บทที่ 3
วิธีดำเนินการวิจัย

3.1 นิยามกฎและความผิดปกติของกฎ

เมื่อพูดถึงรูปแบบทั่วไปของกฎไฟร์วอลล์ประกออบด้วย 2 ส่วน คือ ส่วนของเงื่อนไขและส่วนการตัดสินใจ ให้ R เป็นกฎไฟร์วอลล์ C เป็นส่วนเงื่อนไขและ A เป็นส่วนการตัดสินใจ

กฎไฟร์วอลล์จะมีรูปแบบติอไปนี้

$$R\text{: } C \Rightarrow A \tag{1}$$

ในความเป็นจริงกฎไฟร์วอลล์มีกฎมากกว่าหนึ่งกฎเสมอ ดังนั้นสมการแรก (1) จะต้องมีการแก้ไขในสมการที่สองดังนี้

$$R_i\text{: } C_i \Rightarrow A_i \tag{2}$$

โดยที่ C_i และ A_i เป็นเงื่อนไขและการตัดสินใจของกฎ R_i (กฎไฟร์วอลล์ใด ๆ) โดย $i \in [1, n]$ และ n เป็นจำนวนเต็มที่ไม่ติดลบ f_i แทนโดเมนของจำนวนเต็มบวกเป็นช่วง เขียนแทนด้วย $D\ (f_i)$ ตัวอย่าง เช่น โดเมนของที่อยู่ต้นทางและปลายทางในแพ็คเก็ต IP (SIP และ DIP) คือ $[0, 2^{32} - 1] D(f_1)$ และ $D\ (f_2)$ พอร์ตตน้ทางและพอร์ตปลายทาง (SP และ DP) คือ [0, 2^{16} - 1] และ โปรโตคอล คือ [0, 2^8 - 1] กำหนดให้ C_i เป็นชุดข้อมูล f_1 ถึง f_d ที่ระบุเป็น $f_1 \in F_1 \wedge f_2 F_2 \wedge \ldots \wedge f_d \in F_d$ โดยที่ F_i เป็นส่วนย่อของ $D\ (f_i)$ สว่ น A_i คือ การยอมรับหรือปฏิเสธ (Accept และ Deny) สำหรับแต่ละกฎ หากเงื่อนไขทั้งหมด (f_i) ใน C_i เป็นจริงจะส่งผลให้มีเพียงหนึ่ง

ในการตัดสินใจระหว่างยอมรับและปฏิเสธ (Accept และ Deny) ขึ้นอยู่กับผู้ดูแลจะจัดการ เช่น:

$$R_1\text{: } (f_1 \wedge f_2 \wedge f_3 \wedge \ldots \wedge f_{d1}) \Rightarrow accept$$
$$or\ R_1\text{: } (f_1 \wedge f_2 \wedge f_3 \wedge \ldots \wedge f_{d1}) \Rightarrow deny$$

รับ Pi เป็นแพ็คเก็ต IP บนฟิลด์ $f_1, \ldots, f_d$, P_i เป็น tuple ของ $d\ (p_1, p_2, \ldots, p_d)$ โดยที่แต่ละอัน p_i ($1 \leq i \leq d$) เป็นองค์ประกอบของ $D\ (f_i)$ IP แพ็คเก็ต $(p_1, p_2, \ldots, p_d)$ ตรงกับ R_i ถ้าหากเงื่อนไข $p_1 \in$

$f_1 \wedge p_2 \in f_2 \wedge \ldots \wedge p_d \in f_d$ ชุดของกฎ $(R_1, \ldots, R_i)$ จะใช้ได้เมื่อมีกฎอย่างน้อยหนึ่งกฎจับคู่กับ Pi เพื่อให้แน่ใจว่ากฎไฟร์วอลล์ทำงานอย่างถูกต้อง เงื่อนไขของกฎสุดท้ายในไฟร์วอลล์มักจะถูกระบุเป็น $f1 \in D(f_1) \wedge \ldots \wedge f_d \in D\ (f_d)$ โดยที่ทุกแพ็คเก็ตจะต้องจับคู่ตามที่แสดงใน R_3 ตัวอย่างของกฎสามข้อ เกี่ยวกับเงื่อนไขสามข้อ$C\ (f_1, f_2, f_3)$ โดยที่ $D\ (f_1) = D\ (f_2) \in [1, 100]$ และ$D\ (f_3) \in [1, 50]$

R_1: $f_1 \in [25{,}50 \wedge f_2 \in [40{,}60] \wedge f_3 \in [5{,}25] \Rightarrow$ accept
R_2: $f_1 \in [35{,}70 \wedge f_2 \in [30{,}90] \wedge f_3 \in [10{,}25] \Rightarrow$ accept
R_3: $f_1 \in [1{,}100 \wedge f_2 \in [1{,}100] \wedge f_3 \in [1{,}50] \Rightarrow$ deny

R_1 และ R_2 ซ้ำซ้อน (redundant) เนื่องจาก แพ็กเก็ตของกฎทั้งสองสามารถจับคู่กันได้และมีการกระทำ (action) เหมือนกัน คือ accept (ยอมรับ) นอกจากนี้ R_1 และ R_2 ยังขัดแย้งกับ R_3 เพราะทั้ง R_1 และ R_2 เป็นเซตย่อยของ R_1 ในขณะที่ R_3 มี action ที่ตา งจาก R_1 และ R_2
วิธีแก้ปัญหาในการแก้ไขความขัดแย้งดังกล่าว โดยทั่วไปแล้วไฟร์วอลล์จะเลือกกฎที่ตรงกับแพ็กเก็ตที่พิจารณาก่อน เรียกว่า วิธีจับคู่แรก ความผิดปกติของกฎไฟร์วอลล์สามารถแบ่งออกเป็น 6 ประเภท ให้ [x, y] แทนประเภทของความผิดปกติแต่ละทฤษฎีบท ดังนี้

1. Shadow anomaly

R_x ถูกบังโดย R_y ถ้าหากว่าจุดตัดของ R_x และ R_y มีค่าเท่ากับ Rx และมี actions ที่แตกต่างกัน

R_x: $C_x \Rightarrow A_x$

R_y: $C_y \Rightarrow A_y$

$$R_xR_y \in R_{db} \wedge \neg(A_x \Leftrightarrow A_y) \wedge (C_x \cap C_y = C_x) \quad (3)$$

โดยที่ R_{db} เป็นฐานข้อมูลของกฎทั้งหมดและ Ry เป็นกฎที่ดำเนินการก่อนหน้า Rx

2. Correlation Anomaly

R_x และ R_y ใน R_{db} มีความสัมพันธ์กันหากจุดตัดของ R_x และ R_y ไม่เท่ากับ $\emptyset, R_x - R_y \neq \emptyset, R_y - R_x \neq \emptyset$ และ R_x และ R_y มี actions ที่แตกต่างกัน

R_x: $C_x \Rightarrow A_x$

R_y: $C_y \Rightarrow A_y$

$$R_xR_y \in R_{db} \wedge \neg(A_x \Leftrightarrow A_y) \wedge (C_x \cap C_y = \emptyset) \wedge (C_x - C_y \neq \emptyset) \wedge (C_y - C_x \neq \emptyset) \quad (4)$$

3. Generalization Anomaly

R_x ถูก generalized โดย R_y ถ้าหากว่าจุดตัดของ R_x และ R_y มีค่าเท่ากับ R_y และมี actions โดยที่ R_y คือกฎที่จับคู่ก่อนหน้า R_x

$$R_x: C_x \Rightarrow A_x$$

$$R_y: C_y \Rightarrow A_y$$

$$R_x R_y \in R_{db} \wedge \neg(A_x \Leftrightarrow A_y) \wedge (C_x \cap C_y = C_y) \quad (5)$$

4. Redundancy Anomaly

R_x ซ้ำซ้อนกับ R_y ถ้าหากว่าจุดตัดของ R_x และ R_y ไม่เท่ากับ $\emptyset$ และมี actions เหมือนกัน ภาพประกอบที่ 3.1 **(d)**

$$R_x: C_x \Rightarrow A_x$$

$$R_y: C_y \Rightarrow A_y$$

$$R_x R_y \in R_{db} \wedge \neg(A_x \Leftrightarrow A_y) \wedge (C_x \cap C_y = \emptyset) \quad (6)$$

5. Irrelevance Anomaly

ความผิดปกติที่ไม่เกี่ยวข้องกันเกิดขึ้นในไฟร์วอลล์ หากไม่มีแพ็กเก็ตใดที่สามารถจับคู่กับกฎทั้งหมดได้ในไฟร์วอลล์ ความผิดปกตินี้เกิดจากความเข้าใจผิดเกี่ยวกับการเชื่อมต่อในเครือข่ายของผู้ดูแลระบบเอง

$$R_1: C_1 \implies A_1 B_2: C_2 \Rightarrow A_2$$

$$\vdots$$

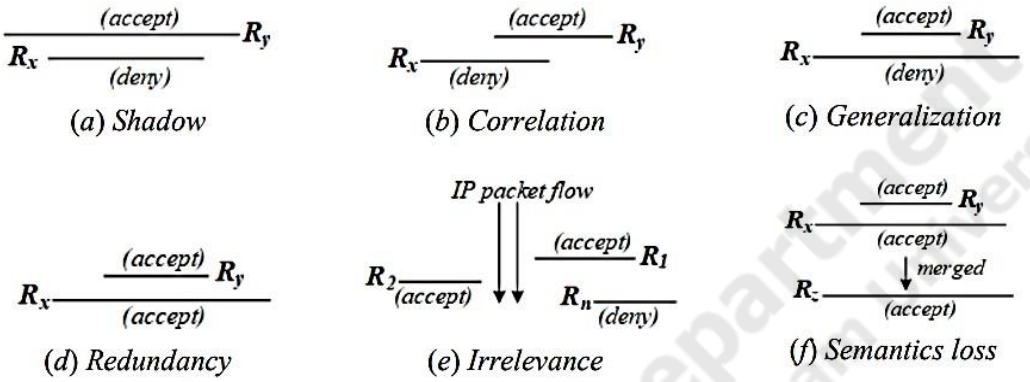
$$R_n: C_n \Rightarrow A_n$$

$$R_1, \dots, R_n \in R_{db} \wedge P_i \notin C_1, \dots C_n \quad (7)$$

โดยที่ P_i เป็นแพ็กเก็ต IP ที่ดำเนินการโดยไฟร์วอลล์

6. Semantics loss Anomaly

การสูญเสียความหมายของ Khummanee เกิดขึ้นจาก R_x และ R_y ถูกรวมเข้ากับ R_z โดยความหมายของกฎเก่าทั้งสองมีการเปลี่ยนแปลงหรือแทนที่ความหมายใหม่ ความผิดปกตินี้ส่วนใหญ่เกิดจากกฎซ้ำกัน ดังแสดงใน**รูปที่ 1**



ภาพประกอบที่ 3.1 ความผิดปกติของกฎไฟร์วอลล์

3.1.1 การปรับคุณสมบัติ Min-Max

การปรับคุณสมบัติ Min-Max (หรือเรียกว่าการปรับสภาพข้อมูล) เป็นวิธีมาตรฐานที่ใช้ในการปรับช่วงของข้อมูล เนื่องจากช่วงของค่าข้อมูลอาจแตกต่างกันมาก ดังนั้นจึงเป็นขั้นตอนที่จำเป็นในการประมวลผลข้อมูลล่วงหน้าก่อนประมวลผลในขั้นตอนถัดไป โดยปกติจะใช้เพื่อปรับขนาดช่วงข้อมูลใด ๆ ให้อยู่ในช่วง [0, 1] เรียกว่าการทำให้เป็น unity-based normalization

$$m' = \frac{m - r_{min}}{r_{max} - r_{min}} \; x(t_{max} - t_{min}) + t_{min} \qquad (8)$$

ให้ m' หมายถึง ค่าที่พิจารณาถูกทำให้เป็นมาตรฐานโดย $m \in [r_{min}\, r_{mz=ax}], r_{min}$ และ r_{max} แสดงถึงค่าต่ำสุดและสูงสุดของช่วงการวัด t_{min} และ t_{max} เป็นช่วงต่ำสุดและสูงสุดของช่วงเป้าหมายที่จะถูกปรับสัดส่วน

3.1.2 ทฤษฎีของเบย์

ทฤษฎีบทของเบย์ (หรือเรียกอีกอย่างว่ากฎเบย์) เป็นสูตรที่อธิบายถึงวิธีการอัพเดทความน่าจะเป็นของสมมติฐานเมื่อได้รับหลักฐาน มันเป็นเครื่องมือที่มีประโยชน์สำหรับการคำนวณความน่าจะเป็นตามเงื่อนไข ทฤษฎีบทของเบย์สามารถนิยามได้ดังต่อไปนี้:

ให้ $A1 \quad 2, A, \dots, A_k$ เป็นเหตุการณ์ที่แบ่งพื้นที่ตัวอย่าง $s, i.e, S = A_1 \cup A_2 \cup \dots \cup A_k$ และ $A_i \cap A_j = \emptyset$ เมื่อ i ≠ j และให้ B และเหตุการณ์ในพื้นที่นั้น $(B) > 0$ จากนั้นทฤษฎีเบย์ คือ :

$$P_r(A_i|B) = \frac{P_r(A_j)P_r(B|A_j)}{\sum_{j=1}^{k} P_1(A_j)P_r(B|A_j)} \tag{9}$$

สูตรนี้สามารถใช้เพื่อย้อนกลับความน่าจะเป็นตามเงื่อนไข หากเราทราบความน่าจะเป็นของเหตุการณ์ A_j และความน่าจะเป็นตามเงื่อนไข $\mathrm{P}_r(B|A_j), j = 1, \dots, k$, สูตรสามารถใช้ในการคำนวณความน่าจะเป็นตามเงื่อนไข $\mathrm{P}_r(A_j \mid B)$

3.1.3 Moving Average (MA)

ค่าเฉลี่ยเคลื่อนที่ (MA) [] เป็นตัวบ่งชี้ที่ใช้กันอย่างแพร่หลายสำหรับการวิเคราะห์แนวโน้มข้อมูล ช่วยให้การดำเนินการข้อมูลราบรื่นขึ้นโดยกรองการรบกวนจากความผันผวนของข้อมูลระยะสั้น ค่าเฉลี่ยเคลื่อนที่มีอยู่ 2 ประเภทซึ่งเป็นที่นิยมและใช้กันอย่างแพร่หลายคือ Simple Moving Average (SMA) และ Exponential Moving Average (EMA) **SMA** คำนวณค่าเฉลี่ยของข้อมูลล่าสุด เมื่อ n แสดงถึงจำนวนของช่วงเวลาที่เรา ต้องการเค่าเฉลี่ย:

$$SMA = \frac{A_1+A_2+A_3+\dots+A_n}{n} \tag{10}$$

โดยที่ A คือ ค่าเฉลี่ยในช่วงเวลา n และ n คือ จำนวนของช่วงเวลา

EMA เป็นค่าเฉลี่ยถ่วงน้ำหนักของข้อมูลล่าสุด โดยที่น้ำหนักของข้อมูล ที่ซึ่งค่าน้ำหนักเกิดขึ้นแบบ exponential กับช่วงของข้อมูล. เมื่อกล่าวอีกนัยหนึ่ง มันเป็นสูตรที่ใช้หาค่าน้ำหนักข้อมูล แบบ exponential นั่นเอง ซึ่งมีสมการดังนี้

$$EMA_t = \left[V_t \times \left(\frac{s}{1+d}\right)\right] + EMA_y \times \left[1 - \left(\frac{s}{1+d}\right)\right] \tag{11}$$

โดยที่: EMA_t = EMA วันนี้, V_t = ข้อมูลปัจจุบัน, EMA_y = EMA เมื่อวาน, S = ความราบเรียบของ ข้อมูล, d คือ จำนวนวัน

3.1.4 การแปลงที่อยู่ IP ให้เป็นจำนวนเต็มไม่ติดลบ

ที่อยู่อินเทอร์เน็ตโพรโทคอล (ที่รู้จักกันในชื่อที่อยู่ IP) เป็นที่อยู่เฉพาะที่อุปกรณ์เครือข่าย เช่น เราเตอร์, สวิตช์ และ คอมพิวเตอร์ ใช้เพื่อระบุตัวเองและสื่อสารผ่านอุปกรณ์อื่น ๆ ในเครือข่ายคอมพิวเตอร์ ที่อยู่ IPv4 (IP รุ่น 4) มีค่าเท่ากับ 32 บิตตั้งแต่ 0 ถึง $2^{32} - 1$ โดยปกติจะแบ่งออกเป็น 4 ส่วนแต่ละส่วน (8 บิต = ออคเต็ต) คั่นด้วยจุดเช่น $A_1.\ A_2.\ A_3.\ A_4$ โดยที่ $A1-4 \in [0, 255]$ ที่อยู่ IPv4 สามารถแปลงเป็น จำนวนเต็มใด ๆ ที่ไม่ตดิลบด้วยสมการต่อไปนี้:

$$\text{IPv4}' = (A_1 \times 2^{24}) + (A_2 \times 2^{16}) + (A_3 \times 2^{8}) + (A_4 \times 2^{0}) \quad (12)$$

โดยที่ IPv4′ เป็นที่อยู่ IP ใหม่ ที่จะแปลง ตัวอย่างเช่น 1.2.3.4 จะถูกแปลงเป็น:

$$\text{IPv4}' = (1 \times 2^{24}) + (2 \times 2^{16}) + (3 \times 2^{8}) + (4 \times 2^{0}) = 16{,}909{,}060$$

3.2 จุดเด่นของโครงงาน

เนื่องจากความผิดปกติของกฎเกิดขึ้นผ่านไฟร์วอลล์ อำนาจการตัดสินใจในการแก้ไขความผิดปกติส่วนใหญ่ขึ้นอยู่กับดุลยพินิจของผู้ดูแลระบบ อย่างไรก็ตามการตัดสินใจที่เกิดขึ้นมักจะส่งผลให้เกิดข้อผิดพลาดหรือช่องโหว่มากกว่ากฎที่มีอยู่หากผู้ดูแลระบบไม่สามารถเข้าใจความสัมพันธ์ระหว่างกฎความขัดแย้งทั้งหมด ดังนั้นจึงจำเป็นต้องพัฒนาระบบสนับสนุนการตัดสินใจ สำหรับผู้ดูแลระบบเพื่อช่วยในการตัดสินใจในกรณีที่มีความผิดปกติเกิดขึ้น ระบบประกอบด้วย 4 ขั้นตอน:

1) ประการแรกเตรียมข้อมูลต่าง ๆ ให้พร้อมก่อนดำเนินการ
2) การวิเคราะห์และตรวจสอบความผิดปกติของกฎโดย Path Selection Tree (PST)
3) การคำนวณความน่าจะเป็น (Bayesian) ของแต่ละกฎโดยพิจารณาจากความถี่ของแพ็ตเก็ตที่จับคู่ กับกฎหลักฐานของการสร้างกฎ, ความเชี่ยวชาญในการสร้างกฎ และลำดับความสำคัญของ โปรโตคอลเพื่อช่วยผู้ดูแลระบบตัดสินใจก่อนปรับกฎ

4) สุดท้ายปรับความผิดปกติหรือกฎความขัดแย้งตามความน่าจะเป็น

3.3 การออกแบบระบบ

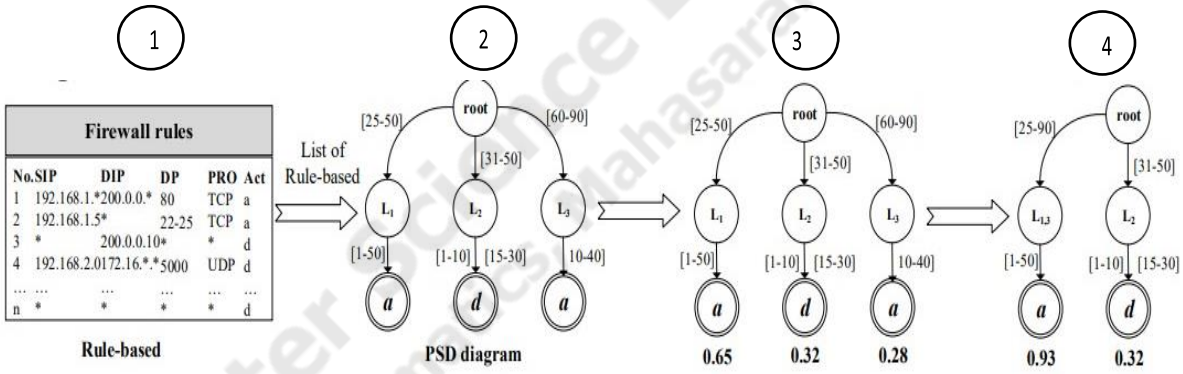
การออกแบบระบบมี 4 ขั้นตอน

ขั้นตอนที่ 1 การเตรียมกฎไฟร์วอลล์

ขั้นตอนที่ 2 วิเคราะห์ความผิดปกติ

ขั้นตอนที่ 3 คำนวณความน่าจะเป็น

ขั้นตอนที่ 4 การปรับปรุงกฎ



ภาพประกอบที่ 3.2 ภาพรวมของการออกแบบระบบ

3.3.1 การเตรียมไฟร์วอลล์ตามกฎ (ขั้นตอนที่ 1)

A. เงื่อนไข (C_i) และการตัดสินใจ (A_i) ของแต่ละกฎ

อ้างถึง R_i ในสมการ **(2)** โดยทั่วไปสมาชิกของ C_i มี 5 ฟิลด์ ($f_1 \wedge \ldots \wedge f_5$) โดยที่ f_1 = ที่อยู่ IP ต้นทาง (SIP), f_2 = ที่อยู่ IP ปลายทาง (DIP), f_3 = พอร์ตต้นทาง (SP), f_4 = พอร์ตปลายทาง (DP) และ f_5 = โปรโตคอล (PRO) ตามลำดับดังแสดงในตารางที่ 1

ตารางที่ 3.1 ฟิลด์สมาชิกพื้นฐานของ Ci และ Ai

R_i	$C_i(f_1 \wedge f_2 \ldots \wedge f_5)$					A_i
	f_1 (SIP)	f_2 (DIP)	f_3 (SP)	f_4 (DP)	f_5 (PRO)	accept/deny
R_1	1.2.3.0 - 1.2.3.10	0.0.0.1 - 0.0.1.0	*	80	TCP/UDP	accept

ตามที่ R_1 ของตารางที่ 1 กระบวนการจัดทำกฎไฟร์วอลล์ เริ่มต้นด้วยการแปลงที่อยู่ IP ของ f_1 และ f_2 เป็นช่วงของจำนวนเต็มบวกโดยสมการ (12) ดังนั้น f_1 และ f_2 จะถูกแปลงเป็นตัวเลขต่อไปนี้: $f_1 \in [16909056,16909056$ และ $f_2 \in [1,256]$ ข้อมูลที่แปลงในลำดับถัดไปคือ f_3 และ f_4 ซึ่งมีตัวเลขตั้งแต่ 0 ถึง $2^{16} - 1$: $f_3 \in [0,65535]$ และ $f_4 \in [80,80]$ โดยที่ * หมายถึงตัวเลขทั้งหมดในโดเมนดังกล่าว ฟิลด์ f_5 เป็นทั้งโปรโตคอล TCP และ UDP ดังนั้นจึงถูกแปลงเป็น: $f_5 \in [6,17]$, โดยที่ TCP = 6 และ UDP = 17 ในกรณีของฟิลด์การตัดสินใจ (A_i) มันจะถูกเปลี่ยนเป็นจำนวนเต็มบวกทั้ง 0 หรือ 1 เช่น $A_i \in \{0,1\}$ โดยที่ยอมรับ = 1 และ ปฏิเสธ = 0 จากการคำนวณทั้งหมดเหล่านี้ R_1 จะถูกแปลงเป็น:

$$R_1: (f_1 \in [16909056, 16909066] \wedge f_2 \in [16909056,16909066] \wedge f_3 \in [0,65535] \wedge f_4 \in [80,80] \wedge f_5 \in \{6,17\}) \Rightarrow 1$$

B. การคำนวณความน่าจะเป็นของแต่ละกฎข้อมูลพิเศษ

ในการคำนวณความน่าจะเป็นของแต่ละกฎในตัวแบบนี้ จะมีการเพิ่ม 4 ฟิลด์เพิ่มเติมรวมถึง ความถี่ของแพ็คเก็ตที่ตรงกับกฎ (FPM), หลักฐานการสร้างกฎ (ECR), ความเชี่ยวชาญของผู้สร้างกฎ (ERC), และลำดับความสำคัญของโปรโตคอล (PRI) ให้ P(e_1), P(e_2), P(e_3) และ P(e_4) คือความ น่าจะเป็นของ FPM, ECR, ERC และ PRI ตามลำดับ ดังนั้นผลรวมของความน่าจะเป็นของกฎ R_i เท่ากับ สมการ (13)

$$P(E_i) = \sum_{j=1}^{4} P(e_j) \quad (13)$$

โดยที่ $P(E_i)$ คือความน่าจะเป็นของ R_i ตัวอย่างเช่น ข้อมูลพิเศษของ R_i ดังแสดงในตารางที่ 2

ตารางที่ 3.2 ข้อมูลพิเศษของ Ri

R_i	$E_i = (e_1 + \cdots + e_4)$			
	$e_1(FPM)$	$e_2(ECR)$	$e_3(ERC)$	$e_4(PRI)$
R_1	2125	3	2	4
R_1'	$e_1' = 0.24$	$e_2' = 1.0$	$e_3' = 0.666$	$e_4' = 0.625$

จาก R_1 ในตารางที่ 2 อัตราการจับคู่ (e_1) ระหว่างแพ็คเก็ตและ R_1 เท่ากับ 2125 ครั้ง e_2, e_3 และ e_4 คือ 3, 2 และ 4 ตามลำดับอธิบายรายละเอียดเพิ่มเติมในส่วนถัดไป ข้อมูลพิเศษทั้ง 4 นี้ถูกคำนวณเป็นรูปแบบ ความน่าจะเป็นที่ข้อมูลอยู่ในช่วงตั้งแต่ 0.0 ถึง 1.0 โดยการปรับขนาดคุณลักษณะ Min-Max ในสมการ (8) **ในกรณี e_1**: มันเป็นความถี่ของแพ็คเก็ตที่ตรงกับกฎใด ๆ ของไฟร์วอลล์กระบวนการเริ่มนับจากเวลาที่กฎถูก สร้างขึ้นมาจนถึงปัจจุบัน ตัวอย่างเช่นหากจำนวนการจับคู่สูงสุดและต่ำสุดของกฎใด ๆ ในไฟร์วอลล์คือ 5,000 และ 1200 ตามลำดับดังนั้น e_1' ในที่นี้จะเท่ากับ:

$$e_1' = \frac{m - r_{min}}{r_{max} - r_{min}} \times (t_{max} - t_{min}) + t_{min} = \frac{2125 - 1200}{5000 - 1200} \times (1.0 - 0.0) + 0.0 \quad = 0.423$$

โดยที่ m = 2125, r_{min} = 1200, r_{max} = 5,000, t_{min} = 0.0 และ t_{max} เท่ากับ 1.0

อย่างไรก็ตามการบันทึก e_1 ในไฟร์วอลล์ต้องใช้สมการ (10 และ 11) เพื่อทำให้ข้อมูลราบรื่นขึ้นเนื่องจากข้อมูลที่บันทึกไว้อาจเป็นข้อมูลการแกว่งที่เกิดจากการโจมตีเครือข่าย พฤติกรรมผู้ใช้ หรือการใช้เครือข่ายในช่วงชั่วโมงเร่งด่วน เป็นต้น ช่วงเวลาสำหรับการคำนวณข้อมูลด้วยวิธี EMA จะขึ้นอยู่กับความเหมาะสมของแต่ละองค์กร สำหรับการวิจัยนี้ e_1 จะถูกบันทึกทุกชั่วโมงต่อวันดังตัวอย่างต่อไปนี้:

รับ e_1 จาก R_1 ในแต่ละชั่วโมงต่อวัน: 1300, 1500, 1200, 1300, 1400, 1500, 1800, 4500, 6000,6300, 5500, 1,000, 2400, 2800, 2600, 2600, 2400, 1900, 1500, 1200 , 1,000, 800, 700, 600 ครั้งจากนั้นสามารถคำนวณ EMA ของ e_1 โดยใช้ห้าชั่วโมงแรกดังต่อไปนี้

$$SMA\ of\ 5^{th}hour = 1340.00 + \frac{(1300+1500+1200+1300+1400)}{5} = 1340$$

$$SMA\ of\ 6^{th}hour = 1340.00 + \frac{2}{(5+1)}(1500 - 1340.00) = 1840.00$$

$$SMA\ of\ 7^{th}hour = 1840.00 + \frac{2}{(5+1)}(1500 - 1840.00) = 1940.00$$

$$SMA\ of\ 8^{th}hour = 1940.00 + \frac{2}{(5+1)}(4500 - 1940.00) = 2840.00$$

$$SMA\ of\ 9^{th}hour = 2840.00 + \frac{2}{(5+1)}(6000 - 2840.00) = 3340.00$$

$$SMA\ of\ 10^{th}hour = 3340.00 + \frac{2}{(5+1)}(6300 - 3340.00) = 3440.00$$

$$SMA\ of\ 11^{th}hour = 3440.00 + \frac{2}{(5+1)}(5500 - 3340.00) = 3173.00$$

$$SMA\ of\ 12^{th}hour = 3173.00 + \frac{2}{(5+1)}(1000 - 3340.00) = 1673.00$$

$$SMA\ of\ 13^{th}hour = 1673.00 + \frac{2}{(5+1)}(2400 - 3340.00) = 2140.00$$

$$SMA\ of\ 14^{th}hour = 2140.00 + \frac{2}{(5+1)}(2800 - 3340.00) = 2273.00$$

$$SMA\ of\ 15^{th}hour = 2273.00 + \frac{2}{(5+1)}(2600 - 3340.00) = 2206.00$$

$$SMA\ of\ 16^{th}hour = 2206.00 + \frac{2}{(5+1)}(2600 - 3340.00) = 2206.00$$

$$SMA\ of\ 17^{th}hour = 2206.00 + \frac{2}{(5+1)}(2400 - 3340.00) = 2140.00$$

$$SMA\ of\ 18^{th}hour = 2140.00 + \frac{2}{(5+1)}(1900 - 3340.00) = 1973.00$$

$$SMA\ of\ 19^{th}hour = 1973.00 + \frac{2}{(5+1)}(1500 - 3340.00) = 1840.00$$

$$SMA\ of\ 20^{th}hour = 1740.00 + \frac{2}{(5+1)}(1200 - 3340.00) = 1740.00$$

$$SMA\ of\ 21^{th}hour = 1740.00 + \frac{2}{(5+1)}(1000 - 3340.00) = 1673.00$$

$$SMA\ of\ 22^{th}hour = 1673.00 + \frac{2}{(5+1)}(800 - 3340.00) = 1606.00$$

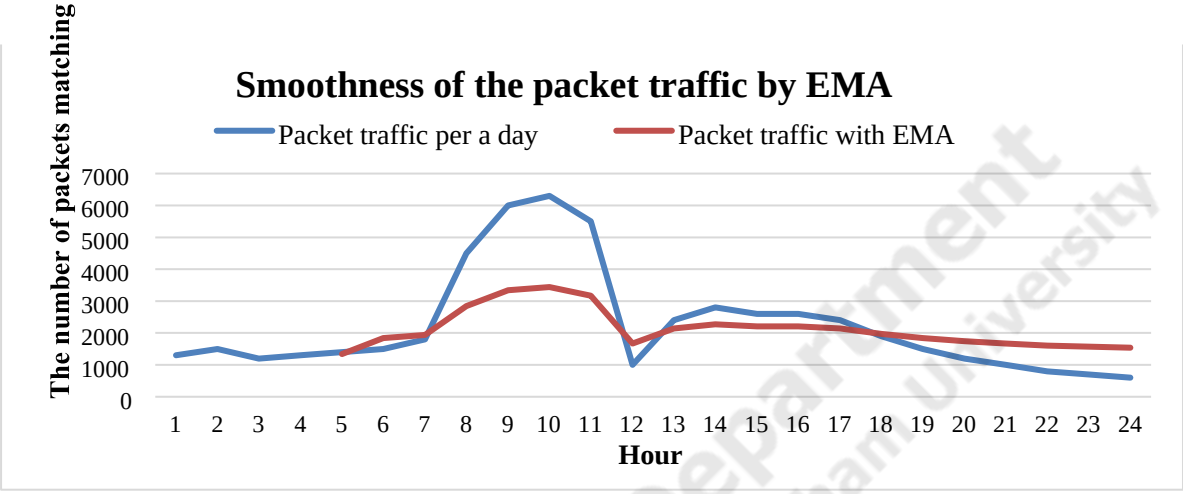
$$SMA\ of\ 23^{th}hour = 1606.00 + \frac{2}{(5+1)}(700 - 3340.00) = 1573.00$$

$$SMA\ of\ 24^{th}hour = 1573.00 + \frac{2}{(5+1)}(600 - 1573.00) = 1540.00$$

การคำนวณ EMA เป็นเวลา 24 ชั่วโมง ดังแสดงในรูปที่ 3 ผลลัพธ์จะถูกคำนวณด้วย SMA อีกครั้ง เพื่อหาค่าเฉลี่ยของแต่ละวันซึ่งจะคำนวณค่าที่คำนวณได้ใน recorded1 ในตารางที่ 2 เพื่อคำนวณค่าเฉลี่ยของ e_1 ของแต่ละวันดังนี้:

$$e_1 = \frac{\sum_{i=5}^{n} EMA_i}{n} = \frac{(1340.00+1840.00+1940.00+\cdots+1540.00)}{20} = 2125.00$$

โดยที่ n คือจำนวนชั่วโมง/วัน ลบด้วยจำนวนชั่วโมงที่ใช้ในอดีต



ภาพประกอบที่ 3.3 การปรับทราฟฟิกแพ็กเก็ตให้ราบรื่นยิ่งขึ้นด้วย EMA

ในกรณี e_2: หมายถึง เอกสารหรือกระดาษที่ใช้เพื่อยืนยันว่ากฎดังกล่าวได้รับการอนุมัติ ในบทความนี้หลักฐานการสร้างกฎถูกแบ่งออกออกเป็น 4 ระดับ:

1. ไม่มีหลักฐานการอนุมัติ
2. ผู้ดูแลระบบไฟร์วอลล์เป็นผู้อนุมัติ
3. หัวหน้าแผนกเป็นผู้อนุมัติ
4. เจ้าขององค์กรเป็นผู้อนุมัติ

โดยการหาน้ำหนักของหลักฐานตามลำดับความสำคัญของผู้อนุมัติเอกสาร มีดังนี้

ไม่มีหลักฐาน = 0

ผู้ดูแล = 1

หัวหน้าแผนก = 2

เจ้าขององค์กร = 3

ตามลำดับ หากน้ำหนักของเอกสารที่ได้รับถูกคำนวณโดยสมการ Min-Max (8) ผลลัพธ์จะเป็น e_2 ให้เจ้าขององค์กรได้รับการอนุมัติให้สร้างกฎ R_1 ผลลัพธ์ของการคำนวณเท่ากับ:

$$e_2' = \frac{m - r_{min}}{r_{max} - r_{min}} \times (t_{max} - t_{min}) + t_{min} = \frac{3 - 0}{3 - 0} \times (1.0 - 0.0) + 0.0 = 1.0$$

โดยที่ m = 3, r_{min} = 0, r_{max} = 3, t_{min} = 0.0 และ t_{max} เท่ากับ 1.0

ในกรณี e_3: ในกรณีของหลักฐานความเชี่ยวชาญในการสร้างกฎนั้นแบ่งออกเป็น 4 ระดับ:

1. ผู้ดูแลระบบมือใหม่
2. ผู้ดูแลที่มีความเชี่ยวชาญเพียงพอ
3. ผู้ดูแลมืออาชีพ
4. ผู้ดูแลระบบที่เชี่ยวชาญมาก

ผู้ดูแลระบบมือใหม่ หมายถึง ผู้ที่เพิ่งได้รับมอบหมายให้กำหนดค่าระบบไฟร์วอลล์ด้วยประสบการณ์น้อยที่สุด เมื่อผู้ดูแลระบบได้กำหนดค่าไฟร์วอลล์สักระยะหนึ่งพวกเขาจะมีความเชี่ยวชาญมากกว่าซึ่งควรมีเวลาทำงานอย่างน้อย 3 - 5 ปีซึ่งเรียกว่าผู้ดูแลระบบปกติสำหรับผู้ที่มีประสบการณ์ในการฝึกอบรมหรือปรับแต่งไฟร์วอลล์เป็นจำนวนมากด้วยเวลาทำงาน 5 - 10 ปีพวกเขาจะเป็นผู้ดูแลระบบมืออาชีพ สำหรับผู้ที่ได้รับการฝึกฝนมากมายและมีใบรับรองเกี่ยวกับไฟร์วอลล์พวกเขาจะเป็นผู้ดูแลระบบที่เชี่ยวชาญ ในบทความนี้กำหนดน้ำหนักของความเชี่ยวชาญดังต่อไปนี้ e_3:

ผู้ดูแลระบบมือใหม่ = 0

ผู้ดูแลระบบปกติ = 1

ผู้ดูแลระบบมืออาชีพ = 2 ผู้ดูแลระบบความเชี่ยวชาญมาก = 3

ให้ผู้ดูแลระบบมืออาชีพสร้างกฎ R_1 ผลลัพธ์จะถูกคำนวณดังนี้:

$$e_2' = \frac{2 - 0}{3 - 0} \times (1.0 - 0.0) + 0.0 = 0.666$$

ในกรณี e_4 : โปรโตคอลที่สื่อสารบนเครือข่ายคอมพิวเตอร์มักจะให้ความสำคัญเช่นการประชุมทางวิดีโอจะต้องราบรื่นในระหว่างการประชุมตลอดเวลา ในทางกลับกันการส่งจดหมายอิเล็กทรอนิกส์นั้นไม่จำเป็นต้องส่งและรับทันที การจัดลำดับความสำคัญของโปรโตคอลสามารถทำได้ขึ้นอยู่กับนโยบายของแต่ละองค์กร ในการวิจัยนี้การจัดลำดับความสำคัญของโปรโตคอลขึ้นอยู่กับลำดับความสำคัญจาก 3 GPP QoS Class Identification [xxx] โดย IP Multimedia มีความสำคัญสูงสุด (1 = สูงสุด) และ Chat, FTP และ P2P มีลำดับ ความสำคัญต่ำสุด (9 = ต่ำสุด) จาก e_4 ในตารางที่ 2 เป็นแอปพลิเคชันการประชุมทางไกลซึ่งเป็นความสำคัญ ลำดับที่ 4 เมื่อประมวลผลในรูปแบบของความน่าจะเป็นโดยใช้ Min-Max Scaling ผลลพันธ์จะเท่ากับ:

$$e_2' = \frac{6-1}{9-1} \times (1.0 - 0.0) + 0.0 = .625$$

โดยที่ m = 4, r_{min} = 1, r_{max} = 9, t_{min} = 0.0 และ t_{max} = 1.0 สังเกตว่าลำดับความสำคัญของโปรโตคอลที่คำนวณจะต้องกลับลำดับความสำคัญเช่นจาก 9 เป็น 1 และจาก 1 เป็น 9 ตัวอย่างเช่นลำดับความสำคัญของ 4 จะกลับเป็น 6

สุดท้าย ในภาพประกอบที่ 3 แสดงถึงตัวอย่างของกฎไฟร์วอลล์ที่ประกอบด้วยความผิดปกติทั้งหมดที่กล่าวถึงก่อนหน้านี้ และกฎเหล่านี้จะถูกดำเนินการในขั้นตอนต่อไป

$R_1: f_1 \in [1,100] \wedge f_2 \in [1,100] \wedge f_3 \in [0,65535] \wedge f_4 \in [80,85] \wedge f_5 \in [6,17] \Rightarrow 1$
$R_2: f_1 \in [10,50] \wedge f_2 \in [20,60] \wedge f_3 \in [0,65535] \wedge f_4 \in [80,80] \wedge f_5 \in [6,17] \Rightarrow 0$
$R_3: f_1 \in [20,40] \wedge f_2 \in [30,70] \wedge f_3 \in [0,65535] \wedge f_4 \in [80,90] \wedge f_5 \in [6,17] \Rightarrow 0$
$R_4: f_1 \in [20,30] \wedge f_2 \in [20,30] \wedge f_3 \in [0,65535] \wedge f_4 \in [80,82] \wedge f_5 \in [6,17] \Rightarrow 0$
$R_5: f_1 \in [1,100] \wedge f_2 \in [1,100] \wedge f_3 \in [0,65535] \wedge f_4 \in [30,90] \wedge f_5 \in [6,17] \Rightarrow 1$
$R_6: f1 \in [1,100] \wedge f_2 \in [1,100] \wedge f_3 \in [0,65535] \wedge f_4 \in [0,65535] \wedge f_5 \in [6,17] \Rightarrow 0$

ข้อมูลพิเศษของแต่ละกฎเมื่อผ่านกระบวนการจัดเตรียมข้อมูลจะสร้างผลลัพธ์ต่อไปนี้ ($R_i \longrightarrow R_i'$):

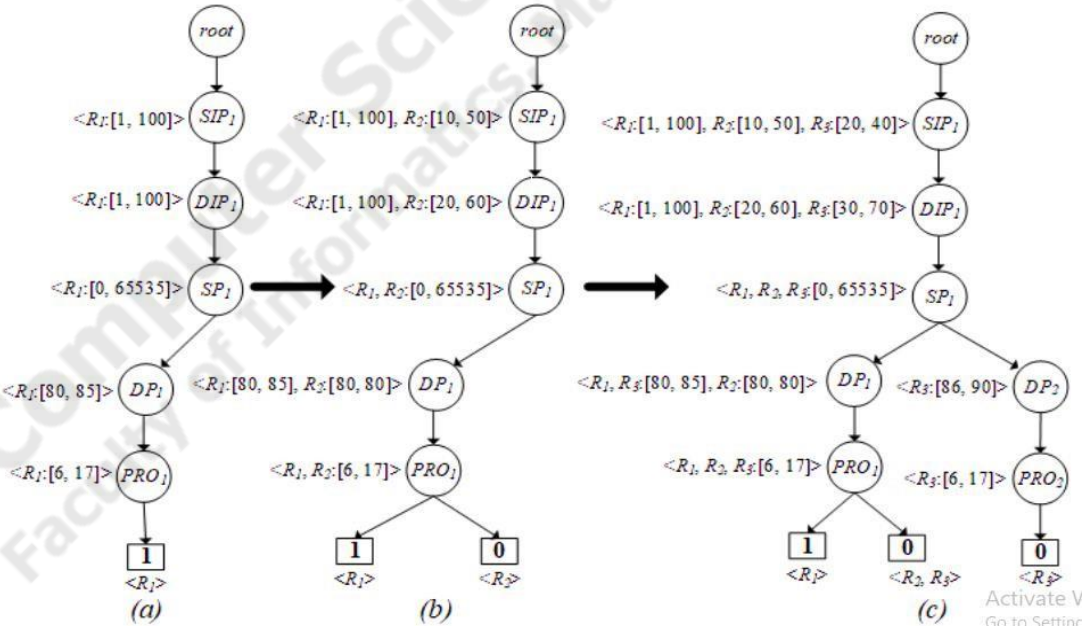
$R_1: e_1 = 2500, e_2 = 1, e_3 = 2, e_4 = 6 \rightarrow R_1': e_1' = 0.26, e_2' = 0.33, e_3' = 0.67, e_4' = 0.67$
$R_2: e_1 = 1500, e_2 = 3, e_3 = 3, e_4 = 3 \rightarrow R_1': e_1' = 0.06, e_2' = 1.00, e_3' = 1.00, e_4' = 0.33$
$R_3: e_1 = 2000, e_2 = 2, e_3 = 1, e_4 = 8 \rightarrow R_1': e_1' = 0.16, e_2' = 0.67, e_3' = 0.33, e_4' = 0.33$
$R_4: e_1 = 3200, e_2 = 1, e_3 = 2, e_4 = 5 \rightarrow R_1': e_1' = 0.40, e_2' = 0.33, e_3' = 0.67, e_4' = 0.56$

$R_5: e_1 = 1200, e_2 = 0, e_3 = 3, e_4 = 2 \rightarrow R_1': e_1' = 0.00, e_2' = 1.00, e_3' = 0.00, e_4' = 0.22$

$R_6: e_1 = 5000, e_2 = 0, e_3 = 3, e_4 = 9 \rightarrow R_1': e_1' = 0.76, e_2' = 0.00, e_3' = 1.00, e_4' = 1.00$

3.3.2 การวิเคราะห์และตรวจจับความผิดปกติ (ขั้นตอนที่ 2)

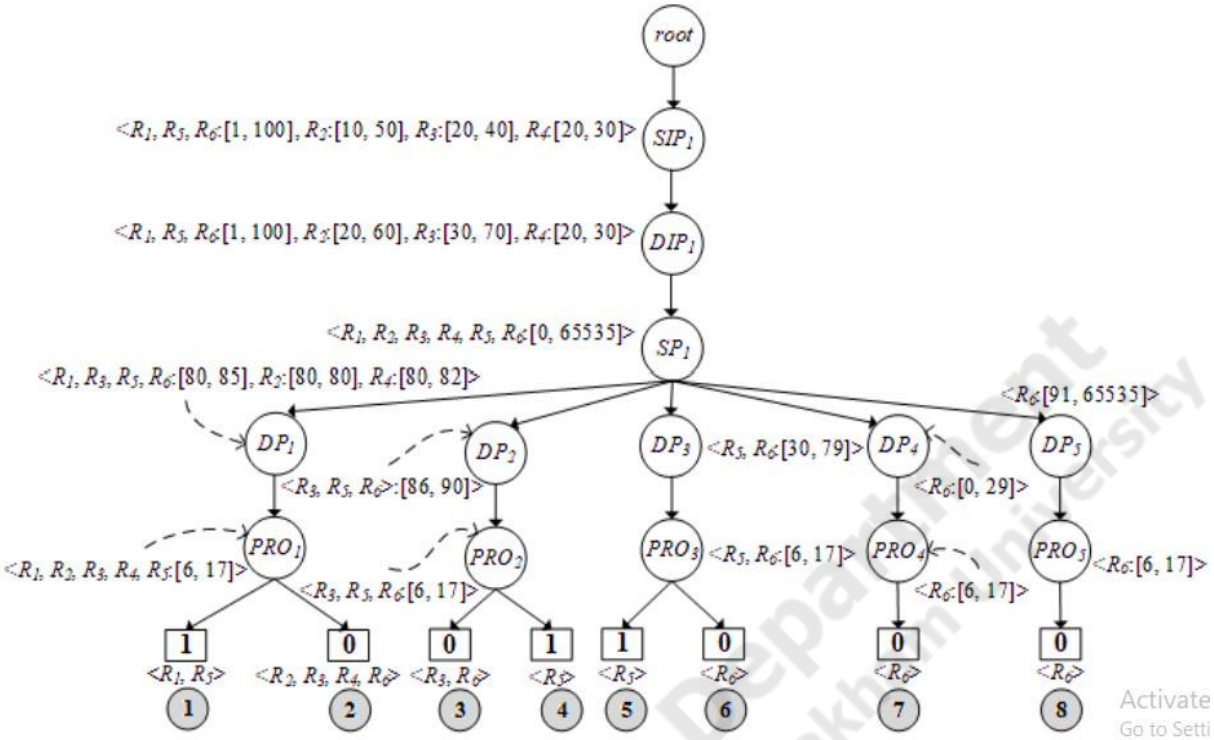
ในขั้นตอนนี้กฎจากขั้นตอนที่ 1 จะใช้ในการสร้างโครงสร้างต้นไม้ที่เรียกว่า เส้นทางการเลือกต้นไม้ (PST) เพื่อวิเคราะห์ความผิดปกติ อัลกอริทึมเริ่มต้นด้วยการสร้างรูดโหนดของ PST หลังจากนั้นฟิลด์ f_1 ของกฎแรกจะถูกสร้างขึ้นเป็นโหนดแรกบนทรี คือ SIP1 ดังแสดงในรูปที่ 4 (a) ในโหนดนี้จะบันทึก IP ต้นทาง (SIP) ของ R_1 ให้เป็น $< R_1$: [1, 100] > โดยที่ [1, 100] = f_1 โหนดถัดไป (DIP1) จะจัดเก็บช่วงของ IP ปลายทาง (DIP) (f_2) ของ R_1 ตั้งแต่ 1 ถึง 100 ถัดไปเป็นโหนดที่บันทึก พอร์ตต้นทาง (SP) ของช่วงตั้งแต่ 0 ถึง 65535 เรียกว่า (SP1) โหนดถัดไปเป็น DP1 โหนดนี้มีกลุ่มของ พอร์ตปลายทาง (DP) f_4 ระหว่าง 80 และ 85 ($< R_1$: [80, 85] >) ฟิลด์สุดท้าย f_5 ของ R_1 เป็น PRO1 ซึ่งเก็บช่วงของโปรโตคอล TCP และ UDP ในการตัดสินตามกล่องสี่เหลี่ยมด้านล่างในทรีประกอบด้วยการตัดสินใจยอมรับ (1) ของ R_1 ในตอนท้ายของต้นไม้มัน Record สิ่งที่กฎเป็นสมาชิกของเส้นทางนี้เช่น $< R_1 >$



ภาพประกอบที่ 3.4 การสร้างกฎ R_1(a), R_2(b) และ R_3(c) ลงใน PST

ในลำดับถัดไปกฎข้อที่สอง R_2 จะถูกนำเข้าสู่ PST ดังแสดงในรูปที่ 4 (b) ขั้นตอนแรก f_1 ของกฎ $R_2 \subset f_1$ ของ

R_1,ดังนั้น กฎฟิล์ดที่ 1 ของ R_2 นี้ (f_1) ใช้เส้นทางเดียวกับ $R_1(f_1)$ และยังบันทึก < R_2: [10, 50] > ลงในโหนด SIP1 เช่นเดียวกัน $R_2(f_1) \subset R_1(f_2)$ จะถูกบันทึกไปยังโหนดเดียวกัน ($DIP1 = < R_1: [1, 100], R_2: [20, 60] >$) และเดินทางข้ามแบบเดียวกันกับ R_1 คล้ายกับ $R_2(f_3)$ ซึ่งเท่ากับ $R_1(f_3)$ ดังนั้น $R_2(f_3)$ จึงถูกต่อท้ายในโหนด SP1 เป็น $< R_1, R_2: [0, 65536] >$ ในกรณีของ $R_2(f_4)$ และ $R_1(f_4), R_2(f_4)$ เป็นชุดย่อยของ $R_1(f_4)$ ดังนั้นข้อมูลของ DP1 จึงถูกอัปเดตเป็น $< R_1: [80, 85], R_2: [80, 80] >$ และ PRO1 ได้รับการอัปเดตเป็น $< R_1, R_2: [6, 17] >$ ในทางกลับกันการตัดสินใจของ R_1 และ R_2 ไม่ เหมือนกัน ดังนั้นเส้นทางการตัดสินใจจะต้องแยกออกจากกันโดยที่ $< R_1 > = 1, < R_2 > = 0$ สำหรับ การเพิ่มกฎ $R_3(c)$ ลงใน PST คล้ายกับการใส่กฎ R_2 ซึ่งมีความแตกต่างกันเพียงเล็กน้อย คือ ในตำแหน่งระดับโพรโทคอลในทรี เนื่องจาก $R_3(f_4)$ เป็น Superset ของ $R_1(f_4)$ และ $R_2(f_4)$ พอร์ตปลายทางบางแห่งของ $R_3(f_4)$ จะต้องแยกออกเป็นโหนดอื่นของต้นไม้ คือ DP2 ซึ่งเก็บพอร์ตปลายทางตั้งแต่ 86 ถึง 90 ($R_3(f_4) - R_1(f_4)$) เช่น $< R_3: [86, 90] >$ พอร์ตปลายทางที่เหลืออยู่รวมกับ DIP1 ใน เส้นทางแรกพร้อมกับ R_1 และ R_2 เป็น $< R_1, R_3: [80, 85], R_2: [80, 80] >$ การตัดสินใจของ R_3 ไม่ได้รับอนุญาตในทั้งสองเส้นทางโดยที่ $< R_3 > = 0$ กฎไฟร์วอลล์ที่เหลืออยู่ (R_4, R_5, R_6) จะถูกเรียกใช้งาน เช่นกฎก่อนหน้า (R_1, R_2, R_3) หากกฎทั้งหมดได้รับการดำเนินการอย่างประสบความสำเร็จผ่าน PST ผลลัพธ์ที่แสดงจะแสดงในรูปที่ 5



ภาพประกอบที่ 3.5 โครงสร้าง PST สมบูรณ์หลังจากรวบรวมกฎทั้งหมด

ในกระบวนการตรวจสอบความผิดปกติของกฎอัลกอริทึมจะใช้ข้อมูลที่บันทึกไว้ในแต่ละโหนดเพื่อตรวจจับความผิดปกติโดยใช้ผลคูณคาร์ทีเซียนของโหนดทั้งหมดที่แยกจากชั้นโปรโตคอล (PRO_i) และมองกลับจากด้านล่างถึงรูดดังนี้

กลุ่ม 1: เส้นทางหมายเลข ① และ ② ภายใต้โหนด PRO1

$$CP(< R_1,R_5 >) = (R_1,R_5) \quad (14)$$

$$\mathrm{CP}(< \mathrm{R_2,R_3,R_4,R_6} >) = (\mathrm{R_2,R_3}),(\mathrm{R_2,R_4}),(\mathrm{R_2,R_6}),(\mathrm{R_3,R_4}),(\mathrm{R_4,R_6}) \quad (15)$$

$$CP(< R_1,R_5 >, < R_2, R_3,R_4,R_6 >) =$$

$$(R_1,R_2), (R_1,R_3),(R_1,R_4), (R_1,R_5),(R_5, R_2),(R_5,R_3), (R_5,R_4),(R_5, R_6) \quad (16)$$

กลุ่ม 2: เส้นทางหมายเลข ① และ ② ภายใต้โหนด PRO1

$$CP(< R_3,R_6 >) = (R_3,R_6) \quad (17)$$

$$CP(< R_3,R_6 >, < R_5 >) = (R_3,R_5), (R_6,R_5) \quad (18)$$

กลุ่ม 3: เส้นทางหมายเลข ① และ ② ภายใต้โหนด PRO1

$$CP(< R_5, R_6 >) = (R_5, R_6) \quad (19)$$

ตำแหน่ง CP คือ ผลคูณคาร์ทีเซียน

ผลลัพธ์ของผลคูณคาร์ทีเซียนแต่ละคู่จะคำนวณโดยสมการ (3) ถึง (7) เพื่อค้นหาประเภทของความผิดปกติ ตัวอย่างเช่นในสมการ (14) ของกลุ่ม 1, (R_1, R_5) มีการตัดสินใจเดียวกัน (1) ดังนั้นจึงดำเนินการโดยสมการ (6) ผลลัพธ์ของการดำเนินการ คือ ความผิดปกติซ้ำซ้อน (Redundancy) ตัวอย่างถัดไปในสมการ (15) ประกอบด้วย (R_2, R_3), (R_2, R_4), (R_2, R_6), (R_3, R_4) และ (R_4, R_6) โดยกฎทุกคู่มีการตัดสินใจเหมือนกัน ดังนั้นทั้งหมดจะถูกดำเนินการโดยสมการ (6) เช่นเดียวกับสมการ (14) ผลลัพธ์ของผลคูณคาร์ทีเซียนในสมการ (16) คือ (R_1, R_2), (R_1, R_3), (R_1, R_4), (R_1, R_6), (R_5, R_2), (R_5, R_3), (R_5, R_4) และ (R_5, R_6) ทั้งคู่มีการตัดสินใจที่แตกต่างกันดังนั้นกฎเหล่านี้จะดำเนินการโดยสมการ (3), (4) และ (5) ตามลำดับ ผลลัพธ์ที่คำนวณได้จะแสดงดังต่อไปนี้:

(R_2,R_3) = Redundancy and Semantics loss (executed by equation (6)),

(R_2,R_4)= Redundancy and Semantics loss (6),

(R_2,R_6)= Redundancy and Semantics loss (6),

(R_3,R_4) = Redundancy and Semantics loss (6),

(R_4,R_6) = Redundancy and Semantics loss (6),

(R_1,R_2) = Shadowing (3),

(R_1,R_3) = Correlation (4),

(R_1,R_4) = Shadowing (3),

(R_1,R_6) = Generalization (5),

(R_5,R_2) = Generalization (5),

(R_5,R_3) = Generalization (5),

(R_5,R_4) = Generalization (5), (R_5,R_6) = Generalization (5).

ผลลัพธ์ที่ได้จากการคำนวณหมายเลขกลุ่ม 2 และ 3 ในสมการที่ 17 ถึง 19:

(R_3,R_6) = Redundancy and Semantics loss (executed by equation (6)),

(R_3,R_5) = Generalization (5), (R_6,R_5) = Generalization (5).

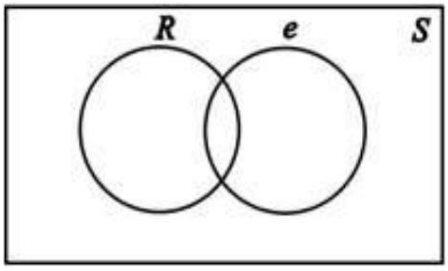
การสูญเสียความหมายของกฎมักจะเกิดขึ้นจากกฎที่ซ้ำซ้อนกัน ดังนั้นสมาชิกทุกตัวในสมการที่ 14, 15 และ 17 จึงเป็นไปได้ที่จะสูญเสียความหมายเช่นกัน

3.3.3 การคำนวณความน่าจะเป็นของแต่ละเส้นทางของ PST (ขั้นตอนที่ 3)

PST ที่ได้รับจากขั้นตอนก่อนหน้านี้ใช้ในการคำนวณความน่าจะเป็นของแต่ละเส้นทางเพื่อให้คำแนะนำแก่ ผู้ดูแลระบบในการตัดสินใจเกี่ยวกับการเพิ่มประสิทธิภาพกฎไฟร์วอลล์อย่างมีประสิทธิภาพซึ่งมีขั้นตอนต่อไปนี้:

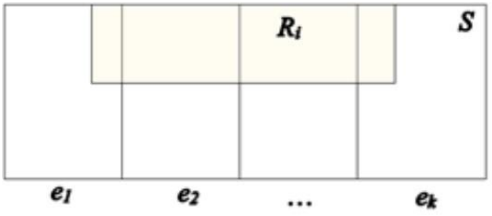
ให้ R เป็นกฎไฟร์วอลล์, e เป็นแอ็ตทริบิวต์คุณสมบัติสำหรับกฎไฟร์วอลล์และ S เป็นขอบเขตของตัวอย่างที่ กำลังพิจารณา แล้วความน่าจะเป็นแบบมีเงื่อนไขของ R ที่รู้ค่า e: แสดงได้สมการ (20) และแสดงในภาพประกอบที่ 3.6

$$P(R|e) = \frac{P(\mathrm{R} \cap e)}{P(e)} \tag{20}$$



ภาพประกอบที่ 3.6 ความน่าจะเป็นแบบมีเงื่อนไขของ R ที่รู้ค่า e แสดงในเวนน์ไดอะแกรม

ตามภาพประกอบที่ 3.7, ให้ R_i เป็นกฎใด ๆ , e_k เป็นคุณลักษณะใด ๆ (ข้อมูลพิเศษ) ของ R_i:



ภาพประกอบที่ 3.7 ความน่าจะเป็นที่มีเงื่อนไขของ R_i ที่ได้รับ e_k

$e_i \cap e_k = \emptyset$; $\forall_i, k$

$e_1 \cup e_2 \cup e_3 \cup \ldots \cup e_k = S = 1$

$$P(R_i) = P(R_i \cap e_1) \cup P(R_i \cap e_2) \cup P(R_i \cap e_3) \cup \ldots \cup P(R_i \cap e_k) \quad (21)$$

จากสมการ (20 $P(R|e) = \frac{P(\mathrm{R} \cap e)}{P(e)}$, ดังนั้น

$$P(R \cap e) = P(e)P(R|e) or P(R)P(e|R)$$

เนื่องจากเราทราบค่าของ $P(e)$ แล้วเราเลือก $P(R \cap e) = P(e)\ P(R \mid e)$ และแทน i และ k ลงในสมการดังนี้

$$P(R_i \cap e_k) = P(e_k)P(R_i|e_k) \quad (22)$$

การใช้สมการ (22) แทนสมการ (21):

$$P(R_i) = P(e_1')P(R_i|e_1) \cup P(e_2)P(R_i|e_2) \cup \ldots \cup P(e_k)P(R_i|e_k) \quad (23)$$

$$p(e_k|R_i) = \frac{P(e_1')P(R_i|e_k)}{P(R_i)} = \frac{P(e_k)P(R_i|e_k)}{\sum_{i,k=1}^{n} P(R_i|e_k)} \quad (24)$$

กำหนดให้ e_k เป็นคณุสมบตั ิตัวใด ๆ เมื่อรู้ค่าความน่าจะเป็นของกฎใด ๆ ที่กำลังพิจารณา จากนั้นแทนค่าดังกล่าวไปยังกฎของ Bayes ซึ่งสามารถอนุมานของเบย์ได้ดังนี้

$$p(e_k|R_i) = \frac{P(R_i \cap e_k)}{P(R_i)} = \frac{P(e_k)P(R_i|e_k)}{\sum_{i,k=1}^{n} P(e_k)P(R_i|e_k)} \quad (25)$$

จากตัวอย่างของฟิลด์คุณสมบัติกฎไฟร์วอลล์ (ฟิลด์พิเศษ) ในขั้นตอนที่ 2 มี 4 ฟิลด์ (e_1', e_2', e_3', e_4') โดยที่ e_1' = ความถี่ของแพ็กเก็ตที่จับคู่กับกฎ (FPM), e_2' = หลักฐานของการสร้างกฎ (ECR), e_3'= ความเชี่ยวชาญของผู้สร้างกฎ (ERC) และ e_4'= ลำดับความสำคัญของโปรโตคอล (PRI) สมมติว่าน้ำหนักของแต่ละคุณลักษณะ e_i'เท่ากัน ดังนั้น $P(e_1'), P(e_2'), P(e_3')$ และ P(e_4') เท่ากับ 25% (0.25) แทนที่ค่าต่าง ๆ ในสมการ (24) และ (25), ผลการคำนวณ:

$$P(e_1') = 0.25, P(e_2') = 0.25, P(e_3') = 0.25, P(e_4') = 0.25$$

ในกรณีของ $P(R_1')$:

$$P(R_1|e_1') = 0.26, P(R_1|e_2') = 0.33, P(R_1|e_3') = 0.67$$

$$P(R_i) = P(e_1')P(R_1|e_1') + P(e_2')P(R_1|e_2') + P(e_3')P(R_1|e_3') + P(e_4')P(R_1|e_4') +$$

$$= P(e_1')P(R_2|e_2') + P(e_2')P(R_2|e_2') + P(e_3')P(R_2|e_3') + P(e_4')P(R_2|e_4') +$$

$$\vdots$$

$$= P(e_1')P(R_2|e_2') + P(e_2')P(R_6|e_2') + P(e_3')P(R_6|e_3') + P(e_4')P(R_6|e_4')$$

$$P(R_i) = ((0.25 * 0.26) + 0.25 * 0.33) + (0.25 * 0.67) + (0.25 * 0.67)) +$$

$$= ((0.25 * 0.06) + (0.25 * 1) + (0.25 * 1) + (0.25 * 0.33)) +$$

$$\vdots$$

$$= ((0.25 * 0.76) + (0.25 * 0) + (0.25 * 1) + (0.25 * 1))$$

$$= 3.07$$

$$P(e_1'|R_1) = \frac{P(e_1')P(R_1|e_1')}{P(R_i)} = \frac{0.25*0.26}{3.07} = 0.0211$$

$$P(e_1'|R_1) = \frac{P(e_1')P(R_1|e_1')}{P(R_i)} = \frac{0.25*0.33}{3.07} = 0.0268$$

$$P(e_1'|R_1) = \frac{P(e_1')P(R_1|e_1')}{P(R_i)} = \frac{0.25*0.67}{3.07} = 0.0545$$

$$P(e_1'|R_1) = \frac{P(e_1')P(R_1|e_1')}{P(R_i)} = \frac{0.25*0.67}{3.07} = 0.0545$$

$$\therefore P(R_1') = P(e_1'|R_1) + P(e_2'|R_1) + P(e_3'|R_1) + P(e_4'|R_1) = 0.157$$

ในกรณีของ $P(\boldsymbol{R_2}')$:

$$P(R_2|e_1') = 0.06, P(R_2|e_2') = 1.0, P(R_2|e_3') = 1.0, P(R_2|e_4') = 0.33$$

$$P(e_1'|R_2) = 0.0488, P(e_2'|R_2) = 0.0814, P(e_1'|R_2) = 0.0814, P(e_1'|R_2) = 0.0268$$

$$\therefore P(R_2') = P(e_1'|R_2) + P(e_2'|R_2) + P(e_3'|R_2) + P(e_4'|R_2) = 0.194$$

ในกรณีของ $P(\boldsymbol{R_3}')$:

$$P(R_3|e_1') = 0.16, P(R_3|e_2') = 0.67, P(R_3|e_3') = 0.33, P(R_3|e_4') = 0.89$$

$$P(e_1'|R_3) = 0.0130, P(e_2'|R_3) = 0.0545, P(e_1'|R_3) = 0.0268, P(e_1'|R_3) = 0.0724$$

$$\therefore P(R_3') = P(e_1'|R_3) + P(e_2'|R_3) + P(e_3'|R_3) + P(e_4'|R_3) = 0.194$$

ในกรณีของ $P(\boldsymbol{R_4}')$:

$$P(R_4|e_1') = 0.40, P(R_4|e_2') = 0.33, P(R_4|e_3') = 0.67, P(R_4|e_4') = 0.56$$

$$P(e_1'|R_4) = 0.0325, P(e_2'|R_4) = 0.268, P(e_1'|R_4) = 0.0545, P(e_1'|R_4) = 0.0456$$

$$\therefore P(R_4') = P(e_1'|R_4) + P(e_2'|R_4) + P(e_3'|R_4) + P(e_4'|R_4) = 0.159$$

ในกรณีของ $P(\boldsymbol{R_5}')$:

$$P(R_5|e_1') = 0.0, P(R_5|e_2') = 1.0, P(R_5|e_3') = 0.0, P(R_5|e_4') = 0.22$$

$$P(e_1'|R_5) = 0.0, P(e_2'|R_5) = 0.0814, P(e_1'|R_4) = 0.0, P(e_1'|R_4) = 0.0179$$

$$\therefore P(R_5') = P(e_1'|R_5) + P(e_2'|R_5) + P(e_3'|R_5) + P(e_4'|R_5) = 0.999$$

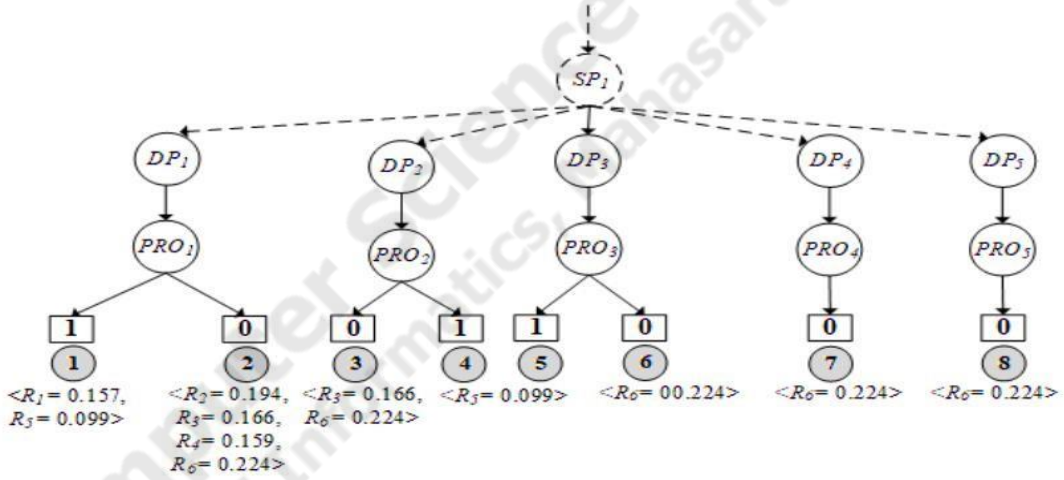
กรณีสุดท้ายของ $P(\boldsymbol{R_6}')$:

$$P(R_6|e_1{}') = 0.76, P(R_6|e_2{}') = 0.0, P(R_6|e_3{}') = 1.0, P(R_6|e_4{}') = 1.0$$

$$P(e_1{}'|R_6) = 0.0618, P(e_2{}'|R_6) = 0.0, P(e_1{}'|R_6) = 0.0814, P(e_1{}'|R_6) = 0.0814$$

$$\therefore P(R_6{}') = P(e_1{}'|R_6) + P(e_2{}'|R_6) + P(e_3{}'|R_6) + P(e_4{}'|R_6) = 0.224$$

ตามน้ำหนักของคุณสมบัติกฎแต่ละกฎ ผู้ดูแลระบบสามารถปรับน้ำหนักของแต่ละคุณสมบัติได้ตามต้องการ เช่น $P(e_1{}') = 35\%\ (0.35), P(e_2{}') = 15\%\ (0.15), P(e_3{}') = 25\%\ (0.25)$ และ $P(e_4{}') = 25\%\ (0.25)$ ขึ้นอยู่กับแต่ละองค์กร เพื่อให้น้ำหนักกับคุณสมบัติของกฎเหล่านั้น หลังจากการคำนวณค่าความน่าจะเป็นทั้งหมดเรียบร้อยแล้วอัลกอริทึมจะแทรกความน่าจะเป็นเหล่านี้ไปยังแต่ละเส้นทางของ PST ดังแสดงในภาพประกอบที่ 3.8



ภาพประกอบที่ 3.8 การใส่ความน่าจะเป็นของแต่ละ Ri ลงใน PSD

3.3.4 การเพิ่มประสิทธิภาพความผิดปกติของกฎ (ขั้นตอนสุดท้าย)

ความผิดปกติที่เกิดขึ้นจากกฎของไฟร์วอลล์มีวิธีแก้ไขปัญหาที่แตกต่างกันเช่นความผิดปกติซ้ำซ้อน (Redundant) ถูกแก้ไขโดยการรวมกฎเข้าด้วยกัน อย่างไรก็ตามวิธีนี้อาจส่งผลให้เกิดการสูญเสียความหมายแทนความผิดปกติอื่น ๆ เช่น Shadowing, Correlation และ Generalisation ไม่ควรใช้วิธีการรวมกันเพราะการตัดสินใจแตกต่างกัน บางครั้งผู้ดูแลระบบเลือกที่จะแก้ไขปัญหาด้วยการสลับกฎ แต่พวกเขาไม่แน่ใจว่าจะ

เกิดอะไรขึ้นในอนาคต ดังนั้นงานวิจัยนี้ใช้ความน่าจะเป็นที่คำนวณได้ในแต่ละกฎเพื่อช่วยผู้ดูแลระบบตัดสินใจว่าจะดำเนินการกับความผิดปกติอย่างไรเพื่อให้เกิดประสิทธิภาพและความสมเหตุสมผลสูงสุด ตัวอย่างเช่น หมายเลขเส้นทาง 1 (1) ของรูปที่ 5, R_1 และ R_5 คือ ความซ้ำซ้อน (Redundant) หากผู้ดูแลระบบตัดสินใจรวมกฎทั้งสองเข้าด้วยกันผลลัพธ์ก็คือ

$$R_1 \Phi R_5 = R_1(f_1) \cup R_5(f_1), R_1(f_2) \cup R_5(f_2) \cup \ldots \cup R_1(f_5) \cup R_5(f_5)$$

$$R_1 \Phi R_5 = ([1{,}100] \cup [1{,}100]), ([1{,}100] \cup [1{,}100]), ([0{,}65535] \cup [0{,}65535]), ([80{,}85] \cup [80{,}85]), ([6{,}17] \cup [6{,}17])$$

$$R_{new} = f_1 \in [1{,}100] \wedge {}^{f_2} \in [1{,}100] \wedge {}^{f_3} \in [0{,}65535] \wedge {}^{f_4} \in [80{,}85] \wedge {}^{f_5} \in [6{,}17] \Rightarrow 1$$

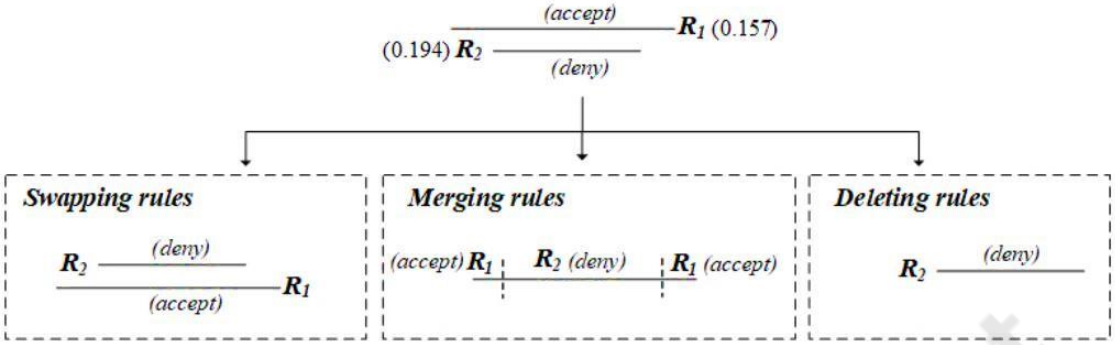
ข้อมูลคุณสมบัติของ R_1: $e_1 = 2500, e_2 = 1, e_3 = 2, e_4 = 6$; และ R_5 คือ $e_1 = 1200, e_2 = 3, e_3 = 0, e_4 = 2$ ดังนั้น, $R_1(e_i) \emptyset R_5(e_i)$:

$$R_1(e_1) \emptyset R_5(e_i) = R_1(e_1) + R_5(e_1), Max\big(R_1(e_2), R_5(e_2)\big), Max\big(R_1(e_2), R_5(e_3)\big), max\big(R_1(e_4), R_5(e_i)\big)$$

$$= R_1(2500) + R_5(1200), Max(R_1(1), R_5(3)), Max(R_1(2), R_5(0)), Max(R_1(6), R_5(2))$$

$$R_{new} \rightarrow (e_1 = 3700, e_2 = 3, e_3 = 2, e_2 = 6)$$

โดยที่ R_{new} เป็นกฎการรวมใหม่ Φ เป็นตัวดำเนินการผสาน $Max()$ เป็นฟังก์ชันที่คำนวณค่าสูงสุด ในลักษณะเดียวกับ $(R_2, R_3), (R_2, R_4), \ldots,$ และ (R_3, R_6) ซึ่งเป็นความขัดแย้งซ้ำซ้อน เพื่อให้สามารถแก้ไขปัญหาได้โดยการรวมกฎเช่น (R_1, R_5) วิธีการแก้ไขความผิดปกติที่เหลืออยู่ (Shadowing, Correlation และ Generalization) สามารถทำได้สามวิธี คือ การยุบรวม การสลับ และการลบกฎ อย่างไรก็ตามผู้ดูแลระบบต้องมีทักษะสูงและตระหนักถึงผลที่ตามมานักวิจัยเกือบทั้งหมดไม่แนะนำให้ใช้วิธีการเหล่านี้และมักจะผลักภาระให้ดุลยพินิจของผู้ดูแลระบบแทน หากผู้ดูแลระบบเลือกวิธีใดวิธีหนึ่งจากสามวิธีพวกเขาสามารถทำได้โดยตรวจสอบความน่าจะเป็นของแต่ละกฎ หากความน่าจะเป็นของกฎใด ๆ นั้นสูงสุดหมายความว่าผู้ดูแลระบบมีโอกาสแก้ไขความผิดปกติให้มีประสิทธิภาพมากขึ้น ตัวอย่างเช่น (R_1, R_2) เป็นความผิดปกติแบบ shadowing หากผู้ดูแลระบบจำเป็นต้อง ลบ รวม หรือสลับกฎ ผู้ดูแลระบบควรให้ความสำคัญกับ R_2 มากกว่า R_1 เพราะ R_2 มีความน่าจะเป็นที่สูงกว่าของ R_1 $(R_1 = 0.157, R_2 = 0.194)$ ดังแสดงในภาพประกอบที่ 3.9



ภาพประกอบที่ 3.9 การแก้ไขเงาโดยการสลับการรวมและการลบกฎ

การอัปเดตคุณสมบัติของ R_1 และ R_2 ไม่จำเป็นในกรณีการเปลี่ยนและการลบกฎ แต่ในกรณีของการรวมมีรายละเอียดดังต่อไปนี้

$R_2 \emptyset R_1 = R_2(f_1) - R_1(f_1), R_2(f_2) - R_1(f_2) \cup \cdots \cup R_2(f_5) - R_1(f_5)$

R^2: $f_1 \in [10,50] \wedge f_2 \in [20,60] \wedge f_3 \in [0,65535] \wedge f_4 \in [80,80] \wedge f_5 \in [6,17] \Rightarrow 0$

R_{new}: $f_1 \in [1,9], [51,100] \wedge f_2 \in [1,19], [61,100] \wedge f_3 \in [0,65535] \wedge f_4 \in [81,85] \wedge f_5 \in [6,17] \Rightarrow 1$

แม้ว่า $R_2(f_3) - R_1(f_3)$ และ $R_2(f_5) - R_1(f_5)$ เท่ากับ $\emptyset$ อย่างไรก็ตามสำหรับโมเดลนี้ทั้งคู่ไม่เท่ากับ $\emptyset$ เนื่องจากเส้นทางการแชร์ในทรี ข้อมูลคุณสมบัติของ R_1: $e_1 = 2500, e_2 = 1, e_3 = 2, e_4 = 6$; และ R_2 คือ $e_1 = 1500, e_2 = 3, e_3 = 3, e_4 = 3$ ดังนั้น, $R_1(e_i) \emptyset R_2(e_i)$:

$R_1(e_i) \emptyset R_2(e_i) = R_1(2500) + R_2(1500), Max(R_1(1), R_2(3)), Max(R_1(2), R_2(3)),$

$Max(R_1(6), R_2(3))$

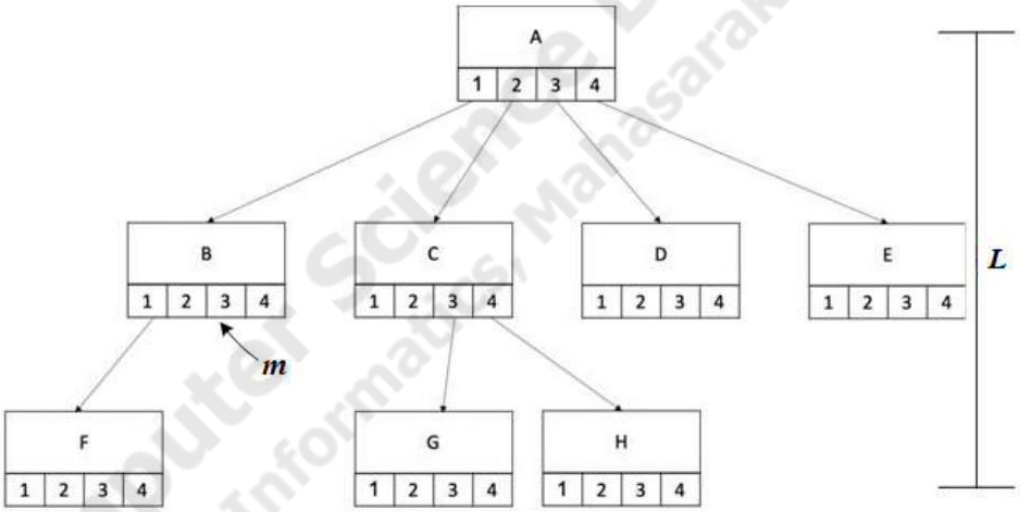
$R_2 \rightarrow (e_1 = 4000, e_2 = 3, e_3 = 3, e_4 = 6)$

$R_{new} \rightarrow (e_1 = 2500, e_2 = 1, e_3 = 2, e_4 = 6)$

โปรดทราบว่าในขณะที่อัปเดตกฎที่ขัดแย้งกันแต่ละครั้งโครงสร้างโครงสร้าง PST จะมีการเปลี่ยนแปลงซึ่งทำให้มีความเป็นไปได้ที่จะคำนวณใหม่ทุกครั้งเมื่อแก้ไขข้อขัดแย้ง

3.4 การใช้งาน PST และการประเมินผลการปฏิบัติงาน

PST ใช้โครงสร้างต้นไม้ k-ary (หรือเรียกอีกอย่างว่าต้นไม้ m-ary หรือ k-way) เพื่อพัฒนา ดังนั้นความเร็วใน การประมวลผลคือ $O(n)$ โดยที่ n คือจำนวนโหนดของต้นไม้ k-ary ที่กำหนด สมมติว่าจำนวนระดับของต้นไม้ k-ary ที่มีอยู่คือ L ความลึกของ k-ary ในกรณีที่แย่ที่สุดคือ N - 1 โดยที่ N คือจำนวนโหนดในต้นไม้ ต้นไม้ k-ary ยังสามารถเก็บไว้ในลำดับความกว้างก่อนเป็นโคร้างสร้างข้อมูลโดยนัย ในแบบตัวชี้แต่ละโหนดจะมีอาร์เรย์ ภายในสำหรับการเก็บพอยน์เตอร์ให้แต่ละตัว ของ childeren m ดังแสดงในภาพประกอบที่ 3.10 ดังนั้นความซับซ้อนของพื้นที่ของโครงสร้างต้นไม้ k-ary คือ $O(m * n)$ การสำรวจเส้นทางต้นไม้ k-ary นั้นคล้ายคลึงกับการสำรวจเส้นทางผ่านต้นไม้ไบนารีนอกจากนี้ความซับซ้อนของเวลา กรณีที่เลวร้ายที่สุดโดยรวมคือ 0(logmn) ในทางปฏิบัติแล้ว PST นั้นถูกใช้งานโดยภาษา Python เวอร์ชัน 3.7 (64 บิต) ทำงานบน MS-Windows และโครงสร้างของทรีเหมือนภาพประกอบที่ 3.10



ภาพประกอบที่ 3.10 การนำตัวชี้ไปยังต้นไม้ k-ary โดยที่ m = 4, L = 2

3.5 สรุป

ในทางปฏิบัติการแก้ไขความผิดปกติของกฎไฟร์วอลล์ค่อนข้างซับซ้อน ขึ้นอยู่กับมุมมองและประสบการณ์ของผู้ดูแลระบบ การแก้ไขข้อผิดพลาดอาจนำไปสู่ความผิดปกติอื่น ๆ ตัวอย่างเช่นเมื่อแก้ไขความผิดปกติซ้ำซ้อนมันอาจกลายเป็นการสูญเสียความหมายของกฎ เพื่อลดผลกระทบของข้อผิดพลาดในการแก้ไขความผิดปกติของผู้ดูแลระบบ ดังนั้นบทความนี้ได้ออกแบบและพัฒนาระบบเพื่อช่วยในการตัดสินใจของผู้ดูแลระบบโดยใช้ความน่าจะเป็นพร้อมกับคุณสมบัติเพิ่มเติม 4 ประการของกฎ คือ ความถี่ของการจับคู่ระหว่าง

แพ็กเก็ต, หลักฐานของการสร้างกฎ, ความเชี่ยวชาญของผู้สร้างกฎ แต่ละกฎคำนวณความน่าจะเป็นตามคุณลักษณะทั้ง 4 นี้ หากความน่าจะเป็นของกฎใด ๆ สูงแสดงว่ากฎมีลำดับความสำคัญสูง ในขณะที่กฎใด ๆ ในไฟร์วอลล์มีข้อขัดแย้งกฎที่มีค่าความน่าจะเป็นสูงจะถือเป็นอันดับแรกเสมอ จากการทดสอบระบบผู้ดูแลระบบสามารถตัดสินใจได้อย่างแม่นยำมากขึ้นเกี่ยวกับกฎข้อขัดแย้งในไฟร์วอลล์ สำหรับประสิทธิภาพโดยรวมของระบบความซับซ้อนของเวลาในการสร้างระบบ (PST) เท่ากับ $O(n)$ เวลาค้นหาผ่าน PST คือ $O(logmn)$ และความซับซ้อนคือ $O(m * n)$ อย่างไรก็ตามระบบยังมีข้อ จำกัด ในการสร้างโครงสร้างต้นไม้ใหม่ ในขณะที่ การแก้ไขความผิดปกติใด ๆ ของกฎในแต่ละช่วงเวลานั้นต้องการโครงสร้างต้นไม้ PST ทั้งชุด