บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา

จากการพัฒนาอย่างรวดเร็วของโครงข่ายอินเทอร์เน็ตนั้น ทำให้คนส่วนใหญ่หันมาตระหนักถึงการรักษาความปลอดภัยกันมากขึ้น โดยปกติแล้วการบุกรุกจะเน้น 3 ด้านคือ การละเมิดความเป็นส่วนตัวหรือความลับ การแก้ไขความถูกต้องของข้อมูล และการทำให้ไม่สามารถใช้งานระบบคอมพิวเตอร์ได้ วิธีการหนึ่งที่นิยมนำมาใช้ในการสร้างความปลอดภัยให้กับระบบเครือข่ายคอมพิวเตอร์ คือ การตรวจจับการบุกรุก (Intrusion Detection) ซึ่งการตรวจจับการบุกรุก สามารถแบ่งออกได้เป็น 2 ชนิดคือ ระบบการตรวจจับการบุกรุกแบบโฮสเบส (host-based intrusion detection systems) และระบบตรวจจับการบุกรุกแบบเน็ตเวิร์คเบส (network-based intrusion detection systems) โดยที่ระบบตรวจจับการบุกรุกแบบเน็ตเวิร์คเบสนั้นจะติดตั้งที่ระบบเครือข่ายเพื่อทำการตรวจสอบและวิเคราะห์ชุดข้อมูลที่ใช้งานบนเครือข่าย ซึ่งมีความแตกต่างจากระบบการตรวจจับการบุกรุกแบบโฮสเบส ที่จะทำงานอยู่บนระบบเพื่อตรวจสอบและวิเคราะห์ชุดคำสั่งเพื่อระบุการทำงานที่น่าสงสัย วิธีการของการตรวจจับการบุกรุกสามารถแบ่งออกได้เป็น 2 ชนิด คือ วิธีการตรวจจับการบุกรุกแบบอโนมาลี (anomaly intrusion detection method) และวิธีการตรวจจับการบุกรุกแบบมิสยูส (misuse intrusion detection method) โดยที่วิธีการตรวจจับการบุกรุกแบบอโนมาลี นั้นเป็นวิธีการหาผู้บุกรุกโดยการวิเคราะห์การใช้งานของผู้ใช้งาน หรือตัวระบบเองที่เบี่ยงเบนไปจากระดับการใช้งานโดยปกติ ส่วนการตรวจจับการบุกรุกแบบมิสยูสนั้น เป็นวิธีการหาผู้บุกรุกโดยการเปรียบเทียบข้อมูลที่เข้ามากับรูปแบบของผู้บุกรุกที่มีอยู่เดิม ซึ่งทั้งสองวิธีนี้มีจุดแข็งและจุดอ่อนที่แตกต่างกัน ปัญหาที่เด่นชัดที่สุดของการตรวจจับการบุกรุกแบบมิสยูส คือ ไม่สามารถตรวจจับการบุกรุกแบบใหม่ หรือการบุกรุกที่ไม่มีในชุดรูปแบบของผู้บุกรุกที่มีได้ ส่วนการตรวจจับการบุกรุกแบบอโนมาลีนั้น จะระบุว่าการใช้งานที่ตรวจสอบนั้นเป็นผู้บุกรุกหรือไม่นั้นจะตรวจสอบจากการใช้งานนั้นว่ามีการเบี่ยงเบนจากกิจกรรมปกติมากหรือไม่ ดังนั้นการตรวจจับการบุกรุกแบบอโนมาลีจะสามารถตรวจจับการบุกรุกจากผู้บุกรุกที่ไม่มีในฐานข้อมูลการบุกรุกได้ แต่ปัญหาที่สำคัญในการตรวจจับการบุกรุกแบบอโนมาลีคือ ทำอย่างไรถึงจะสร้างเค้าโครงของการใช้งานปกติที่ดีได้ เพราะถ้าสร้างเค้าโครงการใช้งานปกติกว้างไป การบุกรุกบางชนิดอาจจะไม่สามารถถูกตรวจพบ ซึ่งเป็นผลให้การตรวจจับมีประสิทธิผลค่อนข้างต่ำ แต่ในทางกลับกันถ้ากำหนดเค้าโครงการใช้งานปกติแคบเกินไป อาจจะทำให้การใช้งานปกติบางอย่าง ถูกตรวจพบว่าเป็นการบุกรุกได้ เป็นผลทำให้มีโอกาสเกิดข้อผิดพลาดมากและ อาจทำให้ลดประสิทธิภาพการทำงานของระบบโดยรวมได้

จากปัญหาที่พบข้างต้นนั้น ได้มีความพยายามที่จะพัฒนาประสิทธิภาพของการตรวจจับการบุกรุกโดยนำวิธีการต่างๆ เพื่อมาช่วยในการแทนข้อมูลและการรู้จำหรือระบุผู้บุกรุก โดยมุ่งเน้นที่ความสัมพันธ์ 2 ด้านคือ การระบุผู้บุกรุกที่ถูกต้องและมีประสิทธิภาพที่ดี และการหาวิธีการแทนข้อมูลที่ดี โดยทั่วๆ ไปมีหลายวิธีถูกนำมาสร้างเป็นต้นแบบเพื่อระบุผู้บุกรุก โดยใช้องค์ประกอบหลักหรือคุณลักษณะเด่นในการศึกษาและการตรวจจับผู้บุกรุก ตัวอย่างเช่น การวิเคราะห์ค่าตัวแปรแบบเบย์ (Bayesian Parameter Estimation) การรวมข้อมูล (Data Fusion) และเครือข่ายประสาทเทียม (Neural Network) ทำให้ปัญหาการตรวจจับการบุกรุกสามารถพิจารณาได้ในลักษณะเดียวกับปัญหาการแบ่งกลุ่ม (Classification Problem) โดยที่ปัญหาการแบ่งกลุ่มนั้นวิธีการที่ให้ผลลัพธ์ที่ดี คือ วิธีการเครือข่ายประสาทเทียม แต่อย่างไรก็ตามการตรวจจับการบุกรุกนั้นควรจะประมวลผลข้อมูลที่ต้องการตรวจสอบทั้งที่เป็นกรณีที่เป็นการบุกรุก และกรณีที่ไม่ใช่การบุกรุก ซึ่งการตรวจสอบดังกล่าวทำให้ข้อมูลที่ตรวจสอบมีปริมาณมากทั้งจำนวนข้อมูล และจำนวนคุณลักษณะของข้อมูล เป็นผลทำให้เกิดความล่าช้าในการระบุผู้บุกรุก และอาจเป็นสาเหตุให้การบุกรุกบางชนิดสามารถบุกรุกเข้าสู่ระบบเครือข่ายได้

จากที่ได้กล่าวมาทั้งหมดนั้น ผู้วิจัยได้แสดงให้เห็นแล้วว่าการสกัดคุณลักษณะของชุดข้อมูลบนเครือข่าย มีความสำคัญต่อการพัฒนาการระบุผู้บุกรุกเป็นอย่างมาก จึงจำเป็นที่จะต้องหาวิธีการที่ดีในการสกัดคุณลักษณะของชุดข้อมูลบนเครือข่าย เพื่อให้ได้ตัวแทนชุดคุณลักษณะของชุดข้อมูลที่เหมาะสมเพื่อใช้ในการระบุผู้บุกรุกโดยอาศัยวิธีการแบบผสมในการสกัดคุณลักษณะของชุดข้อมูลเครือข่าย ซึ่งขั้นตอนนี้จะมีความซับซ้อนมากกว่าวิธีการสกัดคุณลักษณะของชุดข้อมูลแบบปกติทั่วไป แต่อาจจะสามารถเพิ่มความสามารถในการระบุผู้บุกรุกได้เหมาะสมมากกว่า การพัฒนาการสกัดคุณลักษณะชุดข้อมูลเครือข่าย ประกอบไปด้วย 2 ขั้นตอน คือ 1.ขั้นตอนการหาคุณลักษณะของชุดข้อมูลที่สามารถแทนข้อมูลได้และมีจำนวนคุณลักษณะที่เหมาะสม และขั้นตอนที่ 2 การรู้จำรูปแบบการบุกรุกเพื่อระบุผู้บุกรุกจากชุดข้อมูลบนเครือข่าย จากคุณลักษณะที่ได้จากการสกัดคุณลักษณะของชุดข้อมูล โดยวัดประสิทธิภาพจากอัตราความเร็วในการตรวจจับผู้บุกรุก และเปอร์เซ็นต์ความผิดพลาดของการตรวจจับผู้บุกรุก

1.2 วัตถุประสงค์ของโครงการวิจัย

1. เพื่อศึกษาเทคนิคการสกัดคุณลักษณะเด่นของชุดข้อมูล
2. เพื่อศึกษาวิธีการแบบผสมสำหรับการสกัดคุณลักษณะเด่นของชุดข้อมูล
3. เพื่อศึกษาการรู้จำคุณลักษณะของการระบุผู้บุกรุกเครือข่าย
4. เพื่อศึกษาและพัฒนาโปรแกรมที่ใช้วิธีการแบบผสมสำหรับการสกัดคุณลักษณะของชุดข้อมูลบนเครือข่าย เพื่อนำไปใช้กับการระบุผู้บุกรุกในสถานการณ์จริง

5. เพื่อให้ผู้ที่สนใจสามารถนำแนวความคิดที่นำเสนอ ไปศึกษาเพื่อทำการพัฒนาหรือประยุกต์ใช้ในงานวิจัยของตนเองต่อไป

1.3 ขอบเขตของโครงการวิจัย

การวิจัยครั้งนี้มุ่งที่จะศึกษาและพัฒนาการนำวิธีการแบบผสมสำหรับการสกัดคุณลักษณะชุดข้อมูลเครือข่าย การรู้จำเพื่อระบุผู้บุกรุกเครือข่าย มาใช้งานร่วมกันเพื่อสกัดคุณลักษณะของข้อมูลและการรู้จำผู้บุกรุกที่เป็นลักษณะของการบุกรุกเครือข่าย โดยมีขอบเขตดังต่อไปนี้

ชุดข้อมูลเครือข่าย เป็นข้อมูลมาตรฐานที่ใช้ในการทำแบบทดลองการตรวจจับการบุกรุก สามารถดาวน์โหลดได้เว็บไซต์ http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

ลักษณะของข้อมูลที่นำมาทำแบบทดสอบ จะประกอบด้วยข้อมูล 5 กลุ่มหลัก ซึ่งได้แก่ Normal, DoS, Probe, R2L และ U2R และมีรูปแบบของฟีเจอร์ (Feature) ดังแสดงตามรูปที่ 1-1

```
0,tcp,pop_3,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,211,6,0.00,0.00,1.00,1.00,0.03,0.07,0.00,255,6,0.02,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.
0,tcp,pop_3,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,231,16,0.00,0.00,1.00,1.00,0.07,0.06,0.00,255,16,0.06,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.
0,tcp,pop_3,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,232,5,0.00,0.00,1.00,1.00,0.02,0.06,0.00,255,5,0.02,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.
0,tcp,pop_3,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,254,15,0.00,0.00,1.00,1.00,0.06,0.07,0.00,255,15,0.06,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.
0,tcp,pop_3,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,252,6,0.00,0.00,1.00,1.00,0.02,0.07,0.00,255,6,0.02,0.08,0.00,0.00,0.00,0.00,1.00,1.00,neptune.
0,tcp,pop_3,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,272,16,0.00,0.00,1.00,1.00,0.06,0.06,0.00,255,16,0.06,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.
0,tcp,pop_3,SH,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1.00,1.00,0.00,0.00,1.00,0.00,0.00,255,1,0.00,1.00,1.00,0.00,1.00,1.00,0.00,0.00,nmap.
0,tcp,pop_3,SH,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1.00,1.00,0.00,0.00,1.00,0.00,0.00,255,1,0.00,1.00,1.00,0.00,1.00,1.00,0.00,0.00,nmap.
5,tcp,pop_3,SF,6,151,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,511,1,0.07,0.00,0.91,0.00,0.00,1.00,0.00,255,1,0.00,1.00,0.00,0.00,0.07,0.00,0.90,0.00,satan.
40339,tcp,pop_3,RSTR,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,1.00,1.00,1.00,0.00,0.00,255,2,0.01,0.44,0.86,0.00,0.00,0.00,0.86,1.00,portsweep.
```

รูปที่ 1-1 ตัวอย่างของข้อมูลที่ใช้ในงานวิจัย

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ขั้นตอนวิธีการสกัดคุณลักษณะของชุดข้อมูลเครือข่าย โดยอาศัยวิธีการแบบผสมสำหรับการสกัดคุณลักษณะชุดข้อมูลเครือข่าย และการรู้จำการระบุผู้บุกรุก ซึ่งเป็นการสกัดคุณลักษณะในรูปแบบใหม่ที่แตกต่างจากการสกัดคุณลักษณะแบบเดิม โดยการนำการสกัดแบบเดิมผสมกับตัวแบบโครงข่ายประสาทเทียม
2. สามารถนำไปพัฒนาระบบตรวจจับการบุกรุกบนระบบเครือข่ายได้ ซึ่งจะทำให้ช่วยรักษาความปลอดภัยเครือข่ายและให้การใช้งานเครือข่ายมีประสิทธิภาพมากขึ้น
3. ขั้นตอนวิธีที่นำเสนอจะสามารถใช้เป็นต้นแบบในการศึกษาขั้นสูงต่อไป

1.5 ระยะเวลาทำการวิจัยและแผนการดำเนินงานตลอดโครงการวิจัย

ดำเนินการวิจัยแบบ 2 ปีต่อเนื่อง (ปีงบประมาณ 2554-2555)

แผนการดำเนินงานปีที่ 1 (ปีงบประมาณ 2554-2555)

แผนการดำเนินงานวิจัย	เดือนที่											
	1	2	3	4	5	6	7	8	9	10	11	12
การจัดเตรียมเครื่องมือ อุปกรณ์และวัสดุวิจัย	⇨											
การศึกษาการสกัดคุณลักษณะชุดข้อมูล	→	→	→	→	→	→	→					
การจัดทำรายงานความก้าวหน้าครั้งที่ 1							⇨					
การศึกษาวิธีการแบบผสมสำหรับการสกัดคุณลักษณะชุดข้อมูล							→	→	→	→	→	→
การจัดทำรายงานความก้าวหน้าครั้งที่ 2												⇨

แผนการดำเนินงานปีที่ 2 (ปีงบประมาณ 2555-2556)

แผนการดำเนินงานวิจัย	เดือนที่											
	1	2	3	4	5	6	7	8	9	10	11	12
การศึกษาระบบโครงข่ายประสาทเทียมเพื่อการรู้จำและระบุผู้บุกรุก	→	→	→	→	→	→	→					
การจัดทำรายงานความก้าวหน้าครั้งที่ 3							⇨					
พัฒนาเป็นโปรแกรม Application สำหรับผู้ใช้ทั่วไป								→	→	→	→	→
จัดส่งรายงาน												⇨