Case 4

1. Problem Statement

ระบบรักษาความปลอดภัยของบริษัท Cyber Defend ตอนนี้ใช้การ Signature-based ซึ่งมีจุดอ่อนร้ายแรงคือ ถ้าเจอ Zero-day attack ระบบจะไม่สามารถ detected ได้ทันที นอกจากนี้บริษัทมีข้อมูลวิ่งเข้ามาระดับพันล้านรายการต่อวัน ทำให้ระบบเดิมรับมือไม่ไหวและมักจะ False Positives จนทีมงานทำงานไม่ได้ บริษัทจึงต้องหาระบบใหม่ที่เปลี่ยนจากการจำหน้าตาไวรัส มาเป็นการสังเกต anormaly

2. Research Question

2.1 การเปรียบเทียบประสิทธิภาพของสถาปัตยกรรมการเรียนรู้ในรูปแบบต่างๆ ในการตรวจจับ Zero- day Attack เพื่อค้นหาแนวทางที่เหนือกว่าการทำ Hybrid Machine Learning แบบปกติ?

3. Objectives

3.1 วิเคราะห์รูปแบบพฤติกรรม ของ Zero-day attack ที่ระบบ Signature-based ไม่สามารถตรวจจับ ได้

3.2 ศึกษาขีดความสามารถ ของ Machine Learning รูปแบบต่างๆ ในการระบุภัยคุกคามที่ไม่เคย ปรากฏมาก่อน

3.3 เปรียบเทียบประสิทธิภาพ ระหว่าง Hybrid Machine Learning กับสถาปัตยกรรมทางเลือกอื่นๆ ที่ ค้นพบ

3.4 คัดเลือกและนำเสนอ โมเดลต้นแบบที่แม่นยำสูงและลดการแจ้งเตือนผิดพลาด (False Positive) ได้ดีที่สุด

4. Method

1: เตรียมข้อมูล (Data Preparation) นำชุดข้อมูลมาตรฐานระดับสากลด้านความปลอดภัย (เช่น CICIDS2017/2018 ซึ่งเป็น Gold Standard ที่มีเฉลยว่าใครคือแฮกเกอร์) มาใช้เป็นข้อมูลหลักในการฝึกสอนระบบ ร่วมกับข้อมูลจำลองของบริษัท

2: สร้างแผนผังเครือข่าย (Graph Construction) แปลงข้อมูล Log ให้เป็นภาพกราฟ โดยให้ IP/อุปกรณ์คอมพิวเตอร์ เป็น Node และ Edge โดยออกแบบระบบให้วาดจุดและเส้นอย่างเป็นระเบียบ ชัดเจน และต้องไม่มีเส้นทับซ้อนกัน เพื่อให้ระบบคอมพิวเตอร์สามารถคำนวณและมองเห็นโครงสร้างที่ผิดปกติได้อย่างรวดเร็วและแม่นยำที่สุด

3: ดักจับ Anomaly Detection ให้ระบบคอยคำนวณหารูปร่างแผนผังที่ผิดธรรมชาติ เช่น โหนดหนึ่งมีการแตกเส้นเชื่อมออกไปเป็นแฉกๆ (Star-topology) อย่างฉับพลัน ซึ่งเป็นพฤติกรรมปกติของไวรัสที่กำลังสแกนเครือข่าย ระบบจะดักจับสิ่งนี้ได้ทันทีโดยไม่ต้องรู้จักชื่อไวรัส

4: สอบเทียบมาตรฐาน (Evaluation & Benchmarking) นำผลลัพธ์ที่ระบบจับได้ ไปเทียบกับชุดข้อมูล Gold Standard เพื่อพิสูจน์ให้เห็นเป็นตัวเลขชัดเจนว่า ระบบใหม่นี้จับไวรัสได้แม่นยำแค่ไหน และลดการแจ้งเตือนมั่วได้ดีกว่าระบบเดิมของบริษัทจริงหรือไม่