บทคัดย่อ

ปัจจุบันมีการใช้ internet เป็นจำนวนมาก และจำนวนผู้ใช้มีขนาดเพิ่มขึ้นในลักษณะก้าวกระโดด (exponential) การรักษาความปลอดภัยของข้อมูลถือว่าเป็นปัญหาที่สำคัญยิ่ง ระบบการตรวจจับการบุกรุกในเครือข่ายเป็นระบบที่คอยตรวจพฤติกรรมการใช้ที่ผิดปกติหรือผิดไปจากกฎเกณฑ์ที่ตั้งไว้จึงเป็นระบบที่สำคัญพื้นฐานสำหรับงานรักษาความปลอดภัยของข้อมูล ด้วยปริมาณการใช้ที่เพิ่มขึ้น ปัญหาคือต้องพยายามหาระบบการตรวจจับการบุกรุกในเครือข่ายที่มีการประมวลผลที่รวดเร็ว และมีความผิดพลาดในการแจ้งเตือนน้อยที่สุด ระบบการตรวจจับการบุกรุกในเครือข่ายส่วนใหญ่จะใช้คุณลักษณะทั้งหมดจำนวน 41 ตัวแปรเพื่อการหาวิธีใหม่ๆ สำหรับการจัดประเภทของการบุกรุกในเครือข่ายโดยมีการปรับค่าน้ำหนักของแต่ละคุณลักษณะ แต่มีส่วนน้อยที่เริ่มสนใจคัดเลือกเฉพาะคุณลักษณะที่เด่นเท่านั้นเพื่อการลดเวลาในการประมวลผลและเพิ่มประสิทธิภาพการตรวจจับการบุกรุกในเครือข่าย ซึ่งขั้นตอนวิธีในการคัดเลือกคุณลักษณะเด่นนี้ยังซับซ้อน และประมวลผลช้าอยู่ ผู้วิจัยนำเสนอ factor analysis ในการจัดกลุ่มคุณลักษณะทั้งหมดเหล่านี้ แล้วเลือกเฉพาะลักษณะเด่นในแต่ละกลุ่มมาเป็นตัวแทนมาคัดเลือกคุณลักษณะที่เหมาะสมที่สุดสำหรับระบบการตรวจจับการบุกรุกในเครือข่าย ที่สามารถลดเวลาในการประมวลผลและเพิ่มประสิทธิภาพความแม่นยำการตรวจจับการบุกรุกในเครือข่าย ผลการทดลองกับฐานข้อมูลจำนวน 494,020 รายการ จาก KDD Cup 1999 ปรากฏว่าวิธีที่นำเสนอ เลือกคุณลักษณะเด่นเพียง 13 ตัว จาก 41 ตัว ประมาณ 31.71 % ทำให้ประหยัดเวลาในการประมวลผลโดยยังคงความถูกต้องเทียบเท่าการใช้คุณลักษณะทั้งหมด 41 ตัว

Abstract

The internet and local area networks are growing larger in recent years. As a great variety of people all over the world connecting to the internet, they are unconsciously encountering the number of security threats such as viruses, worms and attacks from hackers. Therefore, intrusion detection is becoming a more and more important technology which follows up network traffic and identifies network intrusion such as anomalous network behaviors, unauthorized network access, and malicious attacks to computer systems. Mostly, researches use 41 features to create new methods for classifying network intrusion. Rarely researches focus on feature selection. This paper uses the knowledge of factor analysis to group features, and then picks the main feature in each group representing for main features. The way reduces only 13 main features from 41 features (31.71 %). It is obvious that the reducing features can reduce processing time. With 494,020 records dataset from KDD Cup 1999, the experimental results show that the novel feature selection gain the accuracy greater than or equal to the accuracy from the whole features.