บทที่ 3
ขั้นตอนวิธีที่นำเสนอ

การวิจัยนี้ใช้เทคนิคการวิเคราะห์องค์ประกอบ (Factor Analysis) โดยใช้โปรแกรม SPSS for Windows ในการวิเคราะห์ข้อมูล ดังนี้

1. ทำการตรวจสอบว่าตัวแปรต่างๆ มีความสัมพันธ์กันหรือไม่ หากมีความสัมพันธ์กันอย่างมีนัยสำคัญจึงควรใช้เทคนิคการวิเคราะห์องค์ประกอบ ซึ่งในที่นี้ใช้สถิติ KMO (ค่า KMO หรือ Kaiser-Meyer-Olkin ปกติจะอยู่ที่ค่า 0 ถึง 1 หากมีค่า KMO สูงที่ใกล้ 1 ก็ควรใช้เทคนิคการวิเคราะห์องค์ประกอบ หาก ค่า KMO <0.5 ก็ไม่ควรใช้เทคนิคการวิเคราะห์องค์ประกอบ)
2. ทำการสกัดปัจจัย เพื่อหาจำนวนปัจจัย (Factor) ที่สามารถใช้แทนข้อมูลทั้งหมดได้ โดยใช้วิธี Principal Component Analysisหรือ PCA
3. หมุนแกนปัจจัย เพื่อให้ค่า Factor loading ของตัวแปรมีค่ามากขึ้นหรือลดลงจนกระทั่งทำให้ทราบว่าตัวแปรนั้นๆ ควรอยู่ในปัจจัยใดในการวิจัยนี้จะใช้ Orthogonal rotation
4. ตัวแปรแรกในแต่ละปัจจัยจะเป็นตัวแปรที่มีคุณลักษณะเด่น ที่ได้รับเลือกเป็นตัวแทน
5. เมื่อได้คุณลักษณะเด่นก็นำคุณลักษณะเหล่านั้นไปจัดประเภทด้วย อัลกอริทึม C4.5
6. ทำการเปรียบเทียบค่า Truth positive rate ที่ได้กับ ค่า Truth positive rate จากวิธีต่างๆ เช่นจากข้อมูลทั้งหมด (full features) จากวิธีของคุณ Chi-Hoon Lee และคณะ (ซึ่งจะย่อว่า GA) และจากวิธีของคุณคุณ Kok-Chine Khor และคณะ(ซึ่งจะย่อว่า Kok)

ชุดข้อมูลที่ใช้ทดสอบเป็นชุดข้อมูลที่เป็นมาตรฐาน KDD CUP 1999 จาก MIT Lincoln Labsซึ่งฐานข้อมูลมีจำนวน 494,020 รายการ มีตัวแปรคุณลักษณะทั้งหมด 41 ตัวแปร เป็นชุดข้อมูลที่ได้ถูกใช้และถูกอ้างถึงมากที่สุดในงานวิจัยด้านการตรวจจับการบุกรุกเครือข่าย ซึ่งฐานข้อมูลดังกล่าวนอกจากจะมีคุณลักษณะ 41 ตัวแปรแล้ว ยังมีตัวแปรสุดท้ายเป็นตัวแปรที่บอกประเภทซึ่งมีค่าเป็นปกติ หรือชนิดการบุกรุกดังนี้ normal, Dos, Probing, R2L, U2R โดยรายละเอียดของตัวแปรคุณลักษณะแสดงไว้ในตารางที่ 1

ตารางที่ 1: รายละเอียดตัวแปรคุณลักษณะทั้งหมด 41 ตัว

ตัวแปร	ชื่อคุณลักษณะ	ตัวแปร	ชื่อคุณลักษณะ
V1	duration	V2	protocol_type
V3	service	V4	flag
V5	src_bytes	V6	dst_bytes
V7	land	V8	wrong_fragment
V9	urgent	V10	hot
V11	num_failed_logins	V12	logged_in
V13	num_compromised	V14	root_shell
V15	su_attempted	V16	num_root
V17	num_file_creations	V18	num_shells
V19	num_access_files	V20	num_outbound_cmds
V21	is_host_login	V22	is_guest_login
V23	count	V24	serv_count
V25	serror_rate	V26	srv_serror_rate
V27	reerror_rate	V28	srv_rerror_rate
V29	same_srv_rate	V30	diff_srv_rate
V31	srv_diff_host_rate	V32	dst_host_count
V33	dst_host_srv_count	V34	dst_host_same_srv_rate
V35	dst_host_diff_srv_rate	V36	dst_host_same_src_port_rate
V37	dst_host_srv_diff_host_rate	V38	dst_host_serror_rate
V39	dst_host_srv_serror_rate	V40	dst_host_rerror_rate
V41	dst_host_srv_rerror_rate		

ตาราง 2: รายละเอียดประเภทของการบุกรุกเครือข่าย

ประเภทการบุกรุก	รายละเอียด
DoS	back, land, Neptune, pod, smurf, teardrop
Probing	ipsweep, nmap, portsweep, satan
U2R	buffer_overflow, loadmodule, perl, rootkit
R2L	ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster