**บทที่ 3**

**วิธีการดำเนินการวิจัย**

**3.1. การศึกษาแนวคิดการดำเนินงานวิจัยเก่าเพื่อการต่อยอดงานวิจัยใหม่**

3.1.1.การนำกฎของไฟร์วอลล์มาใช้อ้างอิงในการสร้างชุดข้อมูลฝึกสอน

ไฟร์วอลล์โดยทั่วไปจะมีกฎไฟร์วอลล์เพื่อใช้สำหรับคัดกรองข้อมูลแพ็คเกต ซึ่งจะเป็นเรื่องที่ง่ายกว่าถ้าหากเรานำกฎของไฟร์วอลล์นี้มาเป็นตัวอ้างอิงในการสร้างชุดข้อมูลฝึกสอนซึ่งมีมาตรฐานและมีความน่าเชื่อถือมากกว่าการนำข้อมูลการโจมตีที่เคยเกิดขึ้นมาก่อน โครงงานเดิมนั้นได้มีการนำกฎไฟร์วอลล์แต่ละกฎมาสร้างชุดข้อมูลฝึกสอนโดยการอ่านและประมวลผลทีละกฎไฟร์วอลล์เพื่อสร้างชุดข้อมูลฝึกอบรมที่มาจากกฎนั้นๆ โดยจำนวนชุดข้อมูลที่สร้างขึ้นจะแบ่งออกเป็น จำนวนชุดข้อมูลฝึกสอนต่อ 1 กฎไฟร์วอลล์ และจำนวนชุดข้อมูลฝึกสอนทั้งหมดของกฎไฟร์วอลล์ทั้งหมดรวมกัน

3.1.2. ขั้นตอนการดำเนินงานวิจัยเก่าโดยรวม

ชุดข้อมูลฝึกสอนโมเดลที่สร้างขึ้นที่มีการอ้างอิงจากกฎของไฟร์วอลล์จะถูกสร้างขึ้นมาหลายชุดด้วยกันเพื่อนำไปฝึกฝนโมเดล โดยข้อมูลฝึกสอนแต่ละชุดนั้นสำหรับโมเดลจะมีจำนวนแพ็คเกตฝึกสอนต่อ 1 กฎไฟร์วอลล์ (Sample per Rule) ที่แตกต่างกัน หลังจากได้โมเดลที่ผ่านการเรียนรู้แล้วให้นำโมเดลมาเปรียบเทียบประสิทธิภาพในด้านต่างๆด้วยชุดข้อมูลทดสอบชุดเดียวกันเพื่อหาว่าโมเดลที่เรียนรู้จากชุดข้อมูลฝึกสอนที่จำนวนใดสามารถให้ค่าที่ดีที่สุด

3.1.3. การออกแบบขอบเขตของแพ็คเกตและโครงสร้างที่ใช้

โครงสร้างของข้อมูลฝึกสอนที่ใช้ทุกชุดจะประกอบไปด้วย 2 ส่วน ได้แก่ ส่วนของการตัดสินใจ (Decision Field หรือ Action Field) เป็นส่วนที่ทำให้โมเดลทราบได้ว่าแพ็คเกตถูกยอมปล่อยให้ผ่านหรือไม่สามารถปล่อยให้ผ่านไปได้ โดยตัดสินจากส่วนของข้อมูลดิบ (Data Field) ที่มีการจำลองอยู่ในรูปแบบโครงสร้างของ Packet Header แปลงให้อยู่ในรูปแบบของเลขฐานสอง โมเดลจะมีการคำนวณและแปลงข้อมูลดิบให้อยู่ในรูปแบบของค่าน้ำหนัก เพื่อหาความสัมพันธ์ระหว่างกันเพื่อหาคำตอบที่ตรงกับ Action Field มากที่สุด

3.1.4. อัลกอริทึ่มสำหรับการสร้างชุดข้อมูลฝึกสอน

3.1.5. หลักการออกแบบโครงสร้างของโมเดล

3.1.6. การออกแบบชุดข้อมูลทดสอบ

3.1.7. การวัดค่าประสิทธิภาพของแต่ละโมเดล

**3.2. การวิเคราะห์ปัญหาที่เกิดขึ้นและจุดอ่อนของงานวิจัยเก่า**

3.2.1. ปัญหาความไม่เท่าเทียมกันในการแบ่งชุดข้อมูลฝึกสอน

**3.3. วัฏจักรการพัฒนางานวิจัยใหม่หลังได้รับการแก้ไข**

ในการพัฒนาวิจัยนี้จะเป็นทั้งการต่อยอดและทำใหม่เพื่อแก้ไขในแง่ของแนวคิดและปัญหาที่เราได้พบให้ดีขึ้น โดยจะเป็นการพิจารณาหาโมเดลโครงข่ายประสาทเทียมเชิงลึกหรือ Deep Neural Network ใหม่ที่ดีกว่าเดิม รูปแบบการดำเนินงานส่วนใหญ่จะมีความคล้ายคลึงไปกับงานวิจัยเดิม แต่ให้มีการปรับเปลี่ยนอัลกอริทึ่มที่ใช้ในการสร้างชุดข้อมูลฝึกสอนเพื่อให้สามารถตอบโจทย์สมมติฐานที่สร้างขึ้นได้ ซึ่งในการฝึกโมเดล DNN 1 รอบจนไปถึงการทดสอบหาผลลัพธ์ของโมเดลจะมีการดำเนินงานที่คล้ายคลึงกันเป็นวัฎจักร ดังนี้

Diagram

Description automatically generated

จากรูปภาพ Block Diagram ข้างต้น สามารถแบ่งกระบวนการทำงานออกเป็นขั้นตอนได้ 6 ขั้นตอน ดังนี้

* + การกำหนดขอบเขตของข้อมูล Data Field ที่จะพิจารณา และการกำหนดกฎของไฟร์วอลล์
  + การสร้างชุดข้อมูลสำหรับการฝึกสอนโมเดล
  + การนำโมเดลไปผ่านการเรียนรู้ด้วยชุดข้อมูลสำหรับฝึกสอน
  + การสร้างชุดข้อมูลสำหรับการทดสอบโมเดล
  + การนำโมเดลไปประมวลผล ทำนายผลลัพธ์จากชุดข้อมูลสำหรับทดสอบ
  + บันทึกผลลัพธ์จากการทดสอบโมเดล

**3.3.1. ขั้นตอนที่ 1 การกำหนดขอบเขตของ Data Field ที่จะพิจารณา และการกำหนดกฎไฟร์วอลล์**

**Diagram

Description automatically generated****รูปที่ 3.2** Block Diagram การกำหนดขอบเขตของข้อมูลทั้งหมดที่จะศึกษา

เป็นขั้นตอนที่สำคัญสุดของงานวิจัย เป็นการชี้ประเด็นที่จะศึกษาและแนวทางของผลลัพธ์ที่จะเป็น โดยเริ่มจากการทำการทดลองอิงจากงานวิจัยเก่า ทดลองตั้งสมมติฐาน นำไปต่อยอดและสรุปเป็นประเด็นใหม่ที่สามารถพิสูจน์ได้

เงื่อนไขหลักของการวิจัยคือการสร้างชุดข้อมูลฝึกสอนจากกฎของไฟร์วอลล์ เพื่อให้ได้ระบบการทำงานคัดกรองข้อมูล Packet ที่ได้มาตรฐานและเรียนรู้ได้เองอย่างมีประสิทธิภาพ มีความแม่นยำสูง สิ่งที่ต้องทำในส่วนแรกคือการกำหนดขอบเขตความเป็นไปได้ที่ข้อมูลจะสามารถเกิดขึ้นในเครือข่าย และการกำหนดกฎของไฟร์วอลล์เพื่อให้สามารถสร้างชุดข้อมูล Packet ที่จะนำไปฝึกสอนให้กับโมเดล สร้างชุดข้อมูลทดสอบโมเดลที่สามารถเปรียบเทียบความถูกต้องของผลลัพธ์ที่ได้จากโมเดลหลังผ่านการเรียนรู้แล้ว

**3.3.1.1. การกำหนด Default Pool และ Data Field ที่จะใช้พิจารณา**

การกำหนดขอบเขตของ Packet ที่สามารถเกิดขึ้นหรือการกำหนด Default เองเป็นอีกหนึ่งขั้นตอนที่สำคัญ เพื่อลดปัญหาในการใช้ Workload และลดเวลาที่ใช้ในการทดลองของคอมพิวเตอร์ที่มากเกินจำเป็นในการคำนวณหา Sample Space เพราะ Packet ที่เกิดขึ้นจริงมีจำนวนมหาศาล แม้มีข้อมูลภายใน Field เพียงชุดเดียวที่แตกต่างกัน ชุดข้อมูลนั้นจะถูกสรุปเหมือนเป็นชุดข้อมูลใหม่ แต่ถึงกระนั้นการลดจำนวน Default จะต้องไม่น้อยเกินไปและยังสามารถสร้างกฎไฟร์วอลล์ที่ใช้ในการทดลองได้