การตรวจจับ SYN Flooding Attack ด้วยแผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ค่าถ่วงน้ำหนักเลขชี้กำลังบนกลุ่มตัวอย่างเวลา
SYN Flooding Attack Detection Using EWMA on Time Sampling

ณิชากร เทียบทอง[1] นิธิ ทะนนท์[1] และ เพ็ญณี หวังเมธีกุล[1*]
Nichagon Teabtong[1] Nithi Thanon[1] and Pennee Wangmaeteekul[1]
[1]สาขาวิทยาศาสตร์การคำนวณ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
[1]Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai Campus
วันที่ส่งบทความ : 22 กรกฎาคม 2565 วันที่แก้ไขบทความ : 28 ตุลาคม 2565 วันที่ตอบรับบทความ : 3 กุมภาพันธ์ 2566
Received: 22 July 2022, Revised: 28 October 2022, Accepted: 3 February 2023

บทคัดย่อ

บทความนี้นำเสนอการตรวจจับ SYN Flooding attack (Denial-of-Service) ด้วยวิธีค่าเฉลี่ยเคลื่อนที่ค่าถ่วงน้ำหนักเลขชี้กำลัง (Exponentially Weighted Moving Average: EWMA) และตรวจสอบสามปัจจัยที่ส่งผลต่อประสิทธิภาพการตรวจจับ คือ 1) การแบ่งแพ็คเก็ต (Sample packet) ในช่วงเวลาที่ต่างกัน 2) อัตราการโจมตี (อัตราการโจมตีต่ำและอัตราการโจมตีสูง) 3) การไม่มีผู้ใช้และมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ (Web server) ขณะโจมตี โดยประเมินประสิทธิภาพผ่านค่าความถูกต้อง (Accuracy) อัตราผลบวกลวง (False positive rate) และอัตราผลลบลวง (False negative rate) ผ่านชุดข้อมูลที่จำลองขึ้น ผลการทดลองพบว่าขั้นตอนวิธีที่นำเสนอสามารถตรวจจับการโจมตีทั้งอัตราการโจมตีต่ำและอัตราการโจมตีสูง นอกจากนี้ผลการประเมินประสิทธิภาพบว่าทั้งสามปัจจัยส่งผลต่อประสิทธิภาพการตรวจจับ อัลกอริทึมทำงานได้ดีภายใต้สภาพแวดล้อมที่จำลอง ในกรณีไม่มีผู้ใช้ (User) ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำและอัตราการโจมตีสูงที่การแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่

คำสำคัญ : การโจมตี SYN Flooding การปฏิเสธการให้บริการ แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ค่าถ่วงน้ำหนักเลขชี้กำลัง การสุ่มตัวอย่างเวลา

Abstract

This article presents a detection of SYN Flooding Attack (Denial-of-Service) using the Exponentially Weighted Moving Average (EWMA) method and examines three factors that affect detecting efficiency: 1) Sample packets at different time intervals 2) Attacking Rate

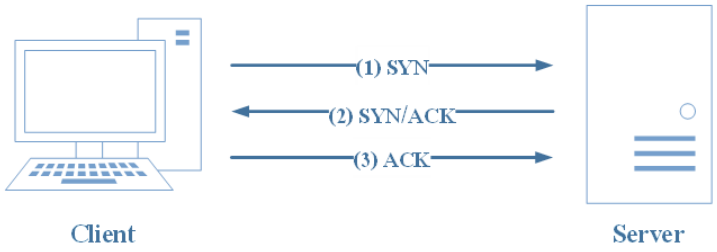
[*]ที่อยู่ติดต่อ E-mail address: pennee.wa@psu.ac.th

(low attack rate and high attack rate) 3) No user and users request information to the Web Server while being attacked. The efficiency has been evaluated via Accuracy formular, False Positive Rate, and False Negative Rate through the simulated datasets. The results show that the proposed algorithm is able to detect both low and high attack rates. Moreover, the experiment results confirm that all three factors cause for the detection performance. The algorithm performs well under the circumstance which there is no user requested information to the Web Server while the attacking rates are low or high with medium or large divided packets size.

Keywords: SYN Flooding attack, Denial-of-Service, Exponentially Weighted Moving Average, Time sampling

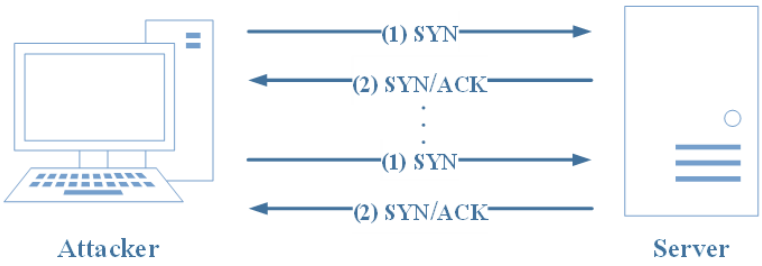
1. บทนำ

ปัจจุบันการโจมตีบนไซเบอร์มีหลากหลายรูปแบบหนึ่งในนั้น คือ การโจมตี SYN Flooding ซึ่งจัดอยู่ในกลุ่มของการปฏิเสธการบริการ DoS (Denial-of-Service) เป็นรูปแบบการโจมตีที่ส่งผลให้เครื่องเซิร์ฟเวอร์ไม่สามารถให้บริการแก่ผู้ใช้ได้ โดยในปี ค.ศ. 2021 การโจมตีรูปแบบ SYN Flooding คิดเป็นร้อยละ 34% ของวิธีการโจมตีแบบ DoS [1] การโจมตีผ่านขั้นตอนการสร้างช่องทางการสื่อสารของโพรโทคอล TCP (Transmission Control Protocol) ที่เรียกว่า Three-Way Handshake [2] แสดงได้ดังรูปที่ 1



รูปที่ 1. Three-Way Handshaking

โดยมี 3 ขั้นตอนดังนี้ 1) เครื่องผู้ใช้ (Client) ส่งแพ็คเก็ต SYN (Synchronize) สำหรับขอเชื่อมต่อไปยังเครื่องเซิร์ฟเวอร์ (Server) 2) เครื่องเซิร์ฟเวอร์ส่งแพ็คเก็ต SYN/ACK (Acknowledge) กลับมายังเครื่องผู้ใช้ 3) เครื่องผู้ใช้ส่งแพ็คเก็ต ACK กลับไปยังเครื่องเซิร์ฟเวอร์ หลังจากนั้นจึงเริ่มรับส่งข้อมูล ซึ่งการโจมตีเครือข่ายแบบ SYN Flooding อาศัยขั้นตอนการทำงานของ Three-Way Handshake แสดงดังรูปที่ 2



รูปที่ 2. SYN Flooding attack

โดย 1) ผู้โจมตี (Attacker) ส่งแพ็คเก็ต SYN ไปยังเซิร์ฟเวอร์ และ 2) เซิร์ฟเวอร์ส่งแพ็คเก็ต SYN/ACK กลับไปยังหมายเลขไอพีของผู้โจมตีซึ่งส่วนใหญ่ไม่มีอยู่จริงทำให้เซิร์ฟเวอร์ต้องรอแพ็คเก็ต ACK [2] ในระยะเวลาหนึ่งส่งผลให้เกิดการสูญเสียทรัพยากร หากผู้โจมตีส่งแพ็คเก็ต SYN จำนวนมากทำให้ half-open connection buffer ของ TCP ฝั่งเครื่องเซิร์ฟเวอร์เต็มไม่สามารถให้บริการการร้องขอใหม่จากเครื่องผู้ใช้ใด ๆ

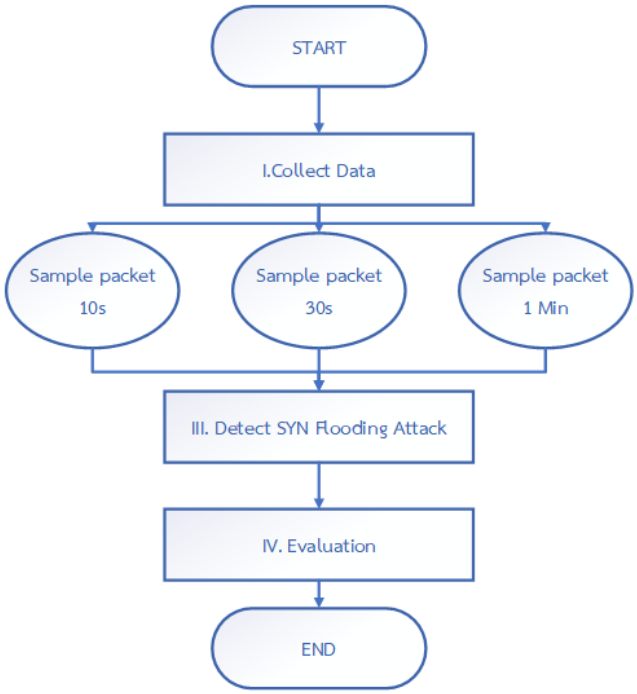
การตรวจจับการโจมตี SYN Flooding มี 2 วิธี คือ การตรวจจับแบบ Misuse detection และการตรวจจับแบบ Anomaly detection [3] โดยวิธี Misuse detection ผู้ตรวจจับจะกำหนดรูปแบบการโจมตีที่รู้จักไว้ในระบบการตรวจสอบการโจมตีเพื่อใช้ตรวจจับพฤติกรรมการใช้งานในเครือข่าย ข้อเสียของวิธีนี้คือหากรูปแบบการโจมตีไม่ตรงกับที่ออกแบบไว้ก็จะไม่สามารถตรวจจับได้ สำหรับวิธี Anomaly detection เป็นการตรวจสอบพฤติกรรมการใช้งานเครือข่ายที่ผิดปกติไปจากผู้ใช้งานทั่วไป โดยจะแยกพฤติกรรมการใช้งานปกติที่ยอมรับได้ออกมาและกำหนดให้พฤติกรรมอื่น ๆ เป็นการใช้งานที่ผิดปกติ [3]-[4]

งานวิจัยนี้นำเสนอการตรวจจับการโจมตีแบบ Anomaly detection โดยนำแผนภูมิการตรวจสอบกระบวนทางสถิติมาปรับใช้ตรวจหาความผิดปกติ เนื่องจากเป็นเครื่องมือสำคัญสำหรับการตรวจสอบระบบตามลำดับที่สามารถทำงานได้อย่างถูกต้อง [5] โดยประยุกต์ใช้แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบเลขชี้กำลัง EWMA (Exponentially Weighted Moving Average chart: EWMA chart) ซึ่งเป็นแผนภูมิติดตามค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบเลขชี้กำลังของค่าเฉลี่ยตัวอย่างก่อนหน้าทั้งหมดและเป็นวิธีการมองการเปลี่ยนแปลงอยู่บนเวลาโดยมีวัตถุประสงค์เพื่อตรวจจับการเปลี่ยนแปลงในค่าเฉลี่ยกระบวนการอย่างรวดเร็ว [6] ในขั้นตอนวิธีการตรวจจับ จากการศึกษางานวิจัยก่อนหน้านี้ Ramkumar และคณะ [2] ได้นำเสนอการตรวจจับและป้องกันการโจมตี SYN Flooding โดยใช้ Adaptive Thresholding Algorithm (ATA) ถูกใช้เพื่อคำนวณ Dynamic threshold เพื่อแก้ปัญหาข้อจำกัดของ Static threshold ที่มีอัตราผลบวกลวงที่สูง ซึ่งทดสอบกับข้อมูลจริง ผลการทดสอบมีประสิทธิภาพในการตรวจจับการโจมตีในอัตราการโจมตีสูง แต่ไม่ได้พิจารณาถึงอัตราการโจมตีต่ำและ IP ปลอม งานวิจัยของ Bouyeddou และคณะ [5] นำเสนอแผนภูมิการตรวจสอบกระบวนการทางสถิติประกอบด้วย 3 รูปแบบ คือ Shewhart chart, Cumulative Sum (CUSUM) Control chart และ EWMA chart ในการตรวจจับการโจมตี SYN Flooding โดยทดสอบกับข้อมูล DARPA 99 และการโจมตีมีทั้งอัตราต่ำและสูง ซึ่งในส่วนของ EWMA สามารถตรวจจับได้ทั้งอัตราการโจมตีที่ต่ำและสูง ส่วนงานวิจัยของ Machaka และคณะ [7] ได้นำเสนอการตรวจสอบประสิทธิภาพของ EWMA สำหรับการตรวจจับการโจมตี SYN Flooding ใน

Internet of things (IoT) ซึ่งเป็นการนำข้อมูลที่ใช้จริงจาก MIT Lincoln มาทดสอบและตรวจสอบพิจารณา SYN แพ็คเก็ตในช่วงเวลา 10 วินาที เพื่อตรวจสอบอัตราการตรวจจับของอัลกอริทึม การแจ้งเตือนผิด ความล่าช้าในการตรวจจับ ทดสอบการปรับพารามิเตอร์ (Parameters) ในสมการของ EWMA และทดสอบความเข้มข้นของอัตราการโจมตีส่งผลต่อประสิทธิภาพการตรวจจับอย่างไร ผลการทดสอบ คือ EWMA สามารถที่จะตรวจจับในอัตราการโจมตีที่สูงได้ผลลัพธ์ที่ดี แต่ในอัตราการโจมตีที่ต่ำได้ผลลัพธ์ที่ไม่ดี และ Nishanth และคณะ [8] นำเสนอการตรวจจับการโจมตี SYN Flooding ใน Mobile Adhoc Network ทดสอบกับข้อมูล DARPA 1999 โดยใช้ Adaptive thresholding algorithm แต่ประสบกับอัตราผลบวกลวงสูงเนื่องจากมีการอัพเดท (Update) ค่าทางสถิติในเครือข่ายของ SAR (SYN Arrival Rates) ทั้งปกติและโจมตี ซึ่งงานวิจัยนี้มีการปรับแต่งอัลกอริทึมโดยการปรับค่า SAR เฉพาะค่าในเครือข่ายที่เป็นปกติ ทำให้ได้ค่าการตรวจจับค่าความถูกต้องที่สูงและอัตราผลบวกลวงที่ต่ำ แต่ไม่ได้พิจารณาถึงอัตราการโจมตีที่ต่างกัน ซึ่งในงานของ Ramkumar และคณะ [2] ไม่ได้พิจารณาถึงอัตราการโจมตีต่ำ Machaka และคณะ [7] ประสบปัญหาในการตรวจจับในอัตราการโจมตีที่ต่ำ และในงานวิจัย Nishanth และคณะ [8] ไม่ได้กล่าวถึงอัตราการโจมตี ขณะที่งานวิจัยของ Bouyeddou และคณะ [5] สามารถตรวจจับการโจมตีได้ทั้งอัตราการโจมตีที่ต่ำและอัตราการโจมตีที่สูง ทำให้มีข้อได้เปรียบกว่างานวิจัยที่ผ่านมาหากแต่ในงานวิจัยของ Bouyeddou และคณะ [5] ไม่ได้กล่าวถึงรายละเอียดของการแบ่งแพ็คเก็ตในเวลาที่ต่างกันเพื่อตรวจสอบ SYN แพ็คเก็ต ในขณะที่ Machaka และคณะ [7] มีการตรวจสอบ SYN แพ็คเก็ตในช่วงเวลา 10 วินาที จึงนำไปสู่คำถามที่ว่า 1) การแบ่งแพ็คเก็ตในเวลาที่ต่างกัน (10 วินาที (กลุ่มขนาดเล็ก) 30 วินาที (กลุ่มขนาดกลาง) และ 1 นาที (กลุ่มขนาดใหญ่) โดยการแบ่งแพ็คเก็ต หมายถึง การตรวจสอบ SYN แพ็คเก็ตตามลำดับการแบ่ง เช่น การแบ่งแพ็คเก็ตเวลา 1 นาที คือ การตรวจสอบ SYN แพ็คเก็ตทุกช่วง 1 นาที อย่างเป็นลำดับ และการแบ่งแพ็คเก็ต 10 วินาที และ 30 วินาที ดำเนินการในทำนองเดียวกัน) 2) อัตราการโจมตีต่ำและอัตราการโจมตีสูง 3) การมีผู้ใช้และไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในขณะโจมตี ทั้งสามปัจจัยจะส่งผลต่อประสิทธิภาพการตรวจจับด้วย EWMA อย่างไร โดยในงานวิจัยนี้ประเมินผลจากค่าความถูกต้อง อัตราผลบวกลวง และอัตราผลลบลวง โดยทดสอบผ่านชุดข้อมูลที่จำลองขึ้นบนสภาพแวดล้อมที่ควบคุม

2. วิธีการทดลอง

งานวิจัยนี้เสนอวิธีการตรวจจับการโจมตี SYN Flooding โดยใช้ EWMA ซึ่งประกอบด้วย 4 ขั้นตอนดังนี้ 1) การเก็บข้อมูล (Collect data) 2) การแบ่งแพ็คเก็ตตามเวลา 10 วินาที 30 วินาที และ 1 นาที (Sample packet 10s 30s and 1 min) 3) การตรวจจับการโจมตี SYN Flooding (Detect SYN flooding attack) และ 4) การประเมินผล (Evaluation) ดังแสดงในรูปที่ 3



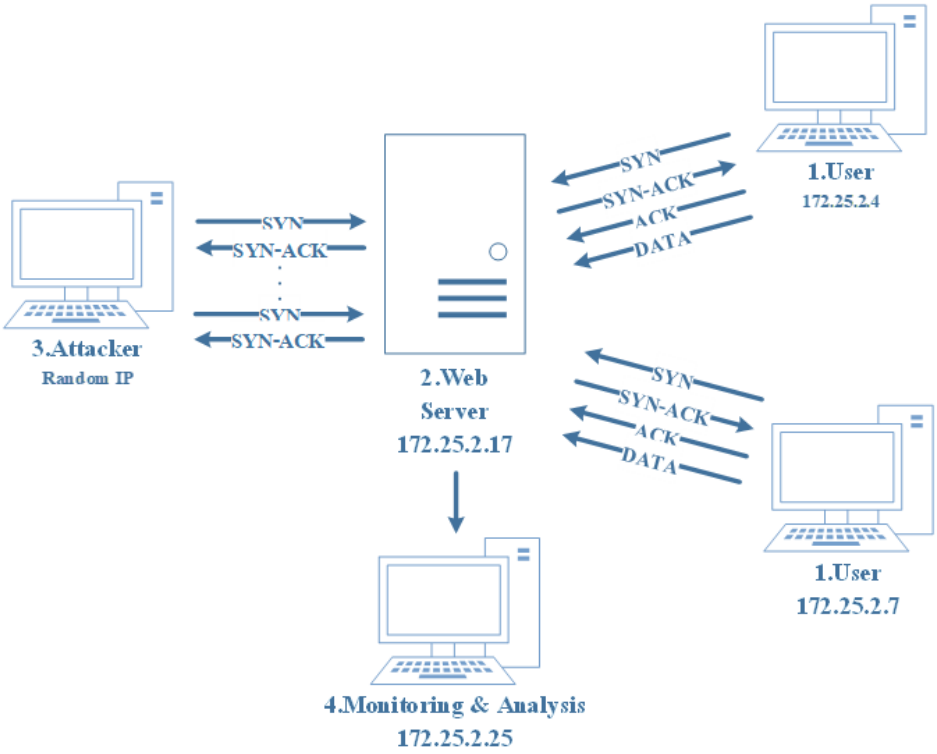
รูปที่ 3. ขั้นตอนการตรวจจับการโจมตี SYN Flooding

2.1 การเก็บข้อมูล (Collect data)

ขั้นตอนการเก็บข้อมูลได้รับการออกแบบสถาปัตยกรรมแวดล้อมเพื่อเก็บชุดข้อมูลจำลองดังแสดงในรูปที่ 4 โดยรูปแบบโครงสร้างข้อมูลที่เก็บ แสดงรายละเอียดในตารางที่ 1 ประกอบด้วย Count (จำนวนแพ็คเก็ต), Times, SourceIP (Source IP), SourcePort (Source Port), DestinationIP (Destination IP), DestPort (Destination Port), Flag (S-SYN, SA-SYN/ACK, A-ACK, PA-PUSH/ACK) และ Data (Raw มีข้อมูล และ NULL ไม่มีข้อมูล) สำหรับสถาปัตยกรรมแวดล้อมของระบบจำลองอยู่ในโปรแกรมจำลอง Virtual box โดยทำงานบนระบบปฏิบัติการ Microsoft Windows 10 และดำเนินการทดลองบนคอมพิวเตอร์ส่วนบุคคล Intel(R) Core(TM) i7-8750H CPU @2.20GHz 2.20 GHz, และ 16GB RAM ได้รับการออกแบบดังนี้

ตารางที่ 1. รูปแบบโครงสร้างข้อมูล

Count	Times	SourceIP	SourcePort	DestinationIP	DestPort	Flag	Data
1	2022-01-07 15:02:47	172.25.2.4	37492	172.25.2.17	443	S	NULL
2	2022-01-07 15:02:47	172.25.2.17	443	172.25.2.4	37492	SA	NULL
3	2022-01-07 15:02:47	172.25.2.4	37492	172.25.2.17	443	A	NULL
4	2022-01-07 15:02:47	172.25.2.4	37492	172.25.2.17	443	PA	Raw
5	2022-01-07 15:02:47	172.25.2.17	443	172.25.2.4	37492	A	NULL



รูปที่ 4. สภาพแวดล้อมสถาปัตยกรรมของระบบ

1. ผู้ใช้ (User) ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ (Web server) เพื่อเข้าใช้เว็บไซต์ โดยทำงานบนระบบปฏิบัติการ Ubuntu Linux

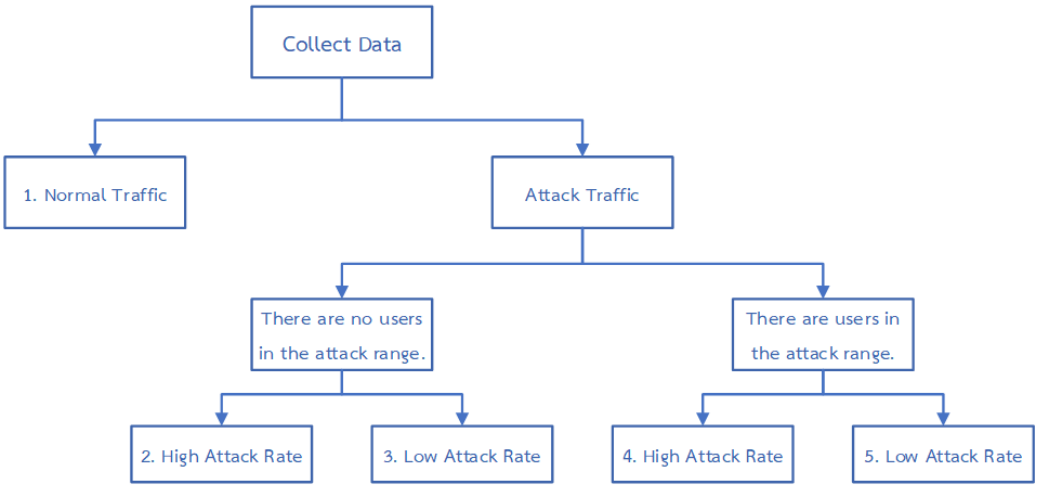
2. เว็บเซิร์ฟเวอร์ เปิดเว็บไซต์เพื่อให้ผู้ใช้เข้าใช้งานและจัดเก็บข้อมูลทราฟฟิคในฐานข้อมูล โดยทำงานบนระบบปฏิบัติการ Ubuntu Linux ใช้โปรแกรม Apache เป็นโปรแกรมจำลองเครื่องเป็นเว็บเซิร์ฟเวอร์ ใช้ MySQL ในการจัดการฐานข้อมูล และโมดูล Scapy ในการ Capture ข้อมูลทราฟฟิค [2]

3. ผู้โจมตี (Attacker) โจมตีทั้งในอัตราการโจมตีต่ำและอัตราการโจมตีสูง โดยใช้เครื่องมือ Hping3 บนระบบปฏิบัติการ Kali Linux จำลองการโจมตี [2]

4. Monitoring and analysis นำข้อมูลจากเว็บเซิร์ฟเวอร์วิเคราะห์ด้วยขั้นตอนวิธีการตรวจจับที่นำเสนอ (Proposed algorithm) โดยทำงานบนระบบปฏิบัติการ Ubuntu Linux และใช้โปรแกรม Jupyter Notebook ในการเขียนโปรแกรม ดังแสดงใน Algorithm 1

การเก็บข้อมูลแบ่งเป็น 2 กรณี คือ Normal Traffic และ Attack Traffic ซึ่งทั้ง 2 กรณีดำเนินการเก็บข้อมูลในช่วงเวลาประมาณ 10 นาที โดยรูปแบบ Normal Traffic คือ ผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เพื่อเข้าใช้เว็บไซต์ในลักษณะปกติและรูปแบบ Attack Traffic คือ ผู้ใช้ร้องขอข้อมูลไปยังเว็บ

เซิร์ฟเวอร์เพื่อเข้าใช้เว็บไซต์แต่มีการโจมตีจากผู้โจมตีทั้งอัตราการโจมตีต่ำหรืออัตราการโจมตีสูงในช่วงเวลาหนึ่ง และในช่วงเวลาการโจมตีทั้งอัตราการโจมตีต่ำหรืออัตราการโจมตีสูงมีผู้ใช้และไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์เพื่อเข้าใช้เว็บไซต์ โดยรูปแบบข้อมูลทั้ง 5 รูปแบบ จำแนกดังแผนภาพในรูปที่ 5



รูปที่ 5. จำแนกรูปแบบข้อมูลทั้ง 5 รูปแบบ

ตามรูปที่ 5 สามารถอธิบายได้ว่า

1. ผู้ใช้ 172.25.2.4 และผู้ใช้ 172.25.2.7 ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เพื่อเข้าใช้เว็บไซต์ประมาณ 5-7 ครั้ง เป็นรูปแบบข้อมูล Normal Traffic

2. ผู้ใช้ 172.25.2.4 และผู้ใช้ 172.25.2.7 ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เพื่อเข้าใช้เว็บไซต์ประมาณ 5-7 ครั้งและทดสอบการโจมตี SYN Flooding ความเข้มสูง (ประมาณ 520 ส่วน/การสังเกต) ในเวลา 2 นาที [5] เป็นรูปแบบข้อมูล High Attack Rate แต่ไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาการโจมตี

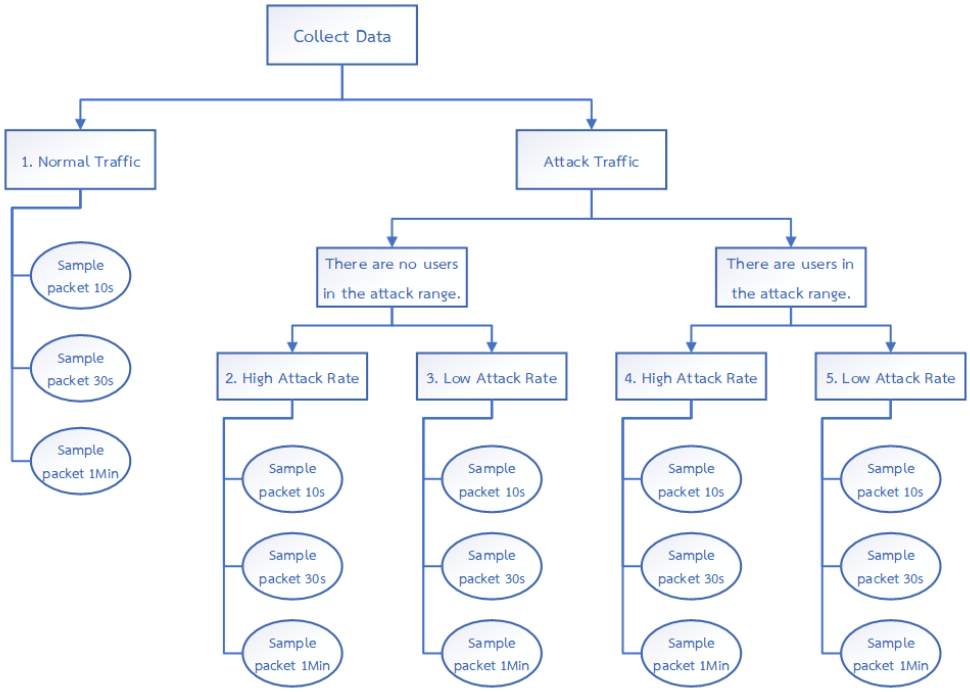
3. ผู้ใช้ 172.25.2.4 และผู้ใช้ 172.25.2.7 ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เพื่อเข้าใช้เว็บไซต์ประมาณ 5-7 ครั้ง และทดสอบการโจมตี SYN Flooding ความเข้มต่ำ (ประมาณ 125 ส่วน/การสังเกต) ในเวลา 2 นาที [5] เป็นรูปแบบชุดข้อมูล Low Attack Rate แต่ไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาการโจมตี

4. ผู้ใช้ 172.25.2.4 และผู้ใช้ 172.25.2.7 ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เพื่อเข้าใช้เว็บไซต์ประมาณ 5-7 ครั้ง และทดสอบการโจมตี SYN Flooding ความเข้มสูง (ประมาณ 520 ส่วน/การสังเกต) ในเวลา 2 นาที [5] และผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาการโจมตี

5. ผู้ใช้ 172.25.2.4 และผู้ใช้ 172.25.2.7 ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เพื่อเข้าใช้เว็บไซต์ประมาณ 5-7 ครั้ง และทดสอบการโจมตี SYN Flooding ความเข้มต่ำ (ประมาณ 125 ส่วน/การสังเกต) ในเวลา 2 นาที [5] และผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาการโจมตี

2.2 การแบ่งแพ็คเก็ตตามเวลา 10 วินาที 30 วินาที และ 1 นาที (Sample packet 10s 30s and 1 min)

ในขั้นตอนนี้นำข้อมูลจากขั้นตอนการเก็บข้อมูลทั้งหมด 5 รูปแบบมาแบ่งแพ็คเก็ตตามเวลา 10 วินาที 30 วินาที และ 1 นาทีตามลำดับ ดังรูปที่ 6 เพื่อแทนแพ็คเก็ตขนาดเล็ก กลางและใหญ่ ทำให้ได้รูปแบบข้อมูลที่แตกต่างกันทั้งหมดจำนวน 15 รูปแบบ เพื่อเป็นข้อมูลนำเข้าสู่กระบวนการตรวจจับการโจมตี SYN Flooding ต่อไป



รูปที่ 6. การแบ่งแพ็คเก็ตตามเวลา 10 วินาที 30 วินาที และ 1 นาที

2.3 การตรวจจับการโจมตี SYN Flooding (Detect SYN flooding attack)

งานวิจัยนี้ประยุกต์ใช้แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบเลขชี้กำลัง ในขั้นตอนวิธีการตรวจจับการโจมตี SYN Flooding เพื่อระบุแพ็คเก็ตเป็นปกติหรือโจมตีจากชุดข้อมูล Attack Traffic เมื่อเทียบกับชุดข้อมูล Normal Traffic

2.3.1 EWMA Control chart

คือ แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบเลขชี้กำลัง (Exponentially Weighted Moving Average chart) หรือ EWMA เป็นแผนภูมิติดตามค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักแบบทวีคูณของค่าเฉลี่ยตัวอย่างก่อนหน้าทั้งหมดและเป็นวิธีการมองการเปลี่ยนแปลงบนเวลามีวัตถุประสงค์เพื่อ

ตรวจจับการเปลี่ยนแปลงในค่าเฉลี่ยตลอดกระบวนการอย่างรวดเร็ว [6] และสำหรับขีดจำกัดควบคุมของแผนภูมิควบคุม EWMA ประกอบด้วยขีดจำกัดควบคุมบน (Upper Control Limit: UCL) และขีดจำกัดควบคุมล่าง (Lower Control Limit: LCL) โดยมีสมการดังนี้

$$EWMA_i = Z_i = \lambda x_i + (1-\lambda)Z_{i-1} \quad (1)$$

$$UCL_i = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \quad (2)$$

$$LCL_i = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \quad (3)$$

x_i แทน Sample ปัจจุบัน μ_0 และ σ แทนค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ x ภายใต้กรณีปกติ และ λ แทนค่าถ่วงน้ำหนัก ($0<\lambda<1$) เมื่อ λ เข้าใกล้ 0 มีแนวโน้มในการตรวจจับการเปลี่ยนแปลงไปที่ข้อมูลก่อนหน้า เมื่อ λ เข้าใกล้ 1 มีแนวโน้มในการตรวจจับการเปลี่ยนแปลงไปที่ข้อมูลปัจจุบัน [5]-[7] และ L คือ ค่าสัมประสิทธิ์ของแผนภูมิควบคุม EWMA [9]

โดยงานวิจัยนี้ค่า x_i แทนค่า SAR ดังสมการที่ (4) ซึ่งเป็นอัตราส่วนของจำนวนแพ็คเก็ต SYN ต่อจำนวนแพ็คเก็ต TCP ทั้งหมดในแต่ละ Sample [8] และ Z_{i-1} แทนค่า EWMA ที่ $i-1$ โดยค่า i คือ ค่าดัชนีที่ชี้ไปที่แพ็คเก็ตที่ i ซึ่งการเก็บข้อมูลจำลองในงานวิจัยนี้อยู่ที่ 10 นาที เมื่อมีการแบ่งแพ็คเก็ตที่ช่วงเวลา 1 นาที ทำให้ค่า i แบ่งกลุ่มออกเป็น 10 กลุ่มนั้นคือ ค่า i = 1, 2, 3, ..., 10 และค่า i ในช่วงเวลา 10 วินาทีแบ่งกลุ่มออกเป็น 60 กลุ่ม และ 30 วินาทีแบ่งกลุ่มออกเป็น 20 กลุ่ม ตามลำดับ

$$SAR = \frac{Number\ of\ SYN\ Packets}{Total\ number\ of\ TCP\ Packets} \quad (4)$$

2.3.2 อัลกอริทึมการตรวจจับการโจมตี SYN Flooding

อัลกอริทึมที่นำเสนอในบรรทัดที่ 2-4 นำชุดข้อมูลทั้ง 5 รูปแบบจากขั้นตอนการเก็บข้อมูลเข้าสู่การแบ่งแพ็คเก็ตตามช่วงเวลา 10 วินาที (กลุ่มขนาดเล็ก) 30 วินาที (กลุ่มขนาดกลาง) และ 1 นาที (กลุ่มขนาดใหญ่) จะได้ข้อมูลทั้งหมด 15 รูปแบบ และคำนวณค่า SAR ในแต่ละชุดข้อมูลดังสมการที่ (4) เมื่อเข้าสู่บรรทัดที่ 5-6 นำค่า SAR ในชุดข้อมูล Normal Traffic จากการแบ่งแพ็คเก็ตเวลา 10 วินาที 30 วินาที และ 1 นาที คำนวณตามสมการ (1), (2), (3) ซึ่งเมื่อคำนวณชุด Normal Traffic ที่ผ่านการแบ่งแพ็คเก็ตในเวลา 10 วินาที 30 วินาทีและ 1 นาที ได้ค่า λ = 0.3, L = 3, μ_0 = 0.018 และ σ = 0.022 เพื่อใช้เป็นเกณฑ์มาตรฐานในการตรวจจับกับชุดข้อมูล Attack Traffic โดยจากตารางที่ 2 เป็นตัวอย่างผลการทดลอง Normal Traffic การแบ่งแพ็คเก็ต 1 นาที ซึ่งกำหนดค่า λ = 0.3, L = 3, μ_0 = 0.018 และ σ = 0.022 ส่งผลให้ค่าสถานะ (Status) ทุกช่วงการแบ่งแพ็คเก็ตแสดงผลทราฟฟิคของแพ็คเก็ตเป็นปกติ หรือหากพิจารณากราฟรูปที่ 7 ซึ่งนำพารามิเตอร์ Z_i, $CL(\mu_0)$, UCL_i และ LCL_i พล็อตในแผนภูมิควบคุม EWMA เมื่อสังเกตค่า Z_i อยู่ในเงื่อนไข $Z_i > LCL_i$ and $Z_i < UCL_i$ จากนั้นกำหนดค่า i = 1 และ

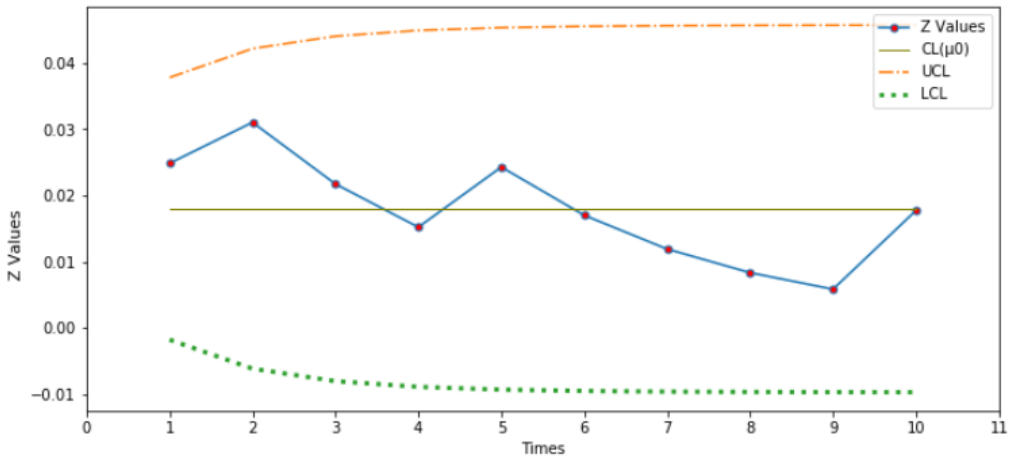
กำหนดค่า num of group ในบรรทัดที่ 8 คือ จำนวนของกลุ่มการแบ่งแพ็คเก็ต (เช่นหากมีการแบ่งแพ็คเก็ตเวลา 1 นาทีทำให้ค่า num of group เท่ากับ 10 และการแบ่งแพ็คเก็ตเวลา 10 วินาทีและ 30 วินาทีดำเนินการในทำนองเดียวกัน) จากนั้นนำชุดข้อมูล Attack Traffic ที่ผ่านการการแบ่งแพ็คเก็ตตามเวลา 10 วินาที 30 วินาที และ 1 นาที ซึ่งมี 12 รูปแบบ นำข้อมูลเข้าสู่ลูป While ทีละรูปแบบ ข้อมูล Attack Traffic ชุดแรกมีการกำหนดค่า x_i เท่ากับ SAR ซึ่งค่า SAR ได้จากบรรทัดที่ 4 ใน Algorithm 1 จากนั้นคำนวณค่า Z_i, LCL_i, และ UCL_i ตามสูตรสมการ (1), (2), (3) ถ้า $Z_i > LCL_i$ และ $Z_i < UCL_i$ ระบุทราฟฟิคของแพ็คเก็ตเป็นปกติ แต่หากค่า $Z_i > UCL_i$ และ $Z_i > Z_{i-1}$ ระบุทราฟฟิคของแพ็คเก็ตเป็นการโจมตี แต่หากค่า $Z_i < Z_{i-1}$ ระบุทราฟฟิคของแพ็คเก็ตเป็นปกติ เมื่อดำเนินการกับข้อมูลชุดแรกเสร็จแล้ว โหลดชุดข้อมูลรูปแบบถัดไปเข้าสู่ลูป While จนกระทั่งดำเนินการครบทั้ง 12 รูปแบบของชุดข้อมูล Attack Traffic

Algorithm 1 Proposed Algorithm

1. Begin
2. Collect Traffic data
3. Sample packet 10s, 30s and 1Min
4. Calculate SAR each sample
5. Find μ_0 and σ for SAR in each sample
6. Determine value in L, λ, and $Z_0 = \mu_0$
7. i = 1
8. While (i $\leq$ num of groups):
9. $x_i = SAR$
10. Find Z_i, LCL_i, UCL_i
11. If $Z_i > LCL_i$ and $Z_i < UCL_i$
12. Traffic = normal
13. If $Z_i > UCL_i$
14. If $Z_i > Z_{i-1}$
15. Traffic = attack
16. If $Z_i < Z_{i-1}$
17. Traffic = normal
18. i = i+1
19. End

ตารางที่ 2. ผลการทดลองรูปแบบ Normal Traffic โดยข้อมูลได้รับการแบ่งแพ็คเก็ตที่เวลา 1 นาที

ID	Time	SAR	Z	UCL	LCL	Status
1	2022-02-02 10:31:00	0.040816	0.024845	0.0378	-0.0018	Normal
2	2022-02-02 10:32:00	0.045455	0.031028	0.042169	-0.006169	Normal
3	2022-02-02 10:33:00	0	0.021719	0.044044	-0.008044	Normal
4	2022-02-02 10:34:00	0	0.015204	0.044915	-0.008915	Normal
5	2022-02-02 10:35:00	0.045455	0.024279	0.045331	-0.009331	Normal
6	2022-02-02 10:36:00	0	0.016995	0.045533	-0.009533	Normal
7	2022-02-02 10:37:00	0	0.011897	0.045631	-0.009631	Normal
8	2022-02-02 10:38:00	0	0.008328	0.045679	-0.009679	Normal
9	2022-02-02 10:39:00	0	0.005829	0.045703	-0.009703	Normal
10	2022-02-02 10:40:00	0.045455	0.017717	0.045714	-0.009714	Normal



รูปที่ 7. ผลจากการทดสอบชุดข้อมูล Normal Traffic และมีชุดข้อมูลการแบ่งแพ็คเก็ตที่เวลา 1 นาที

2.4 การประเมินประสิทธิภาพ (Evaluation)

สำหรับการประเมินประสิทธิภาพของงานวิจัยได้ดำเนินการผ่านหลักการความไวและความจำเพาะ (Sensitivity and specificity) [10] ประเมินจากค่าความถูกต้อง อัตราผลบวกลวงและอัตราผลลบลวง ซึ่งพารามิเตอร์ต่าง ๆ ได้รับการกำหนดดังตารางที่ 3 [8]

ตารางที่ 3. Confusion matrix

		Actual Values: Positive (Attack)	Actual Values: Negative (Normal)
Predict Values	Positive (Attack)	TP	FP
	Negative (Normal)	FN	TN

จากตารางที่ 3 พารามิเตอร์ใน Confusion matrix ประกอบด้วย
True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่า “โจมตี” และ มีค่าความจริงเป็น “โจมตี”
True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่า “ปกติ” และมีค่าความจริงเป็น “ปกติ”
False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่า “โจมตี” แต่มีค่าความจริงเป็น “ปกติ”
False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่า “ปกติ” แต่มีค่าความจริงเป็น “โจมตี”
Accuracy คือ ค่าความถูกต้องที่คาดการณ์ได้ตรงกับสิ่งที่เกิดขึ้นจริง [8] โดยมีสมการดังนี้

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

False Positive Rate (FPR) คือ อัตราส่วนของ FP ต่อผลรวมของ FP และ TN [8]

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{6}$$

False Negative Rate (FNR) คือ อัตราส่วนของ FN ต่อผลรวมของ FN และ TP [11]

$$False\ Negative\ Rate = \frac{FN}{FN + TP} \tag{7}$$

ขั้นตอนการประเมินประสิทธิภาพได้ดำเนินการบนชุดข้อมูลข้างต้น โดยเข้าสู่กระบวนการตรวจจับการโจมตี SYN Flooding เพื่อระบุทราฟฟิคของแพ็คเก็ตเป็น Normal หรือ Attack และนำค่าเข้าสู่ตารางที่ 3 Confusion matrix เพื่อประเมินประสิทธิภาพจากค่าความถูกต้อง อัตราผลบวกลวงและอัตราผลลบลวง ซึ่งตารางที่ 4 แสดงตัวอย่างผลการทดสอบในชุดข้อมูล Attack Traffic กรณีมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาอัตราการโจมตีสูงการแบ่งแพ็คเก็ตที่เวลาหนึ่งนาทีผ่านอัลกอริทึมที่นำเสนอในขั้นตอน 2.3.2 จากนั้นนำแพ็คเก็ตที่ถูกระบุค่าสถานะเป็น Normal หรือ Attack เข้าสู่ตาราง Confusion matrix เพื่อประเมินผล จากตารางที่ 5 ค่าผลบวกจริง (True positive) เท่ากับ 618 แพ็คเก็ตค่าผลบวกลวง (False positive) เท่ากับ 2 แพ็คเก็ต ค่าผลลบลวง(False negative) เท่ากับ 0 แพ็คเก็ตและค่าผลลบ

จริง (True negative) เท่ากับ 5 แพ็คเก็ต เมื่อคำนวณค่าความถูกต้อง อัตราผลบวกลวงและอัตราผลลบลวงดังสมการที่ (5), (6) และ (7) เพื่อประเมินประสิทธิภาพ ซึ่งในกรณีนี้ได้ค่าความถูกต้องเท่ากับ 0.997 อัตราผลบวกลวงเท่ากับ 0.286 และอัตราผลลบลวงเท่ากับ 0 หากพิจารณาผลการทดลองในตารางที่ 4 ประกอบกับกราฟรูปที่ 8 ซึ่งนำพารามิเตอร์ Z_i, CL, UCL_i และ LCL_i พล็อตในแผนภูมิควบคุม EWMA เมื่อสังเกตกราฟในช่วงเวลา 1-3 นาทีค่า Z_i มีค่าประมาณ 0.02 , UCL_i ประมาณ 0.04 และค่า LCL_i ที่มีการลดลงจาก -0.002 จนถึง -0.008 ซึ่งค่า Z_i อยู่ในเงื่อนไข $Z_i > LCL_i$ and $Z_i < UCL_i$ ทำให้สามารถระบุทราฟฟิคของแพ็คเก็ตเป็นปกติ แต่กราฟในช่วงเวลา 4-11 นาที ค่า Z_i อยู่ในเงื่อนไข $Z_i > UCL_i$ หากสังเกตจากกราฟในนาทีที่ 4 ค่า Z_i มีค่าประมาณ 0.1 และค่า UCL_i มีค่าประมาณ 0.045 ซึ่งค่า Z_i สูงกว่า UCL_i และกราฟมีความชันขึ้นอย่างรวดเร็วจนไปถึงนาทีที่ 6 ซึ่งค่า Z_i มีค่าประมาณ 0.22 ซึ่งเป็นค่าสูงสุดทำให้ค่า Z_i ในช่วงเวลา 4-6 อยู่ในเงื่อนไข $Z_i > Z_{i-1}$ และสามารถระบุทราฟฟิคของแพ็คเก็ตเป็นโจมตี แต่กราฟในช่วงเวลา 7-11 ค่า Z_i เริ่มลดลง โดยค่า Z_i ในนาทีที่ 7 มีค่าประมาณ 0.16 ซึ่งต่ำกว่าค่า Z_i ในช่วงเวลาที่ 6 ที่มีค่า Z_i ประมาณ 0.22 และลดลงตามลำดับจนถึงช่วงเวลาที่ 11 โดยมีค่า Z_i ประมาณ 0.05 ทำให้ค่า Z_i ในช่วงเวลา 7-11 อยู่ในเงื่อนไข $Z_i < Z_{i-1}$ และสามารถระบุทราฟฟิคของแพ็คเก็ตเป็นปกติ สำหรับชุดข้อมูล Attack Traffic รูปแบบอื่น ๆ ได้ดำเนินการประเมินประสิทธิภาพในทำนองเดียวกัน

ตารางที่ 4. ผลการทดลองในกรณีมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาอัตราการโจมตีสูง โดยข้อมูลได้รับการแบ่งแพ็คเก็ตที่เวลา 1 นาที

ID	Time	SAR	Z	UCL	LCL	Status
1	2022-02-15 16:15:00	0.040816	0.024845	0.037800	-0.001800	Normal
2	2022-02-15 16:16:00	0	0.017391	0.042169	-0.006169	Normal
3	2022-02-15 16:17:00	0.045455	0.025810	0.044044	-0.008044	Normal
4	2022-02-15 16:18:00	0.306154	0.109913	0.044915	-0.008915	Attack
5	2022-02-15 16:19:00	0.335058	0.177457	0.045331	-0.009331	Attack
6	2022-02-15 16:20:00	0.324649	0.221615	0.045533	-0.009533	Attack
7	2022-02-15 16:21:00	0	0.155130	0.045631	-0.009631	Normal
8	2022-02-15 16:22:00	0.045455	0.122228	0.045679	-0.009679	Normal
9	2022-02-15 16:23:00	0	0.085559	0.045703	-0.009703	Normal
10	2022-02-15 16:24:00	0	0.059891	0.045714	-0.009714	Normal
11	2022-02-15 16:25:00	0.04	0.053924	0.045720	-0.009720	Normal

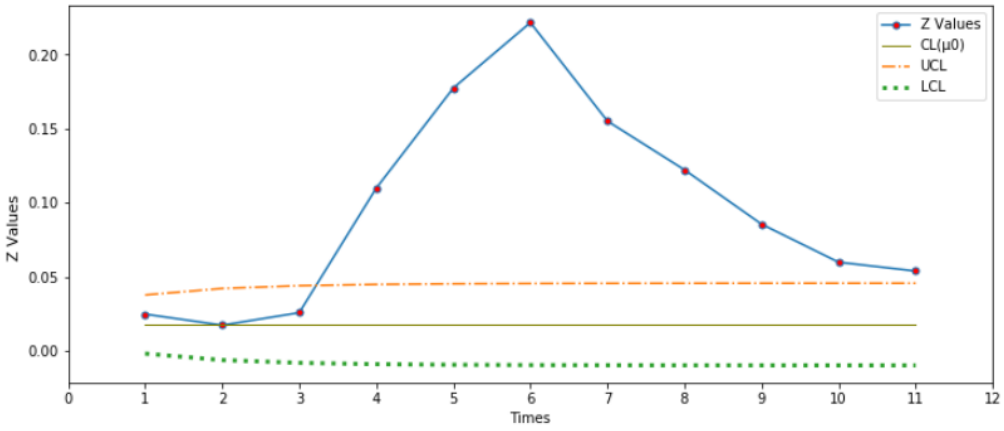
ตารางที่ 5. Confusion matrix จากการทดสอบชุดข้อมูลที่มีผู้ใช้เข้าใช้งานในช่วงเวลาการโจมตีอัตราสูงและมีชุดข้อมูลการแบ่งแพ็คเก็ตที่เวลา 1 นาที

		Actual Values	
		Positive (Attack)	Negative (Normal)
Predict Values	Positive (Attack)	618	2
	Negative (Normal)	0	5

$$\text{Accuracy} = \frac{618+5}{618+5+2} = 0.997$$

$$\text{FPR} = \frac{FP}{FP+TN} = \frac{2}{2+5} = 0.286$$

$$\text{FNR} = \frac{FN}{FN+TP} = \frac{0}{0+618} = 0$$



รูปที่ 8. ผลจากการทดสอบชุดข้อมูลที่มี ผู้ใช้เข้าใช้งาน
ในช่วงเวลาการโจมตีอัตราสูงและชุดข้อมูลมีการแบ่งแพ็คเก็ตที่เวลา 1 นาที

3. ผลการทดลอง

ผลการทดสอบระบบตรวจจับการโจมตี SYN Flooding ด้วย EWMA และการประเมินประสิทธิภาพด้วยชุดข้อมูล Attack Traffic 12 ชุด ทั้งกรณีไม่มีผู้ใช้และมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาการโจมตีทั้งอัตราการโจมตีต่ำและอัตราการโจมตีสูง จำแนกตามการแบ่งแพ็คเก็ตตาม

เวลา 10 วินาที (กลุ่มขนาดเล็ก) 30 วินาที (กลุ่มขนาดกลาง) และ 1 นาที (กลุ่มขนาดใหญ่) ส่งผลต่อค่าความถูกต้อง อัตราผลบวกลวงและอัตราผลลบลวง ดังแสดงในตารางที่ 6

ตารางที่ 6. ผลการทดลองของชุด Attack Traffic ทั้ง 12 ชุด

ID	ผู้ใช้	อัตราการโจมตี	การแบ่งแพ็คเก็ต (Sample Packet)	ความถูกต้อง (Accuracy)	อัตราผลบวกลวง (False positive rate)	อัตราผลลบลวง (False negative rate)
1	ไม่มี	ต่ำ	1 นาที	1	0	0
2			30 วินาที	1	0	0
3			10 วินาที	0.811	0	0.195
4		สูง	1 นาที	1	0	0
5			30 วินาที	0.952	0	0.048
6			10 วินาที	0.827	0	0.174
7	มี	ต่ำ	1 นาที	0.988	0.33	0
8			30 วินาที	0.988	0.33	0
9			10 วินาที	0.735	0	0.274
10		สูง	1 นาที	0.997	0.286	0
11			30 วินาที	0.997	0.286	0
12			10 วินาที	0.946	0.286	0.052

จากทั้งกรณีไม่มีผู้ใช้และมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ พบว่าการแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่ ส่งผลให้ค่าความถูกต้องสูงและอัตราผลลบลวงต่ำ แต่ในขณะที่การแบ่งแพ็คเก็ตในกลุ่มขนาดเล็กส่งผลให้ค่าความถูกต้องมีค่าลดลงและอัตราผลลบลวงมีค่าที่สูงขึ้น หากพิจารณาอัตราผลลบลวงในกรณีมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เมื่อแบ่งแพ็คเก็ตที่กลุ่มขนาดเล็กหรือที่เวลา 10 วินาที อัตราผลลบลวงมีค่าประมาณ 0.27 ในอัตราการโจมตีต่ำและมีค่าประมาณ 0.05 ในอัตราการโจมตีสูง ซึ่งอัตราผลลบลวงในอัตราการโจมตีต่ำมีค่าที่สูงกว่าในอัตราการโจมตีสูง ส่งผลให้อัตราการโจมตีต่ำและอัตราการโจมตีสูงมีผลต่ออัตราผลลบลวงในการแบ่งแพ็คเก็ตในกลุ่มขนาดเล็ก ดังนั้นการแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่ให้ประสิทธิภาพดีกว่าการแบ่งแพ็คเก็ตที่กลุ่มขนาดเล็ก ทั้งในกรณีไม่มีผู้ใช้และมีผู้ใช้ร้องขอข้อมูลในช่วงเวลาที่มีการโจมตี ยกเว้นในกรณีไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีสูง ซึ่งมีการแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่ ค่าความถูกต้องและอัตราผลลบลวงมีค่าที่ใกล้เคียงกัน

กรณีไม่มีผู้ใช้และมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำและอัตราการโจมตีสูง ส่งผลต่ออัตราผลบวกลวงเมื่อแบ่งแพ็คเก็ตที่กลุ่มขนาดกลางและกลุ่มขนาดใหญ่ กรณีมีผู้ใช้มีอัตราผลบวกลวงประมาณ 0.3 ส่งผลให้การไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำและอัตราการโจมตีสูงมีประสิทธิภาพดีกว่ากรณีมีผู้ใช้ร้องขอข้อมูลขณะที่มีการโจมตี

จากผลการทดลองสามารถสรุปได้ว่าปัจจัยที่ส่งผลกระทบต่อประสิทธิภาพการตรวจจับด้วย EWMA คือ 1) ช่วงเวลาการการแบ่งแพ็คเก็ตซึ่งจะส่งผลต่อค่าความถูกต้องและอัตราผลลบลวง เมื่อแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่ให้ค่าความถูกต้องสูงและอัตราผลลบลวงต่ำ ขณะที่การแบ่งแพ็คเก็ตในกลุ่มขนาดเล็กให้ค่าความถูกต้องต่ำและอัตราผลลบลวงสูง ส่งผลให้การแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่มีประสิทธิภาพมากกว่าการแบ่งแพ็คเก็ตในกลุ่มขนาดเล็ก ทั้งนี้เนื่องจากการแบ่งแพ็คเก็ตในกลุ่มขนาดเล็กมีโอกาสส่งผลให้ค่า SAR บางกลุ่มลดลงจากกลุ่มก่อนหน้า เมื่อผ่านอัลกอริทึม EWMA ส่งผลให้ค่าสถานะถูกทำนายเป็น Normal จากความจริงเป็น Attack ส่งผลให้อัตราผลลบลวงสูงและค่าความถูกต้องลดลง 2) อัตราการโจมตีต่ำและอัตราการโจมตีสูงส่งผลต่ออัตราผลลบลวงและอัตราผลบวกลวง ในกรณีมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ เมื่อแบ่งแพ็คเก็ตในกลุ่มขนาดเล็กจะให้อัตราผลลบลวงที่สูงขณะอัตราการโจมตีต่ำ และให้อัตราผลบวกลวงสูงขณะอัตราการโจมตีสูง ทั้งนี้เนื่องจากกรณีมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำ เมื่อแบ่งแพ็คเก็ตในกลุ่มขนาดเล็ก ส่งผลให้ค่า SAR บางกลุ่มลดลงจากกลุ่มก่อนหน้า เมื่อผ่านอัลกอริทึม EWMA ส่งผลให้อัตราผลลบลวงสูง ในทางกลับกันกรณีมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีสูง เมื่อแบ่งแพ็คเก็ตในกลุ่มขนาดเล็ก ส่งผลให้ค่า SAR บางกลุ่มเพิ่มขึ้นจากกลุ่มก่อนหน้า เมื่อผ่าน EWMA ส่งผลให้อัตราผลบวกลวงสูง 3) การมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำและอัตราการโจมตีสูงส่งผลต่ออัตราผลบวกลวงที่สูง เมื่อแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่ ทั้งนี้เนื่องจากการมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำและสูง เมื่อแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและขนาดใหญ่มีค่าผลบวกลวงสูง เพราะค่า SAR แต่ละกลุ่มมีค่าใกล้เคียงกัน เมื่อผ่านอัลกอริทึม EWMA ส่งผลให้ค่า SAR บางกลุ่มเพิ่มขึ้นจากกลุ่มก่อนหน้า ทำให้ผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงเวลาการโจมตีถูกทำนายเป็น Attack จากความจริงเป็น Normal ส่งผลให้อัตราผลบวกลวงสูง

4. สรุปผลการทดลอง

งานวิจัยนี้นำเสนอการตรวจจับการโจมตี SYN Flooding ด้วย EWMA และตรวจสอบปัจจัยที่ส่งผลต่อประสิทธิภาพการตรวจจับ คือ 1) การแบ่งแพ็คเก็ตในเวลาต่างกัน 2) อัตราการโจมตีต่ำและอัตราการโจมตีสูง 3) การไม่มีผู้ใช้และมีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ขณะโจมตี โดยประเมินประสิทธิภาพผ่านค่าความถูกต้อง อัตราผลบวกลวงและอัตราผลลบลวง บนชุดข้อมูลที่จำลองขึ้นภายใต้สภาพแวดล้อมปิด

ผลการทดลองพบว่าขั้นตอนวิธี EWMA ที่นำเสนอสามารถตรวจจับการโจมตี SYN Flooding ได้ทั้งอัตราการโจมตีต่ำและอัตราการโจมตีสูง และทั้งสามปัจจัยข้างต้นส่งผลต่อประสิทธิภาพการตรวจจับ ซึ่งรูปแบบที่เหมาะสำหรับการตรวจจับ คือ กรณีไม่มีผู้ใช้ร้องขอข้อมูลไปยังเว็บเซิร์ฟเวอร์ในช่วงอัตราการโจมตีต่ำและอัตราการโจมตีสูงที่การแบ่งแพ็คเก็ตในกลุ่มขนาดกลางและกลุ่มขนาดใหญ่

ในอนาคตจะดำเนินการทดสอบเพิ่มเติมโดยปรับค่าพารามิเตอร์ในสมการ EWMA และช่วงการแบ่งแพ็คเก็ตในเวลาต่าง ๆ เพื่อตรวจวัดการตรวจจับประสิทธิภาพและนำขั้นตอนวิธีที่นำเสนอในงานวิจัยนี้ไปทดสอบกับชุดข้อมูลอื่น ๆ เพื่อเปรียบเทียบประสิทธิภาพในการตรวจจับการโจมตีว่าเหมือนหรือแตกต่างกันอย่างไร

เอกสารอ้างอิง (References)

[1] Yoachimik, O. and Ganti, V. 2022. DDoS Attack Trends for Q4 2021. Available at: https://blog.cloudflare.com/dDoS-attack-trends-for-2021-q4/. Retrieved 15 January 2022.

[2] Ramkumar, B.N. and Subbulakshmi, T. 2021. TCP SYN flood attack detection and prevention system using adaptive thresholding method. Proceeding ITM Web of Conferences 37, International Conference on Innovative Technology for Sustainable Development (ICITSD 2021), 1-8.

[3] Ransewa, S., Elz, N., Thanon, N. and Intajag, S. 2018. Anomaly detection using Source Port Data with Shannon Entropy and EWMA Control Chart. Proceeding 18th International Conference on Control, Automation and Systems (ICCAS 2018), GangWon, Korea, 596-601.

[4] Al-mansor, M.J. and Gan, K.B. 2018. Intrusion detection systems: principles and perspectives. *Journal of Multidisciplinary Engineering Science Studies*, 4(11), 2266-2270.

[5] Bouyeddou, B., Harrou, F., Sun, Y. and Kadri, B. 2017. Detecting SYN flood attacks via statistical monitoring charts: A comparative study. Proceedings 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), Boumerdes, Algeria,1-5.

[6] Montgomery, D.C. 2009. Introduction to Statistical Quality Control. 6th ed, John Wiley & Sons, New York.

[7] Machaka, P., Bagula, A. and Nelwamondo, F. 2016. Using Exponentially Weighted Moving Average Algorithm to Defend Against DDoS Attacks. Proceedings Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech) Stellenbosch, South Africa, 1-6.

[8] Nishanth, N. and Mujeeb, A. 2021. Application of Adaptive Threshold Algorithm with selected modified parameters for the Detection of flooding based Denial-of-Service (DoS) attack in Mobile Ad Hoc Network. Proceeding International conference on systems energy and environment, GCE Kannur, Kerala, India, 119-123.

[9] ชัชฎาภา ดีวุ่น และเปรมพร เขมาวุฒฆ์. 2561. การศึกษาประสิทธิภาพของแผนภูมิควบคุม p, Ewma และ Isrt p ewma. *วิศวกรรมสารเกษมบัณฑิต*, 8(2), 180-193. [Chatchadapa Dewun and Premporn Khemavuk. 2018. A STUDY OF EFFICIENCY OF P, EWMA AND ISRT P EWMA CONTROL CHARTS. *Kasem Bundit Engineering Journal*, 8(2), 180-193. (in Thai)]

[10] YERUSHALMY, J. 1947. Statistical Problems in Assessing Methods of Medical Diagnosis, with Special Reference to X-ray Techniques. *Public health reports*, 62(40), 1432-1449.

[11] Liu, H. and Lang, B. 2019. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. *Applied Sciences*, 9(20), 1-28.