**บทที่ 4**

**ผลการดำเนินงานวิจัย**

ชุดข้อมูลฝึกสอนที่ทำการพัฒนาขึ้นมีรูปแบบการจำลองมาจาก Packet Header และสร้างขึ้นผ่านโปรแกรม Packet Generator ที่ออกแบบขึ้นเอง โดยชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบออกแบบให้มีความแตกต่างกันตามสมมติฐานที่กำหนด สังเกตกระบวนการทำงานของโมเดล และรูปแบบความสัมพันธ์ของตัวแปรที่ได้หลังโมเดลทำการเรียนรู้และประมวลผล โดยตัวแปรและผลลัพธ์ที่จะนำมาพิจารณาในการวิเคราะห์เชิงเปรียบเทียบในขั้นตอนต่อไป มีดังนี้

ตัวแปรที่มีการตั้งให้มีค่าเท่ากันในทุกการทดลอง

* จำนวน Node ของแต่ละ Layer

Node จะถูกกำหนดค่าไว้ตั้งแต่แรกตามจำนวนขนาดของชุดข้อมูล

* จำนวนรอบการทดสอบ หรือ Epoch

Epoch กำหนดให้มีค่าเท่ากันทุกการทดลอง

* ขนาดของ Batch Size

ค่านี้ตั้งให้เป็น Default ตามที่โมเดลกำหนดไว้ตอนแรก

ตัวแปรที่มีการตั้งให้มีเงื่อนไขแตกต่างกัน ขึ้นอยู่กับประเด็นที่จะศึกษาในสมมติฐาน

* จำนวนและเงื่อนไขของแต่ละกฏไฟร์วอลล์ที่ใช้ภายใน Rule set
* จำนวนของ Packet ที่จะนำเข้าระบบ หรือ Sample(N)

ได้แก่ ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบที่สร้างจากกฎของไฟร์วอลล์

* จำนวนของข้อมูลฝึกสอนที่แต่ละกฎไฟร์วอลล์จะได้รับ
* การนำ Default Rule มาคำนวณด้วย ประกอบด้วย With Default และ Without Default

ผลลัพธ์ที่คาดว่าค่าจะต้องเปลี่ยนแปลงไปตามการทดสอบแต่ละครั้ง

* เวลาที่โมเดลใช้ในการเรียนรู้จากชุดข้อมูลฝึกสอน หรือ Training
* เวลาที่โมเดลใช้ในการตัดสินใจจากชุดข้อมูลทดสอบ หรือ Predict
* ค่าความแม่นยำโดยรวม หรือ Accuracy
* อัตราความผิดพลาดที่อ้างอิงจาก Reference Variant Set

ประเด็นสำคัญที่ต้องนำมาพิจารณา

* Default Rule

นอกจากกฎไฟร์วอลล์ที่กำหนดขึ้นทั่วไป ยังมีกฎของ Default Rule ซึ่งจำเป็นต้องพิจารณาแยกเป็นกรณีพิเศษ เนื่องจากจำนวนความเป็นไปได้ของข้อมูลของกฏไฟร์วอลล์ที่มีการกำหนด ทำให้การทดสอบแบ่งออกเป็น 2 แบบ ได้แก่ With Default Rule และ Without Default Rule ซึ่งเราได้ตั้ง Default Rule เป็นการ Deny ทุกข้อมูลที่นอกเหนือจากไฟร์วอลล์ที่กำหนดไว้

**4.1 สมมติฐานการทดลองที่ 1**

ในสมมติฐานการทดลองที่ 1 เป็นการทดลองใช้ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบที่สร้างขึ้น และเพื่อเป็นการพิสูจน์ว่าโมเดลสามารถประยุกต์ใช้ในงานวิจัยได้จริง มีหลักการทำงานและผลลัพธ์ที่คล้ายคลึงกับปัญญาประดิษฐ์ที่พบได้ทั่วไป โดยวางสมมติฐานเบื้องต้นไว้ ดังนี้

* โมเดลจะสามารถเรียนรู้จากชุดข้อมูลฝึกสอนที่สร้างจากกฎของไฟร์วอลล์และสามารถทำนายผลลัพธ์ได้
* เมื่อโมเดลเรียนรู้จากชุดข้อมูลฝึกสอนที่มีจำนวนมากขึ้นในแต่ละกฎไฟร์วอลล์ โมเดลจะสามารถทำนายผลลัพธ์ได้แม่นยำมากขึ้น
* โมเดลเมื่อมีการเรียนรู้ถึงจุดๆหนึ่งจะไม่สามารถเพิ่มความแม่นยำในการทำนายผลลัพธ์ได้อีก
* โมเดลจะใช้เวลาในการทดสอบประมวลผลข้อมูลเท่าเดิม แม้จะผ่านการเรียนรู้จากข้อมูลฝึกสอนที่มีจำนวนต่างกัน

**การออกแบบเงื่อนไขของชุดข้อมูลสำหรับการทดลองที่ 1**

การจำแนก Data Field ของข้อมูล Packet ที่สามารถเกิดขึ้นได้ทั้งหมด

|  |  |  |
| --- | --- | --- |
| Data Field | ตัวแปรที่ใช้ | จำนวนความเป็นไปได้ภายใน Field |
| Source Address | Subnet 192.168.0.0/16 | 65534 |
| Source Mask | Mask 16, 18 และ 24 | 3 |
| Destination Address | 161.246.34.11 | 1 |
| Destination Mask | 24 | 1 |
| Port | 22, 80 | 2 |
| Protocol | TCP, UDP | 2 |

ดังนั้น ข้อมูล Packet ที่เกิดขึ้นได้ทั้งหมด จะเท่ากับ 65534 \* 3 \* 1 \* 1 \* 2 \* 2 = 786,048

**การจำแนกชุดกฎไฟร์วอลล์ที่จะทำการทดสอบ**

|  |  |
| --- | --- |
| ชุดกฎของไฟร์วอลล์ | จำนวนข้อมูลที่ตรงเงื่อนไข |
| Rule set ที่ 1   * allow 192.168.0.0/16 to 161.246.34.11/24 port 80 tcp * deny 192.168.128.0/18 to 161.246.34.11/24 port 22 udp |  |
| Rule set ที่ 2   * allow 192.168.0.0/16 to 161.246.34.11/24 port 80 tcp * deny 192.168.128.0/18 to 161.246.34.11/24 port 22 udp * allow 192.168.64.0/24 to 161.246.34.11/24 port 22 tcp * deny 192.168.64.0/24 to 161.246.34.11/24 port 80 udp |  |
| Rule set ที่ 3   * allow 192.168.0.0/16 to 161.246.34.11/24 port 80 tcp * deny 192.168.128.0/18 to 161.246.34.11/24 port 22 udp * allow 192.168.64.0/24 to 161.246.34.11/24 port 22 tcp * deny 192.168.64.0/24 to 161.246.34.11/24 port 80 udp * allow 192.168.192.0/18 to 161.246.34.11/24 port 22 udp * allow 192.168.128.0/18 to 161.246.34.11/24 port 22 tcp |  |

การออกแบบชุดข้อมูลในการทดลองที่ 1

* ชุดข้อมูลฝึกสอนในแต่ละกฎไฟร์วอลล์ มีจำนวน N Sample เท่ากันทั้งหมด และเพิ่มขึ้นด้วยจำนวนที่เท่าๆกันในแต่ละครั้ง

**4.2 สมมติฐานการทดลองที่ 2**

ในสมมติฐานการทดลองที่ 2 จะมุ่งเน้นไปที่การแบ่งจำนวนชุดข้อมูลฝึกสอนจากกฎไฟร์วอลล์ที่แตกต่างกัน เพราะในการตั้งกฎไฟร์วอลล์นั้น ชุดข้อมูลที่สามารถเป็นไปได้ทั้งหมดใน Sample Space ไม่จำเป็นต้องมีจำนวนที่เท่ากันเสมอไป จึงเป็นที่มาของอีกสมมติฐานหนึ่งว่า ถ้าหากตั้งเงื่อนไขให้แต่ละกฎไฟร์วอลล์มีจำนวนความเป็นไปได้ของข้อมูลไม่เท่ากัน และแบ่งจำนวนชุดข้อมูลฝึกสอนให้แต่ละกฎให้มีจำนวนไม่เท่ากัน จะส่งผลต่อความแม่นยำของโมเดลอย่างไร โดยการทดลองจะแบ่งออกเป็น 2 รูปแบบ ได้แก่

1. การแบ่งชุดข้อมูลฝึกสอนให้ในแต่ละกฏไฟร์วอลล์ด้วยจำนวน N ชุดที่เท่ากัน
2. การแบ่งชุดข้อมูลฝึกสอนให้ในแต่ละกฏไฟร์วอลล์ด้วยอัตราส่วน Ratio ที่เท่ากัน

ซึ่งการทดลองทั้ง 2 รูปแบบ เพื่อให้สามารถเปรียบเทียบเป็นกราฟได้ง่ายขึ้น จึงจำเป็นต้องใช้จำนวนชุดข้อมูลฝึกสอนทั้งหมด (Total packet) ให้มีจำนวนเท่ากัน และกำหนดเกณฑ์ขั้นต่ำของชุดข้อมูลฝึกสอน ถ้าหากมีจำนวนที่ต่ำเกินไปจะไม่สามารถแบ่งอัตราส่วนชุดข้อมูลฝึกสอนได้

โดยเราได้ตั้งข้อสมมติฐาน และคาดว่าการทดลองจะให้ผลลัพธ์ ดังนี้

* ชุดข้อมูลฝึกสอนที่มีการแบ่งอัตราส่วนแบบ Ratio ตามกฎของไฟร์วอลล์จะให้ค่าความแม่นยำได้สูงกว่าชุดข้อมูลฝึกสอนที่มีการแบ่งด้วยจำนวนที่เท่ากัน

จากสมมติฐานข้างต้นเราได้ออกแบบกฎของไฟร์วอลล์และชุดข้อมูลฝึกสอน ดังนี้