**บทที่ 4**

**ผลการดำเนินงานวิจัย**

การทดลองจะเป็นไปตามวัฎจักรการดำเนินงานวิจัยข้างต้น โดยชุดข้อมูลฝึกสอนที่ทำการพัฒนาขึ้นมีรูปแบบการจำลองมาจาก Packet Header และสร้างขึ้นผ่านโปรแกรม Packet Generator ที่ออกแบบขึ้นเอง ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบจะมีการออกแบบให้มีความแตกต่างกันตามสมมติฐานที่กำหนด สังเกตกระบวนการทำงานของโมเดล และรูปแบบความสัมพันธ์ของตัวแปรที่ได้หลังโมเดลทำการเรียนรู้และประมวลผล และทำการสรุปผลลัพธ์ที่ได้หลังเสร็จสิ้นการทดลอง

**การออกแบบเงื่อนไขของชุดข้อมูล**

ค่าความเป็นไปได้ทั้งหมดที่สามารถเกิดขึ้นได้ในแต่ละส่วนของข้อมูล มีดังนี้

|  |  |  |
| --- | --- | --- |
| Data Field | ตัวแปรที่ใช้ | จำนวนความเป็นไปได้ภายใน Field |
| Source Address | Subnet 192.168.0.0/16 | 65534 |
| Source 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 |

ตารางการจำแนกความเป็นไปได้ของแต่ละ Data Field

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

**การจำแนกชุดกฎไฟร์วอลล์ที่จะทำการทดสอบ**

|  |  |
| --- | --- |
| ชุดกฎของไฟร์วอลล์ | จำนวนข้อมูลที่สามารถเกิดขึ้นตรงตามเงื่อนไขของไฟร์วอลล์ที่กำหนด |
| 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 |  |

ตารางการจำแนกความเป็นไปได้ของแต่ละกฎไฟร์วอลล์

**การนำ Default Rule มาเป็นส่วนหนึ่งของกฎไฟร์วอลล์เพื่อสร้างชุดข้อมูลฝึกสอน**

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

**การออกแบบชุดข้อมูลฝึกสอนในสมมติฐาน**

**รูป**

ประกอบไปด้วย 2 ชุดข้อมูลฝึกสอนใน 1 เซต โดยประกอบไปด้วยชุดข้อมูลฝึกสอนที่มีประเด็นการนำ Default Rule มาใช้ และ ไม่มีการนำ Default Rule มาใช้ในการสร้าง โดยมีตัวแปรสำคัญในการสร้างชุดข้อมูลฝึกสอนเพื่ออิงตามประเด็นศึกษาในสมมติฐาน ดังนี้

* จำนวนและเงื่อนไขของแต่ละกฏไฟร์วอลล์ที่ใช้ภายใน Rule set
* จำนวนของ Packet ที่จะนำเข้าระบบ
* จำนวนของข้อมูลฝึกสอนที่แต่ละกฎไฟร์วอลล์จะได้รับ
* การนำประเด็น Default Rule มาใช้ด้วย ประกอบด้วย With Default และ Without Default

**การออกแบบชุดข้อมูลทดสอบ**

**รูป**

ประกอบไปด้วย 2 ชุดข้อมูลเช่นเดียวกับชุดฝึกสอน โดยประกอบไปด้วยชุดข้อมูลฝึกสอนที่มีประเด็นการนำ Default Rule มาใช้ และ ไม่มีการนำ Default Rule มาใช้ในการสร้าง

**ผลลัพธ์ที่คาดว่าค่าจะต้องเปลี่ยนแปลงไปตามการทดสอบแต่ละครั้ง**

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

**4.1 สมมติฐานการทดลองที่ 1**

ในสมมติฐานการทดลองที่ 1 เป็นการทดลองใช้ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบที่สร้างขึ้น และเพื่อเป็นการพิสูจน์ว่าโมเดลสามารถประยุกต์ใช้ในงานวิจัยได้จริง มีหลักการทำงานและผลลัพธ์ที่คล้ายคลึงกับปัญญาประดิษฐ์ที่พบได้ทั่วไป โดยวางสมมติฐานเบื้องต้นไว้ ดังนี้

* โมเดลจะสามารถเรียนรู้จากชุดข้อมูลฝึกสอนที่สร้างจากกฎของไฟร์วอลล์และสามารถทำนายผลลัพธ์ได้
* เมื่อโมเดลเรียนรู้จากชุดข้อมูลฝึกสอนที่มีจำนวนมากขึ้นในแต่ละกฎไฟร์วอลล์ โมเดลจะสามารถทำนายผลลัพธ์ได้แม่นยำมากขึ้น
* โมเดลเมื่อมีการเรียนรู้ถึงจุดๆหนึ่งจะไม่สามารถเพิ่มความแม่นยำในการทำนายผลลัพธ์ได้อีก
* โมเดลจะใช้เวลาในการทดสอบประมวลผลข้อมูลเท่าเดิม แม้จะผ่านการเรียนรู้จากข้อมูลฝึกสอนที่มีจำนวนต่างกัน

**หลักการออกแบบชุดข้อมูลฝึกสอน**

ชุดข้อมูลฝึกสอนในแต่ละกฎไฟร์วอลล์ จะมีจำนวน N Sample เท่ากันทั้งหมด โดยจำนวนที่ของข้อมูลฝึกสอนที่ใช้ทดสอบ ประกอบไปด้วย 10, 100, 300, 600, 1,000, 3,000, 6,000, 10,000 ในแต่ละกฎไฟร์วอลล์

ผลลัพธ์ที่ได้จากการทดลอง

Table

Description automatically generated  
ตารางผลการทดลอง Rule set ที่ 1 (2 กฎไฟร์วอลล์)

Table

Description automatically generated  
ตารางผลการทดลอง Rule set ที่ 2 (4 กฎไฟร์วอลล์)

Table

Description automatically generated  
ตารางผลการทดลอง Rule set ที่ 3 (6 กฎไฟร์วอลล์)

**4.2 สมมติฐานการทดลองที่ 2**

จากสมมติฐานแรกจะเห็นได้ว่า ความเป็นไปได้ของชุดข้อมูลฝึกสอนในแต่ละกฎไฟร์วอลล์มีจำนวนไม่เท่ากัน ดังนั้นในสมมติฐานนี้จึงเป็นการตั้งข้อสันนิษฐานว่า ถ้าหากตั้งเงื่อนไขให้กฎไฟร์วอลล์แต่ละกฎได้รับจำนวนชุดข้อมูลฝึกสอนไม่เท่ากัน จะส่งผลต่อความแม่นยำของโมเดลอย่างไร มีการพัฒนาโมเดลในทางที่ดีขึ้นหรือแย่ลงอย่างไร

**หลักการออกแบบชุดข้อมูลฝึกสอน**

สมมติฐานนี้จะเป็นการใช้อัตราส่วนมาเป็นหลักเกณฑ์ในการแบ่งจำนวนชุดข้อมูลฝึกสอนที่แต่ละกฎจะได้รับ โดยจำนวนของข้อมูลฝึกสอนที่ใช้ทดสอบ เพิ่มขึ้นด้วยอัตราส่วน Ratio ที่เท่าๆกัน

**การออกแบบเงื่อนไขของชุดข้อมูลสำหรับการทดลองที่ 2**

ในชุดข้อมูลฝึกสอนชุดหนึ่งจะประกอบไปด้วยตัวอย่างข้อมูล Packet ที่ตรงตามเงื่อนไขในแต่ละกฎไฟร์วอลล์ มีวิธีการแบ่งจำนวนตามสมมติฐานที่วางเอาไว้ และจะเพิ่มจำนวนขึ้นเรื่อยๆ นอกจากนี้จะต้องสร้างอีกชุดฝึกสอนหนึ่งที่มีการเพิ่ม Default Rule มาคำนวณด้วย มีการออกแบบอิงตามหัวข้อต่อไปนี้

เพื่อให้การทดสอบได้มาตรฐานว่าการป้อนชุดข้อมูลทดสอบไปสามารถวัดผลที่ให้ไว้ในแต่ละกฎได้จริงหรือไม่ จึงจำเป็นต้องมีหลักการในการออกแบบชุดข้อมูลสำหรับการทดสอบ โดยมีสิ่งที่ต้องคำนึงถึง ดังนี้

* ชุดข้อมูลทดสอบควรจะมีการแบ่งให้ allow และ deny จำนวนพอๆกัน ข้อนี้เป็นหนึ่งหลักการสำคัญในการทดสอบว่าโมเดลได้มาตรฐานหรือไม่ ถ้าหากสร้างโมเดลไม่ได้มาตรฐานเนื่องจากปัญหา overfitting หรือ underfitting เพราะชุดข้อมูลฝึกสอนไม่มีความครอบคลุมมากพอ โมเดลจะไม่สามารถทำนายผลลัพธ์ได้เลย