บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทดลองและผลการทดลองในงานวิจัย การทดลองในงานวิจัยนี้เป็นการทดลองขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่ปรับปรุงใหม่สำหรับประมวลผลในตัวประมวลผลหลายตัว โดยมีรูปแบบการทำงานเป็นการแบ่งงานประมวลผลไปพร้อมๆกัน โดยเปรียบเทียบประสิทธิภาพทางด้านความเร็วในการประมวลผลและความถูกต้องในการจำแนกบอตเน็ตกับสองขั้นตอนวิธี ได้แก่ ขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม [22] ซึ่งขั้นตอนวิธีทั้งสามได้แสดงไว้ในหัวข้อที่ 2.2, 3.3.2 และ 3.3.3 ตามลำดับ

การเปรียบเทียบประสิทธิภาพทางด้านความเร็วในการประมวลผลและความถูกต้องในการจำแนกบอตเน็ตของสามขั้นตอนวิธีดังกล่าว ใช้ชุดข้อมูลของ Botnet Traffic และ Normal Traffic ในการทดลองจำนวนทั้งหมด 100,000 แพ็กเก็ตรวมกัน โดยแบ่งข้อมูลที่นำมาทดลองเป็นจำนวนสี่การทดลองดังต่อไปนี้คือ 10,000 20,000 30,000 และ 40,000 แพ็กเก็ต ตามลำดับ การทดลองสุดท้ายคือ การจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ โดยการทดลองจะทำการเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง และทำการวิเคราะห์เวลาในการประมวลผล ข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic ที่ใช้ในการทดลองมีจำนวน 40,000 แพ็กเก็ตรวมกัน โดยการทดลองทั้งห้าการทดลองพัฒนาขึ้นด้วยภาษาจาวา (JAVA) ข้อมูลที่ใช้ทดลองคือข้อมูล Botnet Traffic ที่สร้างจาก Botnet Traffic Generation [20]

4.1 ระบบที่ใช้ในการทดลอง

1. หน่วยประมวลผลกลาง Intel Core i7 CPU 2.80 GHz 4 GB RAM
2. ซอฟต์แวร์ที่ใช้พัฒนา NetBeans IDE 7.1.1

4.2 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลอง คือข้อมูล Botnet Traffic ที่สร้างจาก Botnet Traffic Generation และข้อมูลที่เป็น Normal Traffic ซึ่งงานวิจัยนี้เก็บข้อมูลภายในแล็บปฏิบัติการ ภาควิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่ โดยข้อมูลทั้งหมดได้ผ่านการวิเคราะห์โดยใช้วิธีการพิจารณา

จากคุณลักษณะ (Feature Analysis) โดยพิจารณาจาก 256 ASCII Characters ใน Packet Payload เพื่อดูค่าความถี่เฉลี่ยในแต่ละไบต์ (Byte) ของตัวอักษรแอสกี (ASCII) ของทราฟฟิกทั้งสองแบบมาใช้เป็นข้อมูลสำหรับขั้นตอนของการจำแนกบอตเน็ตต่อไป

4.3 การเตรียมการทดลอง

การทดลองในงานวิจัยนี้ เป็นการทดลองขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่ปรับปรุงใหม่ สำหรับประมวลผลในตัวประมวลผลหลายตัว โดยมีรูปแบบการทำงานเป็นการแบ่งงานประมวลผลไปพร้อมๆกัน โดยเปรียบเทียบประสิทธิภาพทางด้านความเร็วในการประมวลผลและความถูกต้องในการจำแนกบอตเน็ตกับสองขั้นตอนวิธี ได้แก่ ขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม [22]

การทดลองขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่ปรับปรุงใหม่ ขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม กำหนดให้ใช้ชุดข้อมูลของ Botnet Traffic และ Normal Traffic ในการทดลองจำนวนทั้งหมด 100,000 แพ็กเก็ตรวมกัน โดยแบ่งข้อมูลที่นำมาทดลองเป็นจำนวนสี่การทดลองดังต่อไปนี้คือ 10,000 20,000 30,000 และ 40,000 แพ็กเก็ต ตามลำดับ เพื่อทดสอบประสิทธิภาพทางด้านความเร็วในการประมวลผลและความถูกต้องในการจำแนกบอตเน็ต และการทดลองเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง เพื่อวิเคราะห์เวลาในการประมวลผลของข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวน 40,000 แพ็กเก็ตรวมกัน

4.3.1 การทดลองขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่ปรับปรุงใหม่ ขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม

ในหัวข้อการทดลองนี้ จะทำการทดลองด้วยข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวนทั้งหมด 100,000 แพ็กเก็ตรวมกัน โดยแบ่งข้อมูลที่นำมาทดลองเป็นจำนวนสี่การทดลองดังต่อไปนี้คือ 10,000 20,000 30,000 และ 40,000 แพ็กเก็ต ตามลำดับ เพื่อวัดประสิทธิภาพทางด้านเวลาในการประมวลผลในหน่วยวินาที เมื่อมีการเพิ่มขึ้นของข้อมูล และวัดประสิทธิภาพทางด้านความถูกต้องในการจำแนกบอตเน็ต โดยใช้การคำนวณหาค่าร้อยละของข้อมูล Botnet (BPR) และค่าร้อยละของข้อมูล Normal (NPR)

4.3.2 การทดลองการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ โดยการทดลองเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง

ในหัวข้อการทดลองนี้ จะทำการทดลองด้วยข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวน 40,000 แพ็กเก็ตรวมกัน กับขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอเท่านั้น โดยการทดลองจะทำการเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง และทำการวัดประสิทธิภาพทางด้านเวลา เวลาที่ใช้ในการประมวลผลมีหน่วยเป็นวินาที

4.4 ผลการทดลอง

ในงานวิจัยนี้ทำการทดลองโดยใช้ข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic ที่ใช้ในการทดลองมีจำนวนทั้งหมด 100,000 แพ็กเก็ตรวมกัน โดยแบ่งข้อมูลที่นำมาทดลองเป็นจำนวน 10,000 20,000 30,000 และ 40,000 แพ็กเก็ตตามลำดับ ในการทดลองทำการเปรียบเทียบประสิทธิภาพทางด้านเวลาของการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบเดิม ขั้นตอนวิธีเคมีนแบบขนานในภาพที่ 1 และขั้นตอนวิธีเคมีนแบบขนานในงานวิจัยนี้นำเสนอ เวลาที่ใช้ในการประมวลผลมีหน่วยเป็นวินาที ใช้การคำนวณหาค่าร้อยละของข้อมูล Botnet (BPR) และค่าร้อยละของข้อมูล Normal (NPR) เพื่อวัดประสิทธิภาพทางด้านความถูกต้องในการจำแนกบอตเน็ต

4.4.1 ผลการทดลองขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่ปรับปรุงใหม่ ขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม

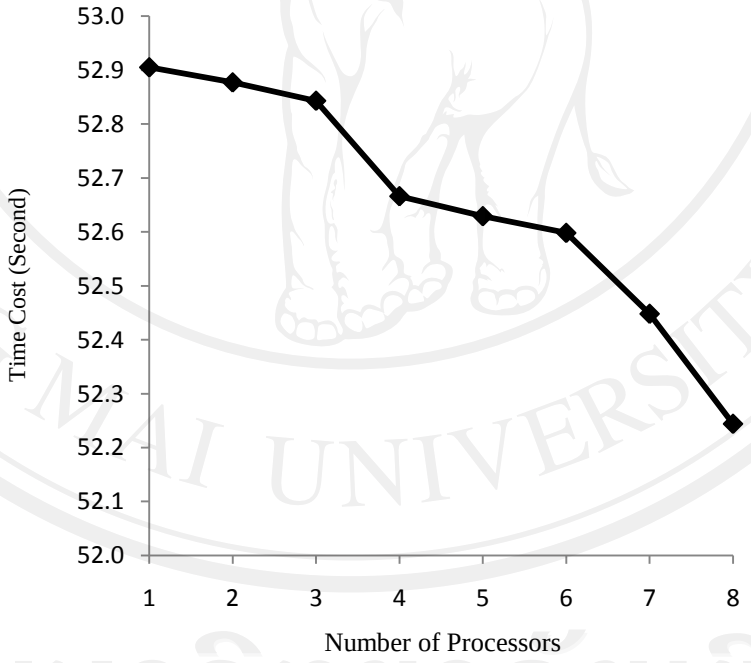
ตารางที่ 4.1 ผลการทดสอบการจำแนกบอตเน็ต

K-means Algorithm			
Data (Packets)	Time (Second)	BPR (%)	NPR (%)
10,000	5.222	98.00	100
20,000	15.842	97.66	100
30,000	31.729	97.00	100
40,000	53.115	96.00	100
ขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วมกัน			
10,000	5.216	98.33	100
20,000	15.716	97.00	100
30,000	31.621	97.00	100
40,000	52.873	97.00	100
วิธีการที่นำเสนอ Parallel K-means Algorithm			
10,000	5.183	98.33	100
20,000	15.639	98.33	100
30,000	31.507	98.00	100
40,000	52.616	97.66	100

ผลการทดลองพบว่า ขั้นตอนวิธีที่นำเสนอมีประสิทธิภาพที่ดีกว่าในด้านเวลาการประมวลผล และเมื่อมีการเพิ่มปริมาณข้อมูลให้มีขนาดใหญ่ขึ้นเวลาที่ใช้ประมวลผลยังคงน้อยกว่าเช่นกัน เนื่องมาจากข้อมูลมีการแบ่งให้มีขนาดเล็กลงขั้นตอนวิธีเคมีนจึงทำงานได้ดีกว่าข้อมูลขนาดใหญ่

ในการทดลองต่อมาคือ การจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ โดยการทดลองจะทำการเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง และทำการพิจารณาเวลาในการประมวลผลของข้อมูลจำนวน 40,000 แพ็กเก็ต

4.4.2 ผลการทดลองการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ โดยการทดลองจะทำการเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง



รูปที่ 4.1 กราฟแสดงเวลาในการประมวลผลเมื่อมีการเพิ่มขึ้นของจำนวนโปรเซสเซอร์ ของข้อมูลแพ็กเก็ต Botnet Traffic และ Normal Traffic จำนวน 40,000 แพ็กเก็ตรวมกัน

ผลการทดลองพบว่า เมื่อเพิ่มจำนวนโปรเซสเซอร์เพิ่มขึ้น เวลาในการประมวลผลจะลดลง เนื่องจากขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอมีการช่วยแบ่งงานกันทำ ทำให้เวลาโดยรวมในการประมวลผลลดลงด้วย ในภาพที่ 4.1 แสดงเวลาที่ใช้ในการประมวลผลทั้งหมดโดยประมาณของจำนวนโปรเซสเซอร์ต่างๆกันที่ทำการทดลอง และเวลาที่ใช้ในการประมวลผลมีหน่วยเป็นวินาที

4.4.3 ผลการทดลองเพิ่มเติม การจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ โดยการทดลองเปรียบเทียบเวลาในการประมวลผลข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวน 10,000 และ 30,000 แพ็กเก็ตรวมกัน เมื่อเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง

ตารางที่ 4.2 ผลการทดลองเปรียบเทียบเวลาในการประมวลผลข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวน 10,000 และ 30,000 แพ็กเก็ตรวมกัน เมื่อเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง

วิธีการที่นำเสนอ Parallel K-means Algorithm		
จำนวนโปรเซสเซอร์	ข้อมูล 10,000 แพ็กเก็ต ใช้เวลาประมวลผล (วินาที)	ข้อมูล 30,000 แพ็กเก็ต ใช้เวลาประมวลผล (วินาที)
1	5.224	31.782
2	5.201	31.776
3	5.193	31.696
4	5.183	31.507
5	5.179	31.461
6	5.160	31.501
7	5.156	31.391
8	5.139	31.327

4.4.4 ผลการทดลองเพิ่มเติม การจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ โดยการทดลองเปรียบเทียบเวลาในการประมวลผลข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวน 20,000 และ 40,000 แพ็กเก็ตรวมกัน เมื่อเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง

ตารางที่ 4.3 ผลการทดลองเปรียบเทียบเวลาในการประมวลผลข้อมูลแพ็กเก็ตของ Botnet Traffic และ Normal Traffic จำนวน 20,000 และ 40,000 แพ็กเก็ตรวมกัน เมื่อเพิ่มจำนวนโปรเซสเซอร์ขึ้นทีละหนึ่ง

วิธีการที่นำเสนอ Parallel K-means Algorithm		
จำนวนโปรเซสเซอร์	ข้อมูล 20,000 แพ็กเก็ต ใช้เวลาประมวลผล (วินาที)	ข้อมูล 40,000 แพ็กเก็ต ใช้เวลาประมวลผล (วินาที)
1	15.842	53.105
2	15.784	52.877
3	15.747	52.843
4	15.639	52.616
5	15.645	52.637
6	15.621	52.598
7	15.614	52.448
8	15.602	52.244

4.5 สรุปผลการทดลอง

ในงานวิจัยนี้ได้ทำการทดลองขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่ปรับปรุงใหม่ สำหรับประมวลผลในตัวประมวลผลหลายตัว โดยมีรูปแบบการทำงานเป็นการแบ่งงานประมวลผลไปพร้อมๆกัน โดยเปรียบเทียบประสิทธิภาพทางด้านความเร็วในการประมวลผลและความถูกต้องในการจำแนกบอตเน็ตกับสองขั้นตอนวิธี ได้แก่ ขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม [22]

จากผลการทดลองทุกการทดลอง พบว่าขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ ใช้เวลาในการประมวลผลน้อยกว่าขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม ในส่วนของการวัดประสิทธิภาพในการจำแนกบอตเน็ตนั้น ทำการวัดจากค่าร้อยละของข้อมูล Botnet (BPR) และค่าร้อยละของข้อมูล Normal (NPR) การทดลองพบว่าขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอมีความถูกต้องในการจำแนก Botnet Traffic และ Normal Traffic มีค่าที่สูงกว่าขั้นตอนวิธีอื่นๆที่ทำการเปรียบเทียบ

ดังนั้นผลการทดลองสามารถสรุปได้ว่า ขั้นตอนวิธีการจำแนกบอตเน็ตโดยใช้ขั้นตอนวิธีเคมีนแบบขนานที่นำเสนอ มีประสิทธิภาพทางด้านความเร็วในการประมวลผลโดยรวมสูงกว่าขั้นตอนวิธีเคมีนแบบเดิม และขั้นตอนวิธีเคมีนแบบขนานโดยใช้จุดศูนย์กลางร่วม [22] มีความถูกต้องในการจำแนก Botnet Traffic และ Normal Traffic มีค่าที่สูงกว่าเช่นกัน