การประเมินผลกระทบจากการโจมตีชนิดซิบิลในระบบการลงคะแนน

นายธีรพล ศิลาวรรณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559

EVALUATION OF EFFECT FROM SYBIL ATTACK IN VOTING SYSTEM

Mr. Teerapol Silawan

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctoral of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2016

หัวข้อวิทยานิพนธ์	การประเมินผลกระทบจากการโจมตีชนิดซิบิล ในระบบการลงคะแนน
โดย	นายธีรพล ศิลาวรรณ์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.เชาวน์ดิศ อัศวกุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ศาสตราจารย์ ดร.วาทิต เบญจพลกุล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.เชาวน์ดิศ อัศวกุล)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ชัยเชษฐ์ สายวิจิตร)

.....กรรมการ
(รองศาสตราจารย์ ดร.นิศาชล ตั้งเสงี่ยมวิสัย)

.....กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.ภูมิพัฒ แสงอุดมเลิศ)

ธีรพล ศิลาวรรณ์ : การประเมินผลกระทบจากการโจมตีชนิดซิบิลในระบบการลงคะแนน (EVALUATION OF EFFECT FROM SYBIL ATTACK IN VOTING SYSTEM) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ. ดร. เชาวน์ดิศ อัศวกุล, 81 หน้า.

ระบบการออกเสียงข้างมากเปราะบางต่อการโจมตีชนิดซิบิลและการสร้างตัวตนปลอมของผู้ไม่หวังดี ผู้ไม่หวังดีหวังว่าการออกเสียงจากตัวตนปลอมเป็นจำนวนมากทำให้ชนะการออกเสียงได้ วิทยานิพนธ์นี้นำเสนอการหาผลกระทบจากการโจมตีชนิดซิบิล คำนวณจากความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในระบบการลงคะแนนเสียงข้างมากที่มีหลายตัวเลือก ผลลัพทธ์ที่ได้จากสูตรที่นำเสนอสอดคล้องกับการจำลองเหตุการณ์แบบมอนติคาร์โล และมีความแม่นยำมากกว่าสูตรที่มีผู้นำเสนออยู่ก่อนแล้วซึ่งอยู่บนพื้นฐานของการประมาณค่าหางของการแจกแจงชนิดทวินาม ความซับซ้อนของการคำนวณสูตรแม่นตรงคือ $O\left((n+S)^k\right)$ เมื่อกำหนดให้ n, k, S คือ จำนวนผู้ใช้งานจริง ตัวเลือก และตัวตนปลอมชนิดซิบิลตามลำดับ สูตรการประมาณค่าที่แม่นยำถูกนำเสนอด้วยระดับความซับซ้อน $O(n)$ โดยใช้การประมาณค่าการแจกแจงปัวส์ซอง และ $O(k)$ โดยใช้การประมาณค่าการแจกแจงปกติ ผลกระทบของพารามิเตอร์ในสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงได้รับการทดสอบเพื่อแสดงให้เห็นถึงประโยชน์ของการใช้สูตร นอกจากนำเสนอการหาสูตรความน่าจะเป็นของการโจมตีชนิดซิบิลแล้ว วิทยานิพนธ์นี้ได้นำเสนอการตรวจจับซิบิลเพื่อปกป้องผู้ใช้งานจริงจากการถูกโจมตี เมื่อระบบการออกเสียงถูกนำเสนอในรูปแบบของกราฟทอพอโลยี ขั้นตอนวิธีการตรวจจับตัวตนปลอมชนิดซิบิลได้ถูกนำเสนอโดยใช้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงและพฤติกรรมที่คาดหวังได้จากกลุ่มซิบิลที่จะมีเส้นเชื่อมต่อจากซิบิลถึงผู้ใช้งานจริงที่เป็นเหยื่อมากกว่าเส้นเชื่อมต่อจากผู้ใช้งานจริงถึงซิบิล สุดท้ายการจำลองเหตุการณ์แบบมอนติคาร์โลถูกใช้ในการพิสูจน์สมรรถนะของขั้นตอนวิธีการตรวจหาตัวตนปลอมชนิดซิบิล ผลการทดสอบพบว่าวิธีการตรวจจับซิบิลที่นำเสนอสามารถใช้ตรวจจับซิบิลได้อย่างมีประสิทธิภาพ

ภาควิชา วิศวกรรมไฟฟ้า **ลายมือชื่อนิสิต**.....
สาขาวิชา วิศวกรรมไฟฟ้า **ลายมือชื่อ อ.ที่ปรึกษาหลัก**.....
ปีการศึกษา 2559

5671415421 : MAJOR ELECTRICAL ENGINEERING
KEYWORDS: ONLINE SOCIAL NETWORK, SUCCESS PROBABILITY, SYBIL ATTACK.

TEERAPOL SILAWAN : EVALUATION OF EFFECT FROM SYBIL ATTACK IN VOTING SYSTEM . ADVISOR: ASSOC. PROF. CHAODIT ASWAKUL, Ph.D., 81 pp.

Majority voting systems are vulnerable to Sybil attacks with malicious bogus identity generation. Malicious users hope that voting from many bogus identities can win voting. This dissertation proposes to derive the effect of Sybil attack, as calculated by the success probability of Sybil attack in systems with the multiple-choice majority voting. The outputs, produced by proposed formulas, are consistent with the Monte-Carlo simulation and more accurate than the existing formula based on the multinomial distribution tail estimate. The computational complexity of exact formulas is $O\left((n+S)^k\right)$ where n, k, S are the number of real users, choices, and Sybil users, respectively. The accurate approximation formula is proposed with $O(n)$ complexity by using a Poisson distribution approximation and $O(k)$ complexity by using a normal distribution approximation available. Effects of parameters on the success probability of Sybil attack have been investigated to highlight usefulness of the formulas. Not only the derivation of success probability of Sybil attack, but also this thesis has presented the detection of Sybil identities to protect real users from the attack. With voting system represented as a graph, Sybil detection algorithms have been proposed by using the success probability of Sybil attack and the expected behaviors of Sybil community, with more attack edges than mistaken edges. The Monte Carlo simulation results are finally reported to verify the effectiveness of this Sybil detection algorithm. The numerical results show that the proposed procedures can effectively detect Sybil identities.

Department : Electrical Engineering
Field of Study : Electrical Engineering
Academic Year : 2016

Student's Signature .....
Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ได้ด้วยดีจากความช่วยเหลือของรองศาสตราจารย์ ดร. เชาวน์ดิศ อัศวกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์และผู้ช่วยศาสตราจารย์ ดร. ชัยเชษฐ์ สายวิจิตร โดยทั้งสองท่านได้กรุณาให้ความรู้ทั้งพื้นฐานและบูรณาการต่อการทำวิทยานิพนธ์ พร้อมทั้งให้คำแนะนำและแนวทางต่าง ๆ ที่เป็นประโยชน์อย่างยิ่ง ทำให้นิสิตมีแนวความคิดและกำลังใจในการผลักดันให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ได้ด้วยดี ผู้วิจัยจึงใคร่ขอกราบขอบพระคุณไว้ ณ ที่นี้ นอกจากนี้ขอขอบพระคุณรองศาสตราจารย์ ดร. วาทิต เบญจพลกุล ประธานกรรมการสอบวิทยานิพนธ์ รวมถึงรองศาสตราจารย์ ดร. ภูมิพัฒ แสงอุดมเลิศ และรองศาสตราจารย์ ดร.นิศาชล ตั้งเสงี่ยมวิสัย กรรมการสอบวิทยานิพนธ์ที่ได้สละเวลาตรวจสอบและให้คำแนะนำเพื่อให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอขอบคุณความช่วยเหลือในทุกเรื่องที่ได้รับจากรุ่นพี่ รุ่นเพื่อน และรุ่นน้องทุกคนในห้องปฏิบัติการวิจัยโทรคมนาคม โดยเฉพาะ Wireless Network and Future Internet Research Unit (WiFuN) ที่ได้ให้กำลังใจและคำปรึกษา จนผู้วิจัยสามารถทำวิทยานิพนธ์นี้ได้เสร็จสมบูรณ์ รวมถึงนางสาวปริยาตม์ บุนนาค ที่คอยให้คำปรึกษาและคำแนะนำที่เป็นประโยชน์อย่างยิ่ง

งานวิจัยนี้ได้รับการสนับสนุนจากหน่วยปฏิบัติการวิจัยโครงข่ายไร้สาย และอินเทอร์เน็ตอนาคต (Wireless Network and Future Internet Research Unit) กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย และได้รับการสนับสนุนทุนการศึกษาจากทุนการศึกษาหลักสูตรดุษฎีบัณฑิต "100 ปี จุฬาลงกรณ์มหาวิทยาลัย" (The 100^{th} Anniversary Chulalongkorn University Fund for Doctoral Scholarship)

ท้ายที่สุดวิทยานิพนธ์นี้สามารถสำเร็จลุล่วงไปได้จนเสร็จสมบูรณ์เพราะมีผู้มีอุปการะคุณที่สำคัญ โดยขอกราบขอบพระคุณ บิดา มารดา และครอบครัว ที่เป็นกำลังใจและกำลังทรัพย์ เป็นผู้สนับสนุนและเชื่อมั่นในผู้วิจัยตลอดมา

สารบัญ

บทที่ หน้า

สารบัญรูป

หน้า

หน้า

หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

1.1.1 ระบบการลงคะแนนและจุดอ่อนของระบบการลงคะแนน

ในระบบการลงคะแนนชื่อเสียง (reputation system) ที่มีผู้ลงคะแนนหลายคน ผู้ลงคะแนนสามารถให้คะแนนกับตัวเลือกหรือให้คะแนนกับผู้ให้คะแนนอื่นได้ ตัวอย่างการลงคะแนนให้กับตัวเลือก ได้แก่ การลงคะแนนชื่อเสียงให้กับสินค้าในระบบการซื้อขายสินค้าออนไลน์ การลงคะแนนให้กับคลิปภาพยนต์ที่เป็นที่นิยม หรือการลงคะแนนให้กับผู้สมัครรับการเลือกตั้งต่าง ๆ เป็นต้น ตัวอย่างการลงคะแนนให้กับผู้ลงคะแนนคนอื่นได้แก่ การลงคะแนนเพื่อเลือกหัวหน้ากลุ่ม การลงคะแนนให้พ่อค้าหรือร้านค้าดีเด่น เป็นต้น เมื่อกำหนดให้ตัวเลือก ผู้ลงคะแนนหรือผู้รับคะแนนเป็นองค์ประกอบ (entity) ของระบบ องค์ประกอบที่มีคะแนนมากจึงหมายถึงเป็นองค์ประกอบที่ได้รับความนิยมมาก หรือเป็นตัวชี้วัดว่าเป็นองค์ประกอบที่เคยทำงานได้ดีในอดีต และได้รับความเชื่อใจให้ทำงานหรือได้รับผลประโยชน์พิเศษมากกว่าองค์ประกอบอื่นของระบบ การให้คะแนนถูกกำหนดกฎเกณฑ์ไว้อย่างชัดเจนตั้งแต่ขั้นตอนการออกแบบระบบซึ่งได้ประกาศให้ทุกองค์ประกอบรับรู้ตั้งแต่ต้น ตัวอย่างรูปแบบการใช้คะแนนชื่อเสียงในการตัดสินใจได้แก่ การซื้อขายสินค้ากับร้านค้าออนไลน์ต่าง ๆ ร้านค้าออนไลน์ที่มีลูกค้าให้คะแนนมากกว่าย่อมมีความน่าเชื่อถือมากกว่า คะแนนดังกล่าวจะถูกใช้เป็นข้อมูลให้ลูกค้ารายใหม่เลือกซื้อสินค้ากับร้านที่มีคะแนนสูงกว่า เป็นต้น หรือในระบบทางวิศวกรรมเช่นโครงข่ายตัวรับรู้ไร้สาย (wireless sensor network) บางระบบที่เลือกส่งข้อมูลให้กับโนดที่มีคะแนนชื่อเสียงมากกว่าก่อน เพราะโนดที่มีชื่อเสียงมากย่อมมีโอกาสส่งผ่านข้อมูลให้ถึงปลายทางได้สำเร็จมากกว่าผ่านโนดอื่น เป็นต้น การใช้ชื่อเสียงในการตัดสินใจจึงเป็นแนวคิดที่น่าสนใจ อย่างไรก็ตามการใช้ชื่อเสียงกลับมีจุดอ่อนในมุมมองการถูกโจมตีจากผู้ไม่หวังดีที่ใช้ชื่อเสียงเป็นเครื่องมือในการกระทำผิดได้เช่นกัน

ระบบการลงคะแนนชื่อเสียงที่มีผู้ลงคะแนนหลายคน ผู้ลงคะแนนแต่ละคนสามารถให้คะแนนได้อย่างอิสระ ระบบดังกล่าวมักจะกำหนดให้แต่ละองค์ประกอบของระบบสามารถถือครองอัตลักษณ์หรือตัวตน (identity: ID) เป็นของตนเองเพียง 1 อัตลักษณ์เท่านั้นเพื่อป้องกันไม่ให้เกิดการหลอกลวงหรือสร้างความสับสนขึ้นในระบบ ผู้ไม่หวังดีจึงอาจใช้ข้อกำหนดดังกล่าวเป็นช่องทางในการโจมตีระบบได้ เช่น ในระบบที่กำหนดให้ผู้ลงคะแนนสามารถเลือกลงคะแนนให้แก่ตัวเลือกเดียวเท่านั้น ผู้ไม่หวังดีอาจสร้างตัวตนปลอมขึ้นมาเป็นจำนวนมาก และใช้ตัวตนปลอมเหล่านั้นออกเสียงเลือกตัวเลือกเดียวกัน เพื่อให้ผลการลงคะแนนถูกบิดเบือนจากผลการลงคะแนนที่ควรจะเป็นหรือเป็นไปตามความต้องการของผู้ไม่หวังดี เป็นต้น ทั้งนี้เรียกการโจมตีในลักษณะดังกล่าวว่า การโจมตีชนิดซิบิล [1] และเพื่อให้สื่อถึงอัตลักษณ์ปลอมชนิดซิบิลดีขึ้น จึงขอเรียกอัตลักษณ์ปลอมที่เกิดจากการโจมตีชนิดซิบิลว่า “ซิบิล” จากนี้เป็นต้นไป

1.1.2 ปัญหาการสร้างตัวตนปลอมและการโจมตีชนิดซิบิล

การโจมตีชนิดซิบิล นอกจากจะทำให้เกิดตัวตนปลอมเป็นจำนวนมากในระบบแล้ว ตัวตนปลอมเหล่านั้นสามารถเปลี่ยนชื่อและข้อมูลเป็นอะไรก็ได้ตามความเหมาะสมและลักษณะเงื่อนไขของระบบ

ที่ถูกโจมตี สามารถปรากฏตัวพร้อมกันหรือผลัดเปลี่ยนกันเข้าและออกจากระบบได้ สามารถโน้มน้าวเหยื่อ (ผู้ถูกโจมตี) ให้เชื่อในสิ่งที่ซิบิลนำเสนอในกรณีที่เหยื่อใช้เกณฑ์การตัดสินใจแบบเสียงข้างมาก (majority vote) ได้ เป็นการโจมตีที่เกิดขึ้นจริงในทางปฏิบัติ ตัวอย่างเช่น เพื่อโน้มน้าวให้สามีซื้อกระเป๋าราคาแพงใบใหม่ให้ในวันเกิด ภรรยาสร้างบัญชี facebook ปลอมขึ้นมา 10 บัญชีและปลอมตัวเข้ามาพูดคุยกับสามีของตน ยุยงให้ซื้อกระเป๋าที่ตนเองต้องการ (โดยแสดงตัวว่าเป็นผู้หญิง 10 คน แตกต่างกัน) หากสามีเห็นว่ามีผู้แนะนำตรงกันเป็นสิบคน (แท้จริงเป็นภรรยาคนเดียว) ว่ากระเป๋าใบนี้ดีมากด้วยสรรพคุณต่าง ๆ จึงซื้อกระเป๋าใบนั้นให้ภรรยาเป็นของขวัญวันเกิด เช่นนี้แล้วถือว่าภรรยาโจมตีสามีด้วยซิบิลและทำให้สามีเสียหายอย่างเป็นรูปธรรม จากตัวอย่างดังกล่าว แม้ว่าการเปิดบัญชี facebook ใหม่จะต้องใช้การยืนยันตัวตนในหลายวิธี เช่น ยืนยันด้วยหมายเลขโทรศัพท์หรือจดหมายอิเล็กทรอนิคส์ก็ตาม ภรรยาก็สามารถลงทะเบียนอย่างถูกต้องได้เช่นกัน เช่นการสมัครจดหมายอิเล็กทรอนิคส์หลายบัญชี การซื้อซิมการ์ดราคาถูกหรือใช้ซิมการ์ดแจกฟรีเพื่อยืนยันตัวเอง เป็นต้น ดังนั้นการยืนยันตัวตนจึงไม่สามารถใช้แก้ปัญหาการโจมตีชนิดซิบิลได้

นอกจากการโจมตีของซิบิลในรูปแบบพื้นฐานดังกล่าวแล้ว การโจมตีชนิดซิบิลสามารถโจมตีร่วมกับการโจมตีรูปแบบอื่น ตัวอย่างเช่น ผู้ไม่หวังดีต้องการให้โครงข่ายไม่สามารถส่งข้อมูลถึงกันได้ หรือต้องการดักฟังข้อมูลทั้งหมดของระบบ จึงได้สร้างซิบิลขึ้นมาจำนวนหนึ่ง แล้วนำซิบิลเหล่านั้นให้คะแนนแก่ซิบิลเป้าหมายตัวตนหนึ่ง ทำให้ข้อมูลส่วนใหญ่ในโครงข่ายถูกส่งมาให้ซิบิลเป้าหมายดังกล่าว และทำให้ซิบิลสามารถดักฟังข้อมูลทั้งหมดของระบบได้ หรือถ้าซิบิลตัดสินใจให้คะแนนกับโนดอื่นที่ไม่เป็นซิบิลจะทำให้ข้อมูลทั้งหมดถูกส่งมาที่โนดเป้าหมายและทำให้โนดเป้าหมายประมวลผลไม่ทันแล้วไม่สามารถใช้การได้ในที่สุด เป็นต้น ทั้งนี้ Kevin Hoffman [2] ได้รวบรวมและแบ่งกลุ่มการโจมตีชนิดซิบิลเป็นหัวข้อ ดังนี้

การโจมตีในรูปแบบของการเพิ่มคะแนนตนเอง (Self-promoting)

การโจมตีในรูปแบบของการเพิ่มคะแนนตนเอง [1] ทำได้โดยผู้โจมตีทั้งแบบเดี่ยวและแบบกลุ่มเกิดขึ้นในระบบที่มีการตอบกลับแบบบวก (positive feedback) หรือระบบที่มีการยืนยันตัวตน (authentication) ไม่ดี และถึงแม้จะเป็นระบบที่มีการยืนยันตัวตนที่ดีก็ตาม การโจมตีชนิดซิบิลก็สามารถเพิ่มคะแนนให้พวกพ้องของตัวเองได้อยู่ดี เนื่องจากระบบส่วนใหญ่ไม่ได้เก็บที่มาหรือเหตุผลของการให้คะแนนบวกระหว่างโนดต่าง ๆ ในระบบ แนวทางในการแก้ไขปัญหาเบื้องต้นได้แก่ การตรวจสอบพฤติกรรมโนดต่าง ๆ ในระบบว่ามีการให้คะแนนเฉพาะในกลุ่มของตัวเองมากเกินไปหรือไม่ รวมถึงการตรวจสอบที่มาของคะแนนบวก เป็นต้น

การโจมตีในรูปแบบของการลบล้างคะแนนเชิงลบของตนเอง (Whitewashing)

การโจมตีในรูปแบบของการลบล้างคะแนนเชิงลบของตนเอง [3] เกิดจากผู้โจมตีได้ทำพฤติกรรมอันไม่เหมาะสม และได้รับคะแนนลบเป็นจำนวนมาก ทำให้เกิดความยากลำบากในการโจมตีต่อไป ดังนั้นการลบล้างคะแนนลบของตนเองจึงเป็นกลยุทธ์หนึ่งที่ผู้โจมตีสามารถกระทำได้ โดยหากรู้เกณฑ์ในการนำคะแนนลบออกหรือทราบสูตรในการคำนวณคะแนนลบจะสามารถนำคะแนนลบของตนเองออกได้ แม้กระทั่งทำการเพิ่มคะแนนตนเอง ให้พวกพ้องของตนเองเพื่อให้หักล้างกับคะแนนลบเดิมก็สามารถกระทำได้เช่นกัน หรืออีกวิธีหนึ่งที่ง่ายกว่า คือ การออกจากระบบแล้วเข้ามาใหม่ ทำให้คะแนนเดิมถูกล้างออกทั้งหมด เสมือนว่าไม่เคยมีการให้คะแนนเกิดขึ้นและทำพฤติกรรมไม่ดีต่อระบบต่อไปได้ดังนั้นการปรับแต่งสูตรการคำนวณการให้คะแนนให้มีความแตกต่างระหว่างผู้เข้าระบบใหม่กับผู้มีพฤติกรรมดีที่เก็บสะสมคะแนนตั้งแต่อดีต จะสามารถลดผลกระทบของการโจมตีใน

รูปแบบของการลบล้างคะแนนเชิงลบของตนเองเบื้องต้นได้

การโจมตีในรูปแบบของการใส่ร้ายผู้อื่นในระบบ (Slandering)

การโจมตีในรูปแบบของการใส่ร้ายผู้อื่นในระบบ [2] มีลักษณะคล้ายกับการโจมตีในรูปแบบของการเพิ่มคะแนนตนเอง แต่เป็นในเชิงลบ ระบบที่ไม่เก็บข้อมูลของที่มาของการให้คะแนนเป็นเป้าโจมตีของพฤติกรรมนี้ ความอ่อนไหว (sensitivity) ของสูตรที่ใช้ในการคำนวณและตัดสินใจลงโทษผู้กระทำผิดในระบบเป็นสิ่งที่ต้องคำนึงเป็นอย่างมาก เพราะถ้าหากปรับสูตรให้มีความอ่อนไหวน้อยจะเปิดโอกาสให้ผู้ไม่หวังดีทำการโจมตีระบบเป็นระยะเวลานานได้ แต่หากปรับสูตรให้มีความอ่อนไหวมากเกินไปจะทำให้ผู้บริสุทธิ์ถูกลงโทษอย่างรวดเร็วเมื่อถูกผู้ไม่หวังดีใส่ร้าย แนวทางการป้องกันเบื้องต้นได้แก่ การตรวจพิสูจน์ว่าการให้คะแนนลบเกิดจากการกระทำผิดจริงหรือไม่เท่านั้น

การโจมตีในรูปแบบของการร่วมมือกันระหว่างผู้ไม่หวังดีหลายกลุ่ม (Orchestrated)

การโจมตีในรูปแบบของการร่วมมือกันระหว่างผู้ไม่หวังดีหลายกลุ่ม [4] แต่ละกลุ่มอาจมีพฤติกรรมที่ไม่เหมือนกันในแต่ละช่วงเวลาเพื่อให้ระบบป้องกันตรวจจับได้ยาก ตัวอย่างเช่น ผู้ไม่หวังดีแบ่งตัวเองออกเป็นหลายกลุ่ม แต่ละกลุ่มมีพฤติกรรมที่ไม่เหมือนกัน ในช่วงเวลาขณะหนึ่ง บางครั้งจะแสดงตัวเป็นผู้มีพฤติกรรมดีในขณะที่กลุ่มอื่นแสดงพฤติกรรมทำร้ายระบบ กลุ่มที่แสดงตัวเป็นผู้มีพฤติกรรมดีสร้างชื่อเสียงให้กับตัวเองให้มากกว่าอัตราการลดชื่อเสียงของกลุ่มที่แสดงพฤติกรรมที่ไม่ดี กลุ่มที่แสดงพฤติกรรมที่ไม่ดีจะสร้างความเสียหายให้กับระบบมากที่สุดเท่าที่จะทำได้ จนกว่าชื่อเสียงของตนเองจะน้อยเกินไป เมื่อถึงเวลานั้นพฤติกรรมของผู้โจมตีแต่ละกลุ่มจะเปลี่ยนไป กลุ่มที่แสดงพฤติกรรมไม่ดีแต่เดิมก็จะแสดงพฤติกรรมดีแทน และกลุ่มที่แสดงพฤติกรรมดีเดิมก็จะเริ่มทำร้ายระบบ การทำร้ายระบบดังกล่าวอาจหมายรวมถึง Self-promoting หรือ Slandering ด้วยก็ได้ นอกจากนั้นหากผู้โจมตีแบ่งกลุ่มการโจมตีเป็น 3 กลุ่มขึ้นไป อาจใช้กลุ่มที่เหลือสร้างความเสียหายในรูปแบบอื่นประกอบได้ กลยุทธ์อื่นของการโจมตีแบบ Orchestrated ได้แก่ การแบ่งกลุ่มการโจมตีออกเป็นอย่างน้อย 2 กลุ่ม ถ้ากลุ่มไหนถูกให้คะแนนลบจากโนดอื่นในระบบ โนดที่ให้คะแนนลบจะถูกผู้โจมตีกลุ่มอื่นให้คะแนนลบไปด้วยทันที เป็นต้น

การโจมตีในรูปแบบของการทำลายการให้บริการของระบบ (Denial of Service)

การโจมตีในรูปแบบของการการทำลายการให้บริการของระบบ [5] คือพฤติกรรมการทำให้ระบบใช้งานไม่ได้ โครงข่ายหรือระบบที่ไม่มีการสำรองข้อมูลไว้หลายแห่งเป็นเป้าของโจมตีนี้ วิธีการทำให้ระบบใช้การไม่ได้ ได้แก่ การทำให้ทรัพยากรการคำนวณของศูนย์กลางระบบถูกใช้งานเกินพิกัด ซึ่งการที่ระบบใช้การคำนวณชื่อเสียงขององค์ประกอบของระบบไม่ได้ จะทำให้ระบบต้องทำงานโดยขาดข้อมูลของชื่อเสียงขององค์ประกอบของระบบ เป็นเหตุให้ผู้โจมตีสามารถทำอะไรก็ได้โดยไม่ต้องกลัวว่าจะมีใครให้คะแนนลบ

นอกจากรูปแบบการโจมตีชนิดซิบิลจะมีความหลากหลายแล้ว โครงข่ายที่ถูกโจมตีโดยซิบิลมีความหลากหลาย อาทิ

โครงข่ายเฉพาะกิจ (Ad Hoc Network)

โครงข่ายเฉพาะกิจ คือ โครงข่ายที่มีวิธีการส่งข้อมูลเป็นทอดระหว่างอุปกรณ์ด้วยกันเอง จากต้นทางจนถึงปลายทาง ซึ่งในการส่งข้อมูลแต่ละครั้งอาจส่งข้อมูลด้วยเส้นทางเดิมเสมอหรือเลือกเส้น-

ทางใหม่ก็ได้ ขึ้นอยู่กับกฎเกณฑ์ที่กำหนดไว้ตั้งแต่ต้น การโจมตีชนิดซิบิลที่พบในโครงข่ายเฉพาะกิจมีหลายรูปแบบขึ้นอยู่กับจุดประสงค์ของผู้ไม่หวังดี และพฤติกรรมของผู้โจมตี ตัวอย่างเช่น ในกรณีที่ซิบิลต้องการสร้างความยุ่งยากให้กับระบบ จำนวนซิบิลที่มีมากจะทำให้โครงข่ายใช้เวลานานในการคำนวณหาเส้นทางที่ดีที่สุด และสำหรับกรณีที่โนดต้นทางต้องการส่งข้อมูลที่มีความปลอดภัยสูง โดยใช้วิธีแบ่งข้อมูลออกเป็นหลายส่วน ส่งข้อมูลหลายเส้นทางไม่ซ้ำกัน ถ้าเส้นทางทั้งหมดที่ข้อมูลผ่านมีซิบิลอยู่ด้วย จะทำให้ซิบิลสามารถประกอบข้อมูลเหล่านั้นได้สำเร็จ หรือแม้กระทั่งการออกเสียงเพื่อเลือกเส้นทางที่ดีที่สุดในสถานการณ์ต่าง ๆ ว่าควรส่งข้อมูลไปเส้นทางไหนจึงจะเหมาะสม ถ้าผู้ออกเสียงมีซิบิลปะปนอยู่ด้วยเป็นจำนวนมาก การส่งข้อมูลจะถูกส่งไปเส้นทางที่ซิบิลเสนอ และเกิดความเสียหายได้ในที่สุด เป็นต้น การแก้ไขปัญหาซิบิลในโครงข่ายเฉพาะกิจมีหลายวิธี ตัวอย่างเช่น การกำหนดโปรโตคอลเฉพาะทางร่วมกับอุปกรณ์สื่อสารข้างถนนสำหรับโครงข่ายเฉพาะกิจบนท้องถนน [6] หรือในโครงข่ายเฉพาะกิจเคลื่อนที่ [7] เป็นต้น

โครงข่ายสัญญาณไร้สาย (Wireless Network)

โครงข่ายสัญญาณไร้สาย คือโครงข่ายที่ประกอบด้วยอุปกรณ์สื่อสารด้วยสัญญาณไร้สาย โดยมีการเชื่อมต่อระหว่างกันด้วยรูปแบบเฉพาะตามความเหมาะสมและการใช้งานระบบ ทั้งที่เป็นแบบรวมศูนย์ หรือแบบไม่รวมศูนย์ อย่างไรก็ตามอุปกรณ์ไร้สายมักจะมีข้อจำกัดของการส่งข้อมูล เช่น ช่องสัญญาณที่ใช้ในการส่งข้อมูล ความเร็วในการส่ง จำนวนข้อมูลที่อนุญาตให้ส่งได้ในช่วงเวลาที่กำหนด เป็นต้น ในกรณีที่อุปกรณ์ไร้สายถูกกำหนดให้ได้ทรัพยากรเท่ากันทั้งหมด ผู้ไม่หวังดีที่ต้องการใช้ทรัพยากรมากจึงสร้างตัวตนปลอมขึ้นมาหลายตัวตน [8] และทำการลงทะเบียนด้วยเทคนิคซึ่งตรวจสอบได้ยากว่าเป็นตัวตนปลอม เพื่อให้ได้ทรัพยากรที่มากกว่าผู้ใช้บริการรายอื่น เป็นต้น

โครงข่ายตัวตรวจวัด (Sensor Network)

โครงข่ายตัวตรวจวัด คือ โครงข่ายที่ประกอบด้วยตัวตรวจวัดหลายตัว แต่ละตัวต่างส่งข้อมูลที่ตัวเองตรวจวัดได้ให้ศูนย์ควบคุมกลางหรือระหว่างกันเองเพื่อประมวลผล ในกรณีที่โครงข่ายตัวตรวจวัดถูกโจมตีแบบซิบิลเมื่อซิบิลสร้างตัวตนปลอมเป็นจำนวนมากและส่งข้อมูลที่ไม่เป็นจริงให้กับศูนย์ควบคุมกลาง จะทำให้การประมวลผลของศูนย์ข้อมูลกลางผิดพลาดได้ นอกจากนั้นข้อมูลของเวลาที่ตัวตรวจวัดแต่ละตัวส่งข้อมูลให้ศูนย์ควบคุมกลางเป็นเรื่องสำคัญ ถ้าซิบิลมีมากเพียงพอและส่งข้อมูลเกี่ยวกับเวลาการส่งข้อมูลที่ไม่ถูกต้องให้กับศูนย์ควบคุมกลางหรือให้กับตัวตรวจวัดอื่นเป็นจำนวนมาก จะทำให้ระบบไม่สามารถคำนวณเวลาในการส่งข้อมูลที่ถูกต้องได้ และทำให้ระบบไม่สามารถใช้ข้อมูลทั้งหมดได้ [9]

อินเทอร์เน็ตของสรรพสิ่ง (Internet of Things) [10]

อินเทอร์เน็ตของสรรพสิ่ง คือ ระบบการสื่อสารแบบอัตโนมัติ ที่สามารถทำให้ตัวตรวจวัด (sensor) กับอุปกรณ์ต่าง ๆ ส่งข้อมูลที่เกี่ยวกับอุปกรณ์ ณ ขณะใด ๆ ออกมาเพื่อตัดสินใจ ประมวลผล หรือทำสิ่งอื่นใดที่ได้กำหนดขั้นตอนไว้แล้วอย่างอัตโนมัติ ระบบดังกล่าวสามารถประยุกต์ใช้ได้ในโครงข่ายสังคม (social network) โครงข่ายยานพาหนะ (vehicular network) ระบบตรวจวัดความผิดปกติในร่างกายมนุษย์ (human body) ระบบอัจฉริยะต่าง ๆ ที่เกี่ยวกับความเป็นอยู่ของมนุษย์ เช่น ระบบที่อยู่อาศัยอัจฉริยะ (smart home), ระบบโครงข่ายไฟฟ้าอัจฉริยะ (smart grid), ระบบสังคมอัจฉริยะ (smart community), ระบบเมืองอัจฉริยะ (smart city) เป็นต้น อย่างไรก็ตามความอัจฉริยะต่าง ๆ คงมีความเสี่ยงต่อการถูกโจมตีโดยผู้ไม่หวังดี โดยเฉพาะการโจมตีแบบซิบิลซึ่งผู้ไม่หวัง

ดีสามารถสร้างตัวตนปลอมเป็นจำนวนมาก เป็นผลให้ระบบ Internet of Things รวนได้ เช่น การใช้ตัวตนปลอมในการส่ง spam, การส่งข้อมูลที่ผิดพลาด, การลดความเป็นส่วนตัวของผู้ใช้งานระบบ, การส่งโฆษณา, malware, การโจมตีชนิด fishing เป็นต้น

ระบบซื้อขายสินค้าออนไลน์ (E-Commerce)

ระบบซื้อขายสินค้าออนไลน์เป็นที่นิยมในปัจจุบัน อย่างไรก็ตามผู้ซื้อไม่สามารถจับ ทดลอง หรือได้เห็นของจริงก่อนสั่งซื้อได้สำหรับการซื้อขายในลักษณะดังกล่าว การสั่งซื้อจึงต้องตัดสินใจจากข้อมูลเพียงรูปภาพและสรรพคุณที่ถูกกล่าวอ้างจากผู้ขายเท่านั้น ดังนั้นจึงได้มีการคิดค้นการให้คะแนนชื่อเสียงของร้านค้าขึ้น โดยผู้ซื้อสามารถให้คำแนะนำติชมร้านค้าได้ โดยคำแนะนำติชมดังกล่าวต้องสามารถนำแสดงได้ในที่สาธารณะ เช่น หน้าเว็บไซต์ของร้านค้า เป็นต้น ทั้งนี้ลูกค้ารายใหม่ที่จะเข้ามาซื้อของในร้านค้าจะสามารถใช้คำแนะนำติชมของลูกค้าที่เคยซื้อของกับร้านค้าที่ตนสนใจมาก่อนประกอบการพิจารณาได้ ระบบการให้คะแนนจึงเป็นเป้าของการโจมตีชนิดซิบิลซึ่งผู้ไม่หวังดีอาจจะเป็นลูกค้าที่ไม่พอใจร้านค้าหรือเป็นฝ่ายร้านค้าเองก็ได้ ถ้าผู้ไม่หวังดีเป็นฝ่ายลูกค้าที่เคยไม่พอใจร้านค้า ลูกค้าดังกล่าวจะสร้างบัญชีตัวตนปลอมขึ้นมาหลายบัญชี และเข้าไปที่หน้าเว็บไซต์ของร้านค้า กล่าวติและให้ร้ายร้านค้าเป็นจำนวนมาก เมื่อผู้ซื้อรายใหม่เข้ามาเยี่ยมชมที่หน้าเว็บไซต์ จะเห็นข้อความให้ร้ายของผู้ไม่หวังดีหลายตัวตนพร้อมกัน ทำให้คิดว่าคนส่วนใหญ่ได้รับบริการไม่ดีจากร้านค้าดังกล่าวและไม่กล้าซื้อของจากร้านนั้นในที่สุด เป็นต้น แต่หากผู้ไม่หวังดีเป็นฝ่ายร้านค้าเอง ร้านค้าจะสร้างบัญชีปลอมขึ้นมาหลายบัญชี และใช้บัญชีเหล่านั้นกล่าวชมร้านค้าตัวเองเสมือนว่ามีลูกค้าเป็นจำนวนมากเข้ามาชมและให้กำลังใจร้านค้า เมื่อมีลูกค้าใหม่เข้ามาเยี่ยมชมที่หน้าเว็บไซต์ของร้านค้า จะหลงเข้าใจผิดคิดว่าร้านค้ามีความน่าเชื่อถือสูงและทำการซื้อขายกับร้านค้าดังกล่าวในที่สุด เป็นต้น ตัวอย่างงานวิจัยที่มุ่งเน้นที่จะตรวจจับซิบิล ได้แก่ งานวิจัย [11] ซึ่งใช้ความเหมือนของพฤติกรรมของผู้ใช้งานระบบในการตัดสินว่าผู้ใช้งานไหนเป็นซิบิล เป็นต้น

เครือข่ายสังคมออนไลน์ (Social Network)

เครือข่ายสังคมออนไลน์ คือ ระบบเครือข่ายที่ผู้ใช้งานระบบสามารถติดต่อสื่อสารกับผู้ใช้งานอื่นในระบบได้ ผ่านทางระบบอินเทอร์เน็ต ในรูปแบบของเว็บไซต์หรือโปรแกรมประยุกต์ (application) โดยสามารถพูดคุย กดถูกใจหรือบอกต่อสิ่งที่ผู้อื่นนำเสนอ นำเสนอข้อมูลของตนเองไม่ว่าจะเป็นแบบส่งให้เพื่อนคนเดียวหรือส่งให้กับเพื่อนกลุ่มใหญ่ก็ตาม รวมถึงการรู้จักเพื่อนใหม่ ไม่ว่าจะเป็นเพื่อนที่รู้จักในชีวิตจริงหรือการค้นหาเพื่อนในเครือข่ายสังคมออนไลน์โดยตรงก็ตาม เป็นต้น เนื่องจากสังคมออนไลน์หลายแห่งเปิดให้ทราบว่าเพื่อนของผู้ใช้งานระบบแต่ละคนมีใครบ้าง (ถ้าผู้ใช้งานระบบยินยอม) ดังนั้นการรู้จักเพื่อนของเพื่อนเป็นเรื่องปกติในสังคมออนไลน์ ผู้ใช้งานระบบเพียงแต่ส่งคำร้องขอเป็นเพื่อนและผู้ใช้ระบบปลายทางกดปุ่มยอมรับคำร้องขอดังกล่าว ผู้ใช้งานระบบทั้งสองก็จะสามารถติดต่อสื่อสารกันได้ทันที การรู้จักเพื่อนใหม่ต่อไปเรื่อย ๆ เช่นนี้จะทำให้เครือข่ายมีการสานต่อความสัมพันธ์ที่ซับซ้อนขึ้นเรื่อย ๆ และกลายเป็นสังคมในโลกของอินเทอร์เน็ตในที่สุด ดังนั้นเครือข่ายสังคมออนไลน์จึงเป็นระบบที่อำนวยความสะดวกให้กับผู้ใช้งานระบบในปัจจุบันมาก อย่างไรก็ตามสำหรับระบบที่มีระบบป้องกันการโจมตีที่ไม่ดีเพียงพออาจทำให้สังคมออนไลน์เป็นเป้าโจมตีของผู้ไม่หวังดีได้ [12], [13] การโจมตีในสังคมออนไลน์มีหลายรูปแบบ ทั้งการโจรกรรมข้อมูล การปลอมข้อมูล การแสดงตัวตนที่ผิดพลาด การมีหลายตัวตน และการหลอกลวงต่าง ๆ โดยถึงแม้จะมีมาตรการรักษาความปลอดภัยในเบื้องต้นแล้ว แต่การโจมตีสามารถกระทำได้โดยง่าย ตัวอย่าง เช่น ผู้โจมตีสามารถมีหลายเบอร์โทรศัพท์เพื่อให้การยืนยันตัวตนไม่เป็นผลผ่านหมายเลขโทรศัพท์

จริงของผู้ใช้งานระบบ หรือการมีหลายบัญชีจดหมายอิเล็กทรอนิกส์ (E-mail) เพื่อให้ระบบรักษาความปลอดภัยสามารถตรวจสอบได้เสมือนว่าเป็นผู้ใช้งานระบบจริง เป็นต้น นอกจากนั้นการโจมตีในสังคมออนไลน์ไม่จำเป็นต้องมีความรู้เชิงเทคนิคทางวิศวกรรมมากเหมือนในระบบอื่น โดยเฉพาะการโจมตีชนิดซิบิลที่ผู้ใช้งานระบบคนหนึ่งลงทะเบียนสมัครเป็นผู้ใช้งานระบบสังคมออนไลน์หลายตัวตนปลอมตัวเป็นคนต่าง ๆ ที่อาจมีตัวตนอยู่จริงหรือไม่มีตัวตนอยู่จริงก็ได้ และใช้ตัวตนปลอมดังกล่าวสร้างความเสียหายให้ผู้ใช้ระบบรายอื่น เช่น โน้มน้าวหรือส่งข้อมูลให้ผู้ ใช้งานระบบคนอื่นเข้าใจอะไรบางอย่างผิดไปเพื่อเอื้อประโยชน์ให้ตนเองบางประการได้ เป็นต้น

เนื่องจากปัญหาการโจมตีชนิดซิบิลมีความหลากหลาย จึงมีความจำเป็นต้องหากลยุทธ์ในการแก้ไขปัญหาการโจมตีชนิดซิบิลอย่างเป็นรูปธรรมต่อไป

1.1.3 กลยุทธ์การแก้ไขปัญหาซิบิลเบื้องต้น

การแก้ไขปัญหาซิบิลในอดีตมีหลายแนวคิด แต่ละแนวคิดมีความเหมาะสมแตกต่างกัน ขึ้นอยู่กับธรรมชาติของระบบและรูปแบบการโจมตี โดยปัจจุบันไม่มีการแก้ไขปัญหาของซิบิลอย่างสมบูรณ์แบบและเป็นประเด็นที่น่าสนใจอยู่ในขณะนี้ กลยุทธ์การแก้ไขปัญหาซิบิลในอดีตแบ่งออกเป็นหลายแนวคิด [14] ดังนี้

กลยุทธ์การตรวจสอบทรัพยากร (Resource testing) [12]

การตรวจสอบทรัพยากรของอุปกรณ์แต่ละเครื่องที่ได้รับการจัดสรรหรือกำหนดเองโดยผู้ใช้งานระบบเป็นอีกกลยุทธ์หนึ่งในการแก้ปัญหาซิบิลโดยอุปกรณ์ที่ใช้ทรัพยากรหรือมีความสามารถ เช่น ความสามารถในการคำนวณ ความสามารถในการเก็บข้อมูล ปริมาณการใช้ช่องสัญญาณต่ำกว่าปกติมากอาจมีจุดประสงค์เพียงเพื่อใช้ชื่อหรือเพียงให้มีตัวตนในระบบเท่านั้น ซึ่งวิธีการตรวจสอบทรัพยากรสามารถใช้ร่วมกับกลยุทธ์อื่นในการตรวจจับซิบิลได้

กลยุทธ์การใช้เอกสารรับรองความน่าเชื่อถือ (Trusted certification) [15]

การลงทะเบียนกับหน่วยงานกลางเพื่อยืนยันตัวตนเป็นกลยุทธ์การแก้ไขปัญหาเบื้องต้น ด้วยกลไกบางอย่างที่ได้กำหนดไว้ เช่น การให้คะแนนจากโนดอื่นในระบบหรือการยืนยันตัวตนที่ถูกวิธี จะสามารถยืนยันได้ว่าอุปกรณ์แต่ละอุปกรณ์จะมีตัวตนเพียงหนึ่งเดียวเท่านั้นได้ อย่างไรก็ตามหน่วยงานกลางจะเป็นเป้าโจมตีของผู้ไม่หวังดีได้ นอกจากนั้นการใช้ตัวตนปลอมลงทะเบียนอย่างถูกต้องเป็นเรื่องที่สามารถกระทำได้ นอกจากนั้นการมีหน่วยงานกลางเพียงแห่งเดียวมีปัญหาเรื่องของความใหญ่ขนาดของโครงข่าย ในกรณีที่หน่วยงานกลางไม่สามารถรับภาระการสื่อสารทั้งหมดได้ เป็นต้น

กลยุทธ์การใช้อุปกรณ์ที่มีความน่าเชื่อถือ (Trusted devices) [16]

ในบางครั้งการลงทะเบียนกับหน่วยงานกลางอาจไม่เพียงพอต่อการป้องกันการโจมตีจากผู้ไม่หวังดี การเข้ารหัสที่ได้จากหน่วยงานกลางสำหรับทุกข้อมูลที่ส่งไปปลายทาง เป็นวิธีหนึ่งในการเพิ่มความปลอดภัยให้กับข้อมูล อย่างไรก็ตามการลงทะเบียนกับหน่วยงานกลางหรือการเข้ารหัสข้อมูลก็ไม่เพียงพอต่อการป้องกันการโจมตีได้ทั้งหมดเนื่องจากผู้โจมตีก็สามารถขโมยรหัสได้ในบางกรณีเช่นกัน

กลยุทธ์การเก็บค่าธรรมเนียม (Recurring costs and fees) [17]

เพื่อไม่ให้มีการสร้างตัวตนที่มากเกินไป ในบางครั้งจึงมีการคิดค่าใช้จ่ายในการลงทะเบียนยืนยัน

ตัวตนในระบบ ซึ่งซิบิลจะไม่กล้าสร้างตัวตนมากเกินไปเนื่องจากต้องลงทุนสูง อย่างไรก็ตามวิทยานิพนธ์นี้นำเสนอผลการคำนวณที่ให้ผลบ่งชี้ว่าการโน้มน้าวให้เหยื่อเชื่อในสิ่งที่ซิบิลนำเสนอไม่จำเป็นต้องใช้ซิบิลเป็นจำนวนมากเสมอไปในกรณีที่ผู้ออกเสียงจริงเลือกออกเสียงอย่างเป็นอิสระต่อกันและแต่ละตัวเลือกมีความน่าจะเป็นเท่ากันที่จะถูกเลือกโดยผู้ออกเสียงจริง โดยมีรายละเอียดในบทที่ 3

1.1.4 ความจำเป็นของสูตรคำนวณหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

ถึงแม้งานวิจัยในอดีตจะสามารถกำจัดตัวตนปลอมส่วนใหญ่ได้แล้วก็ตาม แต่ตัวตนปลอมส่วนที่เหลืออาจส่งผลกระทบต่อผู้ถูกโจมตีได้ ซึ่งเท่าที่สำรวจงานวิจัยในอดีตเกี่ยวกับการโจมตีชนิดซิบิลมีงานวิจัยที่คำนึงถึงความรุนแรงของการโจมตีชนิดซิบิลอยู่น้อยมาก ตัวอย่างเช่น งานวิจัย [18] ที่กำหนดให้ผลกระทบของการโจมตีชนิดซิบิลมีค่าเท่ากับจำนวนซิบิลที่ไม่สามารถตรวจจับได้ หรืองานวิจัย [19] ที่กล่าวว่าผลกระทบของการโจมตีชนิดซิบิลมีค่าเป็นจำนวนตัวตนปลอมชนิดซิบิลที่ไม่ลงทะเบียนยืนยันตัวตนในระบบ แต่ผลกระทบดังกล่าวไม่มีการระบุอย่างเป็นรูปธรรมว่าซิบิลที่หลุดรอดจากการตรวจจับหรือไม่ได้ลงทะเบียนสามารถสร้างความเสียหายให้กับเหยื่อหรือระบบได้มากเพียงใด และไม่พบงานวิจัยใดที่สามารถระบุถึงผลกระทบของการโจมตีชนิดซิบิลได้ในรูปแบบที่แน่นอน เช่น ในรูปแบบของสูตรการคำนวณทางคณิตศาสตร์ หรือแม้กระทั่งระบุได้ว่าการเพิ่มขึ้นของซิบิลจะส่งผลกระทบให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพิ่มขึ้นในลักษณะใด และเท่าใด ทั้งนี้มีงานวิจัยที่สนใจในเรื่องที่ใกล้เคียงกับการประเมินผลกระทบของการโจมตีชนิดซิบิล อาทิ งานวิจัยที่คิดค้นโพรโทคอล (protocol) เพื่อลดความสามารถของการโจมตีของซิบิลชนิดบัตรเลือกตั้งปลอมในโครงข่ายวิทยุที่มีความคิด (cognitive radio networks) ถูกนำเสนอโดย [20] โดยใช้สมมุติฐานว่าช่องสัญญาณคลื่นวิทยุมีอยู่เป็นจำนวนจำกัด โพรโทคอลดังกล่าวแบ่งผู้ใช้งานออกเป็นกลุ่มย่อยและเลือกหัวหน้าของกลุ่มย่อยอย่างมีกลยุทธ์ โดยผู้ที่ถูกเลือกให้เป็นหัวหน้ากลุ่มย่อยจะต้องมีความน่าจะเป็นที่จะเป็นซิบิลน้อยที่สุด งานวิจัยที่ออกแบบกลไกการลงคะแนนเสียงเพื่อตรวจจับซิบิลในโครงข่ายสังคมออนไลน์ได้ถูกนำเสนอโดย [21] อย่างไรก็ตามกลไกดังกล่าวไม่มีส่วนของการประเมินความน่าจะเป็นที่ซิบิลจะโจมตีเหยื่อสำเร็จ งานวิจัยที่ประเมินความน่าจะเป็นที่ซิบิลจะโจมตีเหยื่อสำเร็จได้รับการคำนึงถึงใน [22] โดยใช้ซิบิลต่อสู้กับการโจมตีชนิดบอทเน็ตที่ไม่มีหัวหน้า แต่ความน่าจะเป็นที่ซิบิลจะโจมตีสำเร็จในงานวิจัยดังกล่าวก็ไม่ได้อยู่ในรูปแบบของสูตรการคำนวณ นอกจากนั้นมีงาน วิจัยที่นำเสนอกลไกการออกเสียงเพื่อลดความน่าจะเป็นที่ซิบิลจะโจมตีสำเร็จในระบบเสนอแนะ (recommendation systems) ถูกนำเสนอใน [23] ซึ่งอยู่ในรูปแบบของสูตรคณิตศาสตร์แบบปิด (closed form formulas) อย่างไรก็ตามวิธีและสูตรที่ได้นำเสนอนั้นต้องใช้ข้อมูลการลงคะแนนเสียงในอดีตที่ยาวนานในการคำนวณซึ่งในความเป็นจริงระบบอาจถูกโจมตีก่อนที่กระบวนการป้องกันจะเก็บข้อมูลและประมวลผลให้ทันกาลได้ และมีงานวิจัยที่ศึกษาและสร้างแบบจำลองทางสถิติเพื่อคำนวณความน่าจะเป็นที่ซิบิลจะโจมตีสำเร็จถูกนำเสนอในระบบอื่น ๆ อาทิ โครงข่ายการแบ่งปันไฟล์ (file sharing networks) [24] โครงข่ายตัวรับรู้ไร้สาย (wireless sensor networks) [25] และระบบที่ใช้การแจกจ่ายตารางรหัสลับ (distributed hash tables) [26] แต่กลไกการลงคะแนนเสียงก็ไม่มีการนำเสนอในงานวิจัยเหล่านั้น นอกจากจะคำนึงถึงการโจมตีและการมีอยู่ของซิบิลในโครงข่ายแล้ว ในทางกลับกันสำหรับโครงข่ายที่ไม่มีการโจมตีชนิดซิบิลหรือมีการโจมตีชนิดซิบิลอยู่น้อยมากจนสามารถละเลยได้ หากไม่มีวิธีหรือสูตรการคำนวณเพื่อประเมินผลกระทบของการโจมตีชนิดซิบิลก่อน อาจส่งผลกระทบให้สูญเสียทุนทรัพย์สำหรับการลงทุนในระบบกำจัดตัวตนปลอมชนิดซิบิล ดังนั้นสูตรสำหรับประเมินผลกระทบจากการโจมตีชนิดซิบิลด้วยสูตรคณิตศาสตร์ซึ่งในที่นี้กำหนดให้มีค่าเท่ากับความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจึงยิ่งมีความจำเป็น

เนื่องจากกลยุทธ์ของฝ่ายป้องกันได้แก่การเชิญชวนให้ผู้มีสิทธิลงคะแนนจริงออกมาใช้สิทธิลงคะแนนให้มากที่สุด หรือการเปลี่ยนแปลงจำนวนตัวเลือก ในขณะที่กลยุทธ์ของฝ่ายโจมตีได้แก่การเพิ่มจำนวนซิบิลซึ่งจำนวนผู้ออกเสียงจริง จำนวนซิบิล และจำนวนตัวเลือกในการคำนวณ ส่งผลกระทบโดยตรงต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง วิทยานิพนธ์นี้จึงนำเสนอการประเมินผลกระทบจากการโจมตีชนิดซิบิลในระบบการลงคะแนนจากการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่มีกลไกการตัดสินการลงคะแนนแบบเสียงข้างมาก ซึ่งสูตรดังกล่าวใช้ความสัมพันธ์ของจำนวนผู้ออกเสียงจริง จำนวนซิบิล และจำนวนตัวเลือกในการคำนวณ โดยอยู่บนสมมุติฐานว่ารูปแบบการออกเสียงของผู้ออกเสียงจริงมีการกระจายตัวแบบพหุนาม (multinomial distribution) ความแม่นยำของสูตรที่วิทยานิพนธ์นี้นำเสนอได้รับการประเมินผลโดยการจำลองเหตุการณ์แบบมอนติคาร์โล (Monte Carlo simulation) รวมถึงเปรียบเทียบกับสูตรการคำนวณการประมาณค่าหางของการแจกแจงชนิดพหุนาม [27] ที่มีอยู่แล้วในอดีต

1.2 แนวทางของวิทยานิพนธ์

วิทยานิพนธ์นี้นำเสนอการประเมินผลกระทบจากการโจมตีชนิดซิบิลในระบบการลงคะแนนจากการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่มีกลไกการตัดสินการลงคะแนนแบบเสียงข้างมาก ทั้งกรณีทั่วไปและกรณีที่ทุกตัวเลือกมีความน่าจะเป็นเท่ากันที่จะถูกผู้ออกเสียงจริงเลือก ทั้งสองกรณีแบ่งออกเป็นสูตรแบบแม่นตรงและแบบการประมาณค่า และนำผลการคำนวณจากสูตรดังกล่าวไปเปรียบเทียบกับการจำลองเหตุการณ์แบบมอนติคาร์โลเพื่อตรวจสอบความแม่นยำของสูตร โดยจำนวนครั้งที่ทดสอบแต่ละจุดของการทดลองไม่ต่ำกว่าจุดละ 20,000 ครั้ง ในโครงข่ายขนาดย่อย (จำนวนผู้ออกเสียงรวมไม่เกิน 170 คน) ทั้งนี้ความคาดหวังที่มีต่อสูตรการคำนวณแบบแม่นตรงคือต้องการให้ผลของสูตรการคำนวณแบบแม่นตรงมีความคลาดเคลื่อนไม่เกินระดับ 10^{-3} เมื่อเทียบกับการจำลองเหตุการณ์การเลือกตั้งแบบมอนติคาร์โล แต่สำหรับผลของสูตรการคำนวณแบบประมาณค่าควรให้ผลที่คลาดเคลื่อนไม่มากเกินไป รวมถึงต้องมีความซับซ้อนของสูตรที่น้อยกว่าสูตรแบบแม่นตรง และรองรับจำนวนผู้ออกเสียงเป็นจำนวนที่มากกว่าสูตรแบบแม่นตรงอย่างมีนัยสำคัญ โดยเน้นศึกษาผลกระทบจากจำนวนผู้ออกเสียงจริง จำนวนตัวเลือก จำนวนซิบิลและจำนวนผู้ออกเสียงจริงที่เลือกตัวเลือกเดียวกันกับซิบิล ทั้งนี้ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ดังกล่าวที่ถูกนำเสนอโดยสูตร จะสามารถนำไปเป็นข้อมูลพื้นฐานสำหรับออกแบบการออกเสียงหรือป้องกันการโจมตีชนิดซิบิลในระบบการลงคะแนนชื่อเสียงที่มีผู้ลงคะแนนหลายคนต่อไป

1.3 วัตถุประสงค์ของวิทยานิพนธ์

หาสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพื่อใช้เป็นเครื่องมือพื้นฐานในการประเมินความรุนแรงของการโจมตีแบบซิบิลและนำไปประยุกต์ใช้กับวิธีการหาซิบิลในโครงข่ายรวมถึงสามารถระบุแนวทางการแก้ไขปัญหาการโจมตีแบบซิบิลจากสูตรที่นำเสนอ

1.4 ขอบเขตของวิทยานิพนธ์

1. หาสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเมื่อรู้จำนวนของผู้ออกเสียงจริง จำนวนตัวเลือก และจำนวนซิบิลในสถานการณ์ที่มีผู้ออกเสียงจริงไม่เกิน 170 คน

2. นำเสนอการประยุกต์ใช้สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง เพื่อแยกแยะสถานะของผู้ใช้งานระบบว่าเป็นผู้ออกเสียงจริงหรือเป็นซิบิล

1.5 ขั้นตอนและวิธีการดำเนินการ

1. ศึกษารูปแบบการโจมตีชนิดซิบิลในระบบต่าง ๆ พร้อมทั้งศึกษาวิธีการตรวจจับซิบิลจากงานวิจัยในอดีต
2. ระบุเป้าหมายของสูตรรวมถึงข้อมูลที่จำเป็นต้องใช้ในการคำนวณ
3. สร้างสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง
4. นำเสนอการประยุกต์ใช้สูตรที่นำเสนอในระบบต่าง ๆ

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ใช้สูตรที่วิทยานิพนธ์นี้นำเสนอพิจารณาร่วมกับงานวิจัยในอดีตในการตรวจจับซิบิลรวมถึงสามารถคำนวณความรุนแรงของการโจมตีชนิดซิบิลได้อย่างแม่นยำเพื่อเป็นข้อมูลพื้นฐานในการป้องกันการโจมตีชนิดซิบิลได้อย่างมีประสิทธิภาพในอนาคต

1.7 ประมวลวิทยานิพนธ์

บทที่ 1 บทนำ: กล่าวถึงระบบการลงคะแนน การโจมตีชนิดซิบิล กลยุทธ์การแก้ไขปัญหาเบื้องต้น และความจำเป็นของสูตรคำนวณหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

บทที่ 2 ทฤษฎีและความรู้พื้นฐาน: ปูพื้นฐานเกี่ยวกับคณิตศาสตร์ที่ต้องใช้ในการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

บทที่ 3 สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเมื่อรู้จำนวนของผู้ออกเสียงจริง จำนวนตัวเลือก และจำนวนซิบิล: นำเสนอสูตรและที่มาของการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง ทั้งแบบแม่นตรงและแบบประมาณค่า รวมถึงประเมินความถูกต้องและของเขตความสามารถของสูตรที่นำเสนอ

บทที่ 4 การประยุกต์ใช้สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง เพื่อแยกแยะสถานะของผู้ใช้งานระบบ: การนำสูตรคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงไปประยุกต์ใช้

บทที่ 5 บทสรุปและข้อเสนอแนะ: สรุปงานวิจัยทั้งหมดในวิทยานิพนธ์ฉบับนี้และเสนอแนวทางในการพัฒนางานวิจัยต่อไป

บทที่ 2

ทฤษฎีและความรู้พื้นฐาน

ในการหาสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงกำหนดให้มีค่าเท่ากับอัตราส่วนของจำนวนวิธีที่ซิบิลชนะการออกเสียงต่อจำนวนจำนวนวิธีทั้งหมดในการออกเสียงและกำหนดให้การออกเสียงของผู้ออกเสียงจริงแต่ละคนมีพฤติกรรมที่เหมือนกันและเป็นอิสระต่อกัน จำนวนวิธีทั้งหมดในการออกเสียงหาได้จากการแจกแจงพหุนามทั่วไป แต่จำนวนวิธีที่ซิบิลชนะการออกเสียงจะต้องหาด้วยวิธีพิเศษกว่านั้นซึ่งจะกล่าวถึงในบทที่ 3 อย่างไรก็ตามวิธีการดังกล่าวจะต้องใช้การหาสัมประสิทธิ์ของพหุนามที่เกิดจากการบวกกันของเอกนามหลายพจน์และทุกเอกนามมีเลขชี้กำลังของตัวแปรเป็นจำนวนเต็มทั้งหมด จึงจำเป็นต้องมีความรู้เกี่ยวกับการหาสัมประสิทธิ์บางพจน์ของพหุนามซึ่งจะกล่าวในหัวข้อ 2.1 ซึ่งสูตรที่นำเสนอจะถูกนำไปเปรียบเทียบกับสูตรอื่นที่ใกล้เคียงที่นักวิจัยในอดีตได้นำเสนอไว้ก่อนแล้ว ดังรายละเอียดในหัวข้อ 2.2 ถึงแม้สูตรที่นำเสนอในวิทยานิพนธ์นี้จะมีความถูกต้องแม่นยำกว่าสูตรในอดีต (รายละเอียดอยู่ในบทที่ 3) แต่ก็มีข้อบกพร่องอยู่ที่สูตรที่วิทยานิพนธ์นี้นำเสนอติดฟังก์ชันแฟคทอเรียลหลายตำแหน่งซึ่งยากแก่การคำนวณในทางปฏิบัติ การแจกแจงแบบปัวส์ซองจึงถูกใช้เพื่อประมาณค่าบางพจน์ของสูตรแบบแม่นตรงซึ่งจะกล่าวในหัวข้อที่ 2.3 ต่อไป

2.1 การหาสัมประสิทธิ์ของพหุนามที่เกิดจากการบวกกันของเอกนามหลายพจน์และทุกเอกนามมีเลขชี้กำลังของตัวแปรเป็นจำนวนเต็ม

สำหรับฟังก์ชันพหุนาม $f(x)$ ใด ๆ สามารถหาสัมประสิทธิ์ของพจน์ x^n ได้โดยกำหนดให้ $f(x)$ เป็นฟังก์ชันพหุนาม จะได้ว่า

$$f(x)=\sum_{i=0}^{\infty}a_i x^i$$

เมื่อกำหนดให้ a_i คือสัมประสิทธิ์ของพจน์ x^i จาก $\sum_{i=0}^{\infty}a_i x^i = f(x)$ ดังนั้น

$$\frac{d^n}{dx^n}\sum_{i=0}^{\infty}a_i x^i = n!a_n + \sum_{j=n+1}^{\infty}\frac{j!}{(j-n)!}a_j x^{j-n} = \frac{d^n}{dx^n}f(x)$$

เมื่อหารสมการตลอดด้วย $n!$ จะได้ว่า

$$a_n + \sum_{j=n+1}^{\infty}\frac{j!}{n!\,(j-n)!}a_j x^{j-n} = \frac{1}{n!}\frac{d^n}{dx^n}f(x)$$

กำหนดให้ $x=0$ จะได้ว่า

$$a_n = \lim_{x\to 0}\frac{1}{n!}\frac{d^n}{dx^n}f(x)$$

จึงสรุปว่าสัมประสิทธิ์ของพหุนาม $f(x)$ พจน์ที่มี x^n สามารถหาได้จากสูตร

$$a_n = \frac{1}{n!}\lim_{x\to 0}\frac{d^n}{dx^n}f(x)$$

2.2 สูตรการหาความน่าจะเป็นส่วนหางของการแจกแจงอเนกนาม

สำหรับการเลือกตอบคำถามหลายตัวเลือกครั้งหนึ่ง กำหนดให้มีผู้ตอบคำถามทั้งสิ้น n คน และมีตัวเลือกในการตอบคำถามทั้งหมด K ตัวเลือก [27] นำเสนอสูตรสำหรับคำนวณความน่าจะเป็นที่ความถี่สูงสุดของตัวเลือกใด ๆ จะมีค่าไม่เกินค่าคงที่ค่าหนึ่งที่กำหนดไว้ โดยสรุปเป็นสูตรได้ว่า

$$P\left(\frac{f_{\max}-\mu\left(1+\varepsilon\right)}{\sqrt{\frac{n}{2K\log K}}}+\frac{1}{2}\log 4\pi\leq z\right)\rightarrow e^{-e^{-z}}$$

เมื่อกำหนดให้ z คือ ค่าคงที่ค่าหนึ่ง $f_{\max}$ คือ ความถี่สูงสุด μ คือ ตัวแปรตัวหนึ่งที่กำหนดให้มีค่าเท่ากับ $\frac{n}{K}$ และ ε คือ ตัวแปรตัวหนึ่ง ประมาณค่าได้เป็น $\sqrt{\frac{K\log K}{n}}$ เมื่อกำหนดให้ $m=\frac{z-\frac{1}{2}\log\log 4\pi}{\varepsilon\sqrt{2}}+\mu\left(1-\varepsilon\right)$ และย้ายข้างสมการแล้วจะได้ว่า

$$P\left(f_{\max}\leq m\right)\rightarrow e^{-e^{-\left((m-\mu(1-\varepsilon))\varepsilon\sqrt{2}+\frac{1}{2}\log 4\pi\right)}} \tag{2.1}$$

และเมื่อกำหนดให้ F_α คือ m ที่ทำให้ $P\left(f_{\max}\leq m\right)\approx 100\left(1-\alpha\right)\%$ จะได้ว่า

$$F_\alpha\approx\frac{n}{K}+\sqrt{\frac{n\log K}{K}}-\left(\log\log\frac{1}{1-\alpha}+1.266\right)\sqrt{\frac{n}{2K\log K}} \tag{2.2}$$

และเสนอสูตรการคำนวณจำนวน m เพื่อให้ $P\left(f_{\max}\leq m\right)\approx 99\%$ ไว้ คือ

$$F_{.01}\approx\frac{n}{K}+\sqrt{\frac{n\log K}{K}}+2.358\sqrt{\frac{n}{K\log K}} \tag{2.3}$$

โดยความแม่นยำ การนำไปใช้และผลของสูตรเหล่านี้จะนำเสนอในบทถัดไป

2.3 การแจกแจงชนิดปัวส์ซอง

การแจกแจงปัวส์ซอง (Poisson Distribution) [28] คือการแจกแจงเต็มหน่วยแบบหนึ่ง ใช้สำหรับหาความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจเป็นจำนวน x ครั้งในช่วงเวลาหนึ่งและโอกาสที่จะเกิดเหตุการณ์ในแต่ละขณะเวลามีค่าน้อยมาก เมื่อกำหนดให้ในช่วงเวลาที่สนใจมีค่าเฉลี่ยในการเกิดเหตุการณ์เท่ากับ μ สามารถเขียนเป็นสูตรได้ว่า

$$f_p\left(x,\mu\right):=e^{-\mu}\frac{\mu^x}{x!}$$

ซึ่งการแจกแจงดังกล่าวจะมีความแปรปรวนเท่ากับ μ ด้วย

เมื่อนำการแจกแจงดังกล่าวมาวาดเป็นกราฟ โดยกำหนดให้แกนนอนเป็น x และแกนตั้งเป็น $p(x;\mu)$ กราฟจะมีลักษณะเป็นรูประฆังคว่ำแบบไม่ต่อเนื่อง มีความสมมาตรแบบซ้ายขวาเมื่อ μ มีค่ามาก โดยมีสมการแกนสมมาตรเป็น $x=\mu$

2.4 วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว

การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว (fast modularity optimization) [29] เป็นวิธีการแบ่งโครงข่ายออกเป็นกลุ่มย่อย ซึ่งเมื่อแบ่งกลุ่มย่อยแล้วจะทำให้สภาพมอดุลาร์ (modularity) มีค่าเหมาะที่สุด (ในวิทยานิพนธ์นี้กำหนดให้ความเหมาะที่สุดคือมีค่าสูงสุด) โดยสภาพมอดุลาร์ (Q_M) สามารถคำนวณได้โดย

$$Q_M = \sum_{i=1}^{c} \left(e_{ii} - a_i^2\right) \tag{2.4}$$

เมื่อกำหนดให้ c คือ จำนวนกลุ่มย่อย e_{ij} คือ จำนวนเส้นเชื่อมต่อจากโนดในกลุ่ม i ถึงโนดในกลุ่ม j (e_{ii} จึงหมายถึงจำนวนเส้นเชื่อมต่อภายในกลุ่มย่อย i) และ a_i คือจำนวนเส้นเชื่อมต่อรวมจากโนดในกลุ่ม i ถึงโนดในกลุ่ม j โดย $i \neq j$ จะได้ว่า $a_i = \sum_j e_{ij}$ นอกจากนั้นนิยามให้ค่าสภาพมอดุลาร์ที่เปลี่ยนไปเมื่อรวมกลุ่มย่อย i และกลุ่มย่อย j เข้าด้วยกัน $\Delta Q_{M_{ij}}$ ให้มีค่าเท่ากับ $2\left(e_{ij} - a_i a_j\right)$ จากการนิยามสภาพมอดุลาร์และตัวแปรที่เกี่ยวข้องดังกล่าว สามารถหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วนำเสนอโดย [29] ได้ โดยมีขั้นตอนดังนี้

1. แบ่งโครงข่ายออกเป็นกลุ่มย่อย กลุ่มย่อย 1 กลุ่มประกอบด้วยโนด 1 โนด ดังนั้นในขั้นตอนแรกนี้จะมีกลุ่มย่อยเป็นจำนวนเท่ากับโนดในโครงข่าย
2. คำนวณค่า Q_M ของโครงข่าย
3. กำหนดให้ i และ j เป็นกลุ่มย่อยในโครงข่าย ลองสร้างกลุ่มย่อยใหม่ที่เกิดจากการรวมกันของกลุ่มย่อย i และกลุ่มย่อย j แล้วคำนวณค่า $\Delta Q_{M_{ij}}$ ซึ่งเป็นค่าสภาพมอดุลาร์ของกลุ่มย่อยใหม่ที่เกิดจากการรวมกันของกลุ่มย่อย i และกลุ่มย่อย j
4. คำนวณค่า $\Delta Q_{M_{ij}}$ ทุกคู่กลุ่มย่อย i และกลุ่มย่อย j แล้วรวมกลุ่มย่อย i และกลุ่มย่อย j ที่ทำให้ $\Delta Q_{M_{ij}}$ มีค่าสูงที่สุดเข้าด้วยกัน
5. ทำซ้ำข้อ 2 ถึงข้อ 4 จนกระทั่งค่า Q_M ไม่เพิ่มขึ้นอีกแล้ว

วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วถูกใช้เป็นขั้นตอนหนึ่งในวิธีการตรวจจับซิบิลที่จะนำเสนอในหัวข้อที่ 4.2.1 ต่อไป

2.5 สรุป

การหาความน่าจะเป็นที่ซิบิลชนะการออกเสียงที่วิทยานิพนธ์นี้นำเสนอเป็นการคำนวณทางคณิตศาสตร์ทั้งสิ้น โดยต้องมีพื้นฐานเกี่ยวกับวิธีการนับมาก่อน รวมถึงการแจกแจงชนิดต่าง ๆ โดยเฉพาะการแจกแจงแบบปัวส์ซอง นอกจากนั้นสูตรที่นำเสนอถูกประเมินด้วยการเปรียบเทียบกับงานวิจัยใกล้เคียงที่ได้มีผู้นำเสนอแล้วในอดีต ทั้งนี้ความน่าจะเป็นส่วนหางของการแจกแจงอเนกนามเป็นสิ่งที่งานวิจัยในอดีตใช้ ดังนั้นความรู้พื้นฐานเกี่ยวกับการหาสัมประสิทธิ์ของพหุนามที่เกิดจากการบวกกันของเอกนามหลายพจน์และทุกเอกนามมีเลขชี้กำลังของตัวแปรเป็นจำนวนเต็ม สูตรการหาความน่าจะเป็นส่วนหางของการแจกแจงอเนกนามและการแจกแจงชนิดปัวส์ซองจึงถูกนำเสนอในบทนี้ซึ่งความรู้พื้นฐานดังกล่าวจะได้ใช้ในบทที่ 3 นอกจากนั้นบทนี้ได้นำเสนอวิธีการหาค่าเหมาะที่สุดของสภาพมอ

ดุลาร์อย่างเร็วเพื่อใช้เป็นความรู้พื้นฐานเพื่อใช้เป็นขั้นตอนหนึ่งในวิธีการตรวจจับซิบิลที่จะนำเสนอในหัวข้อที่ 4.2.1 ต่อไป

บทที่ 3

สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเมื่อรู้จำนวนของผู้ออกเสียงจริง จำนวนตัวเลือก และจำนวนซิบิล

3.1 สูตรการคำนวณแบบแม่นตรง

3.1.1 กรณีทั่วไป

พิจารณาระบบที่มีการลงคะแนนและมีความเสี่ยงต่อการถูกโจมตีโดยซิบิล กำหนดให้ n คือ จำนวนผู้ลงคะแนน S คือ จำนวนซิบิลและ k คือจำนวนตัวเลือกสำหรับการลงคะแนนครั้งหนึ่ง จุดประสงค์ของบทนี้คือต้องการหาความสัมพันธ์ระหว่างจำนวนผู้ลงคะแนน จำนวนซิบิล และจำนวนตัวเลือก และผลกระทบของการโจมตีชนิดซิบิลซึ่งกำหนดให้มีค่าเท่ากับความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง $P_{sw}(n,k,S)$

กำหนดให้ผู้มีสิทธิลงคะแนนเสียงสามารถลงคะแนนเสียงได้เพียงตัวเลือกเดียวเท่านั้น จากการทำโพลทำให้สามารถรู้ถึงความนิยมของผู้มีสิทธิลงคะแนนเสียงที่มีต่อตัวเลือกแต่ละตัว กำหนดให้เวกเตอร์ $P = [p_1, p_2, ..., p_k]$ เป็นเวกเตอร์แทนความนิยมของตัวเลือกแต่ละตัว โดยความน่าจะเป็น p_i แสดงถึงความนิยมของผู้มีสิทธิลงคะแนนเสียงที่มีต่อตัวเลือก i เพื่อที่จะหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจึงกำหนดสมมุติฐานให้พฤติกรรมของผู้ลงคะแนนจริงเป็นอิสระต่อกันและเหมือนกัน (independently and identically distributed) ทั้งนี้สำหรับกรณีที่พฤติกรรมของผู้ลงคะแนนจริงเป็นอย่างอื่น จำนวนคะแนนเสียงในตัวเลือกที่มีความนิยมสูงสุดจะเพิ่มขึ้น ทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงมีค่าน้อยลง ดังนั้นสูตรคณิตศาสตร์ที่นำเสนอบนสมมุติฐานที่พฤติกรรมของผู้ลงคะแนนจริงเป็นอิสระต่อกันและเหมือนกันจึงสามารถใช้เป็นสูตรเพื่อคำนวณหาค่าสูงสุดของความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงได้ ในทางกลับกันซิบิลจะต้องเลือกตัวเลือกเดียวกันทั้งหมดเพื่อให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงมีค่ามากที่สุดในกรณีที่มีซิบิลอยู่เป็นจำนวนจำกัด กำหนดให้ตัวเลือกที่ซิบิลเลือกถูกเรียกว่า "ตัวเลือกของซิบิล" และตัวเลือกอื่นที่ซิบิลไม่ได้เลือกถูกเรียกว่า "ตัวเลือกของผู้ออกเสียงจริง" กำหนดให้ผลการลงคะแนนของผู้ออกเสียงจริงถูกนำเสนอเป็นเวกเตอร์ $V = [v_1, v_2, \ldots, v_k]$ โดย v_i แทนจำนวนคะแนนเสียงในตัวเลือก i กำหนดให้ซิบิลเลือกตัวเลือก k เสมอโดยไม่เสียนัยสำคัญ ดังนั้นจำนวนคะแนนเสียงในตัวเลือกของซิบิลจึงมีทั้งสิ้น $S + v_k$ คะแนนเสียง ซิบิลจะชนะการออกเสียงเมื่อจำนวนคะแนนเสียงในตัวเลือกของผู้ออกเสียงจริงที่มากที่สุดมีค่าน้อยกว่าจำนวนคะแนนเสียงในตัวเลือกของซิบิล กำหนดให้ $\mathcal{V}$ เป็นเซตของผลการลงคะแนนของผู้ออกเสียงจริง V ที่มีเงื่อนไขว่าซิบิลชนะการออกเสียง หรือกล่าวอีกนัยหนึ่งคือ $\mathcal{V} = \{V | \max(v_1, \ldots, v_{k-1}) \leq v_k + S - 1 \text{ และ } 0 \leq v_k \leq n\}$ ดังนั้น $P_{sw}(n,k,S|P)$ สามารถหาได้จากการรวมของความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงใ น ทุกรูปแบบการออกเสียง $V \in \mathcal{V}$ จะได้ว่า

$$P_{sw}(n,k,S|P) = \sum_{V \in \mathcal{V}} P(n,k|V,P) \tag{3.1}$$

เมื่อตัวเลือกที่ 1 ถึงตัวเลือกที่ k มีความนิยมเป็น $p_1, p_2, \ldots, p_k$ และมีผลคะแนนการออกเสียง

เป็น $v_1, v_2, \ldots, v_k$ จากผู้ออกเสียงจริงที่เลือกออกเสียงอย่างเป็นอิสระต่อกันทั้งสิ้น n คน จะได้ว่า

$$P\left(n,k|V,P\right)=\left(\begin{array}{c} n \\ v_1,v_2,\ldots,v_k \end{array}\right)p_1^{v_1}p_2^{v_2}\ldots p_k^{v_k} \tag{3.2}$$

โดย $\sum_{i=1}^{k} v_i = n$ และ $\sum_{i=1}^{k} p_i = 1$ เนื่องจาก

$$\left(\begin{array}{c} n \\ v_1,v_2,\ldots,v_k \end{array}\right)=\left(\begin{array}{c} n \\ v_k \end{array}\right)\left(\begin{array}{c} n-v_k \\ v_1,v_2,\ldots,v_{k-1} \end{array}\right) \tag{3.3}$$

แทนค่าสมการที่ (3.3) ลงในสมการที่ (3.2) จะได้ว่า

$$P\left(n,k|V,P\right)=\left(\begin{array}{c} n \\ v_k \end{array}\right)p_k^{v_k}\left(n-v_k\right)!\left(\prod_{i=1}^{k-1}\frac{p_i^{v_i}}{v_i!}\right) \tag{3.4}$$

หากกำหนดให้ a คือค่าคงที่ค่าหนึ่ง และ b คือฟังก์ชันที่อิสระต่อตัวแปร x และ $\Omega\left(f(x),a\right)$ เป็นสัมประสิทธิ์ของพจน์ x^a ในฟังก์ชันพหุนาม $f(x)$ เช่น $b=\Omega\left(bx^a,a\right)$ จะได้ว่า

$$\prod_{i=1}^{k-1}\frac{p_i^{v_i}}{v_i!}=\Omega\left(\left(\prod_{i=1}^{k-1}\frac{p_i^{v_i}}{v_i!}\right)x^{n-v_k},n-v_k\right) \tag{3.5}$$

เนื่องจาก $n-v_k=\sum_{i=1}^{k-1}v_i$ ดังนั้น

$$\begin{aligned}\prod_{i=1}^{k-1}\frac{p_i^{v_i}}{v_i!}&=\Omega\left(\left(\prod_{i=1}^{k-1}\frac{p_i^{v_i}}{v_i!}\right)x^{v_1}x^{v_2}\cdots x^{v_{k-1}},n-v_k\right)\\&=\Omega\left(\left(\prod_{i=1}^{k-1}\frac{(p_ix)^{v_i}}{v_i!}\right),n-v_k\right)\end{aligned} \tag{3.6}$$

แทนค่าสมการที่ (3.6) ใน (3.4) จะได้ว่า

$$\begin{aligned}&P\left(n,k|V,P\right)\\&=\left(\begin{array}{c} n \\ v_k \end{array}\right)p_k^{v_k}\left(n-v_k\right)!\Omega\left(\left(\prod_{i=1}^{k-1}\frac{(p_ix)^{v_i}}{v_i!}\right),n-v_k\right)\end{aligned} \tag{3.7}$$

แทนค่าสมการที่ (3.7) ในสมการที่ (3.1) จะได้ว่า

$$\begin{aligned}&P_{sw}\left(n,k,S|P\right)\\&=\sum_{V\in\mathcal{V}}\left(\begin{array}{c} n \\ v_k \end{array}\right)p_k^{v_k}\left(n-v_k\right)!\Omega\left(\left(\prod_{i=1}^{k-1}\frac{(p_ix)^{v_i}}{v_i!}\right),n-v_k\right)\end{aligned} \tag{3.8}$$

เซตของเหตุการณ์ที่ซิบิลจะชนะการออกเสียงจะหาได้จากเซตของเวกเตอร์ V ที่สอดคล้องกับเงื่อนไข $0\leq v_k\leq n$ และ $\max(v_1,v_2,...,v_{k-1})\leq v_k+S-1$ หรืออีกนัยหนึ่งคือ $0\leq v_i\leq v_k+S-1$ สำหรับทุก $i=1,\ldots,k-1$ จะได้ว่า

$$
\begin{aligned}
&P_{sw}\left(n,k,S|P\right)\\
&=\sum_{v_k=0}^{n}\sum_{v_1=0}^{v_k+S-1}\cdots\sum_{v_{k-1}=0}^{v_k+S-1}\left(\begin{array}{c} n \\ v_k \end{array}\right)p_k^{v_k}\left(n-v_k\right)!\\
&\quad\Omega\left(\left(\prod_{i=1}^{k-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right),n-v_k\right)\\
&=\sum_{v_k=0}^{n}\left(\begin{array}{c} n \\ v_k \end{array}\right)p_k^{v_k}\left(n-v_k\right)!\\
&\quad\sum_{v_1=0}^{v_k+S-1}\cdots\sum_{v_{k-1}=0}^{v_k+S-1}\Omega\left(\left(\prod_{i=1}^{k-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right),n-v_k\right)
\end{aligned}
\tag{3.9}
$$

โดยการใช้คุณสมบัติของฟังก์ชัน $\Omega\left(f(x),a\right)$ กล่าวคือ $\Omega(f_1(x),a)+\Omega(f_2(x),a)=\Omega(f_1(x)+f_2(x),a)$ โดยที่ $f_1(x)$ และ $f_2(x)$ เป็นฟังก์ชันพหุนามตัวแปร x จะได้ว่า

$$
\begin{aligned}
&\sum_{v_1=0}^{v_k+S-1}\cdots\sum_{v_{k-1}=0}^{v_k+S-1}\Omega\left(\left(\prod_{i=1}^{k-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right),n-v_k\right)\\
&=\Omega\left(\sum_{v_1=0}^{v_k+S-1}\cdots\sum_{v_{k-1}=0}^{v_k+S-1}\left(\prod_{i=1}^{k-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right),n-v_k\right)\\
&=\Omega\left(\prod_{i=1}^{k-1}\left(\sum_{v_i=0}^{v_k+S-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right),n-v_k\right)
\end{aligned}
\tag{3.10}
$$

เพราะว่า $\Omega(f(x),a)=\frac{1}{a!}\lim\limits_{x\to 0}\frac{d^a}{dx^a}f\left(x\right)$ ดังนั้น

$$
\Omega\left(\prod_{i=1}^{k-1}\left(\sum_{v_i=0}^{v_k+S-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right),n-v_k\right)=\frac{1}{(n-v_k)!}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\prod_{i=1}^{k-1}\left(\sum_{v_i=0}^{v_k+S-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right)\right]
\tag{3.11}
$$

จากสมการที่ (3.9) ถึง (3.11) ทำให้ได้สูตรการหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเป็น

$$
P_{sw}\left(n,k,S|P\right)=\sum_{v_k=0}^{n}\left(\begin{array}{c} n \\ v_k \end{array}\right)p_k^{v_k}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\prod_{i=1}^{k-1}\left(\sum_{v_i=0}^{v_k+S-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right)\right]
\tag{3.12}
$$

3.1.2 กรณีที่แต่ละตัวเลือกมีความน่าจะเป็นเท่ากันที่ผู้ลงคะแนนเสียงจริงจะเลือก

ในกรณีที่ไม่ทราบความนิยมของตัวเลือกแต่ละตัว ถ้ากำหนดให้แต่ละตัวเลือกมีความน่าจะเป็นเท่ากันที่ผู้ออกเสียงจริงจะเลือก จะได้ว่า $\forall i,p_i=\frac{1}{k}$ ดังนั้นสมการที่ (3.12) จะถูกลดรูปเป็น

$$
\begin{aligned}
&P_{sw}\left(n,k,S\right)\\
&=\left(\tfrac{1}{k}\right)^n\sum_{v_k=0}^{n}\left(\begin{array}{c} n \\ v_k \end{array}\right)\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right]
\end{aligned}
\tag{3.13}
$$

ทั้งนี้ความซับซ้อนของสูตรที่ (3.12) และ (3.13) คำนวณจากจำนวนพจน์ที่ใช้ในสูตร เนื่องจาก $\sum\limits_{v_i=0}^{v_k+S-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}$ มีจำนวนพจน์เท่ากับ v_k+S พจน์ ดังนั้น $\prod\limits_{i=1}^{k-1}\left(\sum\limits_{v_i=0}^{v_k+S-1}\frac{\left(p_ix\right)^{v_i}}{v_i!}\right)$ จึงมีจำนวน

พจน์เท่ากับ $(v_k+S)^{k-1}$ พจน์ ทั้งนี้การหาอนุพันธ์ของพหุนามที่มีเลขชี้กำลังเป็นจำนวนเต็มไม่ติดลบจะไม่ทำให้จำนวนพจน์ของพหุนามเพิ่มขึ้น ดังนั้นเมื่อพิจารณาร่วมกับเครื่องหมาย $\sum_{v_k=0}^{n}$ ที่อยู่ซ้ายสุดของฝั่งขวาของสมการที่ (3.12) แล้วจะได้ว่าสมการที่ (3.12) มีจำนวนพจน์อยู่ไม่เกิน $(n+1)(v_k+S)^{k-1}$ พจน์ แต่เนื่องจาก $n+1$ และ v_k+S ต่างมีค่าน้อยกว่า $n+S$ ดังนั้น $(n+1)(v_k+S)^{k-1} \leq (n+S)(n+S)^{k-1} = (n+S)^k$ ทั้งนี้จำนวนพจน์ของสูตรที่ (3.13) จะมีค่าไม่เกินจำนวนพจน์ของสูตรที่ (3.12) ดังนั้นความซับซ้อนของสูตรที่ (3.12) และ (3.13) จึงมีค่าเท่ากับ $O\left((n+S)^k\right)$

การคำนวณความน่าจะเป็นที่ซิบิลชนะการออกเสียงกรณีที่มีตัวเลือกเพียง 2 ตัวเลือก
แทนค่า $k=2$ ในสมการ (3.13) จะได้ว่า

$$
\begin{aligned}
P_{sw}(n,2,S) &= \left(\tfrac{1}{2}\right)^n \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) \lim_{x\to 0} \tfrac{d^{n-v_k}}{dx^{n-v_k}} \left[\sum_{v_i=0}^{v_k+S-1} \tfrac{x^{v_i}}{v_i!} \right] \\
&= \left(\tfrac{1}{2}\right)^n \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) \lim_{x\to 0} \left[\sum_{v_i=0}^{v_k+S-1-(n-v_k)} \tfrac{x^{v_i}}{v_i!} \right] \\
&= \left(\tfrac{1}{2}\right)^n \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) \lim_{x\to 0} \left[1 + \sum_{v_i=1}^{v_k+S-1-(n-v_k)} \tfrac{x^{v_i}}{v_i!} \right] \\
&= \left(\tfrac{1}{2}\right)^n \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) \left[1 + \lim_{x\to 0} \sum_{v_i=1}^{v_k+S-1-(n-v_k)} \tfrac{x^{v_i}}{v_i!} \right] \\
&= \left(\tfrac{1}{2}\right)^n \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) [1+0] \\
&= \left(\tfrac{1}{2}\right)^n \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) \\
&= \sum_{v_k=0}^{n} \left(\begin{array}{c} n \\ v_k \end{array} \right) \left(\tfrac{1}{2}\right)^{v_k} \left(\tfrac{1}{2}\right)^{n-v_k}
\end{aligned}
\tag{3.14}
$$

สูตรดังกล่าวตรงกับการแจกแจงแบบไบโนเมียล (Binomial distribution) ซึ่งสนับสนุนว่าสูตร (3.13) ที่นำเสนอได้ครอบคลุมถึงกรณีที่มี 2 ตัวเลือกแล้ว ทั้งนี้การออกเสียงที่มี 2 ตัวเลือกเป็นสิ่งที่เกิดขึ้นจริงในทางปฏิบัติ ตัวอย่างเช่น การเลือกตั้งประธานาธิบดีของประเทศสหรัฐอเมริกาที่มีเพียง 2 พรรคการเมือง หรือการรับร่างรัฐธรรมนูญซึ่งมีเพียง 2 ทางเลือกคือรับหรือไม่รับ การถามความคิดเห็นเพื่อนในโครงข่ายสังคม หรือการตัดสินใจว่าจะส่งข้อมูลไปปลายทางหรือไม่ในกรณีที่โครงข่ายมีโอกาสที่จะถูกโจมตีสูง เป็นต้น

การคำนวณความน่าจะเป็นที่ซิบิลชนะการออกเสียงกรณีที่ไม่มีผู้ลงคะแนนจริงคนใดเลยเลือกตัวเลือกของซิบิล

การคำนวณความน่าจะเป็นที่ซิบิลชนะการออกเสียงกรณีที่ไม่มีผู้ลงคะแนนจริงคนใดเลยเลือกตัวเลือกของซิบิลเลยสามารถทำได้โดยกำหนดให้ $v_k=0$ ในสมการที่ (3.12) สำหรับกรณีทั่วไป และในสมการที่ (3.13) ในกรณีที่ทุกตัวเลือกมีความน่าจะเป็นเท่ากันที่จะมีผู้ออกเสียงจริงเลือก ดังนั้นสมการที่ (3.13) ในกรณีที่ไม่มีผู้ลงคะแนนจริงคนใดเลยเลือกตัวเลือกของซิบิลจึงมีลักษณะดังนี้

$$P_{sw}\left(n,k,S\right)=\left(\frac{1}{k}\right)^{n}\lim_{x\to 0}\frac{d^{n}}{dx^{n}}\left[\left(\sum_{v_i=0}^{S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right] \tag{3.15}$$

3.2 การประมาณค่าแบบที่ 1 (การประมาณค่าด้วยการแจกแจงปัวส์ซอง)

ถึงแม้สูตรการคำนวณหาความน่าจะเป็นที่ซิบิลชนะการออกเสียงจะถูกนำเสนอไปแล้ว แต่ความซับซ้อนของสูตรการคำนวณดังกล่าวอาจเป็นอุปสรรคในการใช้งานในทางปฏิบัติเพราะข้อจำกัดของเครื่องคำนวณ ดังนั้นวิทยานิพนธ์นี้จึงนำเสนอการประมาณค่า $P_{sw}\left(n,k,S\right)$ ดังนี้

เมื่อเปลี่ยนรูปแบบของ $\left(\begin{array}{c} n \\ v_k \end{array}\right)$ ให้เป็น $\left(\frac{1}{v_k!}\right)\left(n!\right)\left(\frac{1}{(n-v_k)!}\right)$ จะทำให้สมการที่ (3.13) กลายเป็น

$$P_{sw}\left(n,k,S\right)=\left(\frac{1}{k}\right)^{n}\sum_{v_k=0}^{n}\left(\frac{1}{v_k!}\right)\left(n!\right)\left(\frac{1}{(n-v_k)!}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right]\right) \tag{3.16}$$

กำหนดให้ $f_p\left(x,\mu\right)$ คือฟังก์ชันมวลความน่าจะเป็นแบบปัวส์ซองที่มีพารามิเตอร์เป็น μ (ความน่าจะเป็นที่จะเกิดเหตุการณ์ x ครั้งในช่วงเวลาหนึ่ง ในกรณีที่ในระยะเวลาเท่ากันจะมีการเกิดเหตุการณ์ที่สนใจโดยเฉลี่ย μ ครั้ง) มีสูตรการคำนวณว่า

$$f_p\left(x,\mu\right):=e^{-\mu}\frac{\mu^{x}}{x!}$$

เมื่อ e เป็นค่าคงที่ธรรมชาติ มีค่าประมาณ 2.7183 และ $f_p\left(x,1\right)$ คือ ฟังก์ชันมวลของความน่าจะเป็นชนิดปัวส์ซอง ที่มีพารามิเตอร์ (ค่าเฉลี่ยและความแปรปรวน) เป็น 1 เมื่อกำหนดให้ a เป็นค่าคงที่ใด ๆ จะได้ว่า $\frac{1}{a!}=e\left[e^{-1}\frac{1^{a}}{a!}\right]=e\left(f_p\left(a,1\right)\right)$ หรือกล่าวอีกนัยหนึ่งคือ $a!=\frac{1}{e(f_p(a,1))}$ ดังนั้น $\frac{1}{v_k!}=e\left[e^{-1}\frac{1^{v_k}}{v_k!}\right]=e\left(f_p\left(v_k,1\right)\right)$ และ $n!=\frac{1}{e(f_p(n,1))}$ ทำให้สมการที่ (3.16) กลายเป็น

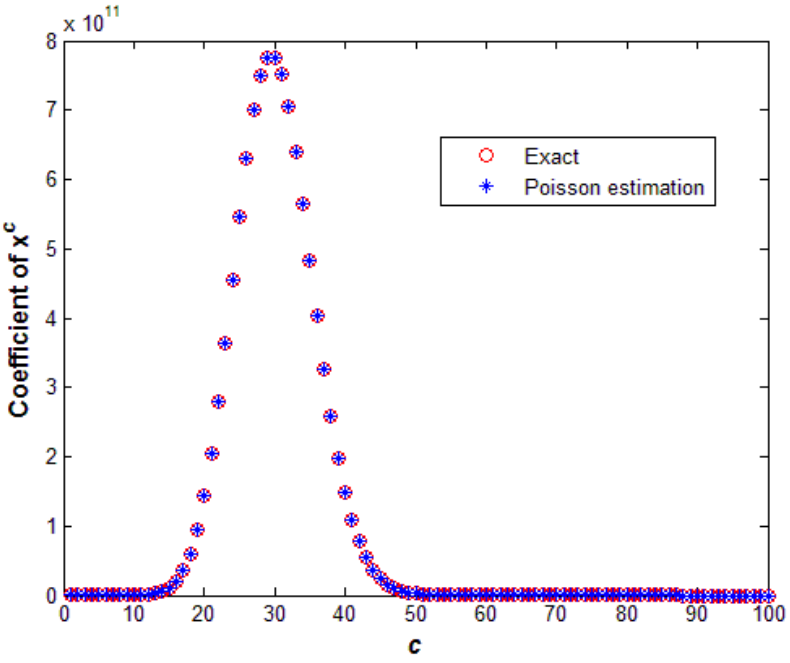
$$P_{sw}\left(n,k,S\right)=\left(\frac{1}{k}\right)^{n}\sum_{v_k=0}^{n}\frac{f_p\left(v_k,1\right)}{f_p\left(n,1\right)}\left(\frac{1}{(n-v_k)!}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right]\right) \tag{3.17}$$

นอกจากนั้นจากผลการนำเสนอด้วยตัวเลข (numerical result) พบว่าเมื่อกำหนดให้ผลรวมคะแนนเสียงในตัวเลือกของผู้ใช้งานจริง $(n-v_k)$ เป็นพารามิเตอร์ในฟังก์ชัน $G\left(n,k,S,V,x\right)=\frac{1}{(n-v_k)!}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right]$ ในกรณีที่จำนวนคะแนนเสียงในตัวเลือกของซิบิลและจำนวนตัวเลือกมีค่ามากจะทำให้ฟังก์ชันดังกล่าวมีลักษณะเป็นรูประฆังคว่ำเกือบสมมาตร อีกทั้งมีค่าเฉลี่ยและความแปรปรวนใกล้เคียงกันซึ่งมีค่าใกล้เคียงกับจำนวนตัวเลือกของผู้ใช้

งานจริง $(k-1)$ นอกจากนั้นพื้นที่ใต้กราฟของฟังก์ชันดังกล่าวมีค่าใกล้เคียงกับ $\left(\sum_{v_i=0}^{v_k+S-1}\frac{1}{v_i!}\right)^{k-1}$ จึงประมาณให้

$$G\left(n,k,S,V,x\right)\approx\left(\sum_{v_i=0}^{v_k+S-1}\frac{1}{v_i!}\right)^{k-1}f_p\left(n-v_k,k-1\right) \tag{3.18}$$

ดังตัวอย่างในรูปที่ 3.1 เมื่อกำหนดให้จำนวนคะแนนเสียงที่มากที่สุดที่ซิบิลชนะการออกเสียง (v_k+S-1) มีค่าเท่ากับ 80 จำนวนตัวเลือกของผู้ใช้งานจริง $(k-1)$ เท่ากับ 30 และกำหนดให้จำนวนคะแนนรวมทุกตัวเลือกของผู้ใช้งานจริง $(c=n-v_k)$ เป็นพารามิเตอร์ในแกนนอนของกราฟ วงกลมสีแดงคือค่าที่แท้จริงของฟังก์ชัน $\frac{1}{(n-v_k)!}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right]$ และดอกจันทร์สีน้ำเงินคือค่าประมาณจากสูตร (3.18) พบว่ามีค่าใกล้เคียงกันและสามารถนำมาใช้แทนกันได้ในกรณีจำนวนผู้เลือกตัวเลือกซิบิลและจำนวนตัวเลือกมีมาก อย่างไรก็ตามการประมาณค่าดังกล่าวมีความคลาดเคลื่อนซึ่งจะได้นำเสนอวิธีการชดเชยความคลาดเคลื่อนดังกล่าวในภายหลัง



รูปที่ 3.1: ตัวอย่างการเปรียบเทียบการหาสัมประสิทธิ์ของพจน์ x^c ในพหุนาม $\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}$ ด้วยวิธีแม่นตรงและวิธีการประมาณค่า ในกรณีที่จำนวนคะแนนเสียงที่มากที่สุดที่ซิบิลชนะการออกเสียง (v_k+S-1) มีค่าเท่ากับ 80 จำนวนตัวเลือกของผู้ใช้งานจริง $(k-1)$ เท่ากับ 30

แทนค่าการประมาณที่ได้จากสมการที่ (3.18) ลงในสมการที่ (3.17) เมื่อกำหนดให้ $\widehat{P}\left(n,k,S\right)$ คือความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงโดยการประมาณค่าด้วยการแจกแจงปัวส์ซองแต่ไม่ได้ชดเชยส่วนต่างของการประมาณค่าจะได้ว่า

$$\widehat{P}\left(n,k,S\right)=\left(\frac{1}{k}\right)^{n}\sum_{v_k=0}^{n}\frac{f_p\left(v_k,1\right)}{f_p\left(n,1\right)}\left(\sum_{v_i=0}^{v_k+S-1}\frac{1}{v_i!}\right)^{k-1}f_p\left(n-v_k,k-1\right) \tag{3.19}$$

เนื่องจาก $\frac{1}{v_i!}=e\left[e^{-1}\frac{1^{v_i}}{v_i!}\right]=e\left(f_p\left(v_i,1\right)\right)$ ดังนั้น

$$\sum_{v_i=0}^{v_k+S-1}\frac{1}{v_i!}=\sum_{v_i=0}^{v_k+S-1}e\left(f_p\left(v_i,1\right)\right)=e\sum_{v_i=0}^{v_k+S-1}f_p\left(v_i,1\right) \tag{3.20}$$

กำหนดให้ $F_p\left(x,\mu\right)$ คือฟังก์ชันการแจกแจงสะสม (cumulative distribution function: cdf) ชนิดปัวส์ซองที่มีพารามิเตอร์เป็น μ จะได้ว่า

$$e\sum_{v_i=0}^{v_k+S-1}f_p\left(v_i,1\right)=eF_p\left(v_k+S-1,1\right) \tag{3.21}$$

แทนค่าสมการที่ (3.21) ลงในสมการที่ (3.19) จะได้ว่า

$$\widehat{P}\left(n,k,S\right)=\left(\frac{1}{k}\right)^{n}\sum_{v_k=0}^{n}\frac{f_p\left(v_k,1\right)}{f_p\left(n,1\right)}\left(eF_p\left(v_k+S-1,1\right)\right)^{k-1}f_p\left(n-v_k,k-1\right) \tag{3.22}$$

อย่างไรก็ตามการประมาณค่าในสมการที่ (3.18) มีความคลาดเคลื่อนอยู่มากในกรณีที่ v_k มีค่าน้อย สูตรที่ (3.22) จึงถูกปรับปรุงเพื่อชดเชยความคลาดเคลื่อนดังกล่าว และคงเงื่อนไข $\lim_{S\to\infty}P_{sw}\left(n,k,S\right)=1$ จึงปรับสูตรให้แม่นยำขึ้นตามอัตราส่วนที่ควรจะเป็น กล่าวคือ $\frac{1-\widehat{P}(n,k,S)}{1-\widehat{P}(n,k,1)}=\frac{1-P_{sw}(n,k,S)}{1-P_{sw}(n,k,1)}$ และ $P_{sw}\left(n,k,1\right)\to\frac{1}{k}$ ดังนั้น

$$P_{sw}\left(n,k,S\right)=1-\left(1-\widehat{P}\left(n,k,S\right)\right)\frac{1-\frac{1}{k}}{1-\widehat{P}\left(n,k,1\right)} \tag{3.23}$$

สมการที่ (3.23) เป็นสูตรการคำนวณหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงแบบประมาณค่าที่มีความซับซ้อนอยู่ในระดับ $O\left(n\right)$ เท่านั้น ทั้งนี้ความถูกต้องของสมการที่ (3.23) จะได้กล่าวในหัวข้อที่ 3.4 ต่อไป

3.3 การประมาณค่าแบบที่ 2 (การประมาณค่าด้วยการแจกแจงปกติ)

3.3.1 กรณีทั่วไป

จากความซับซ้อนของสูตร (3.12) และ (3.13) ซึ่งสามารถใช้ได้กับโครงข่ายขนาดเล็ก ทำให้ต้องมีการประมาณค่าของสูตรในกรณีที่จำนวนผู้ลงคะแนนเสียงและจำนวนตัวเลือกเป็นจำนวนมาก ในกรณีที่ซิบิลชนะการลงคะแนนเสียง จะได้ว่า $P_{sw}\left(n,k,S\right)\approx$

$P\left(\max\left(v_1, v_{2,\ldots}v_{k-1}\right) - v_k \leq S-1\right)$ สำหรับตัวเลือก i การตัดสินใจของผู้ลงคะแนนเสียงจริงคนหนึ่งที่จะเลือกตัวเลือกนั้นเป็นตัวแปรสุ่มแบบเบอร์นูลี (Bernoulli random variable) ดังนั้นจำนวนคะแนนเสียงจากผู้ลงคะแนนเสียงจริงทั้งหมดในข้อ i จึงเป็นตัวแปรสุ่มแบบทวินาม (binomial random variable) ที่มีพารามิเตอร์ p_i เป็นจุดปฏิบัติการ (operating point) ที่มี np_i และ $np_i\left(1-p_i\right)$ เป็นค่าเฉลี่ยและความแปรปรวนของจำนวนคะแนนเสียงในข้อที่ i ตามลำดับ ยิ่งไปกว่านั้น เมื่อ $n \to \infty$ ความไม่เป็นอิสระของจำนวนคะแนนเสียงในแต่ละตัวเลือกจะอ่อนลง ทำให้ $P\left(\max\left(v_1, v_{2,\ldots}v_{k-1}\right) - v_k \leq S-1\right) \approx \prod_{i=1}^{k-1} P\left(v_i - v_k \leq S-1\right)$ และเมื่อ $n \to \infty$ ทำให้ $v_i - v_k \sim \mathcal{N}\left(np_i - np_k, np_i\left(1-p_i\right) + np_k\left(1-p_k\right)\right)$ จะได้ว่า

$$P_{sw}\left(n,k,S|P\right) \approx \prod_{i=1}^{k-1} \Phi\left(\frac{S-1-\left(np_i - np_k\right)}{\sqrt{np_i\left(1-p_i\right) + np_k\left(1-p_k\right)}}\right) \tag{3.24}$$

เมื่อ $\Phi\left(\right)$ เป็นฟังก์ชันการกระจายตัวสะสม (cumulative distribution function) ของการกระจายตัวแบบปกติมาตรฐาน ความซับซ้อนของสูตรที่ (3.24) มีค่าเท่ากับ $O\left(k\right)$ ซึ่งสามารถนำไปใช้ได้ในทางปฏิบัติมากกว่าสูตรที่ (3.12) เนื่องจากมีความซับซ้อนที่ต่ำกว่า

จากสมการที่ (3.24) สามารถแบ่งเป็นกรณีย่อยได้ดังนี้

1. กรณีที่ทุกตัวเลือกมีความน่าจะเป็นเท่ากันที่ผู้ใช้งานจริงจะเลือก (รวมทั้งตัวเลือกของซิบิลด้วย) หาได้โดยกำหนดให้ $\forall i, p_i = \frac{1}{k}$ และ $p_k = \frac{1}{k}$ ในสมการที่ (3.24) จะได้ว่า

$$P_{sw}\left(n,k,S\right) \approx \Phi^{k-1}\left(\frac{S-1}{\sqrt{\frac{2n}{k}\left(1-\frac{1}{k}\right)}}\right) \tag{3.25}$$

2. กรณีพิเศษที่มีเพียง 2 ตัวเลือกรวมถึงทุกตัวเลือกของผู้ใช้งานจริงมีความนิยมเท่ากัน หาได้โดยกำหนดให้ $k=2$ ในสมการที่ (3.25) จะได้ว่า

$$P_{sw}\left(n,2,S\right) \approx \Phi\left(\left(S-1\right)\sqrt{\frac{2}{n}}\right) \tag{3.26}$$

3. กรณีที่ไม่มีผู้ใช้งานจริงคนใดเลือกตัวเลือกของซิบิลเลย หาได้โดยกำหนดให้ $p_k = 0$ ในสมการที่ (3.24) จะได้ว่า

$$P_{sw}\left(n,k,S|P\right) \approx \prod_{i=1}^{k-1} \Phi\left(\frac{S-1-np_i}{\sqrt{np_i\left(1-p_i\right)}}\right) \tag{3.27}$$

4. กรณีที่ไม่มีผู้ใช้งานจริงคนใดเลือกตัวเลือกของซิบิลเลยและตัวเลือกของผู้ใช้งานจริงที่เหลือมีความน่าจะเป็นเท่ากันที่ผู้ใช้งานจริงจะเลือก หาได้โดยแทนค่า $p_k = 0$ และ $\forall i \neq k, p_i = \frac{1}{k-1}$ ลงในสมการ (3.27) จะได้ว่า

$$P_{sw}\left(n,k,S\right) \approx \Phi^{k-1}\left(\frac{S-1-\frac{n}{k-1}}{\sqrt{\frac{n}{k-1}\left(1-\frac{1}{k-1}\right)}}\right) \tag{3.28}$$

สูตรที่ (3.28) จะมีค่าเข้าใกล้สูตรที่ (3.13) ในกรณีที่ k มีค่ามาก ซึ่งในกรณีดังกล่าวสูตรที่ (3.28) มีความเหมาะสมในการใช้งานมากกว่าสูตรที่ (3.13) เนื่องจากมีความซับซ้อนอยู่ในระดับ $O(1)$ เท่านั้น

ทั้งนี้สูตรคำนวณหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงทั้งแบบแม่นตรงและแบบประมาณค่าด้วยการแจกแจงปกติได้รับการยอมรับและตีพิมพ์แล้ว [30] โดย IEEE Communications Letters (2017)

3.4 การประเมินความถูกต้องและขอบเขตความสามารถของสูตร

สูตรที่นำเสนอถูกประเมินความถูกต้องด้วยการจำลองเหตุการณ์ด้วยวิธีมอนติคาร์โล โดยกำหนดให้ความน่าจะเป็นที่ซิบิลชนะการออกเสียงมีค่าเท่ากับอัตราส่วนจำนวนครั้งที่ซิบิลชนะการออกเสียงต่อจำนวนครั้งที่มีการออกเสียงทั้งหมดโดยผู้ออกเสียงจริงแต่ละคนตัดสินใจเลือกตัวเลือกอย่างเป็นอิสระซึ่งกันและกัน กำหนดให้มีตัวเลือกทั้งสิ้น k ตัวเลือก และโดยไม่เสียนัยสำคัญกำหนดให้ซิบิลเลือกตัวเลือกที่ k เสมอ สำหรับการจำลองเหตุการณ์การออกเสียง 1 ครั้งกำหนดให้เวกเตอร์ขนาด $1 \times n$ แทนผลการออกเสียงของผู้ออกเสียงจริงแต่ละคน โดยตำแหน่งที่ h ของเวกเตอร์คือตัวเลือกที่ผู้ออกเสียงจริงคนที่ h เลือก จากนั้นจึงนับความถี่ของทุกตัวเลือก ตั้งแต่ตัวเลือกที่ 1 ถึงตัวเลือกที่ $k-1$ ถ้าความถี่สูงสุดของทุกตัวเลือกที่ซิบิลไม่ได้เลือกดังกล่าวมีค่าน้อยกว่าคะแนนเสียงในข้อที่ซิบิลเลือก ถือว่าซิบิลชนะการออกเสียงในครั้งนั้น ทำการจำลองเหตุการณ์เป็นจำนวนมากจึงได้ค่าความน่าจะเป็นที่ซิบิลชนะการออกเสียงในที่สุด ดังตัวอย่างต่อไปนี้

ตัวอย่าง

ในการออกเสียงครั้งหนึ่ง มีตัวเลือกทั้งหมด 4 ตัวเลือก มีผู้ออกเสียงจริงทั้งหมด 10 คน มีซิบิลทั้งหมด 3 ตัวปะปนอยู่ด้วย และซิบิลเลือกตัวเลือกที่ 4 จำลองเหตุการณ์การออกเสียงทั้งหมด 20 ครั้ง สามารถหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงได้ ดังตารางที่ 3.1 จากการทดลองออกเสียงทั้งหมด 20 ครั้ง พบว่าซิบิลชนะเป็นจำนวน 14 ครั้ง คิดเป็น 70% เป็นต้น

ครั้งที่ออกเสียง	ผลการออกเสียง คนที่									
	1	2	3	4	5	6	7	8	9	10
1	4	1	2	4	3	4	4	1	3	2
2	1	2	4	2	1	3	1	3	3	2
3	1	2	4	3	1	2	3	3	1	1
4	1	2	1	2	4	1	2	4	3	3
5	1	4	2	1	3	2	2	3	4	3
6	3	3	1	3	4	2	3	4	2	3
7	1	1	2	1	4	1	2	2	3	4
8	1	4	1	2	3	1	1	2	2	1
9	4	1	4	2	3	3	3	4	4	3
10	2	2	3	4	2	3	4	2	3	2
11	4	1	2	4	1	2	3	3	3	1
12	1	4	1	2	2	4	4	3	3	3
13	1	4	2	3	1	4	2	2	3	1
14	1	2	1	2	3	4	2	1	1	3
15	4	1	4	1	4	1	1	2	2	2
16	1	4	2	4	3	1	2	3	4	1
17	4	2	2	1	2	3	2	3	2	4
18	1	2	1	3	1	1	1	4	4	4
19	2	2	4	2	2	3	3	3	3	4
20	4	2	2	2	3	4	3	3	3	2

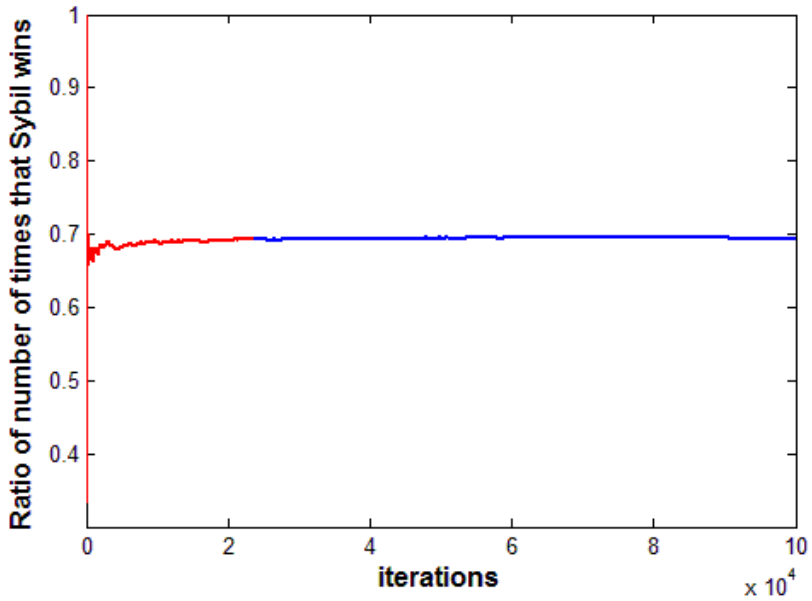
ตารางที่ 3.1: ตัวอย่างผลการออกเสียงจากการจำลองเหตุการณ์แบบมอนติคาร์โล

ครั้งที่ออกเสียง	ความถี่ของแต่ละตัวเลือก					สถานการณ์ชนะของชิบิล
	1	2	3	ความถี่สูงสุดของ 3 ตัวเลือกแรก	4	
1	2	2	2	2	4 + 3 = 7	ชนะ
2	3	3	3	3	4 + 3 = 7	ชนะ
3	4	2	3	4	1 + 3 = 4	แพ้
4	3	3	2	3	2 + 3 = 5	ชนะ
5	2	3	3	3	2 + 3 = 5	ชนะ
6	1	2	5	5	2 + 3 = 5	แพ้
7	4	3	1	4	2 + 3 = 5	ชนะ
8	5	3	1	5	1 + 3 = 4	แพ้
9	1	1	4	4	4 + 3 = 7	ชนะ
10	0	5	3	5	2 + 3 = 5	แพ้
11	3	2	3	3	2 + 3 = 5	ชนะ
12	2	2	3	3	3 + 3 = 6	ชนะ
13	3	3	2	3	2 + 3 = 5	ชนะ
14	4	3	2	4	1 + 3 = 4	แพ้
15	4	3	0	4	3 + 3 = 6	ชนะ
16	3	2	2	3	3 + 3 = 6	ชนะ
17	1	5	2	5	2 + 3 = 5	แพ้
18	5	1	1	5	3 + 3 = 6	ชนะ
19	0	4	4	4	2 + 3 = 5	ชนะ
20	0	4	4	4	2 + 3 = 5	ชนะ

ตารางที่ 3.2: ความถี่ของแต่ละตัวเลือกของตารางที่ 3.1

จำนวนครั้งการจำลองเหตุการณ์ที่เหมาะสม

การจำลองเหตุการณ์สำหรับงานวิจัยนี้เลือกใช้การจำลองเหตุการณ์แบบมอนติคาร์โล อย่างไรก็ตามการจำลองเหตุการณ์ด้วยวิธีดังกล่าวต้องมีจำนวนครั้งในการทดลองที่เพียงพอเพื่อให้ได้ผลลัพธ์ที่แม่นยำและเชื่อถือได้ จึงได้มีการหาจำนวนครั้งในการจำลองเหตุการณ์ เมื่อมีการจำลองเหตุการณ์การออกเสียงเป็นจำนวน i ครั้ง จะได้ว่ามี j ครั้งที่ซิบิลชนะการออกเสียง ดังนั้นความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจึงมีค่าเท่ากับอัตราส่วนจำนวนครั้งที่ซิบิลชนะการออกเสียงหารด้วยจำนวนครั้งการทดลองซึ่งมีค่าเท่ากับ $\frac{j}{i}$ นำค่าความน่าจะเป็นดังกล่าวมาสังเกตแนวโน้มความเปลี่ยนแปลง ดังตัวอย่างในรูปที่ 3.2 โดยแกนนอนคือ i และแกนตั้งคือ $\frac{j}{i}$ เพื่อหาจำนวนครั้งที่เหมาะสมในการคำนวณหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง ในกรณีที่มีตัวเลือกทั้งหมด 4 ตัวเลือก มีผู้ออกเสียงจริงทั้งหมด 10 คน มีซิบิลทั้งหมด 3 ตัว หากอัตราส่วนที่คำนวณได้มีความแตกต่างกันไม่เกิน 0.001 ใน 1000 การทดลองสุดท้ายถือว่าอัตราส่วนจำนวนครั้งที่ซิบิลชนะลู่เข้าสู่ความน่าจะเป็นที่ซิบิลชนะแล้ว และสามารถใช้เป็นตัวกำหนดจำนวนครั้งในการจำลองเหตุการณ์

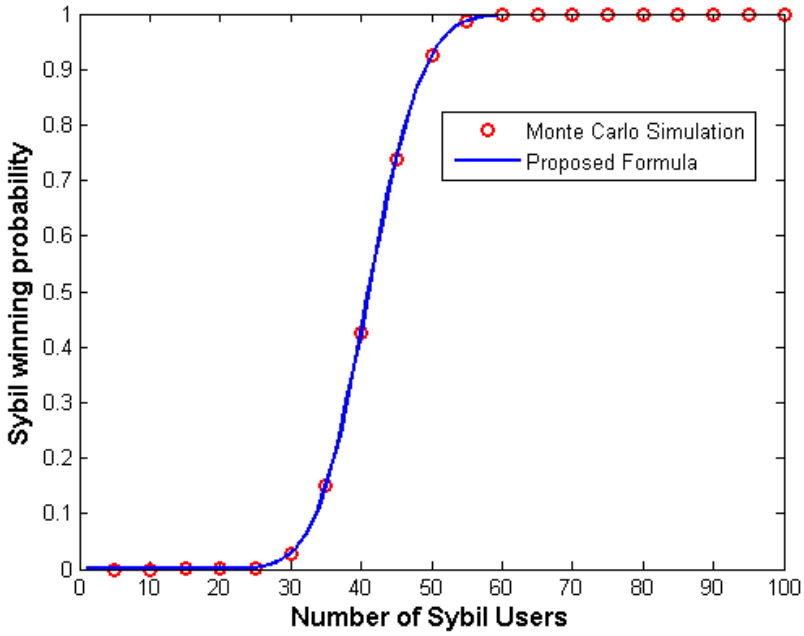


รูปที่ 3.2: ตัวอย่างการจำลองเหตุการณ์แบบมอนติคาร์โล

จากรูปที่ 2 แสดงให้เห็นว่าจำนวนเหตุการณ์ไม่ต่ำกว่า 23,710 ครั้งในการทดลองจะสามารถให้ผลที่เที่ยงตรงโดยมีความคลาดเคลื่อนไม่เกิน 0.1% ได้ จากตัวอย่างให้ผลการคำนวณที่ 23,710 ครั้งการทดลองเป็น 0.6927 เมื่อลองเปรียบเทียบกับผลการจำลองเหตุการณ์ที่ 1 ล้านครั้งจะให้ผลเป็น 0.6925 และผลจากสูตรให้คำตอบเป็น 0.6926 อย่างไรก็ตามการทดลองนี้มีจุดประสงค์เพื่อหาจำนวนครั้งการจำลองที่น้อยที่สุดที่จะให้ผลการจำลองเหตุการณ์คลาดเคลื่อนน้อยที่สุดเท่านั้น แต่สำหรับงานวิจัยนี้ใช้การจำลองเหตุการณ์ 1 ล้านครั้งต่อผลการจำลองเหตุการณ์ 1 จุดเพื่อความมั่นใจในผลการจำลองเหตุการณ์ และการจำลองเหตุการณ์นี้ใช้เวลาน้อยมากอยู่แล้ว (ได้ผลการจำลองเหตุการณ์ 1 ล้านครั้งภายในไม่กี่วินาที) จึงทำให้สามารถเพิ่มจำนวนครั้งการจำลองเหตุการณ์ให้มากได้

3.4.1 การประเมินความถูกต้องของสูตรแม่นตรง

สูตรที่ (3.12) และ (3.13) เป็นสูตรแบบแม่นตรงที่ใช้สำหรับคำนวณหาความน่าจะเป็นที่ซิบิลชนะการออกเสียง สูตรดังกล่าวถึงแม้จะมีความความซับซ้อนของสูตรอยู่ในระดับ $O\left((n+S)^k\right)$ แต่ก็มีความแม่นยำค่อนข้างสูง ดังแสดงตัวอย่างในรูปที่ 3.3 ซึ่งเป็นการเปรียบเทียบผลการคำนวณจากสูตรที่ (3.12) กับการจำลองเหตุการณ์จริง ในกรณีที่มีผู้ใช้งานจริงจำนวน 100 คน จำนวนตัวเลือก 3 ตัวเลือก ตัวเลือกแรกมีความน่าจะเป็นที่ผู้ออกเสียงจริงจะเลือกเป็น 0.4 ตัวเลือกที่สองมีความน่าจะเป็นที่ผู้ออกเสียงจริงจะเลือกเป็น 0.5 และตัวเลือกสุดท้ายมีความน่าจะเป็นที่ผู้ออกเสียงจริงจะเลือกเป็น 0.1 โดยกำหนดให้ซิบิลทุกตัวเลือกตัวเลือกที่ 3 แกนนอนของกราฟคือจำนวนซิบิลและแกนตั้งของกราฟเป็นความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

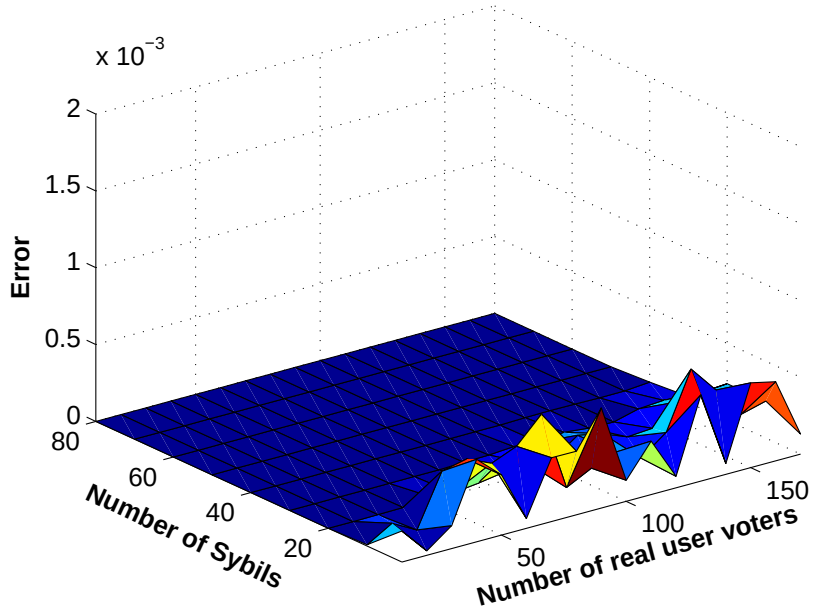


รูปที่ 3.3: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรแม่นตรง

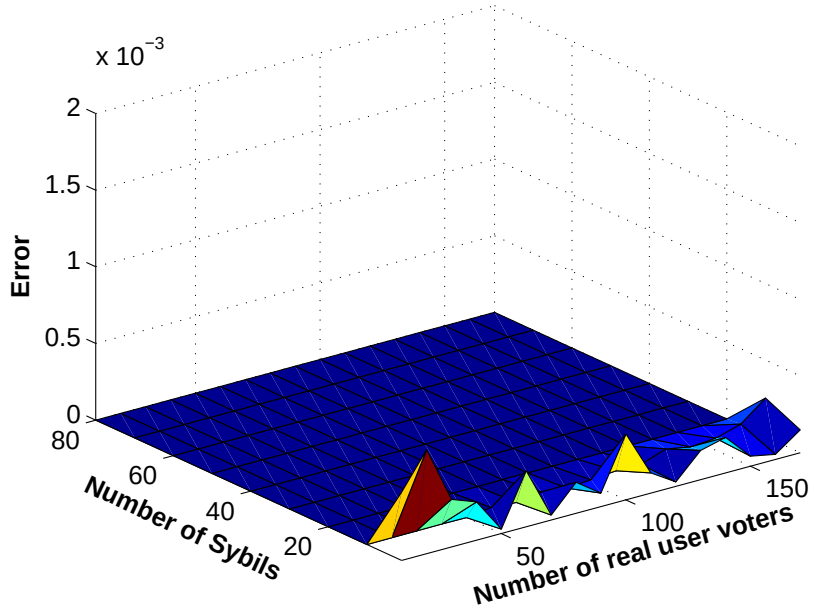
สูตรที่ (3.13) ถูกเปรียบเทียบกับการจำลองเหตุการณ์จริงในกรณีที่แต่ละตัวเลือกมีความนิยมเท่ากันพบว่ามีความแม่นยำสูงไม่ต่างไปจากสูตรที่ (3.12) ผลต่างระหว่างการจำลองเหตุการณ์จริงกับสูตรที่ (3.13) เมื่อกำหนดให้มีจำนวนผู้ออกเสียงจริงไม่เกิน 170 คน และมีจำนวนซิบิลไม่เกิน 80 คน ถูกนำเสนอในรูปที่ 3.4 และรูปที่ 3.5 ในกรณีที่มีจำนวนตัวเลือกน้อย $(k = 3)$ และกรณีที่มีจำนวนตัวเลือกมาก $(k = 30)$ ตามลำดับ ผลการเปรียบเทียบดังกล่าวแสดงให้เห็นว่าสูตรที่นำเสนอกับการจำลองเหตุการณ์จริงมีความใกล้เคียงกันมาก โดยมีความคาดเคลื่อนกันมากที่สุดไม่เกิน 0.05% เท่านั้น

3.4.2 การประเมินความถูกต้องของการประมาณค่าแบบที่ 1 (การประมาณค่าด้วยการแจกแจงปัวส์ซอง)

สูตรที่ (3.23) ถูกเปรียบเทียบกับการจำลองเหตุการณ์จริงในกรณีที่แต่ละตัวเลือกมีความนิยมเท่ากัน ผลต่างระหว่างการจำลองเหตุการณ์จริงกับสูตรที่ (3.23) เมื่อกำหนดให้มีจำนวนผู้ออกเสียง

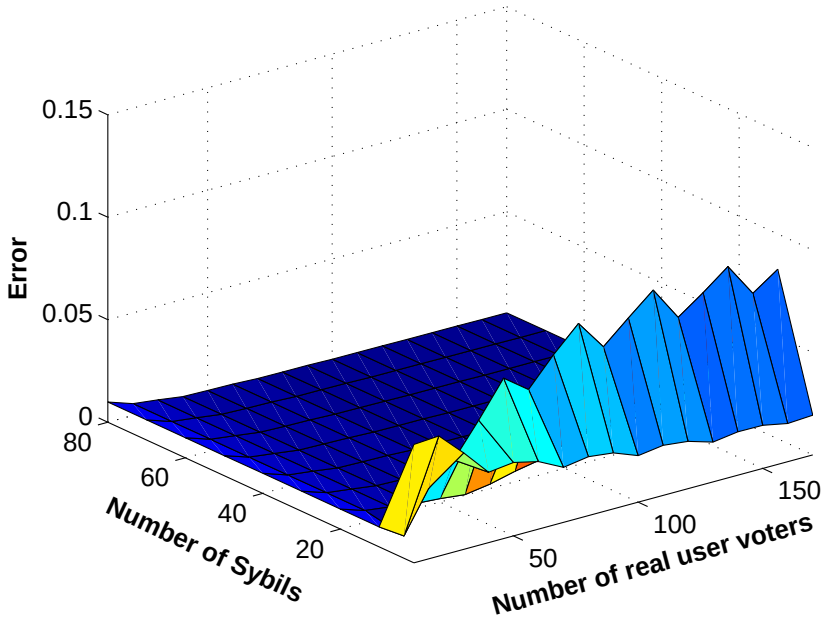


รูปที่ 3.4: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรแม่นตรงกรณีที่มีตัวเลือกน้อย $(k = 3)$



รูปที่ 3.5: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรแม่นตรงกรณีที่มีตัวเลือกมาก $(k = 30)$

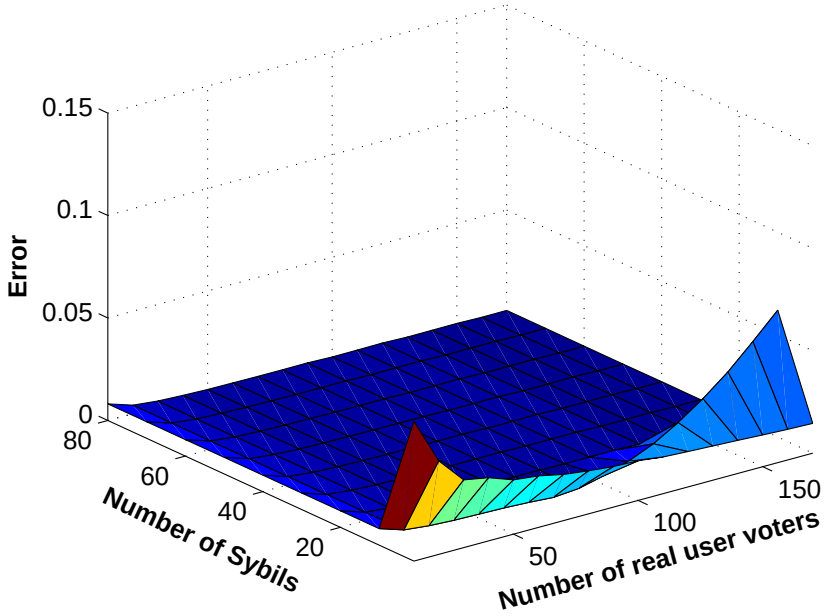
จริงไม่เกิน 170 คน และมีจำนวนซิบิลไม่เกิน 80 ตัวตน ถูกนำเสนอในรูปที่ 3.6 และรูปที่ 3.7 ในกรณีที่มีจำนวนตัวเลือกน้อย $(k = 3)$ และกรณีที่มีจำนวนตัวเลือกมาก $(k = 30)$ ตามลำดับ ผลการเปรียบเทียบดังกล่าวแสดงให้เห็นว่าสูตรที่นำเสนอกับการจำลองเหตุการณ์จริงมีความคลาดเคลื่อนค่อนข้างสูงเมื่อเทียบกับสูตร (3.13) เพราะการประมาณค่าจากสูตรที่ (3.18) สามารถใช้ได้ในกรณีที่จำนวนคะแนนเสียงในตัวเลือกของซิบิล $v_k + S$ มีค่ามากเท่านั้น เพื่อความเข้าใจถึงสาเหตุของความผิดพลาดดังกล่าว จึงเปลี่ยนรูปสมการที่ (3.18) ดังนี้



รูปที่ 3.6: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรการประมาณค่าด้วยการแจกแจงปัวส์ซองกรณีที่มีตัวเลือกน้อย $(k = 3)$

$$f_p\left(n - v_k, k - 1\right) - \frac{\frac{1}{(n-v_k)!}\lim_{x\to 0}\frac{d^{n-v_k}}{dx^{n-v_k}}\left[\left(\sum_{v_i=0}^{v_k+S-1}\frac{x^{v_i}}{v_i!}\right)^{k-1}\right]}{\left(\sum_{v_i=0}^{v_k+S-1}\frac{1}{v_i!}\right)^{k-1}} \approx 0 \tag{3.29}$$

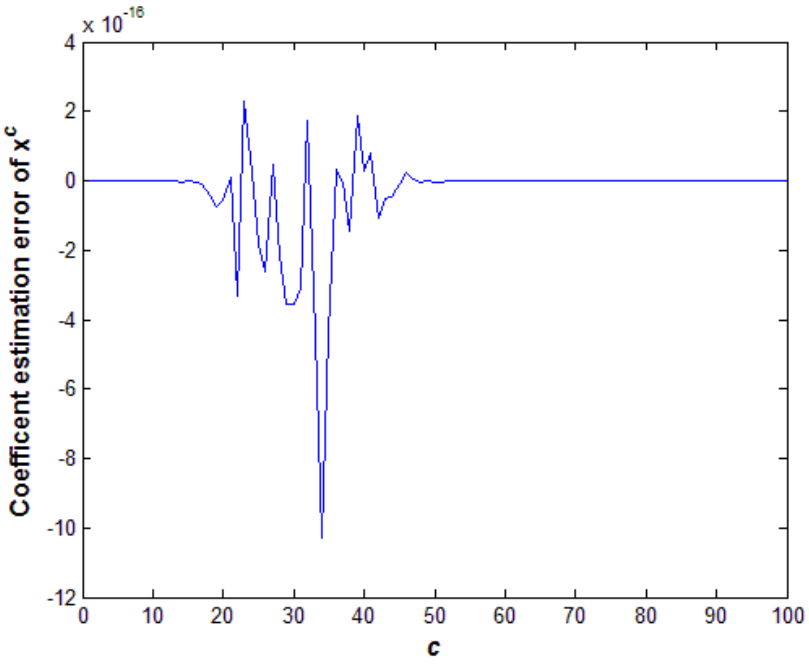
เนื่องจากทั้งฟังก์ชันปัวส์ซองและเศษส่วนในการประมาณค่าปัวส์ซองต่างเป็นฟังก์ชันที่มีจุดเปลี่ยนโค้งที่เดียว ทำให้ผลลบในอสมการที่ (3.29) ควรมีกราฟที่มีจุดยอดเพียง 2 จุดเท่านั้น แต่จากตัวอย่างผลการทดสอบเมื่อกำหนดให้จำนวนคะแนนเสียงที่มากที่สุดที่ซิบิลชนะการออกเสียง $(v_k + S - 1)$ มีค่าเท่ากับ 80 จำนวนตัวเลือกของผู้ใช้งานจริง $(k - 1)$ เท่ากับ 30 และกำหนดให้จำนวนคะแนนรวมทุกตัวเลือกของผู้ใช้งานจริง $(c = n - v_k)$ เป็นพารามิเตอร์ในแกนนอนของกราฟ ดังรูปที่ 3.8 พบว่าผลลบกลับให้ค่าที่มีหลายจุดยอดและอยู่ในระดับ 10^{-16} ซึ่งตรงกับระดับความคลาดเคลื่อนในการคำนวณตัวเลขของ MATLAB version R2009a 64-bit ที่ใช้ในวิทยานิพนธ์



รูปที่ 3.7: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรการประมาณค่าด้วยการแจกแจงปัวส์ซองกรณีที่มีตัวเลือกมาก $(k = 30)$

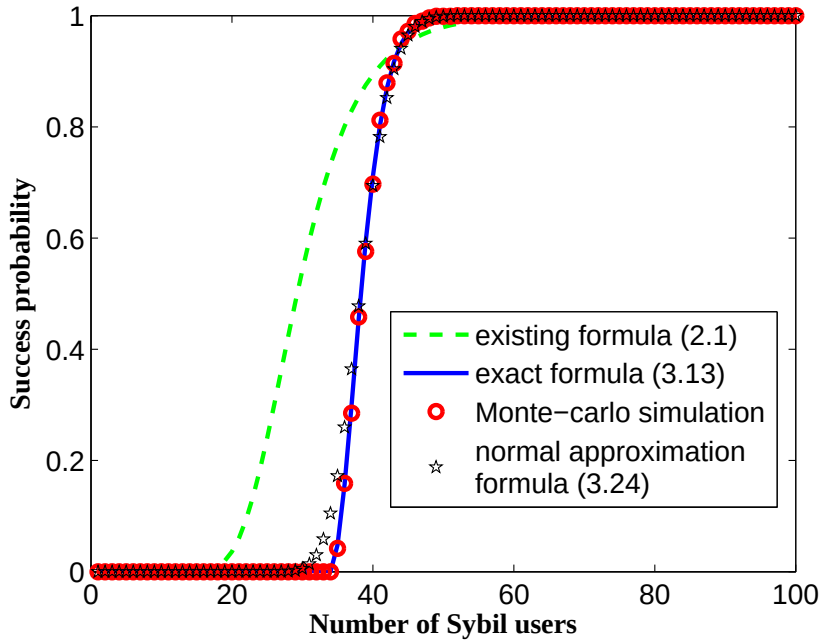
นี้ ดังนั้นจึงสรุปได้ว่าค่าความคลาดเคลื่อนดังกล่าวไม่ได้เกิดจากการประมาณค่าทางคณิตศาสตร์ แต่เกิดจากความคลาดเคลื่อนเครื่องคำนวณหรือซอฟท์แวร์ที่ใช้คำนวณ ทั้งนี้ความคลาดเคลื่อนแม้เพียงเล็กน้อยของการประมาณค่าในสูตรที่ (3.18) จะได้รับการขยายผลเมื่อต้องคูณกับพจน์อื่น ทำให้เกิดความคลาดเคลื่อนมากขึ้นดังรูปที่ 3.6 และรูปที่ 3.7 อย่างไรก็ตามความซับซ้อนของสูตรการประมาณค่าแบบที่ 1 นี้อยู่ในระดับ $O(n)$ ซึ่งมีความเหมาะสมในทางปฏิบัติมากกว่าสูตรที่ (3.13) ในกรณีที่มีตัวเลือกและจำนวนซิบิลเป็นจำนวนมาก

สำหรับกรณีที่ไม่มีผู้ออกเสียงจริงคนใดเลือกตัวเลือกของซิบิลเลย จะได้ว่าการคำนวณความน่าจะเป็นที่ซิบิลชนะการออกเสียงกรณีที่ทุกตัวเลือกมีความน่าจะเป็นเท่ากันที่จะถูกผู้ออกเสียงจริงเลือก แต่ตัวเลือกที่ซิบิลเลือกไม่มีผู้ออกเสียงจริงเลือกด้วยเลยสามารถประยุกต์จากสูตรที่มีอยู่แล้วในอดีต [27] ดังรายละเอียดของสูตรในสมการที่ (2.1) – (2.3) ในหัวข้อ 2.2.2 ซึ่งสามารถเปรียบเทียบกับสูตรที่วิทยานิพนธ์นี้นำเสนอได้ โดยกำหนดให้ $v_k = 0$ เสมอ การเปรียบเทียบระหว่างสูตรจาก [27] กับสูตรที่นำเสนอตามสมการ (3.15) แสดงในรูปที่ 3.9 ในกรณีตัวอย่างนี้มีจำนวนผู้ออกเสียงจริงมีทั้งสิ้น 100 คน และมีจำนวนตัวเลือก 5 ตัวเลือก โดยแกนนอนคือจำนวนของซิบิลและแกนตั้งเป็นความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง ทั้งนี้สังเกตว่าทั้งสองวิธีแสดงแนวโน้มลักษณะของกราฟที่เหมือนกัน แต่จากกราฟดังกล่าวสามารถชี้ชัดถึงความเหมาะสมในการใช้งานของสูตรที่ไม่เหมือนกัน ในกรณีที่จำนวนผู้ออกเสียงจริงมี 100 คน มีจำนวนตัวเลือกทั้งหมด 5 ตัวเลือก ตัวเลือกที่มีผู้ออกเสียงจริงมากที่สุด จะมีคะแนนเสียงต่ำกว่า 20 ไม่ได้ ดังนั้นในกรณีที่ซิบิลมีจำนวนไม่ถึง 20 ตัว และข้อที่ซิบิลเลือกไม่มีผู้ออกเสียงจริงเลือกด้วยเลย จะไม่มีทางที่ซิบิลจะชนะการออกเสียงได้ จากกราฟพบว่าสูตรที่ได้ประยุกต์มาจากอดีตบ่งชี้ว่าซิบิลมีโอกาสชนะการออกเสียงในกรณีดังกล่าว ซึ่งไม่ถูกต้อง อย่างไรก็ตามสูตรที่ประยุกต์มาจากอดีตเป็นสูตรที่ไม่ติดฟังก์ชันแฟคทอเรียล ดังนั้นจึง



รูปที่ 3.8: ตัวอย่างผลต่างของการคำนวณสัมประสิทธิ์ของ x^c ในพหุนาม $\frac{\left(\sum_{j=0}^{v_k+S-1} \frac{x^j}{j!}\right)^{k-1}}{\left(\sum_{j=0}^{v_k+S-1} \frac{1}{j!}\right)^{k-1}}$ และการประมาณค่าจากการคำนวณโดย $f_p\left(c, k-1\right)$ ในกรณีที่จำนวนคะแนนเสียงที่มากที่สุดที่ซิบิลชนะการออกเสียง $(v_k + S - 1)$ มีค่าเท่ากับ 80 จำนวนตัวเลือกของผู้ใช้งานจริง $(k - 1)$ เท่ากับ 30

สามารถคำนวณผลลัพธ์ออกมาได้แม้ในกรณีที่จำนวนผู้ออกเสียงจริงมีเป็นจำนวนมากก็ตาม ซึ่งต่างกับสูตรที่ได้นำเสนอในวิทยานิพนธ์นี้ซึ่งมีความถูกต้องสูง แต่ติดฟังก์ชันแฟคทอเรียลทำให้ไม่สามารถคำนวณในกรณีที่ผู้ออกเสียงจริงมีเป็นจำนวนมากได้ดี ทั้งนี้ขึ้นอยู่กับความสามารถของเครื่องคำนวณและซอฟท์แวร์ที่ใช้คำนวณ โดยเครื่องคำนวณและซอฟท์แวร์ที่ใช้ในวิทยานิพนธ์นี้สามารถคำนวณผลลัพธ์ในกรณีที่ผู้ออกเสียงจริงมีเป็นจำนวนไม่เกิน 170 คนเท่านั้น

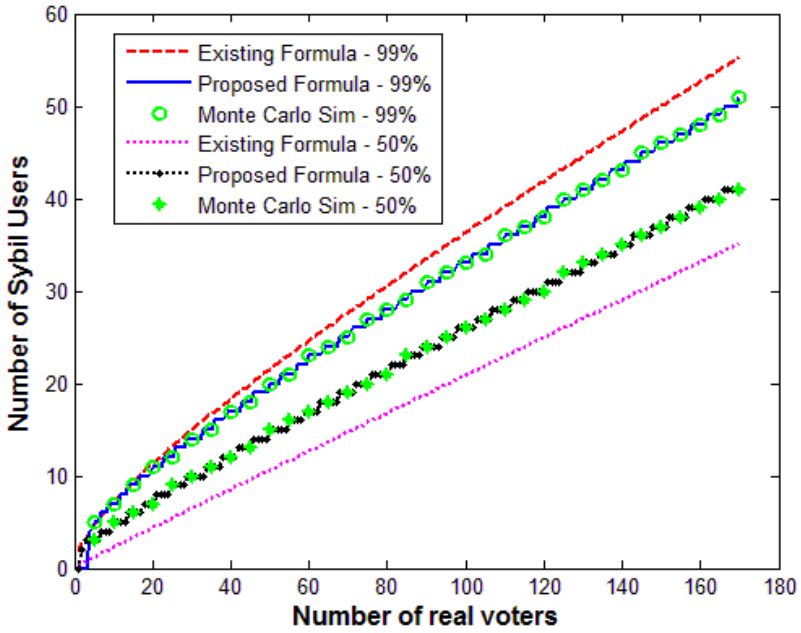


รูปที่ 3.9: ตัวอย่างการเปรียบเทียบการคำนวณความน่าจะเป็นที่ซิบิลชนะการออกเสียงกรณีที่ทุกตัวเลือกมีความน่าจะเป็นเท่ากันที่จะถูกผู้ออกเสียงจริงเลือก และมีตัวเลือกหนึ่งที่ซิบิลเลือกไม่มีผู้ออกเสียงจริงเลือกด้วยเลยระหว่างสูตรที่ประยุกต์จากอดีตกับสูตรที่นำเสนอ

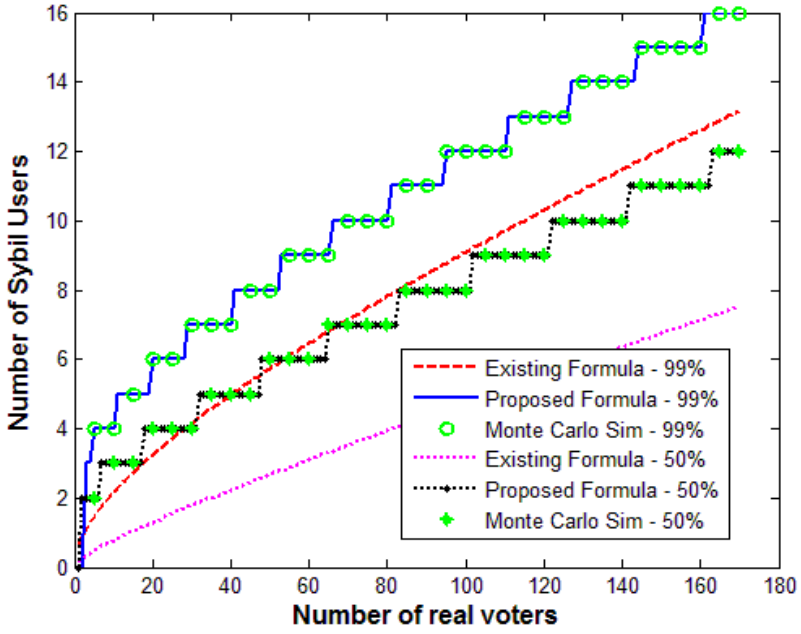
สูตรที่นำเสนอในอดีต (2.1) นอกจากจะใช้สำหรับหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงแล้ว งานวิจัย [27] ได้นำเสนอสูตรการคำนวณหาจำนวนซิบิลที่ต้องใช้เพื่อให้ได้ความน่าจะเป็นที่ต้องการ เช่นที่ 99% และ 50% ไว้ ดังสูตรในสมการ (2.3) ทั้งนี้ผลการคำนวณหาจำนวนซิบิลจากสูตรในสมการ (2.3) และสูตรที่นำเสนอในวิทยานิพนธ์นี้ ดังสมการ (3.15) โดยที่กรณี 50% ใช้สมการ (2.2) โดยแทนค่า $\alpha = 0.5$ และกรณี 99% ใช้สมการ (2.3) เปรียบเทียบกันดังแสดงในรูปที่ 3.10 และ 3.11 โดยแกนนอนคือจำนวนของผู้ออกเสียงจริง และแกนตั้งคือจำนวนของซิบิลพบว่าสูตรที่นำเสนอในวิทยานิพนธ์นี้มีความถูกต้องกว่าสูตรที่นำเสนอโดย [27]

3.4.3 การประเมินความถูกต้องของสูตรการประมาณค่าแบบที่ 2 (การประมาณค่าด้วยการแจกแจงปกติ)

สูตรที่ (3.25) ถูกเปรียบเทียบกับการจำลองเหตุการณ์จริงในกรณีที่แต่ละตัวเลือกมีความนิยมเท่ากัน พบว่าผลต่างระหว่างการจำลองเหตุการณ์จริงกับสูตรที่ (3.25) เมื่อกำหนดให้มีจำนวนผู้ออกเสียงจริงไม่เกิน 170 คน และมีจำนวนซิบิลไม่เกิน 80 คน ถูกนำเสนอในรูปที่ 3.12 และรูปที่ 3.13 ในกรณีที่มีจำนวนตัวเลือกน้อย $(k = 3)$ และกรณีที่มีจำนวนตัวเลือกมาก $(k = 30)$ ตามลำดับ ผลการ

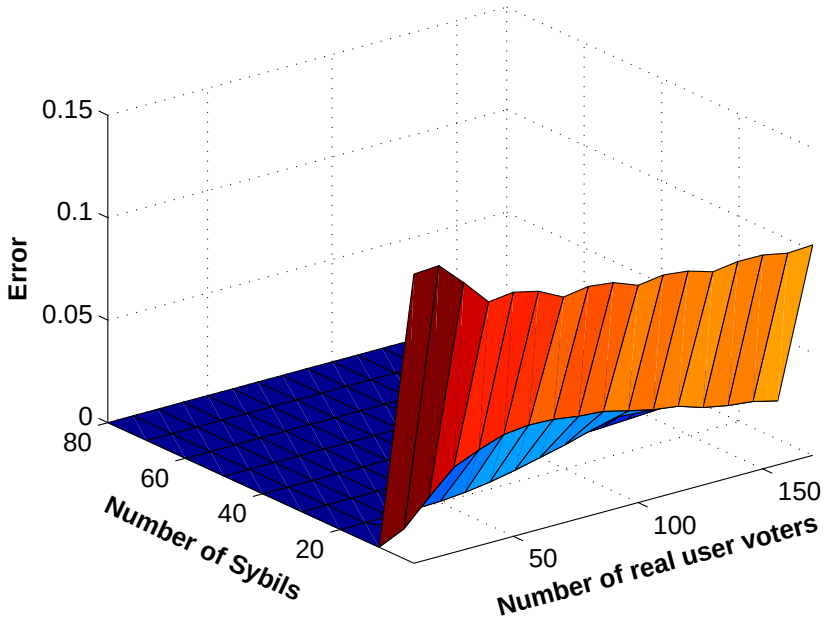


รูปที่ 3.10: จำนวนซิบิลอย่างน้อยที่สุดที่จะทำให้ความน่าจะเป็นที่ซิบิลชนะการออกเสียงไม่ต่ำกว่า 99% หรือ 50% ในกรณีที่มีจำนวนตัวเลือก 5 ตัวเลือก



รูปที่ 3.11: จำนวนซิบิลอย่างน้อยที่สุดที่จะทำให้ความน่าจะเป็นที่ซิบิลชนะการออกเสียงไม่ต่ำกว่า 99% หรือ 50% ในกรณีที่มีจำนวนตัวเลือก 30 ตัวเลือก

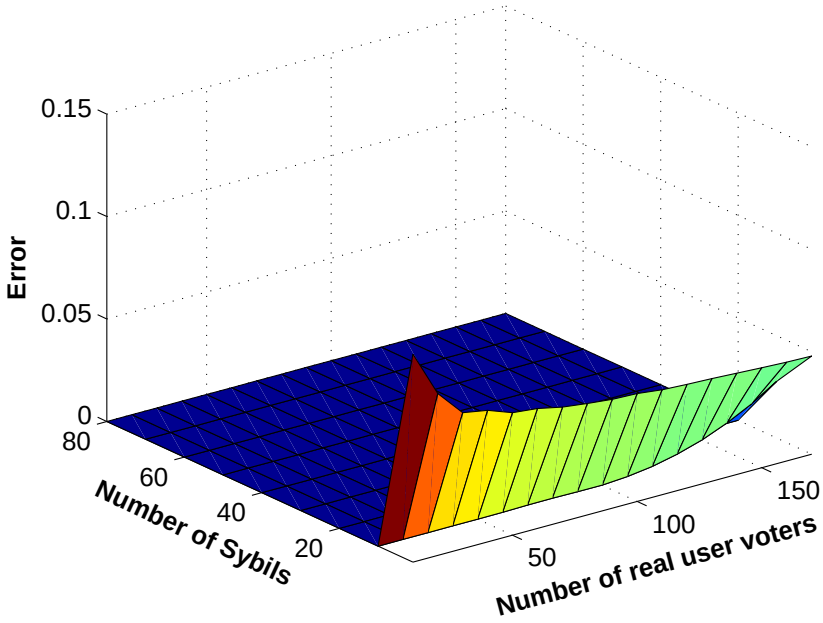
เปรียบเทียบดังกล่าวแสดงให้เห็นว่าสูตรที่นำเสนอกับการจำลองเหตุการณ์จริงมีความคลาดเคลื่อนมากกว่าสูตรแบบแม่นตรง เนื่องจากมีการผ่อนผันสมมุติฐานที่ว่าคะแนนเสียงแต่ละตัวเลือกไม่มีความอิสระต่อกันออกไป ทั้งนี้ความซับซ้อนของสูตรการคำนวณแบบประมาณค่าแบบที่ 2 นี้อยู่ในระดับ $O(1)$ เท่านั้น อีกทั้งความคลาดเคลื่อนดังกล่าวแตกต่างจากสูตรแบบประมาณค่าแบบที่ 1 ไม่มาก ทำให้สูตรการประมาณค่าแบบที่ 2 นี้เป็นอีกทางเลือกหนึ่งที่อาจจะนำไปประยุกต์ใช้จริงในทางปฏิบัติได้



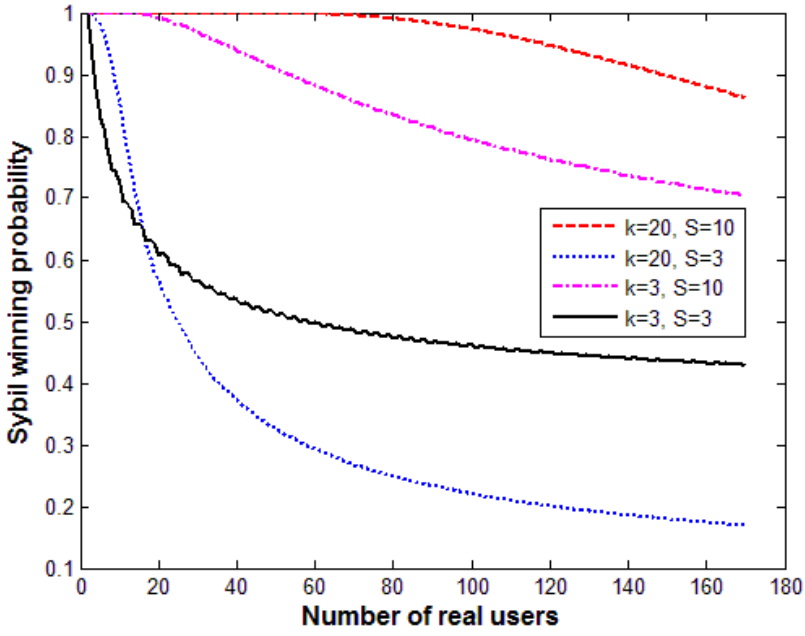
รูปที่ 3.12: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรการประมาณค่าด้วยการแจกแจงปกติกรณีที่มีตัวเลือกน้อย $(k = 3)$

3.5 ผลกระทบของตัวแปรต่าง ๆ

ในหัวข้อที่ 3.4 ได้กล่าวถึงการประเมินความถูกต้องของสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงไปแล้ว โดยสรุปสูตรแบบแม่นตรงให้ความแม่นยำสูงโดยมีความคลาดเคลื่อนจากการจำลองเหตุการณ์จริง 1 ล้านครั้ง ไม่เกิน 0.1% เมื่อเทียบกับผลการจำลองเหตุการณ์จริง สำหรับหัวข้อที่ 3.5 นี้จะกล่าวถึงผลกระทบของตัวแปรต่าง ๆ ได้แก่ จำนวนผู้ออกเสียงจริง (n) จำนวนตัวเลือก (k) จำนวนซิบิล (S) และจำนวนผู้ออกเสียงจริงที่เลือกออกเสียงเป็นตัวเลือกเดียวกันกับซิบิล (v_k) ที่มีต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง ทั้งนี้อ้างอิงหัวข้อที่ 3.4 ที่รับรองความถูกต้องและความแม่นยำของสูตรการคำนวณแบบแม่นตรงแล้ว จึงไม่ขอนำเสนอกราฟการจำลองเหตุการณ์จริงหรือสูตรแบบประมาณค่าอีก แต่จะเป็นการนำสูตรแบบแม่นตรงที่นำเสนอมาใช้เพื่อการประยุกต์ต่อไป



รูปที่ 3.13: ตัวอย่างผลการจำลองเหตุการณ์จริงกับสูตรการประมาณค่าด้วยการแจกแจงปกติกรณีที่มีตัวเลือกมาก $(k = 30)$



รูปที่ 3.14: ผลกระทบของจำนวนผู้ออกเสียงจริงต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

3.5.1 ผลกระทบของจำนวนผู้ออกเสียงจริง (n)

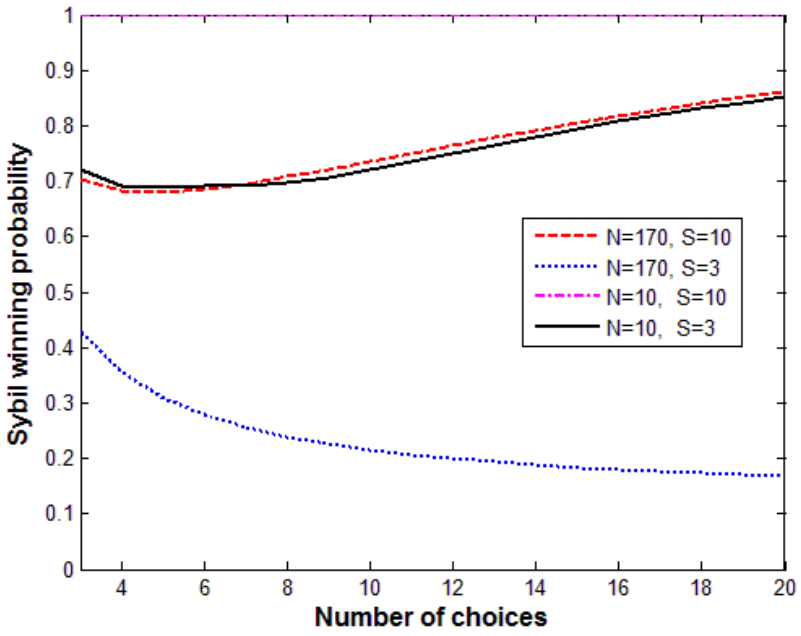
จำนวนผู้ออกเสียงจริง จำนวนตัวเลือก และจำนวนซิบิลมีความสัมพันธ์กันดังรูปที่ 3.14 โดยแกนนอนคือจำนวนผู้ออกเสียงจริง และแกนตั้งคือความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง กราฟประกอบด้วย 4 เส้น แสดงความสัมพันธ์ใน 2 มิติ ทั้งมิติเรื่องจำนวนตัวเลือกและจำนวนซิบิลกำหนดให้จำนวนตัวเลือกมากคือมีตัวเลือก 20 ตัวเลือกและจำนวนตัวเลือกน้อยกำหนดให้มี 3 ตัวเลือก และกำหนดให้จำนวนซิบิลมากมี 10 ตัว และจำนวนซิบิลน้อยมี 3 ตัว ภาพรวมของกราฟทั้ง 4 เส้นให้ผลที่มีแนวโน้มแบบเดียวกันคือ ยิ่งมีจำนวนผู้ออกเสียงจริงมากยิ่งส่งผลให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงมีน้อย เมื่อสังเกตเส้นกราฟที่มีจำนวนตัวเลือกเท่ากันพบว่าความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงมีค่าสูงเมื่อมีซิบิลเป็นจำนวนมาก กล่าวคือสำหรับกรณี 20 ตัวเลือก เส้นสีแดงอยู่สูงกว่าเส้นสีน้ำเงินเสมอ และกรณี 3 ตัวเลือกเส้นสีชมพูอยู่เหนือเส้นสีดำเสมอในทุกจำนวนของผู้ออกเสียงจริง ในมุมมองที่กำหนดให้มีจำนวนซิบิลเท่ากันพบว่าจำนวนตัวเลือกไม่ได้ส่งผลกระทบโดยตรงต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง สังเกตว่าเส้นสีแดงอยู่เหนือเส้นสีชมพูเสมอ ดังนั้นจึงสรุปได้ว่าการเพิ่มขึ้นของจำนวนตัวเลือกจะส่งผลกระทบให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพิ่มขึ้นไปด้วยในกรณีที่ซิบิลมีเป็นจำนวนมาก เพราะการมีจำนวนตัวเลือกเป็นจำนวนมากจะทำให้คะแนนเสียงเฉลี่ยของแต่ละตัวเลือกลดลงและความแปรปรวนของคะแนนเสียงต่ำลงด้วยทำให้ความน่าจะเป็นที่ตัวเลือกที่ซิบิลไม่ได้เลือกมีคะแนนเทียบเท่ากับตัวเลือกที่ซิบิลเลือกน้อยลง ส่งผลให้ความน่าจะเป็นที่ซิบิลชนะการออกเสียงมากขึ้น

ในกรณีที่มีจำนวนซิบิลน้อย (เส้นสีน้ำเงินและเส้นสีดำ) แม้ว่าทั้งสองกราฟจะมีแนวโน้มเป็นฟังก์ชันลดในภาพรวมและเส้นสีน้ำเงินควรอยู่ใต้เส้นสีดำตลอดเวลา แต่จากผลการทดลองพบว่าเส้นสีน้ำเงินยกตัวสูงขึ้นเหนือเส้นสีดำในกรณีที่มีซิบิลเป็นจำนวนน้อยและมีจำนวนผู้ออกเสียงจริงเป็นจำนวนน้อยเมื่อเทียบกับจำนวนตัวเลือก สาเหตุเกิดจากเมื่อมีจำนวนผู้ออกเสียงเป็นจำนวนน้อยเมื่อเทียบกับจำนวนตัวเลือก ความน่าจะเป็นที่ผู้ออกเสียงจริงจะชนะซิบิลจะน้อยมากเนื่องจากค่าเฉลี่ยของคะแนนเสียงของผู้ออกเสียงจริงต่อหนึ่งตัวเลือกมีค่าไม่ถึง 1 ซึ่งน้อยกว่าจำนวนซิบิลอยู่มาก ทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจึงเพิ่มขึ้นเป็นพิเศษ กราฟเส้นสีน้ำเงินที่ควรจะอยู่ใต้เส้นสีดำตลอดจึงมีบางส่วนที่ยกตัวสูงขึ้นเหนือเส้นสีดำได้ โดยจากผลการทดลอง ช่วงที่มีจำนวนผู้ออกเสียงจริงไม่เกิน 16 คนเส้นสีน้ำเงินอยู่สูงกว่าเส้นสีดำ และช่วงที่มีผู้ออกเสียงเกิน 16 คนเส้นสีดำอยู่สูงกว่าเส้นสีน้ำเงิน

ประเด็นที่น่าสนใจอีกประเด็นหนึ่งคือลักษณะของเส้นสีดำที่เป็นขยัก ทำให้ทราบว่ากรณีที่จำนวนตัวเลือกและจำนวนซิบิลมีจำนวนน้อย การเพิ่มจำนวนผู้ออกเสียงจริงอาจทำให้ความน่าจะเป็นที่ซิบิลชนะการออกเสียงเพิ่มขึ้นได้ในบางครั้ง สาเหตุเนื่องจากผู้ออกเสียงจริงที่เพิ่มขึ้นมีโอกาสสูงที่จะเลือกข้อเดียวกันกับซิบิลเพราะมีจำนวนตัวเลือกน้อย รวมถึงมีความไวต่อความเปลี่ยนแปลงสูงเนื่องจากมีจำนวนซิบิลน้อย

3.5.2 ผลกระทบของจำนวนตัวเลือก (k)

จำนวนตัวเลือกเป็นปัจจัยหนึ่งที่ส่งผลกระทบต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง รูปที่ 3.15 แสดงผลกระทบของจำนวนตัวเลือกที่มีผลต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง แกนนอนคือจำนวนตัวเลือกและแกนตั้งคือความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง กราฟประกอบด้วย 4 เส้น แสดงความสัมพันธ์ใน 2 มิติ ทั้งมิติเรื่องจำนวนผู้ออกเสียงจริงและจำนวนซิบิลกำหนดให้จำนวนผู้ออกเสียงจริงมีค่ามากคือมีผู้ออกเสียงจริง 170 คนและจำนวนผู้ออกเสียงจริงน้อยกำหนดให้มีผู้ออกเสียงจริง 10 คน และกำหนดให้จำนวนซิบิลมากมี 10 ตัว และจำนวนซิบิลน้อยมี



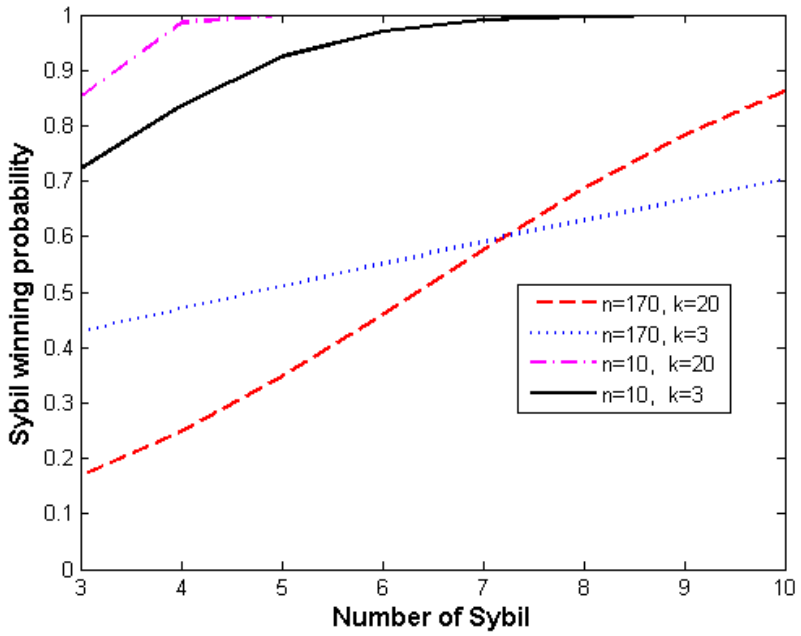
รูปที่ 3.15: ผลกระทบของจำนวนตัวเลือกต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

3 ตัว จากกราฟเส้นสีชมพูซึ่งเป็นกรณีที่จำนวนผู้ออกเสียงจริงกับจำนวนซิบิลมีเท่ากันเป็นกรณีเด่นชัด (obvious solution) ที่ความน่าจะเป็นที่ซิบิลชนะการออกเสียงเป็น 1 เพราะไม่ว่าจะมีกี่ตัวเลือกก็ตาม ก็ไม่มีทางที่จะมีตัวเลือกใดมีจำนวนเสียงมากกว่าจำนวนเสียงในข้อที่ซิบิลเลือกได้ ในกรณีที่มีซิบิลเป็นจำนวนน้อยเมื่อเทียบกับจำนวนผู้ออกเสียงจริง (เส้นสีน้ำเงิน) พบว่ายิ่งเพิ่มจำนวนตัวเลือกมาก ก็ยิ่งลดความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง เพราะยิ่งมีจำนวนตัวเลือกมากก็ยิ่งมีโอกาสสูงที่จะมีอย่างน้อย 1 ข้อที่มีคะแนนเสียงมากกว่าข้อที่ซิบิลเลือก

สำหรับกรณีที่เหลือ คือกรณีที่ไม่สุดขั้วคือจำนวนผู้ออกเสียงจริงกับจำนวนซิบิลต่างกันไม่มากเกินไป การเพิ่มจำนวนตัวเลือกจะส่งผลให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงลดลงเมื่อจำนวนตัวเลือกน้อย และจะเพิ่มขึ้นเมื่อเพิ่มจำนวนตัวเลือก เพราะการเพิ่มตัวเลือกจะทำให้โอกาสที่อย่างน้อยหนึ่งตัวเลือกมีเสียงมากกว่าตัวเลือกที่ซิบิลเลือกในกรณีที่จำนวนตัวเลือกมีน้อย แต่จะเพิ่มความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในกรณีที่มีจำนวนตัวเลือกมาก เพราะคะแนนเสียงเฉลี่ยในแต่ละตัวเลือกจะลดลง

3.5.3 ผลกระทบของจำนวนซิบิล (S)

จำนวนของซิบิลมีผลกระทบต่อความน่าจะเป็นที่ซิบิลชนะการออกเสียง โดยความน่าจะเป็นที่ซิบิลชนะการออกเสียงจะมากขึ้นตามจำนวนของซิบิล อย่างไรก็ตามจำนวนของซิบิลก็ไม่ได้เป็นตัวแปรเดียวที่ส่งผลกระทบต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง ดังแสดงในรูปที่ 3.16 ที่แสดงผลกระทบของจำนวนซิบิลต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง แกนนอนคือจำนวนซิบิลและแกนตั้งคือความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง เส้นประสีแดงและเส้นจุดสีน้ำเงินแทนกรณีที่มีจำนวนผู้ออกเสียงจริงเป็นจำนวนมาก (170 คน) เส้นประสลับจุดสีชมพูและเส้นทึบดำแทนกรณีที่มีผู้ออกเสียงจริงเป็นจำนวนน้อย (10 คน) เส้นประสีแดงและเส้นประสลับจุดสีชมพู

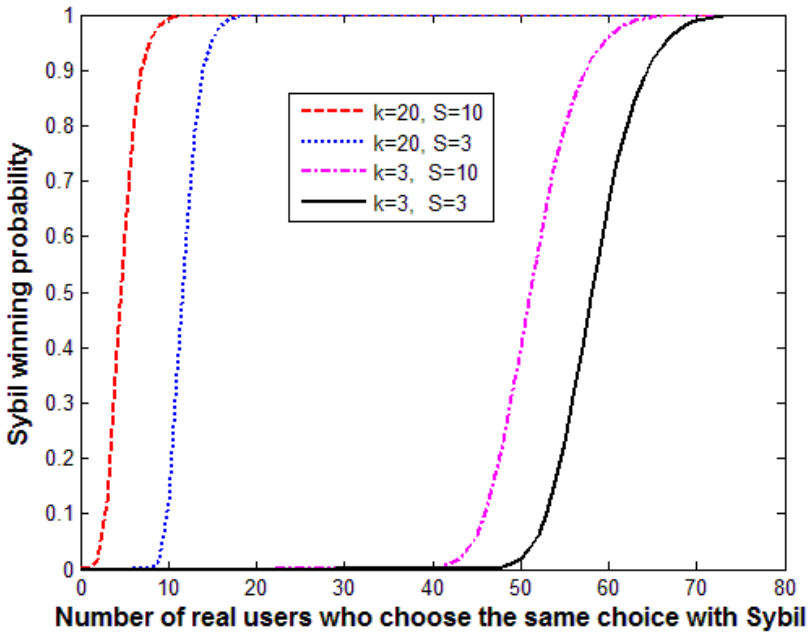


รูปที่ 3.16: ผลกระทบของจำนวนซิบิลต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

แทนกรณีที่มีจำนวนตัวเลือกมาก (20 ตัวเลือก) เส้นจุดสีน้ำเงินและเส้นทึบดำแทนกรณีที่มีจำนวนตัวเลือกน้อย (3 ตัวเลือก) สังเกตว่าเส้นกราฟสีชมพูและสีแดงมีความชันมากกว่าเส้นสีดำและเส้นสีน้ำเงินก่อนถึงจุดอิ่มตัว (ค่าความน่าจะเป็นที่ซิบิลชนะการออกเสียงเป็น 1) ทำให้ทราบว่าความไว (Sensitive) ของอัตราการเปลี่ยนแปลงความน่าจะเป็นที่ซิบิลชนะการออกเสียงขึ้นอยู่กับจำนวนตัวเลือกด้วย โดยเมื่อมีจำนวนตัวเลือกมาก ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจะเพิ่มขึ้นอย่างรวดเร็วกว่าในกรณีที่มีจำนวนตัวเลือกน้อย สำหรับปัจจัยเรื่องจำนวนผู้ออกเสียงจริง สังเกตว่าเส้นสีแดงและเส้นสีน้ำเงิน (จำนวนผู้ออกเสียงจริงมาก) อยู่ต่ำกว่าเส้นสีชมพูและเส้นสีดำ (จำนวนผู้ออกเสียงน้อย) ดังนั้น ในกรณีที่มีจำนวนผู้ออกเสียงจริงมากจะมีความน่าจะเป็นที่ซิบิลชนะการออกเสียงน้อยกว่ามีผู้ออกเสียงจริงน้อย เพราะจำนวนผู้ออกเสียงจริงยิ่งมากก็จะยิ่งสามารถลดผลกระทบจากการโจมตีได้ กล่าวคือหากมีผู้ออกเสียงจริงมากซิบิลก็ยากที่จะโน้มน้าวให้ผลการออกเสียงเป็นไปตามที่ซิบิลเลือกได้

3.5.4 ผลกระทบของจำนวนผู้ออกเสียงจริงที่เลือกตัวเลือกเดียวกันกับซิบิล (v_k)

จำนวนผู้ออกเสียงจริงที่เลือกตัวเลือกเดียวกันกับซิบิลเป็นอีกตัวแปรหนึ่งที่ส่งผลกระทบต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง ซึ่งหากมีผู้ออกเสียงจริงเลือกตัวเลือกข้อเดียวกันกับซิบิลมาก ก็จะยิ่งทำให้ความน่าจะเป็นที่ซิบิลชนะการออกเสียงมากตามไปด้วย อย่างไรก็ตามเมื่อพิจารณาพร้อมกับจำนวนผู้ออกเสียงจริงทั้งหมดและจำนวนซิบิลแล้ว จะสังเกตได้ว่ามีผลกระทบเกี่ยวกัน จากรูปที่ 3.17 สังเกตว่าเส้นกราฟสีแดงและสีน้ำเงิน (กรณีที่มีจำนวนตัวเลือกมาก) จะอยู่ฝั่งซ้ายของเส้นกราฟสีชมพูและสีดำ (กรณีที่มีจำนวนตัวเลือกน้อย) ทำให้ทราบว่าจำนวนตัวเลือกทำให้จำนวนผู้ออกเสียงจริงที่เลือกตอบตัวเลือกเดียวกันกับซิบิลส่งผลกระทบได้เร็วขึ้น



รูปที่ 3.17: ผลกระทบของจำนวนผู้ออกเสียงจริงที่เลือกตัวเลือกเดียวกันกับซิบิลต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

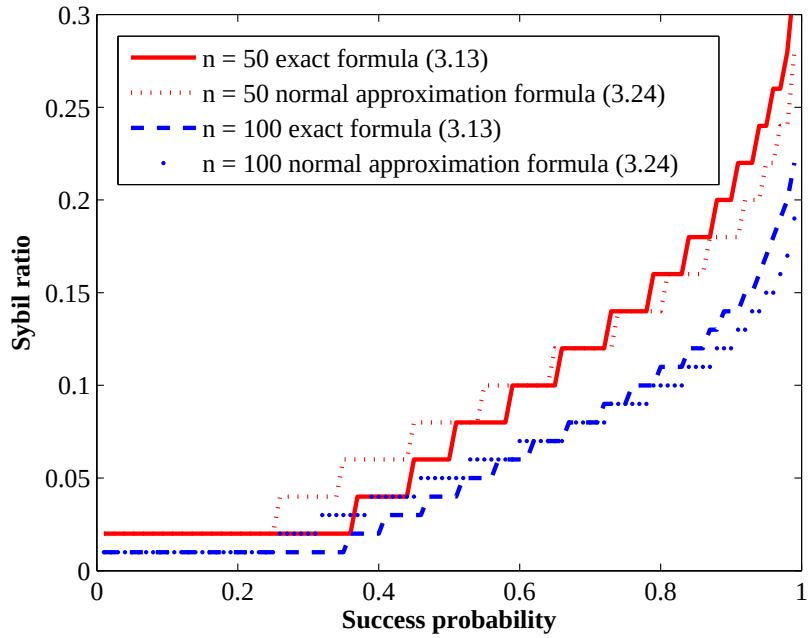
3.5.5 ผลกระทบของอัตราส่วนของจำนวนซิบิลต่อจำนวนของผู้ใช้งานจริง ($\frac{S}{n}$)

อัตราส่วนของจำนวนซิบิลต่อจำนวนของผู้ใช้งานจริงที่เพียงพอจะชนะการออกเสียงถูกนำเสนอในส่วนนี้ (ต่อจากนี้ไปจะเรียกอย่างสั้นว่า “อัตราส่วนซิบิล”) รูปที่ 3.18 และ รูปที่ 3.19 แสดงอัตราส่วนซิบิลเมื่อเปลี่ยนระดับความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงและเมื่อเปลี่ยนจำนวนผู้ใช้งานจริงเมื่อกำหนดให้ $k = 3$ และ $(p_1 = p_2 = p_3 = \frac{1}{3})$ จากรูปที่ 3.18 พบว่าการเพิ่มขึ้นของความน่าจะเป็นที่ซิบิลชนะการออกเสียงจะทำให้อัตราส่วนซิบิลเพิ่มขึ้นแบบเลขชี้กำลัง (exponentially increasing) เนื่องจากการเพิ่มขึ้นของซิบิล 1 ตัวตนมีผลกระทบต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงลดน้อยลงเมื่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพิ่มขึ้น

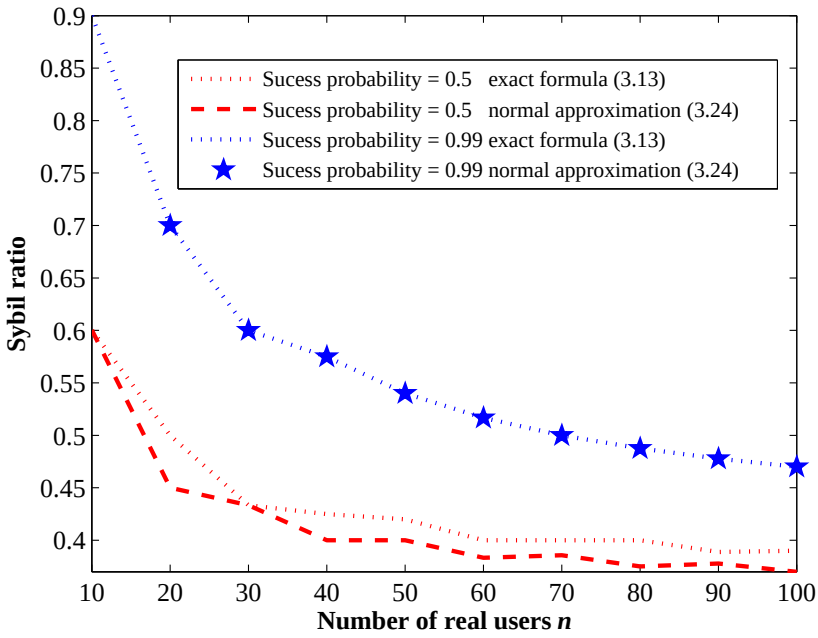
อัตราส่วนซิบิลได้รับผลกระทบจากการเพิ่มขึ้นของจำนวนผู้ใช้งานจริง ความสัมพันธ์ระหว่างอัตราส่วนซิบิลกับจำนวนผู้ใช้งานจริงในกรณีที่มี 3 ตัวเลือกถูกนำเสนอในรูปที่ 3.19 พบว่าเมื่อกำหนดให้จำนวนผู้ใช้งานจริงเป็นพารามิเตอร์พบว่าอัตราส่วนซิบิลเป็นฟังก์ชันไม่เพิ่ม (non-increasing function) เมื่อกำหนดให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเป็นค่าคงที่ สังเกตว่าการเพิ่มขึ้นของอัตราส่วนซิบิลเพียง 5% และ 25% ทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพิ่มขึ้นถึง 50% และ 99% ตามลำดับเมื่อจำนวนของผู้ใช้งานจริงมีมากกว่า 100 คน ดังนั้นจึงสรุปว่าการเพิ่มขึ้นของอัตราส่วนซิบิลเล็กน้อยสามารถส่งผลกระทบต่อผลการออกเสียงอย่างมากได้

3.6 สรุป

บทที่ 3 นี้นำเสนอวิธีการคำนวณหาผลกระทบของการโจมตีชนิดซิบิลโดยกำหนดให้มีค่าเท่ากับความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง วิธีการคำนวณดังกล่าวแบ่งออกเป็น 3 ส่วนหลักด้วยกัน



รูปที่ 3.18: ผลกระทบของความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงต่ออัตราส่วนซิบิล



รูปที่ 3.19: ผลกระทบของจำนวนผู้ใช้งานจริงต่ออัตราส่วนซิบิล

คือ สูตรแบบแม่นตรง และสูตรที่มีการประมาณค่าอีก 2 สูตร โดยการประมาณค่าแบบที่ 1 จะเป็นการประมาณค่าบางพจน์ในสูตรแบบแม่นตรงด้วยสูตรทางคณิตศาสตร์เกี่ยวกับการกระจายตัวซนิดปัวส์ซอง และการประมาณค่าแบบที่ 2 เป็นการประมาณค่าทั้งสูตรใหม่ โดยผ่อนผันสมมุติฐานว่าจำนวนคะแนนเสียงในแต่ละตัวเลือกไม่เป็นอิสระต่อกัน ทั้งนี้ความแม่นยำของสูตรทั้งสามถูกประเมินด้วยการจำนวนเหตุการณ์จริงแบบมอนติคาร์โลเป็นจำนวน 1 ล้านตัวอย่างต่อจุด ผลการประเมินแสดงให้เห็นว่า สูตรการคำนวณแบบแม่นตรงให้ผลการคำนวณที่แม่นยำที่สุด โดยมีความคลาดเคลื่อนอยู่ในระดับ 0.1% เท่านั้น แต่สูตรการคำนวณแบบแม่นตรงดังกล่าวกลับมีความซับซ้อนของสูตรการคำนวณที่ค่อนข้างมาก คือ อยู่ในระดับ $O\left((n+S)^k\right)$ ซึ่งต่างจากสูตรประมาณค่าแบบที่ 1 และแบบที่ 2 ซึ่งมีความซับซ้อนของสูตรเพียง $O(n)$ และ $O(1)$ ตามลำดับ ดังนั้นจึงสรุปว่า หากอยู่ในสถานการณ์ที่มีจำนวนผู้ออกเสียงจริง จำนวนซิบิล และจำนวนตัวเลือกที่ค่อนข้างน้อยควรใช้สูตรแบบแม่นตรงเพราะมีความแม่นยำสูงในขณะที่หากมีจำนวนซิบิลและจำนวนตัวเลือกมากควรใช้สูตรแบบประมาณค่าเพื่อให้ง่ายต่อการคำนวณและลดการใช้ทรัพยากรในการคำนวณ

บทที่ 4

การประยุกต์ใช้สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพื่อแยกแยะสถานะของผู้ใช้งานระบบ

ปัญหาที่เกิดจากการโจมตีชนิดซิบิลเกิดขึ้นได้ในทุกโครงข่าย โดยเฉพาะโครงข่ายที่มีการให้คะแนนชื่อเสียงในระบบ อาทิ โครงข่ายอินเทอร์เน็ตของสรรพสิ่ง (Internet of Things: IoT) ซึ่งเป็นระบบที่อนุญาตให้วัตถุสามารถเชื่อมต่อหรือโต้ตอบกับวัตถุอื่นผ่านทางโครงข่ายการสื่อสารในหลายรูปแบบ [31] ตัวอย่างเช่น โครงการอินเทอร์เน็ตของสรรพสิ่งสำหรับเมืองอัจฉริยะในอนาคต ที่ศูนย์ชุมชนสามารถเชื่อมต่อกับเกตเวย์ของแต่ละบ้านเพื่อบริหารจัดการเกี่ยวกับสาธารณูปโภค ขั้นตอนการบริหารจัดการตามปกติต้องอาศัยระบบการลงคะแนนโดยการนับจำนวนของผู้ร้องขอสิ่งอำนวยความสะดวก พื้นที่ที่มีผู้ร้องขอสิ่งอำนวยความสะดวกมากที่สุดอาจจะได้รับการจัดสรรสิ่งอำนวยความสะดวกก่อน การนับจำนวนผู้ร้องขอสิ่งอำนวยความสะดวกในระบบอินเทอร์เน็ตของสรรพสิ่งดังกล่าวเป็นจุดอ่อนที่จะถูกโจมตีโดยซิบิล [1] โดยผู้ไม่หวังดีจะสร้างตัวตนปลอมในการลงคะแนนเพื่อให้มีคะแนนเสียงมากกว่าพื้นที่อื่นและจะทำให้ได้รับสิ่งอำนวยความสะดวกมากกว่าหรือก่อนพื้นที่อื่น ดังนั้นผู้ดูแลระบบจึงต้องพยายามตรวจจับซิบิลเพื่อปกป้องผู้ใช้งานจริงจากการโจมตีในรูปแบบดังกล่าว

เมื่อกำหนดให้โนดหนึ่งโนดหรือจุดหนึ่งจุดเป็นสัญลักษณ์แทนผู้ใช้งานหนึ่งตัวตน (ทั้งที่เป็นตัวตนของผู้ใช้งานจริงและตัวตนของซิบิล) และกำหนดให้เส้นเชื่อมต่อระหว่างโนดแทนความสัมพันธ์ของผู้ใช้งาน จะได้ว่าโครงข่ายสามารถนำเสนอในรูปแบบกราฟได้ อย่างไรก็ตามความสัมพันธ์ระหว่างโนดมีหลายรูปแบบ อาทิ ความสัมพันธ์ทางเดียว หรือความสัมพันธ์สองทาง ตัวอย่างของความสัมพันธ์ทางเดียว ได้แก่ การเพิ่มเพื่อนในโครงข่ายสังคมออนไลน์ กล่าวคือ เมื่อนาย A ส่งคำร้องขอเป็นเพื่อนถึงนาย B และนาย B ตอบรับคำขอดังกล่าว สามารถเขียนเส้นเชื่อมต่อ (ลูกศร) จากโนด A ถึงโนด B ได้ ทั้งนี้เมื่อนาย A และนาย B เป็นเพื่อนกันแล้ว นาย B จะไม่สามารถส่งคำขอเป็นเพื่อนกลับไปให้นาย A ได้อีก ดังนั้นโครงข่ายการเพิ่มเพื่อนจึงเป็นโครงข่ายทางเดียว ตัวอย่างของความสัมพันธ์สองทาง ได้แก่ การพูดคุยในกระทู้ต่าง ๆ หากในอดีตนาย A เคยตั้งกระทู้และนาย B มาตอบกระทู้นั้น และในเวลาต่อมานาย B ตั้งกระทู้ใหม่ และนาย A มาตอบกระทู้ของนาย B จะเห็นว่าความสัมพันธ์ดังกล่าวเป็นความสัมพันธ์แบบสองทางเพราะสามารถตอบกลับไปมาได้ ทั้งนี้กำหนดให้เส้นเชื่อมต่อจากโนด A ถึงโนด B ในกราฟทอพอโลยีหมายถึง โนด A มีความสัมพันธ์ถึงโนด B และเส้นเชื่อมต่อ 2 ทิศทางระหว่างโนด A และโนด B หมายถึงโนด A มีความสัมพันธ์ถึงโนด B และโนด B มีความสัมพันธ์ถึงโนด A เช่นกัน สถานะของโนดในโครงข่ายที่วิทยานิพนธ์นี้พิจารณามี 2 ชนิด ได้แก่ ผู้ใช้งานจริงและซิบิล ดังนั้นกำหนดให้ความนิยมของโนด A มีค่าเท่ากับความน่าจะเป็นที่ผู้ใช้งานจริงใด ๆ ในโครงข่ายจะมีเส้นเชื่อมต่อถึงโนด A ซึ่งจะมีค่าลู่เข้าสู่จำนวนผู้ใช้งานจริงที่มีเส้นเชื่อมต่อถึงโนด A หารด้วยจำนวนผู้ใช้งานจริงทั้งหมดในระบบในกรณีที่มีผู้ใช้งานจริงในระบบเป็นจำนวนมาก วิทยานิพนธ์นี้จึงนำเสนอการประยุกต์ใช้ประโยชน์จากสูตรการคำนวณผลกระทบของการโจมตีชนิดซิบิลที่ได้นำเสนอในบทที่ 3 มาตรวจจับซิบิลในโครงข่ายที่มีการให้คะแนนชื่อเสียงในระบบทั้งระบบที่มีความสัมพันธ์ทางเดียวและความสัมพันธ์สองทาง ในกรณีที่เป็นความสัมพันธ์สองทางกำหนดให้โนดผู้ใช้งานจริงแต่ละโนดมีความนิยมทั้งเท่ากันและไม่เท่ากัน

ความสัมพันธ์ระหว่างโนดสองโนดขึ้นอยู่กับสถานะของโนดด้วย กล่าวคือ ความสัมพันธ์ระหว่างผู้ใช้งานจริงของระบบ 2 คนจะขึ้นอยู่กับธรรมชาติของการติดต่อในระบบนั้น ถ้าเป็นโครงข่ายที่ผู้ใช้

งานมีการติดต่อสื่อสารกันมาก จำนวนเส้นเชื่อมต่อระหว่างโนดจะหนาแน่นกว่าโครงข่ายที่ผู้ใช้งานมีการติดต่อสื่อสารระหว่างกันน้อย ในขณะที่ความสัมพันธ์ระหว่างซิบิล 2 ตัวที่ถูกสร้างขึ้นมาโดยผู้ไม่หวังดีคนเดียวกัน จะเป็นรูปแบบที่ผู้ไม่หวังดีออกแบบไว้เพื่อจุดประสงค์เฉพาะกิจเท่านั้น ดังนั้นความหนาแน่นของเส้นเชื่อมต่อระหว่างโนดที่อยู่ในกลุ่มเดียวกันจึงมีความไม่แน่นอน ขึ้นอยู่กับธรรมชาติของระบบ อย่างไรก็ตามความหนาแน่นของเส้นเชื่อมต่อระหว่างกลุ่มกลับมีลักษณะเฉพาะเจาะจง กล่าวคือ เส้นเชื่อมต่อจากผู้ใช้งานจริงถึงซิบิลจะมีเบาบางมากเมื่อเทียบกับความสัมพันธ์อื่น เนื่องจากผู้ใช้งานจริงไม่รู้จักซิบิลเป็นการส่วนตัวและซิบิลถูกสร้างมาเพื่อโจมตีผู้ใช้งานจริงที่เป็นเป้าหมายเท่านั้นดังนั้นโดยปกติซิบิลจึงไม่โพสต์ข้อความเพื่อให้ผู้ใช้งานจริงแสดงความคิดเห็น เป็นต้น ในขณะที่เส้นเชื่อมต่อจากซิบิลถึงผู้ใช้งานจริงที่เป็นเป้าหมายการโจมตีจะมีความหนาแน่นมากเป็นพิเศษ แต่ซิบิลจะไม่มีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอื่นที่ไม่ใช่เป้าหมายเลย นอกจากนั้นผู้ใช้งานจริงที่ซิบิลมีเส้นเชื่อมต่อถึงจะมีจำนวนจำกัดเสมอ เพราะผู้ใช้งานจริงที่ซิบิลมีความสัมพันธ์ถึงคือเหยื่อของการโจมตี เมื่อกำหนดให้เหยื่อหมายถึงผู้ใช้งานจริงที่ถูกโจมตีโดยซิบิล อีกทั้งจำนวนเส้นเชื่อมต่อจากซิบิลไปเหยื่อจะต้องมีจำนวนมากเพียงพอที่จะโน้มน้าวให้เหยื่อเชื่อในสิ่งที่ซิบิลชักจูงได้ ดังนั้นสูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่ได้นำเสนอไปแล้วในบทที่ 3 จะถูกใช้เพื่อตรวจจับซิบิลในขั้นตอนนี้ และเนื่องจากความสัมพันธ์ระหว่างโนดในระบบมีรูปแบบเฉพาะที่แตกต่างกัน ดังนั้นความสัมพันธ์ระหว่างโนดจึงถูกใช้เป็นเครื่องมือในการตรวจจับซิบิลในระบบ อย่างไรก็ตามวิทยานิพนธ์นี้กำหนดให้มีซิบิลเพียง 1 กลุ่มเท่านั้นเพื่อหลีกเลี่ยงความสับสนเกี่ยวกับการเชื่อมต่อระหว่างซิบิลต่างกลุ่มกัน

เพื่อทดสอบวิธีในการตรวจจับซิบิลจึงต้องสร้างโครงข่ายจำลองอย่างสุ่มขึ้นมาก่อนเป็นอันดับแรก โดยมีการกำหนดพารามิเตอร์ต่าง ๆ ดังนี้

1. จำนวนผู้ใช้งานจริงในระบบ (n)
2. จำนวนซิบิล (S)
3. จำนวนเหยื่อ (v)
4. ความน่าจะเป็นที่ผู้ใช้งานจริงคนหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอีกคนหนึ่ง (Prr)
5. ความน่าจะเป็นที่ผู้ใช้งานจริงคนหนึ่งจะมีเส้นเชื่อมต่อถึงซิบิล (Prs)
6. ความน่าจะเป็นที่ซิบิลตัวหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงคนหนึ่ง (Psr)
7. ความน่าจะเป็นที่ซิบิลตัวหนึ่งจะมีเส้นเชื่อมต่อถึงซิบิลอีกตัวหนึ่ง (Pss)

โดยค่าความน่าจะเป็น Prr, Prs, Psr, Pss จะใช้สำหรับผู้ใช้งานจริงคนใด หรือซิบิลตัวใดบ้างในเซตของโนดทั้งหมดที่แทนผู้ใช้งาน และซิบิลในโครงข่าย จะขึ้นอยู่ กับกรณีย่อย ๆ ของการพิจารณารูปแบบการเชื่อมต่อและความสัมพันธ์แต่ละแบบซึ่งจะกล่าวถึงรายละเอียดต่อไปในหัวข้อ 4.1 กำหนดให้จำนวนผู้ใช้งานจริงในระบบมีมากกว่าจำนวนซิบิล $(n > S)$ เพราะในกรณีที่มีซิบิลมากกว่าผู้ใช้งานจริงการล้มระบบแล้วสร้างระบบใหม่จะมีประโยชน์มากกว่าการตรวจจับซิบิล สำหรับพารามิเตอร์เกี่ยวกับความน่าจะเป็นในการเชื่อมต่อระหว่างโนด กำหนดให้ Prs มีค่าน้อยมาก ๆ เมื่อเทียบกับความน่าจะเป็นอื่น ในขณะที่ Psr มีค่าเป็น 0 เมื่อเชื่อมต่อกับผู้ใช้งานจริงที่ไม่ใช่เหยื่อ แต่จะมีค่ามากเมื่อเชื่อมต่อกับเหยื่อเมื่อเทียบกับพารามิเตอร์ความน่าจะเป็นอื่น

โครงข่ายจำลองเพื่อใช้ทดสอบวิธีตรวจจับซิบิลที่นำเสนอในวิทยานิพนธ์นี้แบ่งออกเป็น 3 รูปแบบ ได้แก่

1. โครงข่ายจำลองที่มีรูปแบบความสัมพันธ์ทางเดียว
2. โครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน
3. โครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน

วิทยานิพนธ์นี้นำเสนอวิธีการตรวจจับซิบิล 2 วิธี โดยวิธีแรกจะใช้แนวคิดที่ว่าโนดต่าง ๆ ในโครงข่ายจะติดต่อสื่อสารอย่างหนาแน่นกับตัวตนชนิดเดียวกันและจะติดต่อสื่อสารอย่างเบาบางกับโนดที่มีสถานะต่างกัน ดังนั้นแนวคิดของการแบ่งโครงข่ายออกเป็นกลุ่มย่อยแล้วตรวจจับซิบิลเป็นกลุ่มดังที่จะนำเสนอในวิทยานิพนธ์นี้น่าจะเหมาะสมกว่าการตรวจจับซิบิลทีละตัว [32]-[34] โดยวิทยานิพนธ์นี้ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว (fast modularity optimization) [29] เพื่อช่วยคัดแยกผู้ใช้งานจริงกับซิบิลออกจากกันได้ จากนั้นวิทยานิพนธ์นี้จึงใช้สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่ได้นำเสนอในบทที่ 3 เป็นตัวชี้วัดว่ากลุ่มย่อยไหนเป็นกลุ่มของซิบิล สำหรับวิธีที่สองจะใช้คุณลักษณะของการโจมตีของซิบิลให้เป็นประโยชน์ โดยซิบิลส่วนใหญ่จะมีเส้นเชื่อมต่อถึงเหยื่อคนเดียวกันเป็นจำนวนมาก ดังนั้นโนดต่าง ๆ ที่มีความสัมพันธ์ถึงโนดเดียวกันมากย่อมมีความน่าจะเป็นสูงที่จะเป็นซิบิลกลุ่มเดียวกัน วิธีตรวจหาซิบิลทั้ง 2 วิธีถูกเปรียบเทียบความถูกต้องด้วยการจำลองเหตุการณ์แบบมอนติคาร์โลและประเมินความซับซ้อนทางเวลา (time complexity) รวมถึงประเด็นความเหมาะสมในการประยุกต์ใช้ตรวจหาซิบิลในโครงข่ายรูปแบบต่าง ๆ ทั้งนี้รายละเอียดเกี่ยวกับโครงข่ายจำลองเพื่อใช้ทดสอบวิธีตรวจจับซิบิล และวิธีการตรวจจับซิบิลจะได้กล่าวในหัวข้อที่ 4.1 ต่อไป

4.1 การสร้างโครงข่ายจำลองเพื่อใช้ทดสอบวิธีตรวจจับซิบิล

โครงข่ายจำลองสามารถเขียนแทนได้ในรูปแบบของเมตริกซ์การเชื่อมต่อระหว่างโนด (adjacent matrix) เมตริกซ์ดังกล่าวประกอบด้วยเลข 0 และ 1 กำหนดให้ i, j เป็นจำนวนนับที่มีค่าไม่เกิน $n + S$ หากสมาชิกของเมตริกซ์การเชื่อมต่อในพิกัด ij มีค่าเป็น 0 หมายถึงไม่มีเส้นเชื่อมต่อจากโนด i ถึงโนด j ในทางกลับกันหากสมาชิกของเมตริกซ์การเชื่อมต่อในพิกัด ij มีค่าเป็น 1 หมายถึงโนด i มีเส้นเชื่อมต่อถึงโนด j เมตริกซ์การเชื่อมต่อสามารถแบ่งออกเป็น 4 ส่วน ได้แก่ Mrr เป็นเมตริกซ์ขนาด $n \times n$ ใช้ระบุถึงการเชื่อมต่อจากผู้ใช้งานจริงคนหนึ่งถึงผู้ใช้งานจริงอีกคนหนึ่งในโครงข่าย Mrs เป็นเมตริกซ์ขนาด $n \times S$ ใช้ระบุถึงการเชื่อมต่อจากผู้ใช้งานจริงถึงซิบิล Msr เป็นเมตริกซ์ขนาด $S \times n$ ใช้ระบุถึงการเชื่อมต่อจากซิบิลถึงผู้ใช้งานจริง และ Mss เป็นเมตริกซ์ขนาด $S \times S$ ใช้ระบุถึงการเชื่อมต่อจากซิบิลตัวหนึ่งถึงซิบิลอีกตัวหนึ่งในกลุ่มเดียวกัน โดยไม่เสียนัยสำคัญกำหนดให้โนดหมายเลข 1 ถึง n เป็นโนดแทนผู้ใช้งานจริงและโนดหมายเลข $n+1$ ถึง $n+S$ แทนซิบิล ดังนั้นเมตริกซ์การเชื่อมต่อระหว่างโนดในโครงข่ายจึงเขียนได้เป็น $\begin{bmatrix} Mrr & Mrs \\ Msr & Mss \end{bmatrix}$ เมตริกซ์การเชื่อมต่อระหว่างโนดแต่ละส่วนมีคุณลักษณะที่แตกต่างกัน โดยสมาชิกของเมตริกซ์ Mrs ส่วนใหญ่จะมีค่าเป็น 0 และมีเพียงส่วนน้อยมากเท่านั้นที่เป็นเลข 1 เนื่องจากผู้ใช้งานจริงส่วนใหญ่ไม่ส่งข้อมูลหรือส่งคำร้องขอเป็นเพื่อนถึงซิบิลเพราะไม่รู้จักเป็นการส่วนตัว สมาชิกของเมตริกซ์ Msr ส่วนใหญ่จะมีค่าเป็น 1 ในแนวตั้ง (column) ที่ระบุถึงเหยื่อและจะเป็น 0 สำหรับแนวตั้งที่ไม่ใช่เหยื่อ Mrr และ Mss จะมีลักษณะขึ้นอยู่กับลักษณะโครงข่ายและความหนาแน่นของโครงข่ายซึ่งกำหนดโดยค่า Prr และ Pss ตามลำดับ รายละเอียดเกี่ยวกับการสร้างโครงข่ายจำลองเพื่อใช้ทดสอบวิธีตรวจจับซิบิลแบ่ง

ตามชนิดโครงข่ายมีดังต่อไปนี้

โครงข่ายจำลองที่มีรูปแบบความสัมพันธ์ทางเดียว

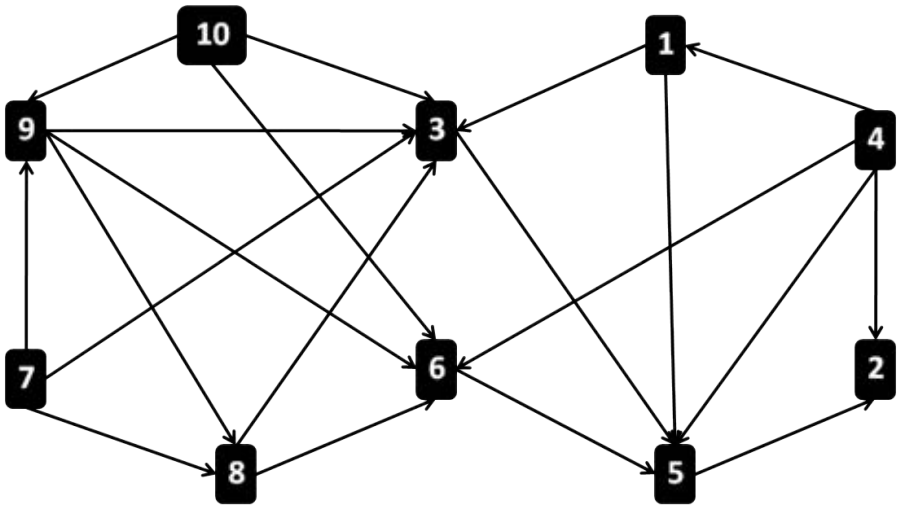
สำหรับโครงข่ายที่มีรูปแบบความสัมพันธ์ทางเดียว เนื่องจากเมื่อโนด A มีเส้นเชื่อมต่อถึงโนด B แล้วโนด B จะมีเส้นเชื่อมต่อถึงโนด A ไม่ได้ ดังนั้นเมตริกซ์การเชื่อมต่อระหว่างโนดในตำแหน่ง ij และตำแหน่ง ji จะมีค่าเป็น 1 พร้อมกันไม่ได้ การสร้างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์ทางเดียวจึงทำตามขั้นตอนวิธีดังนี้

1. กำหนดผู้ใช้งานจริงจำนวน v คนเป็นเป็นสมาชิกของเซต $Victim$ (เซตของผู้ใช้งานจริงที่จะเป็นเหยื่อของการโจมตี)
2. สร้างเมตริกซ์สามเหลี่ยมล่าง $A_{n\times n}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อใช้เป็นตัวตั้งต้นในการสร้างเมตริกซ์ Mrr กำหนดให้ Mrr_{ij} มีค่าเป็น 1 เมื่อ A_{ij} มีค่ามากกว่าหรือเท่ากับ $1-\frac{Prr}{2}$ และ Mrr_{ji} มีค่าเป็น 1 เมื่อ A_{ij} มีค่าน้อยกว่า $\frac{Prr}{2}$ และมีค่าเป็น 0 ในกรณีอื่น
3. สร้างเมตริกซ์ $Mrs_{n\times S}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อกล่าวถึงการเชื่อมต่อจากผู้ใช้งานจริงถึงซิบิล โดย Mrs_{ij} จะถูกเปลี่ยนค่าเป็น 1 เมื่อ Mrs_{ij} มีค่าน้อยกว่า Prs และจะถูกเปลี่ยนค่าเป็น 0 ในกรณีอื่น
4. สร้างเมตริกซ์ $Msr_{S\times n}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อกล่าวถึงการเชื่อมต่อจากซิบิลถึงผู้ใช้งานจริงที่เป็นเป้าหมาย โดย Msr_{ij} จะถูกเปลี่ยนค่าเป็น 1 เมื่อ Msr_{ij} มีค่าน้อยกว่า Psr และ $j \in Victim$ แต่ Msr_{ij} จะถูกเปลี่ยนค่าเป็น 0 ในกรณีอื่น
5. สร้างเมตริกซ์สามเหลี่ยมล่าง $B_{S\times S}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อใช้เป็นตัวตั้งต้นในการสร้างเมตริกซ์ Mss กำหนดให้ Mss_{ij} มีค่าเป็น 1 เมื่อ B_{ij} มีค่ามากกว่าหรือเท่ากับ $1-\frac{Pss}{2}$ และ Mss_{ji} มีค่าเป็น 1 เมื่อ B_{ij} มีค่าน้อยกว่า $\frac{Pss}{2}$ และมีค่าเป็น 0 ในกรณีอื่น
6. ประกอบเมตริกซ์ย่อยให้เป็นเมตริกซ์การเชื่อมต่อระหว่างโนด ดังนี้ $\begin{bmatrix} Mrr & Mrs \\ Msr & Mss \end{bmatrix}$

ตัวอย่างโครงข่ายจำลองที่มีความสัมพันธ์ทางเดียวในรูปแบบของรูปภาพทอพอโลยีและเมตริกซ์การเชื่อมต่อระหว่างโนดในโครงข่ายนำเสนอในรูปที่ 4.1 และรูปที่ 4.2 ตามลำดับ จากตัวอย่างจะเห็นว่าความสัมพันธ์ระหว่างโนดในโครงข่ายจะเป็นเส้นเชื่อมต่อทางเดียวเสมอ

โครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน

1. สร้างเมตริกซ์จัตุรัส $B_{n\times n}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อใช้เป็นตัวตั้งต้นในการสร้างเมตริกซ์ Mrr กำหนดให้ Mrr_{ij} มีค่าเป็น 1 เมื่อ B_{ij} มีค่าน้อยกว่า Prr และมีค่าเป็น 0 ในกรณีอื่น
2. สร้างเมตริกซ์จัตุรัส $C_{S\times S}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อใช้เป็นตัวตั้งต้นในการสร้างเมตริกซ์ Mss กำหนดให้ Mss_{ij} มีค่าเป็น 1 เมื่อ C_{ij} มีค่าน้อยกว่า Pss และมีค่าเป็น 0 ในกรณีอื่น



รูปที่ 4.1: ตัวอย่างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์ทางเดียวในรูปแบบของรูปภาพทอพอโลยี

3. สร้างเมตริกซ์ $D_{n\times S}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อใช้เป็นตัวตั้งต้นในการสร้างเมตริกซ์ Mrs กำหนดให้ Mrs_{ij} มีค่าเป็น 1 เมื่อ D_{ij} มีค่าน้อยกว่า Prs และมีค่าเป็น 0 ในกรณีอื่น

4. สร้างเมตริกซ์ $E_{S\times n}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อใช้เป็นตัวตั้งต้นในการสร้างเมตริกซ์ Msr กำหนดให้ Msr_{ij} มีค่าเป็น 1 เมื่อ E_{ij} มีค่าน้อยกว่า Psr และมีค่าเป็น 0 ในกรณีอื่น

5. ประกอบเมตริกซ์ย่อยให้เป็นเมตริกซ์การเชื่อมต่อระหว่างโนด ดังนี้ $\begin{bmatrix} Mrr & Mrs \\ Msr & Mss \end{bmatrix}$

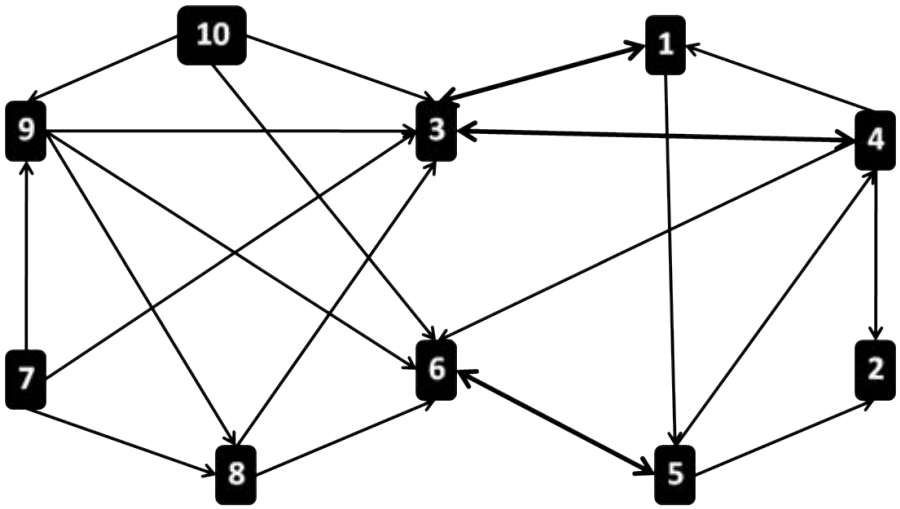
ตัวอย่างโครงข่ายจำลองที่มีความสัมพันธ์สองทาง โดยผู้ใช้งานจริงแต่ละโนดของมีความนิยมเท่ากันในรูปแบบของรูปภาพทอพอโลยีและเมตริกซ์การเชื่อมต่อระหว่างโนดในโครงข่ายนำเสนอในรูปที่ 4.3-4.4 ตามลำดับ โดยโนด 1-6 เป็นผู้ใช้งานจริงและโนด 7-10 เป็นซิบิล ในรูปภาพทอพอโลยีสังเกตว่าเส้นเชื่อมต่อมีทั้งแบบทางเดียวและสองทาง และสำหรับรูปแบบเมตริกซ์การเชื่อมต่อจะเห็นว่าตำแหน่ง ij และ ji มีค่าเป็น 1 พร้อมกันได้ในกรณีที่เป็นเส้นเชื่อมต่อสองทาง ทั้งนี้จำนวนเส้นเชื่อมต่อถึงผู้ใช้งานจริงแต่ละโนดในโครงข่ายมีจำนวนที่ใกล้เคียงกันเพราะผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน

โครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน

เนื่องจากความนิยมของผู้ใช้งานจริง A มีค่าเท่ากับความน่าจะเป็นที่ผู้ใช้งานจริงใด ๆ ในโครงข่ายจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริง A ดังนั้นการสร้างเมตริกซ์ Mrs, Msr, Mss ในกรณีที่ผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน จะมีวิธีการสร้างเหมือนกับกรณีที่ผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน แตกต่างกันแต่เพียงเมตริกซ์ Mrr ซึ่งเป็นเมตริกซ์แทนเส้นเชื่อมต่อจากผู้ใช้งานจริงโนดหนึ่งถึงผู้ใช้งานจริงอีกโนดหนึ่งเท่านั้น ดังนั้นขั้นตอนการสร้างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันจึงมีวิธีดังต่อไปนี้

$$\begin{bmatrix}
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}$$

รูปที่ 4.2: ตัวอย่างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์ทางเดียวในรูปแบบของเมตริกซ์การเชื่อม-ต่อระหว่างโนดในโครงข่าย



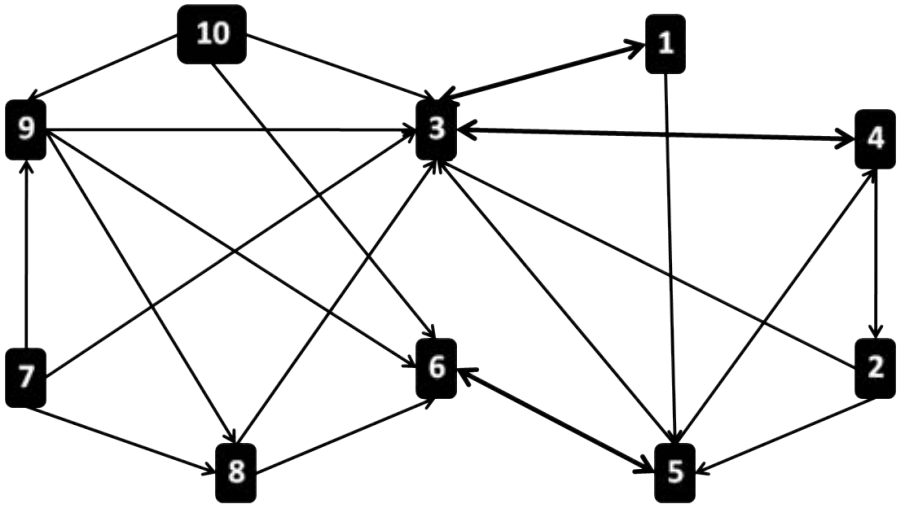
รูปที่ 4.3: ตัวอย่างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันในรูปแบบของรูปภาพทอพอโลยี

$$\begin{bmatrix}
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}$$

รูปที่ 4.4: ตัวอย่างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันในรูปแบบของเมตริกซ์การเชื่อมต่อระหว่างโนดในโครงข่าย

1. กำหนดผู้ใช้งานจริงที่จะถูกซิบิลโจมตี รวมถึงสร้างเมตริกซ์ Mrs, Msr, Mss ในลักษณะเดียวกันกับกรณีโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน
2. สร้างเมตริกซ์ $H_{1\times n}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง Prr เพื่อใช้เป็นความนิยมของผู้ใช้งานจริงแต่ละโนด โดยความนิยมของโนดที่ j มีค่าเท่ากับ H_{1j}
3. สร้างเมตริกซ์ $Mrr_{n\times S}$ อย่างสุ่มแบบเอกรูปต่อเนื่องในช่วง 0 ถึง 1 เพื่อกล่าวถึงการเชื่อมต่อจากผู้ใช้งานจริงถึงซิบิล โดย Mrr_{ij} จะถูกเปลี่ยนค่าเป็น 1 เมื่อ Mrr_{ij} มีค่าน้อยกว่า H_{1j} และจะถูกเปลี่ยนค่าเป็น 0 ในกรณีอื่น
4. ประกอบเมตริกซ์ย่อยให้เป็นเมตริกซ์การเชื่อมต่อระหว่างโนด ดังนี้ $\begin{bmatrix} Mrr & Mrs \\ Msr & Mss \end{bmatrix}$

ตัวอย่างโครงข่ายจำลองที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในรูปแบบของรูปภาพทอพอโลยีและเมตริกซ์การเชื่อมต่อระหว่างโนดในโครงข่ายนำเสนอในรูปที่ 4.5-4.6 ตามลำดับ โดยโนด 1-6 เป็นผู้ใช้งานจริงและโนด 7-10 เป็นซิบิล ซึ่งในส่วนที่เป็นรูปภาพทอพอโลยีสังเกตว่าเส้นเชื่อมต่อมีทั้งแบบทางเดียวและสองทาง และสำหรับรูปแบบเมตริกซ์การเชื่อมต่อจะเห็นว่าตำแหน่ง ij และ ji อาจจะเป็น 1 พร้อมกันก็ได้ในกรณีที่เป็นเส้นเชื่อมต่อสองทาง ทั้งนี้จำนวนเส้นเชื่อมต่อที่เข้าสู่โนดแต่ละโนดในโครงข่ายอาจมีจำนวนแตกต่างกันมากเพราะผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน



รูปที่ 4.5: ตัวอย่างโครงข่ายจำลองที่มีความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในรูปแบบของรูปภาพทอพอโลยี

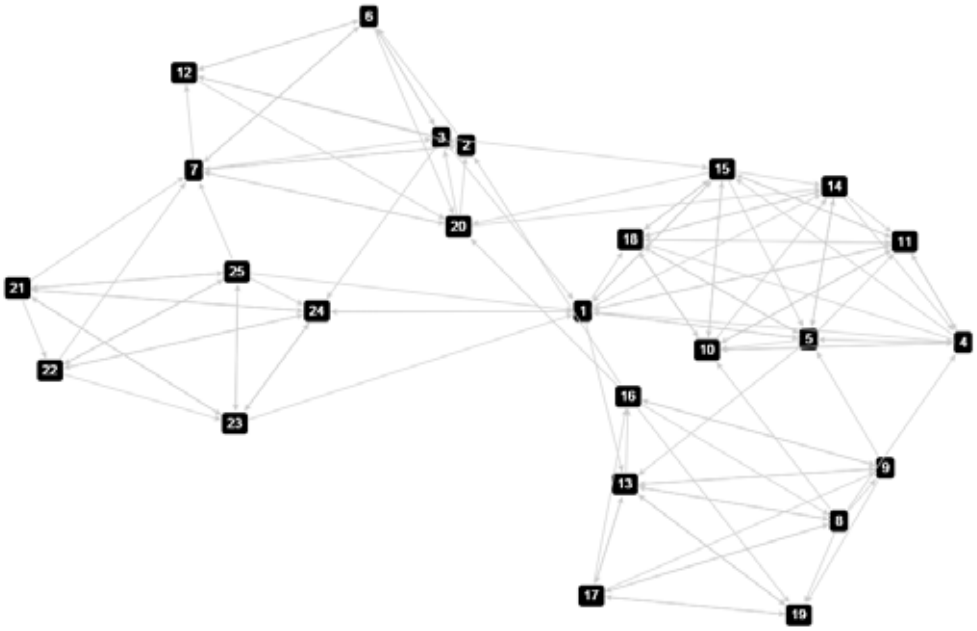
0	0	1	0	1	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0
1	1	1	0	1	1	0	0	0	0
0	1	0	1	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	1	1	0
0	0	1	0	0	1	0	0	0	0
0	0	1	0	0	1	0	1	0	0
0	0	1	0	0	1	0	0	1	0

รูปที่ 4.6: ตัวอย่างโครงข่ายจำลองที่มีความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในรูปแบบของเมตริกซ์การเชื่อมต่อระหว่างโนดในโครงข่าย

4.2 การประยุกต์ใช้สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพื่อแยกแยะสถานะของผู้ใช้งานระบบ

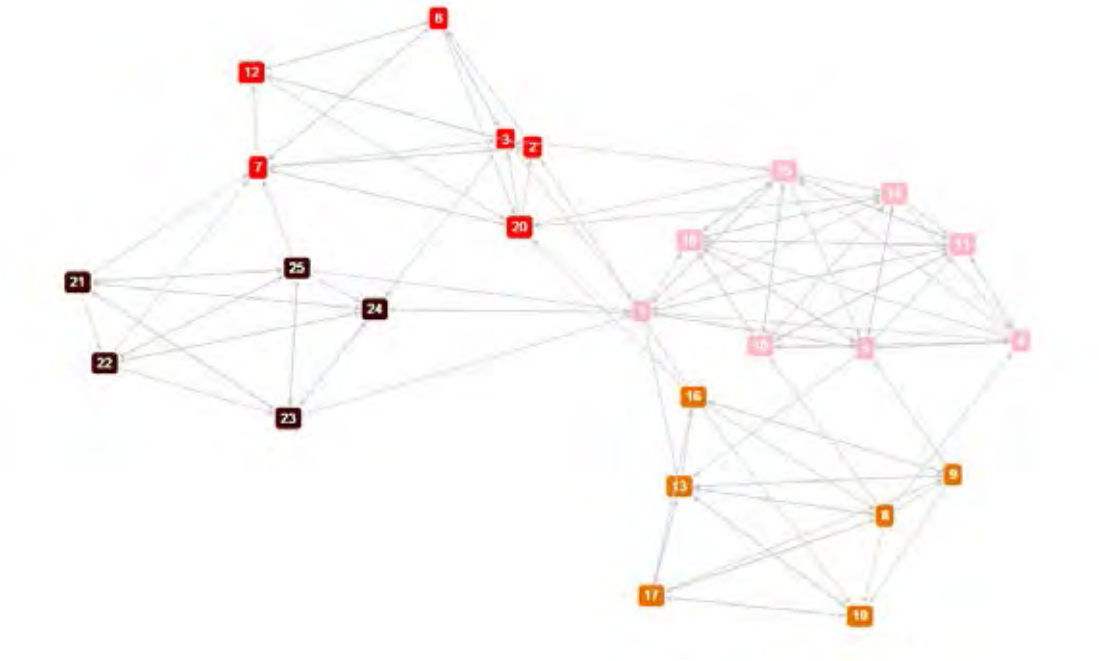
4.2.1 วิธีที่ 1: การแบ่งกลุ่มย่อยแล้วตรวจจับซิบิลเป็นกลุ่ม

ข้อมูลเบื้องต้นคือโครงข่ายที่ประกอบด้วยผู้ใช้งานจริงและซิบิลซึ่งอยู่ปะปนกันโดยไม่ทราบสถานะของโนด ตัวอย่างหนึ่งของโครงข่ายที่ไม่ได้ระบุสถานะของโนดนำเสนอในรูปที่ 4.7 ในอดีตที่ผ่านมามีงานวิจัยที่นำเสนอวิธีแบ่งโครงข่ายออกเป็นกลุ่มย่อยตามคุณสมบัติที่ต้องการอยู่แล้ว คุณสมบัติที่ต้องการสำหรับการตรวจจับซิบิล คือ ผู้ใช้งานจริงและซิบิลจะต้องถูกแบ่งแยกกันและไม่ปะปนกันในกลุ่มย่อยใด ๆ วิทยานิพนธ์นี้เลือกวิธีที่นำเสนอใน [29] โดยวิธีดังกล่าวจะจัดโนดที่มีความสัมพันธ์หนาแน่นระหว่างกันไว้เป็นกลุ่มย่อยเดียวกัน และโนดที่มีความสัมพันธ์เบาบางระหว่างกันเป็นส่วนรอยต่อระหว่างกลุ่มย่อย ผลการทดสอบทางตัวเลขที่แสดงใน [29] ชี้ให้เห็นว่าวิธีดังกล่าวมีสมรรถนะดีเพียงพอที่จะเลือกใช้ เมื่อใช้วิธีดังกล่าวแยกกลุ่มย่อยในโครงข่ายตัวอย่างรูปที่ 4.7 จะได้โครงข่ายที่แบ่งกลุ่มย่อยแล้วดังรูปที่ 4.8 โดยโนดที่อยู่ในกลุ่มย่อยเดียวกันจะใช้สีเดียวกัน ซึ่งกลุ่มย่อยในโครงข่ายรูปที่ 4.8 นี้แบ่งออกเป็น 4 กลุ่มย่อย ได้แก่ กลุ่มย่อยสีชมพู (pink) ได้แก่ โนด 1,4,5,10,11,14,15,18 สีแดง (red) ได้แก่โนด 2,3,6,7,12,20 สีน้ำตาล (brown) ได้แก่โนด 8,9,13,16,17,19 และสีน้ำตาลเข้ม (DarkBrown) ได้แก่โนด 21,22,23,24,25



รูปที่ 4.7: ตัวอย่างโครงข่ายที่ประกอบด้วยผู้ใช้งานจริงและซิบิลซึ่งอยู่ปะปนกันและไม่ทราบสถานะของโนด

สำหรับโครงข่ายที่แบ่งออกเป็นกลุ่มย่อยแล้ว จะพบว่ามีโนดจำนวนหนึ่งที่มีเส้นเชื่อมต่อจากโนดอื่นที่อยู่ต่างกลุ่มย่อยกัน (โนด A มีความสัมพันธ์ถึงโนด B มีความหมายเหมือนกับ โนด B มีความสัมพันธ์จากโนด A) เมื่อโนด X ไม่ได้อยู่ในกลุ่มย่อย c กำหนดให้ $f_{X,c}(\cdot)$ เป็นอัตราส่วนของผลรวมจำนวนเส้นเชื่อมต่อจากโนดในกลุ่มย่อย c ถึงโนด X ต่อจำนวนเส้นเชื่อมต่อจากโนดทุก



รูปที่ 4.8: โครงข่ายที่ถูกแบ่งกลุ่มย่อยแล้วแต่ไม่ทราบสถานะของกลุ่มย่อยแต่ละกลุ่ม

โนดในโครงข่ายถึงโนด X (เรียกอัตราส่วนดังกล่าวอย่างสั้นว่า "อัตราส่วนการโน้มน้าว") เขียนเป็นสูตรได้ว่า

$$f_{X,c}\left(N, n_c\right) = \frac{n_c}{N} \tag{4.1}$$

เมื่อกำหนดให้ N คือ จำนวนเส้นเชื่อมต่อจากโนดทุกโนดในโครงข่ายถึงโนด X และ n_c คือ ผลรวมจำนวนเส้นเชื่อมต่อจากโนดในกลุ่มย่อย c ถึงโนด X จากสมมุติฐานที่ผู้ใช้งานจริงและซิบิลจะอยู่ในกลุ่มย่อยเดียวกันไม่ได้ ในกรณีที่กลุ่ม c เป็นกลุ่มของซิบิล ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจะแปรผันตามอัตราส่วนการโน้มน้าว ดังนั้นวิธีที่ 1 นี้จะใช้อัตราส่วนการโน้มน้าวในการตรวจหากลุ่มของซิบิล [35] ใน 3 ขั้นตอนดังนี้

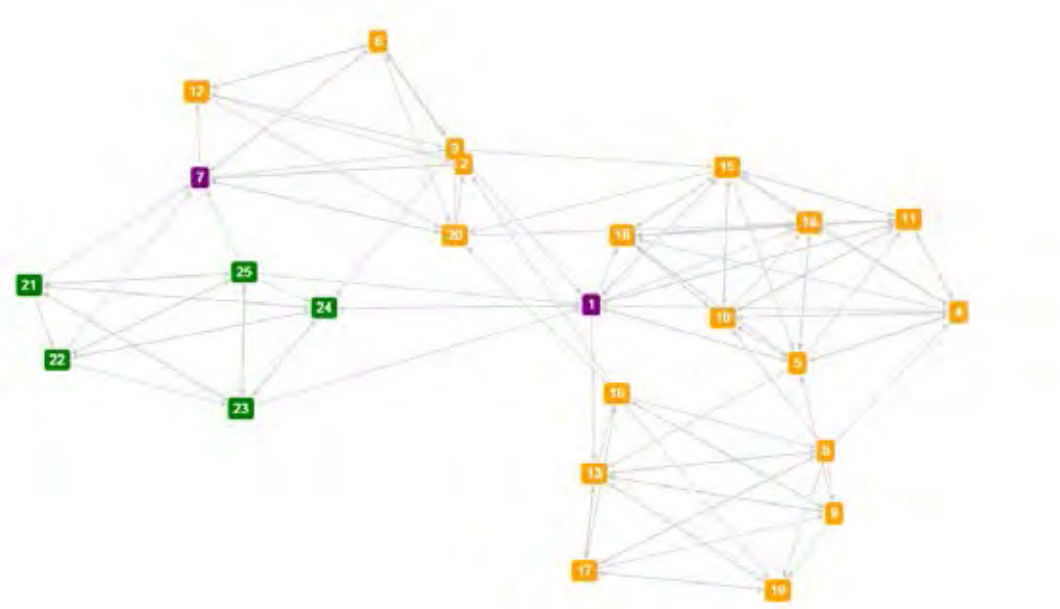
1. แบ่งโครงข่ายออกเป็นกลุ่มย่อยด้วยวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว [29] ดังรายละเอียดในหัวข้อ 2.4

2. คำนวณค่าอัตราส่วนการโน้มน้าวสำหรับโนด X แต่ละโนดในโครงข่ายและกลุ่ม c ทุกกลุ่ม ถ้า $f_{X,c}\left(N, n_c\right) = \frac{n_c}{N} \geq \gamma$ (ระดับการโน้มน้าวที่ยอมรับได้) จะพิจารณาว่าเส้นเชื่อมต่อจากโนดทุกโนดในกลุ่ม c ถึงโนด X เป็นเส้นเชื่อมต่อที่อาจจะเป็นการโจมตีจากกลุ่ม c ทั้งนี้ระดับการโน้มน้าวที่ยอมรับได้ γ เป็นพารามิเตอร์ที่จะถูกกำหนดไว้ก่อนแล้ว

3. ระบุให้กลุ่ม c เป็นกลุ่มของซิบิลเมื่อเส้นเชื่อมต่อที่อาจจะเป็นการโจมตีจากกลุ่ม c มีจำนวนมากกว่าจำนวนเส้นเชื่อมต่อรวมจากโนดนอกกลุ่ม c ถึงโนดในกลุ่ม c

ในตัวอย่างโครงข่ายที่แบ่งกลุ่มย่อยแล้วดังรูปที่ 4.8 มีเส้นเชื่อมต่อ 8 เส้นจากโนดทุกโนดในโครงข่ายถึงโนด 1 $(N = 8)$ ได้แก่ เส้นเชื่อมต่อจากโนด 2,4,5,11,15,23,24,25 แบ่งออกเป็นเส้นเชื่อมต่อ

จากกลุ่มย่อยสีชมพู 4 เส้น ($n_{pink} = 4$) จากโนด 4,5,11,15 เส้นเชื่อมต่อจากกลุ่มย่อยสีแดง 1 เส้น ($n_{red} = 1$) จากโนด 2 เส้นเชื่อมต่อจากกลุ่มย่อยสีน้ำตาลเข้ม 3 เส้น ($n_{DarkBrown} = 3$) จากโนด 23,24,25 จากสมการที่ (4.1) เมื่อกำหนดให้ $\gamma = \frac{1}{3}$ และคำนวณอัตราส่วนการโน้มน้าวโนด 1 จากกลุ่มย่อย c ทั้งหมด พบว่า $f_{1,red}(8,1) = \frac{1}{8}$ และ $f_{1,DarkBrown}(8,3) = \frac{3}{8}$ เส้นเชื่อมต่อจากโนดในกลุ่มย่อยสีแดงถึงโนด 1 ไม่ใช่เส้นเชื่อมต่อที่อาจจะเป็นการโจมตีเพราะอัตราส่วนการโน้มน้าวมีค่าไม่เกินระดับการโน้มน้าวที่ยอมรับได้ (γ) ในขณะที่เส้นเชื่อมต่อจากกลุ่มย่อยสีน้ำตาลเข้มถึงโนด 1 เป็นเส้นเชื่อมต่อที่อาจจะเป็นการโจมตีเพราะอัตราส่วนการโน้มน้าวมีค่าเกินระดับการโน้มน้าวที่ยอมรับได้ เมื่อพิจารณาโนด X ทั้งหมดและกลุ่มย่อย C ทั้งหมดในโครงข่ายแล้ว พบว่ากลุ่มย่อยสีน้ำตาลเข้มมีเส้นเชื่อมต่อที่อาจจะเป็นการโจมตีรวม 6 เส้นเชื่อมถึงโนด 1 (จากโนด 23,24,25) และโนด 7 (จากโนด 21,22,25) แต่กลุ่มย่อยสีน้ำตาลเข้มมีเส้นเชื่อมต่อจากกลุ่มย่อยอื่น 2 เส้น (จากโนด 1,3 ถึงโนด 24) ดังนั้นกลุ่มย่อยสีน้ำตาลเข้มจึงถูกตัดสินว่าเป็นกลุ่มซิบิล เป็นต้น เมื่อทุกขั้นตอนเสร็จสิ้น ผลจากการตรวจจับซิบิลจะมีลักษณะดังรูปที่ 4.9 โดยที่โนด 1 และโนด 7 อาจพิจารณาได้ว่าเป็นเหยื่อของการโจมตีของซิบิล



รูปที่ 4.9: ตัวอย่างโครงข่ายที่ระบุสถานะของกลุ่มแล้ว

วิธีดังกล่าวสามารถประยุกต์ใช้กับโครงข่ายที่มีกลุ่มของผู้ใช้งานจริงหลายกลุ่มได้เป็นอย่างดี เพื่อสะท้อนให้เห็นถึงสมรรถนะของการตรวจจับซิบิลในกรณีที่มีกลุ่มผู้ใช้งานจริงหลายกลุ่ม จึงสร้างโครงข่ายจำลองขึ้นมาเพื่อทดสอบ และกำหนดความน่าจะเป็นในการเชื่อมต่อระหว่างผู้ใช้งานจริงเพิ่มเติม อาทิ P_{inter} คือ ความน่าจะเป็นที่ผู้ใช้งานจริงโนดหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอีกโนดหนึ่งที่อยู่ต่างกลุ่มย่อยกัน P_{intra} คือ ความน่าจะเป็นที่ผู้ใช้งานจริงโนดหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอีกโนดหนึ่งที่อยู่ในกลุ่มย่อยเดียวกัน เนื่องจาก P_{rr} คือ ความน่าจะเป็นที่ผู้ใช้งานจริงโนดหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอีกโนดหนึ่งโดยเฉลี่ยทั้งโครงข่าย ดังนั้น $P_{\text{intra}} \geq Prr \geq P_{\text{inter}}$ เพื่อให้กลุ่มย่อยแต่ละกลุ่มแบ่งออกจากกันได้

ผลกระทบของ $P_{\text{inter}}, P_{\text{intra}}, Psr$ ต่อความแม่นยำของวิธีการตรวจจับซิบิลถูกนำเสนอในรูปแบบของความผิดพลาดเชิงบวกและเชิงลบ โดยกำหนดให้ความผิดพลาดเชิงบวก (positive error) คือ ร้อยละของความผิดพลาดที่ผู้ใช้งานจริงถูกตรวจจับ คำนวณจากจำนวนผู้ใช้งานจริงที่ถูกตรวจจับหารด้วยจำนวนผู้ใช้งานจริงทั้งหมดในโครงข่าย มีค่าเป็น 0 เมื่อไม่มีผู้ใช้งานจริงถูกตรวจจับเลย และมีค่าเป็น 1 เมื่อผู้ใช้งานจริงทั้งหมดถูกตรวจจับ เมื่อกำหนดให้ $\hat{r}$ คือเซตของผู้ใช้งานจริงที่ถูกตรวจจับว่าเป็นซิบิล $\bar{r}$ คือเซตของผู้ใช้งานจริงในโครงข่าย และ E_p คือ ความผิดพลาดเชิงบวก $|A|$ คือจำนวนสมาชิกของเซต A จะได้ว่า

$$E_p = \frac{|\hat{r}|}{|\bar{r}|} \tag{4.2}$$

และกำหนดให้ความผิดพลาดเชิงลบ (negative error) คือ ร้อยละของความผิดพลาดที่ซิบิลไม่ถูกตรวจจับ คำนวณจากจำนวนซิบิลที่ไม่ถูกตรวจจับหารด้วยจำนวนซิบิลทั้งหมดในโครงข่าย มีค่าเป็น 0 เมื่อซิบิลทุกตัวถูกตรวจจับ และมีค่าเป็น 1 เมื่อซิบิลทุกตัวถูกระบุว่าเป็นผู้ใช้งานจริง เมื่อกำหนดให้ $\hat{S}$ คือเซตของซิบิลที่ไม่ถูกตรวจจับ $\bar{S}$ คือเซตของซิบิลทั้งหมดในโครงข่าย และ E_n คือ ความผิดพลาดเชิงลบ จะได้ว่า

$$E_n = \frac{\left|\hat{S}\right|}{\left|\bar{S}\right|} \tag{4.3}$$

ตัวอย่างเช่น เมื่อโครงข่ายมีซิบิลอยู่ 5 โนดได้แก่ $\{s_1, s_2, s_3, s_4, s_5\}$ และมีผู้ใช้งานจริงอยู่ทั้งหมด 30 โนด ได้แก่ $\{r_1, r_2, \ldots, r_{30}\}$ ถ้าวิธีตรวจจับซิบิลระบุว่ากลุ่มโนด $\{s_1, s_2, s_3, r_{16}\}$ เป็นซิบิล จะได้ว่ามีผู้ใช้งานจริงถูกตรวจจับว่าเป็นซิบิล 1 คนจากทั้งหมด 30 คน ดังนั้นความผิดพลาดเชิงบวกจึงมีค่าเท่ากับ $\frac{1}{30} = 3.33\%$ ซิบิลที่ไม่สามารถตรวจจับได้มี 2 ตัว จากทั้งหมด 5 ตัว ดังนั้นความผิดพลาดเชิงลบคือ $\frac{2}{5} = 40\%$ เป็นต้น

อย่างไรก็ตามการตรวจจับซิบิลเป็นรายกลุ่มอาจทำให้เกิดความผิดพลาดสูงได้ในทางปฏิบัติ ตัวอย่างเช่น กลุ่มผู้ใช้งานจริงกลุ่มหนึ่ง กำหนดให้ชื่อว่ากลุ่มย่อย E มีสมาชิก 100 โนด และมีโนดอยู่ 10 โนดในกลุ่มย่อย E ที่มีพฤติกรรมเหมือนกันทุกประการคือมีเส้นเชื่อมต่อถึงผู้ใช้งานจริง X ที่อยู่กลุ่มอื่น และโนด X มีเส้นเชื่อมต่อจากผู้ใช้งานจริงอื่นน้อยมาก จึงทำให้เส้นเชื่อมต่อจากโนดในกลุ่มย่อย E ทั้งหมด 10 เส้นถูกสงสัยว่าเป็นเส้นเชื่อมต่อโจมตี และด้วยความบังเอิญที่จำนวนเส้นเชื่อมต่อรวมจากโนดในกลุ่มย่อยอื่นมาโนดในกลุ่มย่อย E มีน้อยกว่า 10 เส้นจึงสรุปเหมารวมว่าผู้ใช้งานจริงทั้งหมดในกลุ่มย่อย E เป็นซิบิล นำไปสู่ความผิดพลาดเชิงบวกที่สูงได้ ดังนั้นจึงขอเปลี่ยนนิยามกลุ่มของซิบิลใหม่จากกลุ่มย่อยที่สมาชิกทั้งหมดในกลุ่มเป็นซิบิลกลายเป็นกลุ่มย่อยที่มีซิบิลผสมอยู่ และโนดในกลุ่มของซิบิลที่ไม่ได้มีเส้นเชื่อมต่อถึงเหยื่อจะไม่ถือเป็นซิบิลอีกต่อไป ทั้งนี้สมาชิกของกลุ่มของซิบิลจะถูกตรวจจับว่าเป็นซิบิลเมื่อโนดสมาชิกนั้นเป็นต้นกำเนิดเส้นเชื่อมต่อที่ต้องสงสัยว่าจะเป็นเส้นเชื่อมต่อโจมตี

ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงสามารถใช้ในการตรวจจับได้เช่นกัน โดยกำหนดให้ฟังก์ชันระดับความสำเร็จในการโน้มน้าวโนด X จากกลุ่ม c มีค่าเป็น

$$f_{X,c}\left(N, n_c\right) = P_{sw}\left(N - n_c, N - n_c, n_c\right) \tag{4.4}$$

เมื่อกำหนดให้ $P_{sw}\left(\right)$ คือความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่นำเสนอในบทที่ 3 และกำหนดให้กลุ่ม C เป็นกลุ่มซิบิลแต่ไม่ทราบจำนวนซิบิลที่แน่ชัดจึงกำหนดให้จำนวนซิบิลมีค่าเป็น n_c

และกำหนดให้ N คือจำนวนเส้นเชื่อมต่อถึงโนด X ทั้งหมดดังนั้นจำนวนผู้ใช้งานจริงจึงมีค่าเท่ากับ $N - n_c$ นอกจากนั้นจำนวนตัวเลือกจะมีค่าเกินจำนวนผู้ออกเสียงไม่ได้ จึงกำหนดให้จำนวนตัวเลือกมีค่าเท่ากับจำนวนผู้ใช้งานจริง อย่างไรก็ตามเนื่องจากฟังก์ชันระดับความสำเร็จในการโน้มน้าวโนด X จากกลุ่ม c ถูกเปลี่ยนนิยามไป จากอัตราการโน้มน้าวดังสมการที่ (4.1) กลายเป็นความน่าจะเป็นที่ซิบิลชนะการออกเสียงดังสมการที่ (4.4) ทำให้ระดับการโน้มน้าวที่ยอมรับได้ (γ) ต้องถูกเปลี่ยนค่าไปด้วย เนื่องจากการตรวจจับซิบิลโดยใช้สมการที่ (4.1) กำหนดให้ระดับการโน้มน้าวมีค่าเป็น $\frac{1}{3}$ ซึ่งเมื่อเปลี่ยนเป็นความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงแล้วความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจะมีค่าไม่ต่ำกว่า 0.9 ดังนั้น ระดับการโน้มน้าวที่สอดคล้องกับสมการที่ (4.4) จึงถูกตั้งค่าให้เป็น 0.9 เช่นกัน

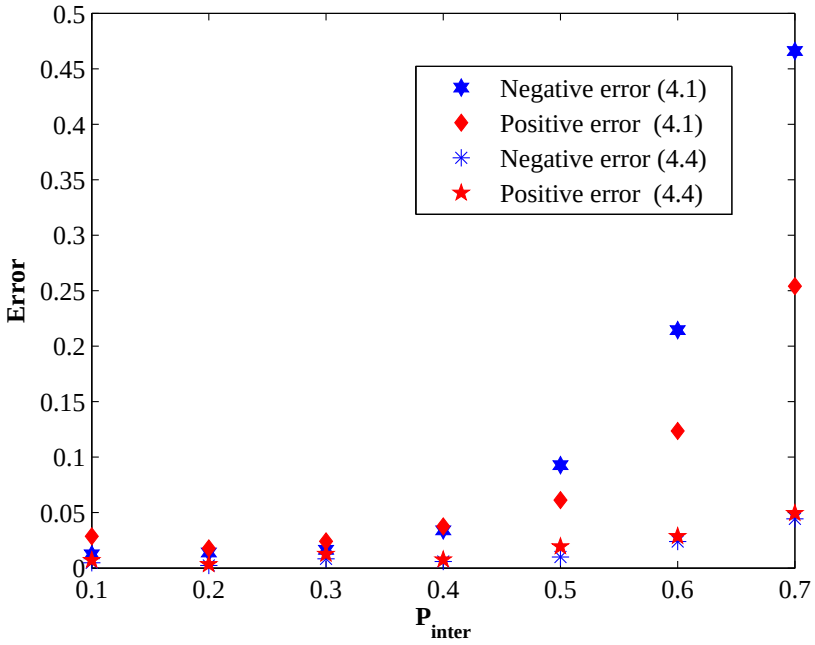
ฟังก์ชันอัตราส่วนการโน้มน้าว (4.1) และฟังก์ชันระดับความสำเร็จในการโน้มน้าว (4.4) มีผลกระทบต่อความแม่นยำของวิธีการตรวจจับซิบิลไม่เท่ากัน ดังแสดงตัวอย่างในรูปที่ 4.10-4.12 เป็นการทดสอบการตรวจจับซิบิลกับโครงข่ายที่มีรูปแบบความสัมพันธ์สองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน โดยทุกรูปกำหนดให้ P_{intra} มีค่าเป็น 0.7, P_{inter} มีค่าเป็น 0.3, Pss มีค่าเป็น 0.9, Psr มีค่าเป็น 0.7, Prs มีค่าเป็น 0.01 จุดแต่ละจุดของกราฟเกิดจากการเฉลี่ยผลการทดลอง 500 ครั้ง รูปแต่ละรูปจะเปลี่ยนแกนนอนของกราฟเพื่อศึกษาผลกระทบของตัวแปรต้น กล่าวคือ รูปที่ 4.10 แสดงตัวอย่างผลกระทบของ P_{inter} เนื่องจากการเชื่อมต่อระหว่างกลุ่มย่อยจะมีอยู่อย่างเบาบางเท่านั้นเมื่อเทียบกับการเชื่อมต่อในกลุ่มเดียวกัน ดังนั้น $P_{\text{inter}} \leq P_{\text{intra}}$ และเนื่องจากกำหนดให้ P_{intra} มีค่าเป็น 0.7 ดังนั้นจึงปรับค่าให้ P_{inter} มีค่าให้อยู่ระหว่าง 0.1-0.7 รูปที่ 4.11 นำเสนอผลกระทบจาก Prs เนื่องจากผู้ใช้งานจริงไม่รู้จักซิบิลเป็นการส่วนตัวทำให้เส้นเชื่อมต่อจากผู้ใช้งานจริงถึงซิบิลมีจำนวนน้อยมากเมื่อเทียบกับจำนวนของเส้นเชื่อมต่อจากผู้ใช้งานจริงถึงผู้ใช้งานจริงอีกโนดหนึ่ง กล่าวคือ $P_{\text{mistaken}} << P_{\text{intra}}$ โดยไม่เสียนัยสำคัญ กำหนดให้ Prs มีค่าน้อยกว่า 1 ใน 4 ของ P_{intra} จึงปรับค่าให้ Prs ซึ่งเป็นแกนนอนของรูปที่ 4.11 อยู่ในช่วงของ 0.02 ถึง 0.2 รูปที่ 4.12 นำเสนอผลกระทบจาก P_{intra} โดยปรับค่าให้ P_{intra} ซึ่งเป็นแกนนอนของรูปที่ 4.12 อยู่ในช่วงของ 0.1-0.7 จากรูปที่ 4.10-4.11 ระดับความผิดพลาดเพิ่มขึ้นเมื่อ P_{inter} และ Prs เพิ่มขึ้น เพราะการเพิ่มขึ้นของ P_{inter} และ Prs จะทำให้โครงข่ายมีรูปแบบที่ความคลุมเคลือมากขึ้น ในทางกลับกันรูปที่ 4.12 แสดงให้เห็น นว่าความผิดพลาดของวิธีการตรวจหาซิบิลที่นำเสนอจะลดลงเมื่อ P_{intra} เพิ่มขึ้นเนื่องจากการเพิ่มขึ้นของ P_{intra} จะทำให้รูปแบบของโครงข่ายมีความชัดเจนมากขึ้น เป็นผลให้ความผิดพลาดของวิธีการตรวจหาซิบิลที่นำเสนอลดลง

จากรูปที่ 4.10-4.12 พบว่าการใช้ฟังก์ชันระดับความสำเร็จในการโน้มน้าว (4.4) ในการตรวจจับซิบิลจะให้ค่าความผิดพลาดต่ำกว่าการใช้อัตราส่วนการโน้มน้าว (4.1) ดังนั้นต่อไปจะใช้ฟังก์ชันระดับความสำเร็จในการโน้มน้าว (4.4) ในการตรวจจับซิบิลวิธีที่ 1 แทนอัตราส่วนการโน้มน้าว (4.1) เพื่อเพิ่มประสิทธิภาพของการตรวจจับซิบิล

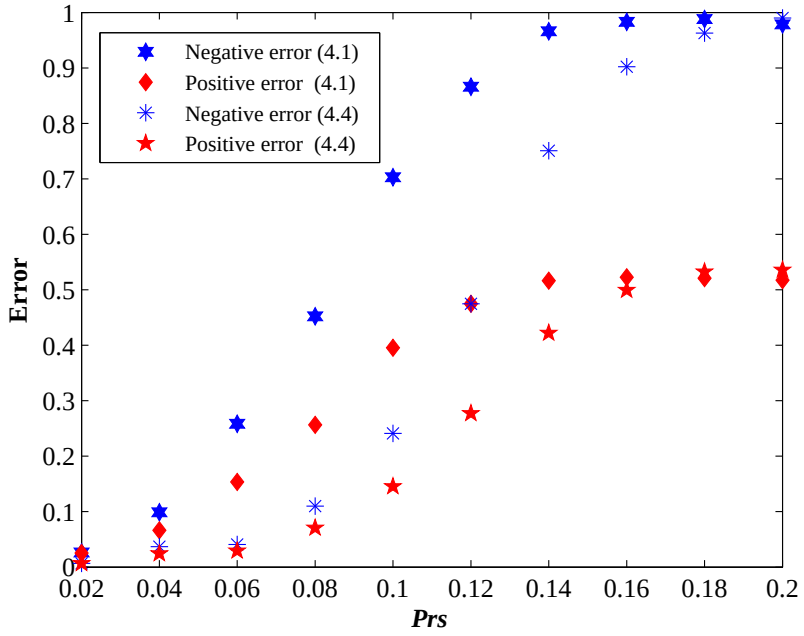
4.2.2 วิธีที่ 2: การใช้ความเหมือนของคู่โนดในโครงข่ายและความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการลดความผิดพลาดการตรวจจับซิบิล

นอกจากจะใช้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการตรวจจับซิบิลแล้ว ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงใช้ในการลดความผิดพลาดในการตรวจจับซิบิลได้ วิทยานิพนธ์นี้จึงนำเสนอวิธีการใช้ความเหมือนของคู่โนดในโครงข่ายและความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการตรวจสอบความถูกต้อง เป็นอีกตัวอย่างวิธีการจับซิบิล

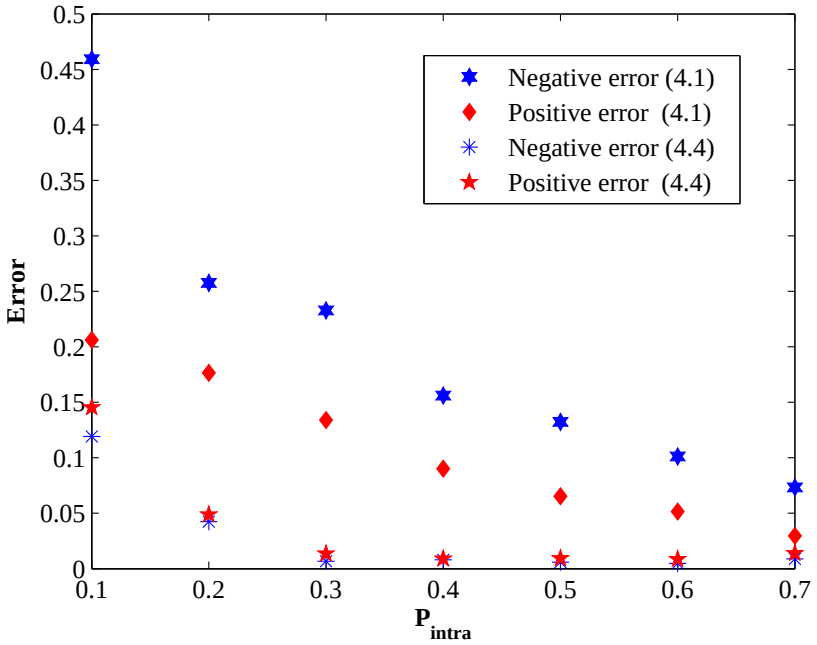
สำหรับโนด x และโนด y ในโครงข่าย กำหนดให้ $Ne(x)$ คือเซตของโนดในโครงข่ายที่โนด



รูปที่ 4.10: ผลกระทบของ P_{inter} ต่อความแม่นยำของวิธีการตรวจสอบซิบิลที่นำเสนอ



รูปที่ 4.11: ผลกระทบของ Prs ต่อความแม่นยำของวิธีการตรวจสอบซิบิลที่นำเสนอ



รูปที่ 4.12: ผลกระทบของ P_{intra} ต่อความแม่นยำของวิธีการตรวจสอบซิบิลที่นำเสนอ

x มีเส้นเชื่อมต่อถึง และ $Ne\,(y)$ คือเซตของโนดในโครงข่ายที่โนด y มีเส้นเชื่อมต่อถึง กำหนดให้ $(sim(x,y))$ คือความคล้าย (similarity) ของโนด x และโนด y โดย $sim(x,y) = \frac{|Ne(x) \cap Ne(y)|}{|Ne(x) \cup Ne(y)|}$ ยิ่งความคล้ายระหว่างโนด x และโนด y มีค่ามาก ยิ่งมีโอกาสสูงที่โนด x และโนด y จะมีสถานะเหมือนกัน อย่างไรก็ตามเพียงความคล้ายกันระหว่างโนดอย่างเดียวไม่สามารถตัดสินสถานะของโนดได้ ตัวอย่างเช่น โนด x มีเส้นเชื่อมต่อถึงโนดอื่น 2 โนด ได้แก่ โนด a_1 และโนด a_2 เช่นเดียวกับโนด y ที่มีเส้นเชื่อมต่อถึงโนดอื่น 2 โนด ได้แก่ โนด a_1 และโนด a_3 จะได้ว่า $sim(x,y) = \frac{1}{3}$ แต่ในสถานการณ์อีกสถานการณ์หนึ่งสมมติว่า โนด x มีเส้นเชื่อมต่อถึงโนดอื่น 18 โนด ได้แก่ โนด $a_1, a_2, a_3, \ldots, a_{18}$ เช่นเดียวกับโนด y ที่มีเส้นเชื่อมต่อถึงโนดอื่น 18 โนด $a_{10}, a_{11}, a_{12}, \ldots, a_{27}$ จะได้ว่า $sim(x,y) = \frac{9}{27} = \frac{1}{3}$ เช่นกัน จะเห็นว่าสถานการณ์ทั้งสองสถานการณ์มีค่า $sim(x,y)$ เท่ากันแต่มีนัยต่างกัน โดยสถานการณ์แรก $|Ne\,(x) \cap Ne\,(y)| = 1$ โดยโนด x และโนด y อาจจะมีความสัมพันธ์ถึงโนด a_1 โดยความบังเอิญได้ ในขณะที่ในสถานการณ์ที่สอง $|Ne\,(x) \cap Ne\,(y)| = 9$ ซึ่งมีค่ามากกว่ากรณีแรก ทำให้มีโอกาสสูงมากกว่าที่โนด x และโนด y จะมีสถานะเหมือนกัน เป็นต้น ดังนั้นจึงควรพิจาณาค่าของ $|Ne\,(x) \cap Ne\,(y)|$ ร่วมด้วย

วิทยานิพนธ์นี้จึงเสนอ ความเหมือน (conformity) ของโนด x และโนด y $(con(x,y))$ คำนวณจาก

$$con(x,y) = 1 - (1 - sim\,(x,y))^{|Ne(x) \cap Ne(y)|} \tag{4.5}$$

ความเหมือนของโนด x และโนด y จะมีค่าตั้งแต่ 0 ถึง 1 โดยมีค่าเป็น 0 เมื่อโนด x และโนด y ไม่มีเส้นเชื่อมต่อถึงโนดเดียวกันเลย แต่จะมีค่าเป็น 1 เมื่อโนด x และโนด y มีเส้นเชื่อมต่อถึงโนดอื่นเหมือนกันทั้งหมด ทั้งนี้ความน่าจะเป็นที่โนด x และโนด y จะมีสถานะเหมือนกันจะแปรผันตรงกับความเหมือนของโนด x และโนด y ดังนั้นจึงใช้ความเหมือนของโนด x และโนด y เป็นส่วนประกอบของวิธีการตรวจจับซิบิลดังนี้

1. แบ่งกลุ่มย่อยในโครงข่ายด้วยการคำนวณ $con(x, y)$ สำหรับคู่โนด x และโนด y ทั้งหมดในโครงข่าย ถ้าโนด x และโนด y มีความเหมือนมากกว่า ε (ระดับความเหมือนขั้นต่ำที่ยอมรับได้) ให้ถือว่าโนด x และโนด y อยู่ในกลุ่มย่อยเดียวกัน (วิทยานิพนธ์นี้กำหนดให้ ε มีค่าเป็น 0.9 เพื่อสื่อว่าความเหมือนของโนดจะต้องมีมากเพียงพอที่จะตัดสินว่าอยู่กลุ่มย่อยเดียวกัน ทั้งนี้ตัวเลขดังกล่าวสามารถปรับเปลี่ยนได้ในอนาคตตามความเหมาะสมของแต่ละโครงข่าย)

2. พิจารณากลุ่มย่อยแต่ละกลุ่ม เรียกกลุ่มย่อยที่กำลังพิจารณาอยู่ว่ากลุ่มย่อย C คำนวณ $con(x, y)$ สำหรับคู่โนด x และโนด y ทั้งหมดในกลุ่มย่อย C โดยครั้งนี้ไม่นับรวมเส้นเชื่อมต่อภายในกลุ่มย่อย C ในการคำนวณ $con(x, y)$ ด้วย ในกรณีที่ $con(x, y)$ ของคู่โนด x และโนด y มีค่าต่ำกว่า ε ให้ถือว่าโนด x และโนด y ไม่เหมือนกัน

3. พิจารณาโนดแต่ละโนด x ในกลุ่มย่อย C ถ้ามีจำนวนโนดที่ไม่เหมือนโนด x อย่างน้อยครึ่งหนึ่งของจำนวนสมาชิกในกลุ่มย่อย C ถือว่าโนด x ไม่ใช่ซิบิลเพราะมีความเหมือนกับโนดอื่นในกลุ่มย่อย C น้อยเกินไป ตัวอย่างเช่น กลุ่มย่อย $C = \{a_1, a_2, \ldots, a_7\}$ และกำหนดให้ $x = a_7$ โดย a_7 มีความเหมือนกับโนด a_1 ถึงโนด a_6 เท่ากับ 0.26, 0.81, 0.94, 0.92, 0.63, 0.11 ตามลำดับ จะได้ว่าโนด a_7 มีความเหมือนกับโนดอื่นในกลุ่มย่อย C มากกว่า ε (กำหนดให้มีค่าเป็น 0.9) เพียง 2 โนดเท่านั้น คือ โนด a_3 และโนด a_4 ดังนั้น a_7 จะถือว่าไม่ใช่ซิบิลและถูกตัดออกจากกลุ่มย่อย C เพราะมีความเหมือนกับโนดอื่นในกลุ่มย่อย C น้อยเกินไป เป็นต้น

4. ทำซ้ำขั้นตอนที่ 3 จนกว่าจะไม่มีโนดใดถูกนำออกจากกลุ่มย่อย C อีก ถือว่าโนดที่เหลือในกลุ่มย่อย C ทั้งหมดเป็นซิบิล

5. พิจารณาเหยื่อทั้งหมด เหยื่อที่มีความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงน้อยกว่า δ (ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่ยอมรับได้ ในที่นี้กำหนดให้มีค่าเป็น 0.5 เพราะผู้ไม่หวังดีต้องคาดหวังว่าจะมีโอกาสชนะการออกเสียงมากกว่าไม่ชนะการออกเสียง ทั้งนี้ตัวเลขดังกล่าวสามารถปรับเปลี่ยนได้ในอนาคตตามความเหมาะสมของแต่ละโครงข่าย) ถือว่าเหยื่อนั้นไม่ใช่เหยื่ออีกต่อไป และซิบิลที่ไม่มีเส้นเชื่อมต่อถึงเหยื่อใด ๆ จะถือว่าไม่เป็นซิบิลอีกต่อไปเช่นกัน

ขั้นตอนสุดท้ายใช้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการคัดกรองผู้ใช้งานจริงออกจากกลุ่มซิบิลเพื่อลดความผิดพลาดเชิงบวกของวิธีการตรวจจับซิบิล

4.3 การทดสอบความแม่นยำและความซับซ้อนเชิงเวลาของวิธีการตรวจจับซิบิล

การทดสอบวิธีการตรวจจับซิบิลที่นำเสนอกับโครงข่ายจำลองเป็นสิ่งจำเป็นเพื่อใช้ในการระบุขอบเขตรวมถึงความเหมาะสมในการนำวิธีการตรวจจับซิบิลทั้ง 2 วิธีที่ได้นำเสนอไปใช้ ตัวชี้วัดที่ใช้ในการทดสอบ ได้แก่ ความแม่นยำของผลการตรวจจับซิบิล (ความผิดพลาดเชิงบวกและเชิงลบ) และความซับซ้อนเชิงเวลาของวิธีการตรวจจับซิบิล

ผลการทดสอบนำเสนอในรูปที่ 4.13-4.30 ซึ่งผลการทดสอบทุกรูปมีการตั้งค่าพารามิเตอร์ให้เหมือนกันทั้งหมด กล่าวคือ แกน x แสดงจำนวนผู้ใช้งานจริง กำหนดให้มีค่าเป็น 4,8,16,32,64,128

แกน y แสดงจำนวนซิบิล กำหนดให้มีค่าเป็น 2,4,8,16,32,64 แกน x และแกน y ถูกกำหนดให้แสดงค่าแบบเลขชี้กำลังฐาน 2 เพื่อให้เห็นผลกระทบของการปรับมาตรา (scaling effect) ของการเปลี่ยนแปลงจำนวนผู้ใช้งานจริงและจำนวนซิบิล และแกน z แสดงความผิดพลาดของวิธีการตรวจจับซิบิล มีค่าตั้งแต่ 0 ถึง 1 กำหนดให้มีผู้ใช้งานจริงที่เป็นเป้าหมายของซิบิล 3 คน ความน่าจะเป็นที่ผู้ใช้งานจริงคนหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอีกคนหนึ่ง (Prr) เท่ากับ 0.4 ความน่าจะเป็นที่ผู้ใช้งานจริงคนหนึ่งจะมีเส้นเชื่อมต่อถึงซิบิล (Prs) เท่ากับ 0.0001 ความน่าจะเป็นที่ซิบิลตัวหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงคนหนึ่ง (Psr) เป็น 0.9 และความน่าจะเป็นที่ซิบิลตัวหนึ่งจะมีเส้นเชื่อมต่อถึงซิบิลอีกตัวหนึ่ง (Pss) เป็น 0.8 จุดของกราฟผลการทดลองแต่ละจุดเกิดจากการทำการทดลอง 100 ครั้งแล้วนำค่าที่ได้มาเฉลี่ย ทั้งนี้ผลที่ได้จากการทดสอบจะถูกวิเคราะห์ถึงความเหมาะสมในสถานการณ์ที่สมควรนำไปใช้ต่อไป

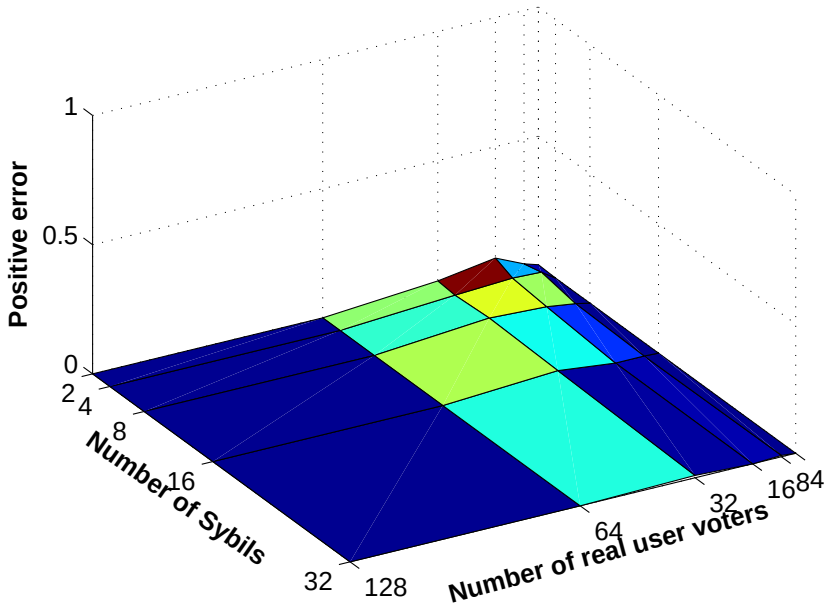
4.3.1 วิธีที่ 1: การแบ่งกลุ่มย่อยแล้วตรวจจับซิบิลเป็นกลุ่ม

การทดสอบในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว

วิธีการแบ่งกลุ่มย่อยและตรวจจับกลุ่มของซิบิลถูกแบ่งออกเป็น 2 ขั้นตอนใหญ่ ได้แก่ ขั้นตอนการแบ่งกลุ่มย่อยโดยใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วที่ได้นำเสนอโดย [29] และขั้นตอนการตรวจจับกลุ่มของซิบิลโดยใช้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงตามสมการที่ (4.4) ดังนั้นความคลาดเคลื่อนในการตรวจจับซิบิลที่นำเสนอจึงเป็นความคลาดเคลื่อนรวมของทั้งสองขั้นตอน อย่างไรก็ตามเพื่อชี้วัดความแม่นยำและความซับซ้อนเชิงเวลาของวิธีการตรวจจับซิบิลของขั้นตอนการตรวจจับกลุ่มของซิบิล (ขั้นตอนหลัง) ให้ชัดเจน ผู้วิจัยจึงทำการทดลองทั้งแบบที่ใช้ค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วในการแบ่งกลุ่มย่อยและแบบที่สมมุติให้มีการแบ่งกลุ่มย่อยอย่างสมบูรณ์แบบ (แบ่งกลุ่มย่อยให้ผู้ใช้งานจริงและซิบิลไม่อยู่ปะปนกัน แต่ไม่ระบุสถานะของโนดตั้งแต่แรกเพื่อใช้ขั้นตอนการตรวจจับกลุ่มของซิบิลเป็นตัวตรวจจับซิบิลและผู้ใช้งานจริงเอง)

รูปที่ 4.13 แสดงความผิดพลาดเชิงบวกของการตรวจจับซิบิลด้วยวิธีการแบ่งกลุ่มย่อยแล้วตรวจจับซิบิลเป็นกลุ่มในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว โดยแกน x แสดงจำนวนผู้ใช้งานจริง กำหนดให้มีค่าเป็น 4,8,16,32,64,128 แกน y แสดงจำนวนซิบิล กำหนดให้มีค่าเป็น 2,4,8,16,32,64 แกน x และแกน y ถูกกำหนดให้แสดงค่าแบบเลขชี้กำลังฐาน 2 เพื่อให้เห็นผลกระทบของการปรับมาตรา (scaling effect) ของการเปลี่ยนแปลงจำนวนผู้ใช้งานจริงและจำนวนซิบิล และแกน z แสดงความผิดพลาดเชิงบวก มีค่าตั้งแต่ 0 ถึง 1 จากผลการทดสอบพบว่าความผิดพลาดเชิงบวกของการตรวจจับซิบิลในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วมากที่สุดไม่เกิน 6.94% และมีค่าเฉลี่ย 1.17% สำหรับความผิดพลาดเชิงลบของการตรวจจับซิบิลในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว ดังแสดงในรูปที่ 4.14 พบว่าความผิดพลาดเชิงลบมีค่าใกล้เคียง 1 เนื่องจากวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วให้ผลการแบ่งกลุ่มย่อยว่าโครงข่ายทั้งโครงข่ายเป็นกลุ่มเดียวกันทั้งหมดและไม่สามารถแบ่งโครงข่ายออกเป็นกลุ่มย่อยได้ ทำให้โนดทุกโนดในโครงข่ายถูกระบุว่าเป็นผู้ใช้งานจริงส่งผลให้ความผิดพลาดเชิงลบมีค่าเป็น 1 สำหรับกรณีการแบ่งกลุ่มสมบูรณ์แบบไม่มีความผิดพลาดเชิงบวกตลอดช่วงทดสอบดังแสดงในรูปที่ 4.15 และมีความผิดพลาดเชิงลบในกรณีที่จำนวนซิบิลน้อยมากเท่านั้นดังแสดงในรูปที่ 4.16 เนื่องจากวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วไม่สามารถตรวจจับซิบิลได้ในกรณีที่มีผลกระทบของการโจมตีต่ำจนสามารถละเลยได้ อย่างไรก็ตามพบว่าความผิดพลาดเชิงบวกและเชิงลบของการตรวจจับซิบิลในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วมีค่ามากกว่าความผิด

พลาดเชิงบวกและเชิงลบของการตรวจจับซิบิลในกรณีการแบ่งกลุ่มย่อยอย่างสมบูรณ์แบบตามลำดับ ทำให้ทราบว่าความผิดพลาดส่วนใหญ่เกิดจากวิธีการแบ่งกลุ่มย่อย ดังนั้นวิธีในการแบ่งกลุ่มย่อยอาจปรับเปลี่ยนได้ตามความเหมาะสมในอนาคต และเมื่อเลือกวิธีในการแบ่งกลุ่มย่อยอย่างดีที่สุดแล้วจะสามารถลดความผิดพลาดในการตรวจจับซิบิลได้ไม่เกินความผิดพลาดของการตรวจจับซิบิลในกรณีการแบ่งกลุ่มย่อยอย่างส มบูรณ์แบบ

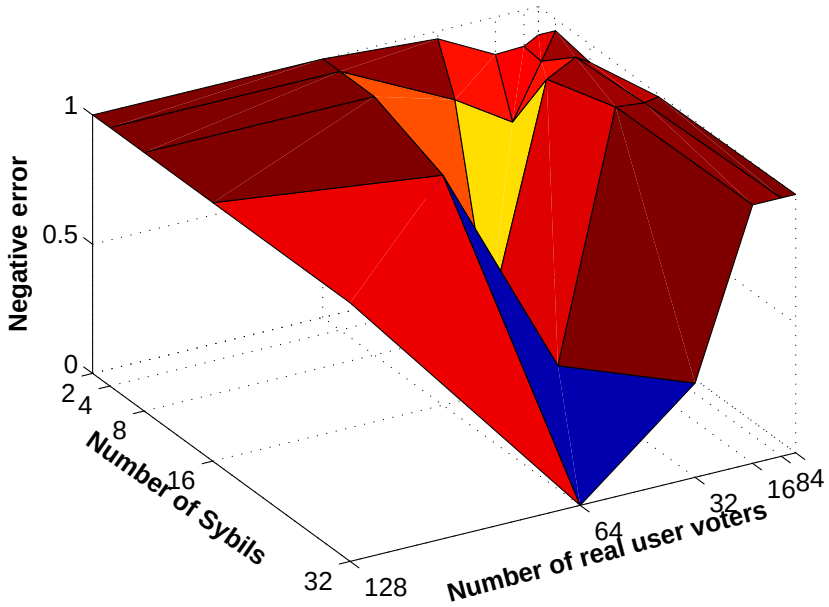


รูปที่ 4.13: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียวในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วในการแบ่งกลุ่มย่อย

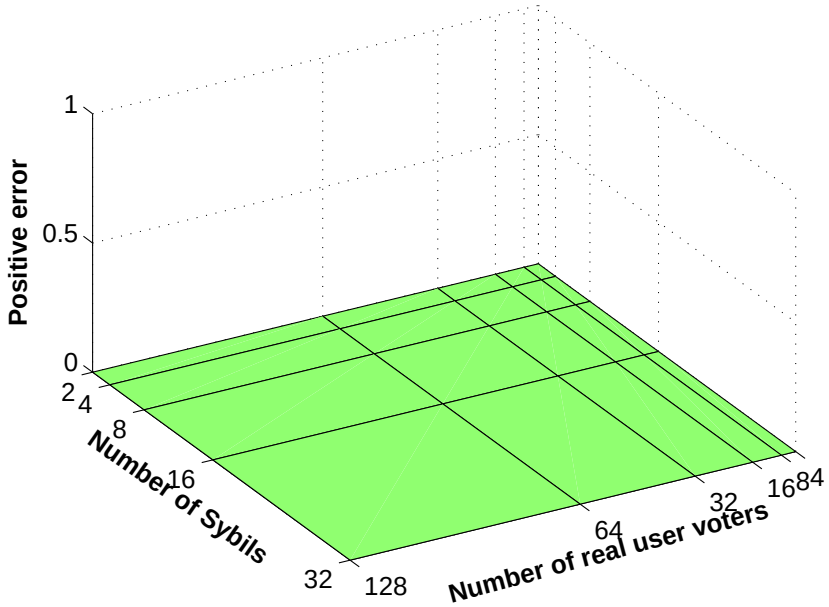
ทั้งนี้สังเกตว่าความผิดพลาดเชิงบวกและเชิงลบไม่ได้เป็นส่วนเติมเต็มซึ่งกันและกันเนื่องจากการตรวจจับซิบิลได้ถูกต้องมากขึ้นไม่ได้ทำให้การระบุผู้ใช้งานจริงผิดพลาดน้อยลง ในทางกลับกันการระบุผู้ใช้งานจริงได้ถูกต้องมากขึ้นไม่ได้หมายความว่าจะตรวจจับซิบิลได้ถูกต้องมากขึ้นเช่นกัน

การทดสอบในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน

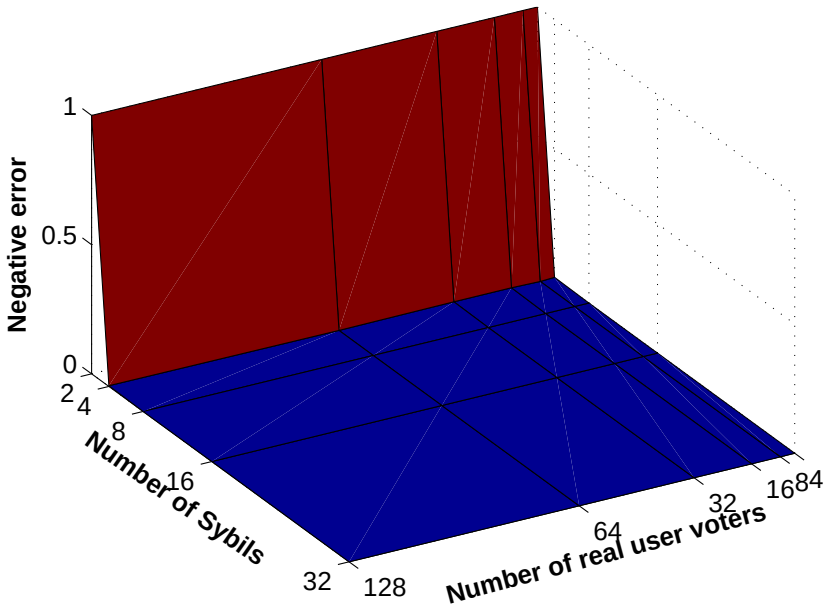
รูปที่ 4.17-4.20 นำเสนอผลการทดสอบวิธีที่นำเสนอกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน พบว่าความผิดพลาดเชิงบวกของการตรวจจับซิบิลในกรณีมากที่สุดไม่เกิน 33.94% และมีค่าเฉลี่ย 9.05% ดังแสดงในรูปที่ 4.17 ซึ่งมีค่าสูงกว่าเมื่อทดสอบกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว เนื่องจากเส้นเชื่อมต่อสองทางจะทำให้เกิดวงจร (cycle) ในโครงข่ายได้ง่ายกว่าโครงข่ายที่มีเพียงเส้นเชื่อมทางเดียวทำให้วิธีการระบุกลุ่มย่อยทำได้ยากกว่าและมีความถูกต้องน้อยกว่าการทดสอบกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว สำหรับรูปที่ 4.18 มีลักษณะสอดคล้องกันกับรูปที่ 4.14 โดยที่มีความผิดพลาดเชิงลบมีค่าใกล้เคียง 1 เนื่องจากวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วให้ผลการแบ่งกลุ่มย่อยว่าโครงข่ายทั้งโครงข่ายเป็นกลุ่มเดียวกันทั้งหมดและไม่สามารถแบ่งโครงข่ายออกเป็นกลุ่มย่อยได้ และรูปที่ 4.19-4.20 ให้ผลการ



รูปที่ 4.14: ความผิดพลาดเชิงลบของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียวในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วในการแบ่งกลุ่มย่อย



รูปที่ 4.15: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียวในกรณีการแบ่งกลุ่มสมบูรณ์แบบ

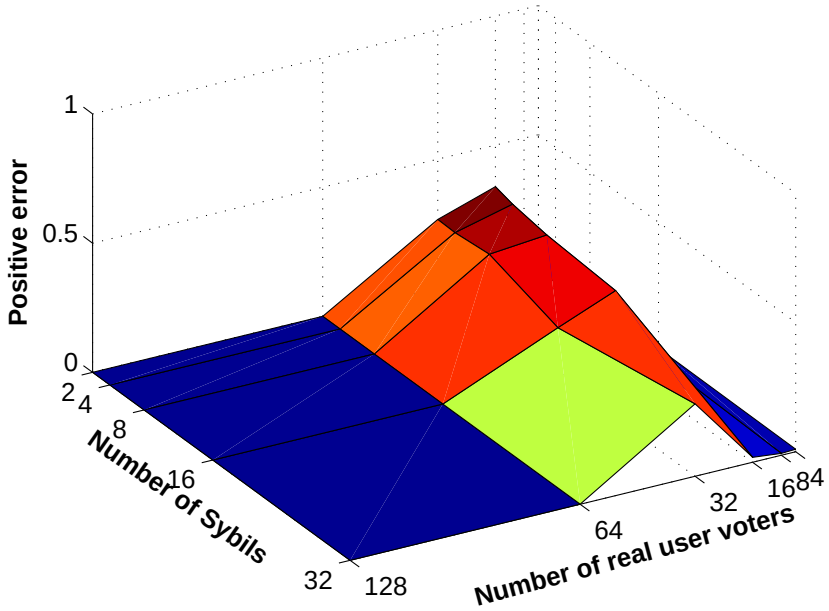


รูปที่ 4.16: ความผิดพลาดเชิงลบของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียวในกรณีการแบ่งกลุ่มสมบูรณ์แบบ

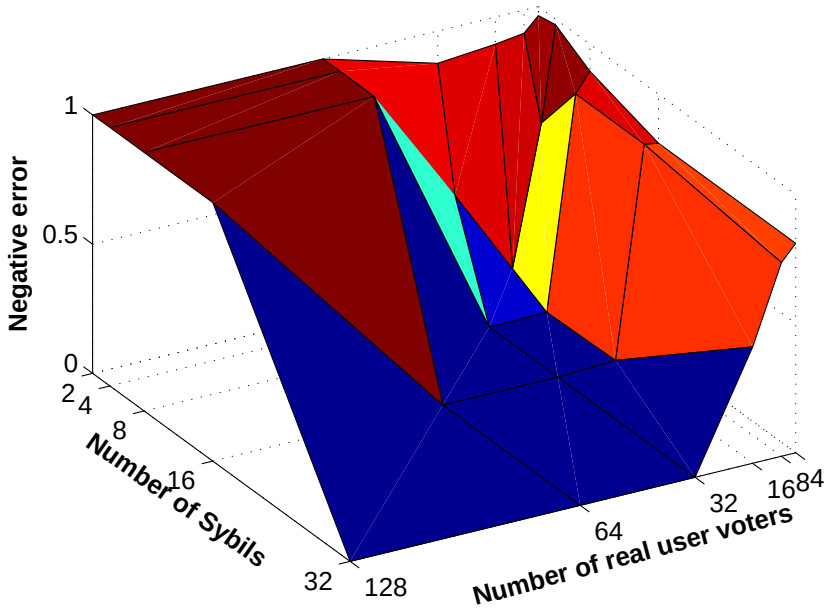
ทดสอบที่เหมือนกับการทดสอบกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว

การทดสอบในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน

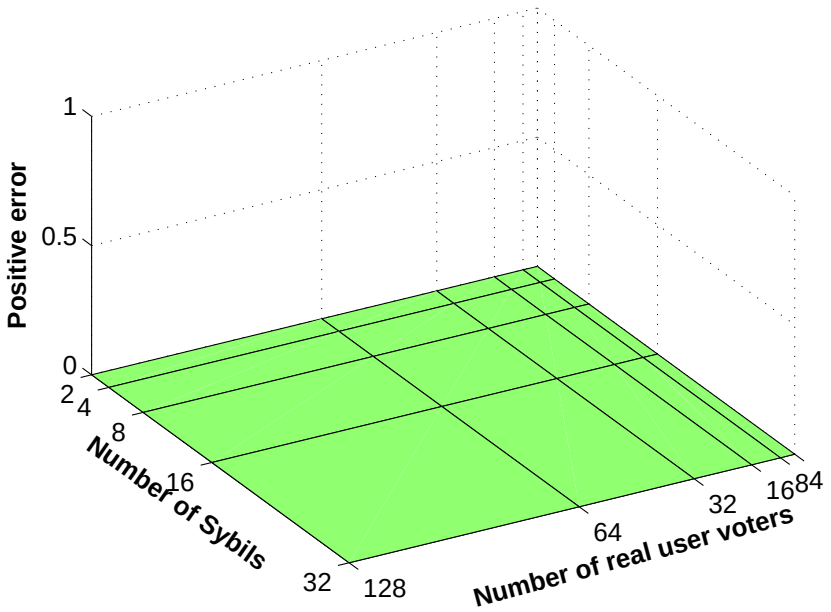
รูปที่ 4.21-4.24 นำเสนอผลการทดสอบวิธีที่นำเสนอกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน พบว่าความผิดพลาดเชิงบวกของการตรวจจับซิบิลในกรณีนี้มีค่ามากที่สุดไม่เกิน 30.75% และมีค่าเฉลี่ย 5.94% ดังแสดงในรูปที่ 4.21 ซึ่งมีค่าสูงกว่าเมื่อทดสอบกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียวแต่ต่ำกว่าเมื่อทดสอบกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน เนื่องจากการตั้งนิยามของ Prr เป็นขอบเขตสูงสุดของความน่าจะเป็นที่ผู้ใช้งานจริงคนหนึ่งจะมีเส้นเชื่อมต่อถึงผู้ใช้งานจริงอีกคนหนึ่ง ทำให้ความน่าจะเป็นที่แท้จริงมีค่าน้อยกว่า Prr และทำให้โครงข่ายในส่วนของผู้ใช้งานจริงมีความคลุมเคลือมากขึ้นในขณะที่เส้นเชื่อมต่อในส่วนอื่นยังชัดเจนเหมือนเดิม ทำให้ผู้ใช้งานจริงบางส่วนถูกจัดให้อยู่ในกลุ่มเดียวกันกับซิบิล จึงทำให้ความผิดพลาดเชิงบวกสูงขึ้น สำหรับรูปที่ 4.22 ลักษณะสอดคล้องกันกับรูปที่ 4.19 โดยที่มีความผิดพลาดเชิงลบมีค่าใกล้เคียง 1 เนื่องจากวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วให้ผลการแบ่งกลุ่มย่อยว่าโครงข่ายทั้งโครงข่ายเป็นกลุ่มเดียวกันทั้งหมดและไม่สามารถแบ่งโครงข่ายออกเป็นกลุ่มย่อยได้ และรูปที่ 4.23-4.24 ให้ผลการทดสอบที่เหมือนกับการทดสอบกับโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว



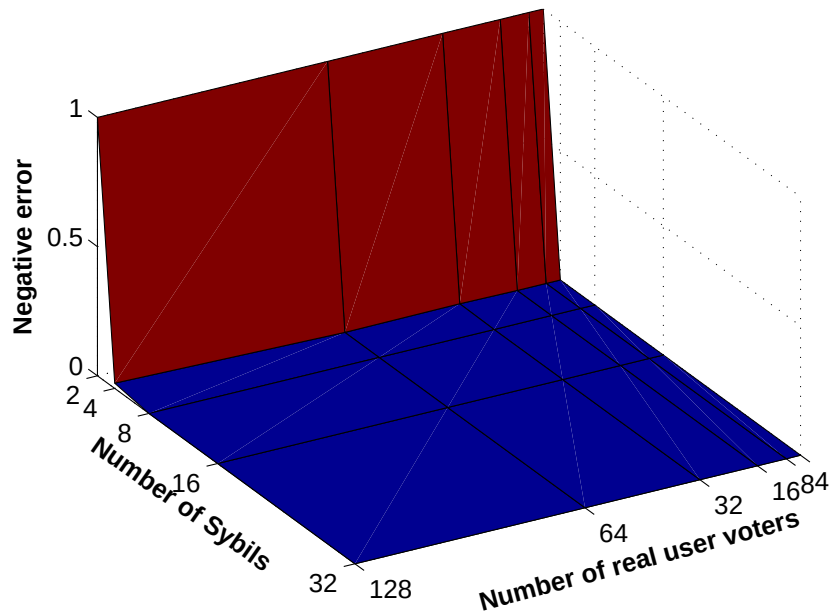
รูปที่ 4.17: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วในการแบ่งกลุ่มย่อย



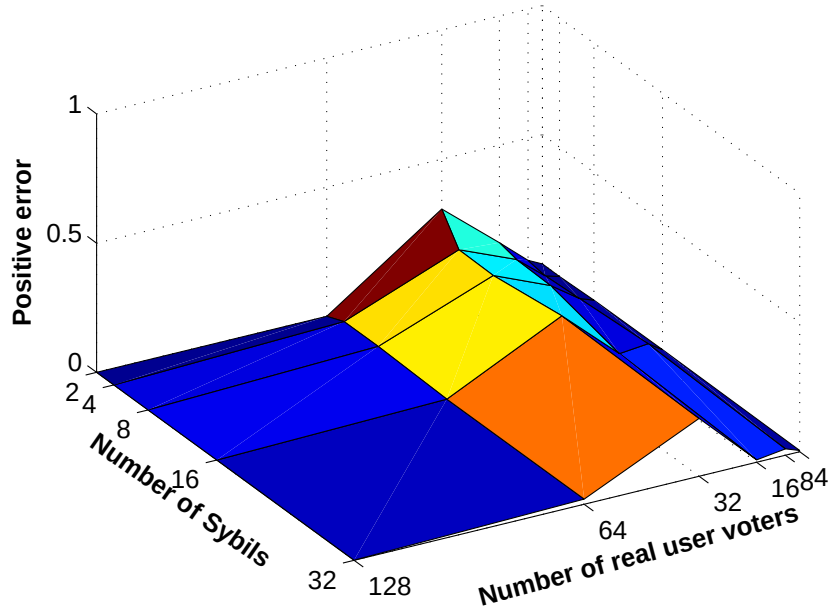
รูปที่ 4.18: ความผิดพลาดเชิงลบของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วในการแบ่งกลุ่มย่อย



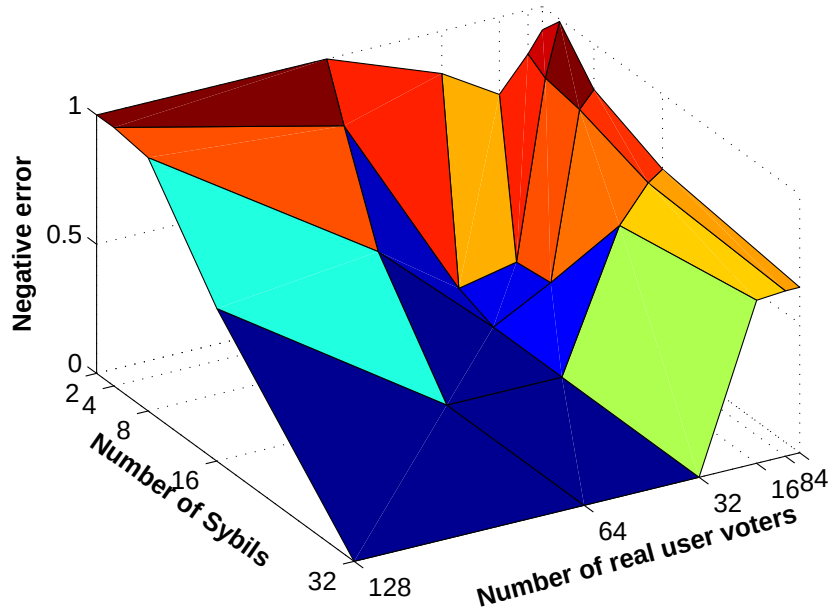
รูปที่ 4.19: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันในกรณีการแบ่งกลุ่มสมบูรณ์แบบ



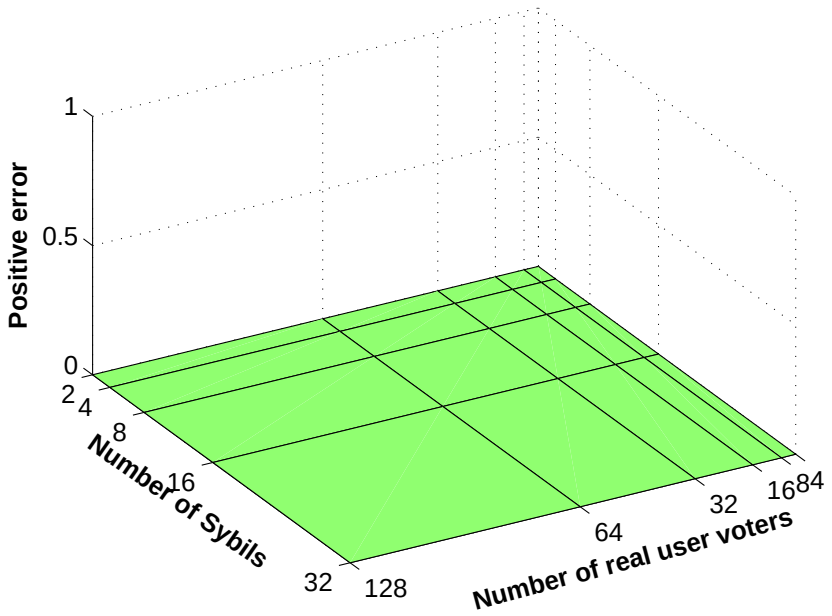
รูปที่ 4.20: ความผิดพลาดเชิงลบของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันในกรณีการแบ่งกลุ่มสมบูรณ์แบบ



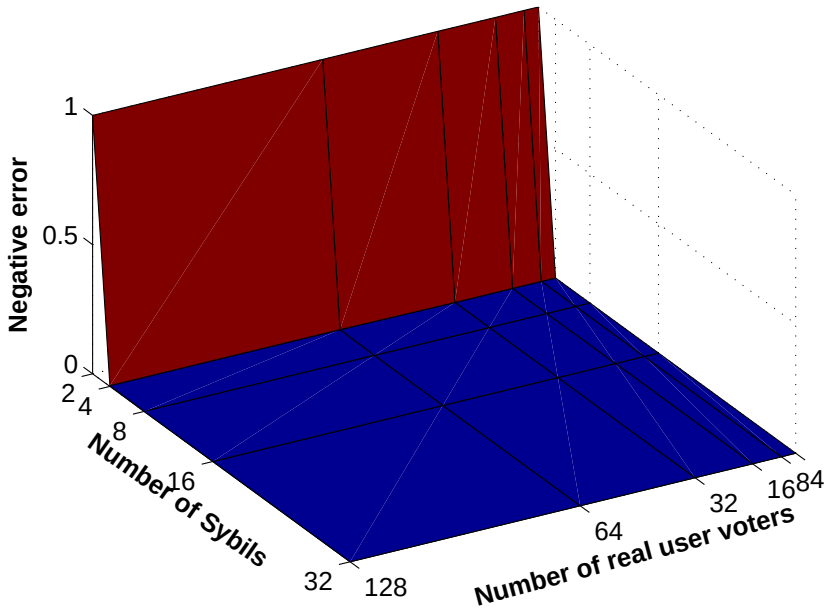
รูปที่ 4.21: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพเป็นส่วนจำเพาะอย่างเร็วในการแบ่งกลุ่มย่อย



รูปที่ 4.22: ความผิดพลาดเชิงลบของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในกรณีที่ใช้วิธีการหาค่าเหมาะที่สุดของสภาพเป็นส่วนจำเพาะอย่างเร็วในการแบ่งกลุ่มย่อย



รูปที่ 4.23: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในกรณีการแบ่งกลุ่มสมบูรณ์แบบ



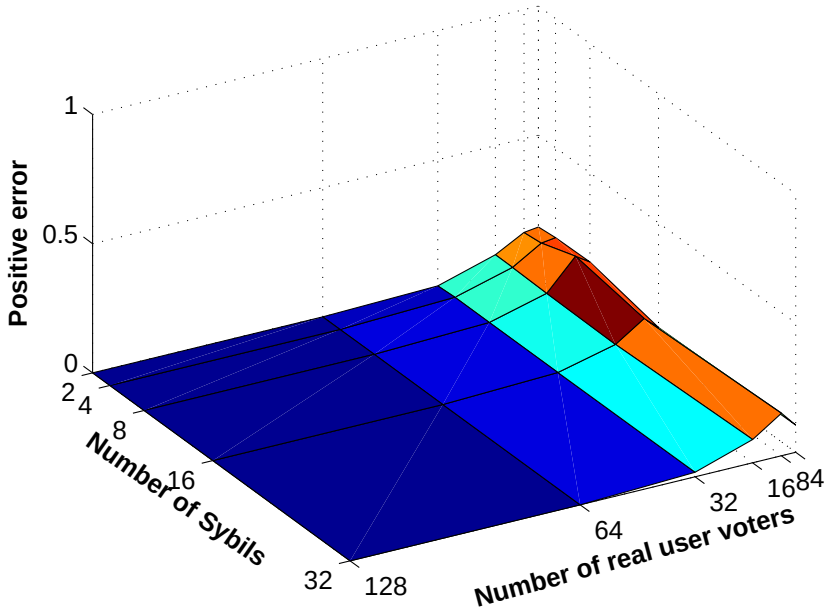
รูปที่ 4.24: ความผิดพลาดเชิงลบของการตรวจจับซิบิลวิธีที่ 1 ในโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันในกรณีการแบ่งกลุ่มสมบูรณ์แบบ

4.3.2 วิธีที่ 2: การใช้ความเหมือนของคู่โนดในโครงข่ายและความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการลดความผิดพลาด

วิธีการใช้ความเหมือนของคู่โนดในโครงข่ายและความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการลดความผิดพลาดถูกทดสอบทั้งในโครงข่ายที่มีรูปแบบความสัมพันธ์ทางเดียวและสองทางมีรายละเอียดดังต่อไปนี้

การทดสอบในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว

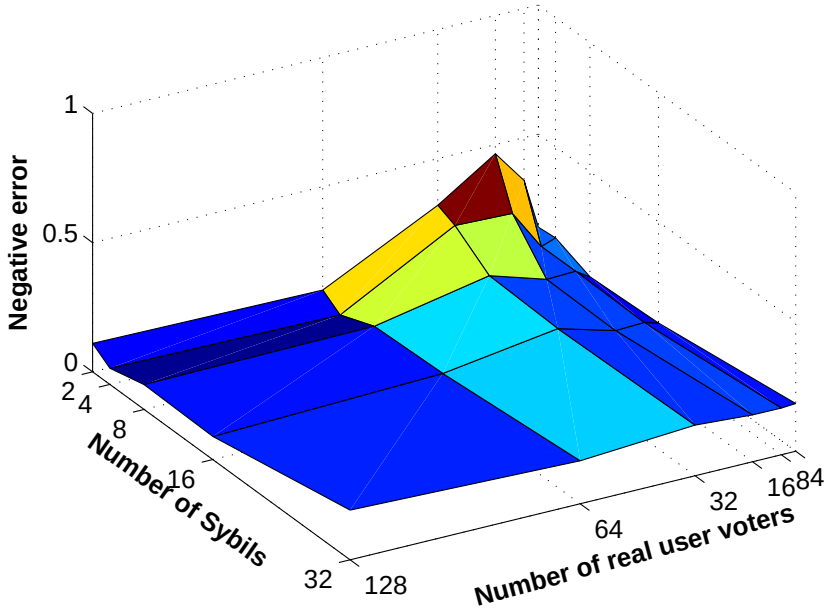
ความผิดพลาดเชิงบวกของการตรวจจับซิบิลในกรณีนี้มีค่ามากที่สุดไม่เกิน 19.13% และมีค่าเฉลี่ย 6.35% ดังแสดงในรูปที่ 4.25 ความผิดพลาดเชิงลบมากที่สุดมีค่าไม่เกิน 46.5% และมีค่าเฉลี่ย 16.72% ดังแสดงในรูปที่ 4.26 จากผลการทดสอบพบว่าความผิดพลาดเชิงบวกและเชิงลบลดลงเมื่อจำนวนผู้ใช้งานจริงและซิบิลเพิ่มขึ้นซึ่งเป็นผลดีในการประยุกต์ใช้ในทางปฏิบัติ



รูปที่ 4.25: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลในโครงข่ายที่มีเส้นเชื่อมต่อทางเดียวโดยใช้วิธีที่ 2

การทดสอบในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน

ความผิดพลาดเชิงบวกของการตรวจจับซิบิลในกรณีนี้มีค่ามากที่สุดไม่เกิน 27.5% และมีค่าเฉลี่ย 5.69% ดังแสดงในรูปที่ 4.27 ความผิดพลาดเชิงลบมากที่สุดในช่วงที่ทดสอบมีค่าไม่เกิน 69% และมีค่าเฉลี่ย 14.52% ดังแสดงในรูปที่ 4.28 จะเห็นว่าสำหรับโครงข่ายที่มีเส้นเชื่อมต่อสองทางและผู้ใช้งานจริงแต่ละคนมีความนิยมเท่ากันมีความผิดพลาดเชิงบวกใกล้เคียงกับการทดสอบกับโครงข่ายแบบเส้นเชื่อมต่อทางเดียว แต่มีความผิดพลาดเชิงลบน้อยกว่าการทดสอบกับโครงข่ายแบบเส้นเชื่อมต่อ

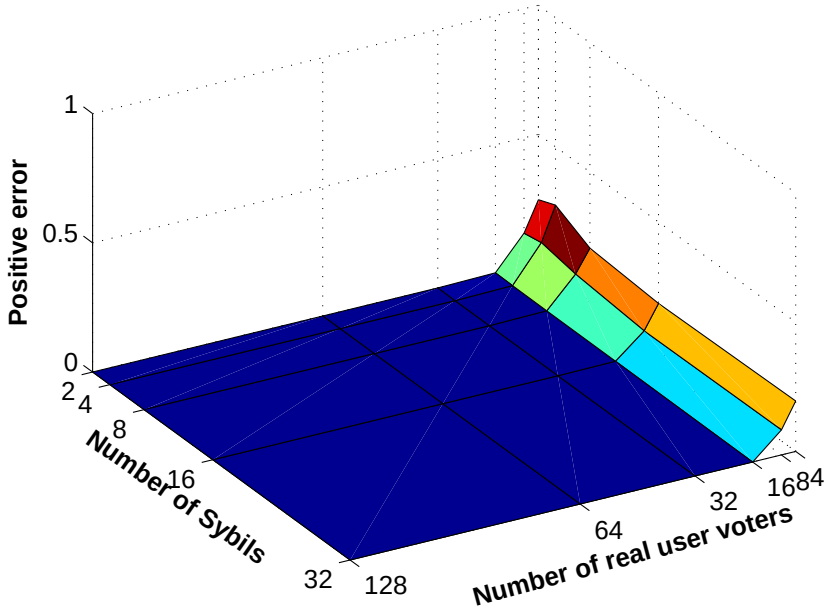


รูปที่ 4.26: ความผิดพลาดเชิงลบของการตรวจจับซิบิลในโครงข่ายที่มีเส้นเชื่อมต่อทางเดียวโดยใช้วิธีที่ 2

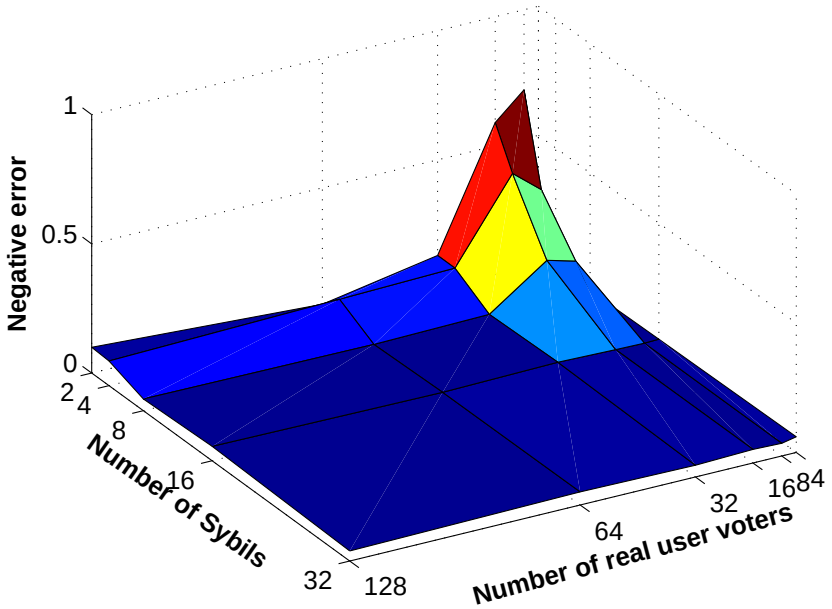
ทางเดียวเนื่องจากเส้นเชื่อมต่อสองทางมีความหมายเหมือนกับเส้นเชื่อมต่อทางเดียว 2 เส้น ทำให้เส้นเชื่อมต่อรวมในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงมีความนิยมเท่ากันจะเสมือนมีเป็นจำนวน 2 เท่าของโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อทางเดียว ทำให้มีข้อมูลมากกว่าและมีความแม่นยำมากกว่าโครงข่ายที่มีเส้นเชื่อมต่อทางเดียว

การทดสอบในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน

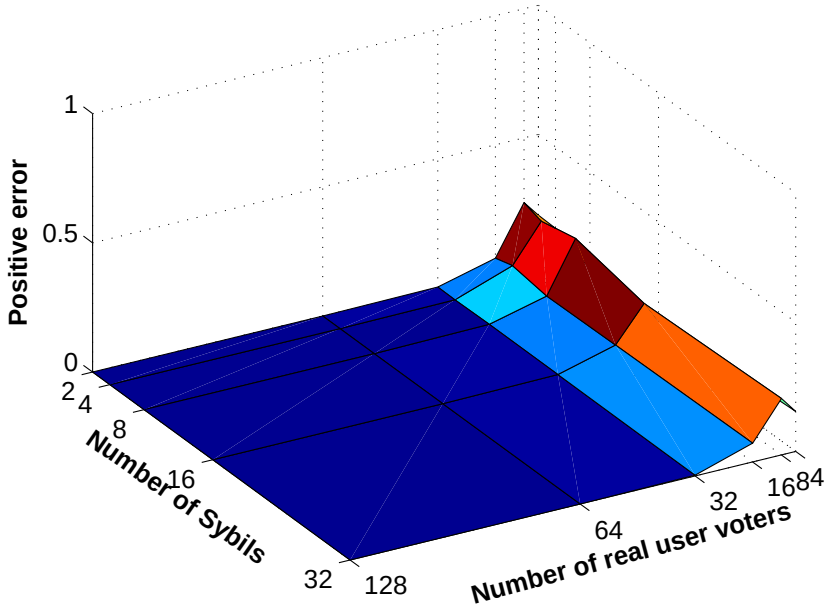
ความผิดพลาดเชิงบวกของการตรวจจับซิบิลในกรณีนี้มีค่ามากที่สุดไม่เกิน 25.62% และมีค่าเฉลี่ย 7.62% ดังแสดงในรูปที่ 4.29 ความผิดพลาดเชิงลบมากที่สุดในช่วงที่ทดสอบมีค่าไม่เกิน 45% และมีค่าเฉลี่ย 11.02% ดังแสดงในรูปที่ 4.30 ซึ่งความผิดพลาดทั้งเชิงบวกและเชิงลบมีค่าสูงกว่าโครงข่ายที่ความน่าจะเป็นที่ผู้ใช้งานจริงคนหนึ่งจะมีเส้นเชื่อมต่อจากผู้ใช้งานจริงอื่นเท่ากันเพราะเมื่อกำหนดให้ความน่าจะเป็น Prr มีค่าเท่ากัน โครงข่ายที่ผู้ใช้งานจริงมีความนิยมไม่เท่ากันจะมีเส้นเชื่อมต่อรวมน้อยกว่าโครงข่ายที่ผู้ใช้งานจริงมีความนิยมเท่ากัน ทำให้ความเหมือนของโนด x และโนด y มีค่าลดลงซึ่งจะทำให้ตรวจจับซิบิลได้ยากขึ้น



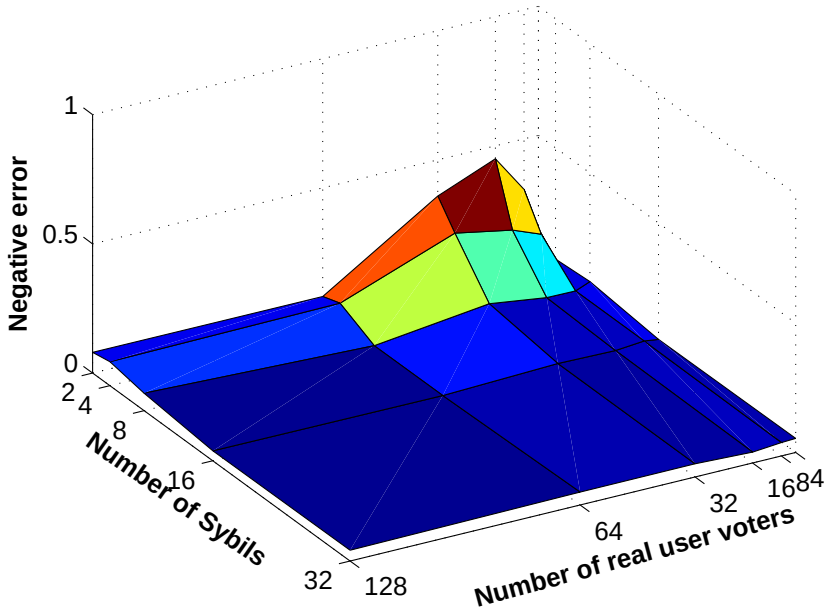
รูปที่ 4.27: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันโดยใช้วิธีที่ 2



รูปที่ 4.28: ความผิดพลาดเชิงลบของการตรวจจับซิบิลในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากันโดยใช้วิธีที่ 2



รูปที่ 4.29: ความผิดพลาดเชิงบวกของการตรวจจับซิบิลในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันโดยใช้วิธีที่ 2



รูปที่ 4.30: ความผิดพลาดเชิงลบของการตรวจจับซิบิลในโครงข่ายที่มีรูปแบบเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากันโดยใช้วิธีที่ 2

4.4 การเปรียบเทียบความซับซ้อนเชิงเวลา ขอบเขตความสามารถของวิธี-การตรวจจับซิบิลที่นำเสนอ และกลยุทธ์การป้องกันซิบิล

ความซับซ้อนเชิงเวลาและขอบเขตความสามารถของวิธีการตรวจจับซิบิลเป็นสิ่งสำคัญที่ต้องคำนึงถึง ความซับซ้อนเชิงเวลาในวิทยานิพนธ์นี้ชี้วัดจากระยะเวลาในการตรวจจับซิบิลในรูปที่ 4.13-4.30 แต่ละรูปประกอบด้วยผลการทดลองจำนวน 30 จุด จุดแต่ละจุดเกิดจากการเฉลี่ยผลการทดลอง 100 ครั้ง ดังนั้นกราฟผลการทดลองแต่ละกราฟจึงเกิดจากการทดลองทั้งสิ้น 3,000 ครั้ง ทั้งนี้กราฟที่ 4.13-4.14 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 1 โดยใช้การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว [29] บนโครงข่ายที่มีเส้นเชื่อมต่อทางเดียว กราฟที่ 4.15-4.16 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 1 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบบนโครงข่ายที่มีเส้นเชื่อมต่อทางเดียว กราฟที่ 4.17-4.18 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 1 โดยใช้การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วบนโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน กราฟที่ 4.19-4.20 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 1 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบบนโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน กราฟที่ 4.21-4.22 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 1 โดยใช้การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วบนโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน กราฟที่ 4.23-4.24 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 1 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบบนโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน กราฟที่ 4.25-4.26 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 2 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบบนโครงข่ายที่มีเส้นเชื่อมต่อทางเดียว กราฟที่ 4.27-4.28 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 2 บนโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมเท่ากัน กราฟที่ 4.29-4.30 เป็นผลการทดสอบวิธีการตรวจจับซิบิลแบบที่ 2 บนโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน ระยะเวลาที่ต้องใช้ในการคำนวณผลสำหรับการท ดลองตรวจจับซิบิลทั้งสิ้น 3,000 ครั้งในกราฟที่ 4.13-4.30 ถูกนำเสนอในตารางที่ 4.1 โดยเปรียบเทียบในมิติของรูปแบบโครงข่าย วิธีการตรวจจับซิบิล และระยะเวลาที่ใช้ตรวจจับซิบิล จากตารางพบว่าการแบ่งกลุ่มย่อยด้วยวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วใช้เวลาในการประมวลผลมากกว่าแบบสมมุติให้มีการแบ่งกลุ่มย่อยอย่างสมบูรณ์แบบอยู่หลายพันเท่า ดังนั้นความซับซ้อนเชิงเวลาส่วนใหญ่เกิดจากวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว สำหรับรูปที่ 4.25-4.30 เป็นกราฟผลการทดลองจากวิธีที่ 2 กราฟแต่ละกราฟใช้เวลาน้อยกว่าระยะเวลาที่ใช้สำหรับวิธีที่ 1 โดยสรุปคือถ้าต้องการความรวดเร็วในการประมวลผล วิธีที่ 2 สามารถนำไปใช้ในทางปฏิบัติมากกว่าวิธีที่ 1

จากผลการทดสอบจึงสรุปว่าวิธีการตรวจจับซิบิลด้วยวิธีที่ 2 มีความผิดพลาดและระยะเวลาในการตรวจจับซิบิลน้อยกว่าวิธีที่ 1 ทั้งนี้วิธีการตรวจจับซิบิลด้วยวิธีที่ 2 สามารถประยุกต์ใช้กับโครงข่ายที่ผู้ใช้งานมีจำนวนมากได้ โดยยิ่งผู้ใช้งานมีจำนวนมากยิ่งจะทำให้ความผิดพลาดทั้งเชิงบวกและเชิงลบน้อยลง

4.5 สรุป

ในบทนี้ได้กล่าวถึงการสร้างโครงข่ายเพื่อใช้ทดสอบวิธีการตรวจจับซิบิล การประยุกต์ใช้สูตรการคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพื่อแยกแยะสถานะของผู้ใช้งานระบบ ซึ่งแบ่งออก

ตารางที่ 4.1: ระยะเวลาที่ต้องใช้ในวิธีตรวจจับซิบิลในโครงข่ายที่มีรูปแบบต่าง ๆ

รูปแบบโครงข่าย	วิธีการตรวจจับซิบิล	ระยะเวลาทั้งหมดในการคำนวณผล 3,000 จุด (วินาที)
เส้นเชื่อมต่อทางเดียว	วิธีที่ 1 โดยใช้การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว	20691
	วิธีที่ 1 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบ	2.6
	วิธีที่ 2	960
เส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน	วิธีที่ 1 โดยใช้การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว	26995
	วิธีที่ 1 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบ	3.1
	วิธีที่ 2	2030
เส้นเชื่อมต่อสองทางโดยผู้ใช้งานจริงแต่ละโนดมีความนิยมไม่เท่ากัน	วิธีที่ 1 โดยใช้การหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็ว	16272
	วิธีที่ 1 โดยสมมุติให้การแบ่งกลุ่มสมบูรณ์แบบ	2.6
	วิธีที่ 2	1087

เป็น 2 วิธี โดยวิธีแรกใช้กลยุทธ์การแบ่งกลุ่มย่อยแล้วตรวจจับซิบิลเป็นกลุ่ม และวิธีที่ 2 ใช้ความคล้ายของคู่โนดในโครงข่ายและความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการลดความผิดพลาด รวมถึงนำเสนอผลการทดสอบวิธีดังกล่าวทั้งในโครงข่ายที่มีเส้นเชื่อมต่อทางเดียวและสองทาง โดยสำหรับโครงข่ายที่มีเส้นเชื่อมต่อสองทางทดสอบทั้งกรณีที่ผู้ใช้งานจริงแต่ละคนมีความนิยมเท่าและไม่เท่ากันที่ผู้ใช้งานจริงคนอื่นจะมีเส้นเชื่อมต่อถึง ทั้งนี้ผลการทดสอบชี้ให้เห็นว่าความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่ได้นำเสนอในบทที่ 3 สามารถนำมาประยุกต์ใช้ในการระบุตัวตนของซิบิลได้ นอกจากนั้นการเปรียบเทียบสมรรถนะของวิธีการตรวจจับซิบิลทั้งสองวิธีได้นำเสนอในมุมมองของความซับซ้อนเชิงเวลา ขอบเขตความสามารถและสถานการณ์ที่เหมาะสมในการประยุกต์ใช้วิธีที่ได้นำเสนอ ผลการทดสอบชี้ให้เห็นว่าวิธีการตรวจจับซิบิลวิธีที่ 2 มีความผิดพลาดและระยะเวลาในการตรวจจับซิบิลน้อยกว่าวิธีที่ 1 เนื่องจากวิธีการแบ่งกลุ่มย่อยด้วยวิธีการหาค่าเหมาะที่สุดของสภาพมอดุลาร์อย่างเร็วไม่สามารถแบ่งกลุ่มย่อยได้ในบางกรณีและการตรวจจับซิบิลวิธีที่ 1 ใช้เวลานานกว่าการตรวจจับซิบิลวิธีที่ 2 อีกด้วย

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุปผลการวิจัย

วิทยานิพนธ์นี้นำเสนอการประเมินผลกระทบจากการโจมตีชนิดซิบิลในระบบการลงคะแนน ซึ่งการประเมินดังกล่าวทำให้ผู้มีภาระ หน้าที่ และความรับผิดชอบในการดูแลระบบการลงคะแนนทราบข้อมูลว่าจะมีซิบิลแฝงตัวอยู่ในระบบหรือไม่ รวมถึงสามารถใช้เป็นส่วนหนึ่งในวิธีตรวจจับซิบิลและใช้ปรับปรุงความถูกต้องของวิธีการตรวจจับซิบิลได้ สูตรการประเมินผลกระทบจากการโจมตีชนิดซิบิลในระบบการลงคะแนนที่นำเสนอถูกคำนวณจากความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง โดยมีจำนวนผู้ใช้งานจริง (n) จำนวนตัวเลือก (k) และจำนวนซิบิล (S) เป็นพารามิเตอร์ในการคำนวณ สูตรที่นำเสนอมีทั้งสูตรแบบแม่นตรงและสูตรแบบประมาณค่า ความแม่นยำของสูตรถูกประเมินโดยการเปรียบเทียบกับเหตุการณ์แบบมอนติคาร์โลจำนวน 1 ล้านครั้ง ผลการทดสอบชี้ให้เห็นว่าสูตรแบบแม่นตรงมีความคลาดเคลื่อนจากการเหตุการณ์จริงไม่เกิน 0.1% ซึ่งต่ำกว่าความเคลื่อนของสูตรใกล้เคียงที่ถูกนำเสนอไว้ในอดีต [27] แต่สูตรแม่นตรงที่นำเสนอมีความซับซ้อนในระดับ $O((n+S)^k)$ ความซับซ้อนดังกล่าวทำให้สูตรแบบแม่นตรงถูกนำไปใช้ในโครงข่ายขนาดใหญ่ได้ยาก วิทยานิพนธ์นี้จึงนำเสนอสูตรแบบประมาณค่าเพิ่มเติมอีก 2 สูตร ได้แก่ สูตรการประมาณค่าด้วยการแจกแจงปัวส์ซอง และสูตรการประมาณค่าด้วยการแจกแจงปกติ ทั้งนี้สูตรการประมาณค่าด้วยการแจกแจงปัวส์ซองสร้างขึ้นบนสมมุติฐานว่ามีผู้ใช้งานจริงเป็นจำนวนมากเลือกตัวเลือกซิบิล ดังนั้นความผิดพลาดของสูตรการประมาณค่าด้วยการแจกแจงปัวส์ซองจึงสูงในกรณีที่จำนวนซิบิลมีค่าน้อย อย่างไรก็ตามความซับซ้อนของสูตรการประมาณค่าด้วยการแจกแจงปัวส์ซองอยู่ในระดับ $O(n)$ ซึ่งน้อยกว่าสูตรแบบแม่นตรงมากทำให้สามารถประยุกต์ใช้ในสถานการณ์จริงได้ สำหรับสูตรการประมาณค่าด้วยการแจกแจงปกติใช้สมมุติฐานที่ว่าคะแนนเสียงของตัวเลือกแต่ละตัวมีความเป็นอิสระต่อกัน ซึ่งสูตรการประมาณค่าด้วยการแจกแจงปกติมีความผิดพลาดสูงกว่าสูตรการประมาณค่าด้วยการแจกแจงปัวส์ซองแต่มีความซับซ้อนในระดับ $O(1)$ ดังนั้นทั้งสูตรแบบแม่นตรงและสูตรแบบประมาณค่าต่างมีข้อดีและข้อเสียที่แตกต่างกันขึ้นอยู่กับว่าต้องการใช้ความแม่นยำหรือความซับซ้อนของสูตรเป็นหลักในการพิจารณาเลือกใช้สูตร

สูตรแม่นตรงถูกใช้ในการอธิบายเส้นเชื่อมต่อระหว่างจำนวนผู้ใช้งานจริง จำนวนตัวเลือก และจำนวนซิบิล ที่มีผลต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง จากผลการทดสอบพบว่าจำนวนผู้ใช้งานจริงส่งผลกระทบต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง กล่าวคือความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจะแปรผกผันกับจำนวนผู้ออกเสียงจริงเนื่องจากจะทำให้คะแนนเสียงเฉลี่ยในตัวเลือกแต่ละตัวเพิ่มขึ้นเป็นผลให้ความแปรปรวนของคะแนนในตัวเลือกแต่ละตัวมีเพิ่มมากขึ้นด้วย ในขณะที่ผลต่างระหว่างตัวเลือกที่ซิบิลเลือกกับตัวเลือกที่ซิบิลไม่ได้เลือกเท่าเดิม ทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงลดลง จำนวนตัวเลือกส่งผลกระทบต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเช่นกัน แบ่งเป็น 2 กรณี คือ กรณีที่มีซิบิลเป็นจำนวนน้อยและมาก สำหรับกรณีที่มีซิบิลจำนวนน้อย การเพิ่มจำนวนตัวเลือกจะลดความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเนื่องจากจะเป็นการเพิ่มโอกาสให้มีอย่างน้อย 1 ตัวเลือกของผู้ใช้งานจริงมีคะแนนเสียงสูงกว่าตัวเลือกของซิบิล แต่สำหรับกรณีที่มีซิบิลจำนวนมาก การเพิ่มจำนวนตัวเลือกกลับทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงมีค่าเพิ่มขึ้นเนื่องจากคะแนนเสียงเฉลี่ยในตัวเลือกของผู้ใช้งานจริงจะลดลงไปตาม

จำนวนตัวเลือก ทั้งนี้ขึ้นกับการกระจายตัวของคะแนนเสียงจากผู้ใช้งานจริงในตัวเลือกแต่ละตัวเลือก การเพิ่มขึ้นของจำนวนซิบิลและการเพิ่มขึ้นของจำนวนผู้ใช้งานจริงที่เลือกตัวเลือกของซิบิลจะทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพิ่มขึ้น นอกจากนั้นผลกระทบจากอัตราส่วนของซิบิลต่อจำนวนผู้ใช้งานจริง (อัตราส่วนซิบิล) ที่มีต่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงถูกนำเสนอในวิทยานิพนธ์นี้ด้วย โดยการเพิ่มขึ้นของซิบิล 1 ตัวมีผลกระทบลดลงเมื่อความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงเพิ่มขึ้น นอกจากนั้นพบว่าหากมีอัตราส่วนซิบิลต่อผู้ใช้งานจริงมีค่าเป็น 5% และ 25% จะทำให้ความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงมีค่าเป็น 50% และ 99% ตามลำดับในกรณีที่มีผู้ใช้งานจริงทั้งหมด 100 คนและมีตัวเลือก 3 ตัว ซึ่งหมายถึงการมีซิบิลปะปนอยู่ในระบบจำนวนน้อยสามารถส่งผลกระทบต่อระบบอย่างมากไ ด้

นอกจากสูตรการประเมินผลกระทบจากการโจมตีชนิดซิบิลแล้ว วิทยานิพนธ์นี้เสนอวิธีการประยุกต์ใช้สูตรการประเมินผลกระทบดังกล่าวในการตรวจจับซิบิลในระบบ แบ่งออกเป็น 2 วิธี ได้แก่ วิธีการแบ่งกลุ่มย่อยแล้วตรวจจับซิบิลเป็นกลุ่ม (วิธีที่ 1) กับวิธีการใช้ความเหมือนของคู่โนดในโครงข่ายและความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในการลดความผิดพลาดการตรวจจับซิบิล (วิธีที่ 2) โครงข่ายที่ใช้ในการทดสอบแบ่งออกเป็น 3 รูปแบบ คือ โครงข่ายที่มีเส้นเชื่อมต่อทางเดียว และโครงข่ายที่มีเส้นเชื่อมต่อสองทางโดยโนดแต่ละโนดในโครงข่ายมีความนิยมทั้งเท่ากันและไม่เท่ากันที่ผู้ใช้งานจริงอื่นจะเข้ามามีเส้นเชื่อมต่อด้วย ทั้งสองวิธีถูกทดสอบความแม่นยำในการตรวจจับซิบิลและความซับซ้อนเชิงเวลา โดยความแม่นยำในการตรวจจับซิบิลถูกคำนวณจากความผิดพลาดเชิงบวกและเชิงลบของผลการตรวจจับซิบิล และความซับซ้อนเชิงเวลาชี้วัดจากระยะเวลาในการตรวจจับซิบิลเป็นจำนวน 3,000 ครั้งเพื่อใช้เปรียบเทียบระหว่างวิธีทั้ง 2 วิธี จากผลการทดสอบพบว่าวิธีที่ 1 มีความแม่นยำสูงกว่าวิธีที่ 2 ในกรณีที่ผู้ใช้งานในระบบมีจำนวนน้อย แต่วิธีที่ 1 มีความแม่นยำต่ำกว่าวิธีที่ 2 มากในกรณีที่ผู้ใช้งานในระบบมีจำนวนมาก เนื่องจากวิธีที่ 1 มีขั้นตอนการแบ่งโครงข่ายออกเป็นกลุ่มย่อย และวิธีในการแบ่งกลุ่มย่อยที่วิทยานิพนธ์นี้เลือกใช้เป็นของ [33] ซึ่งมีความสามารถในการแบ่งโครงข่ายออกเป็นกลุ่มย่อยในกรณีที่โครงข่ายมีวงจร (cycle) เป็นจำนวนน้อยเท่านั้น นอกจากนั้นวิธีที่ 1 ใช้เวลาไม่ต่ำกว่า 7 เท่าของวิธีที่ 2 เนื่องจากขั้นตอนการแบ่งกลุ่มย่อยเช่นกัน ทั้งนี้การเลือกใช้วิธีการตรวจจับซิบิลที่นำเสนอในวิทยานิพนธ์นี้จึงขึ้นอยู่กับความเหมาะสมกับโครงข่ายและกลยุทธ์หลักเบื้องต้นในการป้องกันซิบิลคือ การเพิ่มจำนวนเส้นเชื่อมต่อจากผู้ใช้งานจริงถึงผู้ใช้งานจริงที่รู้จักกันและลดเส้นเชื่อมต่อจากผู้ใช้งานจริงถึงผู้ใช้งานที่ไม่รู้จักตัวตนที่แท้จริงจะสามารถช่วยลดความรุนแรงของการโจมตีชนิดซิบิลได้อย่างเป็นรูปธรรมได้

5.2 ข้อเสนอแนะ

งานวิจัยนี้สามารถต่อยอดวิจัยได้อีกหลายทางในอนาคต ในมุมมองของการเพิ่มสมรรถนะให้กับสูตรและวิธีการตรวจจับซิบิลที่ได้นำเสนอในวิทยานิพนธ์นี้ การประยุกต์ทฤษฎีอื่นที่อาจจะเกี่ยวข้องให้การช่วยตรวจจับซิบิล รวมถึงสามารถประยุกต์ใช้สูตรที่นำเสนอในระบบที่ใช้คะแนนชื่อเสียงได้อีกด้วย โดยมีตัวอย่างดังนี้

5.2.1 การเพิ่มสมรรถนะให้กับสูตรการหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง

สูตรการหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงได้รับการยืนยันแล้วว่ามีความแม่นยำสูงดังผลการทดสอบในบทที่ 3 อย่างไรก็ตามสูตรดังกล่าวสามารถต่อยอดและพัฒนาต่อได้ดังนี้

1. **การลดความซับซ้อนของสูตรแม่นตรง**
   เนื่องจากสูตรแม่นตรงมีความซับซ้อนในระดับ $O((n+S)^k)$ ทำให้สูตรแม่นตรงไม่สามารถคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงในกรณีที่มีผู้ใช้งานจริงและซิบิลเกิน 170 คนบนซอฟท์แวร์รุ่นที่ใช้ทำวิจัยได้ (MATLAB version R2009a 64-bit) การลดความซับซ้อนของสูตรแบบแม่นตรงจะเป็นการเพิ่มขอบเขตความสามารถของสูตรให้สามารถใช้จริงได้มากขึ้น

2. **การลดความซับซ้อนเชิงเวลาของวิธีการตรวจจับซิบิล**
   ความล่าช้าในการตรวจจับซิบิลอาจทำให้ไม่สามารถปกป้องผู้ใช้งานจริงได้ทันกาลและทำให้เกิดความเสียหายได้ การลดระยะเวลาในการตรวจจับซิบิลจึงเป็นประเด็นสำคัญที่สามารถต่อยอดจากวิธีที่นำเสนอในวิทยานิพนธ์นี้ได้

3. **การเพิ่มความแม่นยำให้กับวิธีการตรวจจับซิบิล**
   ความผิดพลาดของวิธีการตรวจจับซิบิลที่นำเสนอสามารถลดลงได้ ดังนั้นจึงควรพัฒนาวิธีการตรวจจับซิบิลให้มีความผิดพลาดลดลงเพื่อให้ซิบิลไม่สามารถโจมตีผู้ใช้งานจริงที่เป็นเหยื่อได้และไม่ทำให้ผู้ใช้งานจริงถูกตรวจจับผิดว่าเป็นซิบิล อันจะทำให้เกิดความเสียหายได้เช่นกัน

5.2.2 การเพิ่มความสมรรถนะให้กับวิธีการตรวจจับซิบิลในสถานการณ์ที่แตกต่างกัน

วิทยานิพนธ์นี้นำเสนอการตรวจจับซิบิลบนโครงข่ายที่มีทอพอโลยีแน่นอน และมีซิบิลในโครงข่ายเพียงกลุ่มเดียวเท่านั้น รวมถึงนับผลการลงคะแนนเพียงครั้งเดียวเมื่อสิ้นสุดการออกเสียง และคะแนนเสียงของผู้ใช้งานแต่ละคนมีความสำคัญเท่ากันเท่านั้น ดังนั้นงานวิจัยที่นำเสนอจึงสามารถต่อยอดในกรณีอื่นได้ ดังนี้

1. **ระบบที่มีทอพอโลยีไม่แน่นอน**
   ในกรณีที่ระบบมีทอพอโลยีไม่แน่นอน เช่น มีผู้ใช้งานใหม่สมัครเข้าใช้งานระบบตลอดเวลาหรือมีการเปลี่ยนตัวตน การลบตนเองออกจากระบบ การตรวจจับซิบิลจะทำได้ยากขึ้น

2. **กรณีมีซิบิลหลายกลุ่ม**
   เมื่อมีผู้ไม่หวังดีหลายกลุ่ม จะยิ่งทำให้การตรวจจับซิบิลยากขึ้นไม่ว่ากลุ่มของซิบิลแต่ละกลุ่มจะร่วมมือกันหรือไม่ก็ตาม

3. **การพิจารณาการลงคะแนนเสียงอย่างเป็นลำดับ**
   เมื่อคะแนนเสียงถูกเปิดเผยก่อนที่การออกเสียงจะเสร็จสิ้น คะแนนเสียงที่ถูกเลือกก่อนอาจส่งผลกระทบต่อการตัดสินใจของผู้ที่ไม่ออกเสียงได้ การหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงหรือการหาวิธีการตรวจจับซิบิลจึงมีความแตกต่างออกไปจากสูตรและวิธีที่นำเสนอในวิทยานิพนธ์นี้

4. **การออกเสียงในกรณีที่ผู้ออกเสียงแต่ละคนมีความสำคัญไม่เท่ากัน**
   เมื่อคะแนนเสียงแต่ละคะแนนมีน้ำหนักไม่เท่ากัน การคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงจะไม่สามารถคำนวณจากสัมประสิทธ์ของฟังกชันอเนกนามอย่างมีเงื่อนไขตามที่วิทยานิพนธ์นี้นำเสนอได้

5.2.3 การประยุกต์ใช้สูตรที่นำเสนอบนระบบการลงคะแนนต่าง ๆ

นอกจากโครงข่ายสังคมออนไลน์ที่พบเห็นการโจมตีชนิดซิบิลได้ชัดเจนแล้ว โครงข่ายชนิดอื่นอาจถูกโจมตีโดยซิบิลได้ด้วยวิธีที่แตกต่างกัน ดังนั้นงานวิจัยนี้จึงสามารถต่อยอดในการตรวจจับซิบิลบนระบบอื่นที่มีรายละเอียดของแต่ละระเบบแตกต่างกัน ดังตัวอย่างเช่น

1. **ตลาดหุ้น**
   ตลาดหุ้นเป็นศูนย์รวมของนักลงทุนที่หวังผลประโยชน์จากการเป็นหุ้นส่วนของบริษัทที่ทำกำไรดีน่าเชื่อถือ โดยราคาของหุ้นขึ้นอยู่กับความต้องการและความไม่ต้องการในหุ้น โดยหากมีผู้ซื้อหุ้นมากราคาของหุ้นจะปรับตัวให้สูงขึ้นในขณะที่ราคาหุ้นจะลดลงเมื่อมีผู้ขายหุ้นเป็นจำนวนมาก ทำให้มีผู้ไม่หวังดีที่มีเงินมากเพียงพอต้องการควรคุมราคาของหุ้น โดยทำการซื้อขายหุ้นซึ่งไม่สะท้อนถึงสถานะทางธุรกิจของบริษัทที่เป็นเจ้าของหุ้น สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ (กลต.) เป็นหน่วยงานกำกับดูแลให้การซื้อขายหุ้นเป็นไปอย่างปกติและไม่ทุจริตพยายามตรวจจับผู้ไม่หวังดีดังกล่าว โดยสังเกตจากพฤติกรรมการซื้อขายหุ้นของนักลงทุนแต่ละคน อย่างไรก็ตามการสมรู้ร่วมคิดของผู้ไม่หวังดีหลายคนอาจตรวจจับได้ยาก ดังนั้นวิธีการตรวจจับซิบิลที่นำเสนอในวิทยานิพนธ์นี้จึงประยุกต์ใช้กับพฤติกรรมการสมรู้ร่วมคิดซื้อขายหุ้นในตลาดหลักทรัพย์ได้ในอนาคต

2. **การลงคะแนนเสียงจากประชาชนบนเวทีประกวด**
   ในการประกวดต่าง ๆ ที่ใช้คะแนนเสียงจากประชาชน เช่น เวทีประกวดร้องเพลง อาจถูกโจมตีโดยซิบิลได้ โดยผู้ไม่หวังดีสร้างตัวตนปลอมขึ้นมาเพื่อลงคะแนนเสียงโดยอาจมีการปลอมหมายเลขโทรศัพท์หรือบัญชีปลอมของโปรแกรมประยุกต์ (application) ที่ใช้ในการลงคะแนนเสียงได้ โดยตัวตนปลอมเหล่านั้นไม่มีการติดต่อสื่อสารกันเลย ทำให้ไม่สามารถตรวจจับซิบิลบนสมมุติฐานว่าซิบิลอยู่รวมกันเป็นกลุ่มได้ ดังนั้นการตรวจจับซิบิลในระบบการลงคะแนนเสียงจากประชาชนจึงเป็นหัวข้อที่น่าวิจัยต่อไป

3. **ร้านค้าออนไลน์**
   คะแนนชื่อเสียงจากลูกค้าเป็นการโฆษณาที่ดีที่สุดที่ร้านค้าออนไลน์ใช้อ้างอิงให้ลูกค้าเชื่อว่าสินค้าและบริการที่ได้รับจากร้านค้านั้น ๆ มีคุณภาพดี อย่างไรก็ตามเนื่องจากร้านค้าเป็นผู้ควบคุมระบบคะแนนชื่อเสียง ทำให้ร้านค้าบางร้านทำการทุจริต สร้างชื่อเสียงปลอมเพื่อแนะนำลูกค้าคนอื่นว่าสินค้าและบริการดี ทั้งนี้ในกรณีที่ผู้ดูแลระบบเป็นผู้ไม่หวังดีเองจะทำให้ยากแก่การตรวจสอบ และเป็นปัญหาที่แก้ไขไม่ได้ในปัจจุบัน

4. **โครงข่ายตัวตรวจวัด**
   โครงข่ายตัวตรวจวัดถูกใช้เพื่อเก็บข้อมูลจากแหล่งข้อมูลต่าง ๆ ในกรณีที่ผู้ไม่หวังดีสามารถโปรแกรมให้ตัวตรวจวัดบางส่วนของโครงข่ายมีพฤติกรรมการส่งข้อมูลเท็จถึงศูนย์กลางจะทำให้เกิดความพิดพลาดในการประมวลผลของศูนย์กลางได้ การปรับปรุงวิธีการตรวจจับซิบิลให้เหมาะสมกับโครงข่ายตัวตรวจวัดที่มีทอพอโลยี วิธีการเก็บข้อมูล และวิธีการประมวลผล จึงเป็นสิ่งท้าทายและสมควรวิจัยเพิ่มเติมเป็นอย่างยิ่ง

5.2.4 การประยุกต์ทฤษฎีที่เกี่ยวข้องในการตรวจจับซิบิล

นอกจากจะใช้ทฤษฎีทางพีชคณิตในการคำนวณหาความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงแล้ว มีทฤษฎีอื่นที่อาจจะช่วยให้การคำนวณความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงสามารถนำไปใช้จริงได้มากขึ้นอีกหลายทฤษฎี ดังตัวอย่างเช่น

1. **ทฤษฎีเกม (Game theory)**
   เมื่อกำหนดให้ผู้ไม่หวังดีเป็นฝ่ายโจมตี และผู้ใช้งานจริงเป็นฝ่ายป้องกัน ฝ่ายโจมตีและฝ่ายป้องกันสามารถสร้างกลยุทธ์ของตนเองได้ โดยฝ่ายโจมตีสามารถเลือกจำนวนซิบิลที่จะสร้างเพื่อโจมตีเหยื่อ ในขณะที่ฝ่ายป้องกันสามารถเพิ่มจำนวนเส้นเชื่อมต่อระหว่างกันได้ และผลของกลยุทธ์ในตารางผลกลยุทธ์สามารถหาได้จากสูตรความน่าจะเป็นที่ซิบิลจะชนะการออกเสียงที่ได้นำเสนอในบทที่ 3 ทั้งนี้เมื่อใช้ทฤษฎีเกมประกอบกับสูตรที่ได้นำเสนอในวิทยานิพนธ์นี้แล้ว ผลที่ได้คือจำนวนซิบิลที่เหมาะสมสำหรับโน้มน้าวเหยื่อ และจำนวนเส้นเชื่อมต่อที่ผู้ใช้งานจริงจะสร้างระหว่างกันได้

2. **การเรียนรู้ของเครื่อง (Machine learning)**
   การทำให้เครื่องจักรเรียนรู้สิ่งต่าง ๆ ได้ต้องมีข้อมูลให้เครื่องจักรเรียนรู้มากเพียงพอ ในกรณีที่ต้องการให้เครื่องจักรเรียนรู้ได้เร็วเป็นพิเศษ (ในทางดี) หรือต้องการให้เครื่องจักรเรียนรู้สิ่งที่ผิด (ในทางไม่ดี) ต้องทำให้เครื่องจักรได้พบกับสถานการณ์ซ้ำเดิมให้มาก ทั้งนี้การคำนวณปริมาณเหตุการณ์ซ้ำเดิมที่ต้องใช้อาจถูกประยุกต์จากความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง โดยกำหนดให้จำนวนเหตุการณ์ซ้ำเท่ากับจำนวนซิบิลและจำนวนเหตุการณ์ตามปกติเป็นจำนวนผู้ใช้งานจริงได้

3. **เหมืองข้อมูล (Data mining)**
   เพื่อให้สามารถเก็บข้อมูลและบริหารจัดการได้ง่ายในระบบเหมืองข้อมูล การวิเคราะห์ปริมาณข้อมูลซ้ำอาจเพิ่มประสิทธิ์ภาพของการบริหารจัดการเหมืองข้อมูลได้ โดยประยุกต์ใช้สูตรความน่าจะเป็นที่ซิบิลจะชนะการออกเสียง เพื่อคำนวณปริมาณข้อมูลซ้ำในระบบ และคัดกรองข้อมูลเพื่อทิ้งและเก็บข้อมูลอย่างมีประสิทธิภาพได้

4. **ทฤษฎีหลักฐาน (Theory of evidence)**
   ในกรณีที่ผู้ใช้งานแต่ละคนในระบบสามารถออกเสียงได้มากกว่า 1 คะแนนเสียงในครั้งเดียว ทฤษฎีหลักฐานสามารถประยุกต์ใช้กับการตรวจจับซิบิลได้ ตัวอย่างเช่น พยานคนที่ 1 บอกว่าผู้ใช้งาน $\{a, b, c\}$ อาจจะเป็นซิบิล ในขณะที่พยานคนที่ 2 บอกว่าผู้ใช้งาน $\{a, c, d\}$ อาจจะเป็นซิบิล จะได้ว่า ผู้ใช้งาน $\{a, c\}$ มีความน่าจะเป็นสูงที่จะเป็นซิบิลมากกว่าผู้ใช้งาน $\{b, d\}$ เป็นต้น ซึ่งจะสามารถทำให้วิธีการตรวจจับซิบิลมีขอบเขตความสามารถและประยุกต์ใช้จริงได้มากขึ้น

5. **ทฤษฎีจำนวน (Number theory)**
   นอกจากมุมมองของการประยุกต์ใช้ในเชิงวิศวกรรมแล้ว สูตรที่นำเสนอสามารถประยุกต์ใช้ในเชิงคณิตศาสตร์ได้อีกด้วย โดยสูตรแม่นตรงที่นำเสนอในบทที่ 3 สามารถใช้คำนวณสัมประสิทธ์ของฟังก์ชันอเนกนามที่มีเงื่อนไขเฉพาะเจาะจงได้

ทั้งนี้ผู้วิจัยหวังเป็นอย่างยิ่งว่างานวิจัยนี้จะเป็นประโยชน์กับผู้ใช้งานและผู้ดูแลระบบการลงคะแนนต่าง ๆ ให้สามารถป้องกันและประเมินผลกระทบจากการโจมตีชนิดซิบิลรวมถึงลดความสูญเสียที่อาจเกิดขึ้นโดยการโจมตีชนิดซิบิลได้

รายการอ้างอิง

[1] Douceur, J. R. The Sybil Attack. International Workshop on Peer-to-peer Systems. Springer Berlin Heidelberg, 2002: 251-260.

[2] Hoffman, K., Zage, D., and Nita-Rotaru, C. A Survey of Attack and Defense Techniques for Reputation Systems. ACM Computing Surveys (CSUR), 42(1), 1, 2009: 1-34.

[3] Lai, K., Feldman, M., Stoica, I., Chuang, J. Incentives for Cooperation in Peer-to-peer Networks. Workshop on Economics of Peer-to-peer Systems, June 2003: 1243-1248.

[4] Srivatsa, M., Xiong, L., Liu, L. TrustGuard: Countering Vulnerabilities in Reputation Management for Decentralized Overlay Networks. Proceedings of the 14th International Conference on World Wide Web. ACM, May 2005: 422-431.

[5] Chaudhary, M. S., Thanvi, M. P. Performance Analysis of Modified AODV Protocol in Context of Denial of Service (Dos) Attack in Wireless Sensor Networks. International Journal of Engineering Research and General Science, 4, 2015: 486-491.

[6] Zhou, T., Choudhury, R. R., Ning, P., Chakrabarty, K. P2DAP—Sybil Attacks Detection in Vehicular Ad Hoc Networks. IEEE Journal on Selected Areas in Communications, 29(3), 2011: 582-594.

[7] Abbas, S., Merabti, M., Llewellyn-Jones, D., Kifayat, K. Lightweight Sybil Attack Detection in MANETs. IEEE Systems Journal, 7(2), 2013: 236-248.

[8] Liu, Y., Bild, D. R., Dick, R. P., Mao, Z. M., Wallach, D. S. The Mason Test: A Defense Against Sybil Attacks in Wireless Networks Without Trusted Authorities. IEEE Transactions on Mobile Computing, 14(11), 2015: 2376-2391.

[9] Dong, W., Liu, X. Robust and Secure Time-synchronization Against Sybil Attacks for Sensor Networks. IEEE Transactions on Industrial Informatics, 11(6), 2015: 1482-1491.

[10] Alsaadi, E., Tubaishat, A. Internet of Things: Features, Challenges, and Vulnerabilities. International Journal of Advanced Computer Science and Information Technology, 4(1), 2015: 1-13.

[11] Wang, G., Musau, F., Guo, S., Abdullahi, M. B. Neighbor Similarity Trust Against Sybil Attack in P2P E-commerce. IEEE Transactions on Parallel and Distributed Systems, 26(3), 2015: 824-833.

[12] Yu, H., Kaminsky, M., Gibbons, P. B., Flaxman, A. D. Sybilguard: Defending Against Sybil Attacks Via Social Networks. IEEE/ACM Transactions on Networking, 16(3), 2008: 576-589.

[13] Gong, N. Z., Frank, M., Mittal, P. Sybilbelief: A Semi-supervised Learning Approach for Structure-based Sybil Detection. IEEE Transactions on Information Forensics and Security, 9(6), 2014: 976-987.

[14] Levine, B. N., Shields, C., and Margolin, N. B. A Survey of Solutions to the Sybil attack. Technical Report, University of Massachusetts Amherst, Amherst, MA(7), 2006.

[15] Zhao, B. Y., Huang, L., Stribling, J., Rhea, S. C., Joseph, A. D., Kubiatowicz, J. D. Tapestry: A Resilient Global-scale Overlay for Service Deployment. IEEE Journal on Selected Areas in Communications, 22(1), 2004: 41-53.

[16] Newsome, J., Shi, E., Song, D., Perrig, A. The Sybil Attack in Sensor Networks: Analysis & Defenses. Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks ACM, April 2004: 259-268.

[17] Maniatis, P., Rosenthal, D. S., Roussopoulos, M., Baker, M., Giuli, T. J., Muliadi, Y. Preserving Peer Replicas by Rate-limited Sampled Voting. ACM SIGOPS Operating Systems Review, 37(5), October 2003: 44-59.

[18] Guette, G., Ducourthial, B. On the Sybil Attack Detection in Vanet. IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS), October 2007: 1-6.

[19] Silva, R. F. E., Silva, E. D., Albini, L. C. P. A Sybil Safe Virtualization-based Public Key Management Scheme for Mobile Ad Hoc Networks. Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), 4(1), 2014.

[20] Chiang, J. T., Hu, Y. C., Yadav, P. Secure Cooperative Spectrum Sensing Based on Sybil-resilient Clustering. IEEE Global Communications Conference (GLOBECOM), December 2013: 1075-1081.

[21] Anagha, P. B., Krishnan, J. Vote Credence: Social Network Sybil Defence by User Behaviour. International Journal of Science and Research (IJSR), 2016: 1500-1503.

[22] Hangxia, Z. Mitigating Peer-to-Peer Botnets by Sybil Attacks. International Conference on Innovative Computing & Communication and Asia-Pacific Conference on Information Technology & Ocean Engineering (CICC-ITOE), January 2010: 241-243.

[23] Yu, H., Shi, C., Kaminsky, M., Gibbons, P. B., Xiao, F. Dsybil: Optimal Sybil-resistance for Recommendation systems. IEEE Symposium on Security and Privacy 30th, May 2009: 283-298.

[24] Shin, K., Joe-Wong, C., Ha, S., Yi, Y., Rhee, I., Reeves, D. S. T-chain: A General Incentive Scheme for Cooperative Computing. IEEE 35th International Conference on Distributed Computing Systems (ICDCS), June 2015: 163-174.

[25] Khan, S. M., Khan, N. M. Low Complexity Signed Response Based Sybil Attack Detection Mechanism in Wireless Sensor Networks. Journal of Sensors, 2016.

[26] Uruena, M., Cuevas, R., Cuevas, A., Banchs, A. A Model to Quantify the Success of a Sybil Attack Targeting RELOAD/Chord Resources. IEEE Communications Letters, 17(2), 2013: 428-431.

[27] DasGupta, A. Exact Tail Probabilities and Percentiles of the Multinomial Maximum. Technical Report, Purdue University, 2009.

[28] Ross, S. M. Introduction to Probability Theory. Academic press, 1981.

[29] Martelot, E. L., Hankin, C. Multi-scale Community Detection Using Stability as Optimisation Criterion in a Greedy Algorithm. arXiv preprint arXiv:1201.3307, KDIR, 2011.

[30] Silawan, T., Aswakul, C. SybilVote: Formulas to Quantify the Success Probability of Sybil Attack in Online Social Network Voting. IEEE Communications Letters, 99, 2017.

[31] Li, X., Lu, R., Liang, X., Shen, J., Lin, X. C. Smart Community: An Internet of Things Application. IEEE Communications Magazine, 49(11), 2011: 68-75.

[32] Ortiz, A. M., Hussein, D., Park, S., Han, S. N., Crespi, N. The Cluster Between Internet of Things and Social Networks: Review and Research Challenges. IEEE Internet of Things Journal, 2014: 206-215.

[33] Yu, H., Gibbons, P. B., Kaminsky, M., Xiao, F. Sybillimit: A Near-optimal Social Network Defense Against Sybil Attacks. IEEE Symposium on Security and Privacy (SP), 2008: 3-17.

[34] Danezis, G., Mittal, P. SybilInfer: Detecting Sybil Nodes Using Social Networks. Presented at the NDSS, 2009.

[35] Silawan, T., Aswakul, C. SybilComm: Sybil Community Detection Using Persuading Function in IoT System. IEEE International Conference on Electronics, Information, and Communications (ICEIC), January 2016: 1-4.

ประวัติผู้เขียนวิทยานิพนธ์

ธีรพล ศิลาวรรณ์ เกิดเมื่อวันที่ 23 เมษายน พ.ศ. 2529 กรุงเทพมหานครฯ เป็นบุตรของนายพิสิทธิ์ ศิลาวรรณ์ และ นางประดับศรี ศิลาวรรณ์ สำเร็จการศึกษาชั้นมัธยมศึกษาจากโรงเรียน-สตรีวิทยา 2 ในปีการศึกษา 2546 จากนั้นได้เข้าศึกษาต่อระดับปริญญา ณ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จนสำเร็จหลักสูตรวิศวกรรมศาสตรบัณฑิตในปีการศึกษา 2550 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตร์มหาบัณฑิตในปีการศึกษาถัดมา ณ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย สังกัดห้องปฏิบัติการวิจัยโทรคมนาคม ได้รับทุนศิษย์ก้นกุฏิขณะศึกษาระดับมหาบัณฑิต และทำหน้าที่ประธานกลุ่มความร่วมมือนิสิตบัณฑิตศึกษาวิศวกรรมไฟฟ้าแห่งจุฬาลงกรณ์มหาวิทยาลัย EEPSA-CU (Electrical Engineering Postgraduate Student Assembly of Chulalongkorn University) จนถึง พ.ศ. 2553 จากนั้นจึงเข้าทำงานเป็นที่ปรึกษาโครงการด้านเทคโนโลยีสารสนเทศให้กับบริษัท C&C International Venture Co., Ltd. เป็นเวลา 2 ปีและได้เข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรดุษฎีบัณฑิต ณ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย สังกัดห้องปฏิบัติการวิจัยโทรคมนาคม ได้รับการสนับสนุนทุนการศึกษาจากทุนการศึกษาหลักสูตรดุษฎีบัณฑิต "100 ปี จุฬาลงกรณ์มหาวิทยาลัย" (The 100th Anniversary Chulalongkorn University Fund for Doctoral Scholarship) ตั้งแต่ปีการศึกษา 2556 จนถึงปีการศึกษา 2559

บทความทางวิชาการจากวิทยานิพนธ์

1. Silawan, T., Aswakul, C. SybilComm: Sybil Community Detection Using Persuading Function in IoT System. IEEE International Conference on Electronics, Information, and Communications (ICEIC), January 2016: 1-4.
2. Silawan, T., Aswakul, C. SybilVote: Formulas to Quantify the Success Probability of Sybil Attack in Online Social Network Voting. IEEE Communications Letters, 21(7), July 2017: 1553-1556.