสนอร์ตนั้น ชุดของกฎ (rules set) เป็นเหมือนข้อมูลที่ให้เปรียบเทียบหรือบ่งบอกให้โปรแกรมใช้ในการระบุแพ็กเก็ตว่าเป็นการบุกรุกหรือไม่ และมีกฎต่างๆ มากมาย และสำหรับการสแกนพอร์ตนั้นอาศัยกฎที่ชื่อ scan.rules ซึ่งจะบอกลักษณะของแพ็กเก็ตของการสแกนพอร์ตประเภทต่างๆ และคำที่ใช้แจ้งเตือนด้วย

2.3 **เทคนิคการสแกนพอร์ต**

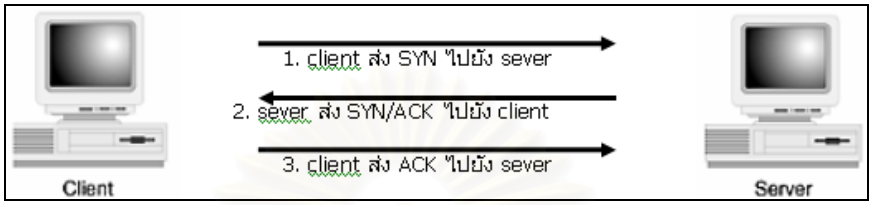
การสแกนพอร์ตเป็นกระบวนการในการติดต่อไปที่พอร์ต TCP หรือ UDP ของเครื่องเป้าหมาย มีจุดประสงค์เพื่อตรวจสอบว่ามีบริการใดบ้างบนระบบที่รอรับการเชื่อมต่อ หรืออยู่ในสถานะที่ให้บริการได้ ซึ่งจัดว่าเป็นขั้นตอนเบื้องต้นของแฮกเกอร์ในการโจมตีไปยังเหยื่อเป้าหมาย และตัวอย่างของเทคนิคการสแกนพอร์ตที่ได้รับความนิยมได้แก่

- The Vanilla TCP connect scan
- The TCP SYN (Half Open) scans
- The TCP FIN scan
- The TCP XMAS scan
- The TCP NULL scan
- The TCP ACK scan
- The TCP Windows scan

โดยมีรายละเอียดดังนี้

2.3.1 The Vanilla TCP connect scan

เป็นเทคนิคการสแกนพอร์ตขั้นพื้นฐานและง่ายที่สุด คือจะใช้ connect system call ของระบบปฏิบัติการไปบนระบบเหยื่อเป้าหมาย ด้วยกลไกมาตรฐานที่เรียกว่า TCP three-way handshake ดังรูปที่ 2.2 เพื่อเปิดการเชื่อมต่อไปยังทุกๆ พอร์ตที่เปิดอยู่ โดยใช้คำสั่ง nmap –sT ซึ่งจะมีลักษณะข้อมูลดังรูปที่ 2.3



รูปที่ 2.2 ลักษณะการส่งข้อมูลของ TCP's three-way handshake

```
00:22:22.542560 192.168.97.81.4982 > 192.168.97.113.837: S 2198837618:2198837618(0)
win 5840 <mss 1460,sackOK,timestamp 4781298 0,nop,wscale 0> (DF)
00:22:22.542659 192.168.97.81.4983 > 192.168.97.113.2038: S
2192554354:2192554354(0)
win 5840 <mss 1460,sackOK,timestamp 4781298 0,nop,wscale 0> (DF)
00:22:22.542724 192.168.97.81.4984 > 192.168.97.113.squid: S
2204737681:2204737681(0)
win 5840 <mss 1460,sackOK,timestamp 4781298 0,nop,wscale 0> (DF)
```

รูปที่ 2.3 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sT

2.3.2 The TCP SYN (Half Open) scans

เทคนิคนี้ผู้โจมตีส่งแพ็กเก็ต SYN ไปยังเป้าหมายและรอการตอบสนอง ถ้าพอร์ตถูกเปิดไว้เป้าหมายจึงส่ง SYN/ACK กลับมา การโจมตีทำได้โดยการใช้คำสั่ง nmap –sS และมีลักษณะข้อมูลดังรูปที่ 2.4

```
16:34:50.602112 192.168.97.81.53613 > 192.168.97.113.346: S 1540627498:1540627498(0)
win 3072
16:34:50.602209 192.168.97.81.53613 > 192.168.97.113.4672: S 1540627498:1540627498(0)
win 3072
16:34:50.602312 192.168.97.81.53613 > 192.168.97.113.49400: S 1540627498:1540627498(0)
win 3072
16:34:50.602361 192.168.97.81.53613 > 192.168.97.113.1517: S 1540627498:1540627498(0)
win 3072
```

รูปที่ 2.4 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sS

2.3.3 The TCP FIN scan

เทคนิคนี้โจมตีโดยการส่งแพ็กเก็ต TCP ที่เซตค่า flag FIN เป็น 1 (TCP FIN) ไปยังระบบของเหยื่อเป้าหมาย สำหรับพอร์ตต่าง ๆ ที่ปิดอยู่จะตอบสนองกลับไปด้วย RST ส่วนพอร์ต

ที่เปิดจะไม่สนใจแพ็กเก็ตเหล่านั้นเลย และทำการโจมตีด้วยคำสั่ง nmap –sF ซึ่งจะมีลักษณะข้อมูลดังรูปที่ 2.5

```
22:56:43.365785 192.168.97.81.55948 > 192.168.97.113.13711: F 0:0(0) win 4096
22:56:43.365942 192.168.97.81.55948 > 192.168.97.113.1499: F 0:0(0) win 4096
22:56:43.366041 192.168.97.81.55948 > 192.168.97.113.511: F 0:0(0) win 4096
22:56:43.366133 192.168.97.81.55948 > 192.168.97.113.1366: F 0:0(0) win 4096
22:56:43.366229 192.168.97.81.55948 > 192.168.97.113.633: F 0:0(0) win 4096
```

รูปที่ 2.5 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sF

2.3.4 The TCP XMAS

ถูกใช้เพื่อหาพอร์ตบนเครื่องเหยื่อเป้าหมายที่อยู่ในสถานะ listening โดยจะส่งแพ็กเก็ตที่ใช้ตัวบ่งชี้ (flag) เป็น URG PSH และ FIN แทน SYN ACK และ RST ในส่วนหัวของ TCP ไปยังพอร์ตของเครื่องเป้าหมาย ทั้งนี้เพื่อหลบหลีกการตรวจจับให้มากที่สุด ซึ่งถ้าพอร์ต TCP ของเครื่องเป้าหมายปิดอยู่ พอร์ตนั้นก็จะส่ง RST กลับมา แต่ถ้าพอร์ตเปิดอยู่ก็จะไม่สนใจแพ็กเก็ตนั้นเลย ใช้คำสั่ง nmap –sX ซึ่งจะมีลักษณะข้อมูลดังรูปที่ 2.6

```
23:44:18.565647 192.168.97.81.54169 > 192.168.97.113.1987: FP 0:0(0) win 2048 urg 0
23:44:18.565742 192.168.97.81.54169 > 192.168.97.113.704: FP 0:0(0) win 2048 urg 0
23:44:18.565796 192.168.97.81.54169 > 192.168.97.113.1348: FP 0:0(0) win 2048 urg 0
23:44:18.565845 192.168.97.81.54169 > 192.168.97.113.1496: FP 0:0(0) win 2048 urg 0
23:44:18.565894 192.168.97.81.54169 > 192.168.97.113.1407: FP 0:0(0) win 2048 urg 0
23:44:18.565942 192.168.97.81.54169 > 192.168.97.113.1432: FP 0:0(0) win 2048 urg 0
23:44:18.565990 192.168.97.81.54169 > 192.168.97.113.720: FP 0:0(0) win 2048 urg 0
```

รูปที่ 2.6 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sX

2.3.5 The TCP NULL scan

เทคนิคนี้จะส่งแพ็กเก็ต TCP ที่มี sequence number แต่ไม่มี flag ออกไปยังเครื่องเป้าหมาย ถ้าพอร์ตปิดอยู่จะส่งแพ็กเก็ต RST กลับมา แต่ถ้าพอร์ตเปิดอยู่ ก็จะไม่สนใจแพ็กเก็ตนั้นเลย ใช้คำสั่ง nmap –sN ซึ่งจะมีลักษณะข้อมูลดังรูปที่ 2.7

```
23:52:42.155569 192.168.97.81.53717 > 192.168.97.113.3052: . win 4096
23:52:42.155771 192.168.97.81.53717 > 192.168.97.113.3900: . win 4096
23:52:42.155906 192.168.97.81.53717 > 192.168.97.113.695: . win 4096
23:52:42.156028 192.168.97.81.53717 > 192.168.97.113.589: . win 4096
23:52:42.156153 192.168.97.81.53717 > 192.168.97.113.853: . win 4096
```

รูปที่ 2.7 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sN

2.3.6 The TCP ACK scan

เป็นเทคนิคที่ใช้ TCP แพ็กเก็ตที่มี flag เป็น ACK ส่งไปยังพอร์ตเครื่องปลายทาง ถ้าพอร์ตเปิดอยู่ เครื่องเป้าหมายจะส่ง RST กลับมา แต่ถ้าปิดอยู่ก็จะไม่สนใจแพ็กเก็ตนั้น ใช้คำสั่ง nmap –sA ซึ่งจะมีลักษณะข้อมูลดังรูปที่ 2.8

```
00:37:02.907732 192.168.97.81.63983 > 192.168.97.113.1496: . ack 3333408271 win 2048
00:37:02.907784 192.168.97.81.63983 > 192.168.97.113.2006: . ack 3333408271 win 2048
00:37:02.907855 192.168.97.81.63983 > 192.168.97.113.967: . ack 3333408271 win 2048
00:37:02.907904 192.168.97.81.63983 > 192.168.97.113.223: . ack 3333408271 win 2048
00:37:02.908018 192.168.97.81.63983 > 192.168.97.113.7010: . ack 3333408271 win 2048
```

รูปที่ 2.8 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sA

2.3.7 The TCP Windows scan

เทคนิคการสแกนนี้จะตรวจสอบพอร์ตที่เปิดอยู่ โดยอาศัยช่องโหว่จากความผิดปกติบางอย่างในการแจ้งค่า TCP Windows Size ของโพรโตคอล TCP/IP ใช้คำสั่ง nmap –sW ซึ่งจะมีลักษณะข้อมูลดังรูปที่ 2.9

```
00:54:05.927433 192.168.97.81.33252 > 192.168.97.113.675: . ack 616330702 win 1024
00:54:05.927523 192.168.97.81.33252 > 192.168.97.113.3049: . ack 616330702 win 1024
00:54:05.927613 192.168.97.81.33252 > 192.168.97.113.668: . ack 616330702 win 1024
00:54:05.927702 192.168.97.81.33252 > 192.168.97.113.672: . ack 616330702 win 1024
00:54:05.927792 192.168.97.81.33252 > 192.168.97.113.224: . ack 616330702 win 1024
00:54:05.927891 192.168.97.81.33252 > 192.168.97.113.725: . ack 616330702 win 1024
```

รูปที่ 2.9 ลักษณะข้อมูลหลังจากใช้คำสั่ง nmap –sW

2.4 ลักษณะของกฎที่เกี่ยวกับการสแกนพอร์ตของโปรแกรมสนอร์ต

ชุดกฎของโปรแกรมสนอร์ตนั้นสามารถเขียนขึ้นมาเองได้ ซึ่งมีหลายๆ ที่ได้จัดเตรียมชุดกฎในการตรวจจับผู้บุกรุกต่างๆ ไว้ให้แล้ว สำหรับชุดกฎของการสแกนพอร์ตที่ใช้ในวิทยานิพนธ์ฉบับนี้ดาวน์โหลดจาก Bleeding Edge Snort [15] และคัดลอกออกมาเฉพาะที่เกี่ยวกับการตรวจจับการบุกรุกประเภทสแกนพอร์ตด้วยโปรแกรม nmap [14] ซึ่งเทคนิคการสแกนพอร์ตที่ไม่มีชุดกฎของโปรแกรมสนอร์ตนั้น เนื่องจากลักษณะของแพ็กเก็ตนั้นเหมือนกับแพ็กเก็ตปกติ แต่การตรวจจับการบุกรุกเหล่านั้นจะใช้วิธีการตรวจจับจากพฤติกรรมของแพ็กเก็ต ตัวอย่างของชุดกฎที่เกี่ยวข้องกับการตรวจจับการบุกรุกประเภทสแกนพอร์ตดังรูปที่ 2.10

```
alert ip $EXTERNAL_NET any -> $HOME_NET any (msg: "BLEEDING-EDGE SCAN NMAP
-sO"; dsize: 0; ip_proto: 21; reference:arachnids,162; classtype:
attempted-recon; sid: 2000536; rev:3; )

alert tcp $EXTERNAL_NET any -> $HOME_NET any (msg: "BLEEDING-EDGE SCAN NMAP
-sS"; fragbits: !D; dsize: 0; flags: S,12; ack: 0; window: 2048;
reference:arachnids,162; classtype: attempted-recon; sid: 2000537; rev:3; )

alert tcp $EXTERNAL_NET any -> $HOME_NET any (msg: "BLEEDING-EDGE SCAN NMAP
-sA (1)"; fragbits: !D; dsize: 0; flags: A,12; window: 1024;
reference:arachnids,162; classtype: attempted-recon; sid: 2000538; rev:4; )

alert tcp $EXTERNAL_NET any -> $HOME_NET any (msg: "BLEEDING-EDGE SCAN NMAP
-sA (2)"; fragbits: !D; dsize: 0; flags: A,12; window: 3072;
reference:arachnids,162; classtype: attempted-recon; sid: 2000540; rev:4; )
```

รูปที่ 2.10 ตัวอย่างกฎที่เกี่ยวกับการตรวจจับการบุกรุกแบบสแกนพอร์ตที่ใช้ในงานวิจัยนี้

2.5 งานวิจัยที่เกี่ยวข้อง

Claffy Polyzos และ Braun [3] ศึกษาหาผลกระทบของพารามิเตอร์ที่เกี่ยวกับการสุ่ม ซึ่งการทดลองนั้นจะประกอบไปด้วยกลุ่มตัวอย่างที่มีขนาดใหญ่ และทดลองตามจุดยึดหลักของวิธีการสุ่มแบบต่างๆ ซึ่งประกอบไปด้วย Systematic เป็นการสุ่มที่จะเลือกเอาสมาชิกตัวแรกของทุกๆ กลุ่มที่แบ่ง Stratified random เป็นการสุ่มเลือกสมาชิกออกจากทุกๆ กลุ่มที่แบ่ง และ Simple random เป็นการสุ่มสมาชิกตามจำนวนที่ต้องการออกจากข้อมูลทั้งหมด เป็นต้น นอกจากนั้นยังยึดหลักตาม วิธีการที่เปลี่ยนแปลงตามเวลา เปรียบเทียบกับเหตุการณ์ ข้อมูลทั้งหมด หรือข้อมูลจากการสุ่มเพียงเล็กน้อย รวมทั้งช่วงเวลาหรือระยะเวลา ที่เก็บกลุ่มตัวอย่างด้วย ซึ่งจากงานวิจัยนี้มาเจาะประเด็นเรื่องวิธีการสุ่มที่ใช้สรุปว่าการสุ่มทั้งสามแบบนั้นมีประสิทธิภาพแตกต่างกันเพียงเล็กน้อยเมื่อใช้วิธียึดหลักตามแพ็กเก็ต ส่วนการใช้เวลาเป็นหลักยึดนั้นได้ผลไม่ค่อยดีนักเนื่องจากบางช่วงเวลาก็มีจำนวนแพ็กเก็ตมากหรือบ้างก็มีน้อย ซึ่งงานวิจัยของ B. Choi J. Park และ Z. Zhang [8] ได้นำวิธีการสุ่มแบบ stratified random ใช้ในการปรับขนาดของโฟล์ว โดยใช้วิธีการจับโฟล์วที่มีขนาดใหญ่ ที่เรียกว่า Elephant Flow ควบคู่กับการสุ่ม

ต่อมา J. Li M. Xu และ Q. Zhao [6] ได้เสนอวิธีการลดขนาดข้อมูลที่ถูกเก็บเพื่อใช้ในการวิเคราะห์หาผู้บุกรุก ซึ่งจะถูกเรียกว่า IP Traceback วิธีการสุ่มนั้นใช้วิธีที่ชื่อ One-bit Random Marking and Sampling (ORMS) ซึ่งอาศัยการแฮชในการย่อยแพ็กเก็ต (Packet Digesting) จากนั้น D. Brauckhoff และคณะ [7] เสนอผลกระทบของการสุ่ม เช่นจำนวนของแพ็กเก็ตหรือโฟล์วเปรียบเทียบกับเวลา หรือค่าของเอนโทรปีเปรียบกับเวลา โดยวิธีการสุ่มนั้นดักแพ็กเก็ตทุกแพ็กเก็ตที่ 10 100 250 และ 1000 ผลจากการศึกษานี้คือ โฟล์วที่มีผลต่อค่าการวัดต่างๆ ถ้าใช้เอนโทรปีจะมีความยืดหยุ่นต่อการสุ่มตัวอย่างมากกว่า และ Ralf Hildebrandt [9] เสนอวิธีการทำงานของ nmap ที่เป็นเครื่องมือในการสแกนพอร์ตของเครื่องเป้าหมายในระบบเครือข่าย และ

สนอร์ต ที่เป็นโปรแกรมตรวจจับผู้บุกรุก ซึ่งจะคอยพยายามช่วยเหลือผู้ดูแลระบบในการป้องกันไม่ให้ถูกบุกรุก นอกจากนี้บทความนี้จะแสดงให้เห็นถึงความเกี่ยวข้องกันระหว่างการสแกนพอร์ตด้วยโปรแกรม nmap และโปรแกรมตรวจจับผู้บุกรุกสนอร์ต เป็นต้น

จากที่ได้กล่าวแล้วนั้นทำให้ทราบถึงทฤษฎีที่นำมาใช้ในงานวิจัยชิ้นนี้ เช่น เอสโฟล์ว ระบบตรวจจับผู้บุกรุก และเทคนิคการสแกนพอร์ต เป็นต้น นอกจากนี้ยังกล่าวถึงผลงานวิจัยที่เกี่ยวข้องกันด้วย จากนั้นในบทถัดไปจะกล่าวถึงผลงานวิจัยชิ้นนี้ ว่ามีแนวทางการดำเนินงานรวมทั้งผลที่ได้เป็นอย่างไร

บทที่ 3
งานวิจัย

บทที่ผ่านกล่าวให้ทราบถึงทฤษฎีต่างๆ รวมทั้งผลงานวิจัยที่เกี่ยวข้อง ในบทนี้จะกล่าวถึงการนำเอาทฤษฎีต่างๆ มาประยุกต์ใช้การพัฒนาโปรแกรมสำหรับสุ่มแพ็กเก็ต เริ่มตั้งแต่เหตุผลการเลือกใช้การสุ่มด้วยวิธี Stratified random และกระบวนการทำงานของโปรแกรมสุ่มแพ็กเก็ต เป็นต้น แล้วจากนั้นจะนำเสนอผลการนำโปรแกรมนี้ไปทดสอบร่วมกับโปรแกรมตรวจจับผู้บุกรุกสนอร์ต รวมทั้งคำแนะนำสำหรับการออกแบบการทดลองเพื่อใช้เป็นข้อมูลในบทต่อไปอีกด้วย

3.1 ศึกษาแนวทางการลดปริมาณข้อมูล

จากบทที่ 1 วิธีแก้ปัญหาการวิเคราะห์หาผู้บุกรุกสำหรับข้อมูลระบบเครือข่ายที่มีปริมาณมาก ที่ประหยัดทรัพยากรของระบบเครือข่าย คือการลดปริมาณข้อมูลด้วยวิธีการสุ่ม ซึ่งจากทฤษฎีเอสโฟล์วที่กล่าวในบทที่ 2 และงานวิจัยนี้เลือกการสุ่มด้วยวิธี Stratified random [3] [8] เนื่องจากการสุ่มวิธีนี้มีการแบ่งข้อมูลเป็นช่วงย่อยๆ และสุ่มข้อมูลออกจากช่วงย่อยๆ เหล่านั้น ทำให้มีการกระจายตัวของการสุ่ม นอกจากนี้ยังมีวิธีการสุ่มอื่นๆ อีกคือ Systematic และ Simple random [3] แต่ทั้ง 2 วิธีนี้มีผลกระทบต่อความเสียหายของข้อมูลเชิงเวลามากกว่าวิธี Stratified random และการกระจายตัวของข้อมูลที่ได้จากการสุ่มนั้นน้อยกว่าอีกด้วย ซึ่งกล่าวคือการสุ่มด้วยวิธี Systematic ที่มีการแบ่งข้อมูลเป็นช่วงๆ และดึงข้อมูลในช่วงแรกของแต่ละช่วงออกมา ซึ่งมีข้อเสียเมื่อมีข้อมูลการบุกรุกอยู่ในช่วงท้ายของแต่ละช่วง ทำให้ไม่สามารถตรวจสอบพบได้ ส่วนวิธี Simple random นั้นมีข้อเสียที่มีโอกาสที่จะสุ่มไม่พบการบุกรุกค่อนข้างสูง เนื่องจากไม่ได้มีการแบ่งเป็นช่วงๆ

ดังนั้นในงานวิจัยนี้จึงเลือกการสุ่มด้วยวิธี Stratified random มาใช้เป็นแนวคิดหลักในงานวิจัยนี้ และนำมาพัฒนาโปรแกรมสำหรับสุ่มแพ็กเก็ต ซึ่งจะกล่าวถึงรายละเอียดในการพัฒนาโปรแกรม รวมทั้งผลการทดสอบใช้งานโปรแกรมในงานวิจัยที่ 3.2 ต่อไป

3.2 การพัฒนาโปรแกรมสำหรับสุ่มแพ็กเก็ต

จากแนวคิดในการลดปริมาณข้อมูลด้วยวิธีการสุ่มแบบ Stratified random ในงานวิจัยนี้จึงได้พัฒนาโปรแกรมสำหรับใช้สุ่มแพ็กเก็ต ซึ่งมีรายละเอียดการพัฒนาดังต่อไปนี้

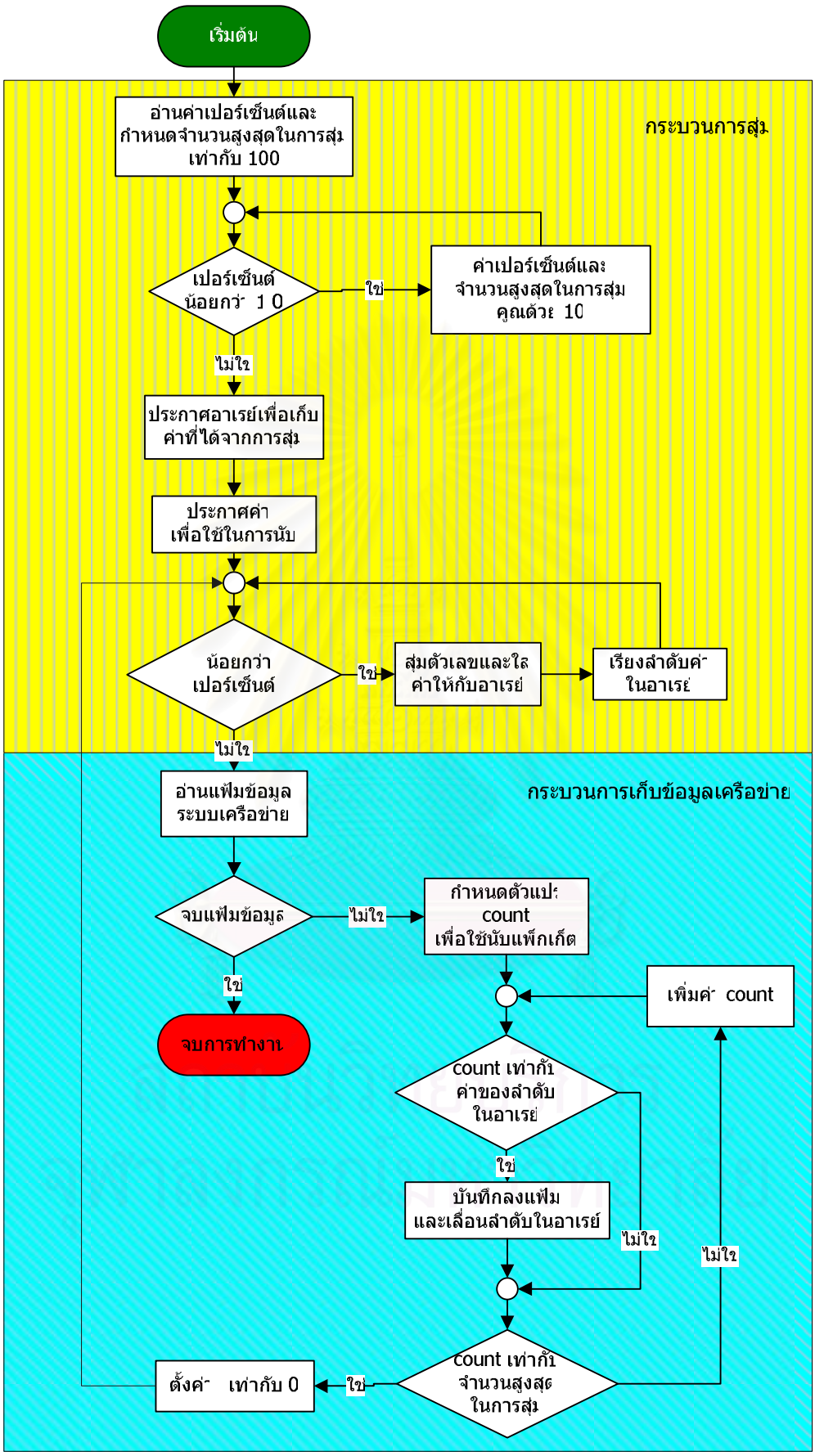
3.2.1 **เครื่องมือที่ใช้ในการพัฒนา**

- เครื่องคอมพิวเตอร์ที่ติดตั้งระบบปฏิบัติการลินุกซ์ Fedora Core 5 และชุดพัฒนาซอฟต์แวร์ภาษาซีของระบบปฏิบัติการ
- คลังโปรแกรม (library) สำหรับการพัฒนาโปรแกรมในการดักเก็บข้อมูลที่ชื่อ Libpcap

3.2.2 **กระบวนการทำงานของโปรแกรม**

จากรูปที่ 3.1 แสดงให้เห็นถึงลักษณะการทำงานของโปรแกรมโดยรวม ซึ่งแบ่งกระบวนการทำงานของโปรแกรมออกเป็นสองส่วนหลักๆ คือ

- กระบวนการสุ่ม เริ่มจากการรับค่าเปอร์เซ็นต์และกำหนดจำนวนในการสุ่ม (เปรียบเหมือนเป็นขนาดกลุ่มที่แบ่งเพื่อจะสุ่ม) โดยให้เท่ากับ 100 จากนั้นตรวจสอบดูว่าค่าเปอร์เซ็นต์ต่ำกว่า 1 หรือไม่ ถ้าต่ำกว่าจำเป็นต้องปรับให้ค่าเปอร์เซ็นต์นั้นเป็นจำนวนเต็มเสียก่อน แต่จำนวนสูงสุดในการสุ่มก็เพิ่มขึ้นไปด้วย เมื่อค่าเปอร์เซ็นต์เป็นเลขจำนวนเต็มแล้ว จึงประกาศอาเรย์สำหรับเก็บค่าที่ได้จากการสุ่ม จากนั้นทำการสุ่มไปเรื่อยๆ จนกว่าจะครบตามเปอร์เซ็นต์ที่ต้องการ ซึ่งในที่ทำการสุ่มนั้นก็จะทำการเรียงลำดับค่าไปด้วย และถ้าหากว่ามีค่าสุ่มที่ซ้ำกัน ทำการสุ่มใหม่ สุดท้ายจะได้ค่าการสุ่มที่ไม่ซ้ำกันและเรียงลำดับจากน้อยไปมาก
- กระบวนการเก็บข้อมูลเครือข่าย เนื่องจากโปรแกรมในเวอร์ชันนี้พัฒนาให้สุ่มแพ็กเก็ตจากแฟ้มข้อมูลระบบเครือข่ายที่เก็บไว้ก่อนแล้ว ดังนั้นในกระบวนการเก็บข้อมูลเครือข่ายจึงเริ่มด้วยการอ่านแฟ้มข้อมูลระบบเครือข่าย และกำหนดค่าตัวแปรสำหรับการนับแพ็กเก็ต เมื่อมีการนับแพ็กเก็ตตรงกับเลขที่ถูกสุ่มก็จะดึงแพ็กเก็ตดังกล่าวบันทึกลงแฟ้มข้อมูลใหม่ จากนั้นก็เลื่อนลำดับของอาเรย์ไป อ่านและเก็บข้อมูลไปเรื่อยๆ จนครบจำนวนสูงสุดในการสุ่ม จึงทำการสุ่มตัวเลขชุดใหม่



รูปที่ 3.1 กระบวนการทำงานของโปรแกรมสุ่มดักแพ็กเก็ต

3.2.3 ผลการทดสอบโปรแกรม

การทดสอบโดยการนำข้อมูลจำนวน 32,634,031 แพ็กเก็ต ทำการสุ่มทั้งหมด 10 ชุดเปอร์เซ็นต์ จำนวน 5 ครั้งต่อชุดเปอร์เซ็นต์ จากนั้นใช้โปรแกรมสนอร์ตนับจำนวนแพ็กเก็ตในชุดนั้นๆ แล้วนำมาหาค่าเฉลี่ย ซึ่งได้ผลดังตารางที่ 3.1

ตารางที่ 3.1 จำนวนแพ็กเก็ตที่ถูกสุ่มด้วยโปรแกรม

% การสุ่ม	จำนวนแพ็กเก็ต				
	สุ่มครั้งที่ 1	สุ่มครั้งที่ 2	สุ่มครั้งที่ 3	สุ่มครั้งที่ 4	สุ่มครั้งที่ 5
100	32,634,031	32,634,031	32,634,031	32,634,031	32,634,031
90	29,370,548	29,370,550	29,370,546	29,370,549	29,370,544
80	26,107,152	26,107,153	26,107,154	26,107,149	26,107,150
70	22,843,757	22,843,760	22,843,757	22,843,757	22,843,759
60	19,580,365	19,580,362	19,580,367	19,580,368	19,580,364
50	16,316,968	16,316,969	16,316,970	16,316,969	16,316,970
40	13,053,576	13,053,581	13,053,576	13,053,578	13,053,579
30	9,790,182	9,790,185	9,790,184	9,790,182	9,790,184
20	6,526,789	6,526,786	6,526,785	6,526,791	6,526,787
10	3,263,396	3,263,397	3,263,394	3,263,393	3,263,395

จากตารางที่ 3.1 แสดงให้เห็นว่าจำนวนแพ็กเก็ตโดยเฉลี่ยลดลงตามเปอร์เซ็นต์ที่สุ่มโดยที่ 100 เปอร์เซ็นต์ เป็นค่าตั้งต้นก่อนเริ่มทำการสุ่ม และในแต่ละช่วงเปอร์เซ็นต์ที่สุ่มนั้นมีจำนวนแพ็กเก็ตที่ถูกสุ่มแต่ละครั้งใกล้เคียงกัน การวิจัยต่อไปจะกล่าวถึงการนำข้อมูลที่ได้จากการสุ่มไปตรวจสอบหาการบุกรุกด้วยโปรแกรมสนอร์ต ซึ่งจะทดสอบกับการบุกรุกที่ถูกตรวจพบด้วยฟังก์ชันการวิเคราะห์พฤติกรรมของโปรแกรมสนอร์ต

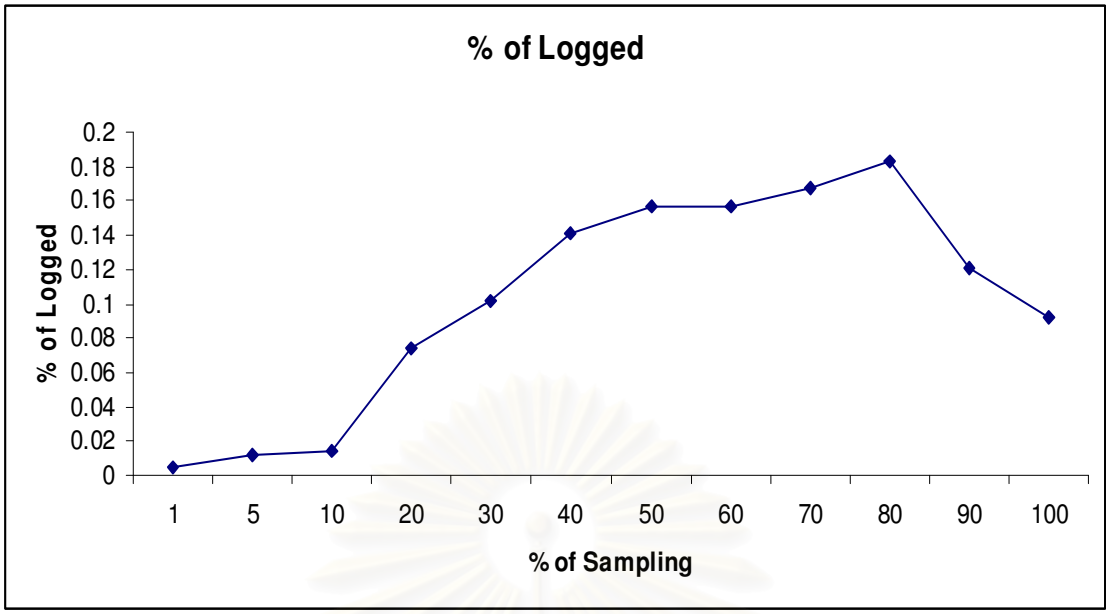
3.3 การหาจำนวนการบุกรุกที่ตรวจพบด้วยการวิเคราะห์พฤติกรรม

หลังจากพัฒนาโปรแกรมสำหรับสุ่มแพ็กเก็ตแล้วจึงนำมาใช้งานควบคู่กับการทำงานของโปรแกรมสนอร์ต ซึ่งในการทดลองครั้งแรกคือนำข้อมูลที่เก็บได้จากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จำนวน 13,822,523 แพ็กเก็ต พบว่ามีผลการแจ้งเตือนที่สูงขึ้นอย่างผิดปกติ ดังผลการทดลองในตารางที่ 3.2 และวาดกราฟได้ดังรูปที่ 3.2 จากนั้นจึงตรวจสอบการแจ้งเตือนพบว่ามีการบุกรุกแบบ Open Port เท่านั้นที่เพิ่มขึ้นอย่างผิดปกติ ได้ผล

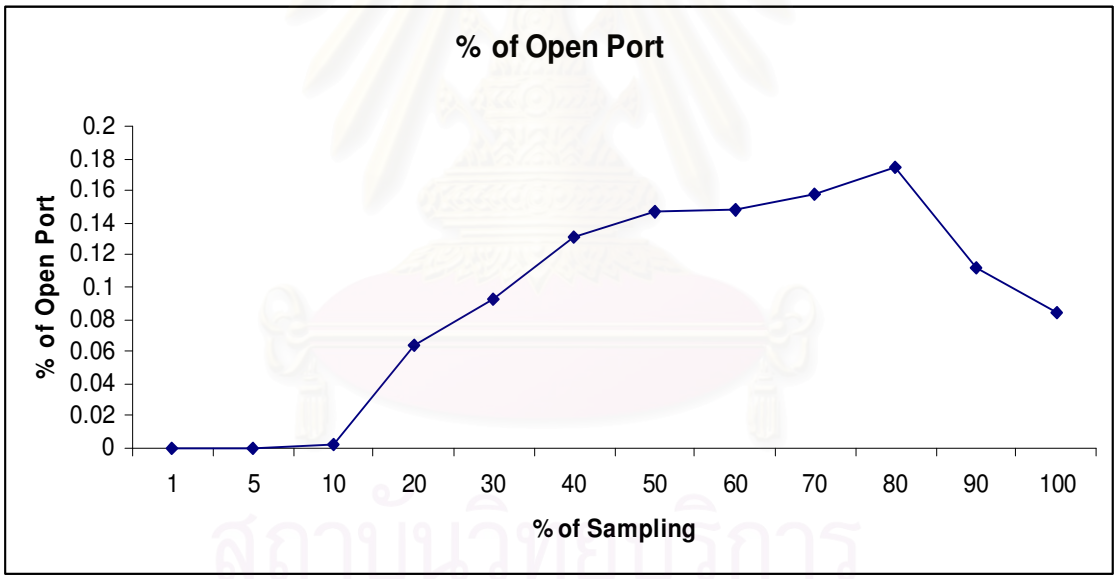
การทดลองดังรูปที่ 3.3 ดังนั้นจึงตั้งสมมติฐานว่าข้อมูลระบบเครือข่ายที่เก็บมานั้นมีความเสียหายหรือไม่สมบูรณ์ ทำให้มีผลต่อการแจ้งเตือนที่สูงผิดปกติ จึงทำการทดลองซ้ำอีกครั้ง โดยมีการเก็บข้อมูลจากระบบเครือข่ายแห่งหนึ่งที่มีจำนวนผู้ใช้งานประมาณ 200-300 คน จำนวน 4,695,061 แพ็กเก็ต และเพื่อเพิ่มความแม่นยำในการตรวจสอบจึงเพิ่มจำนวนการสุ่มเป็น 5 ครั้งต่อเปอร์เซ็นต์การสุ่ม แล้วนำมาหาค่าเฉลี่ย ซึ่งได้ผลดังตารางที่ 3.3 และนำไปวาดกราฟดังรูปที่ 3.4

ตารางที่ 3.2 ผลการทดลองหาเปอร์เซ็นต์การแจ้งเตือนและการพบ Open Port ของข้อมูลจากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

% การสุ่ม	จำนวนแพ็กเก็ต	จำนวนการแจ้งเตือน	% การแจ้งเตือน	จำนวนการพบ Open Port	% การพบ Open Port
1	138225	6	0.004341	0	0
5	691125	86	0.012443	0	0
10	1382255	195	0.014107	41	0.002966168
20	2764506	2050	0.074154	1771	0.064062078
30	4146729	4240	0.102249	3835	0.092482533
40	5528974	7788	0.140858	7256	0.131235922
50	6911212	10810	0.156413	10151	0.146877277
60	8293454	13060	0.157474	12295	0.148249451
70	9675704	16217	0.167605	15314	0.158272721
80	11057947	20232	0.182963	19290	0.174444678
90	12440190	14987	0.120472	13923	0.111919512
100	13822523	12760	0.092313	11584	0.08380525



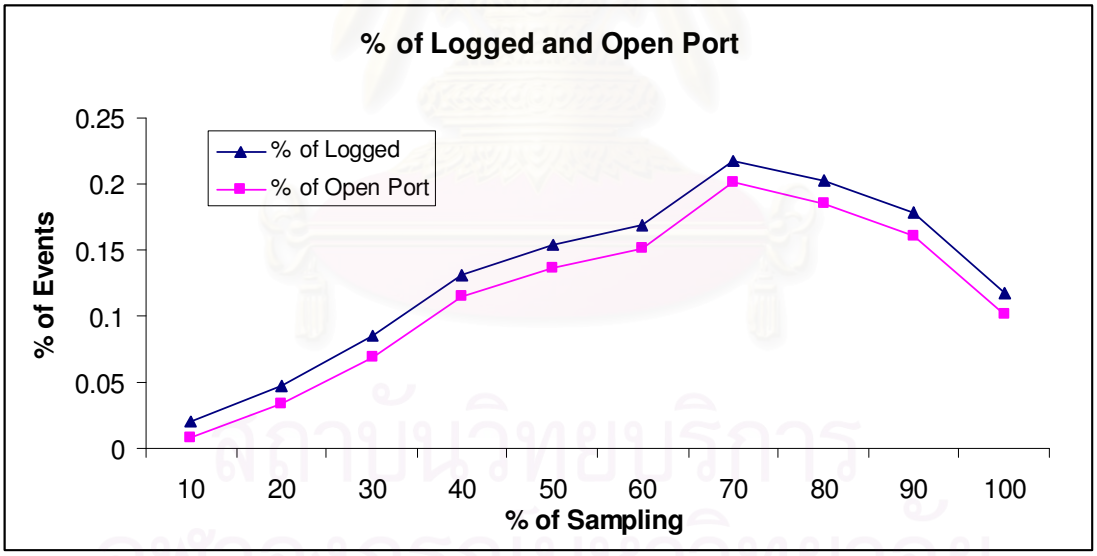
รูปที่ 3.2 ความสัมพันธ์ระหว่างการแจ้งเตือนและระดับการสุ่ม
โดยใช้ข้อมูลที่เก็บจากภาควิชาฯ



รูปที่ 3.3 ความสัมพันธ์ระหว่างการพบ Open Port และระดับการสุ่ม
โดยใช้ข้อมูลที่เก็บจากภาควิชาฯ

ตารางที่ 3.3 ผลการทดลองหาเปอร์เซ็นต์การแจ้งเตือนและการพบ Open Port ของข้อมูลจากระบบเครือข่ายแห่งหนึ่ง

% การสุ่ม	จำนวนแพ็กเก็ต	จำนวนการแจ้งเตือน	% การแจ้งเตือน	จำนวนการพบ Open Port	% การพบ Open Port
10	469,499.8	98.2	0.020916	37.4	0.007966
20	938,993.75	438.5	0.046699	317	0.03376
30	1,408,491.2	1192	0.08463	968.2	0.06874
40	1,877,989.2	2459	0.130938	2159.6	0.114995
50	2,347,484.2	3604	0.153526	3205.8	0.136563
60	2,816,982.8	4752	0.168691	4279.4	0.151914
70	3,286,480.6	7149	0.217528	6608.6	0.201084
80	3,755,976.2	7605	0.202477	6944.2	0.184884
90	4,225,474	7515	0.17785	6801.4	0.160962
100	4,695,061	5527	0.117719	4737	0.100893



รูปที่ 3.4 ความสัมพันธ์ระหว่างการแจ้งเตือนและการพบ Open Port ที่ระดับการสุ่มต่างๆ โดยใช้ข้อมูลจากระบบเครือข่ายแห่งหนึ่ง

จากการทดลองทั้งสองพบว่าเปอร์เซ็นต์การพบการบุกรุกประเภท Open Port ที่ทำให้จำนวนการแจ้งเตือนโดยรวมนั้นสูงขึ้นในช่วงเปอร์เซ็นต์การสุ่มที่ 100 ถึง 70 จากนั้นจะลดลงเรื่อยๆ ดังข้อมูลในตารางที่ 3.2 และ 3.3 ซึ่งนำมาวาดกราฟได้ดังรูปที่ 3.2 ถึง 3.4 ซึ่งในงานวิจัยนี้ทำให้สามารถสรุปได้ว่า การสุ่มนี้ไม่สามารถใช้ในการหาจำนวนการบุกรุกแบบ Open Port ได้

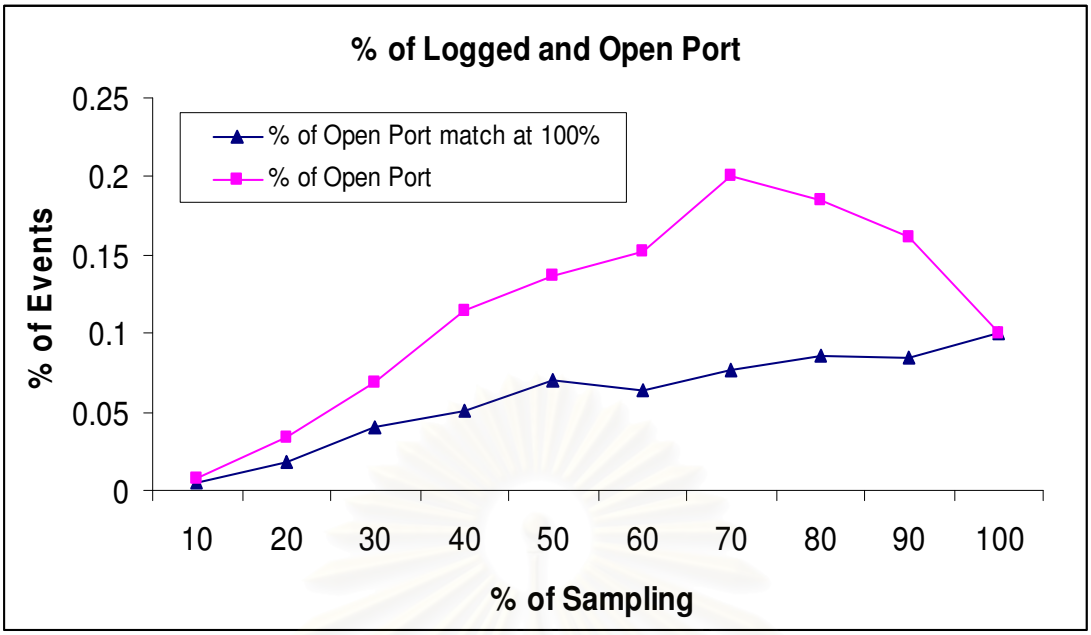
เนื่องจากเปอร์เซ็นต์การพบการบุกรุกนั้นไม่คงที่ นอกจากนี้เพื่อเป็นการยืนยันว่าผลการตรวจพบที่ผิดปกตินี้เกิดขึ้นกับทุกๆ ชุดข้อมูล จึงทำการทดลองเพิ่มโดยใช้ข้อมูลสำหรับการทดสอบระบบตรวจจับผู้บุกรุกขององค์กร DARPA [18] ซึ่งได้จัดหาข้อมูลเหล่านี้ไว้สำหรับให้ดาวน์โหลดไปใช้ทดลองพบว่าผลการตรวจสอบพบมีลักษณะคล้ายกับกราฟรูปที่ 3.2 ถึง 3.4 อีกด้วย และในงานวิจัยที่ 3.4 จะหาจำนวนการพบ Open Port ของชุดการสุ่มที่เปอร์เซ็นต์ต่างๆ ที่เหมือนกันกับ Open Port ที่ 100 เปอร์เซ็นต์

3.4 การหาจำนวนการแจ้งเตือนที่เหมือนกันกับ 100 เปอร์เซ็นต์

จากผลการวิจัยที่ 3.3 สามารถตั้งสมมติฐานเพิ่มเติมว่า เปอร์เซ็นต์การพบ Open Port ของชุดการสุ่มที่เปอร์เซ็นต์ต่างๆ ที่เหมือนกับเปอร์เซ็นต์การพบ Open Port ที่ 100 เปอร์เซ็นต์ นั้นอาจมีจำนวนคงที่ ในงานวิจัยนี้จึงนำข้อมูลที่เก็บได้จากระบบเครือข่ายแห่งหนึ่ง ที่มีจำนวนผู้ใช้งานประมาณ 200-300 คน จำนวน 4,695,061 แพ็กเก็ต มาวิเคราะห์หาจำนวน Open Port ที่ซ้ำกันเมื่อเปรียบเทียบกับ 100 เปอร์เซ็นต์ ดังตารางที่ 3.4 และนำค่าในตารางมาวาดกราฟได้ดังรูปที่ 3.5

ตารางที่ 3.4 ผลการตรวจพบ Open Port ที่ใช้แพ็กเก็ตตรงกับข้อมูลที่ไม่ได้สุ่ม

% การสุ่ม	จำนวนแพ็กเก็ต	จำนวนการพบ Open Port	% การพบ Open Port	จำนวนการพบ Open Port ที่ตรงกับ 100 %	% การพบ Open Port ที่ตรงกับ 100 %
10	469,499.8	37.4	0.007966	23.6	0.005027
20	938,993.75	317	0.03376	172.5	0.018371
30	1,408,491.2	968.2	0.06874	560.4	0.039787
40	1,877,989.2	2159.6	0.114995	965.8	0.051427
50	2,347,484.2	3205.8	0.136563	1651	0.070331
60	2,816,982.8	4279.4	0.151914	1797.2	0.063799
70	3,286,480.6	6608.6	0.201084	2529.2	0.076958
80	3,755,976.2	6944.2	0.184884	3228.8	0.085964
90	4,225,474	6801.4	0.160962	3599.4	0.085183
100	4,695,061	4737	0.100893	4737	0.100893



รูปที่ 3.5 ความสัมพันธ์ระหว่างการพบ Open Port ที่เหมือนกับข้อมูลที่ไม่ได้สุ่มเทียบกับเปอร์เซ็นต์การพบ Open Port

จากรูปที่ 3.5 พบว่าจำนวนการบุกรุกแบบ Open Port ที่เหมือนกับการตรวจสอบที่ 100 เปอร์เซ็นต์ นั้นลดลงอย่างต่อเนื่อง กล่าวคือมีจำนวน Open Port ที่เพิ่มขึ้นมาใหม่เป็นจำนวนมากเมื่อมีการสุ่มข้อมูล

จากงานวิจัยที่ 3.3 และ 3.4 แสดงให้เห็นว่าจำนวนเปอร์เซ็นต์การตรวจพบการบุกรุกประเภทการสแกนพอร์ตแบบ Open Port นั้นไม่คงที่ ซึ่งการบุกรุกประเภทนี้ต้องใช้ฟังก์ชันการตรวจสอบการบุกรุกจากพฤติกรรมแพ็กเก็ตของโปรแกรมสนอร์ต ดังนั้นจึงสรุปได้ว่าการสุ่มนั้นไม่เหมาะที่จะนำมาใช้ตรวจสอบหาจำนวนการบุกรุกประเภทการสแกนพอร์ตด้วยวิธีการวิเคราะห์จากพฤติกรรมของแพ็กเก็ต นอกจากนี้ยังตั้งสมมติฐานต่อไปว่า การสุ่มข้อมูลแพ็กเก็ตนั้นยังคงทำให้สามารถตรวจพบการบุกรุกประเภทการสแกนพอร์ตด้วยฟังก์ชันการเปรียบเทียบชุดกฎ ซึ่งจะกล่าวถึงรายละเอียดในงานวิจัยที่ 3.5 ต่อไป

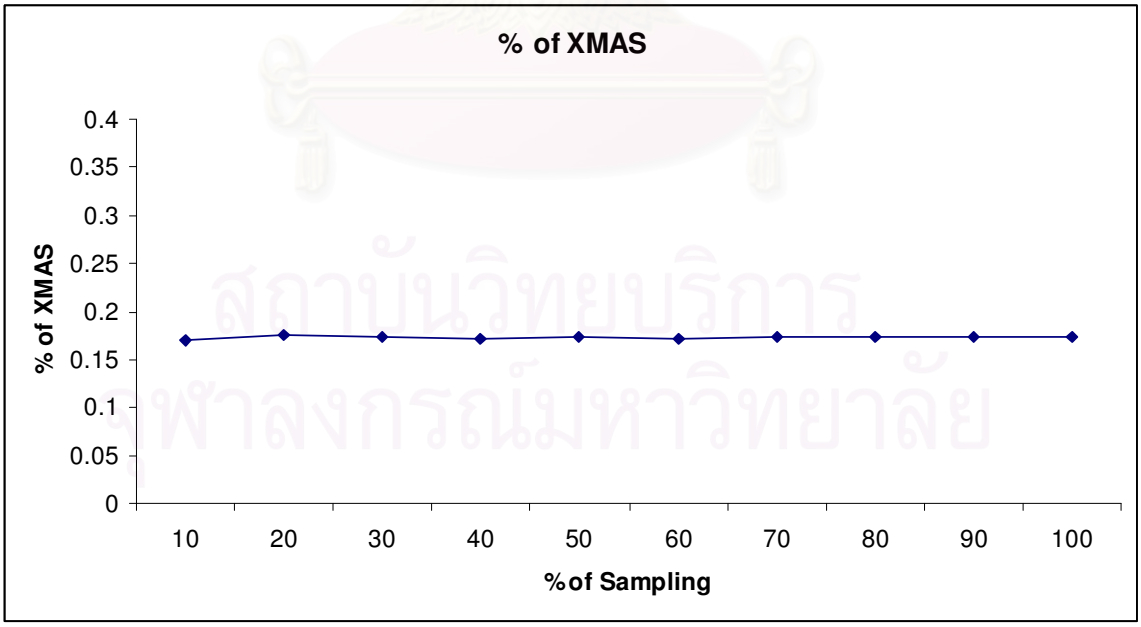
3.5 การหาจำนวนการบุกรุกที่ตรวจพบด้วยการเปรียบเทียบกับชุดกฎ

จากการวิจัยที่ 3.3 และ 3.4 ทำให้สรุปได้ว่าการสุ่มนั้นมีผลกระทบต่อการตรวจจับการบุกรุกที่ถูกตรวจสอบด้วยการวิเคราะห์พฤติกรรม ดังนั้นในการวิจัยนี้จึงวิเคราะห์ผู้บุกรุกด้วยการเปรียบเทียบกับชุดกฎ ด้วยสมมติฐานที่ว่าการสุ่มนั้นไม่มีผลกระทบต่อการตรวจจับผู้บุกรุกด้วยการเปรียบเทียบชุดกฎ

วิธีการทดลองนั้นได้ทำการดักข้อมูลพร้อมกับทำการสแกนพอร์ต ด้วยโปรแกรมสำหรับสแกนพอร์ตชื่อ nmap ซึ่งในงานวิจัยนี้ได้ลองสแกนพอร์ตด้วยวิธี XMAS scan และมีข้อมูลจำนวนทั้งสิ้น 977,028 แพ็กเก็ต จากนั้นจึงทำการสุ่มชุดเปอร์เซ็นต์ละ 30 ครั้ง เพื่อให้ได้ผลแม่นยำมากยิ่งขึ้น แล้วจึงนำมาหาค่าเฉลี่ยของเปอร์เซ็นต์การพบการบุกรุก ได้ผลดังตารางที่ 3.5

ตารางที่ 3.5 ผลการหาค่าเปอร์เซ็นต์การตรวจพบการสแกนพอร์ตแบบ XMAS

% การสุ่ม	% การพบ XMAS
10	0.170163
20	0.174486
30	0.173026
40	0.170874
50	0.173945
60	0.171831
70	0.173923
80	0.172651
90	0.173117
100	0.173178



รูปที่ 3.6 ความสัมพันธ์ระหว่างการตรวจพบ XMAS และระดับการสุ่ม

ในงานวิจัยนี้แสดงให้เห็นว่าเปอร์เซ็นต์การตรวจพบการบุกรุกประเภทสแกนพอร์ตแบบ XMAS นั้นค่อนข้างคงที่ ในช่วงเปอร์เซ็นต์การสุ่มต่างๆ ดังรูปที่ 3.6 ดังนั้นจึงสรุปได้ว่าการสุ่มข้อมูลนั้นไม่มีผลกระทบต่อการตรวจจับการบุกรุกด้วยการเปรียบเทียบกับชุดกฎ ที่งานวิจัยนี้ใช้การบุกรุกประเภทการสแกนพอร์ตแบบ XMAS scan แทนการบุกรุกที่ถูกตรวจสอบด้วยการเปรียบเทียบกับชุดกฎของโปรแกรมสนอร์ต

จากผลการวิจัยที่ผ่านมาในข้างต้นจึงนำสรุปเป็นข้อแนะนำสำหรับการออกแบบการทดลองซึ่งจะกล่าวต่อไปในงานวิจัยที่ 3.6 และจะนำไปใช้ในการออกแบบการทดลองที่จะกล่าวถึงต่อไปในบทที่ 4 ด้วย

3.6 คำแนะนำสำหรับการออกแบบการทดลอง

จากงานวิจัยที่ผ่านมานั้นแสดงสรุปได้เบื้องต้นว่า ระบบตรวจจับผู้บุกรุกนั้นมีการทำงานสองแบบ คือการวิเคราะห์จากพฤติกรรม และการตรวจสอบจากการเปรียบเทียบด้วยชุดกฎ ซึ่งการสุ่มนั้นมีผลกระทบการตรวจจับการบุกรุกด้วยวิธีการวิเคราะห์พฤติกรรมของแพ็กเก็ต ทำให้ไม่สามารถใช้งานร่วมกันได้ ดังผลการวิจัยที่ 3.3 และ 3.4 แต่การสุ่มนั้นสามารถใช้งานร่วมกับการตรวจจับการบุกรุกด้วยวิธีการเปรียบเทียบชุดกฎ ดังผลการวิจัยที่ 3.5

อย่างไรก็ตามงานวิจัยที่ผ่านมานั้น ไม่ได้มีการออกแบบการทดลองที่ชัดเจน ทำให้บางครั้งการควบคุมตัวแปรต่างๆ ก็อาจมีผลต่อผลการวิจัย ดังนั้นงานวิจัยนี้จะกล่าวถึงคำแนะนำสำหรับทำการทดลองดังต่อไปนี้

3.6.1 วิธีการคำนวนหาเปอร์เซ็นต์การตรวจพบการบุกรุก

ดังสมการที่ 3.1

$$p = (\frac{D}{A}) \times 100 \qquad (3.1)$$

เมื่อ p แทนค่าเปอร์เซ็นต์การตรวจพบการบุกรุก

D แทนจำนวนการบุกรุกที่ตรวจพบ

A แทนจำนวนแพ็กเก็ตทั้งหมด

3.6.2 ข้อมูลสำหรับการทดลอง

จากข้อสมมติฐานเกี่ยวกับข้อมูลที่อาจเกิดความเสียหายหรือไม่สมบูรณ์ตั้งแต่ในขั้นตอนการเก็บข้อมูลนั้น จึงจำเป็นที่จะต้องควบคุมการจัดเก็บข้อมูลด้วย ซึ่งข้อมูลนั้นจะประกอบ

ไปด้วยข้อมูลที่จำลองขึ้นและข้อมูลจริงที่เก็บ และจำเป็นต้องเลือกระบบเครือข่ายที่เหมาะสม กล่าวคือมีจำนวนผู้ใช้งานเป็นจำนวนมากกว่า 100 คนขึ้นไป เก็บในปริมาณที่ค่อนข้างมาก และเวลาที่เก็บนั้นประมาณ 1 วันขึ้นไป

3.6.3 จำนวนครั้งในการสุ่ม

เนื่องจากการสุ่มนั้นอาจจะทำให้ผลการทดลองมีค่าไม่แน่นอน ดังนั้นหากทำการสุ่มจำนวนน้อยครั้งเกินไปอาจจะทำให้ได้ผลการทดลองที่คลาดเคลื่อนสูง ดังนั้นจึงจำเป็นต้องเพิ่มจำนวนครั้งในการสุ่มเพื่อทำให้ค่าที่ได้นั้นเข้าสู่ค่าที่ถูกต้องมากที่สุด ซึ่งจะกำหนดไว้ที่ 30 ครั้งต่อชุดเปอร์เซ็นต์

ในบทนี้ได้กล่าวถึงแนวทางการวิจัย รวมทั้งการทดสอบเบื้องต้นเพื่อหาเปอร์เซ็นต์การตรวจสอบพบการบุกรุก และการออกแบบการทดลอง ซึ่งจะกล่าวถึงรายละเอียดการทดลองและผลการทดลองในบทต่อไป

บทที่ 4
การทดลอง

บทที่ผ่านมานั้นแสดงว่าการสุ่มนั้นมีประโยชน์ในการลดปริมาณข้อมูลเพื่อใช้ในการวิเคราะห์หาการบุกรุก ทำให้ความถูกต้องแม่นยำของการตรวจพบการบุกรุกจึงเป็นอีกปัจจัยหนึ่งที่มีความสำคัญมาก ดังนั้นจึงทำให้เกิดปัญหาที่ว่า เมื่อทำการสุ่ม ณ เปอร์เซ็นต์ต่างๆ นั้นมีผลต่อการตรวจพบการบุกรุกประเภทการสแกนพอร์ตของโปรแกรมตรวจจับผู้บุกรุกหรือไม่

โดยบทนี้กล่าวถึงวัตถุประสงค์ของการทดลอง เครื่องมือ และข้อมูลสำหรับการทดลอง และทำการทดลองเพื่อหาผลกระทบของการสุ่มต่อการตรวจจับการบุกรุก รวมทั้งผลการทดลองกับข้อมูลต่างๆ ที่เลือกนำมาทดลอง โดยมีรายละเอียดของแต่ละขั้นตอนดังนี้

4.1 วัตถุประสงค์การทดลอง

เพื่อหาว่าการสุ่มข้อมูลออกนั้นมีผลต่อการตรวจจับการบุกรุกประเภทสแกนพอร์ตหรือไม่ และมีผลต่อเฉพาะข้อมูลที่ถูกสร้างขึ้นหรือข้อมูลที่ดักได้จริงหรือไม่ ซึ่งฟังก์ชันในการตรวจจับการบุกรุกประเภทสแกนพอร์ตของโปรแกรมสนอร์ตนั้นมีสองฟังก์ชันคือ

4.1.1 การตรวจจับด้วยการวิเคราะห์พฤติกรรม เป็นฟังก์ชันที่โปรแกรมสนอร์ตตรวจจับการบุกรุกได้โดยการวิเคราะห์พฤติกรรมของข้อมูล อาจใช้จำนวนแพ็กเก็ตที่เหมือนกันหรือขนาดของแพ็กเก็ต เป็นต้น ซึ่งมีการปรับแต่งค่าได้โดยการตั้งค่าขีดแบ่งขั้นต่ำ

4.1.2 การตรวจจับด้วยการเปรียบเทียบกับชุดกฎ เป็นฟังก์ชันที่โปรแกรมสนอร์ตจะอาศัยชุดกฎในการเปรียบเทียบว่าแพ็กเก็ตใดเป็นแพ็กเก็ตการบุกรุก ซึ่งชุดกฎเหล่านี้สามารถดาวน์โหลดหรือพัฒนาเพิ่มเติมได้

4.2 เครื่องมือที่ใช้ในการทดลอง

เครื่องมือที่ใช้ในการทดลองนั้น ใช้ซอฟต์แวร์ประเภท open source แล้วพัฒนาซอฟต์แวร์เพื่อเสริมอีกส่วนหนึ่งขึ้นเอง รายละเอียดของซอฟต์แวร์และเครื่องมือต่างๆ มีดังต่อไปนี้

4.2.1 โปรแกรมสำหรับทำการสุ่ม เป็นโปรแกรมที่พัฒนาขึ้นในงานวิจัยนี้ โดยใช้ภาษาซี ที่ทำงานร่วมกับคลัง Libpcap บนระบบปฏิบัติการลินุกซ์ เพื่อใช้ในการสุ่มและเก็บแพ็กเก็ตตามเปอร์เซ็นต์ที่ป้อนให้โปรแกรม

4.2.2 โปรแกรมตรวจจับผู้บุกรุก ใช้โปรแกรมสนอร์ต เวอร์ชั่น 2.6.1 (Build 24) และมีกฎทั้งหมด 5187 ข้อ โดยที่เพิ่มกฎพิเศษจาก Bleeding Edge Snort [15] จำนวน 8 ข้อ

สำหรับการตรวจจับการบุกรุกประเภทการสแกนพอร์ต เพื่อเพิ่มประสิทธิภาพให้โปรแกรมสนอร์ตนั้นสามารถตรวจจับการบุกรุกแบบสแกนพอร์ตได้แม่นยำยิ่งขึ้น

4.2.3 โปรแกรมสำหรับดักจับข้อมูล ใช้โปรแกรม TCPdump [13] ที่ทำงานบนระบบปฏิบัติการลินุกซ์ ดักข้อมูลทราฟฟิกเพื่อใช้เป็นข้อมูลตั้งต้นในการทดลอง

4.2.4 โปรแกรมสำหรับการสแกนพอร์ต ใช้โปรแกรม nmap [14] ที่ทำงานบนระบบปฏิบัติการลินุกซ์ เพื่อใช้สร้างแพ็กเก็ตสำหรับจำลองการบุกรุกประเภทการสแกนพอร์ต

4.2.5 โปรแกรมสำหรับเปรียบเทียบค่า เป็นโปรแกรมที่ถูกพัฒนาขึ้นเองด้วยภาษาซี ทำงานบนระบบปฏิบัติการวินโดวส์ เพื่อใช้ค้นหาคำที่อยู่ในแฟ้มลงบันทึกเข้าออก (log file) และนำมาเปรียบเทียบกับไฟล์ตั้งต้น ว่าแพ็กเก็ตที่มีการแจ้งเตือนนั้นเป็นแพ็กเก็ตเดียวกันเท่าไร

4.2.6 ระบบเครือข่าย และสถานที่ทำการทดลอง ซึ่งได้รับความอนุเคราะห์จากหน่วยปฏิบัติการวิจัยเทคโนโลยีและนวัตกรรมเพื่อความมั่นคงของประเทศ (NSTI) ภายใต้ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) สังกัดกระทรวงวิทยาศาสตร์และเทคโนโลยี

4.2.7 อุปกรณ์คอมพิวเตอร์ ใช้เครื่องคอมพิวเตอร์ส่วนบุคคลจำนวนสองเครื่อง ที่ติดตั้งระบบปฏิบัติการวินโดวส์ สำหรับวิเคราะห์ผลการตรวจสอบ และอีกเครื่องนั้นติดตั้งระบบปฏิบัติการลินุกซ์ สำหรับเป็นเครื่องแม่ข่าย พัฒนาโปรแกรม ดักเก็บข้อมูล และทำการทดสอบ

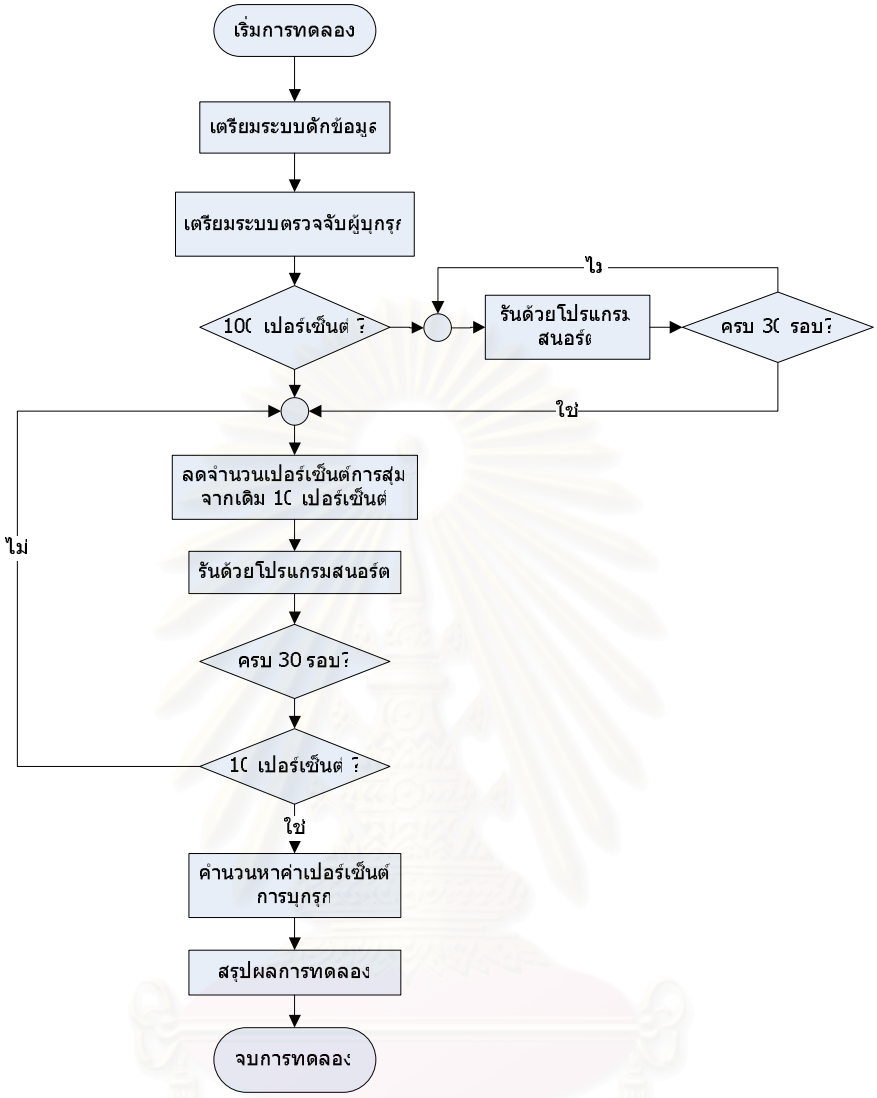
4.2.8 ข้อมูลในการตรวจสอบ ได้จากการดักจับข้อมูลจริง จากระบบเครือข่ายย่อยที่มีจำนวนผู้ใช้งานประมาณ 200-300 คน

4.3 **ข้อมูลที่ใช้ในการทดลอง**

4.3.1 ข้อมูลที่จำลองขึ้น เป็นข้อมูลที่สร้างด้วยโปรแกรม ซึ่งในการทดลองนี้จะใช้โปรแกรม nmap สำหรับการสร้างแพ็กเก็ตที่เกี่ยวข้องกับการสแกนพอร์ต

4.3.2 ข้อมูลจริง เป็นข้อมูลที่เก็บจากระบบเครือข่ายตามหัวข้อ 4.2.6 โดยใช้โปรแกรม TCPdump ทำการดักจับแพ็กเก็ตที่เครื่องแม่ข่าย

4.4 ขั้นตอนการทดลองและคำนวนผล



รูปที่ 4.1 กระบวนการทดลอง

จากรูปที่ 4.1 รายละเอียดกระบวนการทดลองมีดังต่อไปนี้

4.4.1 เตรียมระบบสำหรับทำการดักจับข้อมูล

4.4.2 เตรียมระบบสำหรับการตรวจจับผู้บุกรุก

4.4.3 นำข้อมูลทั้งหมดมาทำการสุ่มที่ระดับค่าเปอร์เซ็นต์ต่างๆ ตั้งแต่ 100 -10 เปอร์เซ็นต์ โดยแต่ละค่าจะลดลงครั้งละ 10 เปอร์เซ็นต์

4.4.4 จากนั้นนำข้อมูลที่สุ่มได้ในแต่ละชุดไปตรวจสอบหาจำนวนการบุกรุกด้วยโปรแกรม Snort ซึ่งจะทดสอบทั้งหมด 30 ครั้งต่อ 1 ชุดข้อมูล

4.4.5 นำจำนวนการบุกรุกที่ได้ ไปผ่านโปรแกรมเพื่อหาจำนวนการสแกนพอร์ตแต่ละแบบที่ต้องการ และเปรียบเทียบกับผลการตรวจสอบที่ 100 เปอร์เซ็นต์ ว่าการบุกรุกที่พบนั้นตรงกันกี่แพ็กเก็ต (เนื่องจากกำหนดให้ที่ 100 เปอร์เซ็นต์พบการบุกรุกถูกต้อง)

4.4.6 นำจำนวนการบุกรุกประเภทสแกนพอร์ตแบบต่างๆ ที่ได้จากข้อ 5 มาหารด้วยจำนวนแพ็กเก็ตทั้งหมดในแต่ละชุดเปอร์เซ็นต์ แล้วคูณด้วย 100 ดังสมการที่ 4.1

$$p = (\frac{D}{A}) \times 100 \qquad (4.1)$$

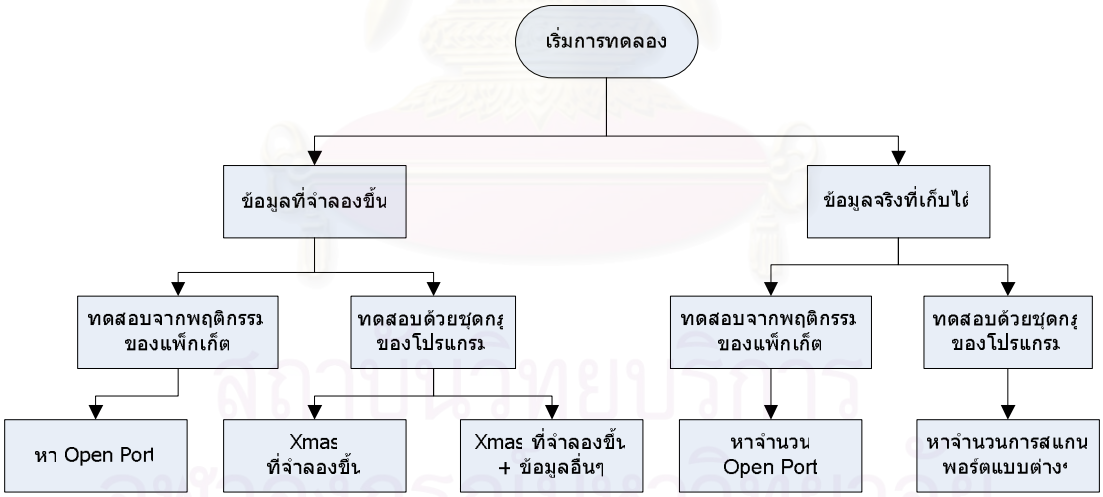
เมื่อ p แทนค่าเปอร์เซ็นต์การตรวจพบการบุกรุก

D แทนจำนวนการบุกรุกที่ตรวจพบ

A แทนจำนวนแพ็กเก็ตทั้งหมด

4.4.7 ทำการหาค่าเฉลี่ยเปอร์เซ็นต์การตรวจพบการบุกรุกประเภทสแกนพอร์ตในแต่ละชุดข้อมูล แล้วจึงสร้างกราฟ และสรุปผล

โดยการทดลองกับกลุ่มข้อมูลจำลองและข้อมูลจริง มีโครงสร้างลำดับ ดังแสดงด้วยแผนภูมิในรูปที่ 4.1



รูปที่ 4.1 แผนภูมิลำดับการทดลอง

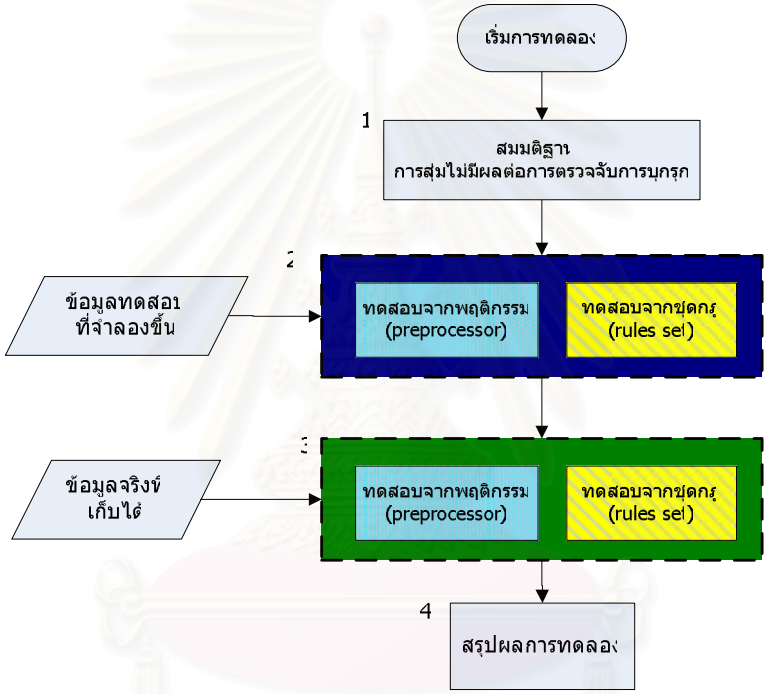
จากรูปที่ 4.1 การทดลองเริ่มด้วยการสร้างชุดข้อมูลจำลองเพื่อทดสอบเบื้องต้นว่า การสุ่มข้อมูลมีผลต่อการตรวจจับการบุกรุกประเภทสแกนพอร์ตหรือไม่ โดยแบ่งการทดสอบทั้ง preprocessor โดยใช้ Open Port เป็นตัวแทนข้อมูลการบุกรุกที่ตรวจสอบได้ด้วยฟังก์ชันนี้ และชุดกฎ โดยใช้ Xmas ที่สร้างขึ้นเป็นตัวอย่างแทนการบุกรุกที่จะถูกตรวจสอบพบได้ด้วยชุดกฎของโปรแกรม หลังจากนั้นจึงนำข้อมูลจริงที่เก็บได้มาทดสอบ เพื่อหาจำนวนการบุกรุกที่เกิดขึ้น ซึ่งใน

ข้อมูลจริงนั้นทำการสแกนพอร์ตในขณะดักจับข้อมูลด้วย เพื่อให้เสมือนว่ามีการบุกรุกเกิดขึ้นจริง ณ ขณะเวลานั้น

ในบทนี้แสดงให้เห็นถึงการเตรียมการทดลอง ซึ่งครอบคลุมวัตถุประสงค์ เครื่องมือที่ใช้ในการทดลอง ข้อมูลที่จะนำมาใช้ และขั้นตอนการทดลอง ซึ่งในบทต่อไปจะกล่าวถึงผลที่ได้จากการทดลอง และสรุปผลการทดลอง ว่าเป็นไปตามวัตถุประสงค์ของการทดลองหรือไม่

บทที่ 5
ผลการทดลองและการสรุปผล

บทที่ผ่านมาเป็นขั้นตอนเตรียมการทดลอง ทั้งการเตรียมเครื่องมือต่างๆ ข้อมูลสำหรับการทดลอง รวมทั้งขั้นตอนของการทดลอง ในบทนี้จะนำข้อมูลที่เตรียมไว้มาทำการทดลอง ตามประเภทของข้อมูล โดยที่จะนำเสนอผลการทดลองนั้นจะเป็นไปตามกระบวนการทดลองในภาพรวมตามรูปที่ 5.1 ดังที่จะกล่าวถึงต่อไป โดยจะได้สรุปผลการทดลองทั้งหมดที่ได้ในหัวข้อ 5.4 ด้านท้ายบท



รูปที่ 5.1 ภาพรวมของกระบวนการทดลอง

5.1 การตั้งสมมติฐาน

สมมติฐานคือ การสุ่มข้อมูลนั้นไม่มีผลต่อการตรวจจับการบุกรุกประเภทสแกนพอร์ต และมีผลต่อเฉพาะข้อมูลที่ถูกจำลองขึ้นหรือต่อเฉพาะข้อมูลที่ดักได้จริงๆ

5.2 การทดลองด้วยข้อมูลจำลอง

การทดลองชุดนี้ต้องการแสดงให้เห็นว่าการสุ่มนั้นไม่มีผลต่อการตรวจพบการบุกรุกประเภทการสแกนพอร์ต และข้อมูลที่นำมาใช้ทดสอบนั้นอาศัยฐานจากข้อมูลที่เกิดจากการดักจับจริงโดยเสริมด้วยข้อมูลที่สร้างขึ้นเองด้วยโปรแกรม nmap [14] ในการสร้างแพ็กเก็ต (packet generator)

ข้อมูลสำหรับการทดลองในชุดนี้ประกอบด้วย

- ข้อมูลสำหรับการตรวจจับด้วยการวิเคราะห์พฤติกรรมจำนวน 32,634,031 แพ็กเก็ต ขนาดของชุดข้อมูลประมาณ 3.03 GB ซึ่งใช้ในขั้นตอนที่ 5.2.1
- ข้อมูลสำหรับการตรวจจับด้วยการเปรียบกฎ แบ่งเป็นสองชุดย่อยๆ คือ
  - ข้อมูลจำนวน 3,352 แพ็กเก็ต ซึ่งมีแพ็กเก็ตของการบุกรุกประเภทการสแกนพอร์ตแบบ XMAS จำนวน 1,692 แพ็กเก็ต สำหรับการทดลองขั้นตอนที่ 5.2.2
  - ข้อมูลจำนวน 6,753,458 แพ็กเก็ต ซึ่งมีแพ็กเก็ตของการทดลองขั้นตอนที่ 5.2.2 รวมอยู่ด้วย สำหรับการทดลองขั้นตอนที่ 5.2.3

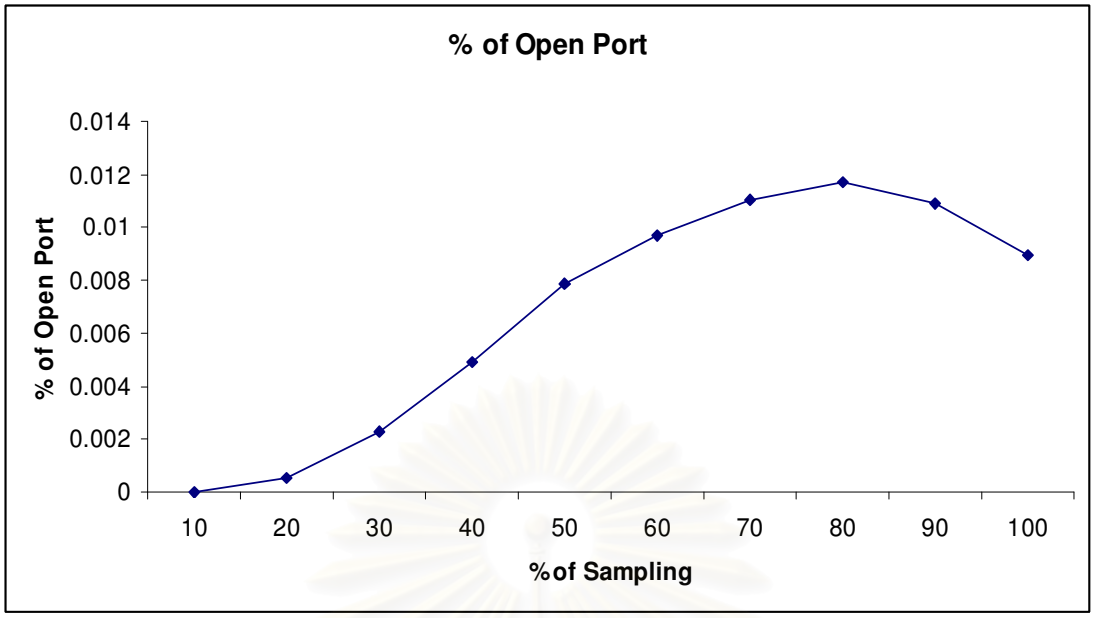
5.2.1 การทดลองสุ่มเพื่อหาค่าเปอร์เซ็นต์การตรวจพบ Open Port

ผลการทดลอง

เปอร์เซ็นต์การตรวจพบการบุกรุกแบบ Open Port นั้นมีแนวโน้มสูงขึ้นในช่วงเปอร์เซ็นต์การสุ่มที่ 100 – 80 และลดลงตั้งแต่เปอร์เซ็นต์ที่ 70 จนถึง 10 ดังตารางที่ 5.1 และนำมาวาดกราฟได้ดังรูปที่ 5.2 เมื่อนำข้อมูลการพบ Open Port ที่เปอร์เซ็นต์การสุ่มต่างๆ มาเปรียบเทียบกับข้อมูลการตรวจสอบที่ 100 เปอร์เซ็นต์ พบว่ามีจำนวนแพ็กเก็ตที่เหมือนกันลดลงเรื่อยๆ ดังข้อมูลในตารางที่ 5.2 ซึ่งนำมาวาดกราฟได้ดังรูปที่ 5.3

ตารางที่ 5.1 ผลการหาค่าเปอร์เซ็นต์การตรวจพบ Open Port

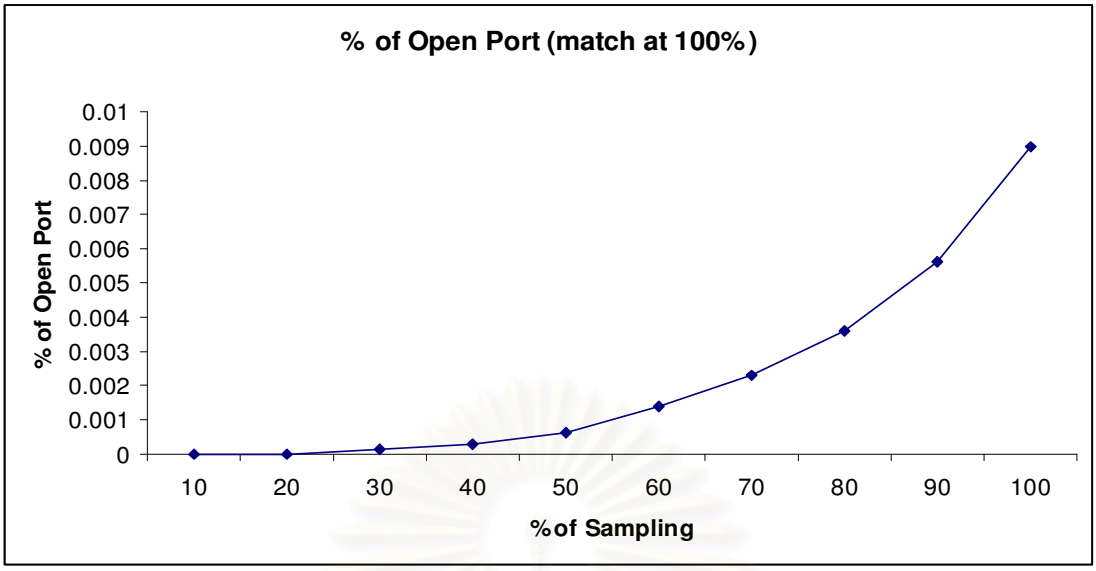
% การสุ่ม	% ที่ตรวจพบ Open Port
10	2.86E-05
20	0.000555659
30	0.002257363
40	0.004910787
50	0.007862775
60	0.009671764
70	0.011068085
80	0.011682112
90	0.01092262
100	0.008978358



รูปที่ 5.2 ความสัมพันธ์ระหว่างการตรวจพบ Open Port และระดับการสุ่ม

ตารางที่ 5.2 ผลการตรวจพบ Open Port ที่ใช้แพ็กเก็ตตรงกับข้อมูลที่ไม่ได้สุ่ม

% การสุ่ม	% ที่ตรวจพบ Open Port
10	0
20	6.64E-06
30	0.000121
40	0.000297
50	0.000637
60	0.001387
70	0.002323
80	0.003622
90	0.005602
100	0.008978



รูปที่ 5.3 ความสัมพันธ์การพบ Open Port ที่แพ็ตเก็ตตรงกับข้อมูลที่ไม่ได้สุ่ม

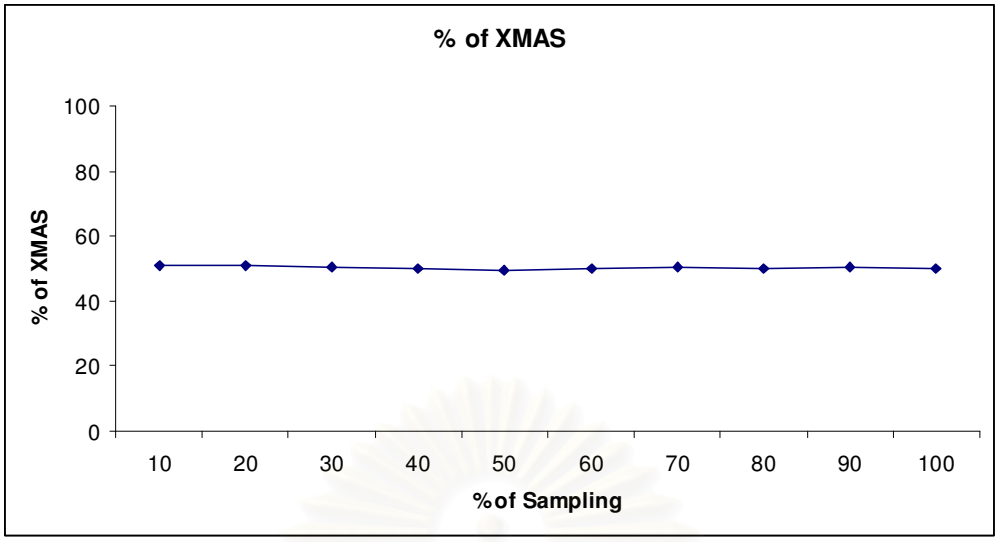
5.2.2 การทดลองสุ่มเพื่อหาเปอร์เซ็นต์การตรวจพบการสแกนพอร์ตด้วยวิธี XMAS ด้วยข้อมูลจำลองขึ้น

ผลการทดลอง

เมื่อนำแพ็กเก็ตการสแกนพอร์ตมาทำการสุ่มแล้วผ่านกระบวนการตรวจจับผู้บุกรุก พบว่าเปอร์เซ็นต์การตรวจพบนั้นค่อนข้างคงที่ แสดงให้เห็นว่าถึงแม้ว่าจำนวนข้อมูลจะลดลง และจำนวนการตรวจพบการบุกรุกนั้นลดลงในอัตราส่วนที่เท่ากัน ดังข้อมูลในตารางที่ 5.3 และนำมาวาดกราฟดังรูป 5.4

ตารางที่ 5.3 ผลการหาค่าเปอร์เซ็นต์การตรวจพบการสแกนพอร์ตแบบ XMAS

% การสุ่ม	% การสแกนแบบ XMAS
10	51.05551805
20	50.88541714
30	50.5803015
40	50.21184531
50	49.67717279
60	49.89339357
70	50.27293042
80	50.15816466
90	50.24182575
100	50.23866348



รูปที่ 5.4 ความสัมพันธ์ระหว่างการตรวจพบการสแกนพอร์ตแบบ XMAS และระดับการสุ่ม

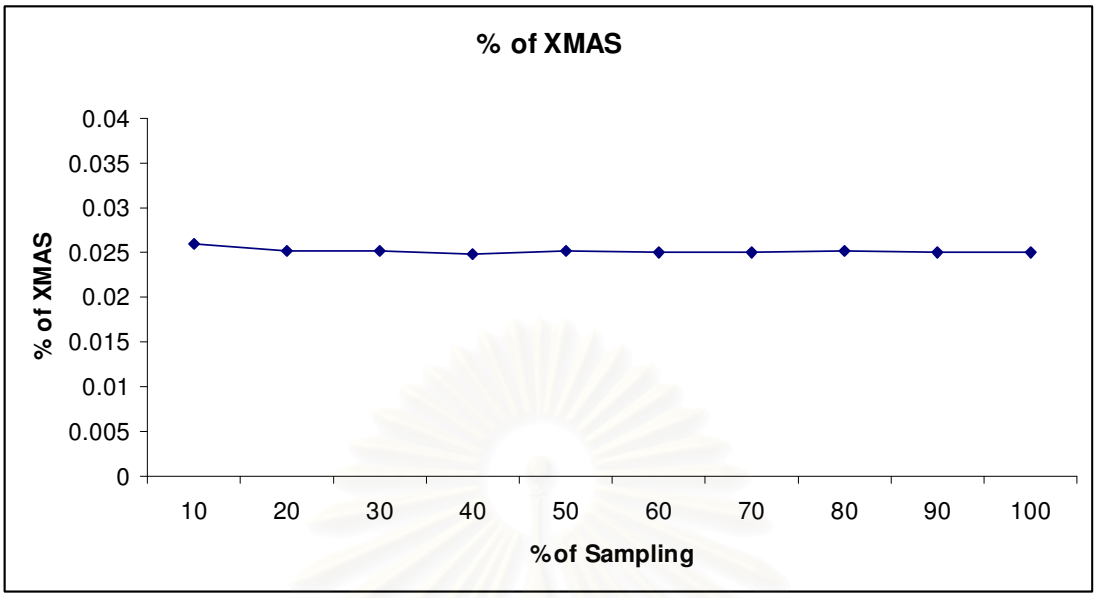
5.2.3 การทดลองสุ่มเพื่อหาค่าเปอร์เซ็นต์การตรวจพบการสแกนพอร์ตแบบ XMAS ด้วยข้อมูลจริง

ผลการทดลอง

เมื่อนำข้อมูลการบุกรุกไปรวมกับข้อมูลอื่น แล้วสุ่มหาปริมาณการบุกรุกประเภทการสแกนพอร์ตแบบ XMAS พบว่าเปอร์เซ็นต์การตรวจสอบนั้นยังคงที่ แต่ต่ำกว่าการทดลองที่ 5.2.2 เนื่องจากปริมาณแพ็กเก็ตในการทดลองนี้มีมากกว่า ดังตารางที่ 5.4 และวาดกราฟได้ดังรูปที่ 5.5

ตารางที่ 5.4 ผลการหาค่าเปอร์เซ็นต์การตรวจพบการสแกนพอร์ตแบบ XMAS

% การสุ่ม	% การสแกนแบบ XMAS
10	0.025933
20	0.025185
30	0.025194
40	0.024756
50	0.025153
60	0.025042
70	0.024972
80	0.025149
90	0.025093
100	0.025054



รูปที่ 5.5 ความสัมพันธ์ระหว่างการตรวจพบการสแกนพอร์ตแบบ XMAS และระดับการสุ่ม

5.3 การทดลองกับข้อมูลจริงที่เก็บได้

ข้อมูลที่ใช้ในการทดลองที่ 5.3 นี้ถูกดักจับข้อมูลเป็นเวลา 1 วัน ในขณะที่ดักจับนั้น ทำการสแกนพอร์ตไปยังเครื่องเซิร์ฟเวอร์ เพื่อให้ดูเหมือนกับมีการพยายามโจมตีเซิร์ฟเวอร์ด้วยโปรแกรม nmap ซึ่งข้อมูลที่ใช้ในชุดการทดลองนี้มีจำนวน 9,302,270 แพ็กเก็ต และมีแพ็กเก็ตการบุกรุกประเภทสแกนพอร์ตดังนี้

- Synscan 2,625 แพ็กเก็ต
- Xmas Scan 1,335 แพ็กเก็ต
- Finscan 441 แพ็กเก็ต
- Ackscan 831 แพ็กเก็ต
- Nullscan 438 แพ็กเก็ต

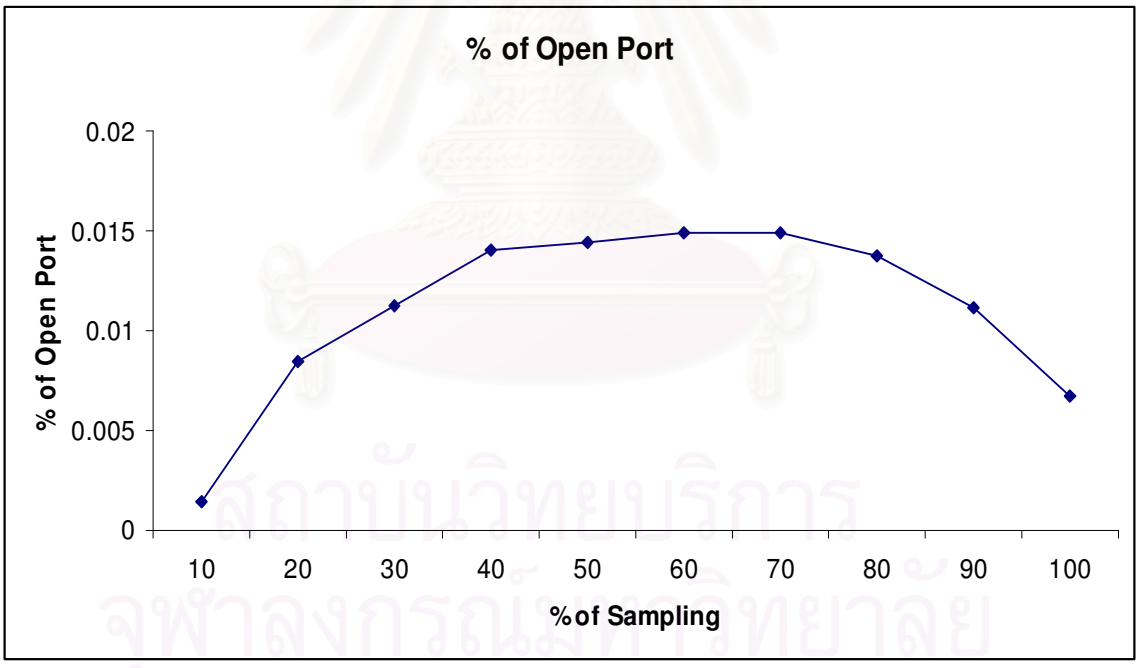
5.3.1 ทดลองหาจำนวนการสแกนพอร์ตแบบ Open Port ในข้อมูลจริง

ผลการทดลอง

เปอร์เซ็นต์การตรวจพบการบุกรุกแบบ Open Port นั้นมีแนวโน้มสูงขึ้นในช่วงเปอร์เซ็นต์การสุ่มที่ 100 – 70 และลดลงตั้งแต่เปอร์เซ็นต์ที่ 60 จนถึง 10 ดังตารางที่ 5.5 และนำมาวาดกราฟได้ดังรูปที่ 5.6 เมื่อนำข้อมูลการพบ Open Port ที่เปอร์เซ็นต์การสุ่มต่างๆ มาเปรียบเทียบกับข้อมูลการตรวจสอบที่ 100 เปอร์เซ็นต์ พบว่ามีจำนวนแพ็กเก็ตที่เหมือนกันลดลงเรื่อยๆ ดังข้อมูลในตารางที่ 5.6 ซึ่งนำมาวาดกราฟได้ดังรูปที่ 5.7

ตารางที่ 5.5 ผลการหาค่าเปอร์เซ็นต์การตรวจพบ Open Port

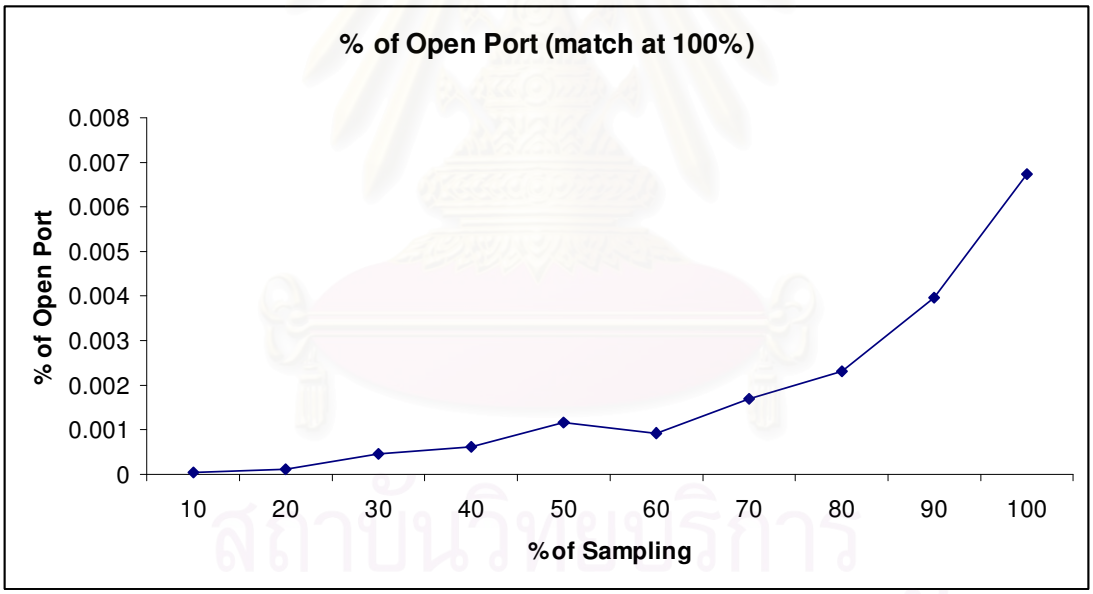
% of Sampling	% of Open Port
10	0.001462016
20	0.008420961
30	0.011233927
40	0.014080031
50	0.014462314
60	0.01487226
70	0.014924817
80	0.013775892
90	0.011120452
100	0.00674029



รูปที่ 5.6 ความสัมพันธ์ระหว่างการตรวจพบ Open Port และระดับการสุ่ม

ตารางที่ 5.6 ผลการตรวจพบ Open Port ที่แพ็กเก็ตตรงกับข้อมูลที่ไม่ได้สุ่ม

% of Sampling	% of Open Port
10	4.30004E-05
20	0.000107502
30	0.00044434
40	0.000619926
50	0.001149897
60	0.000927501
70	0.001707741
80	0.002317557
90	0.003955664
100	0.00674029



รูปที่ 5.7 ความสัมพันธ์การพบ Open Port ที่แพ็ตเก็ตตรงกับข้อมูลที่ไม่ได้สุ่ม

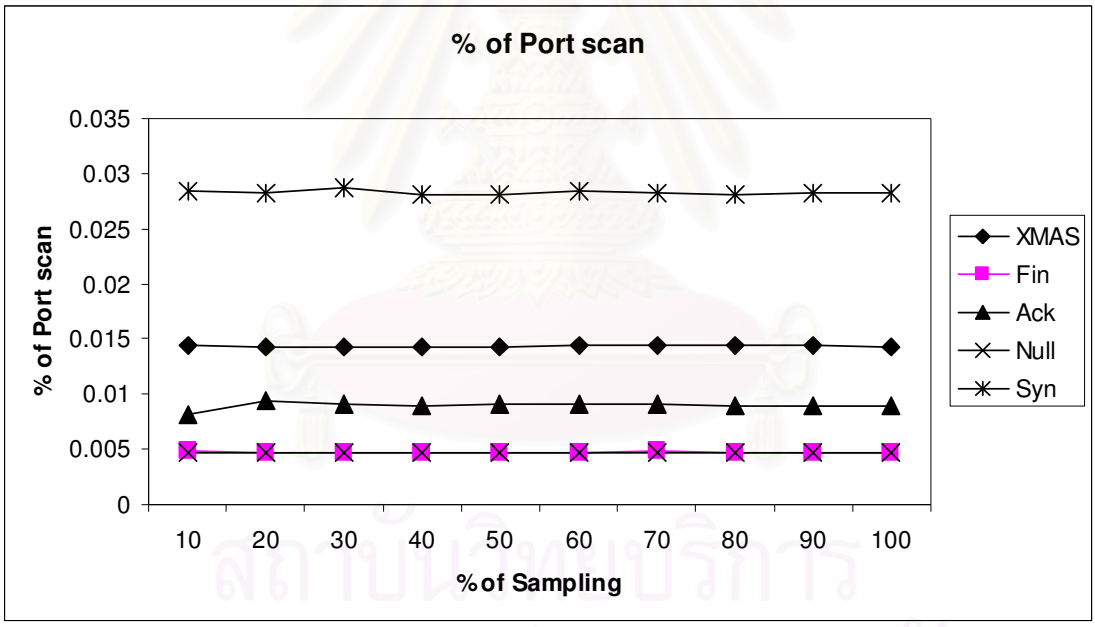
5.3.2 ทดลองหาจำนวนการสแกนพอร์ตแบบต่างๆ ในข้อมูลจริง

ผลการทดลอง

จากตารางที่ 5.7 และรูปที่ 5.8 แสดงให้เห็นว่าไม่ว่าจะมีการบุกรุกประเภทการสแกนพอร์ตแบบใดก็ตาม ก็สามารถตรวจพบได้ในเปอร์เซ็นต์การสุ่มต่างๆ สังเกตได้จากค่าเปอร์เซ็นต์การตรวจพบนั้นค่อนข้างคงที่

ตารางที่ 5.7 ผลการหาค่าเปอร์เซ็นต์การตรวจพบการสแกนพอร์ตแบบต่างๆ

% of Sampling	% of Xmas	% of Fin	% of Ack	% of Null	% of Syn
10	0.014362169	0.00480889	0.008198765	0.004704971	0.028437665
20	0.014288748	0.004687067	0.009393843	0.004644067	0.028208408
30	0.014337142	0.004725296	0.009135255	0.004773075	0.028662338
40	0.014347889	0.004764116	0.008902034	0.004750678	0.028167229
50	0.014307042	0.004761608	0.009102527	0.004711441	0.028169744
60	0.014407614	0.004770686	0.009139435	0.004713949	0.028332069
70	0.01436888	0.004809421	0.009055225	0.004690657	0.028276526
80	0.014384172	0.004727834	0.008910097	0.004709021	0.028143043
90	0.014365808	0.004726889	0.008997492	0.004686277	0.028205653
100	0.014351336	0.004740778	0.008933303	0.004708528	0.028218919



รูปที่ 5.8 ความสัมพันธ์ระหว่างการตรวจพบการสแกนพอร์ตแบบต่างๆ และระดับการสุ่ม

5.4 สรุปผลการทดลอง

จากการทดลองที่ 5.2.1 และ 5.3.1 พบว่า เปอร์เซ็นต์ของการตรวจพบการบุกรุกแบบ Open Port นั้นมีแนวโน้มไม่คงที่ มีการเพิ่มขึ้นเรื่อยๆ ในช่วงการสุ่มตั้งแต่ 100 ถึง 70 จากนั้นก็จะมีแนวโน้มลดลงเรื่อยๆ ดังรูปที่ 5.2 และ 5.6 ซึ่งจะสรุปได้ว่า การสุ่มข้อมูลนั้นทำให้โปรแกรมสนอร์ตไม่สามารถตรวจจับการบุกรุกประเภทสแกนพอร์ตแบบที่ต้องอาศัยพฤติกรรมของข้อมูล เนื่องจากฟังก์ชันการทำงานของการวิเคราะห์พฤติกรรมของข้อมูลนั้นจะอาศัยลักษณะพฤติกรรม

ของข้อมูล เพื่อบ่งบอกว่ามีการบุกรุกหรือไม่ เช่น จำนวนแพ็กเก็ต ขนาดของแพ็กเก็ต เป็นต้น และเมื่อดูจำนวนแพ็กเก็ตที่ตรงกับข้อมูลที่ไม่ได้สุ่มแล้วพบว่ามีจำนวนลดลงเรื่อยๆ แต่ขณะเดียวกันกลับมีแพ็กเก็ตใหม่ที่เพิ่มขึ้นมา ดังกราฟรูปที่ 5.3 และ 5.7

จากการทดลองที่ 5.2.2 ที่นำแพ็กเก็ตที่ได้จากการสแกนแบบ XMAS ซึ่งถูกตรวจจับได้ด้วยชุดกฎของโปรแกรมสนอร์ต พบว่ามีอัตราส่วนระหว่างจำนวนการตรวจพบการสแกนพอร์ตแบบ XMAS ต่อจำนวนแพ็กเก็ตทั้งหมดในแต่ละเปอร์เซ็นต์การสุ่มนั้นค่อนข้างคงที่ ดังรูปที่ 5.4 ซึ่งจากสมมติฐานในขั้นตอนที่ 5.1 ทำให้สรุปได้ว่าการสุ่มนั้นไม่มีผลต่อการตรวจจับการบุกรุกประเภทการสแกนพอร์ตแบบ XMAS ต่อจากนั้นจึงตั้งสมมติฐานเพิ่มเติมว่า ถ้าหากข้อมูลที่ใช้ทดลองมีมากขึ้นแล้วจะมีผลต่อการตรวจสอบจับการบุกรุกแบบ XMAS หรือไม่ จึงทำการทดลองที่ 5.2.3 โดยนำข้อมูลที่ใช้ในการทดลองที่ 5.2.2 ไปรวมกับข้อมูลอื่น เพื่อเพิ่มปริมาณแพ็กเก็ต แล้วนำไปทดลองเหมือนเดิม พบว่าอัตราส่วนระหว่างจำนวนที่พบ XMAS ต่อจำนวนแพ็กเก็ตทั้งหมดในแต่ละเปอร์เซ็นต์การสุ่มนั้นค่อนข้างคงที่เหมือนเดิม แต่เนื่องจากปริมาณแพ็กเก็ตมีมากกว่าส่งผลทำให้ค่าเปอร์เซ็นต์การตรวจพบนั้นมีค่าน้อยกว่าการทดลองที่ 5.2.2 ดังรูปที่ 5.5 อย่างไรก็ตามสามารถสรุปได้ว่าการสุ่มนั้นไม่มีผลต่อเปอร์เซ็นต์การตรวจจับการบุกรุกประเภทการสแกนพอร์ตแบบ XMAS

จากนั้นเมื่อนำข้อมูลจริงที่เก็บได้มาทดลองตามการทดลองที่ 5.3.2 สำหรับการสแกนพอร์ตในรูปแบบต่างๆ นอกเหนือไปจาก XMAS พบว่า การสุ่มยังสามารถใช้ในการตรวจสอบการบุกรุกประเภทพอร์ตสแกนด้วยการเปรียบเทียบกับชุดกฎได้ ดังรูปที่ 5.8 ซึ่งกราฟแสดงเปอร์เซ็นต์ของการบุกรุกแต่ละประเภทค่อนข้างคงที่ จึงสามารถสรุปได้ว่าการตรวจจับการบุกรุกประเภทพอร์ตสแกนประเภทต่างๆ ที่อาศัยชุดกฎในการตรวจสอบนั้น สามารถตรวจพบได้ ณ เปอร์เซ็นต์การสุ่มต่างๆ แต่การสุ่มนั้นมีผลกระทบต่อการตรวจสอบการบุกรุกประเภทการสแกนพอร์ตด้วยการวิเคราะห์พฤติกรรม ดังผลการทดลองก่อนหน้านี้

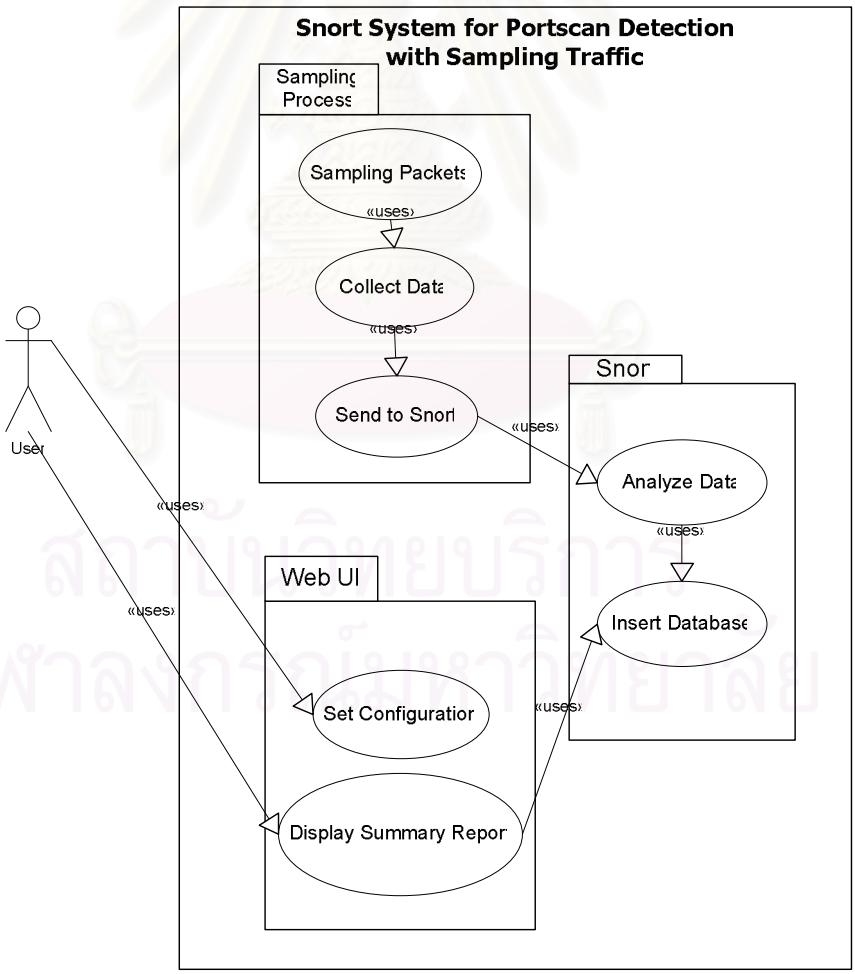
จากการทดลองทั้งหมดที่ผ่านมานั้น พบว่าโปรแกรมสนอร์ตสามารถตรวจจับการบุกรุกได้จริงถึงแม้จะสุ่มลงไปถึง 10 เปอร์เซ็นต์ก็ตาม แต่ใช้ได้เพียงการตรวจจับด้วยวิธีการเปรียบเทียบชุดกฎเท่านั้น ยังไม่สามารถใช้กับการตรวจจับด้วยวิธีการวิเคราะห์พฤติกรรม ซึ่งควรมีการวิจัยเพิ่มเติมต่อไปในอนาคตถึงสาเหตุที่ทำให้การตรวจจับการบุกรุกด้วยวิธีการวิเคราะห์พฤติกรรมจึงไม่สามารถใช้งานร่วมกับการสุ่มได้ นอกจากนี้ยังคงทำการทดลองเพิ่มเติมต่อไปในอนาคตถึงการสุ่มด้วยระดับการสุ่มที่ต่ำมากๆ เช่น 1 หรือ 0.01 เปอร์เซ็นต์ เพื่อให้สามารถใช้กับระบบเครือข่ายขนาดใหญ่ได้ ซึ่งต้องการชุดข้อมูลทดสอบขนาดใหญ่และเครื่องมือที่มีประสิทธิภาพมากกว่านี้

บทที่ 6
การประยุกต์ใช้งาน

จากที่ได้ทำการวิจัยและทดลองที่ผ่านมา แสดงให้เห็นว่าสามารถใช้การสุ่มข้อมูลแล้วยังคงทำให้โปรแกรมสนอร์ตตรวจพบการบุกรุก แต่ก็ยังเป็นโปรแกรมที่แยกกันอยู่ และไม่สามารถทำงานแบบทันที (real time) ได้ ดังนั้นในบทนี้จะกล่าวถึงภาพรวมของระบบ รายละเอียดในการพัฒนาส่วนต่างๆ และตัวอย่างการใช้งานระบบนี้

6.1 ภาพรวมของระบบ

ลักษณะโปรแกรมที่ช่วยตรวจสอบผู้บุกรุกในเครือข่ายขนาดใหญ่โดยการลดปริมาณข้อมูลลงและผู้ดูแลระบบสามารถปรับแต่งตัวรวมทั้งดูผลจำนวนบุกรุกผ่านทางเว็บไซต์ ซึ่งแสดงระบบโดยรวมของระบบดังรูปที่ 6.1

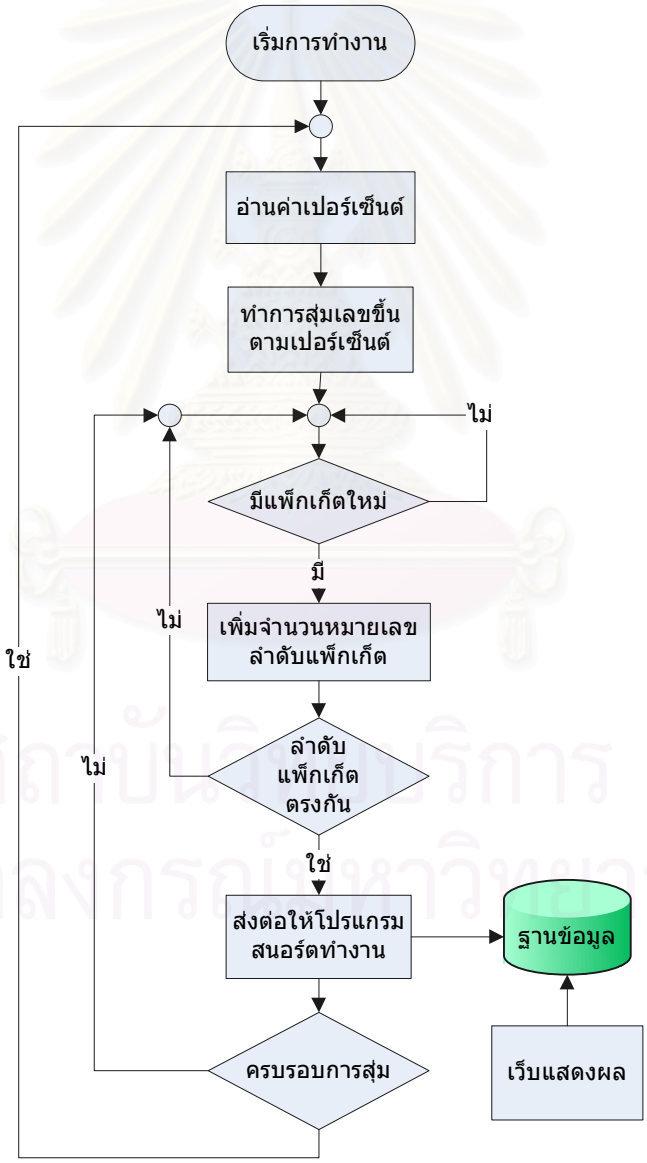


รูปที่ 6.1 ระบบโดยรวมของโปรแกรมประยุกต์ที่ได้พัฒนาขึ้น

จากรูปที่ 6.1 เมื่อระบบเริ่มทำงาน ทำการสุ่มแพ็กเก็ตแล้วจึงส่งข้อมูลที่สุ่มได้ไปยังโปรแกรมสนอร์ต เพื่อตรวจสอบหาผู้บุกรุก เมื่อพบแล้วก็จึงเก็บข้อมูลดังกล่าวเข้าฐานข้อมูลไว้ รอผู้ใช้งานเรียกดูข้อมูลผ่านหน้าเว็บไซต์ นอกจากนี้ผู้ใช้ยังสามารถปรับแต่งค่าเปอร์เซ็นต์การสุ่มผ่านหน้าเว็บไซต์ได้อีกด้วย

6.2 ลักษณะการทำงานของระบบ

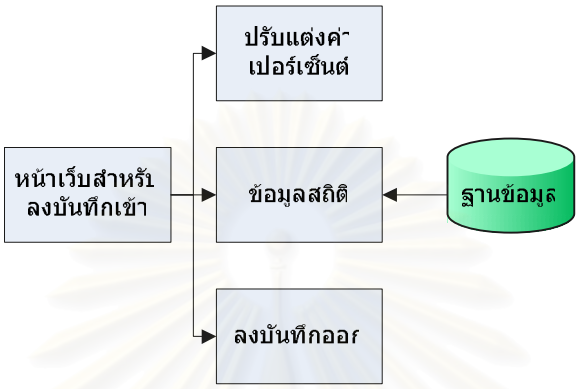
ลักษณะการทำงานของโปรแกรมนี้ จะใช้แนวทางการพัฒนาโปรแกรมสำหรับการสุ่มที่ได้กล่าวไว้ในบทที่ 3 แต่เพียงเปลี่ยนแปลงจากการที่เก็บข้อมูลลงแฟ้มเป็นการส่งข้อมูลต่อไปให้โปรแกรมสนอร์ตเพื่อตรวจหาการบุกรุก ดังรูปที่ 6.2



รูปที่ 6.2 ลักษณะการทำงานของระบบที่พัฒนาขึ้น

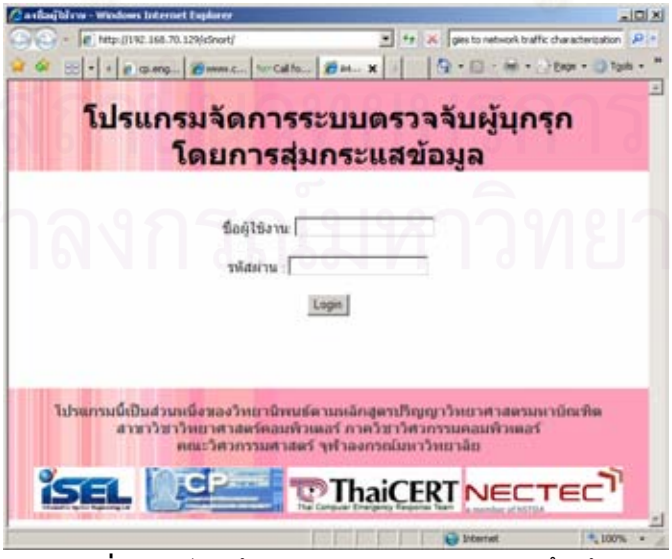
6.3 เว็บไซต์ปรับแต่งค่าและแสดงผล

เนื่องจากการปรับแต่งค่าและการดูผลในโปรแกรมสนอร์ตนั้นทำได้ยาก ทำให้ไม่สามารถระบุปัญหาหรือทราบถึงค่าโดยรวมว่ามีการบุกรุกในระบบเครือข่ายมากน้อยเพียงใด ดังนั้นจึงต้องพัฒนาเว็บไซต์เพื่อช่วยในการปรับแต่งค่าและแสดงผล รายละเอียดหน้าเว็บต่างๆ ดังรูปที่ 6.3

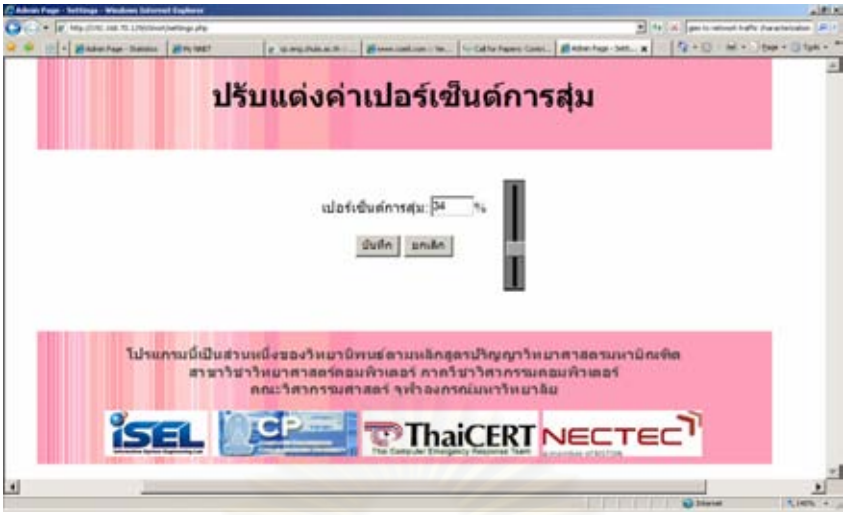


รูปที่ 6.3 ลักษณะหน้าเว็บต่างๆ ของระบบจัดการ

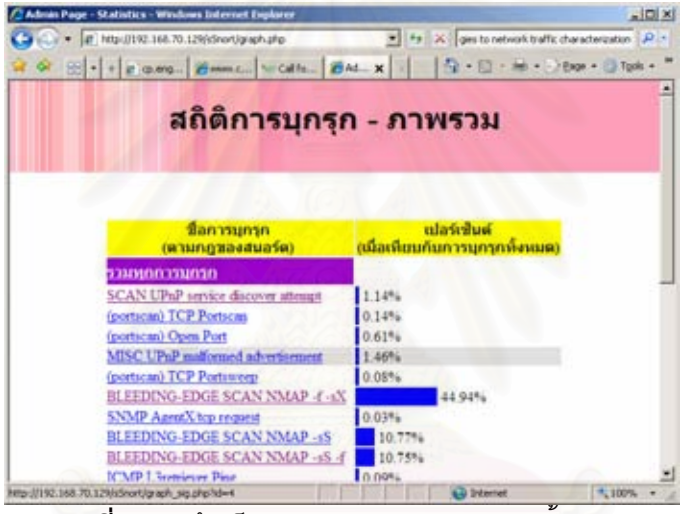
จากรูปที่ 6.3 ลักษณะการทำงานของเว็บไซต์เริ่มต้นด้วยหน้าเว็บสำหรับลงบันทึกเข้าสู่ระบบ ดังรูปที่ 6.4 จากนั้นจะมีหน้าเว็บเพื่อปรับแต่งค่า โดยการปรับแต่งนั้นจะมีเหมือนแถบให้เลื่อนแล้วกดปุ่ม บันทึก เพื่อเปลี่ยนค่า ดังรูปที่ 6.5 ส่วนในหน้าเว็บสำหรับการแสดงผลนั้น เมื่อเข้าไปแล้วจะพบกับหน้าแสดงสรุปการบุกรุกทั้งหมดที่พบ ดังรูปที่ 6.6 จากนั้นถ้าหากต้องการทราบรายละเอียดของแต่ละการบุกรุกก็เลือกเข้าไปจะปรากฏหน้าแสดงวันที่ที่ต้องการแสดงรายละเอียด ดังรูปที่ 6.7 และหาเลือกที่วันแล้วจะได้กราฟแสดงจำนวนการบุกรุกในวันนั้นตลอด 24 ชั่วโมง ดังรูปที่ 6.8



รูปที่ 6.4 เว็บหน้าแรกสำหรับทำการลงบันทึกเข้า



รูปที่ 6.5 หน้าเว็บแสดงแถบปรับแต่งค่า



รูปที่ 6.6 หน้าเว็บแสดงภาพรวมการบุกรุกทั้งหมด

รูปที่ 6.7 หน้าเว็บแสดงรายละเอียดการบุกรุกแยกรายวัน

รูปที่ 6.8 หน้าเว็บแสดงจำนวนรายละเอียดการบุกรุก

จากเนื้อหาในส่วนการประยุกต์ใช้งานนี้ กล่าวถึงภาพรวมของระบบ ลักษณะการทำงานของการสุ่มและการตรวจจับผู้บุกรุกด้วยโปรแกรมสนอร์ต ตลอดจนการแสดงผลผ่านหน้าเว็บไซต์ และในบทต่อไปจะกล่าวถึงบทสรุปของงานวิจัยทั้งหมด

บทที่ 7
สรุปผลงานวิจัยและข้อเสนอแนะ

7.1 สรุปผลการวิจัย

ในงานวิจัยนี้ได้กล่าวถึงการทดลองเพื่อศึกษาผลกระทบของการสุ่มต่อจำนวนการตรวจพบการบุกรุกที่เกิดขึ้น ซึ่งการเปรียบเทียบนั้นอาศัยเปอร์เซ็นต์ในการตรวจพบการบุกรุก เนื่องจากหน่วยของจำนวนการบุกรุกที่ใช้นั้นมีฐานการคิดที่ไม่เท่ากัน ต้องมีการปรับหน่วยให้ตรงกันก่อน ดังนั้นจึงนำหน่วยเปอร์เซ็นต์มาใช้ และในการทดลองนั้นต้องทำการสุ่มใหม่ประมาณ 30 ครั้ง เพื่อให้ได้ผลที่แน่นอนและใกล้เคียงค่าเฉลี่ยมากที่สุด และเนื่องจากการสุ่มในแต่ละครั้งนั้นได้ผลที่ไม่แน่นอนด้วย

ในการทดลองว่าการสุ่มนั้นมีผลกระทบต่อความถูกต้องในการตรวจพบการบุกรุกของโปรแกรมสนอร์ตหรือไม่ ได้ผลคือ การสุ่มมีผลกระทบต่อข้อมูลสำหรับลักษณะการตรวจจับการบุกรุกด้วยวิธีพฤติกรรม เนื่องจากจำนวนการบุกรุกนั้นจะสูงขึ้นในช่วงเปอร์เซ็นต์การสุ่มประมาณ 90 ถึง 70 เปอร์เซ็นต์ และลดลงอย่างต่อเนื่องตั้งแต่ 70 เปอร์เซ็นต์ จนถึง 10 เปอร์เซ็นต์

อย่างไรก็ตาม การสุ่มไม่มีผลต่อการตรวจจับการบุกรุกด้วยวิธีการเปรียบเทียบชุดกฎ โดยเมื่อคำนวนเป็นเปอร์เซ็นต์ที่พบแล้ว พบว่าได้ค่าที่ค่อนข้างคงที่ที่เปอร์เซ็นต์การสุ่มต่าง ๆ

การประยุกต์การใช้งานทำโดยนำวิธีการสุ่มนั้นไปเพิ่มให้โปรแกรมสนอร์ต ประกอบด้วยหน้าเว็บสำหรับการปรับแต่งค่าเปอร์เซนต์การสุ่ม และสามารถดูข้อมูลการบุกรุกที่ถูกตรวจจับได้ในรูปแบบที่เป็นเว็บไซต์ โดยอาศัยข้อมูลจากฐานข้อมูลที่ถูกเก็บด้วยโปรแกรมสนอร์ต

7.2 ปัญหาที่พบจากการวิจัย

งานวิจัยนี้มีผลการวิจัยดังที่ได้กล่าวมาแล้ว อย่างไรก็ตามงานวิจัยนี้ยังมีข้อด้อยหรือข้อจำกัดของงานวิจัยด้วยดังต่อไปนี้

1) ปัญหาในการเก็บข้อมูลสำหรับการทดลอง ปัญหาคือการเก็บข้อมูลนั้นทำได้ยาก เนื่องจากเป็นข้อมูลลับ และต้องได้รับการยินยอมจากผู้มีอำนาจการตัดสินใจหรือผู้ดูแลระบบเครือข่ายนั้นๆ และผู้ดูแลระบบอาจต้องกรองข้อมูลบางส่วนออกก่อน เพื่อความมั่นคงปลอดภัยสำหรับองค์กรที่ให้ข้อมูลด้วย ส่งผลให้ผลการตรวจสอบผิดพลาดได้ บางองค์กรเก็บข้อมูลที่ด้านหลังของไฟร์วอลล์ ซึ่งข้อมูลนั้นถูกตัดส่วนที่เป็นการบุกรุกออกไปบางส่วนแล้ว จึงทำให้ข้อมูลชุดดังกล่าวไม่สามารถนำมาใช้ได้

2) ปัญหาการทำงานของโปรแกรมสนอร์ต จากปัญหาการสุ่มที่ทำให้จำนวนการบุกรุกที่ถูกตรวจพบจากการวิเคราะห์พฤติกรรมของโปรแกรมสนอร์ตนั้นเพิ่มขึ้นอย่างผิดปกติ เนื่องจากฟังก์ชันดังกล่าวถูกฝังไว้ในรหัสต้นฉบับ (source code) ของโปรแกรมสนอร์ต จำเป็นต้องใช้เวลาในการศึกษาอย่างมากในการทำความเข้าใจวิธีการตรวจสอบพฤติกรรมของโปรแกรมสนอร์ตในระดับลึก รวมทั้งแก้ไขโปรแกรมสนอร์ตให้สามารถทำงานควบคู่กับการสุ่มข้อมูลได้

3) ปัญหาเรื่องระยะเวลาที่ใช้ในการทดลอง เนื่องจากจำนวนรอบในการทดลองที่มากและขนาดข้อมูลสำหรับการทดลองที่มีขนาดใหญ่นั้น ทำให้จำเป็นต้องใช้เวลาในการทดลองในแต่ละครั้งค่อนข้างสูง เพื่อให้ค่าเฉลี่ยที่เหมาะสม วิธีการแก้ไขคือการลดจำนวนรอบการทดสอบลง แต่ข้อเสียคือความแปรปรวนของผลการทดลองที่ได้สูง

7.3 **ข้อเสนอแนะ**

จากปัญหาที่พบและข้อจำกัดบางประการ นั้นทำให้งานวิจัยชิ้นนี้ยังมีข้อด้อยอยู่บ้าง ดังนั้นควรมีการวิจัยเพิ่มเติมดังต่อไปนี้

1) ศึกษาหาสาเหตุโดยละเอียดเพิ่มเติมในส่วนของการตรวจจับโดยใช้พฤติกรรม ว่าเพราะเหตุใดการสุ่มทำให้การตรวจจับผู้บุกรุกด้วยวิธีนี้จึงไม่ได้ผล และแก้ไขโปรแกรมสนอร์ตให้สามารถตรวจจับข้อมูลที่ถูกสุ่มโดยใช้วิธีการวิเคราะห์พฤติกรรมของข้อมูลได้

2) ศึกษาการสุ่มมีผลต่อการบุกรุกแบบอื่นๆ ที่นอกเหนือจากการสแกนพอร์ต

3) พัฒนาหน้าเว็บสำหรับการปรับแต่งค่าและแสดงผลให้มีประสิทธิภาพและใช้งานสะดวกยิ่งขึ้น

4) ทำการทดลองสุ่มเพื่อหาเปอร์เซ็นต์การตรวจพบการบุกรุกแบบต่างๆ ด้วยระดับเปอร์เซ็นต์การสุ่มที่ต่ำมากๆ เช่น 1 หรือ 0.01 เปอร์เซ็นต์ เพื่อให้สามารถใช้กับระบบเครือข่ายขนาดใหญ่ได้

จากงานวิจัยและโปรแกรมที่ได้พัฒนาขึ้นสามารถนำไปใช้ตรวจจับหาผู้บุกรุกในระบบเครือข่ายขนาดใหญ่ โดยไม่จำเป็นต้องใช้เครื่องแม่ข่ายที่มีประสิทธิภาพสูงและยังใช้ติดตั้งลงในเครื่องคอมพิวเตอร์ส่วนบุคคลได้อีกด้วย นอกจากนี้ยังมีค่าใช้จ่ายน้อยเนื่องจากโปรแกรมนี้ถูกพัฒนาต่อยอดจากโปรแกรมประเภท open source รวมทั้งใช้งานง่ายเนื่องจากมีเว็บสำหรับปรับแต่งค่าและดูสถิติการบุกรุกที่เกิดขึ้นในระบบเครือข่ายได้ ซึ่งเป็นการอำนวยความสะดวกและลดภาระให้แก่ผู้ดูแลระบบในการเฝ้าระวังระบบเครือข่ายในองค์กรหรือหน่วยงานต่างๆ ซึ่งถ้าหากพบการบุกรุกขึ้นก็สามารถทำการแก้ไขได้อย่างทันท่วงที ทำให้เกิดความปลอดภัยขึ้นในระบบเครือข่ายขององค์กรหรือหน่วยงานที่ดูแลรับผิดชอบอยู่

รายการอ้างอิง

[1] sFlow. (Online). Available from: http://www.sflow.org [18 September 2007].

[2] RFC 3176. (Online). Available from: http://www.ietf.org/rfc/rfc3176.txt [18 September 2007].

[3] Claffy, K. C., Polyzos, G. C., and Braun, H. Application of sampling methodologies to network traffic characterization. In Conference Proceedings on Communications Architectures, Protocols and Applications, San Francisco, California, United States, September 13 - 17, 1993.

[4] Nick Duffield , Carsten Lund , Mikkel Thorup, Estimating flow distributions from sampled flow statistics,In Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, Karlsruhe, Germany, August 25-29, 2003.

[5] Mai, J., Sridharan, A., Chuah, C.-N., Zang, H., and Ye, T. Impact of packet sampling on portscan detection. In IEEE Journal on Selected Areas in Communication, 2006.

[6] Li, J., Sung, M., Xu, J., and Zhao, Q., Large-Scale IP Traceback in High-Speed Internet: Practical Techniques and Theoretical Foundation. In IEEE Symposium on Security and Privacy, Berkeley, CA, May 2004.

[7] Brauckhoff, D., Tellenbach, B., Wagner, A., May, M., and Lakhina, A. Impact of packet sampling on anomaly detection metrics. In Proceedings of the 6th ACM SIGCOMM on internet Measurement, Rio de Janeriro, Brazil, October 25 - 27, 2006.

[8] Choi, B., Park, J., and Zhang, Z. Adaptive packet sampling for flow volume measurement. In SIGCOMM Computer Communication, July, 2002.

[9] Hildebrandt, R. Snot and nmap – two sides of the same coin. Linux Magazine. 4 (2001)

[10] Harper, P. Snort, Apache, SSL, PHP, MySQL, and BASE Install on CentOS 4, RHEL 4 or Fedora Core – with NTOP. (Online). Available from: http://www.snort.org/docs/setup_guides/Snort_Base_Minimal.pdf [18 September 2007].

[11] Packet Capture With libpcap and other Low Level Network Tricks. (Online). Available from: http://www.cet.nau.edu/~mc8/Socket/Tutorials/section1.html [18 September 2007].

[12] Reves, J., Panchen, S., Traffic Monitoring with Packet-Based Sampling for Defense against Security Threats. (Online). Available from: http://www.sflow.org/SamplingforSecurity.pdf [10 February 2007].

[13] The Tcpdump team.TCPdump. (Online). Available from: http://www.tcpdump.org [18 September 2007].

[14] Fyodoor. nmap. (Online). Available from: http://insecure.org/nmap [18 September 2007].

[15] Jonkman, M. Bleeding edge snort. (Online). Available from: http://www.bleedingthreats.net [18 September 2007].

[16] Jung, J., Paxson, V., Berger, A., and Balakrishnan, H. Fast portscan detection using sequential hypothesis testing. In Proceedings of the IEEE Symposium on Security and Privacy, 2004.

[17] Snoeren, A. C. Hash-based IP traceback. In Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications, San Diego, California, United States, 2001.

[18] Defense Advanced Research Projects Agency. Data Sets. (Online). Available from: http://www.ll.mit.edu/IST/ideval/data/data_index.html [26 September 2007].

ประวัติผู้เขียนวิทยานิพนธ์

นายกิติศักดิ์ จิรวรรณกูล เกิดเมื่อวันที่ 14 ตุลาคม พ.ศ. 2523 ที่จังหวัดขอนแก่น สำเร็จการศึกษาระดับปริญญาวิศวกรรมศาสตรบัณฑิต จากภาควิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น ในปีการศึกษา 2544 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2548 ปัจจุบันทำงานอยู่ศูนย์ประสานงานการรักษาความปลอดภัยคอมพิวเตอร์ประเทศไทย (ThaiCERT) ภายใต้ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) สังกัดกระทรวงวิทยาศาสตร์และเทคโนโลยี