การประยุกต์ใช้เทคโนโลยีการระบุสิ่งผิดปกติในระบบตรวจจับการบุกรุกเครือข่าย

Application of Anomaly Detection Technology in Network Intrusion Detection System

ณัฐกานต์ เอี่ยมอ่อน[1] ทศพล บุญเกิน[2] และ นที ปั้นทอง[2]

[1]สำนักวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยแม่ฟ้าหลวง

[2]กองวิชาคณิตศาสตร์และคอมพิวเตอร์ กองการศึกษา โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช

E-Mail: nt.iamon@gmail.com, tossapon_b@rtaf.mi.th, Npantong@rtaf.mi.th

บทคัดย่อ

สถานการณ์ในปัจจุบันอันเกี่ยวเนื่องกับการรักษาความมั่นคงปลอดภัยไซเบอร์ ทั้งภายในประเทศและเหตุการณ์โจมตีเครือข่ายขององค์กรในประเทศต่างๆ ทั่วโลก เป็นการยืนยันถึงภัยคุกคามประเภทใหม่นี้ อีกทั้งยังมีผลกระทบต่อทรัพยากรขององค์กร และส่วนบุคคล ปัญหาดังกล่าวมีแนวโน้มที่จะทวีความรุนแรงยิ่งขึ้น โดยเฉพาะในช่วงการขับเคลื่อนประเทศไทยสู่ยุคเศรษฐกิจดิจิตอล ภาครัฐได้ตระหนักและให้ความสำคัญกับการแก้ไขปัญหาอาชญากรรมนี้อย่างจริงจัง โดยมีการร่างยุทธศาสตร์การวิจัยปี พ.ศ.2556-2560 ในรายประเด็นการรักษาความมั่นคงปลอดภัยไซเบอร์ เป็นแนวทางสนับสนุนงานวิจัยและการพัฒนานวัตกรรม ที่เป็นประโยชน์ต่อสาธารณะ สามารถแก้ไขปัญหาอาชญากรรมไซเบอร์รูปแบบต่างๆ ได้อย่างเป็นรูปธรรมและยั่งยืน หนึ่งในงานวิจัยที่ได้รับการศึกษาอย่างกว้างขวางคือ ระบบตรวจจับการบุกรุกเครือข่าย (network intrusion detection system: NIDS) ซึ่ง ทำหน้าที่วิเคราะห์หน่วยข้อมูลการสื่อสารที่เกิดขึ้น ในเครือข่ายเพื่อหารูปแบบที่สื่อถึงการบุกรุก จากนั้นแจ้งเตือนให้ผู้ใช้ได้ตัดสินตกลงใจต่อไป ระบบส่วนใหญ่ที่ใช้งาน ในปัจจุบันจะรองรับเฉพาะกรณีของรูปแบบที่ได้มีการบันทึกแล้วในอดีต แต่ยังไม่สามารถตรวจจับการโจมตีรูปแบบใหม่ได้ จึงเกิดแนวคิดการประยุกต์ใช้หลักการระบุสิ่งผิดปกติขึ้นเพื่อเพิ่มความยืดหยุ่นของระบบต่อการเปลี่ยนแปลงของสภาพแวดล้อม บทความนี้มีจุดมุ่งหมายที่จะนำเสนอทฤษฎีและความรู้พื้นฐานของการโจมตี ระบบตรวจจับ การบุกรุกและลักษณะของการประยุกต์ใช้การระบุสิ่งผิดปกติเพื่อแก้ไขปัญหาข้างต้น นอกจากนั้นยังได้นำเสนอทิศทางการวิจัยที่จะเกิดขึ้นในอนาคตพร้อมความสอดคล้องของงานวิจัยที่จะเกิดขึ้นกับนโยบายของภาครัฐ ตามลำดับ

คำสำคัญ : ระบบตรวจจับการบุกรุกเครือข่าย, การระบุสิ่งผิดปกติ, การจัดกลุ่มข้อมูล, การรวบรวมผล วิเคราะห์ย่อย

Abstract

Given the rise of recent events related to cyber security, both in Thailand and other countries around the globe, such a threat is imminent, with undesired impacts on organizational and personal resources. The intensity of this problem is likely to increase, especially during the period of promoting Thai digital economy. The government has taken this seriously, as shown by the drafting of national

research strategy 2013-2017 that includes the aforementioned issue. This is to set a guideline for research and innovation development to resolve cyber-security problems. One of the major subjects being investigated widely is a network intrusion detection system or NIDS. In a nutshell, it analyzes network-traffic information to identify possible acts of attack. However, most of the systems developed thus far have focused on the known attack patterns, whist lacking the capability to disclose new or unknown threats. In response, anomaly detection is adopted to provide the flexibility to NIDS. This article is set to provide the review of on network intrusion, NIDS and different applications of anomaly detection to the problem. In addition, it presents the perspective of future research, and its remedy in accordance with the governmental policy.

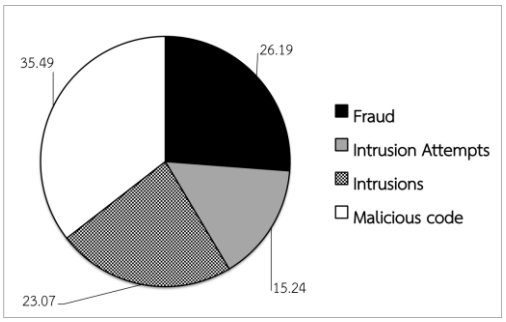
Keywords : intrusion detection system, anomaly detection, data clustering, ensemble.

1. บทนำ

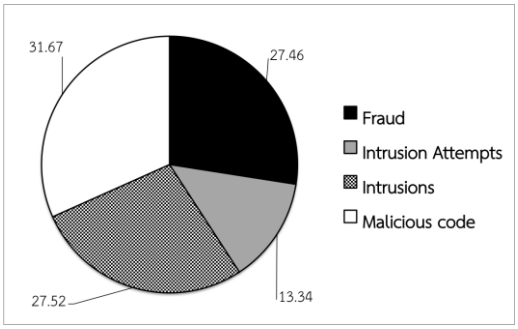
จากผลการสำรวจของสำนักงานสถิติแห่งชาติ [1] พบว่าในปี พ.ศ.2558 ประเทศไทยมีผู้ใช้งานเครือข่ายอินเทอร์เน็ต (มีอายุ 6 ปีขึ้นไป) จำนวน 24.6 ล้านคน คิดเป็นร้อยละ 39.3 ของประชากร ทั้งหมดในประเทศ และในกลุ่มผู้ใช้งานอินเทอร์เน็ตทั้งในประเทศ และต่างประเทศที่มีจำนวนมาก ได้มีบุคคลที่เรียกว่า “อาชญากร ไซเบอร์” ที่แฝงตัวเข้ามาใช้งานระบบเพื่อเข้าถึง ข้อมูลขององค์กร หรือข้อมูลส่วนบุคคล แล้วทำการเปลี่ยนแปลงหรือกระทั่งทำลาย ข้อมูลและอุปกรณ์ต่างๆ ในเครือข่าย รวมถึงการใช้พื้นที่บนโลกไซเบอร์เพื่อแสวงหาผลประโยชน์อย่างผิดกฎหมาย มีการเปิดเผยผลสำรวจความเสียหายจากอาชญากรรมไซเบอร์ที่เกิดขึ้นทั่วโลก ในปี พ.ศ.2557 ว่ามีมูลค่าเฉลี่ยราว 4 แสนล้านดอลลาร์สหรัฐ [2] และยังได้รายงานอีกว่ากว่า 3,000 บริษัทในประเทศสหรัฐอเมริกา ถูกโจมตี นอกจากนั้นกว่า 300,000 เว็บไซต์ในประเทศอินเดียถูกบุกรุกระหว่างปี พ.ศ.2555 และ 2557 อาชญากรไซเบอร์ส่งผลให้จำนวนการโจมตีทางอินเทอร์เน็ตเพิ่มสูงขึ้นถึงร้อยละ 81.3 โดยที่ องค์กรขนาดใหญ่ (พนักงานมากกว่า 2,500 คน) จะถูกโจมตีมากที่สุด มีการตรวจพบการโจมตีและการป้องกันเฉลี่ย 37 ครั้งต่อวัน ทั้งนี้ธุรกิจอาชญากรรม ไซเบอร์ทั่วโลกมีมูลค่ารวมกันมากกว่าหนึ่งล้านล้านเหรียญสหรัฐ ทั้งนี้อุตสาหกรรมที่ผิดกฎหมายดังกล่าวมีภูมิภาคเอเชียเป็นศูนย์กลาง

สถานการณ์ในปัจจุบัน รวมถึงแนวโน้มการเปลี่ยนแปลงของ สภาพแวดล้อมที่เกี่ยวกับประเด็นด้านการรักษาความมั่นคงปลอดภัยไซเบอร์ พบว่าในช่วงหลายปีที่ผ่านมา รัฐบาลได้ตระหนักและให้ความสำคัญกับการแก้ไขปัญหาอาชญากรรมนี้อย่างจริงจัง โดยมีการร่างยุทธศาสตร์การวิจัยปี พ.ศ.2556-2560 ในรายประเด็น ด้านการรักษาความมั่นคงปลอดภัยไซเบอร์ ตอบรับการขับเคลื่อนสู่ยุคเศรษฐกิจดิจิตอล (digital economy) เป็นแนวทางสนับสนุน งานวิจัยรวมถึงการพัฒนานวัตกรรมที่เป็นประโยชน์ต่อสาธารณะ แก้ไขปัญหาอาชญากรรมไซเบอร์รูปแบบต่างๆ เช่น ปัญหาการค้า ภาพอนาจารบนอินเทอร์เน็ต ปัญหาการล่อลวงเด็กและเยาวชนผ่านการสื่อสารทางอินเทอร์เน็ต ปัญหาโจรกรรมข้อมูลบัตรเครดิต ปัญหาการก่อกวนระบบเครือข่ายตลาดหลักทรัพย์ ปัญหาการโจมตีเว็บไซต์หน่วยงานของรัฐและสถาบันทางการเงิน นอกจากนั้นยังได้จัดตั้งหน่วยงานที่ดูแลด้านการรักษาความปลอดภัยไซเบอร์ในประเทศไทย ได้แก่ ศูนย์ประสานการรักษาความมั่นคงปลอดภัยระบบคอมพิวเตอร์ในประเทศ (Thailand Computer Emergency Response Team หรือ ThaiCERT) สำนักงานพัฒนาธุรกรรมทาง

อิเล็กทรอนิกส์ (องค์การมหาชน) กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร โดยหน่วยงานข้างต้นได้ร่วมกันศึกษาข้อมูลและแจ้งเตือนการโจมตีแบบใหม่ ๆ อย่างต่อเนื่อง สถิติอาชญากรรมไซเบอร์ในปี พ.ศ.2558 และ 2559 (จนถึงเดือนพฤษภาคม) สรุปได้ตามรูปที่ 1 และ 2 ตามลำดับ



รูปที่ 1 สัดส่วนของอาชญากรรมไซเบอร์ประเภทต่าง ๆ ในช่วงปี พ.ศ.2558 จากเหตุการณ์โจมตีทั้งสิ้น 4,356 ครั้ง (แหล่งข้อมูล www.thaicert.or.th/statistics/statistics.html)



รูปที่ 2 สัดส่วนของอาชญากรรมไซเบอร์ประเภทต่างๆ ในช่วงปี พ.ศ.2559 จากเหตุการณ์โจมตีทั้งสิ้น 1,806 ครั้ง (แหล่งข้อมูล www.thaicert.or.th/statistics/statistics.html)

จากข้อมูลในภาพทั้งสองนั้น อาชญากรรมไซเบอร์ 4 ประเภทที่มีการรายงานอยู่เป็นประจำ ได้แก่ การปลอมแปลงข้อมูล (fraud) ความพยายามเพื่อบุกรุกเครือข่าย (intrusion attempt) การรุกล้ำเครือข่าย (intrusion) และโปรแกรมที่อันตราย (malicious code) โดยจะสังเกตได้ว่าการบุกรุกเครือข่าย ไม่ว่าจะเป็นความพยายามเจาะระบบหรือการแทรกซึมเข้าระบบได้นั้น เป็นปัญหาที่มีความสำคัญ ดังสัดส่วนที่สูงที่สุดคือร้อยละ 38.31 ในปี พ.ศ.2558 และเพิ่มขึ้นเป็น 40.86 ในปีต่อมา ด้วยเหตุนี้องค์กรต่าง ๆ จึงจำเป็นต้องให้ความสนใจกับภัยคุกคามประเภทนี้ที่อาจส่งผลร้ายแรงต่อทั้งข้อมูล ทรัพย์สินทางปัญญา ทรัพยากรอื่น ๆ รวมถึงความน่าเชื่อถือขององค์กร เทคโนโลยีรักษาความปลอดภัยต่าง ๆ เช่น อุปกรณ์ไฟร์วอลล์ (firewall) มีประสิทธิภาพป้องกันเครือข่ายเพียงแค่ในระดับหนึ่ง หากไม่สามารถปรับเปลี่ยนให้รองรับการโจมตีรูปแบบใหม่ ๆ ได้ ผู้ดูแลระบบเครือข่ายและนักวิจัยในด้านการป้องกันการบุกรุกจึงได้ให้ความสนใจกับนวัตกรรมที่มีชื่อเรียกว่าระบบตรวจจับการบุกรุก (intrusion detection system: IDS) ซึ่งเป็นการตรวจสอบลักษณะของการสื่อสารเครือข่ายที่เกิดขึ้น พร้อมทั้งแจ้งเตือนเมื่อระบุได้ว่าการสื่อสารนั้นมีรูปแบบ หรือมีความน่าจะเป็นสอดคล้องกับการบุกรุก

ระบบตรวจจับการบุกรุกส่วนใหญ่ถูกออกแบบในลักษณะของซอฟท์แวร์ ผลิตและจัดจำหน่ายโดยบริษัทชั้นนำต่าง ๆ เช่น บริษัทซิสโก้ที่นำเสนอระบบในราคาที่มากว่า 500,00 บาท นอกจากค่าใช้จ่ายที่สูงนั้น ระบบตรวจจับการบุกรุกส่วนใหญ่ยังอยู่บนพื้นฐานของการอ้างอิงกับรูปแบบการโจมตีที่มีการบันทึกไว้แล้ว (known attack patterns or signatures) เท่านั้น ไม่สามารถตรวจจับการโจมตีในรูปแบบใหม่ได้ วิวัฒนาการของการโจมตีนั้นเป็นแรงผลักดันของงานวิจัยจำนวนมากในปัจจุบัน ซึ่งมุ่งพัฒนาเทคนิคที่ปรับเปลี่ยนการตรวจจับได้ตามการโจมตีที่เกิดขึ้นในอนาคต ตามแนวทางการตรวจจับเหตุการณ์หรือสิ่งผิดปกติ (anomaly detection) รองรับการตรวจจับเหตุการณ์ที่ส่อพิรุธของการสื่อสารจากทั้งภายนอก (outsider) รวมทั้งภายในเครือข่าย (insider)

จากข้อมูลที่ระบุข้างต้น บทความนี้มีจุดมุ่งหมายที่จะประมวลความรู้พื้นอันเกี่ยวกับระบบตรวจจับการบุกรุก ประเภทการบุกรุกที่พบในปัจจุบัน คุณลักษณะและความสามารถของระบบ รวมถึง

วิธีการที่นิยมใช้ในการพัฒนาระบบ โดยเฉพาะการประยุกต์ใช้การระบุสิ่งผิดปกติกับปัญหาดังกล่าวเป็นแนวทางสำหรับการพัฒนาแนวคิดหรือนวัตกรรมต่อไป โดยเนื้อหาของบทความได้ถูกจัดเรียงดังนี้ หัวข้อที่ 2 จะได้รวบรวมและนำเสนอความรู้พื้นฐานของการบุกรุกเครือข่าย ซึ่งจะได้ขยายผลสู่คุณลักษณะและประเภทของระบบตรวจจับการบุกรุกในหัวข้อที่ 3 การประยุกต์ใช้การระบุสิ่งผิดปกติเพื่อพัฒนาระบบข้างต้นนั้นจะอธิบายไว้ในหัวข้อที่ 4 และตามด้วยบทวิเคราะห์ถึงวิถีการวิจัยต่อไปในอนาคต อีกทั้งบทสรุป ในหัวข้อที่ 5 ตามลำดับ

2. การบุกรุกเครือข่าย

จากวิวัฒนาการของเทคโนโลยีสารสนเทศโดยทั่วไป ระบบเครือข่ายคอมพิวเตอร์และอินเทอร์เน็ต รวมทั้งสถิติของการโจมตีหรือบุกรุกเครือข่ายที่เพิ่มมากขึ้นในช่วงเวลาที่ผ่านมา มีผลทำให้การตรวจจับการบุกรุกเครือข่ายได้กลายประเด็นการวิจัยที่มีความสำคัญยิ่งทั้งต่อทรัพยากรขององค์กรและส่วนบุคคล ถึงแม้ว่าจะมีงานวิจัยจำนวนมากที่ได้นำเสนอแนวคิดและผลการทดลองกับข้อมูลต่าง ๆ ที่อธิบายลักษณะการใช้งานเครือข่าย หากยังมีหัวข้อปัญหาที่น่าสนใจอยู่อีกจำนวนมากที่รอการศึกษาและเผยแพร่ต่อสาธารณะในอนาคต [3] ทั้งนี้การบุกรุกที่กล่าวมาได้ถูกนิยามว่าเป็นการกระทำโดยตั้งใจที่จะเข้าถึงหรือแก้ไขข้อมูลโดยไม่ได้รับอนุญาต ยังรวมถึงการกระทำที่ส่งผลให้ระบบเกิดข้อผิดพลาดไม่สามารถดำเนินงานได้ [4] ตัวอย่าง ประเภทการบุกรุกที่พบได้บ่อยคือการโจมตีด้วยวิธี DoS (denial of service) ที่พยายามจะลดทอนประสิทธิภาพและทรัพยากรของระบบด้วยการส่งข้อมูลที่มีปริมาณมหาศาลเข้าไปที่เป้าหมาย ทำให้การรับส่งข้อมูลเกิดสภาวะคอขวด (congestion) จนไม่สามารถติดต่อสื่อสารกับผู้ใช้งานทั่วไปได้ นอกจากนั้นการโจมตีด้วยโปรแกรมประเภท worm หรือ virus เพื่อเข้าควบคุมและใช้งานทรัพยากรต่าง ๆ ภายในเครือข่าย เป็นอีกหนึ่งความเสี่ยงพื้นฐานในการรักษาความปลอดภัย

อนึ่งการตรวจจับการบุกรุกจากสิ่งหรือเหตุการณ์ที่ผิดปกติเป็นคำจำกัดความปัญหาของการสืบค้นรูปแบบการใช้งานหรือสื่อสารในเครือข่ายที่ผิดแปลกไปจากพฤติกรรมปกติ [5][6] เป็นการประยุกต์ใช้หลักการตรวจจับสิ่งผิดปกติ (anomaly or outlier detection) กับการป้องกันเครือข่าย ซึ่งหลักการข้างต้นยังได้รับความนิยมใช้ในงานประเภทอื่น ๆ อีก เช่น การปลอมแปลงข้อมูลบัตรเครดิต (fraud detection for credit cards) และการตรวจการณ์พฤติกรรมของกำลังทหารฝ่ายตรงข้าม (military surveillance) เป็นต้น กลุ่มนักวิจัยทางด้านสถิติได้ศึกษาปัญหาการตรวจจับสิ่งผิดปกติตั้งแต่ศตวรรษที่ 19 [7] จากนั้นในช่วงยี่สิบถึงสามสิบปีที่ผ่านมา การแก้ปัญหาด้วยหลักการเรียนรู้ของเครื่อง (machine learning) ได้เข้ามามีบทบาทเด่นชัด โดยมีการพัฒนาเทคนิคจำนวนมากเพื่อการตรวจจับการบุกรุกจากเหตุการณ์ที่ผิดปกติ ทั้งในบริบทการใช้งานและคุณลักษณะ เฉพาะระบบ รวมถึงบริบทการแก้ปัญหาทั่วไป [8][9] ก่อนที่จะได้ทำความเข้าในเรื่องดังกล่าว ความรู้พื้นฐานเกี่ยวกับประเภทของการบุกรุกก็เป็นสิ่งที่ต้องให้ความสำคัญในช่วงแรก

ผู้บุกรุกสามารถจำแนกได้เป็นสองกลุ่มคือ ผู้บุกรุกจากภายนอก (external intruder) และผู้บุกรุกจากภายในระบบ (internal intruder) โดยทั่วไปแล้วผู้บุกรุกจากภายนอกนั้น จะไม่มีสิทธิในการเข้าถึงระบบ ซึ่งต่างจากผู้บุกรุกจากภายในระบบที่มีสิทธิในการใช้งานระบบแต่ไม่ได้รับอนุญาตให้เข้าถึงสิทธิของผู้ดูแลระบบ (superuser mode) และมีพฤติกรรมทั้งในแบบสวมรอย (masquerade) โดยใช้สิทธิที่ถูกต้องในการเข้าถึงข้อมูลของผู้อื่น รวมทั้งแบบปิดบังตน (clandestine internal intruder) ที่ลบประวัติการเข้าใช้งานระบบของตน การบุกรุกหรือการโจมตีเครือข่ายคอมพิวเตอร์แบ่งออกได้เป็นประเภทต่างๆ ดังมีรายละเอียดตามตารางที่ 1 [10][11]

3. ระบบตรวจจับการบุกรุกเครือข่าย

การบุกรุกเป็นการดำเนินการเพื่อหลบเลี่ยงมาตรการรักษาความปลอดภัย ให้ได้มาซึ่งความสามารถในการเข้าถึงและควบคุมระบบ อีกทั้งเนื้อหาและความถูกต้องของข้อมูล [12] การกระทำนี้อาจเกิดได้ทั้งจากภายในและภายนอกระบบ ระบบตรวจจับการบุกรุกจึงเป็นหนทางในการป้องกันภัยคุกคามที่เกิดได้ผ่านการรวบรวม วิเคราะห์ ข้อมูลการสื่อสารเครือข่าย และระบุความผิดปกติที่เกิดขึ้นได้

ระบบตรวจจับการบุกรุกต่าง ๆ ได้ถูกคิดค้นเพื่อตอบรับการโจมตี เหล่านี้ และสามารถแบ่งออกได้ตามลักษณะการใช้งานดังต่อไปนี้

- Host-based IDS หรือ HIDS: เป็นระบบที่ตรวจสอบและวิเคราะห์กระบวนการที่เกิดขึ้นภายในระบบ มากกว่าศึกษาพฤติกรรมการสื่อสารและปฏิสัมพันธ์จากนอกระบบ HIDS คำนึงถึงการตรวจจับความผิดปกติกับเหตุการณ์ [13] เช่น โปรแกรมหนึ่งมีการใช้งานทรัพยากรใดและสิทธิในงานนั้น ๆ โปรแกรม word processor ที่พยายามเข้าถึงและปรับปรุงเนื้อหาของฐานข้อมูลรหัสผ่านในระบบ (system password database)

ตารางที่ 1 ประเภทของการบุกรุกหรือโจมตีเครือข่าย

ประเภท	คุณลักษณะ	ตัวอย่าง
Virus	(i) โปรแกรมที่สามารถทำสำเนาของตนเองได้ และเข้ามาอยู่ในระบบโดยไม่จำเป็นต้องมีสิทธิใด (ii) สามารถขยายผลสู่วงกว้างได้เมื่อมีการเข้าถึงโปรแกรมจากคอมพิวเตอร์เครื่องอื่น	Trivial.88.D, Polyboot.B, และ Tuareg
Worm	(i) โปรแกรมที่สามารถทำสำเนาของตนเองได้ และขยายขอบเขตเป้าหมายผ่านการ เรียกใช้การบริการต่าง ๆ ในเครือข่าย (ii) เป็นอันตรายต่อเครือข่าย มีช่วงสัญญาณการสื่อสาร (bandwidth) ในปริมาณสูง	SQL Slammer, Mydoom, และ CodeRed Nimda
Trojan	(i) โปรแกรมที่ไม่สามารถทำสำเนาตนเองได้ แต่ส่งผลร้ายต่อการทำงานของเครือข่าย (ii) ซ่อนตัวในลักษณะของโปรแกรมที่มีประโยชน์ แต่มีส่วนย่อยที่ทำการเปิดช่องโหว่ ของระบบเพื่อดำเนินการโจมตีต่อไป	Example-Mail Bomb และ phishing attack
Denial of Service (DoS)	(i) มีจุดมุ่งหมายที่จะปิดกั้นการเข้าถึงระบบและทรัพยากร มีผลทำให้บริการผู้ใช้ไม่ได้ (ii) มีลักษณะที่ทำให้ระบบปิดอัตโนมัติ หรือมีทรัพยากรไม่เพียงพอใช้งาน	Buffer overflow, ping of death และ teardrop
Network Attack	(i) กระบวนการที่มุ่งจะลดความสามารถของการรักษาความปลอดภัยของระบบตั้งแต่ ในระดับชั้นการสื่อสาร data link layer จนถึง application layer ด้วยเทคนิคต่าง ๆ เช่น การเปลี่ยนรูปแบบการสื่อสารหรือ network protocols (ii) การเข้าใช้บัญชีผู้อื่นเพื่อลบทรัพยากรและช่องทางการสื่อสาร หรือขัดขวางการเข้า ถึงระบบหรือใช้งานทรัพยากรของผู้ใช้ที่มีสิทธิถูกต้อง	Packet injection และ SYN flood
Password Attack	มุ่งครอบครองรหัสผ่าน ซึ่งสังเกตได้จากความล้มเหลวของการเข้าระบบอย่างต่อเนื่อง	Dictionary attack และ SQL injection attack

ตารางที่ 1 ประเภทของการบุกรุกหรือโจมตีเครือข่าย(ต่อ)

ประเภท	คุณลักษณะ	ตัวอย่าง
Physical Attack	มีวัตถุประสงค์จะทำความเสียหายต่อองค์ประกอบทางกายภาพของระบบ	Cold boot และ evil maid
Information Gathering Attack	รวบรวมข้อมูลหรือค้นหาช่องโหว่ของการรักษาความปลอดภัย โดยการทดสอบต่าง ๆ ไปยังคอมพิวเตอร์เฉพาะเครื่องหรือเครือข่ายโดยรวม เพื่อหยั่งการตอบสนอง	SYS scan, FIN scan, และ XMAS scan
Remote to Local (R2L) Attack	(i) สามารถส่ง packets ไปยังระบบอื่น ๆ ที่ไกลออกไปโดยไม่จำเป็นต้องมีบัญชีหรือ สิทธิในการเข้าถึงระบบ เพื่อเข้าใช้บัญชีผู้ดูแลระบบแล้วทำการโจมตีจากภายใน (ii) ทำการโจมตีช่องทางบริการสาธารณะต่าง ๆ เช่น HTTP และ FTP หรือโจมตีใน ขั้นตอนการเชื่อมต่อของช่องทางบริการที่มีการป้องกัน เช่น POP และ IMAP	Warezclient, imap, ftp, warezmaster, write, multihop, phf, และ spy
Probe	(i) ทำการตรวจสอบการสื่อสารภายในเครือข่าย เพื่อค้นหาหมายเลข IP address ที่ใช้ งานได้ รวมทั้งข้อมูลที่เกี่ยวกับเครื่องคอมพิวเตอร์ในระบบ เช่น บริการที่ให้กับผู้ใช้และระบบปฏิบัติการ (ii) ให้ข้อมูลช่องโหว่ต่าง ๆ ของระบบกับผู้บุกรุกใช้ในการโจมตีระบบ	Ipsweep และ portsweep
User to Root (U2R) Attack	(i) มีความสามารถที่จะขยายผลการบุกรุกจากช่องโหว่ เพื่อเข้าถึงบัญชีผู้ดูแลระบบ โดยเริ่มการใช้งานระบบด้วยสิทธิของผู้ใช้งานทั่วไป (ii) ใช้เทคนิค password sniffing, dictionary attack หรือ social engineering	Rootkit, loadmodule, และ perl

ตารางที่ 2 กลุ่มเทคนิคการตรวจจับการบุกรุกเครือข่าย

กลุ่มเทคนิค	คุณลักษณะ
Misuse-based	(i) การตรวจจับนั้นอ้างอิงกับกฎหรือรูปแบบต่าง ๆ ของการโจมตีที่เกิดขึ้นแล้วในอดีต (ii) ต้องทำการสร้างกฎหรือรูปแบบจากข้อมูลในอดีต แต่ปริมาณอาจจะเพิ่มขึ้นเป็นทวีคูณ เนื่องด้วยการโจมตีที่มีวิวัฒนาการอย่างต่อเนื่อง
Anomaly-based	(i) ตั้งอยู่บนสมมุติฐานที่ว่าการบุกรุกนั้นจะแสดงออกถึงความผิดปกติเป็นประจำ (ii) เทคนิคในกลุ่มนี้จะตรวจสอบว่าเหตุการณ์ที่เกิดขึ้นใหม่แตกต่างหรือเบี่ยงเบนจากโมเดลพฤติกรรมปกติอย่างมีนัยยะหรือไม่ แต่อาจส่งผลให้เหตุการณ์ที่ไม่ใช่การบุกรุกแต่มีคุณลักษณะที่แตกต่างออกไปถูกระบุว่าเป็นการโจมตี ก่อให้เกิดข้อผิดพลาดประเภท false positive (iv) ควรประยุกต์ใช้ระดับความมั่นใจประกอบกับการตรวจสอบข้างต้น เพื่อลดผลกระทบจากกรณีที่กล่าวมา
Hybrid	(i) ใช้ประโยชน์จากเทคนิคจากทั้งสองกลุ่มข้างต้น (ii) พยายามตรวจจับรูปแบบการบุกรุกที่รู้จักแล้ว และรูปแบบใหม่ ๆ ไปพร้อมกัน

ตารางที่ 3 ประเภทของความผิดปกติที่เกี่ยวเนื่องกับการรักษาความปลอดภัยเครือข่าย

ประเภท	คุณลักษณะ	ตัวอย่าง
Point anomaly	ข้อมูลจำเพาะที่แตกต่างจากข้อมูลอื่น ๆ	บันทึกการสื่อสารเครือข่ายที่แยกออกจาก ส่วนที่เหลือในห้วงเวลาที่เฉพาะเจาะจง
Contextual anomaly	(i) การตรวจพบความผิดปกติของบันทึกข้อมูลในบริบทที่เฉพาะ และเชื่อมโยงได้กับพฤติกรรม (ii) บริบทนั้นสรุปหรืออนุมานได้จากกลุ่มข้อมูลที่ศึกษา	ช่วงเวลาการสั่งซื้อสินค้า หรือบริการในกรณีของการใช้ข้อมูลเครดิตการ์ดปลอม
Collective anomaly	(i) การตรวจจับความผิดปกติจากชุดของบันทึกข้อมูลที่มีความ เกี่ยวเนื่องกัน เมื่อเปรียบเทียบกับชุดข้อมูลทั้งหมดที่ศึกษา (ii) ชุดของข้อมูลอาจแสดงออกถึงความผิดปกติ แต่หากเมื่อพิจารณาแยกเป็นรายเหตุการณ์ อาจจะไม่สื่อถึงความผิดปกติใด	ลำดับของเหตุการณ์ เช่น . . . http-web, buffer-overflow, http-web, http-web, ftp, http- web, ssh, http-web, ssh, buffer-overflow . . .

- Network-based IDS หรือ NIDS: เป็นระบบที่ตรวจจับการบุกรุกจากการศึกษาข้อมูลสื่อสารในเครือข่าย ในกรณีทั่วไปที่เครือข่ายใด ๆ เชื่อมต่อเข้ากับอินเทอร์เน็ต NIDS จะทำการพิจารณา packets ที่ส่งผ่านเข้ามาเพื่อหารูปแบบที่น่าสงสัย อาทิเช่น การตรวจพบ packets จำนวนมากที่ส่งเข้ามาเพื่อร้องขอการเชื่อมต่อช่องทางการสื่อสาร TCP โดยระบุเลขช่องทาง (port number) ที่แตกต่างกันไปในช่วงเวลาใกล้เคียงกัน เหตุการณ์ดังกล่าวเป็นปัจจัยให้อนุมานได้ว่าเกิดการบุกรุกประเภท port scan กับคอมพิวเตอร์หรือทรัพยากรอื่น ๆ ในระบบ [14] ข้อมูลการสื่อสารที่ใช้งานโดย NIDS แบ่งออกได้เป็นลำดับขั้นตามความละเอียดหรือการสรุปความ ดังเช่นข้อมูลลำดับการ เชื่อมต่อ packet level traces และเรคคอร์ดข้อมูล IPFIX ข้อมูลเหล่านี้มักจะมีมิติสูง (high dimensional) ที่อธิบายคุณลักษณะต่าง ๆ ของการสื่อสาร ทั้งในแบบตัวเลข (numerical data) แบบบัญญัติ (categorical data) และการผสมกันระหว่างข้อมูลทั้งสองแบบ

เทคนิคการตรวจจับการบุกรุกนั้นแบ่งออกเป็นสามแบบ คือ แบบการใช้งานผิดรูปแบบ (misuse-based) แบบการตรวจจับตามความผิดปกติ (anomaly-based) และแบบผสมผสานกันระหว่างสองประเภทที่กล่าวมา (hybrid) รายละเอียดของเทคนิคต่าง ๆ เหล่านี้ได้สรุปไว้ในตารางที่ 2 ในปัจจุบันนักวิจัยให้ความสำคัญกับเทคนิคประเภทการตรวจจับตามความผิดปกติมากกว่าแบบอื่น ๆ ด้วยคุณสมบัติที่ยืดหยุ่นต่อบริบทการโจมตี ซึ่งเปลี่ยนแปลงอย่างต่อเนื่อง อีกทั้งยังครอบคลุมการบุกรุกที่ได้มีการบันทึกรูปแบบมาแล้วในอดีตและแนวทางใหม่ไปพร้อมกัน

4. การประยุกต์ใช้การระบุสิ่งผิดปกติเพื่อพัฒนาระบบ

การตรวจจับการบุกรุกเครือข่ายจากความผิดปกติหรือ ANIDS (Anomaly-based NIDS) เป็นหัวข้อที่นักวิจัยทั่วโลกให้ความสนใจเป็นอย่างมากในปัจจุบัน [15] จากรายงานของ [16] ความผิดปกติในเครือข่าย

แบ่งได้เป็นสองประเภท ได้แก่ ความผิดปกติที่เกี่ยวกับสมรรถนะการดำเนินงาน และความผิดปกติที่สืบเนื่องจากการรักษาความปลอดภัย ตัวอย่างความผิดปกติในกลุ่มแรก คือ broadcast storms, transient congestion, babbling node, paging across the network, และ file server failure ตามลำดับ ส่วนความผิดปกติในกลุ่มที่สองนั้นเกิดจากการบุกรุกเครือข่าย เช่น ตั้งใจเพิ่มปริมาณการสื่อสารข้อมูลในเครือข่ายให้สูงขึ้น ด้วยการส่ง packets จำนวนมากเข้าสู่ระบบ ทั้งนี้ความผิดปกติในกลุ่มนี้สามารถจำแนกย่อยได้เป็นสามแบบ คือ point anomaly, contextual anomaly และ collective anomaly ดังรายละเอียดในตารางที่ 3

4.1 ปัจจัยที่มีผลต่อการพัฒนาระบบ ANIDS

ปัญหาตรวจจับการบุกรุกนั้นสามารถแก้ไขได้ตามแนวทางการจำแนกข้อมูล (classification) หรือการจัดกลุ่มข้อมูล (clustering) โดยมีปัจจัยต่าง ๆ ที่สำคัญต่อการออกแบบระบบ [5] ดังนี้

- ประเภทของข้อมูลที่ศึกษา (types of input data) ข้อมูลนำเข้าสู่การวิเคราะห์นั้นเป็นชุดของหน่วยข้อมูล (ซึ่งในบางครั้งเรียกว่า วัตถุ เรคคอร์ด เวกเตอร์ หรือกรณี) [17] ข้อมูลแต่ละหน่วยอธิบายได้ด้วยคุณลักษณะเฉพาะหรือแอททริบิวต์ (attribute) ซึ่งเป็นข้อมูลประเภทต่าง ๆ กัน เช่น ไบนารี (binary) นามบัญญัติ (categorical) หรือตัวเลข (numeric) ชุดข้อมูลอาจจะประกอบด้วยแอททริบิวต์เดียวหรือว่าหลายแอททริบิวต์ (univariate or multivariate) โดยที่เป็นการผสมผสานระหว่างข้อมูลประเภทต่าง ๆ ได้ คุณสมบัติของชุดข้อมูลที่ศึกษาจะกำหนดแนวทางการประยุกต์เทคนิคการวิเคราะห์ในลำดับถัดไป
- มาตรวัดระดับความแตกต่างหรือระยะห่าง (dissimilarity or proximity measure) นั้นเป็นหลักการพื้นฐานที่สำคัญต่อการรู้จำรูปแบบซึ่งรวมถึงการจำแนกและการจัดกลุ่มข้อมูล ระยะห่าง (distance) คือการวัดระดับความแตกต่างกันระหว่างสองหน่วยข้อมูล [19] วิธีการวัดที่เหมาะสมขึ้นอยู่กับประเภทข้อมูลของแอททริบิวต์ โดยทั่วไปการวัดระยะห่างมักถูกนิยามในแบบฟังชันที่รับค่าแอททริบิวต์ของสองหน่วยข้อมูลแล้วให้ผลลัพธ์เป็นค่าตัวเลข ซึ่งค่าต่ำและค่าสูง แสดงถึงความเหมือนกันในระดับมากและน้อย [18]
- การระบุประเภทหรือคำตอบของหน่วยข้อมูล (labeling of data) ตามลักษณะของแอททริบิวต์ว่าเป็นการสื่อสารปกติ (normal) หรือเป็นเหตุการณ์ที่ผิดปกติ (anomalous) แต่หากขั้นตอนการระบุคำตอบนี้มีค่าใช้จ่ายสูงหรืออาจจะทำไม่ได้เสมอไป เนื่องด้วยต้องอาศัยผู้เชี่ยวชาญให้คำตอบ[5] นอกจากนั้นเหตุการณ์ผิดปกติก็มีวิวัฒนาการไปตามเวลา จึงมีผลให้การระบุคำตอบกับการบุกรุกแบบใหม่ ๆ เป็นไปได้ยาก เมื่อพิจารณาความพร้อมของคำตอบ เทคนิคการตรวจจับ สามารถจำแนกได้เป็นสามประเภท คือ แบบมีผู้สอน (supervised) แบบกึ่งมีผู้สอน (semi-supervised) และแบบไม่มีผู้สอน (unsupervised) โดยเทคนิคแบบมีผู้สอนต้องการชุดข้อมูลที่มีคำตอบทั้งในส่วนของเหตุการณ์ที่ปกติ และผิดปกติ เพื่อทำการพัฒนาโมเดลรู้จำรูปแบบที่ครอบคลุมทั้งสองแบบคำตอบ และใช้ทำนายคำตอบของหน่วยข้อมูลใหม่ที่เกิดขึ้น ในลักษณะที่คล้ายกัน เทคนิคแบบกึ่งมีผู้สอน จะใช้ประโยชน์ของคำตอบในส่วนเหตุการณ์ที่ปกติเท่านั้น โดยคำนึงความยากในการระบุส่วนที่ไม่ปกติล่วงหน้า ทั้งนี้หน่วยข้อมูลใหม่ที่ไม่ได้มีรูปแบบที่สอดคล้องกับเหตุการณ์ที่ปกติจะยังไม่ถูกระบุว่าผิดปกติ แต่เป็นหน้าที่ของผู้เชี่ยวชาญที่จะประเมินความน่าจะเป็นต่อไป [19][21] ในทางตรงกันข้ามกับสองแนวทางที่กล่าวมา เทคนิคแบบไม่มีผู้สอน ไม่ต้องการคำตอบของหน่วยข้อมูลในการพัฒนาโมเดล

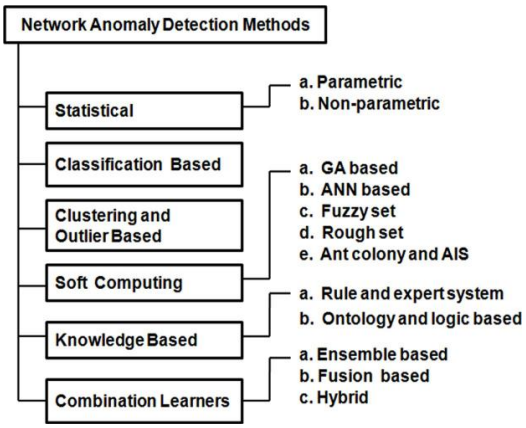
เพื่อระบุประเภทของเหตุการณ์ จึงได้รับความนิยมและใช้งานอย่างแพร่หลาย ทั้งนี้อาศัยการวัดระยะห่างเพื่อประเมินรูปแบบที่เป็นปกติและผิดปกติ เทคนิคนี้ได้ประสบปัญหาของสัดส่วนคำตอบทั้งสองที่แตกต่างกันมาก (โดยที่เหตุการณ์ผิดปกติมีปริมาณน้อยกว่าเหตุการณ์ที่ปกติมาก) น้อยกว่าเทคนิคในสองแบบแรก เหมาะสมกับการใช้งานในบริบทที่มีการเปลี่ยนแปลงบ่อยมากกว่านั่นเอง

- การคัดเลือกชุดแอททริบิวต์ที่สำคัญ (attribute or feature selection) มีบทบาทที่สำคัญและช่วยยกระดับคุณภาพของระบบตรวจจับการบุกรุกให้สูงขึ้น โดยทำการศึกษาถึงความสำคัญและความเกี่ยวเนื่องของแอททริบิวต์ต่าง ๆ กับรูปแบบของเหตุการณ์ที่ระบุได้ด้วยโมเดลการเรียนรู้ของเครื่องแอททริบิวต์ที่ไม่เกี่ยวข้องหรือมีระดับความสัมพันธ์ข้างต้นต่ำจะถูกตัดออกจากกระบวนการพัฒนาโมเดล วิธีจำพวกนี้มีการดำเนินงานในสามขั้นตอน คือ ขั้นการสร้างกลุ่มย่อย (subset) จากแอททริบิวต์ทั้งหมด การประเมินคุณภาพของกลุ่มย่อยที่สร้างขึ้นมาและการตรวจสอบกับข้อมูลใหม่ตามลำดับ [21]
- รูปแบบของรายงานผลการตรวจจับสิ่งที่ผิดปกติ (reporting anomalies) เป็นอีกหนึ่งประเด็นที่สำคัญสำหรับการพัฒนาระบบตรวจจับการบุกรุก [5] ผลลัพธ์ของการตรวจจับนั้นแบ่งได้เป็นสองแบบคือ รูปแบบของคะแนนที่บ่งบอกระดับความมั่นใจของการระบุประเภทเหตุการณ์ และในรูปแบบของสัญลักษณ์ที่ระบุประเภท (ปกติหรือผิดปกติ) ลักษณะของผลลัพธ์ข้างต้นจะมีผลต่อการแปลความหมายและการนำไปประยุกต์ใช้

4.2 เทคนิคที่ใช้ในการพัฒนาระบบ ANIDS

โครงสร้างที่ใช้เพื่อจำแนกเทคนิคสำหรับพัฒนาระบบ ANIDS ได้แสดงไว้ในรูปที่ 3 ซึ่งอยู่บนพื้นฐานของอัลกอริทึมที่ประยุกต์ใช้

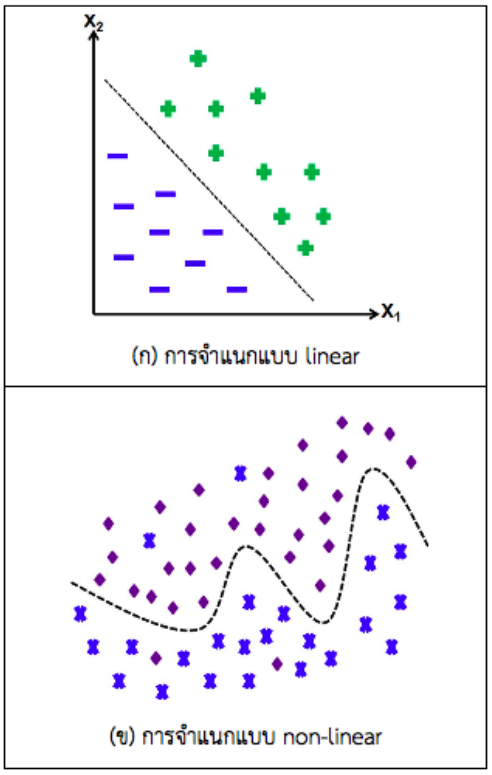
- วิธีการเชิงสถิติ (statistical methods) เหตุการณ์ผิดปกติสามารถระบุได้ด้วยการเทียบกับโมเดลต้นแบบ ซึ่งเป็นแบบสุ่ม [22] ประกอบการอนุมาน (inference) ถึงความปกติหรือผิดปกติของหน่วยข้อมูลที่ทำการวิเคราะห์ วิธีการเชิงสถิติแบ่งออกเป็นสองประเภท คือ แบบอ้างอิงตัวแปรและแบบไม่อ้างอิงตัวแปร (parametric and non-parametric) โดยเทคนิคแบบแรกประเมินค่าตัวแปรสำหรับโมเดลสุ่มจากชุดข้อมูล [23] ในขณะที่อีกแบบไม่อ้างอิงความรู้ใดจากชุดข้อมูล [24] ตัวอย่างของวิธีการเหล่านี้สรุปไว้ในตารางที่ 4



รูปที่ 3 การจำแนกเทคนิคที่ใช้พัฒนาระบบ ANIDS

- วิธีการจำแนกประเภทข้อมูล (classification based methods) เป็นการทำนายประเภทของหน่วยข้อมูลใหม่ โดยอ้างอิงจากการสร้างโมเดลกลุ่มจากชุดข้อมูลที่มีคำตอบประกอบ ตัวอย่างที่แสดงในรูปที่ 4(ก) เป็นกรณีการจำแนกของข้อมูลสองประเภท (+ และ -)

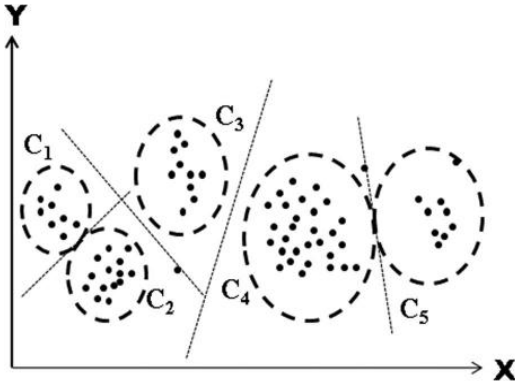
โดยสัญลักษณ์ทั้งสองแทนหน่วยข้อมูลทั้งหมดที่ศึกษา แต่ละหน่วยข้อมูลอธิบายได้ด้วยแอท ทริบิวต์ x_1 และ x_2 การสร้างโมเดลจำแนกประเภทข้อมูลคือ การค้นหาขอบเขตการแบ่งชุดข้อมูลออกเป็นกลุ่มที่เฉพาะสำหรับคำตอบหนึ่ง ๆ ซึ่งการแบ่งนั้นอาจเป็นในลักษณะของสมการเส้นตรง (linear) หรือสมการที่ไม่ ใช่เส้นตรง (non–linear) ตามตัวอย่างในรูปที่ 4 (ก) และ 4(ข) ตามลำดับ



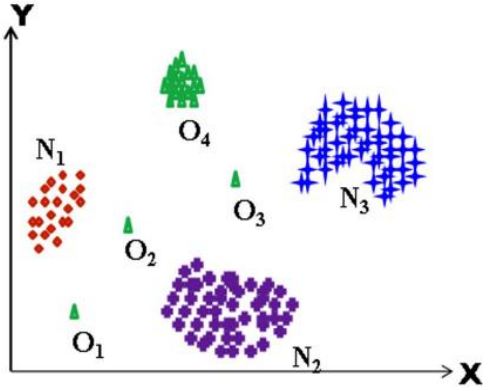
รูปที่ 4 กรณีของการพัฒนาโมเดลจำแนกประเภทข้อมูล

เทคนิคการจำแนกข้อมูลแบบต่าง ๆ ได้ถูกนำมาประยุกต์ใช้ ในการสร้างการตรวจจับการบุกรุก เช่น k-nearest neighbor (KNN), support vector machine (SVM) และ decision tree เป็นต้น ตัวอย่างของงานวิจัยที่เกี่ยวข้องได้สรุปไว้ในตารางที่ 5

- วิธีการจัดกลุ่มข้อมูลและระบุกรณีผิดปกติ (clustering and outlier based methods) ตามหลักการพื้นฐาน การจัดกลุ่มข้อมูลคือ การแบ่งชุดข้อมูลที่ประกอบด้วยหลายเรคคอร์ดเป็นกลุ่มย่อย ๆ หรือคลัสเตอร์ (cluster) โดยให้เรคคอร์ดที่คล้ายกันอยู่ในคลัสเตอร์เดียวกัน และต่างจากเรคคอร์ดที่มี ระยะห่างหรือแตกต่างกันสูง รูปที่ 5 แสดงตัวอย่างของ การแบ่งชุดข้อมูลออกเป็น 5 กลุ่ม ในรูปที่ 6 ได้แสดงอีกหนึ่งตัวอย่างของผลลัพธ์จากการ จัดกลุ่มข้อมูล 3 กลุ่มคือ N_1, N_2 และ N_3 โดยที่ปรากฏกรณีที่ไม่ปกติ (outlier) ซึ่งมีความต่างจากกลุ่มที่ได้พัฒนาขึ้นอย่างมีนัยยะ [39] ได้แก่ O_1, O_2, O_3 และ O_4 ในรูปข้างต้น ตัวอย่างของวิธีการต่าง ๆ ในกลุ่มนี้ ได้สรุปไว้ในตารางที่ 6



รูปที่ 5 ตัวอย่างผลลัพธ์ของการจัดกลุ่มข้อมูล 5 กลุ่ม: C_1, C_2, C_3, C_4 และ C_5



รูปที่ 6 การจัดกลุ่มข้อมูล 3 กลุ่ม (N_1, N_2 และ N_3) และกรณีที่ ไม่ปกติ (O_1, O_2, O_3 และ O_4)

- วิธีการแบบการคำนวณคลุมเครือและโมเดลธรรมชาติ (soft computing methods) เป็น

วิธีการที่ได้รับความนิยมสำหรับการเรียนรู้กับข้อมูลที่คลุมเครือ ซึ่งสอดแทรกมิติของพฤติกรรมเลียนแบบธรรมชาติ ตัวอย่างสรุปไว้ในตารางที่ 7

- วิธีการอ้างอิงองค์ความรู้ (knowledge based methods) จะทำการตรวจหน่วยข้อมูลการสื่อสาร เทียบเคียงกับกฎของการบุกรุกที่แสดงไว้ในรูปแบบที่ครอบคลุมกรณีต่าง ๆ อย่างกว้างขวาง เทคนิคที่นำมาประยุกต์ใช้ ได้แก่ rule-based systems, ontology-based system และ state-transition analysis ตัวอย่างของวิธีการในกลุ่มนี้สรุปไว้ในตารางที่ 8

ตารางที่ 4 ตัวอย่างวิธีการเชิงสถิติ (ประเภทของการบุกรุกที่ครอบคลุม: A1-all attacks, A2-denial of service, A3-probe, A4-user to root, A5-remote to local และ A6-anomalous)

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[23], 2000	ตัวเลข	KDD99	A4	Probability Model
[25], 2002	ตัวเลข	ข้อมูลเฉพาะ	A2, A5	Statistical model with neural network
[27], 2003	ตัวเลข	KDD99	A1	LERAD algorithm
[26], 2003	ตัวเลข	KDD99	A1	Learning Rules
[28], 2004	ตัวเลข	KDD99	A1	Payload-based algorithm
[29], 2007	ตัวเลข	KDD99	A1	Gaussian Mixture Model
[30], 2008	ตัวเลข	ข้อมูลเฉพาะ	A6	FDR method
[31], 2009	ตัวเลข	KDD99	A1	Wavelet Analysis
[32], 2011	ตัวเลข	ข้อมูลเฉพาะ	A2	GLRT Model
[33], 2012	ตัวเลข	ข้อมูลเฉพาะ	A2	Adaptive CUSUM

ตารางที่ 5 ตัวอย่างวิธีการจำแนกประเภทข้อมูล

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[34], 2009	ตัวเลข	KDD99	A1	Probability Model
[35], 2010	ตัวเลข	ข้อมูลเฉพาะ	A2	Statistical model with neural network
[37], 2011	ตัวเลข	ข้อมูลเฉพาะ	A2	LERAD algorithm
[38], 2011	ตัวเลข	KDD99	A1	Learning Rules
[36], 2012	ตัวเลข	DARPA98	A1	Payload-based algorithm

ตารางที่ 6 ตัวอย่างวิธีการจัดกลุ่มข้อมูลและระบุกรณีที่ไม่ปกติ

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[44], 2008	ตัวเลข	ข้อมูลเฉพาะ	A3	AAWP model
[40], 2009	แบบผสม	KDD99	A1	KD algorithm
[42], 2010	ตัวเลข	ข้อมูลเฉพาะ	A6	PAIDS model
[43], 2011	ตัวเลข	KDD99	A1	NADO algorithm
[41], 2012	ตัวเลข	KDD99	A1	UNIDS method

ตารางที่ 7 ตัวอย่างวิธีการแบบการคำนวณคลุมเครือและโมเดลธรรมชาติ

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[47], 2008	ตัวเลข	KDD99	A1	LEM2 and KNN
[48], 2009	ตัวเลข	KDD99	A2	RST-SVM technique
[46], 2011	ตัวเลข	KDD99	A1	Fuzzy ARM-based GNP
[49], 2011	ตัวเลข	ข้อมูลเฉพาะ	A2	Interval type-2 fuzzy set
[45], 2012	ตัวเลข	KDD99	A1	Fuzzy rule-based model

ตารางที่ 8 ตัวอย่างวิธีการแบบอ้างอิงองค์ความรู้

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[50], 2003	ตัวเลข	ข้อมูลเฉพาะ	A2	Markov Chain Model
[51], 2008	ตัวเลข	KDD99	A1	Ontology-based
[52], 2010	ตัวเลข	ข้อมูลเฉพาะ	A2	Incremental KBTA

ตารางที่ 9 ตัวอย่างวิธีการแบบสรุปความจากกลุ่มผลลัพธ์

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[54], 2009	ตัวเลข	KDD99	A1	Min and Max probability
[53], 2010	ตัวเลข	KDD99	A1	Information theoretic
[60], 2011	ตัวเลข	KDD99	A1	Majority voting
[55], 2012	ตัวเลข	UNM	A4	Learn and combine

ตารางที่ 10 ตัวอย่างวิธีการแบบผสานผลลัพธ์

เอกสารอ้างอิง, ปี	ประเภทข้อมูล	ชุดข้อมูลที่ศึกษา	ขอบเขตการบุกรุก	วิธีที่ประยุกต์ใช้
[56], 2008	ตัวเลข	KDD99	A1	dLEARNIN system (data level fusion)
[58], 2010	ตัวเลข	KDD99	A1	IDEA model (data level fusion)
[59], 2011	ตัวเลข	KDD99	A1	HMMPayl model (decision level fusion)
[57], 2012	ตัวเลข	UNM	A2, A3	EWMA model (decision level fusion)

- วิธีการรวมผลจากชุดโมเดลการวิเคราะห์ (combination of learners) สามารถแยกได้เป็นประเภทย่อยดังนี้ (ก) การสรุปความจากกลุ่มผลลัพธ์ (ensemble-based methods) ทำการถ่วงน้ำหนักผลลัพธ์ที่ได้จากโมเดลการวิเคราะห์ที่ต่างกัน แล้วทำการสรุปเป็นคำตอบที่มีความถูกต้องกว่าผลลัพธ์ของโมเดลใด ๆ ในกลุ่มที่นำมาศึกษา นิยมใช้สรุปผลที่ได้จากโมเดลการจำแนก (classifier) ที่สร้างจากการใช้ส่วนย่อยของชุดข้อมูลหรืออัลกอริทึมที่หลากหลาย ตารางที่ 9 ได้สรุปตัวอย่างของวิธีประเภทนี้ (ข) การผสานผลลัพธ์ (fusion-based method) มีจุดมุ่งหมายที่จะยกระดับความแม่นยำของการทำนาย (prediction) ด้วยโมเดลจำแนกข้อมูลพื้นฐาน โดยแบ่งแยกการผสานผลลัพธ์ออกเป็นหลายระดับ ได้แก่ การผสานที่ระดับหน่วยข้อมูล (data level) การผสานในมิติของแอททริบิวต์ (feature level) และการผสานในระดับการตัดสินใจ (decision level) ตามลำดับ ตัวอย่างของวิธีในกลุ่มนี้ได้สรุปไว้ในตารางที่ 10

5. ทิศทางการพัฒนาในอนาคตและบทสรุป

การโจมตีเครือข่ายในอนาคตจะทวีความรุนแรงมากยิ่งขึ้น เพราะการดำเนินธุรกรรมต่าง ๆ ทั้งในส่วนขององค์กรและส่วนบุคคลต่างให้ความสำคัญกับการใช้งานและแบ่งปันข้อมูลผ่านอินเทอร์เน็ต จึงส่งผลให้เกิดช่องทางที่อาจจะเป็นช่องโหว่ซึ่งยากต่อการป้องกัน อีกทั้งรูปแบบการโจมตีที่เปลี่ยนแปลงไปตามกาลเวลาและตอบโต้การป้องกันที่ได้พัฒนาขึ้น ฉะนั้นการตรวจจับการบุกรุกที่เกิดขึ้นในอนาคตต้องอาศัยหลักการระบุสิ่งปกติเพื่อให้ระบบมีความยืดหยุ่นสอดคล้องกับบริบทสิ่งแวดล้อม นอกจากนี้แนวทางการบูรณาการหรือการสรุปผลจากการวิเคราะห์ย่อย (ensemble) นั้นเป็นประเด็นวิจัยที่ได้รับความสนใจจากนักวิจัยทั่วโลก เนื่องด้วยความสามารถที่จะตรวจจับสิ่งผิดปกติได้แม่นยำและครอบคลุมกรณีที่หลากหลายมากกว่าการใช้วิธีการใดวิธีการหนึ่งเพียงอย่างเดียว

5.1 ทิศทางการพัฒนาและวิจัย

หากในปัจจุบันงานวิจัยในหัวข้อที่กล่าวมายังคงใช้หลักการที่มีการใช้งานกับการสรุปผลของโมเดลจำแนกข้อมูลเท่านั้น แต่วิธีการที่ใช้กับการสรุปผลของโมเดลจัดกลุ่มข้อมูล (cluster ensemble หรือ consensus clustering) ยังไม่ได้รับการตอบรับเท่าที่ควร ข้อดีของแนวคิดใหม่นี้ คือ โมเดลดำเนินงานได้โดยไม่ต้องมีตัวอย่างคำตอบ หรือผู้สอน รวมทั้งสามารถแก้ปัญหาการระบุคำตอบกับชุดข้อมูลที่มีขนาดใหญ่ (big data) ได้อีกด้วย

ทฤษฎีการรวบรวมชุดคำตอบย่อยของการจัดกลุ่มที่จะได้นำมาประยุกต์ใช้เพื่อรวบรวมผลของการตรวจจับสิ่งผิดปกตินั้นได้ผ่านการพัฒนาและนำเสนอในเวทีวิชาการนานาชาติมาอย่างต่อเนื่อง อีกทั้งยังมีการนำไปใช้ในการวิเคราะห์ข้อมูลหลากหลายแขนง [61] แม้ว่าจะมีวิธีการจัดกลุ่มมากมาย แต่เป็นที่เข้าใจกันดีว่าไม่มีวิธีใดที่จะสามารถคงคุณภาพที่ดีได้กับทุก ๆ ชุดข้อมูล นั่นคือไม่มีอัลกอริทึมใดที่จะสามารถค้นพบ

ลักษณะของกลุ่มและรองรับโครงสร้างข้อมูลได้ทุกรูปแบบ [62] แต่ละอัลกอริทึมมีข้อดีและข้อจำกัดที่แตกต่างกัน จึงเหมาะกับข้อมูลที่มีลักษณะเฉพาะบางประการเท่านั้น หากได้ใช้อัลกอริทึมที่แตกต่างกันในการจัดกลุ่มชุดข้อมูลใด ๆ หรือแม้กระทั่งการใช้อัลกอริทึมเดียวกัน แต่มีการกำหนดค่าตัวแปรที่แตกต่างกัน ก็มักจะได้ผลของการจัดกลุ่มที่แตกต่างกันด้วย จึงเป็นเรื่องที่ยากมากในการเลือกใช้อัลกอริทึมที่เหมาะสมกับชุดข้อมูลหนึ่ง ๆ โดยทั่วไปผู้ใช้อาจจะต้องทำการจัดกลุ่มข้อมูลชุดนั้น ๆ ด้วยอัลกอริทึมที่ต่าง ๆ กัน หรือการกำหนดค่าตัวแปรที่หลากหลาย จากนั้นทำการเปรียบเทียบผลการจัดกลุ่มที่ได้ว่าแบบใดให้ผลลัพธ์ที่ดีที่สุด

ด้วยเหตุนี้ เทคนิคการจัดกลุ่มข้อมูลแบบรวบรวมชุดคำตอบจึงมีบทบาทในการแก้ปัญหาและข้อจำกัดดังกล่าว พร้อมทั้งสามารถเพิ่มประสิทธิภาพการจัดกลุ่มข้อมูลและคุณภาพกลุ่มข้อมูลที่ได้ เทคนิคการจัดกลุ่มข้อมูลแบบนี้มีวัตถุประสงค์เพื่อรวบรวมผลการจัดกลุ่มที่หลากหลายเข้าด้วยกัน โดยเป็นการรวบรวมเพื่อให้คุณภาพของการจัดกลุ่มดีขึ้นกว่าการใช้อัลกอริทึมใด ๆ เพียงอัลกอริทึมเดียว การจัดกลุ่มข้อมูลแบบรวบรวมชุดคำตอบ สามารถแบ่งได้เป็น 4 ประเภทดังต่อไปนี้

- วิธีการรวบรวมผลแบบคุณลักษณะ (feature-based) นำวิธีการจัดกลุ่มสำหรับข้อมูลประเภทนามบัญญัติมาใช้ โดยจะพิจารณาผลของการจัดกลุ่ม (cluster label) ต่าง ๆ ภายใน ชุดคำตอบเป็นข้อมูลแบบนามบัญญัติ [63]
- วิธีการรวบรวมผลแบบโดยตรง (direct manipulation) ซึ่งจะเทียบเคียงผลระหว่างสมาชิก ในชุดคำตอบโดยตรง พร้อมประมวลความน่า จะเป็นของคำตอบสุดท้าย โดยใช้การนับคะแนนสะสม [64]
- วิธีการรวบรวมผลแบบกราฟ (graph-based) นำเทคนิคการแบ่งกราฟมาประยุกต์ใช้ [65] เพื่อแบ่งข้อมูลที่แสดงในรูปแบบของกราฟออกเป็นส่วนย่อย ๆ
- วิธีการรวบรวมผลตามความเหมือน (pairwise similarity) นำความสัมพันธ์แบบการเกิดร่วมกัน (co-occurrence) ของข้อมูลมาช่วยในการจัดกลุ่ม [66]

5.2 บทสรุป

งานวิจัยเพื่อสนับสนุนการพัฒนาระบบตรวจจับการบุกรุกนั้น มีความสำคัญต่อการดำเนินงานของทั้งภาครัฐและเอกชน มีประเด็นที่ท้าทายเกิดขึ้นตลอดเวลา หนึ่งในนั้นคือ การใช้วิธีระบุสิ่งผิดปกติเพื่อตรวจจับการ บุกรุกหรือโจมตีเครือข่าย โดยเฉพาะแนวคิดการรวบรวมโมเดลแบบไม่มีผู้สอนที่หลากหลายเข้าด้วยกันผ่านเทคนิคการรวมชุดคำตอบของการจัดกลุ่ม ซึ่งตอบรับการใช้งานกับข้อมูลจริงรวมถึงข้อมูลขนาดใหญ่ วิธีตรวจจับการบุกรุกนี้สามารถระบุการโจมตีรูปแบบใหม่ ๆ ได้ และเป็นองค์ความรู้ที่จะเป็นประโยชน์ต่อสาธารณะ ลดการพึ่งพาเทคโนโลยีจากต่างประเทศ อีกทั้งยังสามารถต่อยอดให้ เป็นระบบที่มีราคาต่ำกว่าสินค้าที่ผลิตโดยบริษัทจากประเทศอื่น [67]

การวิจัยและพัฒนาข้างต้นตอบรับนโยบายและยุทธศาสตร์การพัฒนาประเทศไทยเป็นอย่างดี ตามนโยบายของคณะรัฐมนตรีที่ได้ แถลงต่อสภานิติบัญญัติแห่งชาติในวันที่ 12 กันยายน พ.ศ.2557 จะสอดรับกับนโยบายด้านการพัฒนาและส่งเสริมการใช้ประโยชน์จากวิทยาศาสตร์ เทคโนโลยี การวิจัยและพัฒนาและนวัตกรรม วิธีตรวจจับการบุกรุกสามารถนำไปขยายผลสู่การใช้งาน พัฒนาไปสู่การใช้ประโยชน์เชิงพาณิชย์ได้ อีกทั้งส่งเสริมความร่วมมือระหว่างหน่วยงานของรัฐและมหาวิทยาลัยในด้านงานวิจัย ในมิติการพัฒนาบุคคลและองค์ความรู้ที่สำคัญต่อการพัฒนาประเทศ รวมทั้งการใช้งานเพื่อรักษาความมั่นคง [68] ต่อไป

6. เอกสารอ้างอิง

[1] The 2015 Household Survey on the Use of Information and Communication Technology, National Statistical Office, Ministry of Information and Communication Technology, 2015.

[2] Mcafee: Estimating the Global Cost of Cybercrime 2014, www.mcafee.com/us/resources/reports/rp-economic-impact-cybercrime2.pdf [Accessed on 1 June 2016].

[3] A. Sundaram. An introduction to intrusion detection. Crossroads, 2(4): 3–7, 1996.

[4] J.P. Anderson. Computer Security Threat Monitoring and Surveillance. James P Anderson Co, Fort Washington, Pennsylvania, Tech. Rep., April 1980.

[5] V. Chandola, A. Banerjee, and V. Kumar. Anomaly Detection: A Survey. ACM Computing Surveys, 41(3): 15/1–15/58, 2009.

[6] N.K. Ampah, C.M. Akujuobi, M.N.O. Sadiku, and S. Alam. An intrusion detection technique based on continuous binary communication channels. International Journal of Security and Networks, 6(2-3): 174–180, 2011.

[7] F.Y. Edgeworth. On discordant observations. Philosophy Mag., 23(5):364–375, 1987.

[8] A. Patcha and J.M. Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12):3448–3470, 2007.

[9] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, and E. Vazquez. Anomaly-based network intrusion detection: techniques, systems and challenges. Computers & Security, 28(1-2): 18–28, 2009.

[10] H.G. Kayacik, A.N. Zincir-Heywood, and M.I. Heywood. **Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets**. In Proceedings of Annual Conference on Privacy, Security and Trust, 2005.

[11] A.A. Ghorbani, W. Lu, and M. Tavallaee. **Network Intrusion Detection and Prevention: Concepts and Techniques. Advances in Information Security.** Springer-Verlag, 2009.

[12] R. Heady, G. Luger, A. Maccabe, and M. Servilla. The Architecture of a Network Level Intrusion Detection System. Computer Science Department, University of New Mexico, Tech. Rep. TR-90, 1990.

[13] Wikimedia: Intrusion detection system. http://en. wikipedia.org/wiki/Intrusion-detection-system [Accessed on 13 Aug 2016].

[14] M.H. Bhuyan, D.K. Bhattacharyya, and J.K. Kalita. Surveying Port Scans and Their Detection Methodologies. The Computer Journal, 54(10): 1565–1581, 2011.

[15] V. Kumar. Parallel and distributed computing for cyber security. IEEE Distributed Systems Online, 6(10), 2005.

[16] M.Thottan and C. Ji. Anomaly detection in IP networks. IEEE Trans. Signal Process, 51(8): 2191–2204, 2003.

[17] P.N. Tan, M. Steinbach, and V. Kumar. **Introduction to Data Mining**. Addison-Wesley, 2005.

[18] S.H. Cha. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. International Journal of Mathematical Models and Methods in Applied Science, 1(4):300–307, 2007.

[19] M.V. Joshi, R.C. Agarwal, and V. Kumar. Mining needle in a haystack: classifying rare classes via two-phase rule induction. In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 293–298, 2001.

[20] J. Theiler and D.M. Cai. Resampling approach for anomaly detection in multispectral images. In Proceedings of SPIE: 230–240, 2003.

[21] R. Fujimaki, T. Yairi, and K. Machida. An approach to spacecraft anomaly detection problem using kernel feature space. In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery in Data Mining: 401–410, 2005.

[22] F.J. Anscombe and I. Guttman. Rejection of outliers. Technometrics, 2(2): 123–147, 1960.

[23] E. Eskin. Anomaly detection over noisy data using learned probability distributions. In Proceedings of International Conference on Machine Learning: 255–262, 2000.

[24] M. Desforges, P. Jacob, and J. Cooper. Applications of probability density estimation to the detection of abnormal conditions in engineering. In Proceedings of Institute of Mechanical Engineers, 687–703, 1998.

[25] C. Manikopoulos and S. Papavassiliou. Network Intrusion and Fault Detection: A Statistical Anomaly Approach. IEEE Communication Magazine, 40(10):76–82, 2002.

[26] P.K. Chan, M.V. Mahoney, and M.H. Arshad. A machine learning approach to anomaly detection. Department of Computer Science, Florida Institute of Technology, Tech. Rep. CS-2003-06, 2003.

[27] M.V. Mahoney and P.K. Chan. Learning rules for anomaly detection of hostile network traffic. In Proceedings of IEEE International Conference on Data Mining: 601-604, 2003.

[28] K. Wang and S.J. Stolfo. Anomalous Payload-Based Network Intrusion Detection. In Proceedings of Recent Advances in Intrusion Detection: 203–222, 2004.

[29] X. Song, M. Wu, C. Jermaine, and S. Ranka. Conditional Anomaly Detection. IEEE Transactions on Knowledge and Data Engineering, 19: 631–645, 2007.

[30] P. Chhabra, C. Scott, E.D. Kolaczyk, and M. Crovella. Distributed Spatial Anomaly Detection. In Proceedings of IEEE International Conference on Computer Communications: 1705–1713, 2008.

[31] W. Lu and A.A. Ghorbani. Network Anomaly Detection Based on Wavelet Analysis. EURASIP Journal of Advances in Signal Processing, 2009(837601), 2009.

[32] F.S. Wattenberg, J. I.A. Perez, P.C. Higuera, M.M. Fernandez, and I.A. Dimitriadis. Anomaly Detection in Network Traffic Based on Statistical Inference and α-Stable Modeling. IEEE Transactions on Dependable Secure Computing, 8(4): 494–509, 2011.

[33] M. Yu. A Nonparametric Adaptive CUSUM Method And Its Application In Network Anomaly Detection. Int. Journal of Advancements in Computing Technology, 4(1): 280–288, 2012.

[34] W. Lu and H. Tong. Detecting Network Anomalies Using CUSUM and EM Clustering. In Proceedings of International Symposium on Advances in Computation & Intelligence: 297–308, 2009.

[35] M.A. Qadeer, A. Iqbal, M. Zahid, and M.R. Siddiqui. Network Traffic Analysis and Intrusion Detection Using Packet Sniffer. In Proceedings of International Conference on Communication Software and Networks: 313–317, 2010.

[36] I. Kang, M.K. Jeong, and D. Kong. A differentiated one-class classification method with applications to intrusion detection. Expert Systems with Applications, 39(4): 3899–3905, 2012.

[37] C. Wagner, J. Francois, R. State, and T. Engel. Machine Learning Approach for IP-Flow Record Anomaly Detection. In Proceedings of International IFIP conference on Networking: 28–39, 2011.

[38] Z. Muda, W. Yassin, M.N. Sulaiman, and N.I. Udzir. A K-means and naive bayes learning approach for better intrusion detection. Information Technology Journal, 10(3): 648–655, 2011.

[39] M.H. Bhuyan, D.K. Bhattacharyya, and J.K. Kalita. RODD: An Effective Reference-Based Outlier Detection Technique for Large Datasets. Advanced Computing, 133: 76–84, 2011.

[40] C. Zhang, G. Zhang, and S. Sun. A Mixed Unsupervised Clustering-Based Intrusion Detection Model. In Proceedings of International Conference on Genetic and Evolutionary Computing: 426–428, 2009.

[41] P. Casas, J. Mazel, and P. Owezarski. Unsupervised network intrusion detection systems: detecting unknown without knowledge. Computer Communications, 35(7): 772–783, 2012.

[42] Z. Zhuang, Y. Li, and Z. Chen. Enhancing intrusion detection system with proximity information. International Journal of Security and Networks, 5(4): 207–219, 2010.

[43] M.H. Bhuyan, D.K. Bhattacharyya, and J.K. Kalita. NADO: network anomaly detection using outlier approach. In Proceedings of ACM International Conference on Communication, Computing and Security: 531–536, 2011.

[44] Z. Chen and C. Chen. A Closed-Form Expression for Static Worm- Scanning Strategies. In Proceedings of IEEE International Conference on Communications: 1573– 1577, 2008.

[45] F. Geramiraz, A.S. Memaripour, and M. Abbaspour Adaptive Anomaly-Based Intrusion Detection System Using Fuzzy Controller. International Journal of Network Security, 14(6): 352–361, 2012.

[46] S. Mabu, C. Chen, N. Lu, K. Shimada, and K. Hirasawa. An Intrusion-Detection model based on fuzzy class-association-rule mining using genetic programming. IEEE Transactions on System, Man and Cybernetics, Part C, 41(1): 130–139, 2011.

[47] A.O. Adetunmbi, S.O. Falaki, O.S. Adewale, and B.K. Alese. Network Intrusion Detection based on Rough Set and k-Nearest Neighbour. International Journal of Computing and ICT Research, 2(1): 60–66, 2008.

[48] R.C. Chen, K.F. Cheng, Y.H. Chen, and C.F. Hsieh Using Rough Set and Support Vector Machine for Network Intrusion Detection System. In Proceedings of Asian Conference on Intelligent Information and Database Systems: 465–470, 2009.

[49] A. Visconti and H. Tahayori. Artificial immune system based on interval type-2 fuzzy set paradigm. Applied Soft Computing, 11(6): 4055–4063, 2011.

[50] J.M. Estevez-Tapiador, P. Garcya-Teodoro, and J. E. Dyaz-Verdejo. Stochastic protocol modeling for anomaly based network intrusion detection. In Proceedings of International Workshop on Information Assurance: 3–12, 2003.

[51] A. Shabtai, U. Kanonov, and Y. Elovici. Intrusion detection for mobile devices using the knowledge-based, temporal abstraction method. Journal of System Software, 83(8): 1524–1537, 2010.

[52] S.S. Hung and D.S.M. Liu. A user-oriented ontology-based approach for network intrusion detection. Computer Standards & Interfaces, 30(1-2): 78–88, 2008.

[53] K. Noto, C. Brodley, and D. Slonim. Anomaly Detection Using an Ensemble of Feature Models. In Proceedings of IEEE International Conference on Data Mining: 953–958, 2010.

[54] R. Perdisci, D. Ariu, P. Fogla, G. Giacinto, and W. Lee. McPAD: A multiple classifier system for accurate payload-based anomaly detection. Computer Networks, 53(6): 864–881, 2009.

[55] W. Khreich, E. Granger, A. Miri, and R. Sabourin. Adaptive ROC-based ensembles of HMMs applied to anomaly detection. Pattern Recognition, 45(1): 208–230, 2012.

[56] D. Parikh and T. Chen. Data Fusion and Cost Minimization for Intrusion Detection. IEEE Transactions on Information Forensics and Security, 3(3): 381–389, 2008.

[57] R. Yan and C. Shao. Hierarchical Method for Anomaly Detection and Attack Identification in High-speed Network. Journal of Information Technology, 11(9): 1243–1250, 2012.

[58] W. Gong, W. Fu, and L. Cai. A Neural network based intrusion detection data fusion model. In Proceedings of International Joint Conference on Computational Science & Optimization: 410–414, 2010.

[59] D. Ariu, R. Tronci, and G. Giacinto. HMMPayl: An intrusion detection system based on Hidden Markov Models. Computers & Security, 30(4): 221–241, 2011.

[60] H.H. Nguyen, N. Harbi, and J. Darmont. An efficient local region and clustering-based ensemble system for intrusion detection. In Proceedings of Symposium on International Database Engineering & Applications: 185–191, 2011.

[61] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. ACM Computing Survey, 31(3): 264-323, 1999.

[62] A.L. Fred and A.K. Jain. Combining multiple clusterings using evidence accumulation. IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(6): 835-850, 2005.

[63] N. Nguyen and R. Caruana. Consensus clusterings. In Proceedings of IEEE International Conference on Data Mining: 607-612, 2007.

[64] B. Fischer and J. M. Buhmann. Bagging for path-based clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(11): 1411-1415, 2003.

[65] A. Strehl and J. Ghosh. Cluster ensembles: A knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 3: 583-617, 2002.

[66] N. Iam-On and T. Boongoen. Comparative Study of Matrix Refinement Approaches for Ensemble Clustering. Machine Learning, 98(1-2): 269-300, 2015.

[67] T. Meehinkong, P. Praneetpolgrang and N. Chirawichitchai. An Adaptive Real-Time Intrusion Detection System Based on Cybersecurity Knowledge Architecture. NKRAFA Journal of Science and Technology, 10: 71-81, 2014.

[68] P. Sirinam. UAVs from a Cyber Security Perspective: Cyber Attack Vulnerabilities and the Preparation of the RTAFA against Cyber Threats. NKRAFA Journal of Science and Technology, 10: 7-12, 2014.