บทที่ 2

ทฤษฎี หลักการ งานวิจัยที่เกี่ยวข้อง และเครื่องมือที่ใช้

ในบทนี้จะกล่าวถึงทฤษฎี หลักการ งานวิจัยที่เกี่ยวข้อง และเครื่องมือที่นำมาใช้ในการทำวิทยานิพนธ์ โดยที่เนื้อหาในส่วนแรกจะกล่าวถึงภัยคุกคามระบบคอมพิวเตอร์ การบุก-รุกระบบแบบต่าง ๆ หลังจากนั้นจะกล่าวถึงระบบตรวจจับการบุกรุก ประเภทของระบบตรวจจับการบุกรุกและงานวิจัยที่เกี่ยวข้องกับระบบตรวจจับการบุกรุก จากนั้นจะกล่าวถึงซิสเท็มคอลและการแยกประเภทข้อมูลซึ่งเป็นวิธีการทำเหมืองข้อมูล (data mining) ที่นำมาใช้ในงานวิจัยนี้ และในส่วนสุดท้ายของบทนี้จะเป็นเรื่องของเครื่องมือที่นำมาใช้ในวิทยานิพนธ์ชิ้นนี้

2.1 ภัยคุกคาม

ภัยคุกคาม (threat) เป็นเหตุการณ์ กระบวนการ กิจกรรม หรือการกระทำที่กระทำกับจุดอ่อนเพื่อโจมตีทรัพยากรที่มีค่าหรือเป็นสาเหตุให้เกิดอันตรายกับระบบคอมพิวเตอร์ ภัยคุกคามเป็นสิ่งที่นำอันตรายไปสู่เป้าหมายของการโจมตี เช่น ทรัพย์สินต่าง ๆ ในระบบเทคโนโลยีสารสนเทศ ยกตัวอย่างเช่น ข้อมูลสารสนเทศ ฮาร์ดแวร์ ซอฟต์แวร์ เอกสารต่าง ๆ ซึ่งเป็นตัวดึงดูดให้ผู้บุกรุกเข้ามาโจมตีและเป็นจุดอ่อนที่ทำให้โจมตีได้ นอกจากนี้ภัยคุกคามยังเกี่ยวข้องกับตัวกระทำภัยคุกคามคือ คนหรือเหตุการณ์ที่ไม่ปกติที่สามารถทำให้เกิดภัยคุกคามที่ชัดเจน โดยตัวกระทำจะมีลักษณะที่สำคัญ 3 อย่างคือ มีการเข้าถึง (access) มีความรู้ (knowledge) และมีเจตนา (motivation) ในส่วนของเจตนาของการคุกคาม Anderson (Anderson, 1980) ได้อธิบายเจตนาในการคุกคามระบบว่าเป็นความพยายามที่ไม่มีอำนาจซึ่งมีความจงใจที่จะทำสิ่งต่าง ๆ ดังนี้คือ การเข้าถึง การเปลี่ยนแปลงข้อมูล หรือ กระทำให้ระบบไม่น่าเชื่อถือ หรือไม่สามารถใช้งานต่อไปได้ สิ่งที่จะเป็นตัวกระทำได้ เช่น ผู้ใช้งานในระบบ ผู้บุกรุกระบบ บุคคลทั่วไป ผู้ขาย ลูกค้า นอกจากนี้ตัวกระทำภัยคุกคามยังสามารถเป็นภัยธรรมชาติได้อีกด้วย

อีกสิ่งหนึ่งที่เกี่ยวข้องกับภัยคุกคามคือเหตุการณ์ โดยเหตุการณ์จะเป็นกลไกที่ตัวกระทำสามารถทำให้เกิดความเสียหายได้ ซึ่งจะเป็นความเสียหายที่เกิดกับเป้าหมาย โดยตัวกระทำต้องมีความรู้เพียงพอที่จะทำให้เกิดเหตุการณ์ ตัวอย่างของเหตุการณ์ได้แก่ การเข้าถึงโดยไม่ได้รับอนุญาต การดักฟัง การขโมย การเปลี่ยนแปลงข้อมูล เป็นต้น

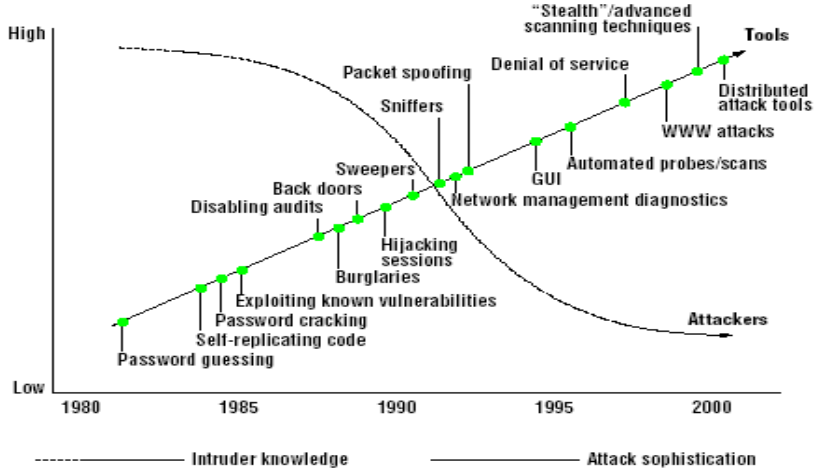
การกระทำที่ก่อให้เกิดภัยคุกคามเหล่านี้เป็นการกระทำที่ทำให้ระบบขาดความปลอดภัย ซึ่งระบบที่มีความปลอดภัยหมายถึงระบบที่มีการให้บริการในลักษณะอย่างน้อย 3 ประการดังนี้คือ

1. การรักษาความลับ (confidentiality) คือการรับรองว่าจะมีการเก็บข้อมูลไว้เป็นความลับ และผู้มีสิทธิเท่านั้นจึงจะเข้าถึงข้อมูลนั้นได้

2. การรักษาความสมบูรณ์ (integrity) คือการรับรองว่าข้อมูลจะไม่ถูกเปลี่ยนแปลงหรือถูกทำลายโดยผู้ที่ไม่มีสิทธิไม่ว่าจะเป็นโดยอุบัติเหตุหรือโดยเจตนา

3. ความพร้อมใช้ (availability) คือการรับรองว่าข้อมูลและการบริการสื่อสารต่าง ๆ พร้อมที่จะใช้ได้ในเวลาที่ต้องการใช้งาน

ในปัจจุบันเครือข่ายและคอมพิวเตอร์ถูกบุกรุกเพิ่มมากขึ้น คอมพิวเตอร์ที่เชื่อมต่อกับอินเทอร์เน็ตจะถูกคุกคามโดยไวรัส หนอนอินเทอร์เน็ต และการบุกรุกจากผู้ประสงค์ร้าย ดังนั้นจึงมีความต้องการที่จะปกป้องระบบคอมพิวเตอร์และเครือข่ายจากการบุกรุกซึ่งมีแนวโน้มที่จะเพิ่มมากขึ้นต่อไป นอกจากภัยคุกคามที่เพิ่มจำนวนมากขึ้นแล้วยังมีเรื่องของความซับซ้อนของเครื่องมือในการบุกรุกที่เพิ่มมากขึ้นด้วยเช่นกัน ในอดีตผู้ที่สามารถบุกรุกระบบได้จะต้องเป็นผู้ที่มีความสามารถสูง แต่ในระยะเวลาตั้งแต่ปี 1995 เป็นต้นมาเครื่องมือในการโจมตีได้รับการพัฒนาให้เป็นแบบ GUI (graphic user interface) ทำให้ใช้งานได้อย่างสะดวก ดังนั้นผู้บุกรุกจึงไม่จำเป็นต้องมีทักษะสูงเหมือนในอดีต หากผู้บุกรุกมีเครื่องมือที่มีประสิทธิภาพก็สามารถโจมตีเป้าหมายที่ต้องการได้ ภาพประกอบ 2.1 แสดงทักษะในการบุกรุกเปรียบเทียบเครื่องมือบุกรุกระบบคอมพิวเตอร์ที่มีประสิทธิภาพเพิ่มขึ้น



ภาพประกอบ 2.1 เปรียบเทียบความสามารถในการโจมตีกับความรู้ทางเทคนิคของผู้บุกรุก [จาก McHugh, Christie and Allen, 2000]

ภัยคุกคามนั้นมีด้วยกันหลายลักษณะและจะมีองค์ประกอบที่เกิดร่วมกันในการเกิดภัยคุกคามที่ได้กล่าวมาแล้ว ในส่วนนี้จะแบ่งภัยคุกคามแยกตามลักษณะของแหล่งที่มา และตามการเข้าถึงและการใช้งาน

2.1.1 ภัยคุกคามแบ่งตามลักษณะของแหล่งที่มา

ภัยคุกคามในระบบรักษาความปลอดภัยแบ่งตามลักษณะของแหล่งที่มาได้เป็น 2 ประเภท ดังนี้คือ

2.1.1.1 ภัยคุกคามจากธรรมชาติ

ภัยคุกคามจากธรรมชาติเป็นสิ่งที่เกิดขึ้นโดยไม่ได้มีการวางแผนและเตรียมการมาก่อน ซึ่งไม่สามารถรู้ถึงระดับของความเสียหายและการทำลายต่อระบบได้ เนื่องจากเป็นความเสียหายที่เกิดขึ้นจากธรรมชาติ เช่น น้ำท่วม แผ่นดินทรุดตัว พายุที่รุนแรงแบบต่าง ๆ เป็นต้น นอกจากนี้ยังอาจเป็นภัยธรรมชาติที่เกี่ยวข้องกับไฟฟ้า เช่น ไฟตก ไฟฟ้าลัดวงจร และการทำลายแบบอื่น ๆ ที่รวมไปถึงการสูญเสียของหน้าที่การทำงานและความสามารถในการจัดการพื้นที่ทำงาน สำหรับภัยคุกคามประเภทนี้จำเป็นต้องมีการเตรียมการรับมือล่วงหน้าที่ดีพอ และทำการสำรองข้อมูลอย่างสม่ำเสมอ นอกจากนี้ควรเก็บสำรองข้อมูลไว้ในสื่อข้อมูลอื่น ๆ เช่น ซีดีรอม แผ่นดิสก์เก็ต เป็นต้น เพื่อให้แน่ใจได้ว่าการทำงานจะสามารถดำเนินต่อไปได้หากเกิดภัยธรรม-ชาติดังกล่าวข้างต้น

2.1.1.2 ภัยคุกคามจากคน

ภัยคุกคามจากคนสามารถแบ่งได้เป็น 2 ชนิดคือ ภัยคุกคามที่ไม่เจตนาที่มีผลกระทบต่อระบบ และภัยคุกคามที่มีความตั้งใจให้เกิดความเสียหายต่อระบบ

2.1.1.2.1 ภัยคุกคามโดยไม่เจตนาหรือไม่ตั้งใจ

ผลที่เกิดขึ้นจากภัยคุกคามประเภทนี้อาจจะเป็นการเปิดเผยหรือเปลี่ยนแปลงวัตถุใด ๆ ซึ่งการเปิดเผยอาจมาจากทั้งฮาร์ดแวร์และซอฟต์แวร์ได้เท่ากับที่อาจมาจากผู้ใช้งานและการดำเนินการที่ผิดพลาด ตัวอย่างเช่น การเปิดเผยที่เกิดจากผู้ใช้งานส่งอิเล็กทรอนิกส์เมล์ (e-mail) ที่เป็นความลับไปให้บุคคลอื่นที่ไม่ได้เป็นผู้รับที่ถูกต้อง

2.1.1.2.2 ภัยคุกคามที่มีเจตนาหรือตั้งใจกระทำ

ภัยคุกคามแบบนี้เป็นภัยคุกคามที่ขึ้นอยู่กับเจตนาของผู้บุกรุกซึ่งมีความเป็นไปได้หลายอย่างดังต่อไปนี้คือ ความตั้งใจในการทำลายความปลอดภัย การเปิดเผยข้อมูลที่เป็นความลับ การเปลี่ยนแปลงทรัพยากรภายในระบบทำให้ทรัพยากรภายในระบบทำงานไม่ถูกต้อง การเพิ่มเติมข้อมูลส่งผลให้ระบบทำงานไม่ถูกต้อง การขัดจังหวะส่งผลให้การทำงานของทั้งระบบคอมพิวเตอร์และระบบเครือข่ายหยุดชะงักลงไม่สามารถทำงานต่อไปได้

ในส่วนของภัยคุกคามที่มาจากบุคคลผู้มีเจตนาร้ายหรือผู้บุกรุกนั้น Anderson ได้แบ่งผู้บุกรุกออกเป็น 2 กลุ่มคือ

ก. ผู้บุกรุกจากภายนอก (external intruder) หรือผู้ซึ่งเป็นผู้ใช้งานที่ไม่ได้รับสิทธิจากเครื่องที่เข้าโจมตี

ข. ผู้บุกรุกจากภายใน (internal intruder) หรือผู้ซึ่งได้รับอนุญาตให้เข้าถึงระบบแต่ไม่ได้รับสิทธิ์ในบางส่วน Anderson ยังแบ่ง internal intruder เพิ่มเติมเป็น masquerade intruder และ clandestine intruder โดยความหมายของผู้บุกรุกแต่ละประเภทเป็นดังนี้คือ

masquerade intruder คือ บุคคลที่ใช้ประโยชน์จากการปลอมแปลงเป็นผู้ใช้งานคนอื่นภายในระบบคอมพิวเตอร์ โดยที่ผู้บุกรุกนั้นมีชื่อและรหัสผ่านของผู้ใช้คนอื่นภายในระบบทำให้ผู้บุกรุกสามารถเข้าใช้งานภายในระบบได้เหมือนกับบุคคลที่ได้รับสิทธิ์เป็นผู้ใช้งานที่ถูกต้อง

clandestine intruder คือ ผู้บุกรุกที่มีอันตรายมากที่สุดเนื่องจากผู้บุกรุกในกลุ่มนี้เป็นผู้ที่มีความสามารถครอบครองการควบคุมสูงสุดของระบบคอมพิวเตอร์ และสามารถดำเนินการกับ audit trail ซึ่งเป็นบันทึกพฤติกรรมหรือการทำงานตามลำดับเวลาของการใช้ทรัพยากรของระบบในการรักษาความปลอดภัยคอมพิวเตอร์เพื่อหลีกเลี่ยงการบันทึกข้อมูล

2.1.2 ภัยคุกคามแบ่งตามลักษณะของการเข้าถึงและการใช้งาน

ภัยคุกคามในระบบรักษาความปลอดภัยแบ่งตามลักษณะของการเข้าถึงและการใช้งานได้เป็น 4 ระดับ ดังนี้คือ

2.1.2.1 ระดับกายภาพ

ภัยคุกคามระดับกายภาพเป็นระดับที่ภัยคุกคามนั้นจะต้องอาศัยการเข้าถึงระบบคอมพิวเตอร์และเครือข่ายในระดับกายภาพและทำงานให้เสียหาย ตัวอย่างเช่น การขโมยอุปกรณ์คอมพิวเตอร์ การตัดสายเคเบิล ไฟฟ้าลัดวงจร รวมถึงภัยคุกคามที่มาจากธรรมชาติ เช่น น้ำท่วม แผ่นดินไหว ความเสียหายของภัยระดับนี้ส่วนใหญ่จะมองเห็นหรือสัมผัสได้ด้วยตาเปล่า

2.1.2.2 ระดับเครือข่าย

ภัยคุกคามระดับเครือข่ายคือ ภัยคุกคามที่มาจากความสามารถเข้าถึงเครือข่ายได้โดยอาศัยช่องทางการสื่อสารที่มีอยู่และการใช้การสื่อสารนั้นเพื่อทำให้เครือข่ายเสียหายใช้งานไม่ได้หรือใช้เครือข่ายเพื่อวัตถุประสงค์ในการบุกรุก โดยการกระทำดังกล่าวใช้เทคนิคในระดับเครือข่าย เช่น อาศัยข้อบกพร่องของโปรโตคอล ภัยคุกคามชนิดนี้ไม่จำเป็นต้องเข้าถึงเป้าหมายในระดับกายภาพ และจะไม่ผูกติดกับโปรแกรมประยุกต์ใด ๆ โดยเฉพาะ แต่จะผูกติดกับโปร-โตคอลและวิธีการสื่อสารเป็นหลัก

2.1.2.3 ระดับโปรแกรมประยุกต์

ภัยคุกคามระดับโปรแกรมประยุกต์คือ ภัยคุกคามที่เกิดขึ้นโดยตรงกับโปรแกรมประยุกต์หรือระบบปฏิบัติการ โดยอาศัยความไม่สมบูรณ์ของโปรแกรมประยุกต์โปรแกรมใดโปรแกรมหนึ่งโดยเฉพาะมาใช้เป็นช่องทางในการบุกรุกก่อกวน หรือทำให้โปรแกรมประยุกต์นั้นเสียหายไม่สามารถให้บริการได้ ตัวอย่างเช่น โปรแกรมประยุกต์ที่มีการตรวจสอบผู้ใช้ไม่รัดกุมพอ (weak authentication) ทำให้ผู้บุกรุกสามารถเข้าไปในระบบได้โดยไม่ต้องป้อนรหัสผ่าน โปรแกรมประยุกต์ที่ตรวจสอบการรับข้อมูลจากผู้ใช้ไม่รัดกุมอาจจะถูกผู้ใช้ป้อนข้อมูลที่อยู่นอกเงื่อนไขการตรวจสอบเข้าไป ทำให้โปรแกรมประยุกต์ทำงานผิดพลาดหรือหยุดการทำงานซึ่งเป็นเทคนิคที่เรียกว่า buffer overflow หรือโปรแกรมประยุกต์ตัวใดตัวหนึ่งอาจจะซ่อนประตูกล (backdoor) ไว้เพื่อให้ผู้อื่นสามารถเข้ามาควบคุมเครื่องคอมพิวเตอร์ได้โดยไม่ต้องผ่านการตรวจ-สอบความปลอดภัย เป็นต้น การมีโปรแกรมประยุกต์ที่มีระบบรักษาความปลอดภัยบกพร่องให้บริการอยู่ไม่ว่าผู้ใช้จะสามารถเข้าถึงโปรแกรมประยุกต์ด้วยวิธีใด เครือข่ายแบบใด โปรโตคอลใดก็สามารถเป็นภัยคุกคามต่อระบบได้โดยไม่แตกต่างกัน

2.1.2.4 ระดับผู้ใช้

ภัยคุกคามในระดับผู้ใช้ได้แก่การเปิดเผยความลับของผู้ใช้เอง เช่น การเปิดเผยรหัสผ่านให้แก่ผู้อื่นเข้าไปใช้งาน การไม่รักษาความลับ การจัดเก็บรหัสผ่านไว้อย่างไม่ปลอดภัย หรือการที่ผู้ใช้ไม่ได้บังคับใช้การรักษาความปลอดภัยที่เหมาะสม ผู้ใช้มีความรู้ไม่เพียงพอ ประมาทเลินเล่อ นอกจากตัวผู้ใช้เองแล้วอาจจะมาจากผู้อื่นที่ใช้เทคนิคทางจิตวิทยาหลอกลวงผู้ใช้เพื่อให้ผู้ใช้เปิดเผยความลับ (social engineering technique) เช่น การแสร้งว่าตนเองเป็นเจ้าหน้าที่ศูนย์คอมพิวเตอร์และทำการหลอกขอรหัสผ่านจากผู้ใช้ หรือหลอกถามข้อมูลจากผู้ใช้โดยทำให้เชื่อว่าเป็นผู้มีอำนาจ เป็นต้น รวมทั้งวิธีการใด ๆ ก็ตามที่มีเป้าหมายไปยังผู้ใช้เพื่อให้ได้มาซึ่งสิทธิ์ในการใช้งานของผู้ใช้นั้น ๆ

2.2 การบุกรุก

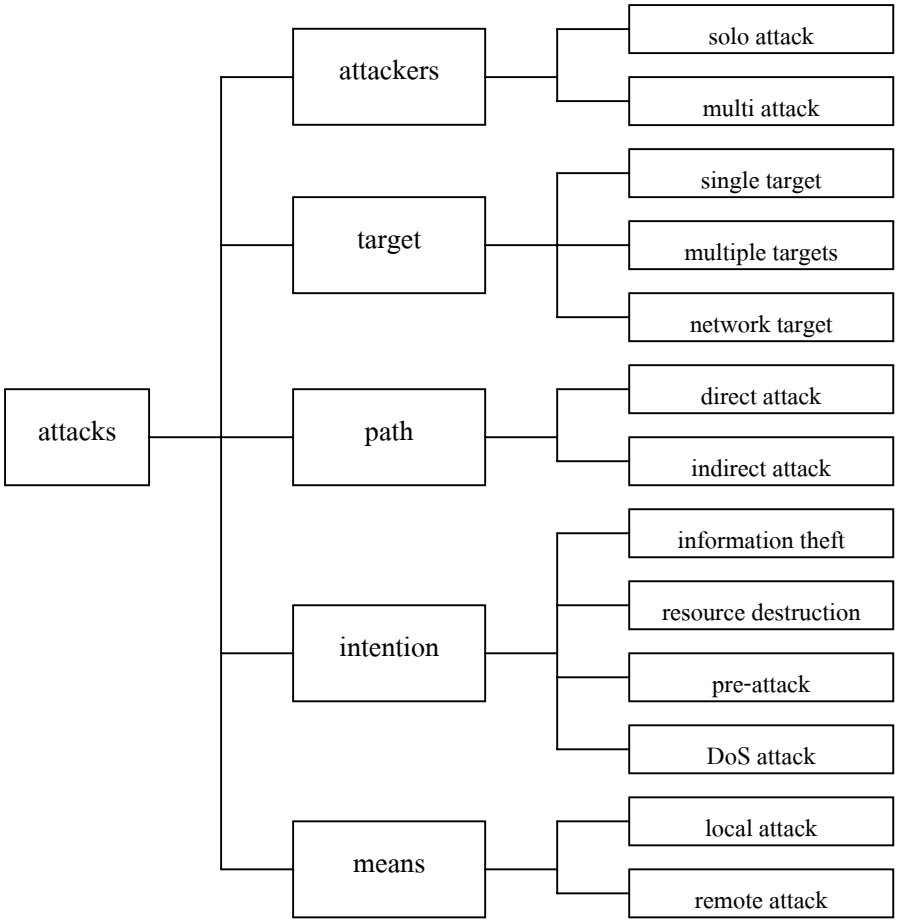
ปัญหาที่สำคัญทางความปลอดภัยของระบบคอมพิวเตอร์และเครือข่ายคือ ปัญหาที่เกิดจากการบุกรุกระบบโดยผู้บุกรุก ในส่วนนี้จะกล่าวถึงประเภทของการบุกรุกและวิธีการบุกรุกซึ่งเป็นที่รู้จักดังรายละเอียดต่อไปนี้

2.2.1 ประเภทของการบุกรุก

วิธีการที่ผู้บุกรุกนำมาใช้บุกรุกระบบเป้าหมายเพื่อให้เกิดความเสียหายหรือเพื่อนำข้อมูลที่ต้องการไปใช้ประโยชน์ต่อไปมีด้วยกันหลายวิธีการซึ่งสามารถแยกประเภทของการบุก-รุกออกเป็นกลุ่มที่สำคัญคือ การบุกรุกโดยใช้ผู้บุกรุกเป็นหลัก การบุกรุกโดยใช้ระบบเป้าหมายเป็นหลัก การบุกรุกโดยใช้เส้นทางของการบุกรุกเป็นหลัก การบุกรุกโดยใช้เจตนาของการบุกรุกเป็นหลัก และการบุกรุกโดยใช้วิธีการในการบุกรุกเป็นหลัก ภาพประกอบ 2.2 เป็นภาพรวมการบุกรุกประเภทต่าง ๆ ที่จะได้กล่าวถึงต่อไป

2.2.1.1 การบุกรุกโดยใช้ผู้บุกรุกเป็นหลัก

การบุกรุกโดยใช้ผู้บุกรุกเป็นหลักสามารถแบ่งการบุกรุกได้เป็น 2 ประเภทดังนี้



ภาพประกอบ 2.2 ประเภทของการบุกรุก

[จาก http://www.kisa.or.kr/edu/edu2001/ET5_20010507_10.pdf]

2.2.1.1.1 solo attack

solo attack คือ การบุกรุกที่ระบบคอมพิวเตอร์เป้าหมายถูกบุกรุกจากผู้บุกรุกเพียงคนเดียว โดยมากการบุกรุกจากผู้บุกรุกเพียงคนเดียวสามารถตรวจสอบได้ง่าย ตัวอย่างเช่น การบุกรุกโดยอาศัยจุดอ่อนของระบบ (system vulnerability attack) การเข้าใช้งานโดยไม่ได้รับอนุญาต (unauthorized access) เป็นต้น ภาพประกอบ 2.3 แสดงการโจมตีแบบ solo attack

ภาพประกอบ 2.3 การโจมตีแบบ solo attack
[จาก http://www.kisa.or.kr/edu/edu2001/ET5_20010507_10.pdf]

2.2.1.1.2 multi attack

multi attack คือ การบุกรุกที่ระบบคอมพิวเตอร์เป้าหมายถูกบุกรุกจากผู้บุกรุกที่ทำงานร่วมกันตั้งแต่ 2 คนขึ้นไป ตัวอย่างเช่น การบุกรุกโดยการปลอมแปลงหมายเลขไอพี (IP spoofing) การบุกรุกแบบกระตุ้นให้ผู้อื่นส่งจดหมายอิเล็กทรอนิกส์จำนวนมากไปยังระบบเดียวกัน (mail bomb) เป็นต้น ภาพประกอบ 2.4 แสดงการโจมตีแบบ multi attack

ภาพประกอบ 2.4 การบุกรุกแบบ multi attack
[จาก http://www.kisa.or.kr/edu/edu2001/ET5_20010507_10.pdf]

2.2.1.2 การบุกรุกโดยใช้ระบบเป้าหมายเป็นหลัก

การบุกรุกโดยใช้ระบบเป้าหมายเป็นหลักสามารถแบ่งการบุกรุกได้เป็น 3 ประเภท คือ

2.2.1.2.1 single target

single target คือ การบุกรุกระบบเป้าหมายซึ่งเป็นระบบคอมพิวเตอร์เพียงระบบเดียว

2.2.1.2.2 multiple targets

multiple target คือ การบุกรุกไปยังระบบคอมพิวเตอร์หลายระบบซึ่งระบบเป้าหมายของการบุกรุกประเภทนี้เป็นได้ทั้งการสุ่มระบบเป้าหมายในการบุกรุกหรือการเลือกระบบเป้าหมายเพื่อบุกรุกโดยเฉพาะ

2.2.1.2.3 network target

network target คือ การบุกรุกทรัพยากรของเครือข่าย หรือตัวเครือข่ายเองที่ถูกบุกรุก

2.2.1.3 การบุกรุกโดยใช้เส้นทางในการบุกรุกเป็นหลัก

การบุกรุกโดยใช้เส้นทางในการบุกรุกเป็นหลักสามารถแบ่งการบุกรุกได้เป็น 2 ประเภท คือ

2.2.1.3.1 direct attack คือ การบุกรุกไปยังระบบเป้าหมายโดยตรง

2.2.1.3.2 indirect attack คือ การบุกรุกที่ได้รับจากเป้าหมายหนึ่งสามารถแบ่งการบุกรุกออกได้อีก 2 แบบคือ

ก. spatial indirect attack คือ การบุกรุกผ่านแม่ข่ายที่มีจุดอ่อนไปยังเป้าหมายการบุกรุกหลัก จุดประสงค์หลักในการบุกรุกแบบนี้เพื่อซ่อนตำแหน่งของผู้บุกรุกขณะทำการบุกรุก

ข. temporal indirect attack คือ การบุกรุกที่มีการควบคุมจังหวะในการบุกรุกเพื่อสร้างความยากลำบากในการตรวจจับ

2.2.1.4 การบุกรุกโดยใช้เจตนาในการบุกรุกเป็นหลัก

การบุกรุกโดยใช้เจตนาในการบุกรุกเป็นหลักสามารถแบ่งการบุกรุกได้เป็น 4 ประเภท คือ

2.2.1.4.1 information theft คือ การบุกรุกเพื่อนำข้อมูลที่ต้องการไปใช้ประโยชน์ต่อไป

2.2.1.4.2 resource destruction คือ การบุกรุกเพื่อทำลายทรัพยากรของระบบที่มีอยู่เพื่อไม่ให้ระบบสามารถใช้ทรัพยากรเหล่านั้นได้อีกต่อไป

2.2.1.4.3 pre-attack คือ การสำรวจก่อนการบุกรุกระบบเพื่อวิเคราะห์เป้าหมายที่ต้องการบุกรุก

2.2.1.4.4 denial of service attack คือ การบุกรุกเพื่อทำให้ส่วนใดส่วนหนึ่งของระบบอัตโนมัติไม่สามารถทำงานได้ตามวัตถุประสงค์ของส่วนนั้น

2.2.1.5 การบุกรุกโดยใช้วิธีการในการบุกรุกเป็นหลัก

การบุกรุกโดยใช้วิธีการในการบุกรุกเป็นหลักสามารถแบ่งการบุกรุกได้เป็น 2 ประเภท คือ

2.2.1.5.1 local attack คือ การบุกรุกที่จะเกิดขึ้นเมื่อผู้บุกรุกได้ล็อกอิน (login) สู่ระบบเป้าหมายเรียบร้อยแล้ว

2.2.1.5.2 remote attack คือ การบุกรุกที่ผู้บุกรุกไม่จำเป็นต้องล็อกอินเข้าไปยังระบบเป้าหมาย แต่สามารถบุกรุกเพื่อขโมยข้อมูลที่ต้องการหรือกระทำให้ระบบอัตโนมัติไม่สามารถทำงานได้ตามวัตถุประสงค์ได้โดยตรง

2.2.2 วิธีการบุกรุก

ในส่วนที่ผ่านมาเราได้รู้จักกับประเภทของการบุกรุกระบบคอมพิวเตอร์ และเครือข่ายแล้ว การบุกรุกประเภทต่าง ๆ ที่ได้กล่าวมานั้นมีวิธีการบุกรุกที่สำคัญ 2 วิธีคือ การบุก-

รุกแบบพาสซีฟ (passive attack) และการบุกรุกแบบแอคทีฟ (active attack) ซึ่งมีรายละเอียดที่จะกล่าวถึงดังนี้

2.2.2.1 การบุกรุกแบบพาสซีฟ

การบุกรุกแบบพาสซีฟเป็นการบุกรุกที่ไม่ก่อให้เกิดความเสียหายแก่ระบบคอมพิวเตอร์โดยตรง ผู้บุกรุกจะใช้วิธีการใดวิธีการหนึ่งเพื่อให้ได้มาซึ่งข้อมูลที่ต้องการ การบุกรุกด้วยวิธีการนี้ได้แก่ packet sniffing, port scanning, wiretap เป็นต้น ในที่นี้จะกล่าวถึงตัวอย่างซึ่งเป็นที่รู้จักกันดีคือ packet sniffing และ port scanning

packet sniffing เป็นการดักจับแพ็คเกตในเครือข่ายโดยการใช้โปรแกรมช่วยดักจับซึ่งโปรแกรมที่เป็นที่นิยมนำมาใช้คือโปรแกรมสนิฟเฟอร์ (sniffer) ตัวอย่างของการดักจับข้อมูลได้แก่ บัญชีรายชื่อและรหัสผ่านของผู้ใช้งานซึ่งผู้บุกรุกสามารถทำได้โดยการดักจับข้อมูลในขณะที่ผู้ใช้งานใส่ชื่อและรหัสผ่านเพื่อติดต่อขอใช้บริการไปยังเครื่องเซิร์ฟเวอร์ใด ๆ ชื่อบัญชีและรหัสผ่านเหล่านั้นจะถูกส่งไปให้ผู้บุกรุกเนื่องจากข้อมูลที่วิ่งไปมาบนระบบเครือข่ายส่วนใหญ่มักจะเป็นข้อมูลที่ไม่ได้เข้ารหัสข้อมูลซึ่งทำให้ผู้บุกรุกสามารถนำข้อมูลไปใช้ประโยชน์ได้

port scanning เป็นเทคนิคที่ผู้บุกรุกใช้ในการค้นหาบริการที่จะสามารถเจาะผ่านเข้าไปยังระบบได้ สำหรับการทำ port scanning นั้น ผู้บุกรุกสามารถค้นหาข้อมูลได้จากระบบเป้าหมาย ได้แก่ บริการอะไรบ้างที่กำลังทำงานให้บริการอยู่ ผู้ใช้คนไหนเป็นเจ้าของบริการเหล่านั้น สนับสนุนการล็อกอินแบบไม่ต้องใช้รหัสผ่านหรือไม่ และบริการด้านเครือข่ายมีการตรวจสอบสิทธิของผู้ใช้งานหรือไม่ การทำ port scanning ทำได้โดยการส่งข้อความหนึ่งไปยังแต่ละพอร์ต ณ เวลาหนึ่ง ๆ ผลลัพธ์ที่ตอบสนองออกมาจะแสดงให้เห็นว่าพอร์ตนั้น ๆ ถูกใช้อยู่หรือไม่ และสามารถทดสอบดูเพื่อหาจุดอ่อนต่อไปได้หรือไม่

2.2.2.2 การบุกรุกแบบแอคทีฟ

การบุกรุกแบบแอคทีฟ เป็นการบุกรุกที่ก่อให้เกิดความเสียหายแก่ข้อมูลในระบบทำให้ระบบหยุดทำงาน หรือทำให้การติดต่อสื่อสารระหว่างเครื่องคอมพิวเตอร์ติดขัด หรือขาดการติดต่อ การบุกรุกประเภทนี้คือการโจมตีที่เรียกว่า denial of service (DoS) ซึ่งการบุกรุกด้วยวิธีการนี้ผู้บุกรุกจะขอใช้บริการที่ระบบเปิดให้บริการอยู่ โดยทำการขอจองทรัพยากรทีมีอยู่ในระบบแบบสะสมด้วยอัตราที่รวดเร็วจนกระทั่งระบบที่ให้บริการนั้นไม่มีทรัพยากรเหลือพอที่จะให้บริการผู้อื่น วิธีการที่นิยมใช้บุกรุกเป็นการสร้างแพ็คเกตขอเชื่อมต่อผ่านทางระบบเครือข่าย

โดยการใช้โปรโตคอลทีซีพีจำนวนมากที่เรียกว่า TCP SYN flooding และ ping of death ซึ่งเป็นวิธีการสร้างแพ็คเกตขนาดใหญ่ส่งไปยังบริการไอซีเอ็มพีด้วยคำสั่ง ping การโจมตีแบบ smurf ก็เป็นการบุกรุกในกลุ่มนี้ซึ่งเป็นการโจมตีไปยังเซิร์ฟเวอร์ในระดับเครือข่าย ผู้บุกรุกจะส่งไอซีเอ็ม-พีแพ็คเกตไปยัง IP broadcast address ซึ่งมี source address ของเป้าหมายที่ถูกปลอมแปลง ถ้าอุปกรณ์ routing ที่ส่ง traffic ไปยัง broadcast address ทำการแจ้ง IP broadcast ไปยังเครือข่ายฟังก์ชันการส่งสัญญาณในชั้นที่ 2 แล้วเซิร์ฟเวอร์ส่วนใหญ่บน IP network นั้น ๆ ก็จะรับ ICMP echo request และตอบกลับไปด้วย echo reply จากแต่ละเครื่องทำให้มีจำนวนการจราจรเป็นทวีคูณด้วยการตอบสนองจากเซิร์ฟเวอร์เหล่านั้นจำนวนมาก ผลที่เกิดจากการโจมตีแบบ smurf ทำให้เกิดความแออัดในการขนส่งแพ็คเกตในระบบและทำให้เครื่องเป้าหมายไม่สามารถให้บริการได้อีก ตัวอย่างอื่น ๆ ของการบุกรุกประเภทนี้ได้แก่ การโจมตีโดยการใช้โปรแกรมประเภทไวรัส (virus program) หนอนอินเทอร์เน็ต (internet worm) ม้าโทรจัน (trojan horse) ประตูกล (backdoor) สปาร์ยแวร์ (spyware) ซึ่งจะขอกล่าวรายละเอียดเฉพาะการโจมตีที่แพร่หลายได้แก่ โปรแกรมประเภทไวรัส หนอนอินเทอร์เน็ต และม้าโทรจัน ดังนี้

โปรแกรมประเภทไวรัส คือ โปรแกรมคอมพิวเตอร์ประเภทหนึ่งที่ถูกออกแบบมาให้แพร่กระจายตัวเองจากแฟ้มหนึ่งไปยังแฟ้มอื่น ๆ ภายในเครื่องคอมพิวเตอร์ การกระจายตัวของไวรัสอาจจะเป็นไปอย่างรวดเร็วไปยังทุก ๆ แฟ้มในเครื่องคอมพิวเตอร์ หรืออาจจะทำให้แฟ้มเอกสารติดเชื้ออย่างช้า ๆ ก็ได้แต่ไวรัสจะไม่สามารถแพร่กระจายจากเครื่องหนึ่งไปยังอีกเครื่องหนึ่งด้วยตัวเอง โดยทั่วไปไวรัสกระจายตัวโดยใช้ผู้ใช้งานเป็นพาหะนำไวรัสจากเครื่องหนึ่งไปยังอีกเครื่องหนึ่ง ยกตัวอย่างเช่น เวลาที่ผู้ใช้ส่งจดหมายอิเล็กทรอนิกส์โดยแนบแฟ้มที่มีไวรัสไปด้วย การทำสำเนาแฟ้มที่ติดไวรัสไปเก็บไว้บนเซิร์ฟเวอร์ การแลกเปลี่ยนแฟ้มที่ติดไวรัสโดยใช้แผ่นดิสก์เก็ต เมื่อผู้ใช้โดยทั่วไปรับแฟ้มหรือดิสก์เก็ตมาใช้งานไวรัสจะแพร่กระจายภายในเครื่องและจะเป็นวงจรในลักษณะเช่นนี้

หนอนอินเทอร์เน็ต โดยทั่วไปมีความคล้ายคลึงกับไวรัสคอมพิวเตอร์ แต่หนอนอินเทอร์เน็ตเป็นโปรแกรมที่ถูกออกแบบมาให้สามารถแพร่กระจายตัวเองจากเครื่องคอมพิวเตอร์เครื่องหนึ่งไปยังอีกเครื่องหนึ่งโดยอาศัยระบบเครือข่ายซึ่งการแพร่กระจายตัวสามารถทำได้ด้วยตัวเองอย่างรวดเร็วและรุนแรงกว่าโปรแกรมประเภทไวรัสมาก

ม้าโทรจัน เป็นโปรแกรมคอมพิวเตอร์ที่ถูกออกแบบมาให้แฝงตัวเองเข้าไปในระบบและทำงานตามที่ผู้เขียนโปรแกรมม้าโทรจันได้กำหนดไว้ เช่น การดักจับเอารหัสผ่านเข้าสู่ระบบต่าง ๆ และส่งกลับไปยังผู้บุกรุกเพื่อเข้าใช้หรือโจมตีระบบในภายหลัง ม้าโทรจันสามารถแฝงเข้ามาในระบบได้หลาย ๆ รูปแบบ อาทิเช่น เกมส์ การ์ดอวยพร หรือจดหมายต่าง ๆ

โปรแกรมม้าโทรจันไม่ได้ถูกออกแบบมาเพื่อทำลายระบบ หรือสร้างความเสียหายต่อระบบคอมพิวเตอร์ โปรแกรมม้าโทรจันต่างจากโปรแกรมประเภทไวรัส และหนอนอินเทอร์เน็ตคือ มันไม่สามารถทำสำเนาตัวเองและแพร่กระจายตัวเองได้ แต่สามารถอาศัยตัวกลางซึ่งอาจเป็นโปรแกรมต่าง ๆ จดหมาย หรือการไปโหลดแฟ้มจากแหล่งต่าง ๆ เมื่อเรียกใช้งานแฟ้มเหล่านี้ม้าโทรจันก็จะทำงานและจะเปิดช่องทางต่าง ๆ ให้ผู้บุกรุกเข้าโจมตีระบบได้

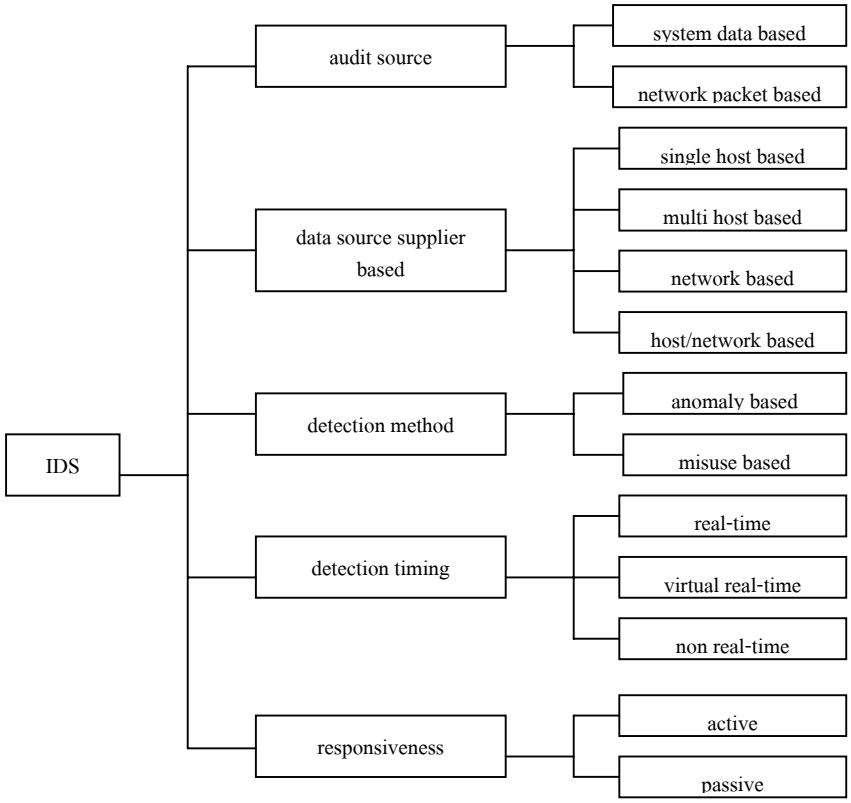
2.3 ระบบตรวจจับการบุกรุก

ระบบตรวจจับการบุกรุก (intrusion detection system (IDS)) คือระบบที่ทำหน้าที่ตรวจจับการคุกคามระบบคอมพิวเตอร์ โดยการรวบรวมข้อมูลจากแหล่งข้อมูลที่ได้มาจากระบบคอมพิวเตอร์หรือเครือข่ายแล้วทำการวิเคราะห์ข้อมูลเพื่อตรวจจับพฤติกรรมที่ไม่เหมาะสมหรือผิดปกติที่ส่งผลต่อความปลอดภัยของระบบคอมพิวเตอร์หรือเครือข่ายคอมพิวเตอร์

การทำงานที่สำคัญของระบบตรวจจับการบุกรุกคือ การตรวจสอบพฤติกรรมที่มี “ความผิดปกติ” นั่นคือพฤติกรรมที่มีความแตกต่างจากพฤติกรรมที่ควรจะเป็น Anderson ได้นำเสนอแนวคิดว่า โดยปกติแล้วพฤติกรรมของผู้ใช้งานจะมีความคล้ายคลึงกันเสมอในการใช้งานระบบในแต่ละครั้ง และเมื่อพฤติกรรมของผู้ใช้งานเปลี่ยนแปลงไปจากเดิมมากอาจจะสรุปได้ว่ามีความผิดปกติเกิดขึ้น ดังนั้นหากสามารถบันทึกพฤติกรรมการใช้งานของผู้ใช้ไว้ได้ก็อาจจะสามารถตรวจหาพฤติกรรมที่ผิดปกติได้โดยการเปรียบเทียบพฤติกรรมเหล่านี้ จากแนวความคิดของ Anderson นำไปสู่การพัฒนาระบบตรวจจับการบุกรุกในเวลาต่อมา

2.3.1 ประเภทของระบบตรวจจับการบุกรุก

เนื่องจากปัญหาของการคุกคามระบบคอมพิวเตอร์เป็นปัญหาที่มีมานาน การจัดการกับปัญหาการบุกรุกระบบจึงมีมาอย่างต่อเนื่องจนพัฒนามาเป็นระบบตรวจจับการบุกรุกซึ่งได้รับการคิดค้นพัฒนาเพื่อให้สามารถตรวจจับการบุกรุกได้อย่างมีประสิทธิภาพ ระบบตรวจจับการบุกรุกสามารถแบ่งออกได้เป็นหลายประเภทโดยใช้การแบ่งแบบต่าง ๆ กันอาทิเช่น การใช้แหล่งข้อมูลในการตรวจสอบ (audit source) การใช้ตัวให้ข้อมูล (data source supplier) การใช้วิธีในการตรวจสอบ (detection method) การใช้เวลาในการตรวจสอบ (detection timing) และการใช้การตอบสนอง (responsiveness) การแยกประเภทระบบตรวจจับการบุกรุกแสดงไว้ในภาพประกอบ 2.5 โดยที่ระบบตรวจจับการบุกรุกแต่ละประเภทมีรายละเอียดดังนี้



ภาพประกอบ 2.5 การแยกประเภทระบบตรวจจับการบุกรุก [จาก http://www.kisa.or.kr/edu/edu2001/ET5_20010507/ET5_20010507_10.PDF, 2001]

2.3.1.1 การใช้แหล่งข้อมูลในการตรวจสอบ

การแยกประเภทระบบตรวจจับการบุกรุกโดยใช้แหล่งข้อมูลการตรวจสอบเป็นหลักเกณฑ์สามารถแบ่งได้เป็น 2 ประเภท คือ

2.3.1.1.1 system data based IDS เป็นระบบตรวจจับการบุกรุกที่ใช้แฟ้มลงบันทึกเข้าออก (log file) ของระบบ และ/หรือ ลำดับของซิสเท็มคอลเพื่อใช้ในการตรวจจับความพยายามบุกรุกที่สำเร็จและล้มเหลว ข้อมูลเหล่านี้สามารถช่วยในการติดตามกิจกรรมภายในระบบได้

2.3.1.1.2 network packet based IDS เป็นระบบตรวจจับการบุกรุกที่ใช้การวิเคราะห์แพ็คเกตของเครือข่ายเพื่อตรวจสอบการบุกรุก ซึ่งแหล่งข้อมูลนี้จะมีความเป็นอิสระจาก

ระบบปฏิบัติการ และเป็นสิ่งที่ยากสำหรับผู้บุกรุกในการซ่อนหรือนำร่องรอยเหตุการณ์ที่เกี่ยวข้องกับความพยายามในการคุกคามออกไป การใช้แพ็คเกตเพื่อตรวจสอบการบุกรุกสามารถทำให้ทราบการโจมตีก่อนที่การโจมตีจะสำเร็จ แต่จะตรวจสอบแพ็คเกตที่เข้ารหัสไว้ได้ยาก

2.3.1.2 การใช้ตัวให้ข้อมูล

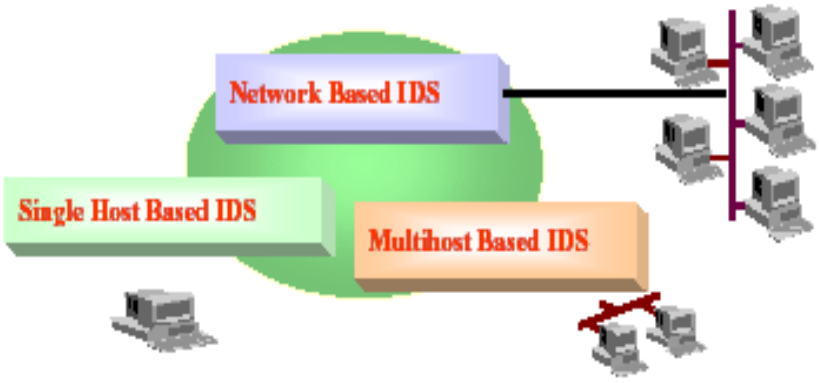
การแยกประเภทระบบตรวจจับการบุกรุกโดยใช้ตัวให้ข้อมูลเป็นหลักเกณฑ์สามารถแบ่งได้เป็น 4 ประเภท คือ

2.3.1.2.1 single host based IDS เป็นระบบตรวจจับการบุกรุกที่ได้รับข้อมูลของการบันทึกหลักฐานและกิจกรรมต่าง ๆ จากการควบคุม นโยบาย และระเบียบปฏิบัติที่จัดตั้งไว้จากเซิร์ฟเวอร์เพียงเครื่องเดียวเท่านั้น อาจมีการตรวจสอบความอ่อนแอของระบบร่วมด้วย แต่จะไม่ได้รับข้อมูลที่เป้าหมายการโจมตีเป็นเครือข่าย การใช้แหล่งข้อมูลที่มาจากเซิร์ฟเวอร์เดี่ยวนั้นเป็นวิธีการที่ง่ายและเสียค่าใช้จ่ายน้อย

2.3.1.2.2 multi host based IDS เป็นระบบตรวจจับการบุกรุกที่ได้รับข้อมูลของการบันทึกหลักฐานและกิจกรรมต่าง ๆ จากการควบคุม นโยบาย และระเบียบปฏิบัติที่จัดตั้งไว้จากเซิร์ฟเวอร์หลาย ๆ เครื่อง วิธีการนี้ทำให้สามารถตรวจสอบการโจมตีแบบผสม (multi-attack) ได้ สำหรับการนำระบบตรวจจับการบุกรุกที่ใช้แหล่งข้อมูลจากเซิร์ฟเวอร์หลาย ๆ เครื่องไปใช้งานจำเป็นต้องใช้เทคโนโลยีที่ก้าวหน้าและซับซ้อนมากขึ้น

2.3.1.2.3 network based IDS เป็นระบบตรวจจับการบุกรุกที่ดูแลอุปกรณ์ทางเครือข่ายและรวบรวมข้อมูลจากการจราจรทางเครือข่าย ข้อมูลต่าง ๆ ที่ได้รับจะถูกจัดให้ network based IDS

2.3.1.2.4 hybrid type of IDS (host/network based IDS) ระบบตรวจจับการบุกรุกแบบนี้เป็นการรวมความสามารถของ host based IDS และ network based IDS เป็นระบบตรวจจับการบุกรุกที่มีความซับซ้อนมากขึ้นแต่มีความสามารถในการทำงานเพิ่มมากขึ้น



ภาพประกอบ 2.6 hybrid type of IDS
[จาก http://www.kisa.or.kr/edu/edu2001/ET5_20010507_10.pdf]

2.3.1.3 การใช้วิธีในการตรวจสอบ

การแยกประเภทระบบตรวจจับการบุกรุกโดยใช้วิธีในการตรวจสอบเป็นหลักเกณฑ์สามารถแบ่งได้เป็นแบบ 2 ประเภท คือ

2.3.1.3.1 anomaly based IDS เป็นวิธีการตรวจจับที่ใช้พฤติกรรมการทำงานของผู้ใช้ซึ่งเก็บรวบรวมไว้เพื่อนำมาเปรียบเทียบในการตรวจจับ (profile) ยกตัวอย่างเช่น ช่วงเวลาการใช้งานของผู้ใช้ผิดไปจากเดิมหรือไม่ หรือการทำงานของผู้ใช้เปลี่ยนแปลงไปหรือไม่ เป็นต้น นอกจากนี้อาจใช้หลักการทางสถิติเข้ามาช่วย อาทิเช่น ผู้ใช้หรือโปรแกรมที่ถูกตรวจสอบมีการใช้เนื้อที่ในหน่วยความจำมากผิดปกติ เป็นต้น ระบบตรวจจับการบุกรุกแบบนี้มีความซับซ้อนเพราะต้องมีการแยกแยะให้ชัดเจนระหว่างสิ่งที่เรียกว่า “ปกติ” และ “ไม่ปกติ” โดยทั่วไปแล้วหากเกิดกลุ่มของพฤติกรรมที่ไม่แน่ใจจะจัดพฤติกรรมนั้นไว้ในกลุ่มพฤติกรรมที่เป็นการบุกรุก หรือกลุ่มที่ไม่ปกติ

2.3.1.3.2 misuse based IDS เป็นระบบตรวจจับการบุกรุกที่วิเคราะห์พฤติกรรมการบุกรุกโดยเปรียบเทียบพฤติกรรมของผู้ใช้ในขณะนั้นกับพฤติกรรมที่กำหนดไว้ หากพฤติกรรมการใช้งานผิดจากพฤติกรรมที่กำหนดไว้จะถือว่าเป็นการบุกรุก ระบบตรวจจับการบุกรุกแบบนี้สามารถตรวจจับการบุกรุกได้ดีกับพฤติกรรมที่ถูกกำหนดไว้แล้ว โดยทั่วไปแล้วระบบตรวจจับการบุกรุกที่ใช้วิธี misuse detection จะใช้กับการบุกรุกที่เกิดขึ้นภายในระบบเอง เช่น ผู้ใช้งานที่มีสิทธิใช้งานระบบแต่ใช้สิทธิของตนในทางที่ผิด เช่น ผู้ใช้งานทั่วไปพยายามเข้าไปแก้ไขแฟ้มรหัสผ่านซึ่งอนุญาตให้ผู้ใช้ที่มีสิทธิสูงสุดเท่านั้นที่สามารถกระทำได้จึงถือว่าเป็นการบุกรุกระบบเนื่องจากผิดจากสิทธิที่ได้กำหนดไว้ให้

2.3.1.4 การใช้เวลาในการตรวจสอบ

การแยกประเภทระบบตรวจจับการบุกรุกโดยใช้เวลาของการตรวจจับเป็นหลักเกณฑ์สามารถแบ่งได้เป็น 3 ประเภท คือ

2.3.1.4.1 real-time IDS เป็นระบบที่สามารถตรวจจับการบุกรุกได้ก่อนที่ผู้โจมตีจะบุกรุกระบบได้สำเร็จ

2.3.1.4.2 virtual real-time IDS เป็นระบบที่ตรวจจับการบุกรุกได้ระหว่างที่การบุกรุกอย่างหนึ่งได้เสร็จสมบูรณ์ไปแล้วและยังมีความพยายามบุกรุกอื่น ๆ เกิดขึ้น

2.3.1.4.3 non real-time IDS เป็นระบบที่มีการตรวจจับการบุกรุกที่ไม่สัมพันธ์กับเวลาที่การบุกรุกได้ดำเนินอยู่

2.3.1.5 การใช้การตอบสนอง

การแยกประเภทระบบตรวจจับการบุกรุกโดยใช้การตอบสนองเป็นหลักเกณฑ์สามารถแบ่งได้เป็น 2 ประเภท คือ

2.3.1.5.1 active IDS เป็นระบบตรวจจับการบุกรุกที่มีการจัดเตรียมแนวทางในการโต้ตอบกับการคุกคามที่ตรวจจับได้

2.3.1.5.2 passive IDS เป็นระบบตรวจจับการบุกรุกที่ทำงานเพียงตรวจสอบการบุกรุกและรายงานผลของการบุกรุกที่ตรวจจับได้เท่านั้น

2.3.2 งานวิจัยที่เกี่ยวข้อง

ระบบตรวจจับการบุกรุกที่ได้รับการพัฒนาขึ้นนั้นมีอยู่เป็นจำนวนมากเพราะการบุกรุกระบบมีการพัฒนาเปลี่ยนแปลงอยู่ตลอดเวลาจึงทำให้เครื่องมือตรวจจับและเทคนิควิธีการที่นำมาใช้เพื่อตรวจจับการบุกรุกได้รับการพัฒนามาอย่างต่อเนื่องเช่นเดียวกัน ด้วยเหตุนี้เองจึงทำให้มีระบบตรวจจับการบุกรุกเป็นจำนวนมาก ในที่นี้จะนำเสนอระบบตรวจจับการบุกรุกบางประเภทเพื่อเป็นตัวอย่างเช่น ระบบตรวจจับการบุกรุกตามวิธีในการตรวจสอบ ระบบตรวจจับการบุกรุกตามแหล่งข้อมูลในการตรวจสอบ และระบบตรวจจับการบุกรุกตามตัวให้ข้อมูล

2.3.2.1 ระบบตรวจจับการบุกรุกตามวิธีในการตรวจสอบ

ระบบตรวจจับการบุกรุกตามวิธีในการตรวจสอบสามารถแบ่งได้เป็นสองประเภท คือ misuse detection system และ anomaly detection system ดังนี้

misuse detection system ระบบตรวจจับการบุกรุกที่ถูกพัฒนาขึ้นโดยใช้วิธีการตรวจจับแบบ misuse detection เป็นระบบตรวจจับการบุกรุกที่ทำงานตรวจจับพฤติกรรมที่ใช้หลักการเปรียบเทียบรูปแบบหรือพฤติกรรมของผู้บุกรุกว่าเหมือนกันกับพฤติกรรมที่เป็นการบุกรุกหรือไม่ (pattern matching) ได้แก่ Netstat ซึ่งเป็นผลงานของ Vigna และ Kermmerer (Vigna and Kermerer, 1998) งานวิจัยในลักษณะเดียวกันได้แก่งานวิจัยของ Ilgun, Kermmerer และ Porras (Ilgun, Kermmerer and Porras, 1995) และงานวิจัยของ Kumar และ Spafford (Kumar and Spafford, 1994) นอกจากนี้ยังมีการใช้ข้อมูลที่บันทึกการใช้งานในระบบ (audit data) มาเปรียบเทียบเพื่อตรวจจับพฤติกรรมที่ถือว่าเป็นการบุกรุกระบบงานวิจัยในลักษณะนี้คือ STATL ซึ่งเป็นการทำงานร่วมกันของ Eckmann, Vigna และ Kemmerer (Eckmann, Vigna and Kemmerer, 2002) P-BEST (production based expert system toolset) ซึ่งถูกนำเสนอโดย Lindqvist และ Porras (Lindqvist and Porras, 1999) และ Bro ซึ่งเป็นงานวิจัยของ Paxon (Paxon, 1998) ตัวอย่างของงานวิจัยที่เกี่ยวข้องอื่น ๆ ได้แก่งานของ Christodorescu และ Jha (Christodorescu and Jha, 2003) เป็นการตรวจจับการบุกรุกที่เน้นในเรื่องการตรวจจับไวรัส การนำเสนอผลงานของ Habib, Hefeeda และ Bhargava (Habib, Hefeeda and Bhargava, 2003) ซึ่งเป็นการตรวจจับการบุกรุกที่เน้นการโจมตีในแบบ DoS และ QoS ซึ่งมีเป้าหมายในการขโมยทรัพยากรทางเครือข่าย เช่น แบนด์วิดธ์ (bandwidth) หรือเพื่อทำให้เกิดการให้บริการงานที่ลดต่ำลง นอกจากนี้ยังมีการตรวจจับการบุกรุกแบบ DDoS (Distributed Denial of Service) ซึ่งถูกนำเสนอโดย Paxson (Paxson, 2001)

anomaly detection system ตัวอย่างของระบบตรวจจับการบุกรุกที่ถูกพัฒนาขึ้นโดยใช้วิธีการตรวจจับแบบ anomaly detection ได้แก่ งานของ Forrest, Hofmeyr และ Somayaji (Forrest, Hofmeyr and Somayaji, 1996; Hofmeyr, Forrest and Somayaji, 1998) ใช้วิธีการตรวจจับที่รวบรวมพฤติกรรมโดยดูจากลำดับการเรียกใช้ซิสเท็มคอลเพื่อนำมาจำลองการทำงานของโปรแกรม นอกจากนี้ยังมีการตรวจจับโดยการนำเอาเรื่องของ FSA (finite state automata) มาใช้ในการติดตามการเรียกใช้งานซิสเท็มคอลซึ่ง Kosoresow และ Hofmeyr (Kosoresow and Hofmeyr, 1997) ได้นำเสนอไว้ นอกจากนี้ยังมีการนำเอาเรื่องของ data mining มาใช้โดยมีการพัฒนาโปรแกรมที่เรียกว่า RIPPER (Repeated Incremental Pruning to Produce Error

Reduction) เพื่อใช้งานในการวิเคราะห์ลำดับของการทำงานของซิสเท็มคอล (Lee, Park and Stolfo, 1999; Lee, Stolfo and Mog, 2003)

2.3.2.2 ระบบตรวจจับการบุกรุกตามแหล่งข้อมูลในการตรวจสอบ

ตัวอย่างของระบบตรวจจับการบุกรุกแบบนี้คือ IDES (Intrusion Detection Expert System) (Anderson, Frivold and Valdes, 1995) เป็นระบบตรวจจับการบุกรุกที่ใช้ audit trail ของระบบซึ่งเป็นบันทึกข้อมูลที่เกิดขึ้นตามลำดับภายในระบบที่เก็บรวบรวมพฤติกรรมการทำงานของผู้ใช้มาใช้งานในการตรวจจับ IDES เป็นระบบผู้เชี่ยวชาญการตรวจจับการบุกรุกที่มีการสร้างแฟ้มข้อมูลพฤติกรรมของผู้ใช้เก็บไว้ก่อน เมื่อมีการทำงานระบบก็จะมีการสร้างแฟ้มพฤติกรรมขึ้นใหม่จากต้นฉบับเดิม เมื่อข้อมูลเหล่านี้ผ่านการประมวลผลแล้วระบบตรวจจับการบุกรุกจะได้ค่าตัวเลขซึ่งใช้เป็นเกณฑ์ในการวัดความผิดปกติของแฟ้มข้อมูลพฤติกรรมในการตัดสินพฤติกรรมที่ผิดปกติจะตัดสินจากการเปรียบเทียบแฟ้มพฤติกรรมปัจจุบันกับแฟ้มพฤติ-กรรมที่เก็บเอาไว้ เนื่องจากการตรวจจับการบุกรุกวิธีนี้ใช้หลักสถิติและการสร้างแฟ้มพฤติกรรมในการตรวจจับจึงไม่สามารถเปลี่ยนแปลงกฎได้โดยอัตโนมัติ ทำให้ไม่สามารถวิเคราะห์การบุกรุกแบบใหม่ ๆ ได้

2.3.3.2 ระบบตรวจจับการบุกรุกตามตัวให้ข้อมูล

การจำแนกระบบตรวจจับการบุกรุกโดยพิจารณาจากตัวให้ข้อมูลที่นำมาวิเคราะห์ในตัวอย่างนี้จะเป็น ระบบตรวจจับการบุกรุกเฉพาะโฮสต์ (host-based intrusion detection system: HIDS) และระบบตรวจจับการบุกรุกทุกสถานีงานในเครือข่าย (network-based intrusion detection system: NIDS)

HIDS ตัวอย่างของระบบตรวจจับการบุกรุกในกลุ่มนี้คือ งานที่ถูกเสนอโดย Okazaki, Sato และ Goto (Okazaki, Sato and Goto, 2002) เป็นระบบตรวจจับการบุกรุกที่เก็บรวมรวมพฤติกรรมของโปรเซสมาสร้างเป็น profile เพื่อใช้งานวิเคราะห์การบุกรุก host-based ADS (automatic defense system) ซึ่งเป็นการทำงานร่วมกันของ Kreidl และ Frazier (Kreidl and Frazier, 2004) เป็นระบบตรวจจับการบุกรุกที่เน้นการตรวจจับการโจมตีของหนอนอินเทอร์เน็ตที่กระทำต่อเว็บเซิร์ฟเวอร์ โดยการเก็บรวบรวมข้อมูลการติดต่อของเครื่องลูกข่าย (client) ที่ขอใช้บริการของเว็บเซิร์ฟเวอร์และการรวบรวมข้อมูลจากเซิร์ฟเวอร์ในส่วนของ user session และการทำงานของสคริปต์ต่าง ๆ บนเซิร์ฟเวอร์

NIDS ระบบตรวจจับการบุกรุกแบบนี้จะรับข้อมูลจากเครื่องทุกเครื่องในระบบเครือข่ายแล้วนำข้อมูลไปวิเคราะห์เพื่อหาความเป็นไปได้ที่อาจจะมีผู้บุกรุกเข้ามาทำลายสถานีงานใด ๆ ในเครือข่ายคอมพิวเตอร์ ระบบนี้อาจจะอาศัยการวิเคราะห์แพ็คเกตที่ส่งไปยังสถานีงานใด ๆ ปริมาณแพ็คเกตที่ได้รับมานี้จะมีเป็นจำนวนมาก ดังนั้นระบบจึงมีความจำเป็นที่จะต้องใช้เทคนิคในการแบ่งข้อมูลเพื่อให้การทำงานมีความรวดเร็วมากยิ่งขึ้น หากตรวจพบความผิดปกติจะต้องรายงานให้ผู้ดูแลระบบทราบ นอกจากนี้อาจจะเป็นการใช้งานร่วมกับระบบไฟร์วอลล์โดยระบบตรวจจับการบุกรุกส่งข้อมูลหรือสัญญาณไปยังระบบไฟร์วอลล์เพื่อให้ทำการปิดการเชื่อมต่อของโฮสต์ที่น่าสงสัยนั้นทันที ด้วยรูปแบบการทำงานลักษณะนี้จะช่วยให้ผู้ดูแลระบบสามารถติดตามการทำงานในระบบเครือข่ายได้ตลอดเวลาโดยไม่จำเป็นต้องติดตั้งระบบตรวจจับการบุกรุกไว้ในทุกสถานีงานเหมือนแบบแรก ตัวอย่างระบบตรวจจับการบุกรุกแบบ NIDS ได้แก่ NADIR (Network Anomaly Detection and Intrusion Report) (Hochberg, *et al*., 1993) เป็นระบบที่ทำงานโดยการเปรียบเทียบการใช้งานบนเครือข่ายของผู้ใช้ในหนึ่งสัปดาห์ โดยการพิจารณาจากกฎที่ได้กำหนดไว้ก่อนล่วงหน้า NSM (Network Security Monitor) นำเสนอโดย Heberlein และคณะ (Heberlein, *et al*., 1990) และงานวิจัยของ Balupari และคณะ (Balupari, *et al*., 2003) เป็นระบบตรวจจับที่วิเคราะห์แพ็คเกตในเครือข่ายโดยพิจารณารูปแบบการใช้งานเครือ-ข่ายตามชุดโปรโตคอลทีซีพี/ไอพี และ DDIS (Distributed Intrusion Detection System) ของ Snap และคณะ (Snap, *et al*., 1991) เป็นระบบตรวจจับการบุกรุกที่มีสถาปัตยกรรมแบบกระจายซึ่งระบบตรวจจับแต่ละระบบจะเชื่อมต่อกันโดยผ่านเครือข่าย ระบบนี้มีส่วนประกอบที่สำคัญสามส่วนคือ DDIS Director, Lan Moitor และ Host Agent

ระบบตรวจจับการบุกรุกที่ได้กล่าวมามีข้อดีและข้อเสียต่าง ๆ กันไปซึ่งสามารถสรุปได้ดังต่อไปนี้คือ ระบบตรวจจับการบุกรุกที่นำแพ็คเกตมาใช้วิเคราะห์การบุกรุกเป็นระบบที่สามารถตอบสนองต่อการบุกรุกได้อย่างรวดเร็ว ไม่ขึ้นกับระบบปฏิบัติการจึงนำไปใช้งานได้ง่าย การนำเฮดเดอร์ (header) ของแพ็คเกตมาใช้ประโยชน์เพื่อตรวจสอบการบุกรุกทำให้สามารถตรวจจับวิธีการที่ผู้บุกรุกใช้โจมตีระบบได้ ตัวอย่างเช่น การโจมตีแบบ DoS การโจมตีที่อาศัยเทคนิคการแตกกระจายแพ็คเกต เช่น teardrop นอกจากนี้การตรวจสอบเฮดเดอร์ของแพ็คเกตยังทำให้สามารถตรวจพบคำสั่งหรือลำดับคำในประโยคที่ฝังเข้ามาเพื่อใช้โจมตีระบบ รวมทั้งการทดสอบช่องโหว่ของระบบ ประโยชน์ที่สำคัญอีกอย่างหนึ่งของการนำแพ็คเกตมาใช้ตรวจจับการบุกรุกคือ ผู้บุกรุกเอาเหตุการณ์การโจมตีระบบออกไปได้ยากเนื่องจากการตรวจจับแบบนี้ทำให้ตรวจพบการกระทำได้ตามเวลาที่เกิดขึ้นจริง (real-time) วิธีการตรวจสอบแพ็คเกตนี้สามารถทำได้กับแพ็คเกตที่ไม่ได้เข้ารหัสเท่านั้น นอกจากนี้แพ็คเกตทางเครือข่ายก็ไม่ได้เก็บรวบรวมพฤติกรรมหรือเหตุการณ์ต่าง ๆ ที่เกิดขึ้นภายในระบบคอมพิวเตอร์เอาไว้จึงเป็นข้อจำกัดสำหรับการนำเอาแพ็คเกตมาใช้ตรวจสอบการบุกรุก ระบบตรวจจับการบุกรุกที่ใช้ audit log ซึ่งเป็น

พฤติกรรมหรือเหตุการณ์ที่เกิดขึ้นตามลำดับเวลาภายในระบบคอมพิวเตอร์สามารถนำมาใช้ประโยชน์เพื่อยืนยันความสำเร็จในการคุกคามระบบได้ นอกจากนี้ระบบตรวจจับบุกรุกที่ใช้การตรวจสอบพฤติกรรมการใช้แฟ้มของผู้ใช้งาน การเปลี่ยนแปลงแก้ไขข้อมูล สิทธิ์ในการเรียกใช้โปรแกรม ตลอดจนการใช้บริการของระบบคอมพิวเตอร์ทำให้ทราบกิจกรรมที่เกิดขึ้นภายในระบบซึ่งทำให้สามารถตรวจพบการโจมตีโดยการใช้ม้าโทรจันและประตูกลได้ อีกทั้งยังทำให้สามารถตรวจพบการโจมตีที่กระทำผ่านการกดแป้นคีย์บอร์ดโดยตรงที่ไม่ได้เป็นการโจมตีผ่านเครือข่ายซึ่งไม่สามารถตรวจจับได้โดยการวิเคราะห์แพ็คเกต แต่การใช้ข้อมูลจากระบบคอมพิวเตอร์เพื่อตรวจจับการบุกรุกจะทำได้ก็ต่อเมื่อข้อมูลที่เกี่ยวข้องกับการบุกรุกได้รับการบันทึกลงไปแล้วเท่านั้นและถ้าผู้บุกรุกมีความสามารถในการเปลี่ยนแปลงหรือทำลายร่องรอยที่ได้บันทึกไว้โดยล็อกของระบบแล้ววิธีการตรวจสอบโดยการใช้ล็อกของระบบก็ไม่สามารถนำมาใช้ประโยชน์ได้ นอกจากนี้ระบบตรวจจับการบุกรุกที่ใช้ข้อมูลจาก audit log อาจจะไม่ได้รับข้อมูลบางอย่างที่เกิดขึ้นในระหว่างการโจมตี เช่น การบุกรุกที่ฝังมาในแพ็คเกตที่ถูกไฟร์วอลล์ทิ้งไปเป็นต้น

อย่างไรก็ตามไม่ว่าระบบตรวจจับการบุกรุกจะถูกออกแบบมาด้วยเทคนิคโดก็ตามปัญหาที่พบส่วนใหญ่ก็คือจะมีกระบวนการอย่างไรที่จะตัดสินได้ว่าพฤติกรรมแบบใดเป็นการบุกรุกและพฤติกรรมใดเป็นพฤติกรรมปกติ ปัญหาที่พบอีกปัญหาหนึ่งคือปัญหาเกี่ยวกับการจัดการกับข้อมูลจำนวนมากในการวิเคราะห์ ข้อมูลเหล่านี้อาจเป็นแพ็คเกตที่รับมาจากเครือข่ายหรือข้อมูลที่ถูกอ่านมาจากแฟ้มบันทึกการทำงานของระบบปฏิบัติการซึ่งโดยทั่วไปจะมีปริมาณมาก จากปัญหาในเรื่องของกระบวนการตัดสินพฤติกรรมการบุกรุก หรือปัญหาเกี่ยวกับขนาดของข้อมูลทำให้เกิดข้อผิดพลาดของการวิเคราะห์การบุกรุก กระบวนการตัดสินที่ผิดพลาดของระบบตรวจจับการบุกรุกมีด้วยกันสองแบบคือ การตัดสินทางบวก (false positive) และการตัดสินทางลบ (false negative) การตัดสินผิดทางบวกคือ การที่ระบบตรวจจับการบุกรุกวิเคราะห์ข้อมูลแล้วตัดสินว่ารูปแบบข้อมูลนั้นเป็นพฤติกรรมการบุกรุกทั้งที่ในความเป็นจริงแล้วรูปแบบข้อมูลนั้นเกิดจากพฤติกรรมการใช้งานตามปกติหรือได้รับอนุญาตจากระบบ อีกกรณีหนึ่งของความผิดพลาดคือการตัดสินทางลบเป็นกรณีที่ระบบตรวจจับการบุกรุกวิเคราะห์ข้อมูลแล้วตัดสินว่าไม่ได้เกิดจากพฤติกรรมหรือการกระทำที่เป็นการบุกรุกระบบทั้งที่ในความเป็นจริงแล้วเหตุ-การณ์หรือพฤติกรรมนั้นเป็นรูปแบบของการบุกรุก การตัดสินที่ผิดพลาดทั้งสองอย่างที่กล่าวมานี้จะเห็นได้ว่าการตัดสินผิดทางลบจะก่อให้เกิดผลเสียหายมากกว่าการตัดสินผิดทางบวกเนื่องจากการบุกรุกนั้นสามารถทำลายระบบได้โดยไม่ถูกตรวจจับ

ในกรณีปัญหาที่เกี่ยวข้องกับปริมาณข้อมูลจำนวนมากที่ระบบตรวจจับการบุก-รุกจะต้องวิเคราะห์นั้นอาจเนื่องมาจากการวิเคราะห์ข้อมูลโดยที่ไม่ได้มีการจัดแบ่งหรือกรองข้อมูล

ก่อน การวิเคราะห์ทำให้การตรวจจับล่าช้าหรือทำงานผิดพลาดได้โดยเฉพาะในระบบตรวจจับการบุกรุกที่ทำงานแบบเวลาจริง (real-time IDS) ซึ่งมีข้อจำกัดของเวลาเป็นตัวแปรสำคัญ ในปัจจุบันนี้ยังไม่มีระบบตรวจจับการบุกรุกแบบใดที่มีประสิทธิภาพในการทำงานดีที่สุดโดยไม่มีข้อผิดพลาด

2.4 distributor ของระบบปฏิบัติการลินุกซ์

บริษัทที่ดำเนินธุรกิจเพื่อรวบรวมและจัดจำหน่ายระบบปฏิบัติการลินุกซ์โดยเฉพาะ (distributor) ได้รวบรวมเอาโปรแกรมต้นฉบับ (source program) ที่พัฒนาและแจกจ่ายโดยทีมงานของ Linus Torvalds มาแปลเป็นภาษาเครื่อง (compile) สำหรับแต่ละแบบหรือแต่ละแพลตฟอร์ม (platform) พร้อมทั้งโปรแกรมอื่น ๆ ที่น่าใช้หรือคิดว่าผู้ใช้จำเป็นในการทำงานซึ่งส่วนมากแล้วมักจะมาด้วยวิธีการเดียวกันคือ จากโปรแกรมต้นฉบับที่มีผู้พัฒนาขึ้นแจกจ่ายแล้วนำมาปรับแต่งการทำงานต่าง ๆ ที่จำเป็นรวมถึงทดสอบเพื่อปรับปรุงแก้ไขข้อผิดพลาดต่าง ๆ หรือแจ้งกลับไปที่ผู้พัฒนา เกิดเป็นโปรแกรมสำหรับแจกจ่ายตามแบบฉบับของแต่ละราย เรียกว่า “ดิสทริบิวชัน (distribution)” ขึ้นมา ระบบปฏิบัติการลินุกซ์มีดิสทริบิวชันต่าง ๆ มากมาย อาทิ เช่น Red Hat, Mandrake, SuSE เป็นต้น (วิภา เพิ่มทรัพย์ และ สาโรจน์ ไพชยนต์ฤทธา, 2543) สำหรับการทำวิทยานิพนธ์นี้ได้นำระบบปฏิบัติการลินุกซ์ดิสทริบิวชันเรดแฮ็ท (Red Hat) มาใช้ซึ่งต่อไปจะเรียกย่อ ๆ ว่าระบบปฏิบัติการลินุกซ์เรดแฮ็ท

2.5 system call

บนระบบปฏิบัติการลินุกซ์เรดแฮ็ท ผู้ใช้สามารถเรียกใช้งานโปรแกรมหรือคำสั่งเพื่อให้ได้ผลลัพธ์ตามที่ต้องการ โดยปกติแล้วโปรแกรมหรือคำสั่งที่กำลังทำงานอยู่ (execute) จะถูกเรียกว่า “โปรเซส” (process) ซึ่งผลลัพธ์ที่เกิดจากการทำงานนั้นขึ้นอยู่กับคำสั่งหรือโปรแกรมที่ถูกเรียกใช้ โปรเซสของคำสั่งหรือโปรแกรมที่ถูกเรียกให้ทำงานจะมีการเรียกใช้งาน “ซิสเท็ม-คอล (system call)” เพื่อติดต่อกับระบบปฏิบัติการ ซิสเท็มคอลมีความสำคัญมากเพราะซิสเท็มคอลคือตัวกลางในการติดต่อระหว่างผู้ใช้งานกับระบบปฏิบัติการและอุปกรณ์ที่ให้บริการเมื่อมีการเรียกใช้คำสั่งหรือโปรแกรมซึ่งหมายความว่าไม่ว่าผู้ใช้จะเรียกใช้คำสั่งที่มีอยู่ในระบบ เรียกใช้งานโปรแกรมที่ดาวน์โหลดมาติดตั้งใช้งาน หรือเขียนโปรแกรมเพื่อทำงานตามความต้องการของผู้ใช้เอง คำสั่งหรือโปรแกรมเหล่านั้นจะมีการเรียกใช้งานซิสเท็มคอลเสมอ

จำนวนของซิสเท็มคอลที่ถูกกำหนดขึ้นบนระบบปฏิบัติการลินุกซ์เรดแฮ็ทมีอยู่เป็นจำนวนมากเพื่อให้ทำงานที่ไม่ซ้ำกัน ลักษณะที่สำคัญของซิสเท็มคอลจะมีชื่อเฉพาะตัวที่ไม่

ซ้ำกันซึ่งแต่ละชื่อนั้นจะมีหมายเลขอ้างอิงเมื่อถูกเรียกใช้งาน ในการทำงานของซิสเท็มคอลต้องมีการกำหนดพารามิเตอร์ (parameter) และเมื่อทำงานแล้วจะส่งค่ากลับคืนมาให้ (return value) ระบบปฏิบัติการลินุกซ์เรดแฮ็ทได้จัดเตรียมคำสั่งของระบบที่สามารถแสดงการทำงานของซิส-เท็มคอลที่โปรเซสเรียกใช้งานออกมาได้ คำสั่งที่กล่าวถึงนี้ชื่อว่า “strace” เมื่อเรียกใช้คำสั่ง strace เพื่อติดตามการทำงานของโปรเซสจะทำให้ทราบลำดับการเรียกใช้งานซิสเท็มคอลของโปรเซสตลอดจนพารามิเตอร์และค่าที่ถูกส่งกลับคืนมาของซิสเท็มคอลด้วย ตัวอย่างการเรียกใช้งานคำสั่ง strace บนระบบปฏิบัติการลินุกซ์เรดแฮ็ท เวอร์ชัน 6.1 เพื่อติดตามการทำงานของโปรเซส ls ซึ่งทำงานในการแสดงรายชื่อสารบบ (directory) และแฟ้มข้อมูลออกมาเป็นดังภาพประกอบ 2.7

execve(*"/bin/ls", ["ls", "-1"], [/* 25 vars */]*) = 0
brk(*0*) = 0x80549b0
open(*"/etc/ld.so.preload", O_RDONLY*) = -1 ENOENT (No such file or directory)
open(*"/etc/ld.so.cache", O_RDONLY*) = 4
fstat(*4, {st_mode=S_IFREG|0644, st_size=18296, ...}*) = 0
mmap(*0, 18296, PROT_READ, MAP_PRIVATE, 4, 0*) = 0x40013000
close(*4*) = 0
open(*"/lib/libc.so.6", O_RDONLY*) = 4
fstat(*4, {st_mode=S_IFREG|0755, st_size=4118299, ...}*) = 0
read(*4, "\177ELF\1\1\1\0\0\0\0\0\0\0\0\0\3\0\3\0\1\0\0\0\250\202"..., 4096*) = 4096
mmap(*0, 993500, PROT_READ|PROT_EXEC, MAP_PRIVATE, 4, 0*) = 0x40018000
...
ioctl(*1, TCGETS, {B38400 opost isig icanon echo ...}*) = 0
write(*1,"Desktop f1 f2 f3 f4 f5 f6" ...,64*) = 64
close(*1*) = 0
munmap(*0#40017000, 4096*)=0
_exit(*0*)=0

ภาพประกอบ 2.7 ตัวอย่างผลที่ได้จากการติดตามการทำงานของคำสั่ง ls ด้วยคำสั่ง strace

จากภาพประกอบ 2.7 แสดงให้เห็นส่วนประกอบที่สำคัญของซิสเท็มคอล 3 ส่วน ดังต่อไปนี้คือ

- ชื่อซิสเท็มคอล คือ ส่วนตัวพิมพ์หนา
- พารามิเตอร์ของซิสเท็มคอล คือ ส่วนที่อยู่ภายในวงเล็บซึ่งเป็นตัวพิมพ์เอียง

- ค่าที่ส่งกลับมาให้ คือ ส่วนที่อยู่หลังเครื่องหมายเท่ากับ

ลักษณะที่สำคัญของโปรเซสบนระบบปฏิบัติการลินุกซ์คือ โปรเซสที่เกิดขึ้นจากคำสั่งหรือโปรแกรมแต่ละอย่างจะมีเอกลักษณ์เฉพาะตัวของโปรเซสนั้น หมายความว่าโปรเซสที่เกิดจากโปรแกรมหรือคำสั่งเดียวกันจะมีการเรียกใช้งานซิสเท็มคอลที่เหมือนกันแต่อาจจะมีความแตกต่างกันบ้างเล็กน้อยขึ้นอยู่กับ option ของโปรแกรมหรือคำสั่งที่ถูกเรียกใช้งานด้วยในการทำงานนั้น

2.6 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (data mining) เป็นกระบวนการของการค้นพบฐานความรู้ของสิ่งที่น่าสนใจ อาทิเช่น รูปแบบ (pattern) ความสัมพันธ์ (association) และการเปลี่ยนแปลง (changes) เป็นต้น จากข้อมูลจำนวนมากที่ถูกเก็บเอาไว้ในฐานข้อมูล (database) โกดังข้อมูล (data warehouse) หรือคลังสารสนเทศแบบอื่น ๆ (Han, 1999)

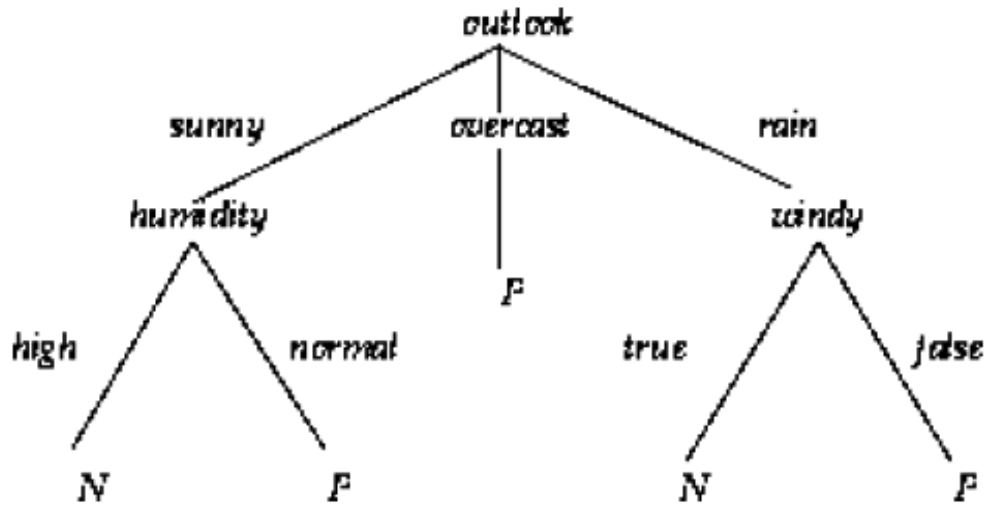
การทำเหมืองข้อมูลมีลักษณะคล้ายกับการค้นพบฐานความรู้ในฐานข้อมูล โดยทั่วไปแล้วกระบวนการในการค้นพบฐานความรู้ประกอบด้วยการดำเนินการซ้ำ ๆ ดังต่อไปนี้

- data cleaning เป็นขั้นตอนจัดการกับสิ่งรบกวน ความผิดพลาด การสูญหาย หรือความไม่เกี่ยวข้องกันของข้อมูล ออกไปจากข้อมูลที่ต้องการ
- data integration เป็นขั้นตอนการนำข้อมูลจากแหล่งต่าง ๆ มารวมเข้าไว้ด้วยกัน
- data selection เป็นขั้นตอนการนำข้อมูลที่มีความสัมพันธ์กันออกมาจากฐานข้อมูลเพื่อใช้ในการวิเคราะห์ข้อมูล
- data transformation เป็นขั้นตอนที่ข้อมูลถูกเปลี่ยนแปลงหรือรวมเข้าด้วยกันเพื่อให้ได้รูปแบบที่มีความเหมาะสมสำหรับการจัดเก็บ
- data mining เป็นขั้นตอนที่มีการประยุกต์ใช้วิธีการเพื่อสกัดข้อมูลที่ต้องการออกมาใช้ประโยชน์
- pattern evaluation เป็นขั้นตอนในการยืนยันความถูกต้องในการแทนรูปแบบของสิ่งที่สนใจ
- knowledge presentation เป็นขั้นตอนของการนำเทคนิคการนำเสนอมาใช้งานเพื่อให้ผู้ใช้สามารถนำความรู้ที่ได้รวบรวมไว้ไปใช้งานต่อไป

องค์ประกอบที่สำคัญของการทำเหมืองข้อมูลมี 3 อย่างคือ classification, association rules และ sequence analysis (Joshi, 1997) ในที่นี้จะกล่าวถึง classification เพียงอย่างเดียวเท่านั้นเพราะเป็นกระบวนการที่นำมาใช้ในงานวิจัยที่ออกแบบไว้

การแยกประเภทข้อมูล (classification) คือ การจำแนกข้อมูลโดยการเรียนรู้จากข้อมูลฝึกสอนซึ่งเป็นชุดของสิ่งที่ทราบชื่อ “คลาส (class)” แล้ว และการสร้างแบบจำลองสำหรับคลาสตามลักษณะที่สำคัญของข้อมูลเพื่อนำมาใช้งานสำหรับการกำหนดคลาสให้กับข้อมูลที่ต้องการทราบคลาส แบบจำลองที่นำมาใช้สำหรับแยกประเภทข้อมูลมีด้วยกันหลายรูปแบบ อาทิเช่น ต้นไม้การตัดสินใจ (decision tree) หรือชุดของกฎการแยกประเภทข้อมูล (classification rules) การใช้สถิติ การใช้ neural network เป็นต้น ในที่นี้จะยกตัวอย่างแบบจำลองเพื่อการแยกประเภทข้อมูลที่เป็นต้นไม้การตัดสินใจเนื่องจากเป็นการแทนความรู้ที่ง่ายและนำมาใช้งานในอัลกอริทึมหลาย ๆ อย่าง

แบบจำลองต้นไม้การตัดสินใจ ต้นไม้การตัดสินใจเป็นแบบจำลองที่ใช้งานเพื่อการทำนาย แบบจำลองนี้มีชื่อเรียกว่าต้นไม้การตัดสินใจเนื่องจากโครงสร้างของแบบจำลองมีลักษณะเหมือนต้นไม้ แต่ละกิ่งของต้นไม้ใช้สำหรับการแยกประเภทคำถาม และใบของต้นไม้เป็นการแยกกลุ่มข้อมูลออกตามคลาส ต้นไม้การตัดสินใจเป็นการแทนความรู้ที่ง่ายและสามารถแยกประเภทข้อมูลได้ตามคลาสที่มีอยู่ บัพ (node) มีชื่อตามชื่อของลักษณะประจำตัว (attribute) เส้นเชื่อม (edge) มีชื่อตามค่าที่เป็นไปได้ของลักษณะประจำตัวนั้น และใบ (leaf) มีชื่อตามคลาสต่าง ๆ สิ่งที่ต้องการแยกประเภทจะถูกแยกออกมาตามแนวระดับของต้นไม้จากระดับบนลงมาสู่ระดับล่างตามความเหมาะสมของค่าในลักษณะประจำตัวของสิ่งที่ต้องการแยกประเภท ภาพประกอบ 2.8 เป็นตัวอย่างของสิ่งที่บอกสภาพบรรยากาศตามเวลาที่กำหนดไว้ซึ่งให้ความรู้ในเรื่องของการมีเมฆมาก ความชื้นสัมพัทธ์ เป็นต้น ตัวอย่างบางสิ่งเป็นตัวอย่างในลักษณะที่ดีซึ่งแสดงด้วย P และตัวอย่างบางสิ่งเป็นตัวอย่างในลักษณะที่ไม่ดีซึ่งแสดงด้วย N เป็นต้น การแยกประเภทตามตัวอย่างทำให้เกิดโครงสร้างของต้นไม้ตามภาพประกอบ 2.8 ซึ่งสามารถนำมาใช้งานเพื่อแยกประเภทสิ่งที่ต้องการได้



ภาพประกอบ 2.8 โครงสร้างต้นไม้การตัดสินใจ [จาก Joshi, 1997]

เพื่อความเข้าใจแนวคิดของต้นไม้การตัดสินใจ พิจารณาตัวอย่างของชุดข้อมูลตัวอย่างฝึกสอนจากตาราง 2.1 และชุดข้อมูลทดสอบจากตาราง 2.2 ชุดข้อมูลทั้ง 2 ชุดต่างมีลักษณะประจำตัว 5 อย่างเหมือนกัน

ตาราง 2.1 ชุดข้อมูลตัวอย่างฝึกสอน

Outlook	Temp (F)	Humidity (%)	Windy	Class
sunny	79	90	true	no play
sunny	56	70	false	play
sunny	79	75	true	play
sunny	60	90	true	no play
overcast	88	88	false	play
overcast	63	75	true	play
overcast	88	95	false	play
rain	78	60	false	play
rain	66	70	false	no play
rain	68	60	true	no play

ที่มา : Joshi, 1997

จากตาราง 2.1 ชุดข้อมูลตัวอย่างฝึกสอนมีลักษณะประจำตัวพิเศษคือ class ใช้สำหรับบอกชื่อคลาส ลักษณะประจำตัว temp (temperature) และ humidity เป็นลักษณะประจำตัวชนิดตัวเลข และลักษณะประจำตัวอื่น ๆ แบ่งออกเป็นประเภทอย่างชัดเจน กลุ่มข้อมูลตัวอย่างฝึกสอนจะถูกนำมาหากฎต่าง ๆ เพื่อวินิจฉัยค่าของ outlook, temperature, humidity และ wind เพื่อตัดสินใจว่าจะเล่นกอล์ฟหรือไม่ ภาพประกอบ 2.9 แสดงตัวอย่างต้นไม้การตัดสินใจ

ในภาพประกอบ 2.9 มีบัพใบ 5 บัพซึ่งแต่ละบัพนำมาใช้แทนกฎการตัดสินใจหนึ่งกฎ กฎที่ได้จากต้นไม้การตัดสินใจตามภาพประกอบ 2.9 เป็นดังนี้

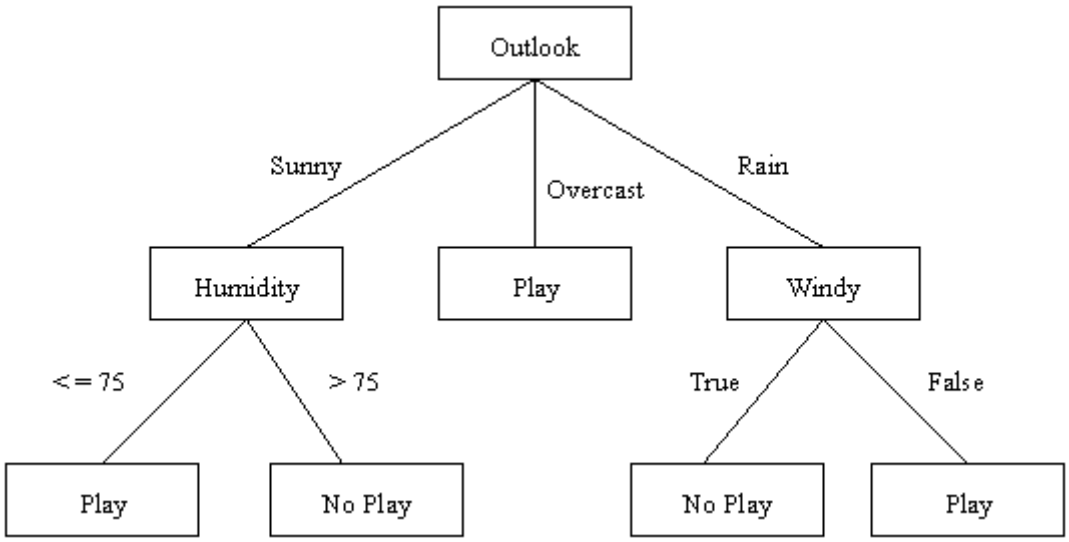
กฎข้อที่ 1 ถ้ามีแดดกล้า (sunny) และความชื้นสัมพัทธ์ (humidity) ไม่เกิน 75% แล้วจะเล่นกอล์ฟ

กฎข้อที่ 2 ถ้ามีแดดกล้าและความชื้นสัมพัทธ์เกิน 75% แล้วจะไม่เล่นกอล์ฟ

กฎข้อที่ 3 ถ้ามีเมฆมากแล้วจะเล่นกอล์ฟ

กฎข้อที่ 4 ถ้ามีฝนตก (rainy) และลมไม่แรง (windy) แล้วจะเล่นกอล์ฟ

กฎข้อที่ 5 ถ้ามีฝนตกและลมแรงแล้วจะไม่เล่นกอล์ฟ



ภาพประกอบ 2.9 ตัวอย่างต้นไม้การตัดสินใจ [จาก Joshi, 1997]

โปรดสังเกตว่ากฎที่ได้เหล่านี้อาจจะไม่ใช่กฎที่ดีที่สุดจากกลุ่มข้อมูลตัวอย่างฝึกสอนที่เตรียมไว้ การแยกประเภทเวกเตอร์นำเข้า (input vector) ที่ไม่มีชื่อถูกกระทำโดยการแวะผ่านจากบัพราก (root node) ไปยังบัพใบ (leaf node) ของต้นไม้ ระเบียนที่ต้องการทราบคลาสถูกนำเข้าไปในต้นไม้ที่ตำแหน่งของบัพราก ที่บัพรากจะมีการทดสอบเพื่อหาบัพลูกให้กับระเบียนนั้น กระบวนการทำงานเช่นนี้จะเกิดขึ้นซ้ำ ๆ กันจนกระทั่งระเบียนมาถึงบัพใบ ระเบียน

ทั้งหมดที่นำมาแยกประเภทจะได้รับการแยกประเภทด้วยวิธีการแบบเดียวกันนี้ซึ่งท้ายที่สุดแล้วจะมาอยู่ที่บัพใบของต้นไม้ มีเส้นทาง (path) เพียงเส้นเดียวจากรากที่จะไปยังใบแต่ละใบได้ เส้นทางนั้นคือกฎการตัดสินใจที่นำมาใช้งานเพื่อแยกประเภทระเบียน

ต้นไม้การตัดสินใจสามารถนำมาใช้เพื่อแยกประเภทระเบียนที่ไม่ทราบคลาสได้ ตัวอย่างเช่น สมมติว่ามีระเบียนอยู่ระเบียนหนึ่งซึ่งทราบลักษณะประจำตัวทั้ง 4 อย่างแล้ว แต่ไม่ทราบว่าระเบียนนั้นมีคลาสชื่อว่าอะไร ดังนี้

outlook= rain; temp = 70; humidity = 65; and windy= true

การหาคลาสให้กับระเบียนนี้โดยใช้ต้นไม้การตัดสินใจเริ่มต้นจากบัพรากเพื่อตรวจสอบค่าลักษณะประจำตัวของระเบียนนั้น โดยทั่วไปแล้วทุก ๆ บัพของต้นไม้การตัดสินใจจะมีลักษณะประจำตัวอย่างหนึ่งซึ่งมีความสัมพันธ์กับบัพนั้นเรียกว่า splitting attribute จากตัวอย่าง outlook คือ splitting attribute ที่บัพราก เนื่องจากลักษณะประจำตัวที่ชื่อ outlook ของระเบียนมีค่าเป็น rain ดังนั้นการแวะผ่านจึงเคลื่อนต่อไปยังบัพลูกทางด้านขวาสุดของบัพรากซึ่งที่บัพลูกนี้ splitting attribute คือ windy และเนื่องจากระเบียนที่ต้องการแยกประเภทมีค่า windy เป็น true ดังนั้นการแวะผ่านจึงเคลื่อนต่อไปยังบัพลูกทางซ้ายซึ่งทำให้ได้ชื่อคลาสคือ “no play”

หมายเหตุ ทุก ๆ เส้นทางจากบัพรากไปยังบัพใบใช้แทนกฎการตัดสินใจ นอกจากนี้ยังอาจสังเกตได้อีกว่าแต่ละใบของต้นไม้การตัดสินใจอาจมีชื่อคลาสเหมือนกันได้ แต่ใบแต่ละใบนั้นไม่ได้ใช้กฎการตัดสินใจอย่างเดียวกัน

ความถูกต้องในการทำงานของตัวแยกประเภทสามารถตัดสินได้จากอัตราร้อยละของชุดข้อมูลทดสอบที่สามารถแยกประเภทได้อย่างถูกต้อง พิจารณาชุดข้อมูลทดสอบในตาราง 2.2

ตาราง 2.2 ชุดข้อมูลทดสอบ

Outlook	Temp (F)	Humidity (%)	Windy	Class
sunny	79	90	true	play
sunny	56	70	false	play
sunny	79	75	true	no play
sunny	50	90	true	no play
overcast	88	88	false	no play

ตาราง 2.2 ชุดข้อมูลทดสอบ (ต่อ)

Outlook	Temp (F)	Humidity (%)	Windy	Class
overcast	63	75	true	play
overcast	88	95	false	play
rain	78	60	false	play
rain	66	70	false	no play
rain	68	60	true	play

ที่มา : Joshi, 1997

เมื่อพิจารณาจากกฎข้อที่ 1 พบว่ามีระเบียน 2 ระเบียนในชุดข้อมูลทดสอบที่เป็นไปตามกฎคือมีลักษณะประจำตัว outlook=sunny และ humidity <= 75 แต่มีระเบียนเพียงระเบียนเดียวเท่านั้นที่ได้รับคลาส “play” อย่างถูกต้อง ดังนั้นความถูกต้องของกฎข้อที่ 1 เป็น 0.5 (50%) ในทำนองเดียวกันความถูกต้องของกฎข้อที่ 2 เป็น 0.5 เช่นเดียวกัน ความถูกต้องของกฎข้อที่ 3 เป็น 0.66 ความถูกต้องของกฎข้อที่ 4 เป็น 0.5 และความถูกต้องของกฎข้อที่ 5 เป็น 0

อัลกอริทึมที่ใช้งานเพื่อสร้างแบบจำลองในการแยกประเภทคลาสมีอัลกอริทึมหลายอย่างด้วยกัน (Joshi, 1997) ตัวอย่างเช่น

- ID3 algorithm อัลกอริทึมนี้เป็นการสร้างต้นไม้การตัดสินใจที่กำหนดการแยกคลาสของข้อมูลโดยการทดสอบค่าของลักษณะประจำตัวของข้อมูลเหล่านั้น อัลกอริทึม ID3 สร้างต้นไม้การตัดสินใจตามการค้นหาแบบจากระดับบนลงไปสู่ระดับล่าง (top-down) เริ่มต้นจากชุดของข้อมูลและการกำหนดลักษณะประจำตัวที่แต่ละบัพของต้นไม้ มีการทดสอบลักษณะประจำตัวและใช้ผลที่ได้รับเพื่อแบ่งชุดของข้อมูล กระบวนการในการสร้างต้นไม้นี้จะถูกทำซ้ำ ๆ เพื่อสร้างต้นไม้ย่อยที่เป็นคลาสเดียวกันออกมาซึ่งกลุ่มแบบเดียวกันจะกลายเป็นบัพใบ บัพแต่ละบัพจะถูกทดสอบลักษณะประจำตัวโดยเลือกตามเกณฑ์ของข้อมูลเพื่อให้ได้ข้อมูลมากที่สุดและให้มีเอนโทรปี (entropy) ต่ำที่สุด
- C4.5 algorithm อัลกอริทึมนี้สร้างต้นไม้การตัดสินใจเพื่อการแยกคลาสสำหรับชุดข้อมูลที่กำหนดให้โดยการทำงานซ้ำ ๆ เพื่อแบ่งข้อมูล ต้นไม้ที่ถูกสร้างขึ้นเป็นการสร้างตามการค้นหาในแนวลึก (depth-first) อัลกอริ

ทึม C4.5 จะพิจารณาการทดสอบที่มีความเป็นไปได้ทั้งหมด เมื่อสามารถแบ่งชุดข้อมูลได้แล้วจะเลือกเอาการทดสอบที่ทำให้ได้รับข้อมูลที่ดีที่สุด

- SLIQ algorithm SLIQ (Supervised Learning In Quest) เป็นต้นไม้การตัดสินใจที่ถูกออกแบบมาเพื่อแยกคลาสข้อมูลตัวอย่างฝึกสอนขนาดใหญ่ อัลกอริทึมนี้ใช้เทคนิค pre-sorting ในการสร้างต้นไม้การตัดสินใจเพื่อช่วยในการลดการสูญเสียเวลาในการจัดเรียงที่เกิดขึ้นในแต่ละบัพ SLIQ แยกรายการลักษณะประจำตัวออกมาเป็น class list หน่วยใน class list เสมือนเป็นชิ้นของข้อมูลและมีชื่อคลาสกับชื่อบัพของต้นไม้การตัดสินใจ SLIQ สร้างต้นไม้การตัดสินใจตามวิธีการค้นหาในแนวกว้าง (breadth-first)

ในงานวิจัยนี้ได้นำเอาโปรแกรมที่ทำงานแยกประเภทข้อมูลมาใช้เพื่อแยกประเภทคลาสในการตรวจจับการบุกรุกซึ่งจะกล่าวถึงโปรแกรมที่นำมาใช้งานนี้ในหัวข้อถัดไป

2.7 เครื่องมือที่นำมาใช้ในงานวิจัย

เครื่องมือที่นำมาใช้ในงานวิจัยนี้แบ่งออกเป็น 2 กลุ่มใหญ่ด้วยกันคือ เครื่องมือที่ใช้แยกประเภทคลาสข้อมูล และเครื่องมือบุกรุกระบบที่สามารถหาได้จากอินเทอร์เน็ต

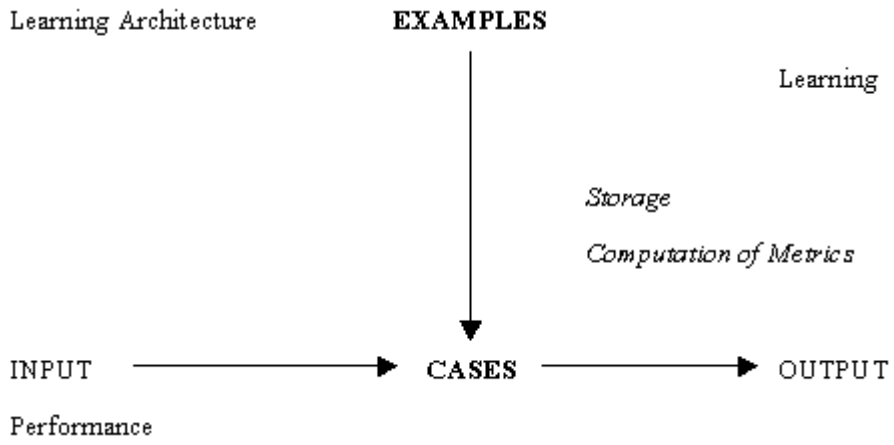
2.7.1 เครื่องมือแยกประเภทข้อมูล

เครื่องมือที่นำมาใช้แยกประเภทข้อมูลนี้เป็นซอฟต์แวร์ที่ทำงานเข้ากันได้ดีที่สุดกับงานวิจัยที่ออกแบบไว้ และเป็นซอฟต์แวร์ที่ไม่เสียค่าใช้จ่ายใด ๆ ด้วย ซอฟต์แวร์นี้มีชื่อว่า “TiMBL (Tiburg Memory-Based Learner)” TiMBL เป็นเครื่องมือที่ได้รับการพัฒนาขึ้นที่ Tilburg University นับตั้งแต่ปลายปี ค.ศ. 1980 TiMBL เป็นผลจากการนำเอาเทคนิคและอัลกอริทึมทางด้าน memory-based learning หลายอย่างมารวมกันซึ่งอัลกอริทึมแบบนี้เป็นการแทนชุดของข้อมูลฝึกสอนระบบไว้ในหน่วยความจำ TiMBL สามารถนำไปใช้กับงานค้นคว้าทางด้าน NLP (Natural Language Processing) และการแยกประเภทที่อาศัยการเรียนรู้จากกลุ่มตัวอย่างซึ่งเป็นงานที่ต้องพิจารณาข้อมูลขนาดใหญ่ มีตัวอย่างของข้อมูลจำนวนมาก และลักษณะประจำตัวของข้อมูลที่ไม่เกี่ยวข้องซึ่งกันและกันจำนวนมากได้

ระบบ MBL (memory-based learning) มีแบบแผนทางความคิดซึ่งแสดงออกมาเป็นส่วนประกอบได้ 2 อย่างคือ ส่วนประกอบทางการเรียนรู้ และส่วนประกอบทางการดำเนินการ ดังภาพประกอบ 2.10

ส่วนประกอบทางการเรียนรู้ (learning component) เป็นการนำตัวอย่างฝึกสอนไปไว้ในหน่วยความจำ (instance base หรือ case base) ตัวอย่างที่ใช้เพื่อฝึกสอนประกอบด้วยลักษณะประจำตัวที่มีจำนวนคงที่ และส่วนของคลาสเพื่อใช้สำหรับการแยกประเภท

ส่วนประกอบทางการดำเนินการ (performance component) ผลที่ได้จากส่วนประกอบทางการเรียนรู้ถูกใช้เป็นหลักเกณฑ์เพื่อจับคู่สิ่งนำเข้า (input) กับสิ่งนำออก (output) ซึ่งเป็นส่วนของกระบวนการแยกประเภท ในระหว่างการแยกประเภท ตัวอย่างทดสอบที่ไม่รู้จักมาก่อนจะถูกนำเข้าสู่ระบบ ความเหมือนกันระหว่างตัวอย่างที่เข้ามาใหม่กับตัวอย่างทั้งหมดที่มีอยู่ในหน่วยความจำจะถูกคำนวณแล้วกำหนดคลาสให้กับตัวอย่างทดสอบที่นำเข้ามาใหม่ให้มีคลาสเป็นคลาสเดียวกันกับคลาสของตัวอย่างที่พบมากที่สุดที่มีลักษณะประจำตัวเหมือนกัน



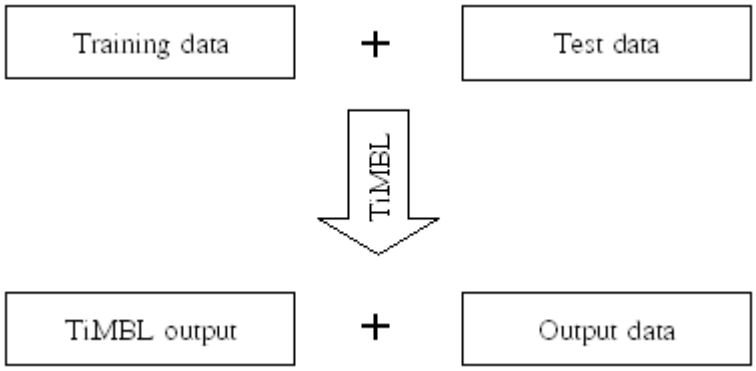
ภาพประกอบ 2.10 สถาปัตยกรรมของระบบ MBL [จาก Daelemans, *et al.*, 2002]

การทำงานของโปรแกรม TiMBL ในการทำงานของโปรแกรม TiMBL จะมีการใช้งานแฟ้มข้อมูล 4 แฟ้ม เป็นแฟ้มข้อมูลนำเข้า (input file) 2 แฟ้ม และแฟ้มข้อมูลนำออก (output file) 2 แฟ้ม ดังนี้

- แฟ้มข้อมูลฝึกสอนระบบ (training data) เป็นแฟ้มข้อมูลนำเข้าที่เก็บพฤติกรรมการทำงานและลักษณะเฉพาะตัวของโปรเซสเพื่อฝึกสอนระบบ
- แฟ้มข้อมูลทดสอบ (test data) เป็นแฟ้มข้อมูลนำเข้าอีกแฟ้มหนึ่งที่เก็บพฤติกรรมการทำงานของโปรเซสที่ต้องการให้ TiMBL กำหนดคลาสที่เหมาะสมให้
- แฟ้มข้อมูลผลลัพธ์ (TiMBL output) เป็นแฟ้มข้อมูลนำออกซึ่ง TiMBL ได้กำหนดคลาสที่เหมาะสมให้กับข้อมูลทดสอบ

- แฟ้มการเปลี่ยนทิศทางจากจอภาพ (output data) เป็นแฟ้มข้อมูลนำออกซึ่งเกิดจากการเปลี่ยนทิศทางการแสดงการทำงานต่าง ๆ ของ TiMBL ซึ่งโดยปกติแล้ว จะถูกแสดงผ่านทางจอภาพให้ไปเก็บไว้ในแฟ้มข้อมูลนี้

แฟ้มข้อมูลที่กล่าวถึงทั้งหมดนี้แสดงไว้ในแผนภาพแฟ้มข้อมูลที่เกี่ยวข้องกับการทำงานของ TiMBL ในภาพประกอบ 2.11



ภาพประกอบ 2.11 แฟ้มข้อมูลที่เกี่ยวข้องกับการทำงานของ TiMBL

ขั้นตอนการทำงานของ TiMBL การทำงานของโปรแกรม TiMBL แบ่งออกเป็น 3 ช่วงดังนี้

ช่วงที่ 1 เป็นช่วงการวิเคราะห์ข้อมูลฝึกสอนระบบ ในช่วงนี้มีการแสดงเวลาเริ่มต้นและเวลาสิ้นสุดที่ใช้ในการวิเคราะห์ข้อมูลฝึกสอนระบบ หลังจากที่วิเคราะห์ข้อมูลฝึกสอนระบบเสร็จแล้วจะรายงานจำนวนตัวอย่างฝึกสอน จำนวนคลาส และเอนโทรปีของข้อมูลฝึกสอน นอกจากนี้ยังมีการแสดงข้อมูลซึ่งเป็นรายละเอียดเกี่ยวกับลักษณะของข้อมูล (feature) แต่ละอย่าง โดยมีตัวแปร 4 ตัวคือ X-square Variance InfoGain และ GainRatio เพื่อใช้สำหรับวัดค่าความสัมพันธ์ของลักษณะข้อมูลซึ่งจะนำมาใช้เพื่อการจัดหน่วยความจำในระหว่างการฝึกสอน และใช้เพื่อถ่วงน้ำหนักความสัมพันธ์ของลักษณะข้อมูลในระหว่างการทดสอบ ในตอนท้ายของช่วงนี้จะมีการกำหนดลำดับแบบเรียงสับเปลี่ยน (permutation) ให้กับลักษณะของข้อมูลเพื่อสร้างดัชนี (tree-index) ให้กับตัวอย่างฝึกสอนที่อยู่ในหน่วยความจำ (case base) ภาพประกอบ 2.12 แสดงตัวอย่างการทำงานในช่วงที่ 1 ของโปรแกรม TiMBL

ช่วงที่ 2 เป็นช่วงการเรียนรู้ ในช่วงนี้ตัวอย่างฝึกสอนทั้งหมดจะถูกนำเข้าไปสู่หน่วยความจำเพื่อใช้ระหว่างการทดสอบ มีการแสดงข้อมูลข่าวสารการจับเวลาอีกครั้งหนึ่ง รวมทั้ง

บอกข่าวสารเกี่ยวกับขนาดของโครงสร้างข้อมูลที่เก็บไว้และการบีบขนาดข้อมูลที่ทำสำเร็จด้วย ภาพประกอบ 2.13 แสดงตัวอย่างการทำงานในช่วงที่ 2 ของโปรแกรม TiMBL

ช่วงที่ 3 เป็นช่วงที่ตัวแยกประเภทข้อมูลฝึกสอนถูกนำไปใช้กับข้อมูลที่นำมาทดสอบ ในการตัดสินความคล้ายคลึงกันของค่าในลักษณะสำคัญได้นำเอา Modified Value Difference Metric (MVDM) มาใช้ โดยการพิจารณาเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าในลักษณะสำคัญกับคลาสเป้าหมาย คลาสของข้อมูลที่นำมาทดสอบถูกตัดสินจากหลักเกณฑ์ของข้อมูลที่มีระยะทางน้อยที่สุดในหน่วยความจำ ไม่มีการคำนวณความสัมพันธ์กันระหว่างลักษณะประจำตัว ในช่วงนี้ได้แสดงเวลาความก้าวหน้าในการทำงานของช่วงการทดสอบให้เห็นด้วย ในส่วนท้ายสุดของช่วงจะมีการบันทึกจำนวนที่ตรงกันอย่างแน่นอนไว้ด้วย ชื่อของแฟ้มข้อมูลนำออกมีความหมายดังนี้ ls.txt.test.IB1.M.nw.k1.out หมายถึง แฟ้มข้อมูลนำออก (.out) ของ ls.txt.test โดยใช้อัลกอริทึม MBL (=IB1) คือ ความคล้ายคลึงกันที่คำนวณตาม Modified Value Difference Metric (.M) ไม่กำหนดน้ำหนักของความสัมพันธ์ (.nw) และจำนวนของรูปแบบของหน่วยความจำที่เหมือนกันมากที่สุดตามคลาสที่ได้รับใช้ค่าเท่ากับ 1 (.k1) ภาพประกอบ 2.14 แสดงตัวอย่างการทำงานในช่วงที่ 3 ของโปรแกรม TiMBL

```
Phase 1: Reading Datafile: /testit/train/tool_train
Start:       0 @ Tue Mar 22 21:56:34 2005
Finished:    3221 @ Tue Mar 22 21:56:36 2005
Calculating Entropy      Tue Mar 22 21:56:36 2005
Lines of data    : 3221
DB Entropy       : 3.6023477
Number of Classes : 53

Feats    Vals      X-square          Variance          InfoGain          GainRatio

  1     1 0.0000000        0.0000000        0.0000000        0.0000000
  2     4 659.32345        0.068231755      0.077659244      0.25048584
  3     1 0.0000000        0.0000000        0.0000000        0.0000000

                              ...

379     1          0.0000000        0.0000000        0.0000000        0.0000000
380     14         2368.3777        0.056560975      0.094091184      0.31752066

Feature Permutation based on GainRatio/Values :

< 290, 154, 188, 72, 140, 142, 276, 222, 86, 148, 186, 166, 18, 32, 120, 76, 92, 46, ...,
299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, ...,
376, 377, 378, 379 >
```

ภาพประกอบ 2.12 ตัวอย่างการทำงานช่วงที่ 1 ของโปรแกรม TiMBL

```
Phase 2: Learning from Datafile: /testit/train/tool_train

Start:       0 @ Tue Mar 22 21:56:36 2005
Finished:    3221 @ Tue Mar 22 21:56:37 2005

Size of InstanceBase = 363434 Nodes, (7268680 bytes), 18.47 % compression
```

ภาพประกอบ 2.13 ตัวอย่างการทำงานช่วงที่ 2 ของโปรแกรม TiMBL

```
Starting to test, Testfile: ls.txt.test
Writing output in:        ls.txt.test.IB1.M.nw.k1.out
Algorithm     : IB1
Global metric : Value Difference, Prestored matrix
Deviant Feature Metrics:(none)

Size of MVDM[1] = 8 Bytes
Size of MVDM[2] = 128 Bytes
Size of MVDM[3] = 8 Bytes
            ...
Size of MVDM[379] = 8 Bytes
Size of MVDM[380] = 1568 Bytes
Total Size of MVDM matrices 1531992 Bytes

 Weighting    : No Weighting

Tested:      1 @ Tue Mar 22 21:56:37 2005
Ready:       1 @ Tue Mar 22 21:56:37 2005
Seconds taken: 1 (1.00 p/s)

0/1 (0.000000), of which 1 exact matches
```

ภาพประกอบ 2.14 ตัวอย่างการทำงานช่วงที่ 3 ของโปรแกรม TiMBL

2.7.2 เครื่องมือที่ใช้ทดสอบระบบตรวจจับการบุกรุก

เครื่องมือที่นำมาทดสอบระบบตรวจจับการบุกรุกในงานวิจัยชิ้นนี้เป็นโปรแกรมบุกรุกระบบที่สามารถหาได้จากอินเทอร์เน็ตจากเว็บไซต์ http://pulh.as/main.php3 http://www.hack.co.za และ http://frapes.org และคำสั่งของระบบปฏิบัติการลินุกซ์เรดแฮ็ท ในการทดสอบได้แบ่งกลุ่มในการทดสอบเป็น 4 กลุ่มคือ กลุ่ม DoS กลุ่ม root compromise กลุ่ม miscellany และกลุ่ม system command ดังนี้

กลุ่ม DoS (denial of service) เป็นโปรแกรมที่ถูกสร้างขึ้นมาเพื่อทำให้ระบบใช้งานทรัพยากรอย่างสิ้นเปลืองจนไม่สามารถให้บริการงานอื่น ๆ ภายในระบบได้ หรือทำให้ระบบทำงานขัดข้อง ตลอดจนทำให้ระบบไม่สามารถทำงานได้อีกต่อไป ตัวอย่างการโจมตีของโปรแกรม เช่น การโจมตีโปรเซส inetd ซึ่งให้บริการการติดต่อทางเครือข่าย การโจมตี DNS ซึ่งทำหน้าที่แปลง IP address เป็นชื่อของโฮสต์ และแปลงกลับชื่อของโฮสต์ให้มาเป็น IP address นอกจากนี้ยังมีการโจมตีแบบอื่น ๆ อาทิเช่น การส่ง ACK packet ย้อนเข้าหาตนเอง การใช้เทคนิคเรียกใช้งานฟังก์ชัน fork() เพื่อสร้างโปรเซสจำนวนมากขึ้นในระบบทำให้มีการใช้งานซีพียูและหน่วยความจำอย่างสิ้นเปลือง หรือการใช้ช่องโหว่ที่มีอยู่ในระบบปฏิบัติการโดยกระทำผ่านซิสเท็มคอล mmap เพื่อรบกวนหน่วยความจำจนระบบไม่สามารถทำงานอื่น ๆ ต่อไปได้

กลุ่ม root compromise เป็นโปรแกรมที่สร้างขึ้นมาโดยมีวัตถุประสงค์เพื่อให้ได้รับสิทธิ์ของผู้ใช้งานสูงสุดในระบบซึ่งในที่นี้คือ สิทธิ์ของ root การบุกรุกระบบเพื่อให้ได้สิทธิ์ของ root จะใช้เทคนิคการล้นของบัฟเฟอร์ (buffer overflow) ที่เกี่ยวเนื่องกับวิธีการโปรแกรม เช่น การใช้ฟังก์ชัน sprintf(), vsprintf() หรือ strcat() โดยไม่มีการตรวจสอบขนาดของอาร์กิวเมนต์ที่ส่งมาโดยผู้บุกรุกจะส่งค่าที่มีขนาดใหญ่กว่าบัฟเฟอร์ที่ผู้พัฒนากำหนดไว้เพื่อให้พอยเตอร์เปลี่ยนตำแหน่งไปอยู่ในที่ซึ่งผู้บุกรุกต้องการและทำงานคำสั่งอื่นที่ผู้บุกรุกกำหนดไว้ นอกจากนี้ยังมีการใช้เทคนิคของการใช้จุดอ่อนของระบบผ่านทางซิสเท็มคอล ptrace ทำให้ผู้บุกรุกได้รับสิทธิ์ของ root ได้

กลุ่ม miscellany เป็นโปรแกรมประเภทประตูกล (backdoor) rootkit และโปรแกรมที่ทำงานเพื่อยกระดับสิทธิ์ในการใช้งานของผู้บุกรุกให้มากขึ้นแต่ไม่ทำให้ได้รับสิทธิ์ของผู้ดูแลระบบ โดยที่โปรแกรมประเภทประตูกลคือ โปรแกรมที่ผู้บุกรุกนำมาใช้งานเพื่อให้สามารถกลับเข้าสู่ระบบได้ง่ายขึ้น ตัวอย่างของประตูกลได้แก่ login backdoor ทำให้ผู้บุกรุกสามารถเข้าใช้งานเป็นผู้ใช้ในระบบโดยใช้รหัสผ่านของประตูกล service backdoor เป็นประตูกลที่นำมาแทนที่บริการของระบบ เช่น FTP, rlogin เป็นต้น โปรแกรมประเภท rootkit เป็นเครื่องมือหรือกลุ่มของเครื่องมือที่ถูกรวบรวมเข้าไว้ด้วยกันเพื่อให้ผู้บุกรุกซ่อนตัวและในขณะเดียวกัน rootkit ก็ทำให้ผู้บุกรุกได้รับสิทธิ์ของผู้ใช้สูงสุดในระบบ

กลุ่ม system command เป็นคำสั่งของระบบปฏิบัติการลินุกซ์เรดแฮ็ทที่เลือกมาใช้งานเป็นกรณีตัวอย่าง

2.8 สรุป

ในบทนี้ได้นำเสนอทฤษฎี หลักการ และตัวอย่างการวิจัยพัฒนาระบบตรวจจับการบุกรุก ตลอดจนเครื่องมือที่นำมาใช้ในงานวิจัยชิ้นนี้ ในบทถัดไปจะนำเสนอในส่วนของการวิเคราะห์และการออกแบบระบบตรวจจับการบุกรุกที่จะพัฒนาต่อไป