ระบบวิเคราะห์ล็อกแบบอัตโนมัติเพื่อการพิสูจน์หลักฐานทางดิจิทัล

An Automated Log Analyzer for Digital Forensics Investigation

มงคล พิรารักษ์[1] และ ศุภกร กังพิศดาร[2]

คณะวิทยาการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีมหานคร

E-mail: mpirarak@yahoo.com[1] และ supakorn@mut.ac.th[2]

บทคัดย่อ – งานวิจัยฉบับได้ทำการศึกษาและพัฒนาระบบเก็บรวบรวม และระบบวิเคราะห์ล็อกไฟล์แบบอัตโนมัติ (Automated Log Analyzer) เพื่อให้ได้มาของข้อมูลที่จะนำมาวิเคราะห์หาผู้กระทำผิดได้ ผู้วิจัยได้วิเคราะห์ข้อดีข้อเสียของโมเดลการวิเคราะห์ล็อกแบบอัตโนมัติที่มีอยู่และได้เสนอโมเดลใหม่สำหรับการวิเคราะห์ล็อกแบบอัตโนมัติที่สามารถแสดงหลักฐานการก่ออาชญากรรมทางคอมพิวเตอร์ได้ นอกจากนี้ผู้วิจัยยังได้สร้างระบบต้นแบบตามโมเดลที่ได้นำเสนอ ผลที่ได้จากการพัฒนาระบบต้นแบบแสดงให้เห็นว่าโมเดลที่นำเสนอสามารถช่วยให้ผู้ใช้สามารถวิเคราะห์และรายงานผลการจัดเก็บหลักฐานเพื่อการพิสูจน์หลักฐานทางดิจิทัลได้

คำสำคัญ – การพิสูจน์หลักฐานทางเครือข่าย, การวิเคราะห์ล็อก, การพิสูจน์หลักฐานทางดิจิทัล, การตรวจจับการบุกรุก

ABSTRACT – In this paper, we propose a study and development of an automated log analyzer in order to acquire critical evidence of crime and find potential suspects. We analyze several existing network forensics models and propose a new model for automated log analyzer that can provide and summarize necessary evidence. We create a prototype of the proposed model. The results of our implementation show that the proposed model can assist users analyze and report evidence for forensics investigation.

KEYWORDS – Network forensics, log analysis, digital forensics, intrusion detection

1. บทนำ

การที่จะสามารถจัดการกับผู้ก่ออาชญากรรมทางคอมพิวเตอร์จำเป็นต้องมีพยานหลักฐานที่น่าเชื่อถือ ซึ่งพยานหลักฐานที่จะนำมาใช้ในการจัดการกับผู้ก่ออาชญากรรมนั้น จะต้องได้พยานหลักฐานมาอย่างรวดเร็วและมีพยานหลักฐานเพียงพอในการเอาผิดกับผู้ก่ออาชญากรรม โดยที่พยานหลักฐานดังกล่าวจะต้องมีความน่าเชื่อถือด้วย อย่างไรก็ตาม ปัญหาที่ทำให้ไม่สามารถนำบทลงโทษไปจัดการกับผู้ก่ออาชญากรรมได้ก็คือ ความยากลำบากในการเก็บรวบรวมและการวิเคราะห์พยานหลักฐาน ซึ่งต้องใช้เวลาในการเก็บรวบรวมและวิเคราะห์อย่างมาก

ที่ผ่านมามีงานวิจัยหลายๆ งาน [1, 2, 3, 4] ที่ศึกษาวิธีการเก็บรวบรวมและการวิเคราะห์พยานหลักฐานเป็นแบบอัตโนมัติโดยอาศัยความสัมพันธ์ของเหตุการณ์หรืออาศัยเทคนิคทางวิทยาศาสตร์ เพื่อจุดประสงค์ในการจัดการกับผู้ก่ออาชญากรรม แต่ระบบที่ถูกนำเสนอโดย Lin *et al.* [2] ยังมีข้อบกพร่องใน

ส่วนของการเก็บรักษาข้อมูลต้นฉบับ (Raw Log) เพื่อป้องกันการแก้ไขหรือลบข้อมูลต้นฉบับ

การเก็บรวบรวมพยานหลักฐาน เป็นสิ่งแรก ๆ ที่ต้องดำเนินการก่อน เพื่อให้ได้มาของข้อมูลที่จะนำมาวิเคราะห์หาผู้กระทำความผิดได้ จึงเป็นเหตุผลของการที่จะต้องได้มาของข้อมูลจากแหล่งข้อมูลหลาย ๆ แหล่ง เพื่อที่จะได้ข้อมูลเพียงพอที่ใช้ในการวิเคราะห์ต่อไป ปัญหาต่อมาคือ ข้อมูลที่ได้มาบางอย่างก็ไม่เกี่ยวข้องกับสิ่งที่ต้องการวิเคราะห์ ด้วยเหตุนี้จึงเป็นสิ่งที่ต้องมีการพัฒนาวิธีการเพื่อลดปัญหาดังกล่าว ให้สอดคล้องกับวัตถุประสงค์ของการปรับปรุงคุณภาพของพยานหลักฐานที่เก็บรวบรวมได้โดยอัตโนมัติ

บทความวิจัยฉบับนี้นำเสนอการออกแบบและพัฒนาระบบเก็บรวบรวมและระบบวิเคราะห์ล็อกไฟล์ เพื่อเป็นการป้องปัญหา การแก้ไขหรือลบข้อมูลต้นฉบับจากผู้ที่ไม่หวังดีหรือจากผู้ก่ออาชญากรรม รวมถึงสามารถนำข้อมูลที่ได้จากวิเคราะห์ข้อมูลแล้วมาพิสูจน์ได้ว่าเป็นข้อมูลที่ได้มาจากข้อมูลต้นฉบับจริง

โครงสร้างของงานวิจัยฉบับนี้มีดังนี้ บทที่ 2 กล่าวถึงแนวคิด บทที่ 3 เป็นงานวิจัยที่นำเสนอที่ได้จากแนวคิดที่ได้กล่าวถึงไปแล้ว บทที่ 4 กล่าวถึงการพัฒนาระบบที่ใช้ในการทดลอง รวมถึงผลการทดลองที่ได้ บทที่ 5 เป็นการวิเคราะห์และอภิปรายในส่วนของระบบที่ได้พัฒนาขึ้น บทที่ 6 เป็นบทสรุปของงานวิจัยที่ได้นำเสนอไปแล้ว

2. แนวคิด

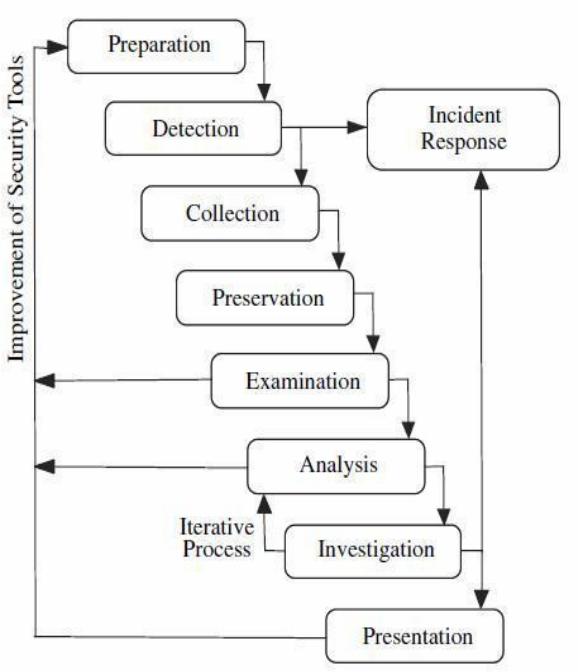
การพิสูจน์หลักฐานทางเครือข่าย (Network Forensics) เป็นส่วนหนึ่งที่สำคัญของการพิสูจน์หลักฐานทางดิจิทัล (Digital Forensics) ซึ่งเป็นหนึ่งในสี่หัวข้อของ Digital Forensic Research Workshop (DFRW) ครั้งที่ 1 ในปี ค.ศ. 2001 การอภิปรายในการประชุมเชิงวิชาการ ได้ก่อให้เกิดคำจำกัดความของการพิสูจน์หลักฐานทางเครือข่ายเช่นเดียวกันกับที่ใช้กันในการพิสูจน์หลักฐานในงานสำคัญด้านอื่น ๆ ที่เกี่ยวข้อง และอาจเป็นคำที่ใช้กันในการวิจัยขั้นพื้นฐานและขั้นประยุกต์ คำจำกัดความของการพิสูจน์หลักฐานทางเครือข่ายคือ การใช้เทคนิคที่พิสูจน์ได้ทางวิทยาศาสตร์ เพื่อเก็บรวบรวม รวมเข้าด้วยกันพิสูจน์ตัวตน สอบสวน การวิเคราะห์ และเก็บรักษาเป็นหลักฐานทางดิจิทัลจากหลาย ๆแหล่งที่ถ่ายทอดและดำเนินการอยู่ โดยมีจุดประสงค์เพื่อเปิดเผยเจตนาที่แท้จริงหรือประสบความสำเร็จในการกำหนดมาตรการสำหรับกิจกรรมที่ไม่ได้รับอนุญาต อาทิ การทำให้ระบบวุ่นวาย ทุจริต และหรือปรับเปลี่ยนส่วนประกอบของระบบ รวมทั้งจัดหาข้อมูลเพื่อช่วยการตอบสนองกับกิจกรรม หรือการกู้กลับคืนจากกิจกรรมเหล่านี้ อ้างถึงงานวิจัยและการนำไปปฏิบัติในปัจจุบันแล้วการพิสูจน์หลักฐานทางเครือข่ายจะเน้นเฉพาะเจาะจงไปที่การเก็บข้อมูลหลักฐานขณะก่ออาชญากรรม (Dynamic Forensics) และการป้องกันเชิงรุก

ในปัจจุบัน หนึ่งในหลายเหตุผลที่ทำให้ไม่สามารถนำบทลงโทษทางกฎหมายไปจัดการกับผู้ก่ออาชญากรรมดิจิทัลบนเครือข่ายคอมพิวเตอร์คือ การสิ้นเปลืองเวลาและความยากลำบากในการเก็บรวบรวมและการวิเคราะห์พยานหลักฐานในกระบวนการพิสูจน์หลักฐานทางเครือข่ายหรือการพิสูจน์หลักฐานการบุกรุก (Intrusion Forensics) นั้น หลักฐานจะได้มาจากการตรวจจับจากการจราจรในเครือข่ายหรือระบบการบันทึกล็อกของระบบ (System Log) ถึงแม้ว่าการวิเคราะห์ล็อก (Log) จะเป็นสิ่งสำคัญของกระบวนการพิสูจน์หลักฐานทางเครือข่าย แต่ก็มีงานวิจัยจำนวนน้อยมากที่เกี่ยวข้องกับการวิเคราะห์ล็อกแบบอัตโนมัติ ในปัจจุบันการวิเคราะห์ล็อกสำหรับกระบวนการพิสูจน์หลักฐานทางเครือข่ายยังคงเป็นแบบที่ใช้มนุษย์เป็นผู้วิเคราะห์และขึ้นอยู่กับประสบการณ์และทักษะของผู้วิเคราะห์ ซึ่งสูญเสียเวลามากและมีแนวโน้มผิดพลาด [2]

รูปที่ 1 แสดงรูปแบบกระบวนการพิสูจน์หลักฐานของเครือข่ายทั่วไป สำหรับการวิเคราะห์เครือข่ายขึ้นอยู่กับการพิสูจน์หลักฐานต่างๆ ที่มีอยู่แบบดิจิทัล [1] รายละเอียดของกระบวนการดังกล่าวมีดังนี้

2.1 การเตรียมการ (Preparation)

การพิสูจน์หลักฐานเครือข่ายสามารถใช้ได้กับสภาพแวดล้อมที่มีเครื่องรับรู้ (Sensor) เช่น ระบบตรวจจับการบุกรุก การวิเคราะห์แฟ้มข้อมูลไฟร์วอลล์ (Firewall) การปรับใช้ซอฟต์แวร์วัดอัตราการไหลของข้อมูล นำไปติดตั้งที่จุดที่สำคัญต่างๆ ในเครือข่าย การอนุมัติที่จำเป็นและใบสำคัญแสดงสิทธิที่จะได้รับตามกฎหมายเพื่อที่จะไม่ละเมิดความเป็นส่วนตัว



รูปที่ 1. แสดงกระบวนการพิสูจน์หลักฐานทางเครือข่าย [1]

2.2 การตรวจพบ (Detection)

การแจ้งเตือนที่เกิดจากเครื่องมือรักษาความปลอดภัยต่างๆ บ่งบอกถึงการละเมิดความปลอดภัยหรือการละเมิดนโยบายไม่ได้รับอนุญาตและความผิดปกติที่สังเกตเห็นจะได้รับการวิเคราะห์ การแสดงตนและธรรมชาติของการโจมตีจะถูกกำหนดจากตัวแปรต่างๆ การตรวจสอบจะกระทำอย่างรวดเร็วเพื่อประเมินและยืนยันการโจมตีที่น่าสงสัย ซึ่งจะอำนวยความสะดวกในการตัดสินใจที่สำคัญว่าจะดำเนินการต่อไปหรือไม่ อย่างไร เครื่องมือที่ช่วยในขอบข่ายในส่วนนี้ ได้แก่ TCPDump [7] Wireshark [8] pads [9] Sebek [10] ntop [11] P0f [12] Bro [13] และ Snort [14]

2.3 การตอบสนองต่อเหตุการณ์ที่เกิดขึ้น (Incident Response)

การตอบสนองต่อการเกิดอาชญากรรมหรือการบุกรุกจะสามารถตรวจพบได้ โดยที่จะต้องเริ่มจากข้อมูลที่รวบรวมเพื่อการตรวจสอบและประเมินเหตุการณ์ที่เกิดขึ้น การตอบสนองจะเกิดขึ้นได้ต้องขึ้นอยู่กับชนิดของการโจมตี ที่สามารถระบุและเกี่ยวข้องกับนโยบายขององค์กรด้านกฎหมายและข้อจำกัดของธุรกิจ แผนปฏิบัติการเกี่ยวกับวิธีการป้องกันการโจมตีในอนาคตและการกู้คืนจากความเสียหายที่มีอยู่จะเริ่มในขณะเดียวกันกับการตัดสินใจว่าจะดำเนินการต่อ โดยที่ข้อมูลที่ได้อาจต้องมีการกระทำบางอย่างในการควบคุมและลดการโจมตี

2.4 การจัดเก็บรวบรวม (Collection)

ข้อมูลที่ได้มาจากเครื่องรับรู้ที่ใช้ในการเก็บรวบรวมข้อมูลจากการเข้ามาใช้งาน ขั้นตอนที่กำหนดไว้อย่างดีโดยใช้เครื่องมือที่เชื่อถือได้และซอฟท์แวร์จะต้องอยู่ในสถานะที่จะรวบรวมหลักฐานได้ จำนวนข้อมูลที่ได้รับมาจากเครื่องรับรู้ จะต้องใช้พื้นที่หน่วยความจำขนาดใหญ่และระบบจะต้องสามารถจัดการกับรูปแบบที่แตกต่างกันของข้อมูลได้อย่างเหมาะสม เครื่องมือที่ช่วยในขอบข่ายในส่วนนี้ ได้แก่ TCPDump Wireshark TCPFlow [15] NfDump [16] pads Sebek และTCPReplay [17]

2.5 การเก็บรักษา (Preservation)

ข้อมูลดั้งเดิมที่ได้รับในรูปแบบของร่องรอยและการบันทึกจะถูกเก็บไว้ในอุปกรณ์สำรองข้อมูลที่สามารถอ่านได้อย่าง

เดียว การเก็บรักษาข้อมูลโดยใช้วิธีการแฮช (Hash) สามารถนำไปใช้ตรวจสอบความถูกต้องของข้อมูลที่ได้ทำสำเนาไปเพื่อใช้ในการวิเคราะห์ข้อมูลจราจรทางเครือข่ายกับข้อมูลดั้งเดิมได้ เครื่องมือที่ช่วยในขอบข่ายในส่วนนี้ ได้แก่ TCPDump Wireshark TCPFlow NfDump pads Sebek TCPReplay Bro และ Snort

2.6 การตรวจสอบ (Examination)

เบาะแสที่ได้จากเครื่องรับรู้ต่างๆ เมื่อมีการบูรณาการการรักษาความมั่นคงปลอดภัยและการรวมเข้าด้วยกันของชุดข้อมูลให้เป็นข้อมูลขนาดใหญ่ที่สามารถดำเนินการวิเคราะห์ มีปัญหาบางอย่างเช่น ข้อมูลที่ซ้ำซ้อนกัน จำเป็นต้องมีการจัดการในส่วนปัญหานี้ด้วย ในบางกรณีก็อาจจะมีการแจ้งเตือนจากแหล่งต่างๆ ที่ขัดแย้งกัน อย่างไรก็ตามขั้นตอนนี้ต้องทำเพื่อให้ข้อมูลที่สำคัญจากแหล่งข้อมูลต่าง ๆ จะได้ไม่สูญหาย หลักฐานที่เก็บรวบรวมอย่างมีระบบจะสามารถค้นหาเพื่อดึงตัวชี้วัดเฉพาะเจาะจงของอาชญากรรมได้ เครื่องมือที่ช่วยในขอบข่ายในส่วนนี้ ได้แก่ TCPDump Wireshark TCPFlow Flow-tool [18] NfDump และ pads

2.7 การวิเคราะห์ (Analysis)

ใช้การแบ่งประเภทและความสัมพันธ์กับรูปแบบการโจมตีที่ได้มีการศึกษามาก่อนแล้ว โดยมีแนวทางการทำฐานข้อมูลที่ใช้ในการเปรียบเทียบข้อมูลให้ตรงกับรูปแบบการโจมตี และอาศัยบางส่วนของตัวแปรที่เกี่ยวข้องกับการเชื่อมต่อเครือข่าย เช่น DNS packet fragmentation โพรโทคอลและระบบปฏิบัติการ เครื่องมือที่ช่วยในขอบข่ายในส่วนนี้ ได้แก่ TCPDump Wireshark TCPFlow Flow-tool NfDump และ pads

2.7 การสืบสวน (Investigation)

เป้าหมายคือ การตรวจสอบเส้นทางจากเครือข่ายที่ถูกโจมตีกลับไปยังจุดที่ของการโจมตี การดักจับแฟ้มข้อมูลและสถิติการใช้งานจะใช้ในการกำหนดลักษณะของการโจมตี ในขอบข่ายงานส่วนนี้อาจต้องใช้คุณลักษณะเพิ่มเติมบางส่วนจากขอบข่ายงานในส่วนของการวิเคราะห์และด้วยเหตุนี้ทั้งสองขอบข่ายงานจะทำงานร่วมจนมาถึงข้อสรุปที่จะสามารถบอกลักษณะของผู้บุกรุก ขอบข่ายงานในส่วนนี้เป็นส่วนที่ยากที่สุดของการพิสูจน์หลักฐานทางเครือข่ายคอมพิวเตอร์ วิธีการที่เรียบง่ายของผู้บุกรุกจะมีการซ่อนตัวหรือมีการใช้ไอพีแอดเดสที่หลอกลวง ขั้นตอนการตรวจสอบข้อมูลของการตอบสนองต่อเหตุการณ์ที่เกิดขึ้นเพื่อใช้ดำเนินคดีตามกฎหมายกับผู้บุกรุก

2.8 การนำเสนอ (Presentation)

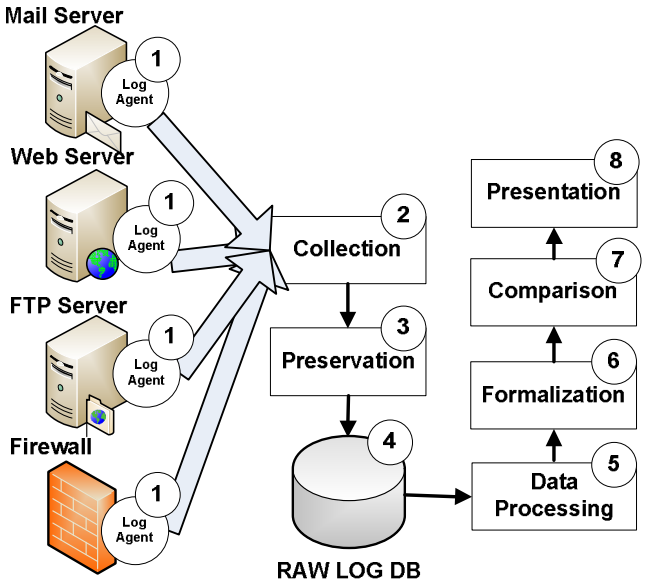
การนำเสนอจะมีการนำเสนอในภาษาที่บุคลากรทางกฎหมายเข้าใจได้ง่าย ในขณะที่คำอธิบายวิธีการต่างๆที่ใช้ในการนำเสนอ จะรวบรวมเป็นเอกสารของระบบเพื่อให้ตรงกับข้อกำหนดทางกฎหมาย ข้อสรุปที่รวบรวมเป็นเอกสารมีผลต่อการตรวจสอบในอนาคตและเพื่อใช้เป็นข้อเสนอแนะเพื่อเป็นแนวทางในการปรับใช้และปรับปรุงเครื่องในการรักษาความจัดการการจราจรของเครือข่ายแบบเรียลไทม์ (Real-time) ขั้นตอนการตรวจสอบจะช่วยในการค้นพบการโจมตีและระยะการเก็บรวบรวมแฟ้มข้อมูลของเครือข่ายสร้างความมั่นใจในความสมบูรณ์ของข้อมูลที่จะสามารถตอบสนองต่อเหตุการณ์ที่เกิดขึ้นอย่างเหมาะสม ข้อมูลที่ถูกสร้างขึ้นและคัดลอกจะทำในอยู่ในส่วนขอบข่ายของการเก็บรักษา (Preservation)

3. งานวิจัยที่นำเสนอ

ระบบเครือข่ายเป็นระบบที่ซับซ้อนด้วยอุปกรณ์ที่แตกต่าง การเข้าสู่ระบบ (Log on) ไปในอุปกรณ์ แต่ละครั้งเป็นการจำกัดให้เห็นเพียงความเป็นเจ้าของ แต่ล้มเหลวในการสะท้อนภาพทั้งหมดของกิจกรรมการบุกรุก เมื่อมนุษย์เป็นผู้วิเคราะห์ล็อกเราจะพบว่าล็อกจากหลาย ๆ แหล่ง จะก่อให้เกิดหลักฐานที่ครอบคลุมและมีประสิทธิภาพมากขึ้น ล็อกที่มีความสัมพันธ์กันจะถูกรวบรวมจากอุปกรณ์หลายแหล่งซึ่งช่วยเพิ่มความน่าเชื่อถือให้กับหลักฐานการบุกรุก ประการแรก ต้องแก้ปัญหาหลักฐานของระบบตรวจจับการบุกรุกหรือล็อก ของระบบอื่น ๆ ที่เกี่ยวข้องกับประเด็นการกระทำที่ผิดกฎหมาย

3.1 ภาพรวมของระบบ

การทำงานของระบบที่นำเสนอแบ่งการทำงานหลักของระบบออกเป็น 2 ส่วนด้วยกัน ส่วนแรกคือ การติดตั้งโปรแกรมเรียกว่า เอเจนต์ (Agent) ไว้ที่แหล่งข้อมูลต่าง ๆ ที่ต้องการเก็บรวบรวมล็อก (Log) และอีกส่วนคือ ระบบที่ทำหน้าที่รับข้อมูลล็อกและนำข้อมูลล็อกมาวิเคราะห์ จัดเก็บและรายงานข้อมูลที่ได้จากการวิเคราะห์ล็อก



รูปที่ 2. แสดงโครงร่าง (Framework) ของงานวิจัย

ในขั้นตอนที่ 1 ทำการติดตั้งเอเจนต์ไว้ที่แหล่งข้อมูลต่าง ๆ ที่ต้องการเก็บรวบรวมล็อก เพื่อใช้ในการรวบรวมล็อกเบื้องต้น ในขั้นตอนที่ 2 โมดูลนี้จะทำการรวมรวบล็อกที่ได้รับมาจาก ล็อกเอเจนต์ ที่ได้ติดตั้งไว้ในแหล่งข้อมูลหลายๆ แหล่งก่อนส่งไปยังโมดูลต่อไป ในขั้นตอนที่ 3 เป็นการรวบรวมข้อมูลและเก็บไว้ตามโครงสร้างที่ได้ออกแบบตำแหน่งที่เก็บข้อมูลและทำการเข้ารหัสเพื่อป้องการการแก้ไขและความถูกต้องของข้อมูลด้วยการบวนการของ AES (Advanced Encryption Standard) [19] และเก็บข้อมูลในแบบ RAW LOG ในขั้นตอนที่ 4 เป็นฐานข้อมูลที่ใช้เก็บรักษาล็อกที่เป็นต้นฉบับดั้งเดิม เพื่อใช้ในการตรวจสอบที่มาที่ไปของข้อมูลที่ได้รับมากับข้อมูลที่นำไปทำการวิเคราะห์หรือสืบสวนแล้วว่า มาจากข้อมูลต้นฉบับจริงหรือไม่ ในขั้นตอนที่ 5 จะทำการถอดรหัสลับของข้อมูล ที่ได้มีการเข้ารหัสลับไว้ในขั้นตอน 3 และมีการตรวจสอบความถูกต้อง รวมถึงการตรวจสอบการแก้ไขข้อมูลด้วย SHA-1 [20] sum ก่อนที่จะนำข้อมูลไปประมวลผลต่อไป ขั้นตอนที่ 6 เป็นการนำล็อกที่ได้จากขั้นตอนที่ 5 มากำหนดรูปแบบที่สามารถเข้าใจได้ง่าย โดยการนำเอาไปแทนที่ช่องว่างด้วยวิธีการ Base64 [21] เพื่อไม่ให้เกิดปัญหาเกี่ยวกับอักขระหรือสัญลักษณ์ในข้อมูลของล็อก ในขั้นตอนที่ 7 คือการนำล็อกที่อยู่ในฐานข้อมูลมาสร้างเป็นไฟล์ใหม่อีกรอบแล้วทำการตรวจสอบความถูกต้องของข้อมูลล็อกต้นฉบับด้วย SHA-1 sum แล้วนำค่าไปเปรียบเทียบกับค่า SHA-1 sum ที่ได้จากขั้นตอนที่ 5 และในขั้นตอนสุดท้าย ขั้นตอนที่ 8 หลังจากที่ได้ทำการวิเคราะห์ข้อมูลล็อกแล้ว จะนำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้น

3.2 การวิเคราะห์ระบบที่นำเสนอ

ผู้ทำการวิจัย จะทำการทดสอบโดยการติดตั้งเครื่องล็อกศูนย์กลาง (Central log) เข้ากับระบบเซิร์ฟเวอร์ (Server system) เพื่อเก็บรวบรวมล็อกจากแหล่งข้อมูลต่าง ๆ ของระบบเซิร์ฟเวอร์ มาทำการวิเคราะห์ข้อมูลล็อก เครือข่ายที่ใช้ทดสอบระบบมีรายละเอียดดังรูปที่ 3 ด้านล่างนี้

จากรูปที่ 3 แสดงระบบเครือข่ายที่ใช้ทดสอบระบบ ที่มีการเชื่อมต่ออินเทอร์เน็ตแบบกำหนดหมายเลขไอพีแอสเตรสแบบคงที่และแบบ ADSL ประกอบด้วยเครื่องเซิร์ฟเวอร์ที่ได้ติดตั้งเอเจนต์ สำหรับส่งข้อมูลล็อกไฟล์ไปยังเครื่องล็อกศูนย์กลางในการจัดการและวิเคราะห์ล็อกไฟล์ที่รับมาแล้วทำการจัดเก็บข้อมูลลงฐานข้อมูล

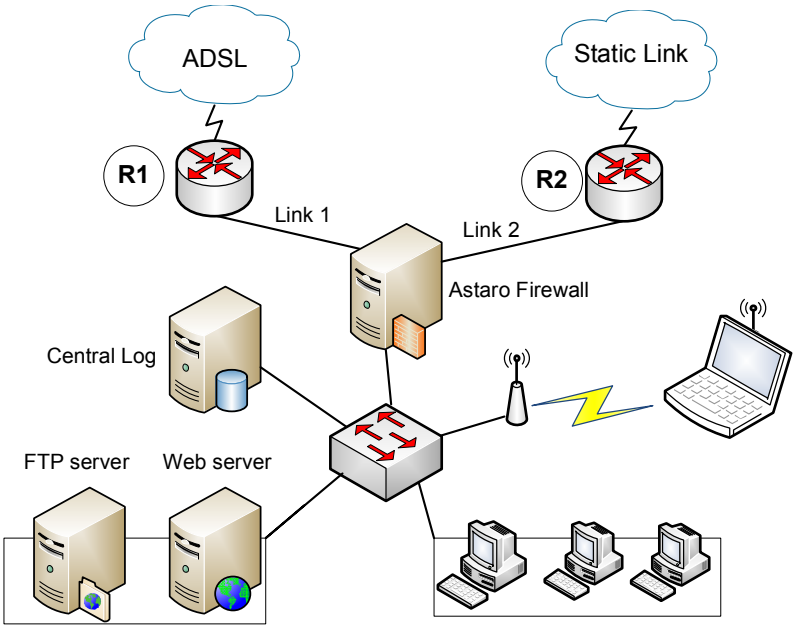
4. การพัฒนาระบบ

เพื่อเป็นการทดสอบว่าแนวคิดที่ได้นำเสนอในบทที่ 3 สามารถนำไปประยุกต์ใช้งานได้จริงกับสภาพแวดล้อมของเครือข่ายโดยทั่วไปนั้น ผู้วิจัยได้ทำการพัฒนาระบบบนระบบเซิร์ฟเวอร์ที่มีการใช้งานอยู่จริง เพื่อทำให้เกิดข้อมูลล็อก จากการเข้ามาใช้งานในระบบเซิร์ฟเวอร์ โดยที่จะทำการวิเคราะห์ล็อกในส่วนของการเข้ามาใช้งานเว็ปไซต์, การโอนถ่ายข้อมูล รวมถึงล็อกของระบบการป้องกันการบุกรุก เช่น การสแกนพอร์ตและการคัดกรองข้อมูลต่างๆ ที่จะเข้ามาในระบบเซิร์ฟเวอร์

4.1 รายละเอียดของระบบที่พัฒนาขึ้น

ระบบวิเคราะห์ล็อกแบบอัตโนมัติที่ได้พัฒนาขึ้น ได้ทำการติดตั้งระบบพร้อมกับระบบเวปเซิร์ฟเวอร์และเอฟทีพีเซิร์ฟเวอร์ โดยมีรายละเอียดดังนี้

- Router	2 หน่วย
- Firewall	1 หน่วย
- FTP and Web Server	1 หน่วย
- Central Log	1 หน่วย
- Access Point	1 หน่วย
- Switch 3 com	1 หน่วย



รูปที่ 3. แสดงระบบที่ใช้ในการทดลอง

Router 2 หน่วย

R1. ADSL router dynamic IP
- Link bandwidth 10Mbps
- PPPOE mode

R2. ADSL router static IP
- Link bandwidth 2Mbps
- Bridge mode

Astaro Firewall 1 หน่วย

1. Interface eth1 (Link 1)
   Mode : NAT
   Type : Ethernet Standard
2. Interface eth2 (Link 2)
   Mode : NAT
   Type : DSL (PPPOE)
3. Interface eth0 (Internal Network)
   Type : Ethernet Standard

4. Policy of Firewall

Packet Filter

Source	Internal Network
Services	23, 25, 53, 443, 110
Destination	Any

Network Address Translation (NAT)

Traffic Source	any
Traffic Service	80,443
Traffic Destination	110.77.147.201
NAT mode	DNAT
Destination	Web Server

Proxy web cache

Mode	Transparent
Source	Internal Network
Services	http, ftp

FTP and Web Server 1 หน่วย

OS	Ubuntu 10.04
Service	FTP (VSFTP)
	httpd, https
Domain name	www.bnsc.ac.th

Central Log 1 หน่วย

OS	CentOS 5.5
Service	Syslog-ng

Access Point 1 หน่วย

Product	Linksys wrt54gl
Mode	AP Mode

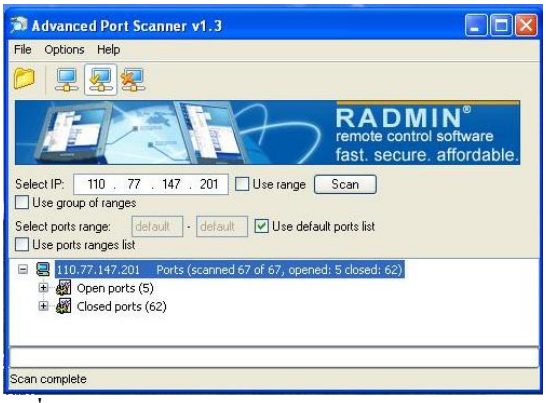
Switch 3 com 1 หน่วย

Product	HP V1405-16G Switch
Role	Layer 2 Access switch

4.2 ผลการทดลอง

ในการทดลองระบบที่ได้พัฒนาขึ้น เราได้ทดลองบนระบบเครือข่ายจริง เพื่อประเมินประสิทธิภาพในการวิเคราะห์ข้อมูลล็อก ที่เป็นแบบอัตโนมัติตามวิธีการที่ได้นำเสนอไปก่อนหน้านี้ โดยการทดลองจะทำการสแกนพอร์ต (Port scan) จากเครือข่ายภายนอกเข้ามายังเครือข่ายภายในและจากเครือข่ายภายในออกไปยังเครือข่ายภายนอกด้วย

จากรูปที่ 4 แสดงการใช้โปรแกรม Advanced Port Scanner v1.3 ที่ใช้ในการทดสอบการทำการสแกนพอร์ตทั้งจากเครือข่ายภายนอกเข้ามายังเครือข่ายภายในและจากเครือข่ายภายในออกไปยังเครือข่ายภายนอก



รูปที่ 4. แสดงโปรแกรมการทำสแกนพอร์ต

ผลการทดลองที่ได้ช่วยให้เราสามารถที่จะเข้าใจความหมายในแต่ละส่วนของล็อกที่เกิดขึ้นได้โดยง่ายและสะดวกรวดเร็ว โดยบ่งบอกถึง จำนวนล็อกที่เกิดขึ้นหรือจำนวนครั้งที่มีการสแกนพอร์ต วันเวลา รูปแบบที่ตรวจจับได้ว่าเป็นการสแกนพอร์ต หรือบ่งบอกถึงไอพีแอดเดรสต้นทางและปลายทางของการสแกนพอร์ต เป็นต้น

Astaro Intrusion Prevention Systems (IPS)

No.	Date	Period
684	2-Apr-2011 21:13:2	04:02-21:13:26
683	2-Apr-2011 21:13:2	04:02-21:13:26
682	2-Apr-2011 12:39:0	04:02-12:39:02
681	1-Apr-2011 17:42:3	04:01-17:42:35
680	1-Apr-2011 17:41:5	04:01-17:41:59
679	1-Apr-2011 17:41:5	04:01-17:41:56
678	1-Apr-2011 17:34:3	04:01-17:34:39
677	1-Apr-2011 17:34:3	04:01-17:34:38
676	1-Apr-2011 17:34:3	04:01-17:34:37
675	1-Apr-2011 17:34:3	04:01-17:34:36
674	1-Apr-2011 17:32:5	04:01-17:32:56

รูปที่ 5. แสดงวันเวลาการสแกนพอร์ต

จากรูปที่ 5 แสดงวันเวลาของตารางล็อกการสแกนพอร์ตที่ได้วิเคราะห์ข้อมูลล็อกแล้ว นำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้น

Log Report

Source	Destination
124.122.195.178	192.168.168.2
124.122.195.178	192.168.168.2
203.209.53.238	192.168.168.2
192.168.168.154	216.254.119.37
192.168.168.154	178.156.140.107
192.168.168.154	183.83.202.163
192.168.168.154	89.37.111.132
192.168.168.154	117.211.83.179
192.168.168.154	78.152.169.41
192.168.168.154	77.242.227.4
192.168.168.154	125.19.148.172

รูปที่ 6. แสดงไอพีแอดเดรสการทำสแกนพอร์ต

จากรูปที่ 6 แสดงไอพีแอดเดรสต้นทางและปลายทาง การสแกนพอร์ตที่ได้วิเคราะห์ข้อมูลล็อกแล้ว นำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้น

Catogory	Detect	Action
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan
ips	portscan detected	portscan

รูปที่ 7. แสดงรูปแบบของการสแกนพอร์ต

จากรูปที่ 7 แสดงรูปแบบ การสแกนพอร์ตที่ได้วิเคราะห์ข้อมูลล็อกแล้ว นำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้น

IN ITF	OUT ITF	SRC Port	DST Port
ppp0	eth0	1481	22
ppp0	eth0	1480	21
ppp0	eth0	55926	21
eth0	eth1	1794	6679
eth0	eth1	1759	6063
eth0	eth1	1758	6738
eth0	eth1	1403	6823
eth0	eth1	1402	7611
eth0	eth1	1401	6636
eth0	eth1	1400	8115
eth0	eth1	1317	5962

รูปที่ 8. แสดงอินเตอร์เฟสและพอร์ทของการสแกนพอร์ต

จากรูปที่ 8 แสดงอินเตอร์เฟสและพอร์ท การสแกนพอร์ตที่ได้วิเคราะห์ข้อมูลล็อกแล้ว นำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้น

SRC MAC	DST MAC	TCI
0:15:17:35:df:18		SYN
0:15:17:35:df:18		SYN
0:15:17:35:df:18		SYN
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	
0:15:af:e6:91:8f	0:15:17:35:df:18	

รูปที่ 9. แสดง MAC ของการสแกนพอร์ต

จากรูปที่ 9 แสดงค่าแม็คแอดเดรสต้นทางและปลายทาง การสแกนพอร์ตที่ได้วิเคราะห์ข้อมูลล็อกแล้ว นำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้น

5. การวิเคราะห์และอภิปราย

5.1 การวิเคราะห์ความมั่นคงปลอดภัย

งานวิจัยฉบับนี้เป็นการติดตั้งเครื่องล็อกศูนย์กลางเข้ากับระบบเซิร์ฟเวอร์ที่มีอยู่แล้ว จึงทำให้ระบบมีความมั่นคงปลอดภัยเทียบเท่ากับระบบเซิร์ฟเวอร์เดิม ส่วนข้อมูลที่มีการจัดเก็บไว้ในเครื่องล็อกศูยน์กลาง มีการป้องกันการเข้าถึงข้อมูลและมีการยืนยันสิทธิตัวบุคลในการเข้าถึงข้อมูลและความถูกต้องของข้อมูลด้วย จึงทำให้ข้อมูลมีความมั่นคงปลอดภัยด้วยเช่นกัน

5.2 การนำข้อมูลไปใช้เพื่อพิสูจน์หลักฐาน

ข้อมูลที่ได้ทำการวิเคราะห์แล้วจะถูกจัดเก็บลงในฐานข้อมูล และรายงานข้อมูลที่ได้จากการวิเคราะห์ล็อกสามารถนำไปตรวจสอบความถูกต้องกับข้อมูลต้นฉบับได้ เพื่อยืนยันว่าเป็นการนำข้อมูลต้นฉบับดังกล่าวมาใช้ในการวิเคราะห์ข้อมูลจริงหรือไม่ จากนั้นจึงนำข้อมูลไปใช้ประกอบเพื่อพิสูจน์หลักฐานหาตัวผู้กระทำผิดต่อไป

5.3 การวิเคราะห์ประสิทธิภาพ

ระบบที่ได้นำเสนอในงานวิจัยฉบับนี้มีประสิทธิภาพที่ดีกว่า ระบบที่ถูกนำเสนอโดย Lin *et al.* [2] เนื่องจากมีการป้องกันปัญหา การแก้ไขหรือลบข้อมูลต้นฉบับจากผู้ที่ไม่หวังดีหรือจากผู้ก่ออาชญากรรม รวมถึงสามารถตรวจสอบที่มาที่ไปของข้อมูลที่ได้รับมากับข้อมูลที่นำไปทำการวิเคราะห์หรือสืบสวนแล้วว่า มาจากข้อมูลต้นฉบับจริงหรือไม่

6. บทสรุป

งานวิจัยฉบับนี้นำเสนอการออกแบบและพัฒนาระบบการวิเคราะห์ล็อกแบบอัตโนมัติ เพื่อใช้ในการเก็บรวบรวมและวิเคราะห์ล็อกได้รวดเร็วยิ่งขึ้น เนื่องจากในปัจจุบันการเก็บรวบรวมและวิเคราะห์ล็อกยังคงเป็นแบบที่ใช้มนุษย์เป็นผู้วิเคราะห์และขึ้นอยู่กับประสบการณ์และทักษะของผู้วิเคราะห์ ซึ่งสูญเสียเวลามากและมีแนวโน้มผิดพลาด

งานวิจัยฉบับนี้สามารถนำไปใช้งานร่วมกับระบบเซิร์ฟเวอร์ที่ใช้ในปัจจุบันได้ โดยติดตั้งเครื่องล็อกศูนย์กลางเพิ่มเข้าไปในระบบและติดตั้งเอเจนต์ไว้ที่แหล่งข้อมูลต่าง ๆ ที่ต้องการเก็บรวบรวมล็อก

แนวทางในการพัฒนางานวิจัยต่อไปนั้น จะทำการพัฒนาระบบให้สามารถวิเคราะห์ล็อคในรูปแบบอื่นๆ ได้อีก เช่น การทำ IPSec VPN, SSL VPN หรือการบุกรุกเข้ามาในระบบรูปแบบอื่นๆ เป็นต้น เพื่อจะนำข้อมูลล็อกมาแสดงผลในรูปแบบของตารางที่สามารถเข้าใจความหมายของแต่ละส่วนของล็อกที่เกิดขึ้นได้โดยง่าย

เอกสารอ้างอิง

[1] E. S. Pilli, R. C. Joshi, and R. Niyogi. Network Forensic Framework: Survey and Research Challenges. Digital Investigation (2010), doi:10.1016/j.diin.2010.02.003..

[2] C. Lin, L. Zhitang, and G. Cuixia. Automated Analysis of Multi-source Logs for Network Forensics. 2009 First International Workshop on Education Technology and Computer Science.

[3] B. J. Nikkel. Generalizing Sources of Live Network Evidence. Digital Investigation (2005) 2, 193-200..

[4] Bruce J. Nikkel. A portable network forensic evidence collector. Digital investigation (2006) 3, 127–135.

[5] Hou Ming, Shen LiZhong. A New System Design of Network Invastion Forensics. 2009 Second International Conference on Computer and Electrical Engineering.

[6] ณัฐพล กิตติรุ่งเรือง, พงษ์สุรีย์ ลิ้มมณีวิจิตร, ศุภกร กังพิศดาร.การออกแบบและพัฒนาล็อคเซิร์ฟเวอร์สำหรับระบบปฏิบัติการวินโดวส์

[7] Tcpdump, http://www.tcpdump.org/

[8] Wireshark, http://www.wireshark.org/

[9] pads, http://www.mentor.com/products/pcb-system-design/design-flows/pads/

[10] Sebek, http://www.honeynet.org/project/sebek

[11] ntop, http://www.ntop.org/news.php

[12] P0f, http://freshmeat.net/projects/p0f/

[13] Bro, http://www.bro-ids.org/

[14] Snort, http://www.snort.org/

[15] TCPFlow, http://www.circlemud.org/jelson/software/tcpflow

[16] NfDump, http://sourceforge.net/projects/nfdump/

[17] TCPReplay, http://tcpreplay.synfin.net/trac/

[18] Flow-tools, http://www.splintered.net/sw/flow-tools

[19] AES (Advanced Encryption Standard), http://en.wikipedia.org/wiki/Advanced_Encryption_Standard

[20] SHA-1, http://en.wikipedia.org/wiki/SHA-1

[21] Base64, http://en.wikipedia.org/wiki/Base64