รายงานการวิจัยฉบับสมบูรณ์

เรื่อง

วิธีการแบบผสมสำหรับการสกัดคุณลักษณะของชุดข้อมูลบนเครือข่ายเพื่อระบุผู้บุกรุกแบบเวลาจริง

(A Hybrid Method for Feature Extraction in Real-time Intrusion Detection)

โครงการวิจัยนี้ได้รับการสนับสนุนทุนวิจัย
จาก
สำนักงานคณะกรรมการวิจัยแห่งชาติ
ปีงบประมาณ พ.ศ. ๒๕๕๔

คณะผู้วิจัย

นายกฤษณะ ชินสาร	หัวหน้าโครงการวิจัย
นางสาวสุวรรณา รัศมีขวัญ	ผู้ร่วมวิจัย
นางสาวสุนิสา ริมเจริญ	ผู้ร่วมวิจัย

ศูนย์วิจัย Knowledge and Smart Technology
คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

รายงานการวิจัยฉบับสมบูรณ์
เรื่อง
วิธีการแบบผสมสำหรับการสกัดคุณลักษณะของชุดข้อมูลบนเครือข่ายเพื่อระบุผู้บุกรุกแบบเวลาจริง

(A Hybrid Method for Feature Extraction in Real-time Intrusion Detection)

โครงการวิจัยนี้ได้รับการสนับสนุนทุนวิจัย
จาก
สำนักงานคณะกรรมการวิจัยแห่งชาติ
ปีงบประมาณ พ.ศ. ๒๕๕๔

คณะผู้วิจัย

นายกฤษณะ ชินสาร หัวหน้าโครงการวิจัย
นางสาวสุวรรณา รัศมีขวัญ ผู้ร่วมวิจัย
นางสาวสุนิสา ริมเจริญ ผู้ร่วมวิจัย

ศูนย์วิจัย Knowledge and Smart Technology
คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

บทคัดย่อ

วิธีการของการตรวจจับการบุกรุกสามารถแบ่งออกได้เป็น 2 ชนิด คือ วิธีการตรวจจับการบุกรุกแบบอโนมาลี (anomaly intrusion detection method) และวิธีการตรวจจับการบุกรุกแบบมิสยูส (*misuse intrusion detection method*) โดยที่วิธีการตรวจจับการบุกรุกแบบอโนมาลีนั้นเป็นวิธีการหาผู้บุกรุกโดยการวิเคราะห์การใช้งานของผู้ใช้งาน หรือตัวระบบเองที่เบี่ยงเบนไปจากระดับการใช้งานโดยปกติ ส่วนการตรวจจับการบุกรุกแบบมิสยูสนั้น เป็นวิธีการหาผู้บุกรุกโดยการเปรียบเทียบข้อมูลที่เข้ามากับรูปแบบของผู้บุกรุกที่มีอยู่เดิม ซึ่งทั้งสองวิธีนี้มีจุดแข็งและจุดอ่อนที่แตกต่างกัน ปัญหาที่เด่นชัดที่สุดของการตรวจจับการบุกรุกแบบมิสยูส คือ ไม่สามารถตรวจจับการบุกรุกแบบใหม่ หรือการบุกรุกที่ไม่มีในชุดรูปแบบของผู้บุกรุกที่มีได้ ส่วนการตรวจจับการบุกรุกแบบอโนมาลีนั้น จะสามารถตรวจจับการบุกรุกจากผู้บุกรุกที่ไม่มีในฐานข้อมูลการบุกรุกได้ แต่ปัญหาที่สำคัญในการตรวจจับการบุกรุกแบบอโนมาลี คือ ทำอย่างไรถึงจะสร้างเค้าโครงของการใช้งานปกติที่ดีได้

ในงานวิจัยนี้ คณะผู้วิจัยได้แสดงให้เห็นแล้วว่าการสกัดคุณลักษณะของชุดข้อมูลบนเครือข่าย มีความสำคัญต่อการพัฒนาการระบุผู้บุกรุกเป็นอย่างมาก ในการได้มาซึ่งตัวแทนชุดคุณลักษณะของชุดข้อมูลที่เหมาะสม เพื่อใช้ในการระบุผู้บุกรุกโดยอาศัยวิธีการแบบผสมในการสกัดคุณลักษณะของชุดข้อมูลเครือข่าย ซึ่งจะเพิ่มความสามารถในการระบุผู้บุกรุกได้เหมาะสมมากกว่า การพัฒนาการสกัดคุณลักษณะชุดข้อมูลเครือข่าย ประกอบด้วย 2 ขั้นตอน คือ 1.การหาคุณลักษณะของชุดข้อมูลที่สามารถแทนข้อมูลได้และมีจำนวนคุณลักษณะที่เหมาะสม และขั้นตอนที่ 2.การรู้จำรูปแบบการบุกรุกเพื่อระบุผู้บุกรุกจากชุดข้อมูลบนเครือข่ายจากคุณลักษณะที่ได้จากการสกัดคุณลักษณะของชุดข้อมูล โดยวัดประสิทธิภาพจากอัตราความเร็วในการตรวจจับผู้บุกรุก และเปอร์เซ็นต์ความผิดพลาดของการตรวจจับผู้บุกรุก

Abstract

Detection of Network Intrusion can be categorized into two groups. The First one is Anomaly Intrusion Detection Method. The second one is Misuse Intrusion Detection Method. For the first method is to inspect the irregular behavior on the usage of the network or on the computer systems. For the second method is to inspect the mis-matching with those patterns store in the database. This brings the discussion of improper way to detect the intrusion as the intruders keep on changing their ways to intrude the networks or computer systems.

In this research report, we have demonstrated how the use of feature selection on those data traffic will help in improving the detection of intrusion more efficient. There are two steps in extract feature on data traffic to detect intrusion. First step is to extract features. And then use pattern recognition to validate whether there is any anomaly behavior for those data traffic. We compare the speed in detecting the intrusion and measure the percentage of the misclassification.

สารบัญ

บทที่ 1 บทนำ **1**

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของโครงการวิจัย 2

1.3 ขอบเขตของโครงการวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับ 3

1.5 ระยะเวลาทำการวิจัยและแผนการดำเนินงานตลอดโครงการวิจัย 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง **5**

2.1 ลักษณะของข้อมูลที่ใช้ในการทำแบบทดลอง 5

2.2 กระบวนการรู้จำทางคอมพิวเตอร์ 5

2.2.1 ระบบโครงข่ายประสาทเทียมแบบวิธีการแพร่กระจายย้อนกลับ (Back propagation Algorithm) 6

2.2.2 วิธีการรู้จำแบบซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) 8

2.3 วิธีการวิเคราะห์องค์ประกอบหลัก 9

2.3.1 การหาค่าไอเกน และไอเกนเวกเตอร์ (Eigen Value and Eigen Vector) 10

2.4 การสกัดคุณลักษณะสำคัญ (Feature Extraction) 11

2.5 การทบทวนวรรณกรรม/สารสนเทศ (Information) ที่เกี่ยวข้อง 11

บทที่ 3 วิธีดำเนินการวิจัย **13**

3.1 การจัดการชุดข้อมูล 13

3.1.1 การสกัดคุณลักษณะชุดข้อมูล 13

3.1.2 การรู้จำด้วยโครงข่ายประสาทเทียม 13

3.1.3 การประเมินระบบ 14

บทที่ 4 ผลการทดลอง **15**

4.1 การสกัดคุณลักษณะข้อมูล 15

4.2 การรู้จำประเภทของผู้บุกรุกเบื้องต้น 16

บทที่ 5 สรุปผลการทดลอง **19**

5.1 สรุปผลการทดลอง 19

5.2 งานที่ต้องทำต่อไปในปีงบประมาณ พ.ศ. 2555 19