19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

การระบุผู้ต้องสงสัยจากรูปแบบในประวัติการโทร
Identifying Suspects From Call Log Pattern

นายยศวีร์ ล่องทอง[1], ดร.รุจ เอกะวิภาต[2] , และ ดร.ลลิตา นฤปิยะกุล[1*]
[1]ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล, นครปฐม
[2]ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพมหานคร
E-mail: may.yossawee@gmail.com[1], ruj.a@ku.th [2], lalita.nar@mahidol.ac.th [3*]

บทคัดย่อ

ปัจจุบันโทรศัพท์มือถือได้กลายเป็นส่วนหนึ่งของการดำเนินชีวิตของทุกคน เป็นอุปกรณ์ที่ใช้ในการติดต่อสื่อสารกันระหว่าง ครอบครัว, เพื่อน, เพื่อนร่วมงาน และคนอื่นๆ อาชญากรได้นำเอาเทคโนโลยีในการติดต่อสื่อสารเข้ามามีส่วนเกี่ยวข้องในการก่อคดี ทำให้การติดตามการใช้โทรศัพท์ของผู้ต้องสงสัยและเครือข่ายมีความสำคัญมาก การเปลี่ยนหมายเลขโทรศัพท์สามารถทำได้ง่าย และไม่มีการระบุเจ้าของ งานวิจัยนี้จึงได้เสนอวิธีการที่ใช้ในการติดตามหมายเลขโทรศัพท์ที่ผู้ต้องสงสัยเปลี่ยนไปเพื่อหลบหนีจากการสืบสวนของเจ้าหน้าที่ โดยการเปรียบเทียบพฤติกรรมการโทรที่ได้จากประวัติการโทรที่เป็นเอกลักษณ์เฉพาะบุคคล ซึ่งพิจารณาจากกลุ่มหมายเลขที่ผู้ต้องสงสัยเคยติดต่อ และความถี่ในการโทร ผลการวิจัยพบว่าวิธีการที่นำมาใช้เปรียบเทียบข้อมูลทั้งสองชุดนั้นให้ผลในการตรวจจับได้มากกว่า 90% ของหมายเลขที่ใช้ในการทดสอบ

คำสำคัญ: วิเคราะห์บันทึกการโทร, การติดตามการใช้โทรศัพท์, รูปแบบการโทร

Abstract

Smart phone is one of technologies that used widely today. Many people often use smart phones to contact their families, friends, colleagues, etc. Criminals use mobile technology to plan and communicate among their groups. Investigators can analyze and obtain suspect's information from call logs, derived from smart phones and monthly phone bills. In many cases, the suspects changed their phone numbers to avoid tracking. The problem is that the investigators have difficulty to track these suspects from their call logs. In this paper, we introduce the process of tracking telephone number based on call logs. Our hypothesis is that each user has a unique calling behavior pattern. Calling pattern is importance for tracking suspect's telephone number. To compare the calling patterns, we consider common contact groups and calling frequencies. The experimental results on real call logs demonstrate that our method can track telephone number more than 90% of correct matching.

Keywords: Call log analysis, Phone tracking, Calling pattern

1. คำนำ

เทคโนโลยีการติดต่อสื่อสารในยุคปัจจุบันมีบทบาทสำคัญอย่างยิ่งในการดำรงชีวิตของมนุษย์ แล้วได้นำความรู้ทางเทคโนโลยีด้านต่างๆ มาพัฒนาให้การติดต่อสื่อสารกันมีความรวดเร็วและสะดวกมากขึ้น โทรศัพท์มือถือเป็นเทคโนโลยีที่นิยมใช้ในการติดต่อสื่อสารกัน ซึ่งส่วนใหญ่จะใช้เป็น Smart Phone เนื่องจากโทรศัพท์มือถือโดยทั่วไปทำได้เพียงโทรออกและรับสายเท่านั้น แต่ Smart Phone มีความสามารถพิเศษเพิ่มเติมในการรับส่งอีเมล์ ถ่ายภาพ ฟังเพลง เป็นต้น ซึ่งสามารถตอบสนองผู้ใช้ได้อย่างดี ประกอบกับรูปลักษณ์ที่ทันสมัย สะดวกต่อการพกพา

โทรศัพท์มือถือเป็นอุปกรณ์สื่อสารที่ใช้ในการติดต่อระยะไกล อาชญากรได้นำเทคโนโลยีทางการสื่อสารด้วยโทรศัพท์มือถือมาเป็นสื่อกลางในการวางแผน นัดหมาย เพื่อการก่อการร้ายและอาชญากรรมในหลากหลายรูปแบบ นอกจากนี้ส่งผลให้การติดตามผู้กระทำผิดทำได้ยาก เพราะการเปลี่ยนหมายเลขโทรศัพท์ทำได้ง่าย บางครั้งเจ้าหน้าที่สามารถสืบสวนจนได้ข้อมูลของผู้ต้องสงสัย และหนึ่งในข้อมูลที่ได้มาคือหมายเลขโทรศัพท์มือถือ แล้วมีหลายกรณีที่ผู้ต้องสงสัยรู้ตัว จึงเปลี่ยนหมายเลขโทรศัพท์มือถือเพื่อหลีกเลี่ยงการถูกติดตามตัว ทำให้เกิดอุปสรรคในการสืบค้นข้อมูลของผู้ต้องสงสัยจากหมายเลขโทรศัพท์ แต่ส่วนใหญ่ผู้ต้องสงสัยเหล่านั้นยังคงติดต่อคนกลุ่มเดิม หรือมีพฤติกรรมการใช้โทรศัพท์คล้ายเดิม ซึ่งหากวิเคราะห์หาความสัมพันธ์ แล้วสามารถทำการเชื่อมโยงความสัมพันธ์ต่างๆ อาจทำให้สืบทราบถึงหมายเลขโทรศัพท์ใหม่ที่ผู้ต้องสงสัยเปลี่ยนไปใช้อีกด้วย

2. ทบทวนวรรณกรรมที่เกี่ยวข้อง

การวิเคราะห์ข้อมูลที่เกี่ยวข้องกับหมายเลขโทรศัพท์ของผู้ต้องสงสัยนั้น เป็นวิธีการหนึ่งที่เจ้าหน้าที่ใช้ในการสืบสวน และติดตามหาตัวผู้

ห้องสมุดวิทยาศาสตร์

ต้องสงสัย โดยเฉพาะอย่างยิ่งเมื่อโทรศัพท์มือถือของผู้ต้องสงสัยถูกทำลาย ปิด หรือดัดแปลง จนไม่สามารถติดตามตัวผู้ต้องสงสัยจากหมายเลข IMEI (International Mobile Equipment) ได้ แม้ว่ามีหลายประเทศยอมรับการใช้หมายเลข IMEI มาใช้ในการสืบสวนติดตามตัวผู้กระทำผิด ด้วยเหตุผลที่ว่าหมายเลข IMEI ในโทรศัพท์มือถือแต่ละเครื่องที่ถูกผลิตขึ้นมาแต่ละเครื่องบนโลกไม่ซ้ำกัน ปัจจุบันนี้อาชญากรสามารถปลอมแปลงหมายเลข IMEI ได้

ในประเทศไทย ผู้ให้บริการเครือข่ายโทรศัพท์มือถือมีการให้บริการหมายเลขโทรศัพท์ 2 ประเภทหลัก คือ ระบบจดทะเบียนรายเดือน และระบบเติมเงิน ซึ่งลูกค้าที่ใช้งานระบบรายเดือนนั้น จะต้องทำการลงทะเบียน และให้ข้อมูลส่วนบุคคลไว้กับผู้ให้บริการ แต่ลูกค้าระบบเติมเงิน ได้มีการลงทะเบียนผ่านระบบอัตโนมัติ ซึ่งอาจจะไม่ใช่ข้อมูลของเจ้าของหมายเลขโทรศัพท์ที่แท้จริง ดังนั้นการติดตาม หรือระบุตัวตนของเจ้าของหมายเลขทำได้ยากกว่าหมายเลขในระบบรายเดือน

ด้วยเหตุนี้จึงจำเป็นที่จะต้องอาศัยหลักการวิเคราะห์ข้อมูลที่ได้จากหมายเลขของผู้ต้องสงสัย และข้อมูลที่เกี่ยวข้องกับหมายเลขโทรศัพท์นั้นๆ เจ้าหน้าที่สืบสวนสามารถวิเคราะห์หาความสัมพันธ์เบื้องต้นโดยอาศัยประสบการณ์ทำงาน แล้วสังเกตจากเหตุการณ์ หรือคดีที่เกี่ยวข้อง ซึ่งงานรูปแบบนี้ต้องอาศัยความเชี่ยวชาญเฉพาะบุคคล หากนำเทคโนโลยีเข้ามาช่วยในการติดตามหมายเลขผู้ต้องสงสัย จะช่วยให้งานสืบสวนทำได้อย่างรวดเร็วมากขึ้น ส่วนเทคโนโลยีที่เจ้าหน้าที่มีการใช้งานอยู่เป็นเพียงการกับบันทึกข้อมูลการใช้โทรศัพท์ที่จะได้ข้อมูลจากโทรศัพท์ Smart Phone มากกว่า และมีโปรแกรมที่แสดงกราฟที่เชื่อมโยงข้อมูลที่ได้มาเท่านั้น แต่ไม่สามารถวิเคราะห์หาความสัมพันธ์จากข้อมูลการใช้โทรศัพท์ได้

ผู้วิจัยได้ศึกษาทฤษฎีที่เกี่ยวข้องและงานวิจัยที่น่าสนใจ เพื่อหาแนวความคิดที่สามารถนำมาปรับใช้กับงานวิจัยได้ Social Network [1] คือการที่ผู้คนสามารถทำความรู้จัก และเชื่อมโยงกันในทิศทางใดทิศทางหนึ่ง ความสัมพันธ์ระหว่างบุคคล (nodes) คือ links แล้วนำมาสร้างเป็นกราฟเพื่ออธิบายการเชื่อมโยงกันระหว่าง nodes

Social Network Analysis [10–11] เป็นการวิเคราะห์ความเชื่อมโยงจากข้อมูลจะช่วยให้เข้าใจถึงโครงสร้างของเครือข่าย และความสัมพันธ์ของสมาชิกในเครือข่ายได้ดียิ่งขึ้น ตัวอย่างเช่น การวิเคราะห์เนื้อหาหรือข้อความที่มีอยู่ในสังคมออนไลน์ เป็นกระบวนการสำหรับการติดตามและเข้าใจในพฤติกรรมของมนุษย์ มีการนำทฤษฎีปัญญาทางสังคม (Social Cognitive Theory) ถูกนำมาใช้เพื่อติดตามข้อมูลใน Twitter เพื่อโอกาสสำหรับเสนอแผนสุขภาพให้กับบุคคลตามความสนใจ [3]

งานวิจัยที่มีการใช้ระยะทางของยูคลิด ในการแก้ปัญหาของการรู้จำใบหน้า [4], การรู้จำการเคลื่อนไหวของมนุษย์ [12] ซึ่งระยะทางของยูคลิด (Euclidean Distance) คือระยะระหว่างจุดสองจุดโดยมีข้อจำกัดว่าเวกเตอร์ทั้งสองต้องมีขนาดเท่ากัน หาได้จากสมการนี้

$$EU_{DIS} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (1)$$

นอกจากนั้นความผิดปกติของข้อมูล และความแปรปรวนของข้อมูลมีผลต่อค่าของ Euclidean Distance อย่างมาก [2] ซึ่งผู้วิจัยได้นำ Euclidean Distance มาประยุกต์ใช้ในการหาความต่างของพฤติกรรมการโทรของผู้ต้องสงสัยกับพฤติกรรมการโทรของหมายเลขอื่นๆ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลการใช้โทรศัพท์ ผู้วิจัยได้ศึกษางานวิจัยของ Rashmi Dutta Baruah และ Plamen Angelov [5] ได้เสนอวิธีการวิเคราะห์ Social Network จากข้อมูลโทรศัพท์มือถือ ใช้ชุดข้อมูลจาก VAST 2008 Challenge ใช้หลักการ Jaccard coefficient คำนวณค่า EI (Evolution Index) ติดตามการเปลี่ยนแปลงของเครือข่าย ซึ่งพิจารณากลุ่มของหมายเลขที่อยู่ในเครือข่ายที่มีร่วมกัน แล้วหาความสำคัญของแต่ละ node ใช้หลักการของ Degree – Centrality Measure พิจารณาจากระยะเวลาการโทร และความถี่ในการโทร ทำให้ทราบว่าในแต่ละวัน node ใดบ้างที่มีปริมาณการโทรสูง (โทรออกและรับสาย) จากนั้นพิจารณาหมายเลขที่มีรูปแบบการโทร ตำแหน่งที่โทรคล้ายกันโดยใช้หลักการ Clustering เข้ามาช่วย จัดกลุ่มหมายเลขโทรศัพท์ (400 หมายเลข) ตามคุณลักษณะช่วงเวลาในการโทร (24 ชั่วโมง) และ cell tower location (6 พื้นที่) ผลลัพธ์ของวิธีการที่ถูกนำเสนอนี้สามารถทำการติดตามได้ทั้งหมด 5 หมายเลขที่เปลี่ยนหมายเลขใหม่ของชุดข้อมูล VAST 2008 Challenge

Zhengbin Dong และคณะ [6] ได้ทำการศึกษาปัญหาใหม่ของ Individual Friendship pattern (IFP) เพื่ออธิบายลักษณะพฤติกรรมร่วมกันของแต่ละบุคคลที่อยู่ในบันทึกการโทรของโทรศัพท์มือถือ IFP แสดงให้ความถี่ของความสัมพันธ์และสิ่งสำคัญในการการดำเนินชีวิตของผู้ใช้งานโทรศัพท์มือถือ ซึ่งเป็นลักษณะเฉพาะของแต่ละบุคคลคล้ายกับลายนิ้วมือ ซึ่งสามารถนำมาประยุกต์ใช้ได้กับการสื่อสารโทรคมนาคม งานวิจัยนี้ทำการทดลองกับข้อมูลจริงของผู้ใช้งานโทรศัพท์มือถือในเมืองหนึ่งของประเทศจีน เพียงแค่ใช้ข้อมูล 50% - 70% ของข้อมูลที่เก็บมาก็สามารถหา IFP ของผู้ใช้งานหมายเลขโทรศัพท์ได้มากกว่า 80%

3. วิธีการดำเนินงาน

ในส่วนนี้จะอธิบายถึงแนวคิดในการวิเคราะห์หารูปแบบของความสัมพันธ์เพื่อติดตามหาหมายเลขโทรศัพท์ที่ได้เปลี่ยนไปใช้หมายเลขใหม่ ซึ่งมีขั้นตอนในการดำเนินงาน 3 ขั้นตอน ได้แก่

1) การเตรียมข้อมูล และการเข้ารหัสหมายเลขโทรศัพท์

2) วิเคราะห์หาพฤติกรรมการโทร (Calling vector)

3) เปรียบเทียบหาหมายเลขโทรศัพท์ที่มีพฤติกรรมคล้ายกับหมายเลขโทรศัพท์ของผู้ต้องสงสัยมากที่สุด

ผู้วิจัยได้เสนอแนวคิดในการติดตามหมายเลขโทรศัพท์ที่ผู้ต้องสงสัยเปลี่ยนไปเพื่อหลบหนีการสืบสวน โดยพิจารณาจากกลุ่มของหมายเลขที่เคยมีการติดต่อกัน และความถี่ของการโทรในช่วงเวลาที่ได้ทำการเก็บข้อมูลมา เนื้อหาในส่วนนี้ได้อธิบาย 5 ส่วน คือ 1) ข้อมูลที่ใช้ใน

ข้อสงวนลิขสิทธิ์

งานวิจัย 2) การเข้ารหัสหมายเลขโทรศัพท์ 3) การเตรียมข้อมูล 4) เวกเตอร์การโทร และ 5) การวิเคราะห์พฤติกรรมการโทร

3.1 ชุดข้อมูลที่ใช้ในงานวิจัย

ผู้วิจัยได้ทำการวิเคราะห์จากชุดข้อมูลทั้งหมด 3 ชุด คือ

1) ชุดข้อมูลตัวอย่างบันทึกการโทร จาก VAST 2008 challenge [8] ประกอบด้วย ข้อมูลการใช้โทรศัพท์ 10 วัน จำนวน 400 หมายเลข 9,834 เรคคอร์ด มีการใช้งานแบ่งออกเป็น 2 กลุ่ม กลุ่มที่ 1 คือ ข้อมูลบันทึกการโทรในช่วง 7 วันแรก และกลุ่มที่ 2 คือข้อมูลในช่วง 3 วันสุดท้าย ข้อมูลชุดนี้มีหมายเลขโทรศัพท์ 5 หมายเลขจากกลุ่มที่ 1 ที่เปลี่ยนหมายเลขการใช้งานในกลุ่มที่ 2

2) ชุดข้อมูลกรณีศึกษาจากกรมสอบสวนคดีพิเศษ ประกอบด้วยข้อมูลบันทึกการโทรจากผู้ให้บริการของหลายคดีผสมกัน ทั้งหมด 13,294 เรคคอร์ด มีข้อมูล 1,983 หมายเลข ในช่วงเวลา 196 วัน ซึ่งข้อมูลหมายเลขโทรศัพท์ถูกเข้ารหัสและแทนที่ทุกหมายเลข และผู้วิจัยไม่ทราบถึงข้อมูลเจ้าของหมายเลขโทรศัพท์ และหมายเลขโทรศัพท์ที่แท้จริง อย่างไรก็ตามข้อมูลที่ถูกเข้ารหัสเหล่านี้สามารถนำมาทำการวิเคราะห์พฤติกรรมการโทรได้

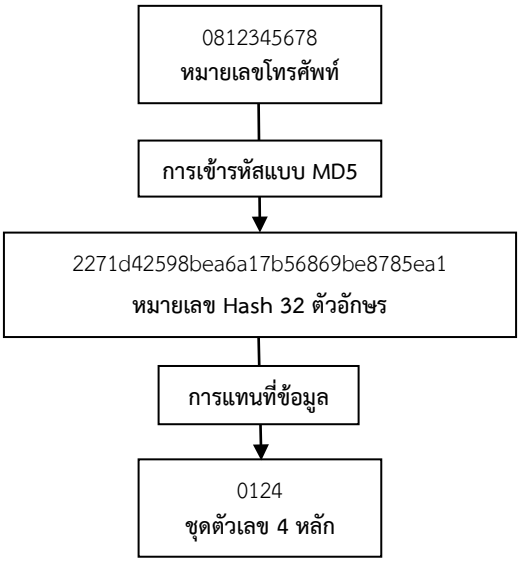
3) ชุดข้อมูลบันทึกการโทรจากนักศึกษากลุ่มตัวอย่าง ข้อมูลชุดนี้ดึงข้อมูลการโทรจากโทรศัพท์มือถือโดยตรงของกลุ่มตัวอย่างนักศึกษา 55 คน และข้อมูลเจ้าของเครื่องทั้งหมดถูกเก็บเป็นความลับและเข้ารหัสไว้ก่อนที่จะนำข้อมูลชุดนี้มาทำการวิจัย ข้อมูลการใช้งานโทรศัพท์ 304 วัน ทั้งหมด 7,675 เรคคอร์ด ประกอบด้วยหมายเลขที่มีการใช้งาน 1,368 หมายเลข, วัน – เวลา และระยะเวลาในการโทร

3.2 การเข้ารหัสหมายเลขโทรศัพท์

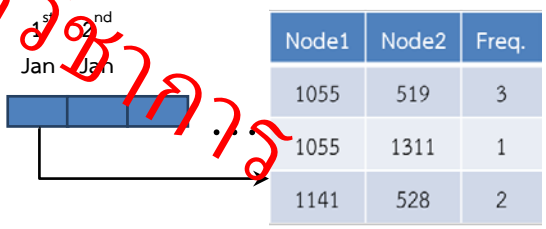
หมายเลขโทรศัพท์จะถูกปกปิดด้วยการเข้ารหัสแบบ MD5 [7] โดยอาศัย Hash function ซึ่งเป็นเทคนิคทางคณิตศาสตร์เพื่อใช้ในการเข้ารหัสข้อมูล และเป็นหนึ่งในวิธีที่ได้รับการยอมรับอย่างแพร่หลาย ดังนั้นหมายเลขต้นทาง และหมายเลขปลายทางจะถูกเข้ารหัส และเนื่องจากว่าการใช้หมายเลข Hash ซึ่งมีความยาว 32 ตัวอักษร จะไม่สะดวกในการวิจัย จึงจำเป็นต้องแทนที่หมายเลข Hash ด้วยชุดตัวเลข 4 หลัก ดังรูปที่ 1 เพราะหากใช้หมายเลข Hash จะใช้เวลานานในการเปรียบเทียบ เนื่องจากข้อมูลที่ใช้ในการวิจัยค่อนข้างมาก จึงใช้ชุดตัวเลข 4 ตัวมาแทนหมายเลข Hash เพื่อลดเวลาในการประมวลผล และรูปแบบผลลัพธ์สามารถจดจำง่ายกว่าหมายเลขที่มีขนาดยาว

3.3 การเตรียมข้อมูล

ข้อมูลที่ได้มาจากกรมสอบสวนคดีพิเศษ และกลุ่มตัวอย่างนักศึกษา ต้องทำการเข้ารหัสหมายเลขทั้งหมดก่อน และนำหมายเลขที่เข้ารหัสแล้วมาแทนด้วยหมายเลขอีกครั้งหนึ่ง ส่วนข้อมูลของ VAST 2008 เป็นข้อมูลที่แทนที่แทนด้วยหมายเลขมาอยู่แล้ว นำข้อมูลแต่ละชุดมานับความถี่ในการติดต่อกันของทุกหมายเลขในแต่ละวัน (เฉพาะโทรออก – รับสายเท่านั้น) เพื่อนำมาใช้ในการสร้างเวกเตอร์การโทร (Calling vector) ของแต่ละหมายเลข ดังรูปที่ 2 โดยกำหนดให้ Node1 คือหมายเลขโทรศัพท์ต้นทาง, Node2 คือหมายเลขโทรศัพท์ปลายทาง, Freq. คือความถี่ในการโทรของแต่ละหมายเลขที่ติดต่อกัน



รูปที่ 1 ตัวอย่างหมายเลขโทรศัพท์ที่ถูกเข้ารหัสแบบ MD5 และการแทนที่ข้อมูล



รูปที่ 2 การเตรียมข้อมูลเพื่อนับความถี่ในแต่ละวันของทุกหมายเลข

3.4 เวกเตอร์การโทร (Calling vector)

ผู้วิจัยได้เสนอแนวคิดในการเปรียบเทียบรูปแบบการโทร โดยสร้างเวกเตอร์การโทรเพื่อแสดงให้เห็นถึงพฤติกรรมการโทรของหมายเลขนั้นๆ ตัวอย่างเวกเตอร์การโทรของผู้ต้องสงสัย ประกอบด้วยหมายเลขที่ผู้ต้องสงสัยติดต่อด้วย และความถี่ในการโทรแต่ละหมายเลข ดังสมการนี้

$$\mathrm{Freq}(s,c) = \mathrm{Number_of_Call}\log(s,c) \qquad (2)$$

เมื่อ $\mathrm{Freq}(s,c)$ หมายถึง จำนวนความถี่ทั้งของการโทรติดต่อกันระหว่าง 2 หมายเลข ในชุดข้อมูล

s หมายถึง หมายเลขของผู้ต้องสงสัย

c หมายถึง หมายเลขที่ผู้ต้องสงสัยติดต่อด้วย

เช่น หมายเลข 0021 ติดต่อกับ 5 หมายเลข ซึ่งติดต่อหมายเลข 0216 จำนวน 1 ครั้ง เป็นต้น ดังตารางที่ 1

Contact Number	0216	0412	0603	0934	0938
0021	1	2	1	1	3

ตารางที่ 1 ตัวอย่างเวกเตอร์การโทร (Calling vector)

เนื่องจากขนาดของข้อมูลที่นำมาเปรียบเทียบกันมีจำนวนวันของข้อมูลไม่เท่ากัน อาจส่งผลกระทบในการหาค่า Euclidean Distance ผู้วิจัยได้ทำการหารข้อมูลความถี่แต่ละหมายเลขด้วยความถี่สูงสุดของเวกเตอร์นั้น จากสมการนี้

$$\text{Weight}(s, c) = \frac{\text{Freq}(s,c)}{\max \text{Freq}(s)} \quad (3)$$

$$\max \text{Freq}(s) = \max (\text{Freq}(s, c_1), \ldots, \text{Freq}(s, c_n)) \quad (4)$$

เมื่อ $\text{Weight}(s, c_i)$ หมายถึง ค่าความถี่ที่ถูกหารด้วยค่าความถี่สูงสุดในเวกเตอร์การโทร

$\max \text{Freq}(s)$ หมายถึง ค่าความถี่สูงสุดในเวกเตอร์การโทร

n หมายถึง จำนวนของหมายเลขที่ผู้ต้องสงสัยติดต่อด้วยทั้งหมด ในชุดข้อมูล

จากตารางที่ 1 ความถี่สูงสุดมีค่าเท่ากับ 3 นำไปหารค่าความถี่ในเวกเตอร์การโทรทุกตัว ตัวอย่างเช่น หมายเลข 0216 ที่มีค่าความถี่เป็น 1 ÷ 3 = 0.33 ค่าของความถี่ที่ได้จะอยู่ในช่วง [0, 1] ดังตารางที่ 2 ซึ่งในการเปรียบเทียบหาหมายเลขผู้ต้องสงสัยใหม่ โดยทั่วไปแล้วหมายเลขใดที่มีความถี่ในการติดต่อกันมาก ผู้ต้องสงสัยน่าจะมีความสัมพันธ์ใกล้ชิดกับเจ้าของหมายเลขนั้น แม้ว่าความถี่ในการโทรจะลดลง แต่หากยังคงติดต่อคนกลุ่มเดิมและลักษณะความถี่ยังคงคล้ายเดิมอยู่ ก็จะสามารถหาพฤติกรรมการโทรของผู้ต้องสงสัยได้

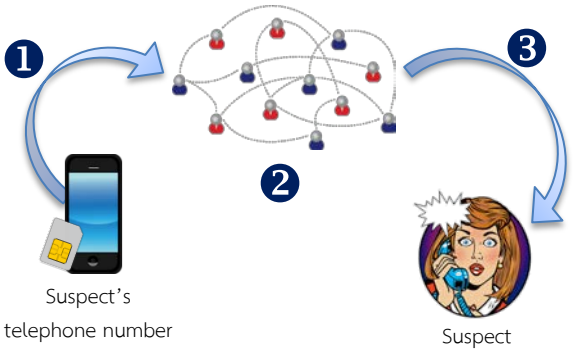
Contact Number	0216	0412	0603	0934	0938
0021	0.33	0.67	0.33	0.33	1.00

ตารางที่ 2 เวกเตอร์การโทรที่หารด้วยความถี่สูงสุดของเวกเตอร์นั้น

3.4 การวิเคราะห์พฤติกรรมการโทร

ผู้วิจัยเลือกคุณลักษณะที่ใช้ในการวิเคราะห์พฤติกรรมการใช้โทรศัพท์จากสมมติฐานว่า ผู้ต้องสงสัยเมื่อเปลี่ยนหมายเลขใหม่ไปแล้วก็คงยังติดต่อคนกลุ่มเดิม หรือมีพฤติกรรมในการโทรคล้ายเดิม โดยผู้วิจัยพิจารณาความถี่ในการโทร และหมายเลขที่ผู้ต้องสงสัยเคยติดต่อ เพื่อใช้ในการเปรียบเทียบหาหมายเลขที่มีพฤติกรรมการโทรคล้ายกับหมายเลขเดิมของผู้ต้องสงสัยมากที่สุด

ชุดข้อมูลจาก VAST 2008 เป็นข้อมูลที่เหมาะสมในการทดสอบสมมติฐานเนื่องจากสามารถเปรียบเทียบผลลัพธ์ที่ได้กับผลลัพธ์ที่มีการเผยแพร่ออกมา ส่วนข้อมูลจากกรมสอบสวนคดีพิเศษและนักศึกษากลุ่มตัวอย่าง ผู้วิจัยได้ทำการจำลองเสมือนว่ามีการเปลี่ยนเป็นหมายเลขใหม่ของหมายเลขเป้าหมายที่กำหนด



รูปที่ 3 กระบวนการในการติดตามหมายเลขโทรศัพท์ของผู้ต้องสงสัย

กระบวนการในการติดตามหาหมายเลขโทรศัพท์ที่ผู้ต้องสงสัยเปลี่ยนไป ดังรูปที่ 3 ประกอบด้วย 3 ขั้นตอน คือ

1) ดึงข้อมูลบันทึกการโทรจากโทรศัพท์มือถือของผู้ต้องสงสัย หรือการขอข้อมูลจากผู้ให้บริการ โดยเจ้าหน้าที่ที่มีอำนาจในการดำเนินการ

2) วิเคราะห์หาพฤติกรรมการใช้โทรศัพท์ เพื่อหารูปแบบการโทรเฉพาะบุคคล (Calling vector)

3) เปรียบเทียบหาหมายเลขที่มีพฤติกรรมการโทรคล้ายกับผู้ต้องสงสัยมากที่สุด

4. ผลการดำเนินงาน

ผู้วิจัยได้ทำการทดสอบการเปรียบเทียบพฤติกรรมการโทร 2 วิธีการ คือ

1) เปรียบเทียบพฤติกรรมการโทร จากความถี่ในการโทรของหมายเลขเป้าหมายเคยติดต่อ โดยหาค่า Euclidean Distance ระหว่างเวกเตอร์การโทร หาหมายเลขที่มีค่า Euclidean Distance น้อยที่สุด

2) เปรียบเทียบจากจำนวนหมายเลขที่เหมือนกับกลุ่มหมายเลขที่หมายเลขเป้าหมายเคยติดต่อด้วย (Common contacts) หมายเลขที่มีพฤติกรรมการโทรคล้ายกันจะมีจำนวนหมายเลขที่ซ้ำกลุ่มเดิมอยู่มากที่สุด

เพื่อทดสอบประสิทธิภาพวิธีการในการติดตามหมายเลขที่เปลี่ยนไป ผู้วิจัยจึงให้ผลลัพธ์ของหมายเลขผู้ต้องสงสัยเพียง 1 หมายเลขเท่านั้น

การทดลองที่ 1 : ผู้วิจัยได้ใช้ข้อมูลจาก VAST 2008 เพื่อทดสอบสมมติฐานเบื้องต้นและตรวจสอบคุณลักษณะที่จะใช้ในการวิเคราะห์พฤติกรรมการโทร ลักษณะของข้อมูลชุดนี้มี 5 หมายเลขที่เปลี่ยนแปลง

ห้ามเผยแพร่ เพื่อการศึกษาเท่านั้น

หมายเลขที่เคยใช้หลังจากวันที่ 7 ดังนั้นข้อมูลก่อนการเปลี่ยนหมายเลขมี 7 วัน และข้อมูลหลังจากเปลี่ยนหมายเลขมี 3 วัน

ทำการสร้างเวกเตอร์การโทรของทั้ง 5 หมายเลขโดยใช้ข้อมูล 7 วันแรกของชุดข้อมูล และสร้างเวกเตอร์ของทุกหมายเลขโดยใช้ข้อมูล 3 วันสุดท้ายของชุดข้อมูลเพื่อเปรียบเทียบหาหมายเลขที่มีพฤติกรรมการโทรคล้ายกับ 5 หมายเลขเป้าหมายมากที่สุด

จากทั้ง 2 วิธีการสามารถติดตามหมายเลขได้ 4 หมายเลข จากทั้งหมด 5 หมายเลข ดังตารางที่ 3

Day 1 – 7 Cell	Day 8 -10 Cell	
	Published Result	Result
0001	0309	0309
0002	0397	0397
0003	0360	0360
0005	0306	0306
0200	0300	✖

ตารางที่ 3 ผลลัพธ์ที่ได้กับผลลัพธ์ที่เผยแพร่จาก VAST 2008 Challenge

จากตารางที่ 3 หมายเลข 0200 ไม่สามารถติดตามหาหมายเลขที่เปลี่ยนไปได้ เนื่องจากกลุ่มของหมายเลขที่เคยติดต่อได้เปลี่ยนหมายเลขไปด้วยเกือบทั้งหมด ทำให้การติดตามหมายเลขเป็นไปได้ยาก ซึ่งวิธีการทั้ง 2 ผู้วิจัยพิจารณาได้จาก 2 คุณลักษณะ คือ หมายเลขที่ติดต่อร่วมกัน และความถี่ในการโทร ก็สามารถติดตามหาหมายเลขที่เปลี่ยนไปได้ ส่วนผลลัพธ์จากงานวิจัยที่ผ่านมา [5] มีวิธีการในการหาซับซ้อนกว่า โดยพิจารณาจาก 5 คุณลักษณะ คือ หมายเลขที่ติดต่อ ความถี่ในการโทร ช่วงเวลา ระยะเวลาในการโทร และcell tower location หากเป็นการดึงข้อมูลมาจากโทรศัพท์เคลื่อนที่ จะไม่ได้ข้อมูล cell tower location ก็จะไม่สามารถใช้วิธีการนี้ได้

การทดลองที่ 2 : ผู้วิจัยทดสอบทั้ง 2 วิธีการ โดยใช้ข้อมูลจากกรมสอบสวนคดีพิเศษ และกลุ่มนักศึกษาตัวอย่าง ซึ่งข้อมูลมีความใกล้เคียงกับสถานการณ์จริง ผู้วิจัยใช้ขนาดของข้อมูลเหมือนกับงานวิจัยที่ผ่านมาคือ ข้อมูลก่อนเปลี่ยนมี 7 วัน และข้อมูลหลังเปลี่ยน 3 วัน ที่สามารถติดตามหาหมายเลขที่เปลี่ยนไปได้จากการทดลองที่ 1 แต่ข้อมูลทั้ง 2 ชุดที่นำมาทำการทดลองไม่ได้มีการเปลี่ยนหมายเลขจริง จึงได้ทำการจำลองว่ามีการเปลี่ยนหมายเลขไป

ผลการทดลองพบว่า วิธีการที่ 1 สามารถติดตามหมายเลขได้ 70% ของหมายเลข ส่วนวิธีการที่ 2 สามารถติดตามได้ 90% ของหมายเลข จะเห็นได้ว่าวิธีการที่ 2 มีประสิทธิภาพในการติดตามหมายเลขของผู้ต้องสงสัยได้ดีกว่า เนื่องจากสามารถติดตามหมายเลขที่เปลี่ยนไปได้มากกว่าวิธีการที่ 1 และพบปัญหาเมื่อมีหมายเลขที่มีพฤติกรรมการโทรคล้ายกับหมายเลขเป้าหมายมากกว่า 1 หมายเลข ซึ่งผู้วิจัยได้เสนอแนวทางการแก้ไขปัญหาดังกล่าว ในการทดลองที่ 3

การทดลองที่ 3 : ผู้วิจัยปรับปรุงวิธีการในการเปรียบเทียบพฤติกรรมการโทร โดยใช้ข้อมูลจากกรมสอบสวนคดีพิเศษ และกลุ่มนักศึกษาตัวอย่าง ขั้นตอนแรกใช้วิธีการที่ 2 ในการเปรียบเทียบหาหมายเลขที่มีพฤติกรรมคล้ายกับผู้ต้องสงสัย หากมีผลลัพธ์มากกว่า 1 หมายเลข ใช้วิธีการที่ 1 เปรียบเทียบพฤติกรรมการโทรอีกครั้ง ขนาดของข้อมูลที่ใช้คือ ข้อมูลก่อนเปลี่ยนมี 7 วัน และข้อมูลหลังเปลี่ยน 3 วัน ตัวอย่างเช่นหมายเลข 0021 เป็นหมายเลขผู้ต้องสงสัยที่ติดต่อกับ 6 หมายเลขก่อนทำการเปลี่ยนหมายเลขไป และมีความถี่ในการโทรแต่ละหมายเลข ดังตารางที่ 4

ใช้วิธีการที่ 2 ในการเปรียบเทียบพฤติกรรมการโทร มีหมายเลขที่มีพฤติกรรมการโทรคล้ายกับหมายเลขผู้ต้องสงสัยอยู่ 5 หมายเลข ซึ่งแต่ละหมายเลขมีเวกเตอร์การโทร ดังตารางที่ 5

Tel. Number	0216	0412	0603	0755	0934	0938
0021	1	2	1	2	1	3

ตารางที่ 4 เวกเตอร์การโทรของหมายเลข 21

นำเวกเตอร์การโทรของแต่ละหมายเลขมาใช้วิธีการที่ 1 เพื่อลดจำนวนหมายเลขที่จะเป็นผลลัพธ์ เหลือเพียงหมายเลข 9999 เท่านั้นที่มีค่า EU_{DIS} = 1.1066 ซึ่งน้อยที่สุด ส่วนที่เหลืออีก 4 หมายเลขมีค่า EU_{DIS} = 1.4142 เท่ากัน เพื่อให้มีความถูกต้องเพิ่มมากขึ้นและลดจำนวนของผลลัพธ์ในบางกรณีให้น้อยลง ซึ่งวิธีการนี้สามารถทำให้การติดตามหมายเลขที่เปลี่ยนไปได้ถูกต้อง 95% ของหมายเลขทั้งหมด

	Tel. Number	0216	0412	0603	0755	0934	0938
Calling vector	9999	0	0	0	0	0	1
	0080	0	0	0	0	1	0
	0223	0	0	0	0	1	0
	0487	0	0	0	0	1	0
	0726	0	0	0	0	1	0

ตารางที่ 5 เวกเตอร์การโทรของหมายเลขที่มีพฤติกรรมการโทรคล้ายกับหมายเลขเป้าหมายมากที่สุด

5. สรุป

ผู้วิจัยได้เสนอแนวคิดและวิธีการในการเปรียบเทียบพฤติกรรมการโทรของผู้ต้องสงสัย ใช้ในการติดตามหมายเลขโทรศัพท์ของผู้ต้องสงสัยที่เปลี่ยนหมายเลขโทรศัพท์เพื่อหลบหลีกการสืบสวนจากเจ้าหน้าที่ ซึ่งใช้ข้อมูลบันทึกการโทรในการวิเคราะห์หาพฤติกรรมการโทรของผู้ต้องสงสัย โดยสร้างเวกเตอร์การโทรแสดงถึงพฤติกรรมการโทรของหมายเลขนั้นๆ

ประกอบด้วยหมายเลขที่หมายเลขนั้นเคยติดต่อด้วย และความถี่ในการโทรติดต่อกับหมายเลขนั้น (โทรออกและรับสาย) การวิเคราะห์จากจำนวนหมายเลขที่มีการติดต่อร่วมกัน ติดตามหมายเลขที่ติดต่อคนกลุ่มเดิมกับหมายเลขเป้าหมายเคยติดต่อด้วยมากที่สุด แต่บางกรณีอาจจะมีหมายเลขที่ต้องติดตามมากกว่า 1 หมายเลข ผู้วิจัยนำหมายเลขเหล่านั้นมาทำการวิเคราะห์หาค่า Euclidean Distance ระหว่างเวกเตอร์การโทรของแต่ละหมายเลขกับหมายเลขเป้าหมาย จากวิธีการนี้สามารถติดตามหมายเลขได้ 95% ของหมายเลขจาก 3 ชุดข้อมูล

บางกรณีที่วิธีการนี้ไม่สามารถติดตามหมายเลขได้เนื่องจากสภาพแวดล้อมและกลุ่มหมายเลขที่เคยติดต่อด้วยเปลี่ยนไปค่อนข้างมาก ส่งผลให้พฤติกรรมการโทรเปลี่ยนไปจากเดิม อาจจะไม่ติดต่อกับคนกลุ่มเดิม หรือกลุ่มเบอร์ที่เคยติดต่อด้วยเปลี่ยนหมายเลขเช่นกัน ซึ่งกรณีนี้เป็นข้อจำกัดของงานวิจัย การพัฒนางานวิจัยในอนาคต หมายเลขที่นำมาเปรียบเทียบกับพฤติกรรมในการโทรของผู้ต้องสงสัย ผลลัพธ์ที่ได้อาจจะบอกเป็นเปอร์เซ็นต์ของความน่าจะเป็นที่จะเป็นหมายเลขโทรศัพท์ที่ผู้ต้องสงสัยเปลี่ยนไปใช้ เนื่องจากในสถานการณ์จริงอาจมีหมายเลขที่เป็นไปได้มากกว่า 1 หมายเลข ซึ่งเจ้าหน้าที่จะติดตามหมายเลขเหล่านั้นไปพร้อมๆ กัน เพื่อสังเกตพฤติกรรมต่อไป

ห้ามส่งวารสารออกนอก

6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากสำนักงานคณะกรรมการวิจัยแห่งชาติ – Cyber Security ประจำปีงบประมาณ 2556

เอกสารอ้างอิง

[1] Hansen, D. L. , Shneiderman, B. , & Smith, M. A. , (2011). *Analyzing social media networks with NodeXL : insights from a connected world*. Burlington, MA : Morgan Kaufmann.

[2] Mitsa, T. , (2010). *Temporal data mining*. Boca Raton, Fla. : Chapman &Hall/CRC.

[3] Yoon, Hong-Jun; Tourassi, Georgia, "Analysis of online social networks to understand information sharing behaviors through social cognitive theory," *Biomedical Science and Engineering Center Conference (BSEC), 2014 Annual Oak Ridge National Laboratory* , pp.1,4, 6-8 May 2014.

[4] Malkauthekar, M.D., "Analysis of euclidean distance and Manhattan Distance measure in face recognition," *Computational Intelligence and Information Technology, 2013. CIIT 2013. Third International Conference on* , pp.503,507, 18-19 Oct. 2013.

[5] Baruah, R.D.; Angelov, P., "Evolving social network analysis: A case study on mobile phone data," *Evolving and Adaptive Intelligent Systems (EAIS), 2012 IEEEConference on*, pp.114, 120, 17-18 May 2012.

[6] Zhengbin Dong; Guojie Song; KunqingXie; Yixian Sun; Jingyao Wang, "Adequacy of Data for Mining Individual Friendship Pattern from Cellular Phone Call Logs," *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on* , vol.5, no., pp.573,577, 14-16 Aug. 2009.

[7] Zhao Yong-Xia; Zhen Ge, "MD5 Research," *Multimedia and Information Technology (MMIT), 2010 Second International Conference on*, vol.2, pp.271, 273, 24-25 April 2010.

[8] Grinstein, G.; Plaisant, C.; Laskowski, S.; O'Connell, T.; Scholtz, J.; Whiting, M., "VAST 2008 Challenge: Introducing mini-challenges," *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on* , 19-24 Oct. 2008.

[9] Zhuang Cheng, "Regional Economic Indicators Analysis Based on Data Mining," *Intelligent Systems Design and Engineering Applications (ISDEA), 2014 Fifth International Conference on* , pp.726,730, 15-16 June 2014.

[10] Khanafeev, D.; Luc, C.; Taehyung Wang, "Social Network Data Mining Using Natural Language Processing and Density Based Clustering," *Semantic Computing (ICSC), 2014 IEEE International Conference on* , pp.250,251, 16-18 June 2014.

[11] Srivastava, A.; Anuradha; Gupta, D.J., "Social Network Analysis: Hardly easy," *Optimization, Reliabilty, and Information Technology (ICROIT), 2014 International Conference on* , pp.128,135, 6-8 Feb. 2014.

[12] Javed, J.; Yasin, H.; Ali, S.F., "Human movement Recognition using Euclidean Distance: A tricky approach," *Image and Signal Processing (CISP), 2010 3rd International Congress on* , vol.1, pp.317,321, 16-18 Oct. 2010.