การตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออกด้วยวิธีเชิงเวฟเล็ต

A Wavelet-Based Anomaly Detection in Outbound Network Traffic

เกรียงไกร ลิ่มทอง และ ธิดารัตน์ ต่อสุข
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยกรุงเทพ
9/1 ม.5 ถ.พหลโยธิน อ.คลองหลวง จ.ปทุมธานี 12120 E-mail: {kriangkrai.l,thidarat.t}@bu.ac.th

บทคัดย่อ

การเฝ้าระวังและตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออก (ข้อมูลที่ส่งจากระบบเครือข่ายภายในออกไปยังระบบเครือข่ายอินเตอร์เน็ต) มีความแตกต่างจากการเฝ้าระวังและตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาเข้า (ข้อมูลที่รับจากระบบเครือข่ายอินเตอร์เน็ตเข้ามายังระบบเครือข่ายภายใน) โดยเฉพาะอย่างยิ่งในกรณีของการโจมตีแบบกระจายเพื่อหยุดการให้บริการ (Distributed Denial of Service Attacks) ชุดข้อมูลสำหรับการโจมตีที่ตรวจจับได้ในข้อมูลจราจรคอมพิวเตอร์ขาเข้า (ที่ระบบเครือข่ายเป้าหมาย) จะมีปริมาณมากกว่าที่ตรวจจับได้ในข้อมูลจราจรคอมพิวเตอร์ขาออก (ที่ระบบเครือข่ายต้นทาง) ดังนั้นการนำวิธีตรวจจับความผิดปกติที่ใช้กับข้อมูลจราจรคอมพิวเตอร์ขาเข้ามาใช้กับข้อมูลจราจรคอมพิวเตอร์ขาออกจึงได้ประสิทธิภาพที่ไม่เท่าเทียมกัน บทความนี้จึงได้นำเสนอวิธีการตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออกด้วยวิธีเชิงเวฟเล็ต เพื่อป้องกันไม่ให้ระบบเครือข่ายคอมพิวเตอร์เข้าไปมีส่วนร่วมในการโจมตีระบบเครือข่ายอื่น ๆ ดังนั้นถ้าผู้ดูแลระบบนำวิธีการดังกล่าวไปใช้ในการเฝ้าระวังและตรวจจับความผิดปกติในระบบเครือข่ายคอมพิวเตอร์ จะช่วยให้สามารถตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออกได้ถูกต้องและแม่นยำมากยิ่งขึ้น

คำสำคัญ: เวฟเล็ต, ข้อมูลจราจรคอมพิวเตอร์, สัญญาณทางเวลา, การตรวจจับความผิดปกติ, ระบบเครือข่ายคอมพิวเตอร์

Abstract

Monitoring and detecting anomalies in outbound network traffic (traffic from a customer network entering the Internet) are different from inbound network traffic (traffic from the Internet entering a customer network). Particularly, in case of Distributed Denial of Service (DDoS) attacks, the number of attack packets in inbound traffic (at the target of attack) is much larger than the number of attack packets in outbound traffic (at the source of attack). Consequently, the accuracy of anomaly detection methods used in inbound traffic is diverse from those in outbound traffic. We propose a wavelet-based anomaly detection in outbound network traffic so as to prevent internal attacker using network system involve attacking or intruding to another network through the Internet. If network administrators employ the proposed method on their network system, it would enable them to monitor and detect various anomalies in outbound network traffic accurately and expeditiously.

Keywords: wavelet, network traffic, time series, anomaly detection, computer network

1. บทนำ

ความผิดปกติที่เกิดขึ้นในระบบเครือข่ายคอมพิวเตอร์เกิดขึ้นได้จากหลายสาเหตุ เราสามารถแบ่งความผิดปกติในระบบเครือข่ายคอมพิวเตอร์ออกตามลักษณะสาเหตุที่เกิดความผิดปกติได้เป็น 2 ประเภท คือ 1) ความผิดปกติที่เกิดจากการบุกรุกหรือการโจมตี ตัวอย่างเช่น การโจมตีเพื่อหยุดการให้บริการ (Denial of Service Attacks), การส่งอีเมลขยะหรือข้อความขยะ (Spamming), การแพร่กระจายของหนอนหรือไวรัสคอมพิวเตอร์ (Worms or Viruses) เป็นต้น และ 2) ความผิดปกติที่เกิดจากความผิดพลาดของอุปกรณ์หรือจากผู้ใช้งาน ตัวอย่างความผิดปกติประเภทนี้ได้แก่ การทำงานผิดพลาดของอุปกรณ์ (Failures), การเปลี่ยนแปลงค่าในระบบทำให้ระบบทำงานผิดพลาด (Misconfigurations) เป็นต้น มีงานวิจัยจำนวนมากได้เสนอแนวทางในการตรวจจับความผิดปกติดังกล่าว ซึ่งสามารถแบ่งออกตามลักษณะของวิธีการตรวจจับได้เป็น 2 ประเภทคือ การตรวจจับการใช้งานที่ผิด (Misuse Detection) และการตรวจจับการใช้งานที่ผิดปกติ (Anomaly Detection)

การตรวจจับการใช้งานที่ผิดจะมีการกำหนดกฎ (Rules) หรือสัญลักษณ์ (Signatures) ที่ระบุถึงรูปแบบของการใช้งานที่ผิด ถ้าตรวจจับได้ว่ามีข้อมูลหรือพฤติกรรมใดที่ตรงตามกฎหรือสัญลักษณ์ที่ได้กำหนดไว้ก็จะส่งสัญญาณเตือนว่ามีการใช้งานที่ผิดเกิดขึ้น แต่ถ้าข้อมูลหรือพฤติกรรมใดไม่ตรงตามกฎหรือสัญลักษณ์ที่ได้กำหนดไว้ ระบบจะตัดสินใจว่าเป็นข้อมูลหรือพฤติกรรมการใช้งานที่ปกติ ซึ่งแตกต่างจากการตรวจจับการใช้งานที่ผิดปกติ โดยวิธีนี้จะมีการเก็บข้อมูลในระบบเครือข่ายเป็นระยะเวลาหนึ่งเพื่อให้ระบบได้เรียนรู้ถึงข้อมูลและพฤติกรรมการใช้งานปกติ โดยผู้ดูแลระบบจะทำการสอนและกำหนดให้ระบบรู้ว่าข้อมูลใดเป็นการใช้งานปกติหรือข้อมูลใดเป็นการใช้งานที่

ผิดปกติ ถ้าตรวจจับได้ว่ามีข้อมูลหรือพฤติกรรมใดที่แตกต่างจากข้อมูลการใช้งานปกติที่ระบบเรียนรู้ก็จะส่งสัญญาณเตือนว่ามีข้อมูลหรือพฤติกรรมที่ผิดปกติเกิดขึ้น ดังนั้นจึงทำให้วิธีการตรวจจับการใช้งานที่ผิดปกตินี้สามารถตรวจจับข้อมูลหรือพฤติกรรมผิดปกติที่ยังไม่เคยเกิดขึ้นมาก่อนได้ ซึ่งแตกต่างจากการตรวจจับการใช้งานที่ผิดที่สามารถตรวจจับได้เฉพาะข้อมูลหรือพฤติกรรมการใช้งานที่ผิดที่เคยเกิดขึ้นและถูกกำหนดเอาไว้ในระบบแล้วเท่านั้น

งานวิจัยนี้ได้นำเสนอแนวทางการตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออกของระบบเครือข่าย (Outgoing Traffic) ซึ่งมีรูปแบบเดียวกันกับการตรวจจับการใช้งานที่ผิดปกติ โดยมีจุดประสงค์เพื่อตรวจจับความผิดปกติในระบบเครือข่ายที่มีผู้บุกรุกใช้เครื่องคอมพิวเตอร์ภายในระบบเครือข่ายโจมตีไปยังระบบเครือข่ายภายนอกผ่านทางอินเตอร์เน็ต วิธีดังกล่าวใช้การเปลี่ยนข้อมูลจราจรคอมพิวเตอร์ให้อยู่ในรูปสัญญาณทางเวลา (Time Series) หลังจากนั้นจึงนำสัญญาณดังกล่าวไปแยกส่วนประกอบด้วยวิธีเชิงเวฟเล็ต แล้วนำส่วนประกอบของสัญญาณดังกล่าวไปเปรียบเทียบกับลักษณะการใช้งานปกติ โดยใช้วิธีการประมวลผลทางสถิติมาช่วยวิเคราะห์และตรวจสอบว่ามีข้อมูลหรือพฤติกรรมในข้อมูลจราจรคอมพิวเตอร์ขาออกที่แตกต่างจากข้อมูลการใช้งานปกติหรือไม่ ถ้าแตกต่างก็จะส่งสัญญาณเตือนว่ามีสิ่งผิดปกติเกิดขึ้น ถ้าไม่แตกต่างก็จะนำข้อมูลดังกล่าวไปประมวลผลทางสถิติเพื่อรวมกับข้อมูลเดิมที่เป็นการใช้งานปกติต่อไป

บทความนี้ประกอบด้วยส่วนต่าง ๆ ตามลำดับดังนี้ งานวิจัยที่เกี่ยวข้อง, ขั้นตอนและหลักการทำงาน, การทดลองและผลการทดลอง, สรุปและแนวทางในการพัฒนาต่อไป

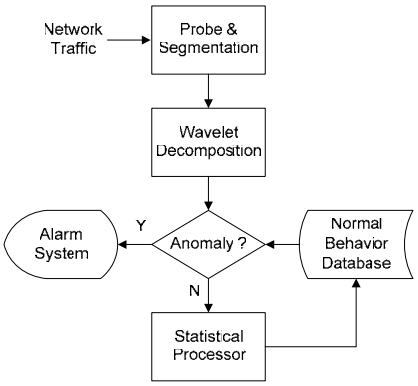
2. งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษารายงานการสำรวจวิธีตรวจจับความผิดปกติในระบบเครือข่ายคอมพิวเตอร์จากงานของ Peng และคณะ [1] พบว่าสามารถแบ่งวิธีในการตรวจจับความผิดปกติได้เป็น 2 ประเภท คือ การตรวจจับการใช้งานที่ผิด (Misuse Detection) และ การตรวจจับการใช้งานที่ผิดปกติ (Anomaly Detection) มีงานวิจัยจำนวนมากได้เสนอวิธีต่าง ๆ ที่เป็นการตรวจจับการใช้งานที่ผิดปกติ เช่น การใช้เหมืองข้อมูล (Data Mining), การใช้วิธีวิเคราะห์เชิงสถิติ (Statistics), หรือแม้แต่การวิเคราะห์ด้วยการประมวลผลสัญญาณ (Signal Processing) เป็นต้น คุณลักษณะเด่นของวิธีตรวจจับการใช้งานที่ผิดปกติคือสามารถตรวจจับการโจมตีรูปแบบใหม่ที่ยังไม่เคยเกิดขึ้นมาก่อนได้ หรือที่เราเรียกการโจมตีดังกล่าวว่า Zero-Day Attacks

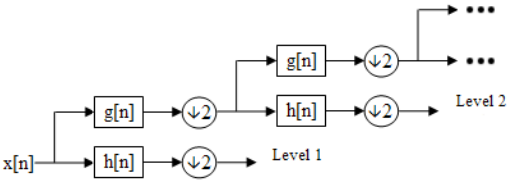
งานวิจัยของ Blazek และคณะ [2] ได้เสนอวิธีการตรวจจับความผิดปกติในระบบเครือข่ายโดยการประยุกต์ใช้กระบวนการทางสถิติมาวิเคราะห์ข้อมูลโปรโตคอลแบบหลายระดับในระบบเครือข่าย เพื่อตรวจจับการเปลี่ยนแปลงของจำนวนชุดข้อมูลต่อช่วงเวลาที่เกิดขึ้นจากการโจมตี วิธีการดังกล่าวยังมีบางกรณีที่มีการโจมตีเกิดขึ้นแต่ไม่สามารถตรวจจับการโจมตีที่เกิดขึ้นได้ (False Negative) งานวิจัยของ Glenn และคณะ [3] จึงเสนอการปรับปรุงวิธีของ Blazek และคณะ [2] โดยใช้การแปลงเวฟเล็ตเข้ามาช่วยในการกรองสัญญาณซึ่งทำให้ False Negative มีจำนวนลดลง งานวิจัยข้างต้นเป็นตัวอย่างการประยุกต์ใช้วิธีการทางสถิติมาวิเคราะห์เพื่อหาความผิดปกติในระบบเครือข่าย

นอกเหนือจากการประยุกต์ใช้วิธีการทางสถิติแล้วยังมีงานวิจัยที่เสนอวิธีประมวลผลสัญญาณ เช่น การแยกส่วนประกอบของสัญญาณด้วยวิธีเชิงเวฟเล็ต เพื่อนำมาวิเคราะห์และหาความผิดปกติในระบบเครือข่าย ตัวอย่างเช่นงานวิจัยของ Barford และคณะ [4] ได้เสนอแนวทางในการตรวจจับความผิดปกติในระบบเครือข่ายโดยการแยกส่วนประกอบความถี่ด้วยวิธีเชิงเวฟเล็ตออกเป็น 3 ความถี่; ความถี่สูง, ความถี่กลาง, และความถี่ต่ำ ซึ่งวิธีการดังกล่าวสามารถตรวจจับความผิดปกติที่เกิดแบบฉับพลัน (Abrupt Changes) ได้ แต่ยังไม่สามารถตรวจจับความผิดปกติที่เกิดแบบค่อยเป็นค่อยไป (Step Changes) ได้ อีกตัวอย่างหนึ่งก็คืองานวิจัยของ Lu และคณะ [5] ได้ศึกษาผลกระทบของเวฟเล็ตแม่แบบต่าง ๆ ต่อการตรวจจับความผิดปกติและได้เสนอวิธีตรวจจับความผิดปกติโดยวิธีเชิงเวฟเล็ตและการประมาณค่าแบบถดถอย นอกจากนั้นงานวิจัยของ Kriangkrai และคณะ [6] ได้เสนอวิธีวิเคราะห์ข้อมูลที่จัดเก็บใน Darknet ด้วยวิธีเชิงเวฟเล็ตเพื่อแยกความผิดปกติที่เกิดจากการโจมตีด้วยชุดข้อมูลจำนวนมากออกจากข้อมูลหรือพฤติกรรมผิดปกติที่เกิดจากสาเหตุอื่น

ในงานวิจัยนี้ได้เสนอวิธีการตรวจจับความผิดปกติในระบบเครือข่ายซึ่งแตกต่างจากงานวิจัยที่กล่าวมาข้างต้น เนื่องจากวิธีที่นำเสนอเป็นการประยุกต์วิธีการทางสถิติและการประมวลผลสัญญาณด้วยวิธีเชิงเวฟเล็ตในการแยกส่วนประกอบของสัญญาณออกเป็นระดับต่าง ๆ รวมทั้งสิ้น 11 ระดับ และศึกษาว่าส่วนประกอบของสัญญาณในระดับใดที่เหมาะสำหรับนำไปใช้ในการตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออก โดยสามารถตรวจจับความผิดปกติที่เกิดแบบฉับพลันและแบบค่อยเป็นค่อยไปได้



รูปที่ 1 ขั้นตอนการตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์



รูปที่ 2 แผนภูมิต้นไม้สำหรับการแยกส่วนประกอบด้วยวิธีเชิงเวฟเล็ต

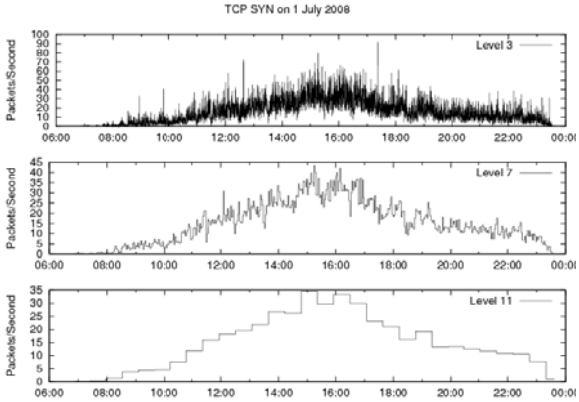
3. ขั้นตอนและหลักการทำงาน

งานวิจัยนี้ได้แบ่งขั้นตอนการตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออกดังแสดงในรูปที่ 1

Probe and Segmentation : มีหน้าที่รับข้อมูลจราจรคอมพิวเตอร์และคัดกรองข้อมูลในส่วนที่ต้องการ หลังจากนั้นจึงทำการแปลงข้อมูลจราจรคอมพิวเตอร์ดังกล่าวให้อยู่ในรูปของสัญญาณทางเวลาที่แสดงถึงจำนวนชุดข้อมูลต่อวินาที การศึกษาครั้งนี้ได้ทำการคัดกรองข้อมูลจราจรคอมพิวเตอร์ตามโปรโตคอลมาตรฐานที่ใช้ในระบบเครือข่าย ได้แก่ ICMP, TCP SYN, TCP SYN/ACK, และ UDP

Wavelet Decomposition : ทำหน้าที่แยกส่วนประกอบของสัญญาณด้วยวิธีการแปลงเวฟเล็ตแบบเต็มหน่วย (Discrete Wavelet Transform : DWT) เพื่อแยกสัญญาณที่ได้รับออกเป็นส่วนประกอบสัญญาณ 2 ส่วน คือ ส่วนประกอบสัญญาณความถี่ต่ำ (Approximation) และส่วนประกอบสัญญาณความถี่สูง (Detail) จากนั้นจึงนำส่วนประกอบความถี่ต่ำมาแยกส่วนประกอบของสัญญาณด้วยวิธีการแปลงเวฟเล็ตแบบเต็มหน่วยอีกครั้ง เราก็จะได้ส่วนประกอบของสัญญาณในระดับต่อไป รูปที่ 2 แสดงให้เห็นถึงแผนภูมิต้นไม้สำหรับการแยกส่วนประกอบของสัญญาณด้วยวิธีเชิงเวฟเล็ตในระดับต่าง ๆ โดย x[n] เป็นสัญญาณทางเวลาที่ได้จากข้อมูลจราจรคอมพิวเตอร์ขาออก g[n] เป็นส่วนประกอบสัญญาณความถี่ต่ำ และ h[n] เป็นส่วนประกอบความถี่สูง ส่วนประกอบทั้งสองจะมีอัตราสุ่ม (Sampling Rate) ของสัญญาณลดลงเหลือครึ่งหนึ่งของอัตราสุ่มเดิมตามกฎของ Nyquist การศึกษาครั้งนี้ใช้เวฟเล็ตแม่แบบ Haar ในการแยกส่วนประกอบของสัญญาณเพราะสามารถประมวลผลได้เร็วกว่าเวฟเล็ตแม่แบบอื่น รูปที่ 3 แสดงตัวอย่างลักษณะส่วนประกอบสัญญาณความถี่ต่ำในระดับที่ 3, 7, และ 11 ของโปรโตคอล TCP SYN ที่ได้จากการแปลงเวฟเล็ตแบบเต็มหน่วยโดยใช้เวฟเล็ตแม่แบบ Haar

Anomaly Classification : ในขั้นตอนนี้จะนำส่วนประกอบสัญญาณความถี่ต่ำที่ได้จากการแยกส่วนประกอบด้วยวิธีเชิงเวฟเล็ตไปเปรียบเทียบกับข้อมูลการใช้งานปกติที่ระดับเดียวกัน ซึ่งถูกจัดเก็บเอาไว้ใน Normal Behavior Database ถ้าข้อมูลจราจรคอมพิวเตอร์ขาออกแตกต่างไปจากค่าฐาน (Threshold) ที่ระบบคำนวณได้นั่นหมายความว่ามีความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออก ระบบจะทำการส่งสัญญาณเตือนไปยัง Alarm System เพื่อเตือนให้ผู้ดูแลระบบทราบว่ามีสิ่งผิดปกติเกิดขึ้นในระบบเครือข่ายคอมพิวเตอร์



รูปที่ 3 ส่วนประกอบสัญญาณความถี่ต่ำในระดับที่ 3 (บน), 7 (กลาง) และ 11 (ล่าง)

Alarm System : เป็นระบบที่ทำหน้าที่แสดงผลให้ผู้ใช้งานหรือผู้ดูแลระบบเครือข่ายทราบถึงสิ่งผิดปกติที่เกิดขึ้น โดยรับผลการตัดสินใจที่ได้จากขั้นตอน Anomaly Classification และประมวลผลข้อมูลดังกล่าวเพื่อแสดงให้ผู้ใช้งานหรือผู้ดูแลระบบทราบว่ามีข้อมูลหรือพฤติกรรมใดที่ผิดปกติ

Statistical Processor : กรณีที่ข้อมูลจราจรคอมพิวเตอร์ขาออกไม่มีสิ่งผิดปกติเกิดขึ้น ที่ขั้นตอนนี้จะทำการประมวลผลทางสถิติและรวมข้อมูลจราจรคอมพิวเตอร์ขาออกเข้ากับข้อมูลการใช้งานปกติที่ถูกจัดเก็บเอาไว้ แต่ถ้าข้อมูลจราจรคอมพิวเตอร์ขาออกมีสิ่งผิดปกติเกิดขึ้น ข้อมูลดังกล่าวจะไม่ถูกรวมเข้ากับข้อมูลการใช้งานปกติ การประมวลผลดังกล่าวจะทำให้ข้อมูลการใช้งานปกติถูกปรับปรุงให้มีความทันสมัยอยู่เสมอเพื่อใช้ในการตรวจสอบความผิดปกติต่อไป ค่าทางสถิติที่ถูกประมวลผลและถูกจัดเก็บในฐานข้อมูลได้แก่ ค่าบันทึกเวลา (t), จำนวนข้อมูล (n), ผลรวมของข้อมูล $(\sum x)$, และผลรวมของข้อมูลยกกำลังสอง $(\sum x^2)$

Normal Behavior Database : คือฐานข้อมูลสำหรับจัดเก็บข้อมูลจราจรคอมพิวเตอร์ขาออกที่มีพฤติกรรมการใช้งานปกติ ในขั้นตอน Anomaly Classification ข้อมูลดังกล่าวจะถูกเรียกไปคำนวณหาค่าฐานเพื่อนำไปเปรียบเทียบกับข้อมูลจราจรคอมพิวเตอร์ขาออกและทำการตรวจสอบว่ามีสิ่งผิดปกติเกิดขึ้นหรือไม่

ค่าฐานที่ถูกนำไปใช้ในการตัดสินใจที่ขั้นตอน Anomaly Classification นั้นถูกคำนวณมาจากค่าทางสถิติทั้ง 4 ค่าที่จัดเก็บอยู่ในฐานข้อมูล โดยคำนวณจากสมการที่ (1) – (3)

$$b_{(j,t)} = \bar{x}_{(j,t)} \pm c\sigma_{(j,t)} \quad (1)$$

$$E(X_{(j,t)}) = \bar{x}_{(j,t)} = \frac{\sum_{i=1}^{n} x_{(i,j,t)}}{n} \quad (2)$$

$$\sigma^2_{(j,t)} = E\left(X^2_{(j,t)}\right) - \left(E\left(X_{(j,t)}\right)\right)^2 \quad (3)$$

จากสมการที่ (1) ค่า b คือค่าฐานที่ถูกนำไปใช้ในการตัดสินใจ คำนวณได้จากค่าเฉลี่ย $\bar{x}$ และค่าเบี่ยงเบนมาตรฐาน σ ของส่วนประกอบสัญญาณระดับ j ที่เวลา t โดยมี c เป็นค่าคงที่สำหรับกำหนดค่าความมั่นใจ (Confidence Interval) การทดลองครั้งนี้กำหนดค่าความมั่นใจเท่ากับ 0.95 ค่าเฉลี่ย $\bar{x}$ สามารถคำนวณได้จากสมการที่ (2) โดย i คือวันที่เก็บข้อมูลเริ่มจากวันที่ 1 จนถึงวันที่ n และค่าเบี่ยงเบนมาตรฐาน σ ในสมการที่ (1) สามารถคำนวณได้จากสมการที่ (3)

4. การทดลองและผลการทดลอง

4.1 ข้อมูลจราจรคอมพิวเตอร์

งานวิจัยนี้ได้ทำการเก็บข้อมูลจราจรคอมพิวเตอร์ขาออกที่มีข้อมูลและพฤติกรรมการใช้งานปกติจากศูนย์บริการสารสนเทศมหาวิทยาลัยเกษตรศาสตร์ (Kasetsart IT Square) วิทยาเขตบางเขน ระหว่างเดือนมิถุนายนถึงสิงหาคม 2551 ศูนย์แห่งนี้มีเครื่องปฏิบัติการคอมพิวเตอร์ลูกข่ายให้บริการจำนวนทั้งหมด 175 เครื่อง ทุกเครื่องมีการติดตั้งโปรแกรมป้องกันไวรัสและมีการปรับปรุงฐานข้อมูลในโปรแกรมป้องกันไวรัสอย่างสม่ำเสมอ มีการติดตั้งโปรแกรมพื้นฐานที่จำเป็นสำหรับการใช้งานทั่วไป ผู้ใช้งานไม่สามารถติดตั้งโปรแกรมอื่นเพิ่มเติมได้ แต่ละวันมีผู้เข้ามาใช้งานประมาณ 1,300 คน ศูนย์แห่งนี้เปิดให้บริการวันจันทร์ถึงวันศุกร์ตั้งแต่เวลา 8.30 ถึง 24.00 น. และวันเสาร์ตั้งแต่เวลา 8.30 ถึง 16.30 น.

4.2 ขั้นตอนการทดลอง

ในการทดลองได้ทำการคัดกรองข้อมูลจราจรคอมพิวเตอร์ขาออกที่เป็นการใช้งานปกติจากโปรโตคอลมาตรฐาน ICMP, TCP SYN, TCP SYN/ACK, และ UDP ทั้งหมด 55 วัน ถูกนำมาสอนให้ระบบเรียนรู้ถึงพฤติกรรมการใช้งานปกติ 39 วัน อีก 16 วันที่เหลือถูกนำมาจำลองความผิดปกติให้เกิดขึ้นในระบบเครือข่ายคอมพิวเตอร์

การจำลองความผิดปกติในระบบเครือข่ายเริ่มด้วยการจำลองปริมาณชุดข้อมูลที่ผิดปกติในระบบเครือข่าย 5 ชุดข้อมูลต่อวินาทีเป็นระยะเวลา 1 นาที การจำลองดังกล่าวทำซ้ำทุก 1 ชั่วโมงตั้งแต่ 9.00 ถึง 23.00 น. รวมทั้งหมด 15 ครั้งใน 1 วัน โดยจำลองความผิดปกติทั้ง 4 โปรโตคอลตามที่กล่าวมาแล้ว จากนั้นให้ระบบตรวจสอบความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ดังกล่าว บันทึกผลการตรวจจับว่าถูกต้องตรงกันกับการจำลองข้อมูลที่ผิดปกติหรือไม่

จากนั้นจึงเพิ่มปริมาณชุดข้อมูลที่ผิดปกติเป็น 10, 15, 20 จนถึง 50 ชุดข้อมูลต่อวินาที ในแต่ละช่วงเพิ่มปริมาณขึ้นครั้งละ 5 ชุดข้อมูลต่อวินาที หลังจากนั้นจึงทำการจำลองความผิดปกติซ้ำแบบเดิมตามที่กล่าวมาแล้วแต่เพิ่มระยะเวลาการจำลองชุดข้อมูลที่ผิดปกติจาก 1 นาทีเป็น 2, 3, 4 จนถึง 10 นาที ในแต่ละช่วงเพิ่มขึ้นครั้งละ 1 นาที เมื่อทำการทดลองครบทั้งหมดจึงนำค่าเฉลี่ยของผลการทดลองที่ได้มาสร้างเป็นกราฟเพื่อดูอัตราความความถูกต้องในการตรวจจับ (Detection Rate) และอัตราความผิดพลาดที่เกิดการแจ้งเตือน (False Positive Rate) ของส่วนประกอบสัญญาณในแต่ละระดับ

4.3 วิธีการวัดผล

งานวิจัยนี้ได้ทำการวัดอัตราความความถูกต้องในการตรวจจับและอัตราความผิดพลาดที่เกิดการแจ้งเตือนของส่วนประกอบสัญญาณตั้งแต่ระดับที่ 1 จนถึงระดับที่ 11 โดยวัดการจำลองชุดข้อมูลผิดปกติที่เกิดขึ้นตั้งแต่ 5-50 ชุดข้อมูลต่อวินาที และวัดการจำลองชุดข้อมูลผิดปกติที่เกิดขึ้นเป็นระยะเวลาตั้งแต่ 1-10 นาที อัตราความถูกต้องในการตรวจจับและอัตราความผิดพลาดที่เกิดการแจ้งเตือนคำนวณได้จากสมการที่ (4) และ (5)

$$detection\ rate = \frac{number\ of\ true\ alarms}{total\ number\ of\ alarms} \quad (4)$$

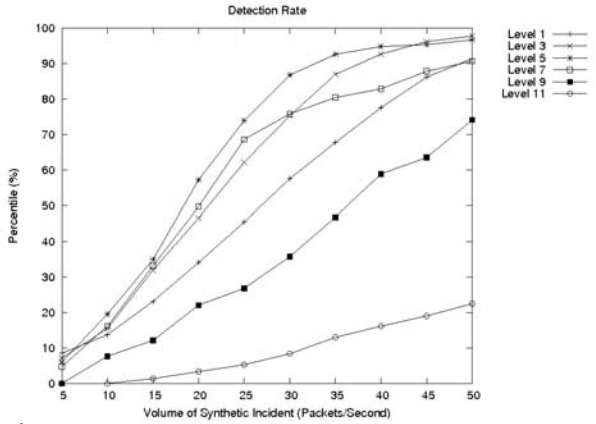
$$false\ positive\ rate = \frac{number\ of\ false\ positive}{total\ number\ of\ alarms} \quad (5)$$

เราแบ่งช่วงเวลา (Time Slot) ทั้งหมดเท่ากับ 57,600 วินาที เริ่มตั้งแต่เวลา 8.00 ถึง 24.00 น. ดังนั้นจะมีจำนวนการแจ้งเตือนทั้งหมด (Total Number of Alarms) เท่ากับ 57,600 ครั้ง ตัวอย่างเช่นถ้าระบบตรวจจับได้ถูกต้องทั้งข้อมูลปกติและผิดปกติ (Number of True Alarms) เป็นจำนวน 50,000 ครั้ง เมื่อคำนวณอัตราความถูกต้องในการตรวจจับตามสมการที่ (4) จะได้เท่ากับ 86.81% และถ้าระบบแจ้งเตือนว่ามีสิ่งผิดปกติเกิดขึ้นแต่ความจริงแล้วไม่มีสิ่งผิดปกติ (Number of False Positive) เป็นจำนวน 7,600 ครั้ง เมื่อคำนวณอัตราความผิดพลาดที่เกิดการแจ้งเตือนตามสมการที่ (5) จะได้เท่ากับ 13.19% เป็นต้น

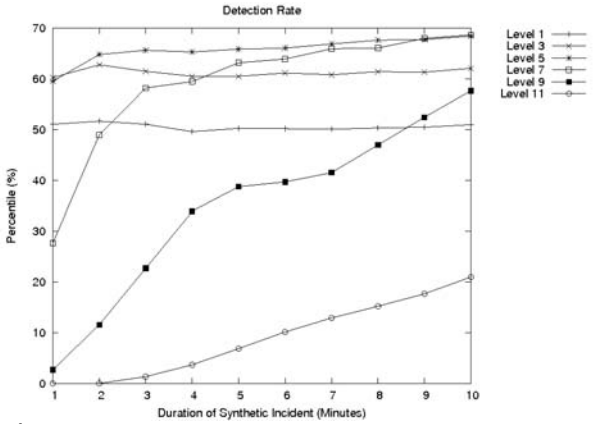
4.4 ผลการทดลอง

รูปที่ 4-6 แสดงถึงอัตราความถูกต้องในการตรวจจับและอัตราความผิดพลาดที่เกิดการแจ้งเตือนเปรียบเทียบกับจำนวนชุดข้อมูลที่ผิดปกติต่อวินาทีและเปรียบเทียบกับระยะเวลาที่เกิดความผิดปกติของส่วนประกอบเวฟเล็ตตั้งแต่ระดับที่ 1-11 ทั้งนี้เพื่อความสะดวกในการพิจารณาผลการทดลองจึงนำมาแสดงเฉพาะระดับที่เป็นเลขคี่เท่านั้น

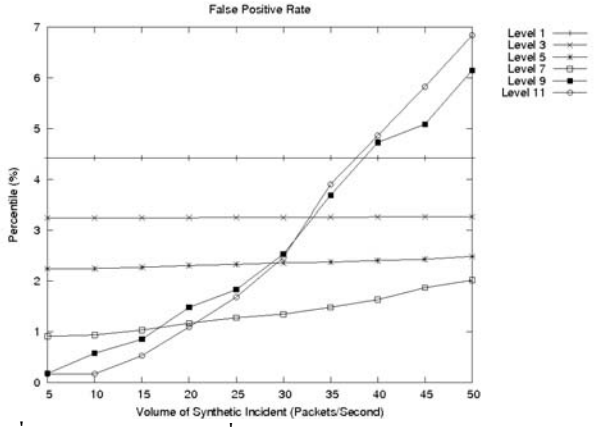
จากการจำลองความผิดปกติในระบบเครือข่าย เมื่อปริมาณชุดข้อมูลที่ผิดปกติเพิ่มมากขึ้นอัตราความถูกต้องในการตรวจจับก็จะเพิ่มมากขึ้นเช่นเดียวกัน เมื่อพิจารณาส่วนประกอบสัญญาณในแต่ละระดับพบว่าอัตราความถูกต้องในการตรวจจับเพิ่มมากขึ้นเมื่อใช้ส่วนประกอบสัญญาณในระดับที่สูงขึ้น ส่วนประกอบสัญญาณระดับที่ 5 จะมีอัตราความถูกต้องในการตรวจจับสูงที่สุด หลังจากนั้นเมื่อใช้ส่วนประกอบสัญญาณที่สูงกว่าระดับที่ 5 อัตราความถูกต้องในการตรวจจับก็จะลดลงตามลำดับ ดังแสดงในรูปที่ 4 เมื่อพิจารณาอัตราความผิดพลาดที่เกิดการแจ้งเตือนพบว่าการใช้ส่วนประกอบสัญญาณระดับที่ 7 มีอัตราความผิดพลาดที่เกิดการแจ้งเตือนน้อยที่สุดเมื่อปริมาณข้อมูลที่ผิดปกติมีขนาดมากกว่า 20 ชุดข้อมูลต่อวินาที ดังแสดงในรูปที่ 5



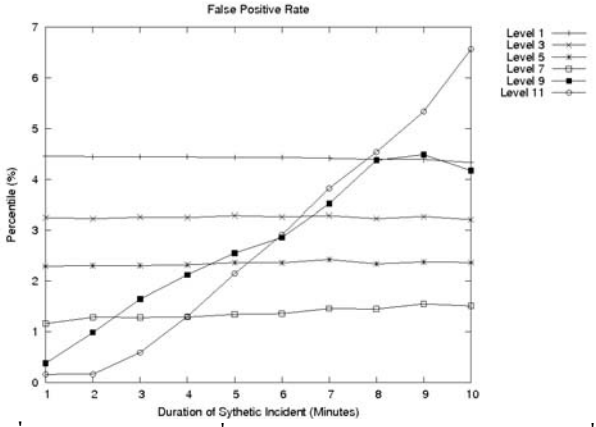
รูปที่ 4 อัตราความถูกต้องในการตรวจจับเปรียบเทียบกับจำนวนชุดข้อมูลที่ผิดปกติต่อวินาที



รูปที่ 6 อัตราความถูกต้องในการตรวจจับเปรียบเทียบกับระยะเวลาที่เกิดความผิดปกติ



รูปที่ 5 อัตราความผิดพลาดที่เกิดการแจ้งเตือนเปรียบเทียบกับจำนวนชุดข้อมูลที่ผิดปกติต่อวินาที



รูปที่ 7 อัตราความผิดพลาดที่เกิดการแจ้งเตือนเปรียบเทียบกับระยะเวลาที่เกิดความผิดปกติ

รูปที่ 6 แสดงให้เห็นว่าเมื่อเพิ่มระยะเวลาการจำลองข้อมูลที่ผิดปกติในระบบเครือข่าย การใช้ส่วนประกอบสัญญาณระดับที่ 5 ก็ยังคงให้อัตราความถูกต้องในการตรวจจับสูงที่สุด แต่ถ้าใช้ส่วนประกอบสัญญาณในระดับที่ต่ำกว่าหรือสูงกว่าระดับที่ 5 อัตราความถูกต้องในการตรวจจับก็จะลดลงตามลำดับเช่นกัน เมื่อพิจารณาอัตราความผิดพลาดที่เกิดการแจ้งเตือนเปรียบเทียบกับระยะเวลาการจำลองข้อมูลที่ผิดปกติในระบบเครือข่ายพบว่าการใช้ส่วนประกอบสัญญาณระดับที่ 7 มีอัตราความผิดพลาดที่เกิดการแจ้งเตือนน้อยที่สุดเมื่อระยะเวลาการจำลองข้อมูลที่ผิดปกติมากกว่า 4 นาที ดังแสดงในรูปที่ 7

5. สรุปและแนวทางในการพัฒนาต่อไป

งานวิจัยนี้ได้เสนอแนวทางในการตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออกเพื่อป้องกันไม่ให้ผู้บุกรุกใช้ระบบเครือข่ายภายในทำการบุกรุกหรือโจมตีระบบเครือข่ายภายนอกผ่านทางอินเตอร์เน็ต โดยวิธีดังกล่าวจะทำการเปลี่ยนข้อมูลจราจรคอมพิวเตอร์เป็นสัญญาณทางเวลา หลังจากนั้นจึงแยกส่วนประกอบของสัญญาณด้วยวิธีการแปลงเวฟเล็ตแบบเต็มหน่วย และนำส่วนประกอบของสัญญาณดังกล่าวไปเปรียบเทียบกับข้อมูลหรือพฤติกรรมการใช้งานปกติที่ระดับเดียวกัน ถ้าส่วนประกอบสัญญาณของข้อมูลจราจรคอมพิวเตอร์ขาออกต่างไปจากค่าฐานของการใช้งานปกติ ก็จะตัดสินใจว่ามีสิ่งผิดปกติเกิดขึ้นในระบบเครือข่าย แต่ถ้าไม่ต่างไปจากค่าฐานของการใช้งานปกติก็จะตัดสินใจว่าไม่มีสิ่งผิดปกติเกิดขึ้นในระบบเครือข่าย หลังจากนั้นก็จะรวมสัญญาณที่ไม่มีสิ่งผิดปกติเกิดขึ้นเข้ากับฐานข้อมูลที่ใช้เก็บข้อมูลหรือพฤติกรรมการใช้งานปกติ เพื่อปรับปรุงฐานข้อมูลให้มีความเป็นปัจจุบัน

ผลการทดลองกับข้อมูลจราจรคอมพิวเตอร์ขาออกที่มีการใช้งานจริงได้แสดงให้เห็นว่าส่วนประกอบของสัญญาณในระดับที่ 5-7 มีความเหมาะสมในการนำไปตรวจจับความผิดปกติในข้อมูลจราจรคอมพิวเตอร์ขาออก ส่วนประกอบของสัญญาณระดับที่ 5 มีอัตราความความถูกต้องในการตรวจจับมากกว่าระดับอื่นเมื่อเปรียบเทียบกับปริมาณและระยะเวลาที่เกิดความผิดปกติที่มีค่าเท่ากัน แต่ส่วนประกอบสัญญาณในระดับที่ 5 ยังมีอัตราความผิดพลาดที่เกิดการแจ้งเตือนมากกว่าส่วนประกอบสัญญาณระดับที่ 7 ในทางตรงกันข้าม ผลการทดลองแสดง

ให้เห็นว่าส่วนประกอบของสัญญาณในระดับที่ 7 มีอัตราความผิดพลาดที่เกิดการแจ้งเตือนน้อยที่สุดเมื่อความผิดปกติมีปริมาณมากกว่า 20 ชุดข้อมูลต่อวินาทีหรือเกิดความผิดปกติเป็นระยะเวลามากกว่า 4 นาที แต่ส่วนประกอบสัญญาณระดับที่ 7 มีอัตราความถูกต้องในการตรวจจับน้อยกว่าระดับที่ 5

เนื่องจากในงานวิจัยนี้ผู้วิจัยได้ทำการเก็บข้อมูล, ทดลองการทำงานของระบบ, และจำลองการทำงานเป็นแบบออฟไลน์ ดังนั้นแนวทางในการพัฒนาต่อไปคือ ผู้วิจัยต้องการให้ระบบดังกล่าวสามารถนำไปใช้งานและตรวจจับความผิดปกติแบบออนไลน์ได้ และนำระบบตรวจจับความผิดปกติแบบออนไลน์ดังกล่าวไปทดสอบกับระบบเครือข่ายคอมพิวเตอร์ที่มีการใช้งานจริง นอกจากนี้ข้อมูลที่ใช้ในงานวิจัยครั้งนี้เป็นข้อมูลที่ได้มาจากขอบของระบบเครือข่าย (Access Network) ซึ่งมีลักษณะของข้อมูลและพฤติกรรมการใช้งานแตกต่างจากข้อมูลที่ได้จากแกนของระบบเครือข่าย (Core Network) ดังนั้นจึงจำเป็นต้องมีการเก็บข้อมูลจากแกนของระบบเครือข่ายและนำมาทดสอบกับวิธีการที่ได้นำเสนอ เพื่อศึกษาว่าวิธีการดังกล่าวสามารถนำไปประยุกต์ใช้ที่แกนของระบบเครือข่ายได้หรือไม่

6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากโครงการพัฒนาอาจารย์ประจำภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ ของมหาวิทยาลัยกรุงเทพ ประจำปี 2552

เอกสารอ้างอิง

[1] P. Tao, et al., "Survey of network-based defense mechanisms countering the DoS and DDoS problems," ACM Comput. Surv., vol. 39, p. 3, 2007.

[2] B. Rudolf, et al., "A Novel Approach to Detection of Denial-of-Service Attacks via Adaptive Sequential and Batch-Sequential Change-Point Detection Methods." Proc. IEEE Workshop Information Assurance and Security. IEEE CS Press, pp. 220 226, 2001.

[3] G. Carl, et al., "Wavelet based Denial-of-Service detection," Computers & Security, vol. 25, pp. 600-615, 2006.

[4] B. Paul, et al., "A signal analysis of network traffic anomalies," presented at the Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment, Marseille, France, 2002.

[5] L. Wei, et al., "Detecting Network Anomalies Using Different Wavelet Basis Functions," 2008, p. 149.

[6] K. Limthong, et al., "Wavelet-Based Unwanted Traffic Time Series Analysis," in Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on, 2008, pp. 445-449.