ระบบบูรณาการสารสนเทศมัลแวร์

MALWARE INFORMATION INTEGRATION SYSTEM

ศูนย์วิจัยเทคโนโลยีซอฟต์แวร์และระบบสารสนเทศ

คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2552

KMITL-2009-EN-SE-SEC-089

ระบบบูรณาการสารสนเทศมัลแวร์

MALWARE INFORMATION INTEGRATION SYSTEM

ชิตนนท์ เผื่อนพิภพ

CHITTANOND PHEUNPHIPHOP

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมสารสนเทศ

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2552

KMITL-2009-EN-M-230-089

MALWARE INFORMATION INTEGRATION SYSTEM

CHITTANOND PHEUNPHIPHOP

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN INFORMATION ENGINEERING

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2009

KMITL-2009-EN-M-230-089

คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ ระบบบูรณาการสารสนเทศมัลแวร์

Thesis Title Malware Information Integration System

นักศึกษา นายชิตนนท์ เผื่อนพิภพ

รหัสประจำตัว 47061137

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชา วิศวกรรมสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ.มยุรี เลิศเวชกุล

หมายเลขวิทยานิพนธ์ KMITL-2009-EN-M-230-089

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
ผศ.ดร.พิทักษ์	ธรรมวาริน	
รศ.ดร.ชวลิต	เบญจางคประเสริฐ	
รศ.ดร.ปิติเขต	สู้รักษา	
ดร.วีระพล	โมนยะกุล	
ผศ.มยุรี	เลิศเวชกุล	

วัน / เดือน/ ปี ที่สอบ วันพฤหัสบดีที่ 28 พฤษภาคม พ.ศ. 2552 เวลา 14.30-16.30 น.

สถานที่สอบ ณ อาคาร A ชั้น 3 ห้องประชุม 2

คณะวิศวกรรมศาสตร์ รับรองแล้ว

(รองศาสตราจารย์ ดร.กอบชัย เดชหาญ)

คณบดี คณะวิศวกรรมศาสตร์

วันที่ 28 พฤษภาคม พ.ศ. 2552

หัวข้อวิทยานิพนธ์	ระบบบูรณาการสารสนเทศมัลแวร์
นักศึกษา	นายชิตนนท์ เผื่อนพิภพ
รหัสนักศึกษา	47061137
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมสารสนเทศ
พ.ศ.	2552
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.มยุรี เลิศเวชกุล

บทคัดย่อ

การปกป้องตนเองจากมัลแวร์ของผู้ใช้นั้นต้องอาศัยความรู้ความเข้าใจเกี่ยวกับมัลแวร์ (Malware) ซึ่งจำเป็นต้องได้รับข้อมูลที่ทันสมัยและเชื่อถือได้จากเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์ แต่เนื่องจากแหล่งข้อมูลอธิบายมัลแวร์โดยการใช้คีย์เวิร์ดและคำศัพท์เทคนิคที่แตกต่างกันซึ่งมักจะสร้างความสับสนให้กับผู้ใช้

เพื่อเป็นการแก้ไขปัญหาดังกล่าวนี้งานวิจัยฉบับนี้ได้นำเสนอ ระบบบูรณาการสารสนเทศมัลแวร์ ซึ่งเป็นการบูรณาการแหล่งข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาไว้ภายใต้มาตรฐานเดียวกันด้วยการแก้ไขปัญหาเรื่องความหลากหลายทางด้านความหมาย (Semantic Heterogeneous) และความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneous) ด้วยการใช้ระบบสหภาพสารสนเทศ (Federation Information System) ซึ่งได้นำเสนอการสร้างสคีมากลาง (Global Schema) และเมตาดาต้า (Metadata) เพื่อให้ผู้ใช้สามารถเข้าถึงแหล่งข้อมูลได้อย่างเป็นหนึ่งเดียวกัน และยังได้กำหนดหมายเลขประจำตัวร่วมมัลแวร์ (Malware Common ID: MCID) เพื่อใช้ในการแก้ปัญหามัลแวร์ที่มีชื่อเรียกแตกต่างกัน ระบบบูรณาการสารสนเทศมัลแวร์ประกอบด้วยโปรแกรมซึ่งสามารถรวบรวมคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาวิเคราะห์ และบันทึกลงในฐานความรู้มัลแวร์โดยอัตโนมัติ และระบบบูรณาการสารสนเทศมัลแวร์ยังมีการออกแบบส่วนเชื่อมต่อผู้ใช้ที่เปิดโอกาสให้ผู้เชี่ยวชาญสามารถแบ่งปันความรู้เกี่ยวกับมัลแวร์ให้กับผู้ใช้ท่านอื่น ๆ ได้อีกด้วย

Thesis Title	Malware Information Integration System
Student	Mr. Chittanond Pheunphiphop
Student ID.	47061137
Degree	Master of Engineering
Program	Information Engineering
Year	2009
Thesis Advisor	Asst.Prof. Mayuree Lertwatechakul

ABSTRACT

In order to protect computer systems from malware, users must have reliable and up-to-date knowledge about malware. Since anti-malware websites could be the best knowledge sources for users. But each website may arbitrarily define different keywords and technical terms to describe the same thing. This may make the information become non-integrity to people's sense.

In order to solve the problem, we have proposed Malware Information Integration System which integrates malware information from various websites into a unified database. We also have solved heterogeneous problems such as semantic heterogeneous and schematic heterogeneous that usually occur to information integration from various sources. By implementing a federation information system which provides global schema and metadata. The system allows users to access each information source as a unified database. Besides, we have defined MCID to cross reference the information of a malware among many sources because the malware names may be differently defined by each sources. Malware information integration composes of malware information gathering tool that retrieve, analyze and store malware description from various websites into a unified malware information integration system. More over malware information integration system provides interface for malware experts to share their knowledge to other users.

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยความกรุณาจากอาจารย์ที่ปรึกษา ผศ. มยุรี เลิศเวชกุล ที่ให้ความช่วยเหลือตลอดจนดูแลเอาใจใส่งานวิจัยของข้าพเจ้าเป็นอย่างดีและพร้อมที่จะให้คำปรึกษาแก่ข้าพเจ้าทุกเมื่อด้วยบรรยากาศที่เป็นกันเองและในโอกาสนี้ข้าพเจ้าขอขอบคุณ รศ.ดร. ปิติเขต สู้รักษา ที่ให้ความช่วยเหลือและคำปรึกษาแก่ข้าพเจ้าในทุก ๆ เรื่อง นอกจากนั้นข้าพเจ้าขอขอบคุณเพื่อน ๆ พี่ ๆ และน้อง ๆ ในห้องปฏิบัติการทุกคนที่ให้ความรู้และความช่วยเหลือเมื่อข้าพเจ้าประสบปัญหา

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดา มารดา และญาติที่เคารพยิ่ง ตลอดจนครูอาจารย์ ที่เคารพทุกท่านที่ได้ประสิทธิประสาทวิชาความรู้และถ่ายทอดประสบการณ์ที่ดีแก่ข้าพเจ้า

ชิตนนท์ เผื่อนพิภพ

สารบัญ

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

สารบัญรูป

สารบัญรูป (ต่อ)

รูปที่ หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันมีผู้ใช้คอมพิวเตอร์จำนวนมากที่ได้รับผลกระทบจากซอฟต์แวร์ที่มีเจตนาร้าย หรือมัลแวร์ (Malware: Malicious Software) ซึ่งนับวันมัลแวร์เหล่านี้จะมีจำนวนเพิ่มมากขึ้น โดยปรกติแล้วผู้ใช้คอมพิวเตอร์ทั่วไปยังขาดความรู้ความเข้าใจด้านมัลแวร์อยู่มาก อันเนื่องมาจากแหล่งข้อมูลเกี่ยวกับมัลแวร์นั้นจะจำกัดอยู่ในเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์เท่านั้น ซึ่งข้อมูลคำอธิบายมัลแวร์เหล่านี้จะถูกจัดเตรียมโดยผู้ผลิตโปรแกรมกำจัดมัลแวร์ค่ายต่าง ๆ โดยข้อมูลที่มีประโยชน์เหล่านี้ถูกเขียนขึ้นและนำเสนอในลักษณะแตกต่างกันออกไป โดยแต่ละเวปไซต์นั้นจะมีการใช้คำศัพท์เทคนิคและโครงสร้างข้อมูลคำอธิบายมัลแวร์ที่กำหนดขึ้นมาเอง แม้แต่ชื่อของมัลแวร์ก็ถูกเรียกแตกต่างกันไปตามแต่ละค่าย ทำให้มัลแวร์หนึ่งตัวอาจจะมีชื่อเรียก (Alias Name) ได้หลายชื่อ ปัญหาเหล่านี้นอกจากจะเป็นอุปสรรคต่อการแลกเปลี่ยนข้อมูลระหว่างผู้ผลิตด้วยกันเองแล้วยังเป็นอุปสรรคที่มีผลต่อการศึกษาข้อมูลมัลแวร์ของผู้ใช้ทั่วไปเป็นอย่างมาก

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

วิทยานิพนธ์ฉบับนี้มุ่งหวังเพื่อแก้ปัญหาที่เกิดขึ้นจากการบูรณาการแหล่งข้อมูลคำอธิบายมัลแวร์ที่หลากหลาย (Heterogeneous Data Sources) ซึ่งประกอบไปด้วย ปัญหาความหลากหลายทางด้านความหมาย (Semantic Heterogeneity) และความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneity) รวมถึงปัญหาชื่อมัลแวร์ที่แตกต่างกันไปในแต่ละแหล่งข้อมูลอีกด้วย ทั้งนี้เพื่อให้สามารถเก็บรวบรวมคำอธิบายมัลแวร์จากแหล่งข้อมูลที่หลากหลายลงในระบบฐานความรู้มัลแวร์ภายใต้มาตรฐานเดียวกัน

1.3 สมมติฐานของการศึกษา

เนื่องจากแหล่งความรู้เกี่ยวกับมัลแวร์ที่มีอยู่ในปัจจุบันนั้นส่วนใหญ่จะเผยแพร่อยู่ตามเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์ โดยโครงสร้างข้อมูล หรือคำศัพท์เทคนิค ก็มักจะถูกกำหนดให้แตกต่างกันออกไป ซึ่งอาจจะทำให้ผู้ใช้เกิดความสับสนได้ ประกอบกับคำอธิบายมัลแวร์จากแต่ละแหล่งข้อมูลนั้นอธิบายมัลแวร์ตัวเดียวกันด้วยข้อมูลรายละเอียดที่แตกต่างกันออกไป บางแหล่งข้อมูลอธิบายการทำงานของมัลแวร์ได้ละเอียดชัดเจน เข้าใจได้ง่าย บางแหล่งข้อมูลอธิบายวิธีการกำจัดมัลแวร์ได้ดีและสามารถนำไปใช้งานได้จริง ถ้าหากว่าเราสามารถ

บูรณาการแหล่งข้อมูลที่หลากหลายเหล่านี้เข้าไว้ในระบบฐานความรู้มัลแวร์เดียวกันและมีมาตรฐานโครงสร้างข้อมูลรูปแบบเดียวกันได้ ก็จะสามารถช่วยให้ผู้ที่สนใจมัลแวร์หรือผู้ใช้ที่ประสบกับปัญหาจากมัลแวร์ หาวิธีแก้ไขหรือทำความเข้าใจกับมัลแวร์ต่าง ๆ ได้สะดวกยิ่งขึ้น

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

ในวิทยานิพนธ์ฉบับนี้ได้นำเสนอวิธีการแก้ไขปัญหาที่เกิดจากการบูรณาการแหล่งข้อมูลมัลแวร์ที่หลากหลาย ซึ่งงานวิจัยนี้ได้นำเสนอ ตารางเปรียบเทียบคำศัพท์เทคนิค (Technical Terms to General Terms Mapping Table) และเมตาดาต้าที่ใช้อธิบายรูปแบบการแสดงผล (Presentation Metadata) คำอธิบายมัลแวร์ที่แตกต่างกันออกไปในแต่ละเวปไซต์เพื่อแก้ไขปัญหาความหลากหลายทางด้านโครงสร้างและยังนำเสนอวิธีการแก้ไขปัญหาที่เกิดจากความหลากหลายด้านความหมายโดยการกำหนดหมายเลขประจำตัวร่วมมัลแวร์ (Malware Common ID) ซึ่งมีส่วนช่วยในการรวบรวมมัลแวร์ที่เป็นตัวเดียวแต่มีเรียกแตกต่างกันไปตามแหล่งข้อมูลให้อยู่ภายใต้ฐานความรู้มัลแวร์เดียวกันและมีเลขประจำตัวร่วมมัลแวร์หมายเลขเดียวกันรวมถึงการจำแนกประเภทมัลแวร์และการประเมินระดับภัยคุกคามที่เกิดจากมัลแวร์ซึ่งสามารถบันทึกข้อมูลประเภทของมัลแวร์และระดับภัยคุกคามที่เกิดจากเกิดจากมัลแวร์ที่มาจากแหล่งข้อมูลที่แตกต่างกันลงในฐานความรู้มัลแวร์ภายใต้มาตรฐานเดียวกันได้

1.5 ขอบเขตการวิจัย

ในวิทยานิพนธ์ฉบับนี้ได้นำเสนอระบบฐานความรู้มัลแวร์ที่ได้บูรณาการจากแหล่งข้อมูลคำอธิบายมัลแวร์ที่หลากหลายจากเวปไซต์ เพื่อให้ผู้ใช้สามารถเข้ามาศึกษาและค้นหาข้อมูลเกี่ยวกับมัลแวร์ต่าง ๆ ได้อย่างง่ายดาย รวดเร็ว และรอบด้าน และยังเสนอการกำหนดหมายเลขประจำตัวร่วมมัลแวร์ (Malware Common ID) เพื่อให้สามารถระบุได้ว่ามัลแวร์ตัวใดบ้างที่เป็นตัวเดียวกันถึงแม้ว่ามัลแวร์เหล่านั้นจะถูกเรียกด้วยชื่อที่ต่างกันไปในแต่ละแหล่งข้อมูล พร้อมทั้งเปิดโอกาสให้ผู้ที่มีความเชี่ยวชาญด้านมัลแวร์สามารถแบ่งปันความรู้หรือประสบการณ์ในการแก้ปัญหาที่เกิดขึ้นจากมัลแวร์กับผู้ใช้ทั่วไปผ่านทางเวปไซต์ที่เราได้ออกแบบไว้ และยังอนุญาตให้ผู้ใช้ที่ได้ลงทะเบียนเป็นสมาชิกกับทางเวปไซต์สามารถลงคะแนนโหวตให้กับคำอธิบายมัลแวร์ที่ตนเองอ่านแล้วเห็นว่ามีประโยชน์หรือมีเนื้อหาค่อนข้างครอบคลุมและละเอียด ได้ด้วยตนเอง นอกจากนี้ระบบฐานความรู้มัลแวร์ยังมีเครื่องมือสำหรับช่วยรวบรวมคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาเก็บลงในฐานข้อมูลอีกด้วย

1.6 ขั้นตอนของการศึกษา

มีขั้นตอนการศึกษาดังนี้

1.6.1 ค้นคว้าเอกสารและข้อมูลต่าง ๆ ที่เกี่ยวข้อง

1.6.2 ค้นคว้าหาทฤษฎีของมัลแวร์แต่ละประเภท

1.6.3 วิเคราะห์คำอธิบายมัลแวร์จากแหล่งข้อมูลที่หลากหลาย

1.6.4 ออกแบบโครงสร้างฐานข้อมูล

1.6.5 ออกแบบโปรแกรม

1.6.6 ทำการทดสอบและพัฒนาโปรแกรม

1.6.7 สรุปและเรียบเรียงวิทยานิพนธ์

1.7 โครงสร้างของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทคือ

บทที่ 1 กล่าวถึงความเป็นมาของงานวิจัย ความมุ่งหมาย วัตถุประสงค์ สมมติฐาน ทฤษฎีที่ใช้ ขอบเขตของการวิจัย และขั้นตอนการศึกษา

บทที่ 2 กล่าวถึงความหมายของมัลแวร์ ประเภทของมัลแวร์ พฤติกรรมของมัลแวร์ ความเสียหายที่เกิดขึ้นจากมัลแวร์ แบบจำลองไนแอม และทฤษฎีการบูรณาการข้อมูลสารสนเทศ

บทที่ 3 กล่าวถึงการออกแบบโครงสร้างฐานข้อมูลของมัลแวร์ การออกแบบตารางเปรียบเทียบคำศัพท์เทคนิค การออกแบบเครื่องมือที่ทำหน้าที่รวบรวมคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาเก็บลงในฐานข้อมูล ระบบลงทะเบียนแหล่งข้อมูลคำอธิบายมัลแวร์

บทที่ 4 ผลการทดลอง

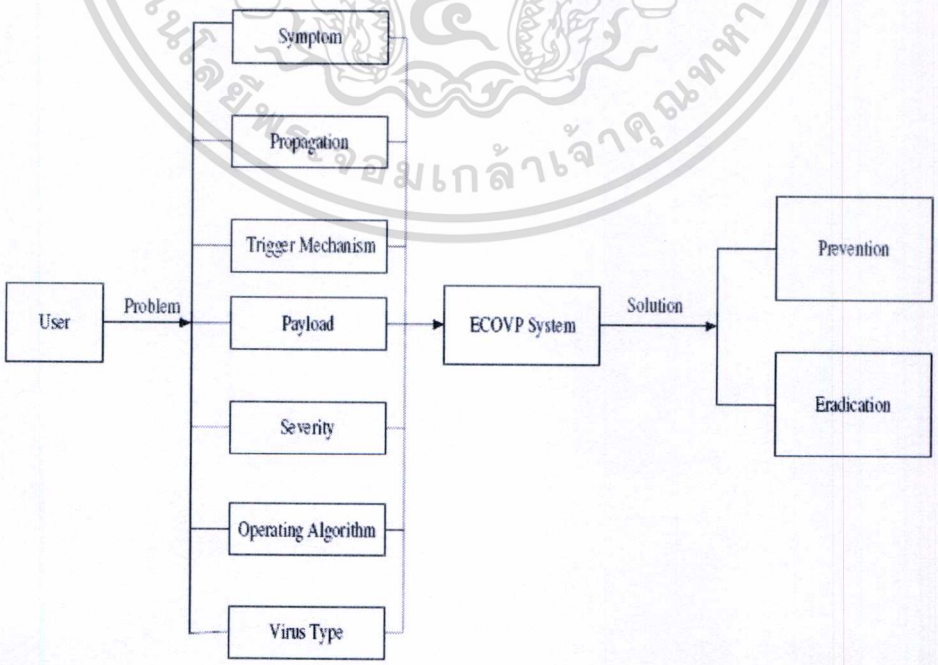
บทที่ 5 เป็นบทสรุปผลการวิจัย

บทที่ 2
ทฤษฎีพื้นฐานที่ใช้ในการวิจัย

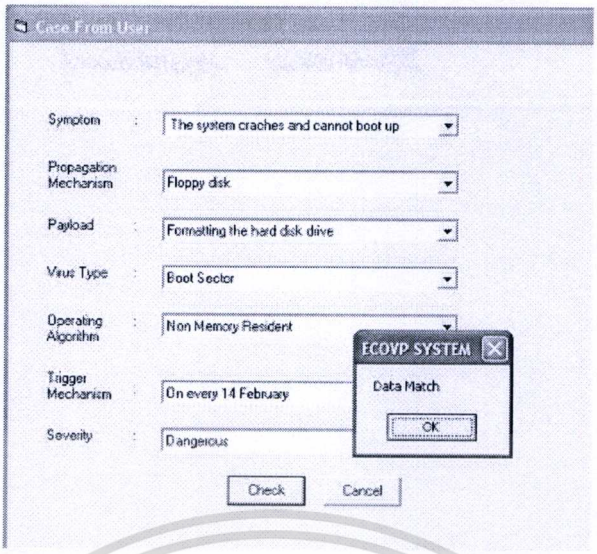
2.1 งานวิจัยที่เกี่ยวข้อง

ในการที่ผู้ใช้จะป้องกันตนเองจากมัลแวร์ได้นั้นสิ่งสำคัญที่สุดคือการมีความรู้ความเข้าใจที่ถูกต้องเกี่ยวกับมัลแวร์ในงานวิจัยของ Saudi Madihah และ Jomhari Nazean [1] จากคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัย Islamic ประเทศมาเลเซียในปี 2006 ได้นำเสนอโครงสร้างความรู้เกี่ยวกับไวรัสที่ได้ออกแบบตามแนวความคิดของ Marko Helenius [2] โดยมีจุดประสงค์ในการสร้างระบบที่สามารถให้คำแนะนำถึงวิธีการป้องกันและวิธีการแก้ไขปัญหาที่เกิดจากไวรัสแต่ละตัวให้กับผู้ใช้ที่ประสบปัญหาเกี่ยวกับไวรัสได้

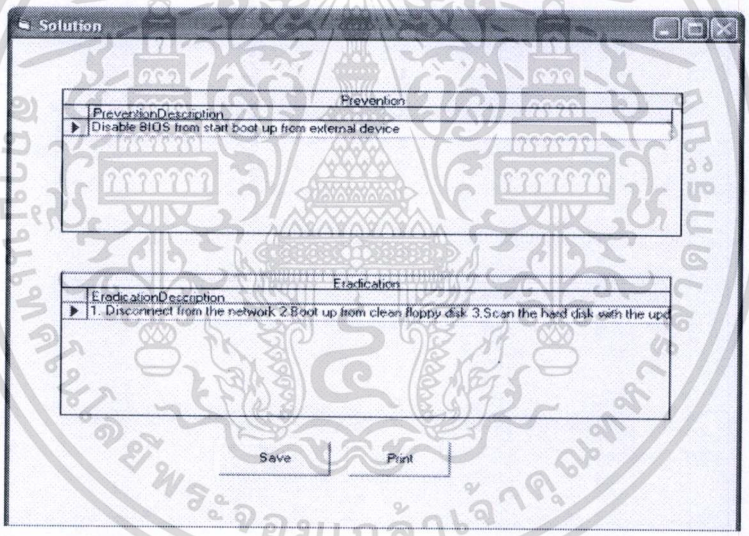
การทำงานของโปรแกรมดังกล่าวนี้จะทำการรับค่ารายละเอียดเกี่ยวกับไวรัสที่ผู้ใช้คอมพิวเตอร์ประสบอยู่ เช่น อาการที่เกิดจากไวรัส ช่องทางการแพร่กระจาย ความเสียหายที่เกิดขึ้น ความร้ายแรง ประเภทของไวรัส เป็นต้น จากนั้นโปรแกรมจะนำรายละเอียดที่ผู้ใช้คอมพิวเตอร์กรอกผ่านอินเตอร์เฟสการทำงานของโปรแกรมไปประมวลแล้วแสดงผลลัพธ์ซึ่งเป็นคำอธิบายถึงวิธีการป้องกันและวิธีการแก้ไขไวรัสตัวดังกล่าว ซึ่งผู้ใช้สามารถนำคำอธิบายดังกล่าวนี้ไปแก้ไขปัญหาที่เกิดจากไวรัสในเบื้องต้นได้ด้วยตนเอง



รูปที่ 2.1 แผนผังการทำงานของระบบของงานวิจัย [1]



รูปที่ 2.2 แบบฟอร์มรับรายละเอียดของไวรัสจากผู้ใช้ของงานวิจัย [1]



รูปที่ 2.3 ผลลัพธ์ที่ได้จากการค้นหาข้อมูลไวรัสของงานวิจัย [1]

อย่างไรก็ตามในทางปฏิบัติแล้วผู้ใช้คอมพิวเตอร์ที่ได้รับผลกระทบจากไวรัสส่วนใหญ่มักจะขาดความรู้ความเข้าใจเกี่ยวกับพฤติกรรมการทำงานของไวรัส ความเสียหาย ประเภทของไวรัส รวมถึงรายละเอียดอื่น ๆ ของไวรัสที่ตนกำลังได้รับผลกระทบอยู่ จากสาเหตุนี้เองจึงทำให้มีผู้ใช้จำนวนมากไม่สามารถใช้งานโปรแกรมนี้ได้อย่างเต็มประสิทธิภาพ เพราะผู้ใช้คอมพิวเตอร์นั้นไม่สามารถกรอกข้อมูลรายละเอียดของไวรัสได้ครบถ้วนเพียงพอต่อการทำงานของระบบ รวมถึงระบบที่ออกแบบมานั้นไม่ได้รองรับการอ้างอิงถึงชื่อไวรัสแต่อย่างไร จึงเกิดความไม่สะดวกในการใช้งานขึ้นเมื่อผู้ใช้คอมพิวเตอร์นั้นทราบเพียงชื่อของไวรัสเพียงอย่างเดียวแต่ไม่ทราบถึงรายละเอียดและการทำงานของไวรัส

นอกจากนั้นโลกอินเทอร์เน็ตในปัจจุบันยังมีการแพร่ระบาดของมัลแวร์ทั้งสามประเภท คือ ไวรัส หนอนคอมพิวเตอร์ และโทรจัน ซึ่งการออกแบบของงานวิจัยดังกล่าวครอบคลุมในส่วนของไวรัสเพียงอย่างเดียวเท่านั้น และข้อมูลสำหรับอธิบายวิธีการแก้ไขและวิธีการป้องกันไวรัสนั้นยังต้องอาศัยบุคคลากรในการกรอกข้อมูลของไวรัสลงในฐานข้อมูลด้วยตนเองซึ่งเป็นกระบวนการที่ใช้เวลาค่อนข้างมากและอาจเกิดข้อผิดพลาดได้ง่าย

เพื่อเป็นการสร้างประโยชน์ในด้านการศึกษาข้อมูลเกี่ยวกับมัลแวร์ ผู้เขียนจึงได้ออกแบบโครงสร้างฐานข้อมูลที่สามารถรองรับข้อมูลของมัลแวร์ทุกชนิดได้ และยังได้ออกแบบโปรแกรมที่ทำหน้าที่รวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาเก็บลงฐานความรู้เดียวกันอีกทั้งการออกแบบหมายเลขประจำตัวร่วมมัลแวร์ก็จะสามารถช่วยแก้ปัญหาในการอ้างอิงถึงมัลแวร์ที่มีชื่อเรียกแตกต่างกันจากหลายแหล่งข้อมูล

2.2 มัลแวร์ (Malware) [3]

คำว่ามัลแวร์ (Malware) เป็นคำที่ย่อมาจาก Malicious Software ซึ่งหมายถึงซอฟต์แวร์ชนิดหนึ่งที่ถูกออกแบบมาด้วยเจตนาที่ประสงค์ร้ายต่อผู้ใช้ ยกตัวอย่างเช่น การแฝงตนเองเข้ามาในเครื่องคอมพิวเตอร์ของผู้ใช้เพื่อขโมยข้อมูลส่วนตัวของผู้ใช้ หรือสร้างความเสียหายให้กับโปรแกรมหรือข้อมูลของผู้ใช้คอมพิวเตอร์เป็นต้น โดยมัลแวร์สามารถแบ่งออกได้เป็น 3 ประเภท ได้แก่ ไวรัส (Virus) หนอนคอมพิวเตอร์ (Computer worm) และโทรจัน (Trojan horse)

2.2.1 ไวรัส (Virus) [4][5][6]

เป็นโปรแกรมคอมพิวเตอร์ที่ถูกออกแบบมาให้บันทึกอยู่ในโฮสต์ไฟล์ (Host file) จำพวกไฟล์กระทำการ (Executable file) โดยไวรัสไม่สามารถอยู่อย่างโดดเดี่ยวได้ (Standalone) จะต้องอยู่บนโฮสต์ไฟล์เสมอ ไวรัสสามารถแพร่กระจายไปยังไฟล์เป้าหมายที่อยู่ในเครื่องคอมพิวเตอร์ที่ถูกติดตั้งไวรัสไปแล้วได้ด้วยตัวของมันเอง ซึ่งข้อจำกัดของไวรัสก็คือมันไม่สามารถแพร่กระจายไปนอกเครื่องคอมพิวเตอร์ของผู้ใช้ได้เว้นเสียแต่ว่าผู้ใช้ทำตัวเป็นพาหะเอง เช่น ผู้ใช้ส่งอีเมลหรือส่งแผ่นซีดีรอมที่มีไวรัสให้กับผู้ใช้คอมพิวเตอร์คนอื่น ๆ

ในการทำงานนั้นไวรัสจะทำการคัดลอกโค้ดของตัวมันเองไปไว้ในโฮสต์ไฟล์เป้าหมาย โดยเมื่อระบบปฏิบัติการหรือผู้ใช้ทำการเรียกใช้โปรแกรมที่ติดไวรัสนั้นก็จะเป็นการสั่งให้โค้ดของไวรัสที่แฝงเอาไว้ในโปรแกรมดังกล่าวทำงานด้วย ซึ่งก็ขึ้นอยู่กับว่าไวรัสแต่ละตัวถูกออกแบบมาให้ทำอย่างไรกับเครื่องเป้าหมายบ้าง

2.2.1.1 ตำแหน่งและชนิดของไฟล์ที่ไวรัสอาศัยอยู่

เนื่องจากไวรัสนั้นต้องอาศัยอยู่บนโฮสต์เสมอ ดังนั้นสามารถแบ่งไวรัสออกได้ตามตำแหน่งและชนิดของไฟล์ที่มันอาศัยอยู่ ได้ดังนี้

- ไวรัสที่อาศัยอยู่ในไฟล์คำสั่ง (File Type) ได้แก่ ไฟล์กระทำการอยู่ในระบบปฏิบัติการในเครื่องของผู้ใช้หรือไฟล์ปฏิบัติการของโปรแกรมต่าง ๆ ตัวอย่างเช่น .EXE .COM เป็นต้น ตัวอย่างของไวรัสประเภทนี้ได้แก่ Jerusalem virus, Cascade virus เป็นต้น ไวรัสประเภทไฟล์รวมถึงไฟล์ที่เป็นสคริปต่าง ๆ ด้วย เช่นไฟล์ .BAT หรือ.VBS เป็นต้น ตัวอย่างของไวรัสประเภทนี้ได้แก่ Virus.VBS.Infi หรือ BAT.Looper เป็นต้น

ซึ่งไวรัสจะใช้วิธีการที่แตกต่างกันในการแพร่กระจายตนเองลงไปในไฟล์ต่าง ๆ เหล่านี้ เช่น วิธีการเขียนโค้ดลงในไฟล์เป้าหมายในส่วนต้นไฟล์ (Prepending Technique) การเขียนไฟล์เป้าหมายในส่วนท้ายไฟล์ (Appending Technique) หรือการเขียนโค้ดลงบนไฟล์เป้าหมายแบบเขียนทับ (Overwriting Technique) เป็นต้น

-ไวรัสที่อาศัยอยู่ในตำแหน่งสำคัญบนหน่วยความจำสำรอง (Disk Type) ไวรัสประเภทนี้จะเขียนโค้ดของตนเองลงในส่วนโปรแกรมเริ่มต้นการทำงานของแผ่นดิสก์เก็ต (Boot Sector) หรือส่วนโปรแกรมเริ่มต้นการทำงานหลักของฮาร์ดดิสก์ (Master Boot Record) ไวรัสประเภทนี้สามารถแพร่กระจายได้โดยการที่ผู้ใช้นำแผ่นดิสก์เก็ตที่ติดไวรัสนั้นไปใส่ในเครื่องของผู้อื่นแล้วทำการบู๊ตเครื่องคอมพิวเตอร์ด้วยดิสก์ที่มีไวรัสอยู่ โดยปรกติแล้วเมื่อเครื่องคอมพิวเตอร์เริ่มการทำงานมันจะค้นหาโปรแกรมเริ่มต้นการทำงานจากแผ่นดิสก์เก็ตหรือฮาร์ดดิสก์ หลังจากนั้นเครื่องจะโหลดคำสั่งการทำงานที่ได้เก็บไว้ในหน่วยความจำของระบบ ซึ่งถ้าแผ่นดิสก์เก็ตมีไวรัสอาศัยอยู่ไวรัสก็จะถูกโหลดเข้าไปไว้ในหน่วยความจำของระบบด้วยเช่นกัน ต่อจากนั้นไวรัสก็จะเริ่มปฏิบัติการตามคำสั่งที่ผู้สร้างไวรัสได้กำหนดเอาไว้ ตัวอย่างของไวรัสประเภทนี้ได้แก่ Stone, Michelangelo เป็นต้น

-ไวรัสที่อาศัยอยู่ในไฟล์เอกสาร (Document Type) [7][8] ไวรัสประเภทนี้จะอาศัยอยู่ในชุดคำสั่งมาโคร (Macro) ซึ่งมีอยู่บนไฟล์เอกสารของผู้ใช้ โดยจะพบเห็นมากในไฟล์เอกสารของโปรแกรมไมโครซอฟท์ออฟฟิศ เช่น ไมโครซอฟท์เวิร์ด (Microsoft Word) ไมโครซอฟท์เอ็กเซล (Microsoft Excel) ไมโครซอฟท์เพาเวอร์พอยต์ (Microsoft PowerPoint) เป็นต้น

คำสั่งต่าง ๆ ที่ใช้ในโปรแกรมไมโครซอฟท์ออฟฟิศนั้นล้วนแล้วแต่เรียกชุดคำสั่งมาโครมาใช้ทั้งสิ้นโดยโปรแกรมไมโครซอฟท์ออฟฟิศจะทำการเรียกชุดคำสั่งมาโครที่มีอยู่ในตัวให้เหมาะกับคำสั่งนั้น ๆ ยกตัวอย่างเช่น การใช้คำสั่งไฟล์/เซฟ (File/Save) จะทำการเรียก ชุดคำสั่งไฟล์เซฟมาโคร หรือการใช้คำสั่งไฟล์เซฟ/แอส (File/SaveAs) โปรแกรมก็จะทำการเรียก ไฟล์/เซฟแอดมาโคร ออกมาทำงาน เป็นต้น

นอกจากชุดคำสั่งมาโครปรกติที่อธิบายมาแล้วนั้นยังมีชุดคำสั่งมาโครชนิดที่เรียกว่า มาโครอัตโนมัติ ซึ่งชุดคำสั่งมาโครชนิดนี้จะถูกเรียกใช้งานทันทีที่มีการเปิดไฟล์เอกสารขึ้นมา ยกตัวอย่างเช่น เมื่อผู้ใช้ทำการเปิดไฟล์เอกสารเวิร์ดขึ้นมาโปรแกรมไมโครซอฟท์เวิร์ดจะตรวจสอบดูว่าเอกสาร ดังกล่าวนั้นมีชุดคำสั่งมาโครอัตโนมัติอยู่ด้วยหรือไม่ ถ้ามีโปรแกรมก็ชุดคำสั่งมาโครเหล่านี้ขึ้นมาทำงานทันที โดยตัวอย่างชุดคำสั่งมาโครอัตโนมัติได้แก่ ชุดคำสั่งมาโครออโต้โอเพ่น (AutoOpen) ชุดคำสั่งมาโครออโต้โคลส (AutoClose) ชุดคำสั่งมาโครออโต้เอ็กซิคิว (AutoExec) เป็นต้น

ซึ่งมาโครไวรัสบางตัวหลังจากที่ถูกเรียกใช้งานแล้วจะทำการสำเนาตัวเองไปไว้ในส่วนคำสั่งมาโครหลัก (Global Macros) ของโปรแกรม ซึ่งเมื่อไรก็ตามที่มีการเรียกไฟล์เอกสารอื่นหรือเปิดไฟล์เอกสารใหม่ขึ้นมาชุดคำสั่งมาโครหลักที่ติดไวรัสไปแล้วนั้นก็จะทำการคัดลอกชุดคำสั่งมาโครที่มีไวรัสลงไปยังไฟล์เอกสารใหม่นั้น ๆ ซึ่งเป็นวิธีการแพร่กระจายของไวรัสประเภทนี้

- ไวรัสที่อาศัยอยู่บนไฟล์ได้หลายชนิด (Multipartite Type)

คือไวรัสที่สามารถแพร่กระจายตัวมันเองไปยังโฮสต์ไฟล์มากกว่าหนึ่งประเภท ยกตัวอย่างเช่น ไวรัสบางตัวสามารถแพร่กระจายได้ทั้งไฟล์ปฏิบัติการของโปรแกรมและโปรแกรมเริ่มต้นการทำงานของดิสก์ เป็นต้น

2.2.1.2 เทคนิคที่ไวรัสใช้ในการแพร่กระจาย (Infection Techniques)

นอกจากนั้นไวรัสแต่ละตัวยังมีเทคนิคที่ใช้ในการแพร่กระจาย (Infection Technique) ตัวมันเองไปยังไฟล์เป้าหมายด้วยวิธีที่แตกต่างกันดังนี้ [9]

- เทคนิคแบบเขียนต้นไฟล์ (Prepending Technique) โดยปรกติแล้วเมื่อไฟล์กระทำการจำพวก .COM หรือ .EXE ถูกเรียกใช้งาน คอมพิวเตอร์จะโหลดไฟล์นั้นเข้าไปเก็บไว้ในหน่วยความจำของเครื่อง จากนั้นคอมพิวเตอร์จะเริ่มอ่านคำสั่งจากจุดเริ่มต้นของไฟล์นั้นเป็นอันดับแรก ซึ่งไวรัสประเภทนี้จะเขียนโค้ดของตัวมันเองลงในตำแหน่งจุดเริ่มต้นของไฟล์เป้าหมาย เพื่อให้เครื่องคอมพิวเตอร์อ่านโค้ดของมันและนำไปประมวลผลก่อนคำสั่งโปรแกรมของไฟล์นั้น ๆ

- เทคนิคแบบเขียนท้ายไฟล์ (Appending Technique) เทคนิคนี้จะตรงกันข้ามกับการเขียนต้นไฟล์ กล่าวคือ ไวรัสจะเขียนโค้ดตัวมันเองลงในส่วนท้ายของไฟล์เป้าหมายเพื่อรอให้คอมพิวเตอร์เรียกไฟล์ที่ติดไวรัสนี้ไปเก็บไว้ในหน่วยความจำและรอให้เครื่องคอมพิวเตอร์อ่านคำสั่งของมัน แม้ว่าไวรัสจะเขียนโค้ดของมันลงในส่วนท้ายของไฟล์เป้าหมายแต่ตัวมันเองต้องทำการแก้ไข

ข้อมูลในส่วนแฮดเดอร์ (Header) ของไฟล์เป้าหมายด้วยเช่นกัน เพื่อเป็นการที่จะแจ้งให้คอมพิวเตอร์ทราบว่าต้องไปอ่านโค้ดของไวรัสที่ตำแหน่งใด

- เทคนิคแบบเขียนทับไฟล์ (Overwriting Technique) เทคนิคนี้เป็นวิธีการที่ไวรัสจะเขียนโค้ดของตัวมันเองทับไฟล์เป้าหมาย ซึ่งไวรัสจะต้องมองหาตำแหน่งในไฟล์เป้าหมายซึ่งสามารถเขียนโค้ดของตัวมันทับลงไปได้โดยที่ไฟล์เป้าหมายยังสามารถเรียกใช้งานได้ตามปรกติ ซึ่งไวรัสที่ใช้เทคนิคนี้มีโอกาสที่ผู้ใช้คอมพิวเตอร์จะพบได้ง่ายกว่าเทคนิคการแพร่กระจายแบบอื่น เนื่องจากว่าหากไวรัสเขียนทับลงไฟล์เป้าหมายในตำแหน่งที่ผิดพลาดจะทำให้ไฟล์ดังกล่าวไม่สามารถใช้งานได้ตามปรกติ ซึ่งผู้ใช้คอมพิวเตอร์จะสังเกตได้ง่ายเนื่องจากโปรแกรมหรือระบบปฏิบัติการที่ติดไวรัสจะหยุดทำงานไม่นานหลังจากติดไวรัส

- เทคนิคแบบแทรกไฟล์ (Inserting Technique) เทคนิคนี้เป็นวิธีการที่ไวรัสจะแทรกโค้ดของตัวมันเองไปในโค้ดของไฟล์เป้าหมาย โดยไวรัสจะต้องค้นหาตำแหน่งที่ว่างในไฟล์เป้าหมายซึ่งอาจจะเป็นพื้นที่ในส่วนเฮดเดอร์ของไฟล์เป้าหมาย หรือเป็นพื้นที่ว่างสำหรับเขียนคำอธิบายโค้ด (Text Area) เป็นต้น เมื่อพบแล้วไวรัสจะทำการแทรกโค้ดของมันลงไปในตำแหน่งที่ว่างดังกล่าว ซึ่งการแทรกโค้ดของมันลงไปนี้ต้องยังคงทำให้โปรแกรมเดิมสามารถทำงานต่อไปได้ด้วย เพื่อเป็นการป้องกันไม่ให้ผู้ใช้ทราบว่าไฟล์ดังกล่าวมีการติดไวรัสแต่อย่างใด ซึ่งวิธีการนี้ค่อนข้างยุ่งยากและซับซ้อนซึ่งมีโอกาสสูงที่จะทำให้ไฟล์เป้าหมายนั้นใช้การไม่ได้

2.2.1.3 เทคนิคการปกปิดตนเอง (Stealth Technique) [10][11]

นอกจากเทคนิคการแพร่กระจายไวรัสที่ได้กล่าวมาแล้ว ไวรัสบางชนิดนั้นสามารถป้องกัน หรือหลบเลี่ยงการตรวจจับจากโปรแกรมกำจัดมัลแวร์ได้โดยอาศัยเทคนิคดังต่อไปนี้เรียกวิธีการเหล่านี้ว่าเทคนิคการปกปิดตนเอง โดยเทคนิคเหล่านี้ได้แก่

- เทคนิคแบบมีหลายรูป (Polymorphic or Encrypted) ไวรัสประเภทนี้ตัวของมันเองจะประกอบด้วย 2 ส่วนด้วยกัน ส่วนแรกจะเป็นส่วนของโค้ดไวรัสที่เข้ารหัสไว้ (Encrypted virus body) และส่วนรูทีนการถอดรหัส (Decryption routine) ซึ่งเมื่อไรก็ตามที่ไฟล์ที่ติดไวรัสประเภทนี้ถูกเรียกใช้ ส่วนรูทีนการถอดรหัสจะถูกเรียกใช้งานทันที โดยรูทีนการถอดรหัสนี้จะถอดรหัสไวรัสที่ถูกเข้ารหัสเอาไว้เพื่อให้ได้ชุดคำสั่งของไวรัสที่พร้อมใช้งานออกมาเพื่อปฏิบัติการตามวัตถุประสงค์ที่ไวรัสตัวนั้นถูกสร้างขึ้นมา ซึ่งการแพร่กระจายตัวมันเองไปยังไฟล์อื่นในแต่ละครั้งนั้นไวรัสจะทำการเปลี่ยนคีย์ที่ใช้เข้ารหัสไปทุก

ครั้งที่มีการแพร่กระจาย แต่ส่วนที่ยังเหมือนเดิมคือส่วนรูทีนการถอดรหัสซึ่งวิธีนี้เป็นวิธีการที่ไวรัสใช้ในการหลบเลี่ยงโปรแกรมกำจัดมัลแวร์ได้ในระดับหนึ่ง

- เทคนิคแบบเปลี่ยนรูป (Metamorphic) ในกรณีของเทคนิคแบบมีหลายรูปนั้นจะเห็นได้ว่าส่วนที่ไม่ได้เปลี่ยนแปลงไปเลยในการแพร่กระจายแต่ละครั้งนั้นจะเป็นส่วนรูทีนการถอดรหัส ซึ่งจากจุดนี้เองทำให้โปรแกรมกำจัดมัลแวร์สามารถตรวจจับไวรัสที่ใช้เทคนิคดังกล่าวได้ ผู้เขียนไวรัสจึงพัฒนาเทคนิคชนิดใหม่ขึ้นมาเพื่อแก้ไขจุดอ่อนดังกล่าว โดยเพิ่มส่วนประกอบส่วนที่ 3 ขึ้นมานั่นก็คือ ส่วน กลไกการเปลี่ยนรูป (Mutation engine) ซึ่งทำหน้าที่สร้างรหัสอย่างสุ่มขึ้นมาให้กับรูทีนการถอดรหัส เพื่อให้การแพร่กระจายไวรัสแต่ละครั้งนั้นมีรูทีนถอดรหัสที่เปลี่ยนแปลงไปด้วย ซึ่งเมื่อไรที่ไฟล์นั้นถูกเรียกใช้งานรูทีนการถอดรหัสก็จะทำงานทันที ถึงแม้ว่าตัวมันจะถูกเปลี่ยนแปลงไปในทุก ๆ ครั้งที่มีการแพร่กระจายก็ตาม โดยมันจะทำหน้าที่ถอดรหัสส่วนที่ไวรัสได้ถูกเข้ารหัสไว้ เพื่อให้ได้ชุดคำสั่งที่พร้อมจะทำงานต่อไป เทคนิคแบบเปลี่ยนรูปนี้เป็นวิธีการที่ค่อนข้างจะซับซ้อนทำให้โปรแกรมตรวจจับไวรัสตรวจจับได้ยากลำบากขึ้น แต่เทคนิคชนิดนั้นก็นับว่าทำได้ยากเหมือนกันเนื่องจากหาก เอนจินการเปลี่ยนรูปทำงานผิดพลาดอาจทำให้ไม่สามารถถอดรหัสของไวรัสตัวนั้นเพื่อทำงานต่อไปได้ ส่งผลให้ไวรัสไม่สามารถแพร่กระจายไปยังไฟล์อื่นได้ตลอดจนอาจทำให้ไฟล์เป้าหมายใช้งานไม่ได้ด้วยเช่นกัน

2.2.2 หนอนคอมพิวเตอร์ (Computer Worms) [12]

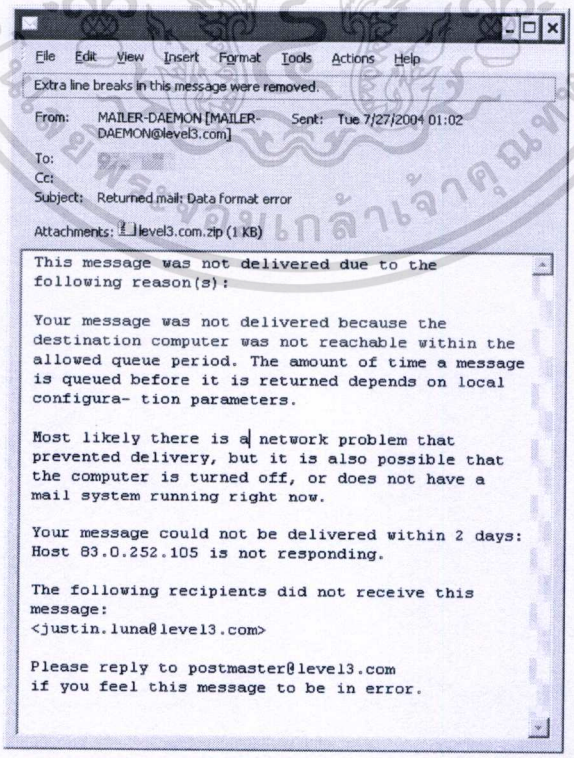
มีลักษณะที่ค่อนข้างจะคล้ายกับไวรัสซึ่งมีผู้ใช้บางส่วนยังสับสนระหว่างหนอนคอมพิวเตอร์กับไวรัสอยู่บ้าง กล่าวคือ หนอนคอมพิวเตอร์นั้นมีลักษะการแพร่กระจายด้วยตนเองได้เหมือนไวรัส แต่ในการแพร่กระจายของหนอนคอมพิวเตอร์นั้นไม่จำเป็นต้องมีโฮสต์ไฟล์เหมือนไวรัส หนอนคอมพิวเตอร์สามารถอยู่อย่างโดดเดี่ยวได้ (Standalone) และสามารถแพร่กระจายไปยังเครื่องคอมพิวเตอร์เครื่องอื่นในเครือข่ายหรือในอินเทอร์เน็ตได้ด้วยตัวของมันเองโดยสามารถแบ่งประเภทของหนอนคอมพิวเตอร์ตามการแพร่กระจาย (Distribution Type) ของมันได้ดังนี้

2.2.2.1 หนอนคอมพิวเตอร์ประเภทอีเมล (Email Type)

หนอนคอมพิวเตอร์ประเภทนี้จะใช้อีเมลเป็นช่องทางในการแพร่กระจายตนเองไปยังคอมพิวเตอร์เครื่องอื่น โดยมันจะทำสำเนาตัวมันเองแล้วส่งไฟล์ดังกล่าวแนบไปกับอีเมลเพื่อส่งไปยังผู้ใช้ที่มีรายชื่อในสมุดที่อยู่ (Address book)

ส่วนใหญ่แล้วมัลแวร์จะใช้กลลวงทางสังคม (Social engineering) เข้ามาเพิ่มโอกาสในการแพร่กระจายตนเอง หนอนคอมพิวเตอร์ประเภทนี้ก็เช่นกัน หนอนคอมพิวเตอร์บางตัวจะใช้หัวข้ออีเมล (Subject) กับข้อความในอีเมล (Messages) ซึ่งเป็นข้อความที่น่าสนใจหรือเป็นข้อความที่สร้างความกังวลให้กับผู้ใช้ที่รับอีเมลดังกล่าวอันเป็นสาเหตุให้ผู้รับต้องเปิดไฟล์นั้น

นอกเหนือจากนั้นหนอนคอมพิวเตอร์ได้ใช้รายชื่อในสมุดที่อยู่ของเหยื่อในการส่งอีเมลต่อไปยังผู้รับรายอื่น ซึ่งเมื่อผู้ใช้รายอื่นได้รับอีเมลดังกล่าวแล้วพบว่ารายชื่อผู้ส่งมาจากบุคคลที่ตนรู้จักหรือเป็นบุคคลที่อยู่ในสมุดรายชื่อของตนอยู่แล้วจึงไม่ระแวงสงสัยอีเมลดังกล่าว ซึ่งเป็นสาเหตุให้หนอนประเภทนี้แพร่กระจายได้อย่างรวดเร็ว ยกตัวอย่างเช่น ผู้ใช้ที่ตกเป็นเหยื่อของหนอนคอมพิวเตอร์ประเภทอีเมลชื่อ W32.Mydoom.BB@mm [13] จะได้รับอีเมลซึ่งมีหัวข้อดังต่อไปนี้ “The original message was included as attachment.”, “Returned mail: see transcript for details.”, “Delivery reports about your e-mail.” เป็นต้น รวมถึงข้อความที่หนอนคอมพิวเตอร์ส่งมา เช่น “Dear User of [recipient domain] We have received reports that your account was used to send a large amount of junk email messages during the week. Probably, your computer had been compromised and now contains a hidden proxy server. Please follow the instruction in the attached file in order to keep your computer safe. Have a nice day, [recipient domain] user support team.” และหนอนคอมพิวเตอร์ยังแนบไฟล์สำเนาของตนเองมากับอีเมลเหล่านี้ด้วยซึ่งไฟล์เหล่านี้มักจะมีนามสกุลดังต่อไปนี้ .BAT,.CMD,.COM,.EXE,.PIF,.SCR,.ZIP เป็นต้น ตัวอย่างของอีเมลที่ได้รับจากหนอนคอมพิวเตอร์ตัวนี้แสดงในรูปที่ 2.4



รูปที่ 2.4 ตัวอย่างอีเมลที่ได้รับจากหนอนคอมพิวเตอร์ W32.Mydoom.BB@mm

2.2.2.2 หนอนคอมพิวเตอร์ประเภทแพร่กระจายผ่านโปรแกรมส่งข้อความแบบทันทีทันใด (Instant Messaging Type)

ในปัจจุบันการสื่อสารด้วยโปรแกรมส่งข้อความแบบทันทีได้รับความนิยมเป็นอย่างมาก เช่น Windows Live Messenger, Yahoo Messenger, ICQ เป็นต้น เนื่องจากความสะดวกในการใช้งานของโปรแกรมเหล่านี้ ซึ่งสามาถส่งข้อความ เสียง ภาพ หรือไฟล์ต่างๆ ให้แก่กันได้อย่างสะดวกและรวดเร็ว นอกจากนั้นผู้ใช้โปรแกรมประเภทนี้ยังสามารถจัดการกับรายชื่อผู้สนทนา (Contact list) ได้อีกด้วย จากความสะดวกในการติดต่อสื่อสารนี้เองก็เป็นช่องทางหนึ่งที่หนอนคอมพิวเตอร์ใช้ประโยชน์ในการแพร่กระจายตนเองไปยังเครื่องคอมพิวเตอร์ของเหยื่อ

วิธีการแพร่กระจายของหนอนคอมพิวเตอร์ประเภทนี้จะทำได้หลายวิธี เช่น หนอนคอมพิวเตอร์อาจส่งไฟล์ไปให้เครื่องปลายทาง หรืออาจส่งเป็น URL ไปให้เครื่องปลายทางที่มีอยู่ในรายชื่อผู้สนทนาและกำลังออนไลน์อยู่ในขณะนั้น โดยข้อความที่ส่งไปพร้อมกับไฟล์หรือ URL นั้นมักจะเป็นข้อความที่ชักชวนให้ผู้ที่ได้รับเปิดขึ้นมาดู ยกตัวอย่างเช่น ในกรณีหนอนคอมพิวเตอร์ W32.MSN.Worm [14] นั้นจะส่งข้อความต่าง ๆ เหล่านี้มาพร้อมกับไฟล์ “Image.zip” ตัวอย่างของข้อความดังกล่าวได้แก่

- LOL, you look so ugly in this picture, no joke...
- Should I put this on facebook/myspace?
- Hey m8, who is this on the right, in this picture...
- Sup, seen the pictures from the other night?

เมื่อผู้ที่อยู่ปลายทางได้รับไฟล์ของหนอนและเรียกไฟล์นั้นขึ้นมาทำงานแล้ว หนอนคอมพิวเตอร์ก็จะทำการติดตั้งตัวเองลงในเครื่องผู้ใช้และค้นหารายชื่อผู้ใช้ที่ออนไลน์อยู่ในขณะนั้นเพื่อทำการแพร่กระจายตนเองต่อไป นอกจากนั้นหนอนคอมพิวเตอร์บางตัวจะส่งข้อความที่เป็น URL ให้กับบุคคลที่อยู่ในรายชื่อผู้สนทนาซึ่งกำลังออนไลน์อยู่ในขณะนั้น ซึ่ง URL ดังกล่าวจะมีหนอนคอมพิวเตอร์ซึ่งรอให้ผู้ใช้ดาวน์โหลดลงไปในเครื่องคอมพิวเตอร์ของตน ซึ่งหนอนคอมพิวเตอร์มักจะใช้ชื่อไฟล์และนามสกุลที่ไม่น่าสงสัย เช่น IMG455.jpg-www.photo.com ถ้าสังเกตให้ดีจะพบว่าไฟล์ดังกล่าวนี้ไม่ใช่ไฟล์ภาพแต่อย่างใด เนื่องจากนามสกุลของไฟล์ดังกล่าวนี้เป็น .COM ซึ่งเป็นไฟล์ปฏิบัติการ วิธีการดังกล่าวนี้เป็นวิธีการที่หนอนคอมพิวเตอร์ประเภทนี้นิยมใช้ ซึ่งจะทำให้ผู้ใช้ที่ไม่ได้สังเกตทำโหลดไฟล์ดังกล่าวมาไว้ในเครื่องคอมพิวเตอร์ของตนอันเป็นสาเหตุให้หนอนคอมพิวเตอร์เหล่านี้สามารถติดตั้งตนเองและแพร่กระจายต่อไปได้ ตัวอย่างของหนอนคอมพิวเตอร์ประเภทนี้ได้แก่ W32.IRCBot.AJY [15]

2.2.2.3 หนอนคอมพิวเตอร์ประเภทแพร่กระจายผ่านโปรโตคอลของเครือข่าย (Network Protocol Type)

หนอนคอมพิวเตอร์ประเภทนี้จะทำการแพร่กระจายตนเองไปในเครื่องข่ายโดยอัตโนมัติ โดยหนอนคอมพิวเตอร์เหล่านี้จะใช้วิธีในการหาเครื่องเป้าหมายแตกต่างกันออกไปซึ่งจะแบ่งเป็นประเภทต่าง ๆ ดังนี้ [16]

- การสแกนนิ่ง (Scanning) เป็นวิธีการที่หนอนคอมพิวเตอร์ใช้ในการค้นหาเครื่องเป้าหมายในเครือข่าย โดยหนอนจะค้นหาเครื่องเป้าหมายจากกลุ่มหมายเลขไอพีแอดเดรส (IP Address) ภายใต้หมายเลขไอพีของเครือข่ายนั้น โดยอาจจะใช้วิธีการสุ่มหมายเลขไอพีแอดเดรส หรือไล่เรียงลำดับหมายเลขไอพีแอดเดรสทีละลำดับ การสแกนหมายเลขไอพีนั้นเป็นวิธีการที่เสียเวลาค่อนข้างมากและยังสามารถถูกตรวจจับได้ง่าย เนื่องจากว่าระบบรักษาความปลอดภัยในเครือข่ายจะตรวจพบทราฟฟิก ในเครือข่ายที่มีปริมาณสูงขึ้นได้อย่างชัดเจน

- การใช้รายชื่อ (Lists) ผู้สร้างหนอนคอมพิวเตอร์บางคนได้ค้นหาหมายเลขไอพีแอดเดรสของคอมพิวเตอร์ที่มีช่องโหว่ (Vulnerable) เอาไว้แต่แรกแล้ว จากนั้นจึงนำหมายเลขเหล่านั้นมาใส่เป็นรายชื่อซึ่งบรรจุเอาไว้ในตัวหนอนคอมพิวเตอร์เพื่อกำหนดเป็นเป้าหมายในการแพร่กระจายต่อไป ซึ่งวิธีการแพร่กระจายแบบนี้จะทำได้ค่อนข้างรวดเร็วกว่าวิธีการสแกนนิ่งและยังทำให้ระบบรักษาความปลอดภัยในเครือข่ายตรวจพบได้ยากขึ้นด้วย

นอกจากนั้นหนอนคอมพิวเตอร์นั้นยังอาศัยโปรโตลคอลที่จะช่วยแพร่กระจายตนเองไปยังเครือข่ายเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ในระบบเครือข่ายที่แตกต่างกันออกไป ยกตัวอย่างเช่น การใช้โปรโตรคอล UDP (User Datagram Protocol) การใช้โปรโตคอล TCP (Transmission Control Protocol) และ P2P (Peer to Peer) ในการแพร่กระจายตนเอง เป็นต้น

2.2.2.4 หนอนคอมพิวเตอร์ประเภทแพร่กระจายผ่านอุปกรณ์เก็บข้อมูลแบบพกพา (Removable Storage Type)

เป็นหนอนคอมพิวเตอร์ที่พึ่งเกิดขึ้นเมื่อไม่นานมานี้ เนื่องจากความนิยมในการใช้อุปกรณ์เก็บข้อมูลแบบพกพาซึ่งสามารถโอนถ่ายข้อมูลจากเครื่องคอมพิวเตอร์มาเก็บไว้ยังตัวอุปกรณ์ได้ง่ายและรวดเร็วผ่านพอร์ต USB (Universal Serial Bus) ซึ่งหนอนคอมพิวเตอร์อาศัยช่องทางนี้ในการแพร่กระจายตนเองไปยังคอมพิวเตอร์เครื่องอื่น ยกตัวอย่างเช่น เมื่อผู้ใช้คอมพิวเตอร์นำอุปกรณ์เก็บข้อมูลแบบพกพาไปเชื่อมต่อเข้ากับเครื่องคอมพิวเตอร์ที่มีหนอนคอมพิวเตอร์ประเภทนี้อยู่ในเครื่อง หนอนคอมพิวเตอร์จะทำการแพร่กระจายตัวมันเองลงในอุปกรณ์ของผู้ใช้ทันทีรวมทั้งแก้ไขไฟล์ Autorun.inf ซึ่งเป็นไฟล์ทำงานอัตโนมัติในอุปกรณ์ของผู้ใช้ให้เรียกใช้งานหนอนตัวนี้ทุกครั้งที่ผู้ใช้นำอุปกรณ์ดังกล่าวไปเชื่อมต่อกับคอมพิวเตอร์เครื่อง

อื่น ซึ่งเมื่ออุปกรณ์ดังกล่าวได้เชื่อมต่อกับเครื่องคอมพิวเตอร์อื่น ๆ แล้ว หนอนคอมพิวเตอร์ก็จะทำการติดตั้งตนเองลงในเครื่องของเหยื่อโดยอัตโนมัติและเข้าสู่กระบวนการแพร่กระจายต่อไป ซึ่งจะเป็นวัฏจักรแบบนี้ต่อไปเรื่อย ๆ ตัวอย่างของหนอนคอมพิวเตอร์ประเภทนี้ได้แก่ VBS.Godzilla [17] VBS.Solow [18] เป็นต้น

2.2.3 โทรจัน (Trojan) หรือม้าโทรจัน (Trojan Horses) [19]

เป็นโปรแกรมที่ถูกออกแบบมาให้ทำหน้าที่เฉพาะทาง โดยโทรจันจะต่างกับไวรัสและหนอนคอมพิวเตอร์ตรงที่มันไม่สามารถแพร่กระจายตนเองไปยังไฟล์และเครื่องคอมพิวเตอร์เครื่องอื่นได้ รูปแบบการทำงานของโทรจันนั้นค่อนข้างจะชัดเจน กล่าวคือ โทรจันจะทำหน้าที่ขโมยข้อมูลข้อมูลส่วนตัวของผู้ใช้ เช่น บัญชีผู้ใช้และรหัสผ่านอีเมล แล้วส่งข้อมูลส่วนตัวเหล่านี้กลับไปยังผู้เขียนโปรแกรมโทรจันขึ้นมา หรือโทรจันอาจจะเข้ามาเปิดช่องทางให้ผู้อื่นสามารถควบคุมเครื่องผู้ใช้จากระยะไกลได้ แล้วสั่งให้เครื่องของผู้ใช้ที่ตกเป็นเหยื่อทำตามคำสั่งของผู้ที่สร้างโทรจันขึ้นมา โทรจันมักจะอยู่ในเวปไซต์ที่อันตรายรอโอกาสให้ผู้ใช้คอมพิวเตอร์ที่รู้เท่าไม่ถึงการณ์ดาวน์โหลดไฟล์โทรจันเข้ามาในเครื่องของตนเอง โดยสามารถแบ่งประเภทโทรจันออกตามการทำงานได้ดังนี้

2.2.3.1 โทรจันประเภทขโมยข้อมูลส่วนตัว (Personal Information Stealing Type)

เป็นโทรจันที่ถูกออกแบบมาเพื่อขโมยข้อมูลส่วนตัวของผู้ใช้คอมพิวเตอร์ไม่ว่าจะเป็น บัญชีผู้ใช้ (Username) และรหัสผ่าน (Password) ของอีเมล ตลอดจนบัญชีผู้ใช้และรหัสผ่านสำหรับทำธุรกรรมทางการเงิน เป็นต้น ตัวอย่างของโทรจันประเภทนี้มีดังต่อไปนี้

-โทรจันประเภทคีย์ล็อกเกอร์ (Keylogger Trojan) โดยปรกติแล้วโทรจันจะสามารถเข้ามาในเครื่องคอมพิวเตอร์ของผู้ใช้ได้โดยผ่านการติดตั้งจากผู้ใช้เองเนื่องจากความรู้เท่าไม่ถึงการณ์ ซึ่งโทรจันนั้นอาจแฝงมากับโปรแกรมที่สามารถดาวน์โหลดได้ฟรีหรือมากับโปรแกรมเกมส์คอมพิวเตอร์ที่มีอยู่บนอินเทอร์เน็ตก็ได้ หลังจากที่โทรจันได้ถูกติดตั้งแล้วมันจะทำการบันทึกข้อมูลจากแป้นพิมพ์ที่ผู้ใช้พิมพ์งานในแต่ละวัน ซึ่งโทรจันจะส่งข้อมูลเหล่านี้ไปยังผู้สร้างโทรจันที่ผ่านเครือข่ายอินเทอร์เน็ต

- โทรจันประเภทสปายแวร์ (Spyware Trojan) [20] จะคล้ายกับโทรจันประเภทคีย์ล็อกเกอร์ แต่โทรจันชนิดนี้จะเก็บรายละเอียดสถิติการเข้าชมเวปไซต์ติดตามเวปไซต์ต่าง ๆ ที่ผู้ใช้เข้าเยี่ยมชมด้วย ซึ่งรายละเอียดเหล่านี้จะถูกส่งไปยังผู้สร้างโทรจันที่เช่นกัน

2.2.3.2 โทรจันประเภทควบคุมผ่านทางไกล (Remote Access Type) เป็นโทรจันถูกออกแบบมาให้สามารถควบคุมเครื่องผู้ใช้ที่ตกเป็นเหยื่อทั้งทางตรงและทางอ้อม เพื่ออาศัยเครื่องคอมพิวเตอร์ของผู้ใช้ที่ตกเป็นเหยื่อเป็นช่องทางในการสร้างผลประโยชน์ให้กับผู้ที่สร้างโทรจันดังกล่าวขึ้นมา โดยโทรจันประเภทนี้จะแบ่งออกเป็น 4 ประเภทดังนี้

- โทรจันประเภทแบ็คดอร์ (Backdoor Trojan) เป็นโทรจันที่ค่อนข้างอันตราย เนื่องจากโทรจันประเภทนี้จะสร้างช่องทางการติดต่ออย่างลับ ๆ ให้กับผู้ที่เขียนโทรจัน ซึ่งส่งผลให้ผู้เขียนโทรจันสามารถควบคุมเครื่องคอมพิวเตอร์ของเหยื่อได้โดยที่ผู้ใช้ไม่รู้ตัว

- โทรจันประเภทดาวน์โหลดเดอร์ (Downloader Trojan) เป็นโปรแกรมโทรจันที่แฝงตัวมาในเครื่องของผู้ใช้แล้วสั่งให้ดาวน์โหลดไฟล์ที่มีอันตรายต่อเครื่องคอมพิวเตอร์ของผู้ใช้จากเซิร์ฟเวอร์หรือเวปไซต์ที่มีมัลแวร์อยู่ เมื่อมัลแวร์เหล่านั้นถูกดาวน์โหลดเข้ามาในเครื่องคอมพิวเตอร์ของผู้ใช้แล้วก็มักจะสร้างความเสียหายให้กับเครื่องคอมพิวเตอร์ของผู้ใช้ด้วย

- โทรจันประเภทคลิ๊กเกอร์ (Clicker Trojan) โทรจันประเภทนี้จะแฝงตัวเข้ามาในเครื่องผู้ใช้แล้วสั่งให้โปรแกรมเวปบราวเซอร์ (web browser) ของผู้ใช้เปิดเวปไซต์ต่าง ๆ ที่โทรจันต้องการด้วยเหตุผลหลายอย่าง เช่น การเพิ่มยอดการคลิ๊กเวปไซต์เพื่อผลประโยชน์ทางการเงิน การอาศัยเครื่องผู้ใช้เป็นเครื่องมือในการโจมตีแบบ (Denial-of-service: DoS) ไปยังเวปไซต์ต่าง ๆ เป็นต้น

- โทรจันประเภทพร็อกซี่ (Proxies Trojan) เป็นโทรจันที่ทำให้เครื่องคอมพิวเตอร์ของผู้ใช้กลายเป็นพร็อกซี่เซิร์ฟเวอร์เพื่อให้บุคคลอื่นสามารถเข้าสู่เครือข่ายอินเทอร์เน็ตโดยผ่านเครื่องของผู้ใช้

2.3 พฤติกรรมของมัลแวร์ (Malware Behavior)

แม้ว่าในปัจจุบันอินเทอร์เน็ตจะมีมัลแวร์ประเภทต่าง ๆ เป็นจำนวนมาก มัลแวร์เหล่านี้มักจะมีขั้นตอนการทำงานซึ่งเป็นพฤติกรรมหลัก ๆ ที่เหมือนกันอยู่เสมอ โดยสามารถแบ่งขั้นตอนการทำงานของมัลแวร์เหล่านี้ได้ดังต่อไปนี้

2.3.1 ขั้นตอนการติดตั้ง (Installation Phase) เป็นโค้ดส่วนที่ทำหน้าที่ให้มัลแวร์ถูกติดตั้งในเครื่องคอมพิวเตอร์ของผู้ใช้ ซึ่งขั้นตอนการติดตั้งนี้อาจจะเกิดขึ้นโดยอัตโนมัติเมื่อมัลแวร์ได้เข้ามาอยู่ในเครื่องของผู้ใช้ทันทีเลย หรือจะเริ่มติดตั้งเมื่อผู้ใช้เปิดไฟล์ที่มีโค้ดของมัลแวร์ขึ้นมาใช้งานก็ได้ ในปัจจุบันมัลแวร์ส่วนใหญ่มักจะติดตั้งตนเองลงในโฟลเดอร์ของระบบปฏิบัติการและปรับแก้รีจิสทรีของระบบปฏิบัติการเพื่อให้มัลแวร์เหล่านี้สามารถทำงานโดยอัตโนมัติทันทีที่เครื่องคอมพิวเตอร์ของผู้ใช้ได้ถูกเปิดขึ้นมา

2.3.2 ขั้นตอนการแพร่กระจาย (Infection & Propagation Phase) เป็นขั้นตอนหลังจากที่มัลแวร์ได้ติดตั้งตัวมันลงในเครื่องคอมพิวเตอร์ของผู้ใช้แล้ว มัลแวร์ก็จะเริ่มทำการแพร่กระจายตัวมัน ตามที่ผู้เขียนมัลแวร์ได้โปรแกรมเอาไว้ ถ้าเป็นกรณีของไวรัสมันจะแพร่โค้ดของมันลงไปบนโฮสต์ไฟล์เป้าหมายในเครื่องเดียวกัน ส่วนหนอนคอมพิวเตอร์นั้นจะทำการสำเนาตนเองไปยังคอมพิวเตอร์เครื่องอื่น ๆ ผ่านทางอีเมล หรือโปรโตรคอลเครือข่ายอื่น ๆ

2.3.3 ขั้นตอนการสร้างความเสียหาย (Damaged Phase) เป็นขั้นตอนการสร้างความเสียหายให้กับผู้ใช้ โดยทั้งนี้อาจเกิดขึ้นตามวัตถุประสงค์และความตั้งใจของผู้สร้างมัลแวร์ หรือแม้กระทั่งเกิดความเสียหายในขณะที่อยู่ในขั้นตอนของการแพร่กระจายของมัลแวร์ ที่กินทรัพยากรของเครื่องคอมพิวเตอร์ที่ติดมัลแวร์ และทรัพยากรของระบบเครือข่าย แต่ไม่ว่าจะเป็นด้วยสาเหตุใดก็ตาม ล้วนแต่นำมาซึ่งความเสียหายที่เกิดขึ้นกับผู้ใช้ทั้งสิ้น

2.4 การจำแนกประเภทของความเสียหายที่เกิดขึ้นจากมัลแวร์ (Classified Damages)

2.4.1 ความหมายของความเสียหาย

ในงานวิจัยนี้จะนิยามความเสียหายที่เกิดขึ้นกับผู้ใช้ไม่ว่าโดยความเสียหายนั้นมาจากความต้องการของผู้สร้างมัลแวร์ก็ดี หรือมาจากผลข้างเคียงของมัลแวร์ก็ดี ถือเป็นความเสียหายต่อผู้ใช้คอมพิวเตอร์ทั้งสิ้น ยกตัวอย่างเช่น มัลแวร์บางตัวไม่ได้สร้างความเสียหายให้กับผู้ใช้โดยตรง แต่การที่มัลแวร์ได้อาศัยอยู่ในเครื่องของผู้ใช้นั้น ก็จะต้องมีการใช้ทรัพยากรของเครื่องคอมพิวเตอร์ เช่น พื้นที่ดิสก์ หน่วยความจำ และยังต้องทำให้หน่วยประมวลผลกลาง (CPU) ทำงานหนักขึ้นกว่าเดิม ซึ่งเป็นภาระที่ผู้ใช้ไม่ควรจะต้องแบกไว้ จึงนับได้ว่าการคงอยู่ของมัลแวร์ในเครื่องคอมพิวเตอร์นั้นก็ถือเป็นความเสียหายอย่างหนึ่งด้วย โดยสามารถแบ่งความเสียหายหลักๆ ได้ดังนี้

2.4.2 ประเภทของความเสียหาย

- ความเสียหายด้านข้อมูล (Data Damaged)

เป็นความเสียหายที่เกิดขึ้นกับไฟล์เอกสารของผู้ใช้งาน หรือไฟล์งานที่ผู้ใช้บันทึกเก็บเอาไว้ ซึ่งอาจเกิดมากจากการที่มัลแวร์นั้นได้ทำการ ลบ หรือแก้ไขข้อมูลดังกล่าว โดยความเสียหายในลักษณะนี้มักจะส่งผลให้ผู้ใช้ไม่สามารถเรียกใช้งานไฟล์เอกสาร หรือไฟล์งานของผู้ใช้ได้ตามปรกติ

- ความเสียหายด้านข้อมูลส่วนตัว (Private Information Damaged)

เป็นความเสียหายที่เกิดขึ้นจากการที่ผู้ใช้ถูกขโมยข้อมูลส่วนตัวซึ่งอาจจะเป็นบัญชีผู้ใช้และรหัสผ่านของอีเมล หรือบัญชีผู้ใช้และรหัสผ่านสำหรับทำธุรกรรมทางการเงินของผู้ใช้ ซึ่งสร้างความเสียหายให้กับผู้ใช้เมื่อถูกขโมยข้อมูลเหล่านี้ไป

- ความเสียหายด้านโปรแกรม (Program Damaged)

เป็นความเสียหายที่มัลแวร์ได้ทำการแก้ไข หรือลบไฟล์โปรแกรมบนเครื่องคอมพิวเตอร์ของผู้ใช้ ซึ่งอาจทำให้โปรแกรมต่างๆ ไม่สามารถทำงานได้ตามปรกติ

- ความเสียหายด้านทรัพยากรเครือข่าย (Network Resource Damaged)

เป็นความเสียหายที่เกิดจากการใช้ทรัพยากรในเครือข่ายซึ่งอาจจะเป็นความจงใจของมัลแวร์เองในการโจมตีระบบเครือข่าย หรือเพื่อสนันสนุนการทำงานของมัลแวร์ ซึ่งมีทั้งการแพร่กระจายตนเองผ่านระบบเครือข่าย รวมถึงการโจมตีแบบ DoS (Denial of Services) ซึ่งจะต้องทำให้สิ้นเปลืองแบนด์วิธ (Bandwidth) จำนวนหนึ่งไปกับการทำงานของมัลแวร์

- ความเสียหายด้านไฟล์ของระบบ (System File Damaged)

เป็นความเสียหายที่เกิดขึ้นกับไฟล์ของระบบปฏิบัติการในเครื่องคอมพิวเตอร์ ซึ่งมัลแวร์ได้ทำการลบ หรือแก้ไขไฟล์ของระบบ ซึ่งส่งผลให้ระบบไม่สามารถทำงานได้ หรืออาจแก้ไขไฟล์ดังกล่าวนั้นเพื่ออำนวยความสะดวกให้กับการทำงานของตัวมัลแวร์เอง

- ความเสียหายด้านความปลอดภัย (Security Damaged)

เป็นความเสียหายที่เกิดจากเจตนาของมัลแวร์ที่จะทำให้เครื่องของผู้ใช้มีความเสี่ยงด้านความปลอดภัย ยกตัวอย่างเช่น มัลแวร์ทำการลบไฟล์สำคัญของโปรแกรมกำจัดมัลแวร์ที่ออกไปจากเครื่องของผู้ใช้เพื่อไม่ให้ผู้ใช้สามารถตรวจพบมัลแวร์ได้ หรือมัลแวร์ทำการปรับเปลี่ยนแก้ไขค่าบางอย่างในไฟล์โปรแกรม ซึ่งทำให้โปรแกรมนั้น ๆ อนุญาตให้มัลแวร์ดังกล่าวทำงานได้ เป็นต้น

- ความเสียหายด้านอินเตอร์เฟส (Interface Damaged)

เป็นความเสียหายที่เกิดขึ้นกับผู้ใช้โดยผ่านอุปกรณ์อินเตอร์เฟสต่าง ๆ เช่น จอแสดงผล เมาส์ คีย์บอร์ด เป็นต้น ยกตัวอย่างเช่น มัลแวร์บางตัวจะทำการเปลี่ยนรูปแบบอักษรบนจอแสดงผล หรือทำให้คีย์บอร์ดไม่สามารถพิมพ์ได้ตามปรกติ

2.5 แบบจำลองแนวความคิดไนแอม (NIAM Model) [21][22]

ไนแอม (NIAM) นั่นย่อมาจาก (Nijssen's Information Analysis Methodology) ซึ่งเป็นแบบจำลองข้อมูลในระดับแนวความคิด (Conceptual Schema Design Procedure) ที่ถูกคิดค้นขึ้นโดย ศาสตราจารย์ G.M. Nijssen จากมหาวิทยาลัยควีนแลนด์ (University of Queensland) และ E.D. Falkenberg (Katholieke Universiteit, Nijmegen) มาตั้งแต่ปี ค.ศ 1977 โดยเป็นการพัฒนาแนวความคิดขึ้นมาจากประโยคภาษาธรรมชาติ แบบจำลองไนแอมนี้ได้มีการแทนสิ่งต่าง ๆ ด้วยออปเจ็กต์ (Objects) อาจเป็นได้ทั้งเอนติตี้หรือข้อมูล และบทบาทต่าง ๆ (Roles) ซึ่งจะทำให้การออกแบบนั้นง่ายขึ้น เนื่องจากเป็นการออกแบบโดยใช้ประโยคหรือบทบาทที่เกิดขึ้นจริงมาแทน

ค่าเพื่อทดสอบความถูกต้องหลาย ๆ ค่าพร้อมกันได้ และไนแอมยังสร้างความเป็นมาตรฐานเดียวกัน เนื่องจากมีกฎข้อบังคับในความสัมพันธ์แน่นอนและชัดเจนซึ่งง่ายต่อการทำความเข้าใจ

2.5.1 ขั้นตอนการสร้างแบบจำลองไนแอม

ในการสร้างแบบจำลองไนแอมเพื่อนำไปใช้งานนั้นสามารถสรุปเป็นขั้นตอนการออกแบบได้ 7 ขั้นตอนดังนี้

- ขั้นตอนที่ 1 แปลงตัวอย่างข้อมูลที่มีให้อยู่ในรูปของข้อเท็จจริงพื้นฐาน

ขั้นตอนแรกนี้นับว่ามีความสำคัญมาก เนื่องจากผู้ออกแบบต้องทำการรวบรวมข้อมูลที่เกี่ยวข้องมาวิเคราะห์และแปลงข้อมูลเหล่านั้นเป็นประโยคภาษาธรรมชาติ โดยตัวอย่างของข้อมูลเหล่านี้นั้นอาจจะวิเคราะห์จากฟอร์มรายงาน หรือแบบฟอร์มสำหรับกรอกข้อมูลหรือเอกสารคำสั่งอื่น ๆ เมื่อรวบรวมข้อมูลได้ครบถ้วนแล้วจึงแปลงข้อมูลเหล่านั้นให้อยู่ในรูปของข้อเท็จจริงพื้นฐาน (Elementary Facts) ซึ่งเป็นประโยคที่ไม่สามารถแยกย่อยลงไปได้อีก ตัวอย่างของข้อเท็จจริงพื้นฐานนั้นได้แก่

ประโยคที่ 1. Tutorial group A “meets” at Time Mon 3 p.m.
ประโยคที่ 2. Tutorial group A “is held in” Room CS-718.
ประโยคที่ 3. Student 302156 “belongs to” Tutorial group A.
ประโยคที่ 4. Student 302156 “has” Name Bloggs FB.

โดยข้อเท็จจริงพื้นฐานนั้นสามารถแบ่งออกได้เป็น 2 ลักษณะดังนี้คือ

- ข้อเท็จจริงประเภทความสัมพันธ์ของออปเจ็กต์เดียว (Unary Fact Type) เป็นลักษณะของข้อเท็จจริงพื้นฐานที่เรียบง่าย มีเพียงออปเจ็กต์เดียวและบทบาทเดียว เช่น Person Ann “smokes” ซึ่งออปเจ็กต์ คือบุคคลที่มีชื่อว่า “Ann” และมีบทบาท “Smokes” เป็นต้น

- ข้อเท็จจริงประเภทความสัมพันธ์หลายออปเจ็กต์ (N-ary Fact Type) เป็นลักษณะของชนิดข้อเท็จจริงพื้นฐานที่เป็นความสัมพันธ์ระหว่างออปเจ็กต์ที่เหมือนกันมากกว่าหนึ่งออปเจ็กต์ขึ้นไปและมีความสัมพันธ์กันโดยมีบทบาทคั่นระหว่างออปเจ็กต์ทั้งสอง ยกตัวอย่างเช่น Person Ann “employs” Person “Bob” จะเห็นได้ว่า “Ann” มีบทบาทเป็นนายจ้างและ “Bob” มีบทบาทเป็นลูกจ้าง เป็นต้นซึ่งจากขั้นตอนข้างต้นทำให้สามารถแยกแยะระหว่างเอนติตี้และบทบาทได้

- ขั้นตอนที่ 2 วาดรูปเค้าร่างของแบบจำลองข้อมูลในระดับแนวความคิดและทดสอบโดยการใส่ตัวอย่างข้อมูลที่ใช้จริงลงไป

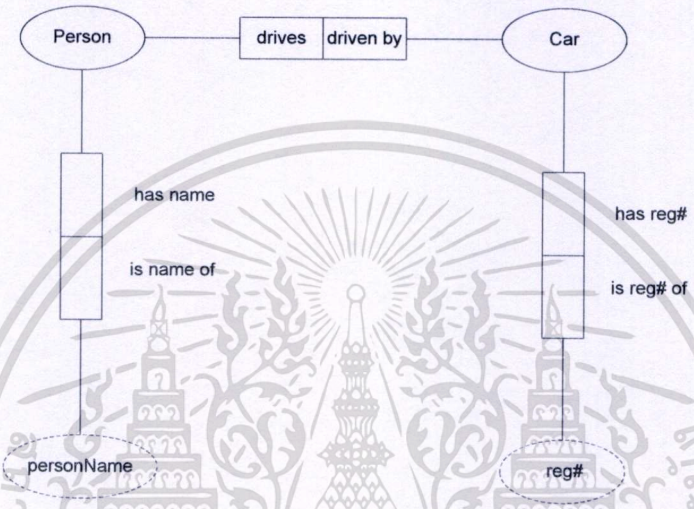
ส่วนสำคัญในขั้นตอนนี้ก็คือการวาดไดอะแกรมซึ่งแสดงข้อเท็จจริงพื้นฐานทั้งหมดที่มี โดยต้องวาดไดอะแกรมเพื่อแสดงเอนติตี้ เลเบลและบทบาทต่างๆ ของระบบ

ออกมาให้ครบถ้วน และจากข้อเท็จจริง 3 ประโยคข้างล่างนี้สามารถนำมาสร้างเป็นแบบจำลองไนแอมได้ดังแสดงในรูปที่ 2.5

ประโยค ที่ 1. The Person with name Adams B “drives” the Car with reg# 235PZN.

ประโยค ที่ 2. The Person with name Jones E “drives” the Car with reg# 235PZN.

ประโยค ที่ 3. The Person with name Jones E “drives” the Car with reg# 108AAQ.

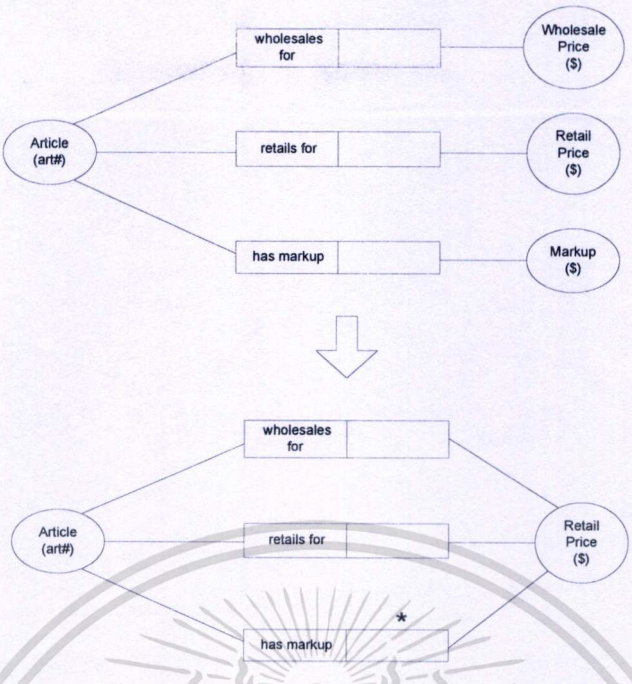


รูปที่ 2.5 ตัวอย่างแบบจำลองไนแอมที่ได้จากข้อเท็จจริงพื้นฐาน

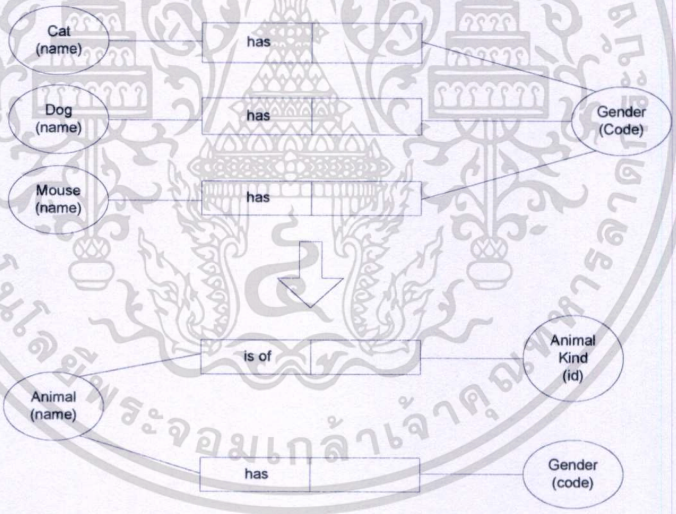
ในแบบจำลองไนแอมนั้นมีความสัมพันธ์ของข้อเท็จจริงที่หลากหลายรูปแบบตามลักษณะการใช้งานในประเภทต่าง ๆ โดยสามารถแบ่งความสัมพันธ์ของข้อเท็จจริงได้เป็นหลายประเภท ยกตัวอย่างเช่น ข้อเท็จจริงประเภทความสัมพันธ์ของออปเจ็กต์เดียว (Unary Fact Type), ข้อเท็จจริงประเภทความสัมพันธ์ระหว่างสองออปเจ็กต์ (Binary Fact Type), ข้อเท็จจริงประเภทความสัมพันธ์ระหว่างสามออปเจ็กต์ (Ternary Fact Type) และข้อเท็จจริงประเภทความสัมพันธ์แบบซับซ้อน (Nested Fact Type) เป็นต้น

- ขั้นตอนที่ 3 กำจัดเอนทิตี้ส่วนที่เกินออกไปและรวมเอนติตี้ที่เหมือนกันไว้ด้วยกัน

ในขั้นตอนนี้จะทำการพิจารณาข้อเท็จจริงต่าง ๆ ที่ได้ออกแบบไว้แล้วว่ามีข้อเท็จจริงใดบ้างที่หาได้จากการคำนวณสามารถใส่เครื่องหมาย “*” แสดงถึงข้อเท็จจริงที่เกิดจากการคำนวณดังแสดงในรูปที่ 2.6 และ 2.7



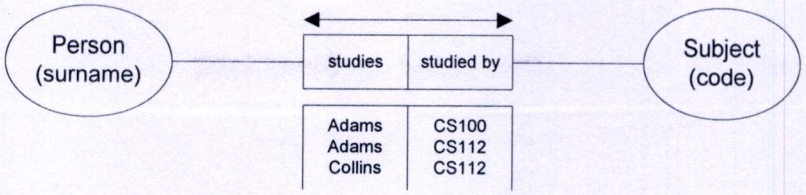
รูปที่ 2.6 ตัวอย่างการรวมเอนติตี้ที่เหมือนกันไว้ด้วยกันแบบที่ 1



รูปที่ 2.7 ตัวอย่างการรวมเอนติตี้ที่เหมือนกันไว้ด้วยกันแบบที่ 2

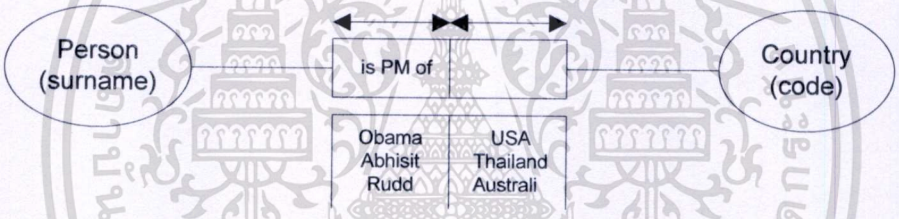
- ขั้นตอนที่ 4 การใส่ข้อบังคับความเป็นหนึ่ง (Uniqueness Constraints)

ซึ่งเป็นเครื่องหมายที่บังคับว่าข้อมูลที่อยู่ในเอนติตี้นั้นจะต้องไม่มีค่าที่ซ้ำกัน สามารถนำเครื่องหมาย ข้อบังคับความเป็นหนึ่งนี้ไปใช้กับความสัมพันธ์ของข้อเท็จจริงรูปแบบต่าง ๆ กันได้ ตัวอย่างของข้อบังคับความเป็นหนึ่งมีดังนี้



รูปที่ 2.8 ข้อบังคับความเป็นหนึ่งของประเภทความสัมพันธ์ระหว่างสองออปเจ็กต์ แบบ Many to Many

จากตัวอย่างจะเห็นได้ว่า นักศึกษาหนึ่งคนสามารถลงทะเบียนได้มากกว่าหนึ่งวิชา และวิชาหนึ่งวิชาสามารถมีคนเลือกลงได้มากกว่าหนึ่งคน ดังแสดงในรูปที่ 2.8 ยกตัวอย่าง เช่น นักศึกษาที่ชื่อ "Adams" สามารถลงทะเบียนได้มากกว่าหนึ่งวิชา และวิชาที่มีรหัสเป็น CS112 สามารถถูกเลือกลงทะเบียนได้โดยนักศึกษามากกว่าหนึ่งคน

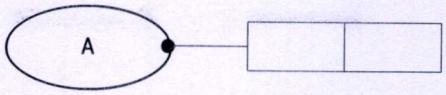


รูปที่ 2.9 ข้อบังคับความเป็นหนึ่งของประเภทความสัมพันธ์ระหว่างสองออปเจ็กต์ แบบ One to One

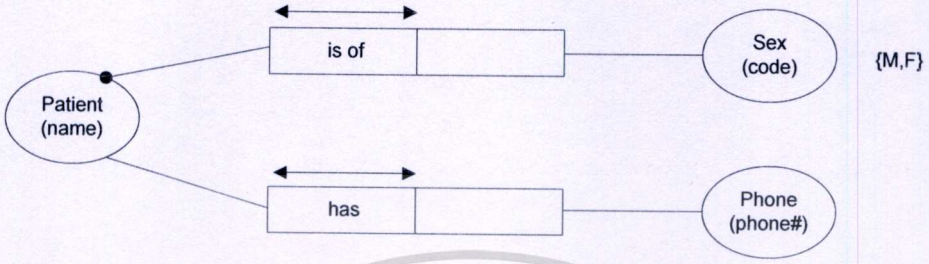
จากความสัมพันธ์แบบ One to One ในรูปที่ 2.9 นั้นสามารถอธิบายได้ว่าบุคคลหนึ่งคนสามารถเป็นนายกรัฐมนตรีได้เพียงประเทศเดียวเท่านั้นและข้อมูลในแต่ละเอนติตี้นั้นจะต้องไม่มีค่าที่ซ้ำกันเกิดขึ้นด้วย

- ขั้นตอนที่ 5 ใส่ประเภทของเอนติตี้และสัญลักษณ์ข้อบังคับความจำเป็น

สัญลักษณ์ข้อบังคับความจำเป็น (Mandatory Roles) นั้น เป็นข้อบังคับที่บอกว่าในเอนติตี้ นั้น มีข้อมูลที่สำคัญและมีความจำเป็นที่จะต้องมีอยู่ในทุก ๆ ตาราง ยกตัวอย่าง เช่น ในรูปที่ 2.11 มีการใส่สัญลักษณ์ข้อบังคับความจำเป็นระหว่างเอนติตี้ของผู้ป่วยและเอนติตี้ของเพศของผู้ป่วย ซึ่งหมายถึง ข้อมูลประวัติผู้ป่วยนั้นจะต้องมีเพศของผู้ป่วยเสมอ แต่อาจจะมีหรือไม่มีเบอร์โทรศัพท์ก็ได้



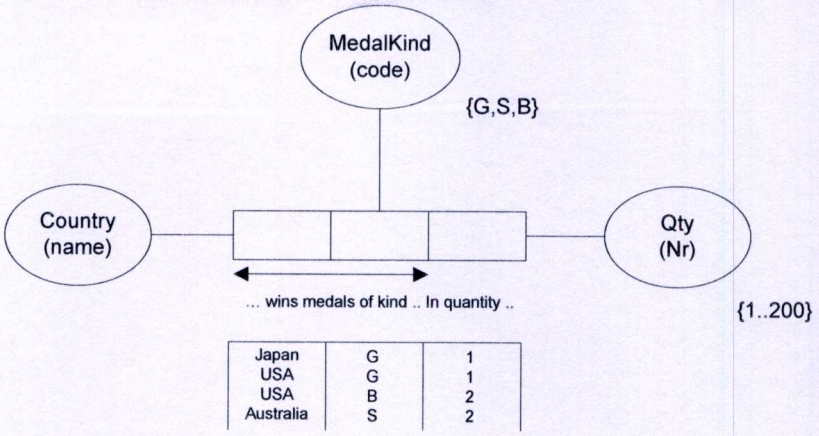
รูปที่ 2.10 ตัวอย่างสัญลักษณ์ข้อบังคับความจำเป็น (Mandatory Roles)



รูปที่ 2.11 ตัวอย่างการใช้สัญลักษณ์ข้อบังคับความจำเป็น

- ขั้นตอนที่ 6 ใส่เครื่องหมายกฎข้อบังคับความถูกต้องอื่น ๆ เช่น ข้อบังคับค่าของข้อมูล (Value Constraint), ข้อบังคับของเซ็ต (Set Comparison) และข้อบังคับของสับไทป์ (Subtype Constraint)

ขั้นตอนนี้เป็นการใส่กฎข้อบังคับความถูกต้องเพิ่มเติมลงไปในแบบจำลองไนแอมที่ได้ออกแบบไว้ โดยข้อบังคับค่าของข้อมูลนั้นจะใช้การกำหนดกลุ่มของค่าของข้อมูลที่สามารถเกิดขึ้นได้ในเอนติตี้นั้น ๆ ดังแสดงในรูปที่ 2.12 เอนติตี้เหรียญรางวัลการแข่งขันนั้นในความเป็นจริงแล้วจะมีได้แค่สามชนิดเท่านั้น คือ เหรียญทอง เหรียญเงิน และเหรียญทองแดง ซึ่งข้อบังคับค่าของข้อมูลนี้จะแทนด้วย {G, S, B} ซึ่งก็คือเหรียญรางวัลที่เป็นไปได้สามชนิดตามลำดับนั่นเอง ส่วนเอนติตี้ จำนวนเหรียญที่แต่ละชาติได้รับสามารถใช้ข้อบังคับค่าของข้อมูลระบุเป็นช่วงได้ เช่น {1...200} หมายถึงค่าที่เกิดขึ้นในเอนติตี้นี้จะต้องมีจำนวนเหรียญได้ตั้งแต่ 1 ถึง 200 เท่านั้น จะไม่สามารถใช้ค่าที่มากกว่า หรือน้อยกว่านี้ได้



รูปที่ 2.12 ข้อบังคับของค่าความถูกต้องชนิดต่าง ๆ

- ขั้นตอนที่ 7 ทำการตรวจสอบความถูกต้อง

ขั้นตอนนี้ถือขั้นสุดท้ายเพื่อให้ได้แบบจำลองแนวความคิดไนแอมที่ครบถ้วนสมบูรณ์ตามที่ผู้สร้างต้องการ ไนแอมมีข้อบังคับหลาย ๆ อย่างซึ่งไม่อาจกล่าวได้หมดในงานวิจัยนี้ โดยจะกล่าวเฉพาะข้อบังคับที่เกี่ยวข้องกับงานวิจัยเท่านั้น เนื่องจากไนแอมมีข้อบังคับและรูปแบบที่หลากหลายจึงสามารถนำไปประยุกต์ใช้กับงานวิจัยที่หลากหลายได้

2.6 การบูรณาการสารสนเทศ (Information Integration)

ในปัจจุบันมีแหล่งข้อมูลสารสนเทศที่มีประโยชน์ต่อผู้ใช้อยู่มากมายบนเครือข่ายอินเทอร์เน็ต แม้ว่าข้อมูลสารสนเทศบางแหล่งข้อมูลนั้นจะมีเนื้อหาสอดคล้องกัน แต่ก็ไม่ได้อยู่บนฐานข้อมูลเดียวกัน ทำให้การเข้าถึงแหล่งข้อมูลเหล่านั้นเป็นไปด้วยความยากลำบาก การบูรณาการแหล่งข้อมูลสารสนเทศนั้นจึงเป็นการรวบรวมแหล่งข้อมูลสารสนเทศที่หลากหลายและกระจายอยู่บนเครือข่ายอินเทอร์เน็ตมารวบรวมไว้ภายใต้ฐานข้อมูลที่มาตรฐานเดียวกัน ซึ่งจะทำให้การเข้าถึงข้อมูลเหล่านั้นทำได้สะดวกและรวดเร็วมากยิ่งขึ้น

การบูรณาการข้อมูลสารสนเทศนั้น เป็นการรวมสารสนเทศจากแหล่งข้อมูลที่ต่างๆ ที่แตกต่างกันออกไป ซึ่งอาจจะแตกต่างทางด้านแนวความคิด (Conceptual), ความหมาย, รูปแบบการแสดงผลข้อมูล โดยการบูรณาการสารสนเทศนั้นจะใช้ในการรวมแหล่งข้อมูลที่หลากหลายนั้นเข้าไว้ด้วยกัน แม้ว่าแหล่งข้อมูลต่างๆ จะไม่ได้มีโครงสร้าง (Unstructured) หรือเป็นแบบกึ่งโครงสร้าง (Semi-Structured) ก็ตาม การบูรณาการสารสนเทศนั้นบางครั้งก็รวมถึงการบูรณาการองค์ความรู้ด้วยเช่นกัน

2.7 ปัญหาที่พบในการบูรณาการสารสนเทศ [23][24][25]

โดยปรกติแล้วเมื่อนำแหล่งข้อมูลสารสนเทศจากแหล่งต่าง ๆ มาบูรณาการเข้าไว้ด้วยกัน มักจะเกิดความขัดแย้งขึ้นอยู่เสมอ เนื่องจากแหล่งข้อมูลสารสนเทศแต่ละแหล่งนั้น มีโครงสร้างฐานข้อมูลและการใช้คำ ชื่อ สัญลักษณ์ รหัส และหน่วยข้อมูลที่แตกต่างกัน โดยปัญหาดังกล่าวนั้นสามารถแบ่งออกได้เป็นสองประเภทดังนี้คือ ความหลากหลายด้านความหมาย (Semantic Heterogeneity) และความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneity) โดยปัญหาทั้งสองประเภทสามารถอธิบายโดยละเอียดได้ดังนี้

2.7.1 ความหลากหลายทางด้านความหมาย (Semantic Heterogeneity)

ความหลากหลายทางด้านความหมาย คือปัญหาที่เกิดขึ้นจากแหล่งข้อมูลแต่ละแหล่งนั้นใช้ คำ ชื่อ สัญลักษณ์ รหัส และหน่วยข้อมูลทีแตกต่างกันออกไปเพื่อแสดงถึงข้อมูลที่มีความหมายเดียวกัน โดยปัญหาที่เกิดจากความแตกต่างทางด้านความหมายนั้นสามารถแบ่งเป็นประเภทต่าง ๆ ได้ดังต่อไปนี้

- **ประเภทการใช้ชื่อที่แตกต่างกัน (Naming Conflicts)**

ปัญหาความหลากหลายทางด้านความหมายประเภทนี้ เกิดจากแหล่งข้อมูลแต่ละแหล่งนั้นใช้ชื่อที่แตกต่างกันออกไปในการแสดงถึงข้อมูลที่มีความหมายเดียวกัน ยกตัวอย่างเช่น United States of America สามารถเขียนได้หลายแบบ เช่น USA, U.S.A. หรืออาจเรียกสั้นๆได้ว่า America เป็นต้น

- **ประเภทการใช้หน่วยข้อมูลที่แตกต่างกัน (Scaling Conflicts หรือ Unit Conflicts)**

เป็นปัญหาที่เกิดขึ้นจากแต่ละแหล่งข้อมูลมีการใช้มาตรฐานของหน่วยวัดที่ใช้แสดงถึงปริมาณของข้อมูลที่แตกต่างกัน ยกตัวอย่างเช่น ข้อมูลรายงานทางด้านการเงินนั้นอาจจะใช้ค่าเงินที่แตกต่างกันไปในแต่ละประเทศ เกรดของนักศึกษาแต่ละมหาวิทยาลัยอาจจะใช้หน่วยที่ไม่เหมือนกัน เช่น บางมหาวิทยาลัยอาจจะใช้ “A”, “B”, “C”, “D” และ “F” หรือบางมหาวิทยาลัยใช้ “Excellent”, “Good”, “Pass”, “Fail” เป็นต้น

- **ประเภทขอบเขตของข้อมูล (Confounding Conflicts)**

ยกตัวอย่างเช่น รายงาน “อัตราแลกเปลี่ยนเงินตรา” นั้นควรจะต้องมีวันที่ที่ประกาศใช้อัตราแลกเปลี่ยนเงินดังกล่าวกำกับไว้เสมอ เนื่องจากการขาดข้อมูลที่กำกับดังกล่าวไปนั้นอาจจะทำให้ผู้ที่อ่านรายงานมีความเข้าใจที่คลาดเคลื่อนได้ หรือ

2.7.2 ความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneity)

โดยปรกติแล้วในการบูรณาการแหล่งข้อมูลสารสนเทศ นอกจากปัญหาที่เกิดจากความหลากหลายทางด้านความหมายแล้วนั้น ยังมีปัญหาที่เกิดจากความหลากหลายทางด้านโครงสร้างด้วยเช่นกัน เนื่องจากแหล่งข้อมูลสารสนเทศส่วนใหญ่นั้นมีโครงสร้างที่ออกแบบมาอย่างอิสระ ซึ่งความแตกต่างทางด้านโครงสร้างนั้นสามารถแบ่งเป็นประเภทต่าง ๆ ได้ดังนี้

- การใช้รูปแบบของข้อมูลที่แตกต่างกัน (Data Type Conflicts)

ยกตัวอย่างเช่น ผู้ออกแบบโครงสร้างฐานข้อมูลในแหล่งข้อมูลสารสนเทศแห่งหนึ่งอาจจะใช้ประเภทของข้อมูลแบบ “String” ในการเก็บข้อมูลวันที่ ในขณะที่ผู้ออกแบบระบบสารสนเทศอีกแห่งหนึ่งอาจจะใช้ประเภทข้อมูลแบบ “Date” ในการเก็บวันที่ดังกล่าว เป็นต้น

- การใช้ชื่อเลเบลที่แตกต่างกัน (Labeling Conflicts)

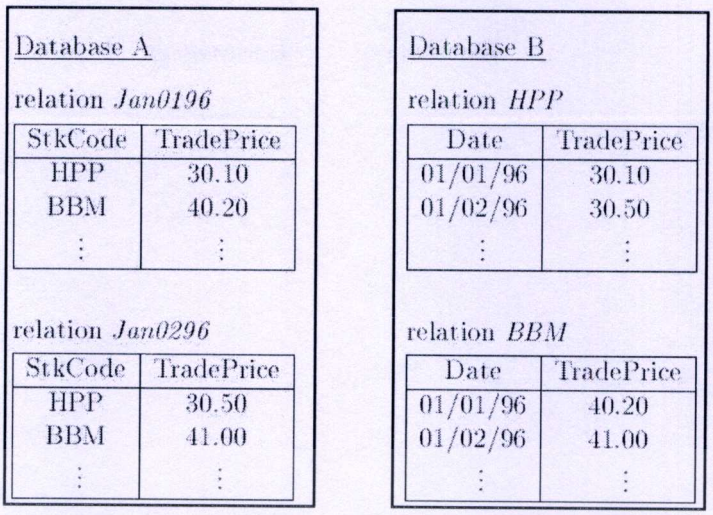
ยกตัวอย่างเช่น แอดทริบิวที่มีความหมายเดียวกันอาจจะใช้ชื่อเลเบลที่แตกต่างกัน (Synonym) กล่าวคือการใช้ชื่อ ในฐานข้อมูลบริษัทแห่งแรกอาจจะใช้ชื่อเลเบลว่า “Earning” แทนเงินเดือนที่พนักงานได้รับ ส่วนบริษัทที่อาจจะใช้ “Salary” แทนเงินเดือนของพนักงาน เป็นต้น ในทางตรงกันข้าม เลเบลเดียวกันอาจจะสื่อความหมายที่แตกต่างกันออกไป (Homonyms) ในแอดทริบิวเหล่านั้นเช่นกัน ยกตัวอย่างเช่น ในบางแหล่งข้อมูลอาจจะใช้ชื่อแอดทริบิวว่า “ราคา” แทนราคาสินค้าที่รวมภาษีเรียบร้อยแล้ว แต่แหล่งข้อมูลสารสนเทศบางแหล่งอาจจะใช้ชื่อแอดทริบิว “ราคา” เหมือนกันแทนราคาสินค้าที่ยังไม่ได้คิดภาษี เป็นต้น

- ความขัดแย้งแบบรวมกลุ่ม (Aggregated Conflicts)

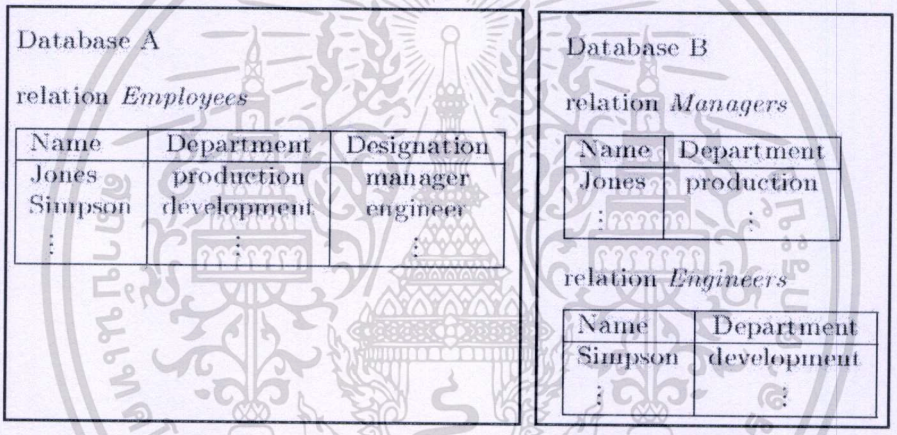
เป็นความขัดแย้งซึ่งเกิดมาจากผู้ออกแบบระบบโครงสร้างฐานข้อมูลแต่ละคนได้ออกแบบตารางข้อมูลแตกต่างกันไปตามความจำเป็นหรือความต้องการของแต่ละองค์กร ยกตัวอย่างเช่น ในรูปที่ 2.13 “ราคาแลกเปลี่ยนสินค้า” นั้น ฐานข้อมูล A ออกแบบตารางข้อมูลสำหรับเก็บราคาดังกล่าวซึ่งเก็บตามรหัสของสินค้า ในขณะที่ฐานข้อมูล B ออกแบบตารางข้อมูลสำหรับเก็บราคาดังกล่าวตามวันที่ เป็นต้น ซึ่งจะทำให้เกิดปัญหาในการบูรณาการได้

- ความขัดแย้งแบบกระจาย (Generalization Conflicts)

ความขัดแย้งรูปแบบนี้จะแตกต่างจะความขัดแย้งแบบรวม ตรงที่การออกแบบฐานข้อมูลแต่ละแหล่งนั้นอาจจะเลือกประเภทของเอนติตี้ที่แตกต่างกันไป ยกตัวอย่างเช่น ในรูปที่ 2.14 นั้น ฐานข้อมูล A ออกแบบตารางซึ่งจะแยกเอนติตี้ “Manager” และ “Engineer” ออกจากกันตามความจำเป็นในแต่ละองค์กร ในขณะที่ระบบฐานข้อมูล B ได้ออกแบบตารางที่มี “Manager” และ “Engineer” ไว้ด้วยกันภายในเอนติตี้เดียวกัน โดยเรียกเอนติตี้นั้นว่า “Employees” เป็นต้น



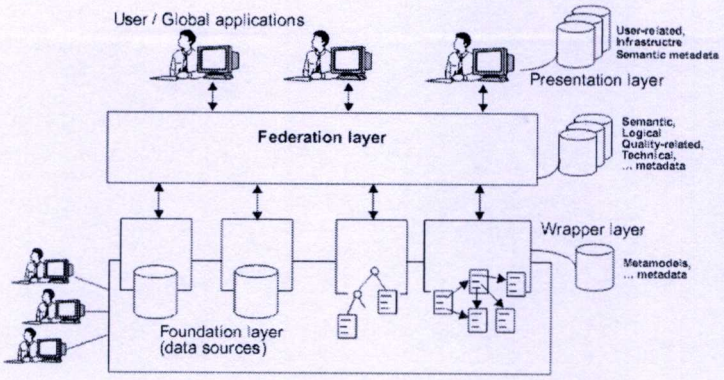
รูปที่ 2.13 ตัวอย่างความขัดแย้งแบบรวมกลุ่ม (Aggregated Conflicts)



รูปที่ 2.14 ตัวอย่างความขัดแย้งแบบกระจาย (Generalization Conflicts)

2.8 ระบบสหภาพสารสนเทศ (Federation Information system) [26]

จากปัญหาเรื่องความหลากหลายทางด้านความหมายและโครงสร้างที่เกิดจากการบูรณาการแหล่งข้อมูลสารสนเทศที่แตกต่างกันดังที่กล่าวไปข้างต้นนั้น ระบบสหภาพสารสนเทศ (Federation Information System) ได้นำเสนอวิธีการแบ่งระดับชั้นสหภาพ (Federation Layer) ซึ่งทำหน้าที่เป็นตัวกลางในการเข้าถึงข้อมูลที่ถูกเก็บไว้ในแหล่งข้อมูลที่แตกต่างกัน โดยระดับชั้นสหภาพนี้จะมีการสร้างสคีมากลาง (Global Schema) ขึ้นมา เพื่อให้ผู้ใช้สามารถเข้าถึงแหล่งข้อมูลสารสนเทศที่หลายหลากได้อย่างเป็นหนึ่งเดียวกัน ในระดับชั้นสภาพนั้นจะใช้พจนานุกรมข้อมูล (Data Dictionary) หรือ เมตาดาต้า (Metadata) เข้ามาช่วยในการแก้ไขปัญหาด้านความหลากหลายนี้ เมตาดาต้านั้นจะทำหน้าที่เป็นตัวอธิบายข้อมูลโครงสร้างของแต่ละแหล่งข้อมูลสารสนเทศต่างๆในระบบ นอกจากนี้เมตาดาต้ายังมีอีกหลายประเภทดังจะได้กล่าวถึงในส่วนต่อไป โดยระบบสหภาพสารสนเทศนั้นแสดงในรูปที่ 2.15



รูปที่ 2.15 ระบบสหภาพสารสนเทศ

2.8.1 เมตาดาต้าในระบบสารสนเทศ

แนวความคิดสำคัญอย่างหนึ่งในการแก้ไขปัญหาความหลากหลายของแหล่งข้อมูลสารสนเทศที่แตกต่างกันไป ก็คือการใช้เมตาดาต้าหรือพจนานุกรมข้อมูลนั้นเอง เมตาดาต้านั้นให้ความกระจ่างในการบริหารจัดการข้อมูลสารสนเทศ โดยการให้คำอธิบายเกี่ยวกับโครงสร้างและอีลิเมนต์ต่าง ๆ ในระบบเพื่อช่วยในการทำงานร่วมกันระหว่างระบบ (Interoperation) ได้ ซึ่งเมตาดาต้านี้มีส่วนช่วยให้การทำงานของระบบเป็นไปด้วยความยืดหยุ่นมากขึ้น

เมตาดาต้าทำหน้าที่อธิบายรายละเอียดและคุณลักษณะของข้อมูล โดยมีโครงสร้างและลักษณะที่เข้าใจได้ง่าย จึงมีการนำเมตาดาต้านี้ไปใช้ในงานในระบบต่าง ๆ ยกตัวอย่างเช่น ในการสืบค้นข้อมูลของหนังสือในห้องสมุดนั้นมีการใช้เมตาดาต้าในการกำหนดมาตรฐานข้อมูลของหนังสือให้เป็นมาตราฐานเดียวกัน โดยมาตรฐานของเมตาดาต้าที่นิยมใช้ได้ห้องสมุดได้แก่ ดับลินคอร์ [27] MARC [28] เป็นต้น นอกจากการใช้เมตาดาต้ากับการสืบค้นหนังสือในห้องสมุดแล้วยังมีการนำเมตาดาต้าไปใช้ในการกำหนดมาตรฐานข้อมูลของไฟล์เพลง mp3 อีกด้วย โดยมาตรฐานเมตาดาต้าที่ใช้ในการกำหนดไฟล์เพลง mp3 คือ ID3 [29]

- **เมตาดาต้าประเภทเทคนิค (Technical Metadata)**

เป็นเมตาดาต้าที่อธิบายถึงรายละเอียดของกระบวนการเข้าถึงข้อมูลในแต่ละแหล่งข้อมูลต่าง ๆ ยกตัวอย่างเช่น โปรโตคอล ความเร็วในการเชื่อมต่อ เป็นต้น ซึ่งมักจะใช้เมตาดาต้าประเภทนี้ร่วมกับเมตาดาต้าประเภทอื่น ๆ

- **เมตาดาต้าประเภทตรรกะ (Logical Metadata)**

เป็นข้อมูลที่เกี่ยวกับสคีมาของแหล่งสารสนเทศต่างๆ และความสัมพันธ์เชิงตรรกะ (Logical relationship) ซึ่งตัวอย่างของเมตาดาต้าประเภทนี้ได้แก่ คำอธิบายข้อมูลใน

RDBMS (Relational database management system) หรือ คลาสไดอะแกรมใน OODBMS (Object oriented database management systems) เป็นต้น

- **เมตาโมเดล (Metamodels)**

เป็นเมตาดาต้าที่สนับสนุนการทำงานระหว่างระบบที่มีรูปแบบข้อมูลที่แตกต่างกัน ซึ่งจะช่วยแก้ไขปัญหาที่เกิดจากการนำข้อมูลสารสนเทศมาบูรณาการเข้าไว้ด้วยกัน

- **เมตาดาต้าประเภทความหมาย (Semantic Metadata)**

เป็นข้อมูลสารสนเทศที่ช่วยอธิบายความหมายของข้อมูล เป็นเมตาดาต้าที่ออกแบบมาเพื่อช่วยแก้ปัญหาในเรื่องความหลากหลายทางด้านความหมาย

- **เมตาดาต้าประเภทคุณภาพข้อมูล (Quality-related Metadata)**

อธิบายรายละเอียดเฉพาะเจาะจงของคุณสมบัติในระบบสารสนเทศแต่ละแหล่งทางด้านคุณภาพ ยกตัวอย่างเช่น ความน่าเชื่อถือ (Reliability), ความถี่ในการอัพเดทข้อมูล เป็นต้น ซึ่งเมตาดาต้าประเภทนี้ยังมีส่วนช่วยในจัดการระบบอีกด้วย

- **เมตาดาต้าประเภทโครงสร้างพื้นฐานที่จำเป็น (Infrastructure Metadata)**

เมตาดาต้าประเภทนี้จะช่วยผู้ใช้ในการค้นหาข้อมูลที่เกี่ยวข้องซึ่งรวมไปถึงการทำหมายเหตุข้อมูล (Bookmarks) เมตาดาต้าประเภทนี้จะถูกเรียกใช้โดยผู้ใช้หรือระบบบริการสารสนเทศในระดับชั้นแอพพลิเคชั่นของระบบสหภาพสารสนเทศ

- **เมตาดาต้าประเภทผู้ใช้ (User-related Metadata)**

เป็นเมตาดาต้าที่เก็บข้อมูลด้านความรับผิดชอบหรือความต้องการของผู้ใช้ที่มีในระบบสารสนเทศ ยกตัวอย่างเช่น ระบบโปรไฟล์ของผู้ใช้ เป็นต้น

2.8.2 การแบ่งประเภทของระบบสหภาพสารสนเทศ

เพื่อแก้ไขปัญหาเรื่องความหลากหลายต่าง ๆ ที่เกิดขึ้นจากการบูรณาการระบบสารสนเทศ จึงมีการใช้ระดับชั้นสหภาพเข้ามาเป็นตัวกลางในการแก้ไขปัญหา โดยระดับชั้นสหภาพนั้นจะแบ่งเป็นสองประเภทด้วยกัน คือระบบสหภาพแบบรัดกุม (Tight Federation) และระบบสหภาพแบบอิสระ (Loose Federation)

2.8.2.1 การออกแบบระบบสหภาพแบบรัดกุม (Tight Federation)

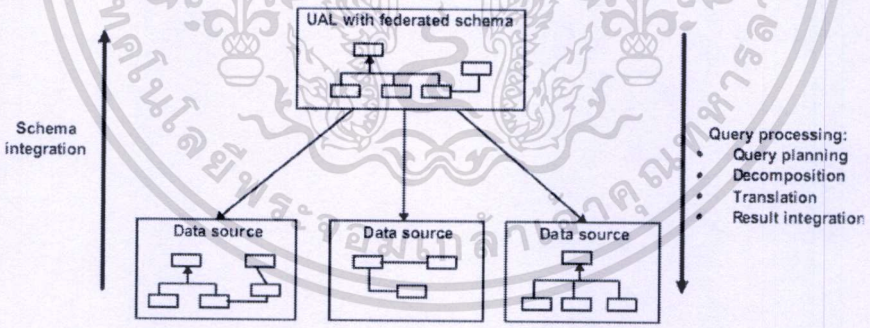
ในการออกแบบระบบสหภาพแบบรัดกุมนั้นจะมีการสร้างสคีมากลาง (Global Schema) หรือ สคีมาแบบบูรณาการ (Integrated or Federated Schema) เพื่อให้ผู้ใช้สามารถเข้าถึงข้อมูลได้ภายใต้โครงสร้างข้อมูลที่เป็นหนึ่งเดียวกัน (Unified) สคีมากลางนี้จะถูกสร้างผ่านกระบวนการบูรณาการแบบอัตโนมัติ [30] [31] หรือกระบวนการที่มีอยู่ทั่วไปก็ได้ และสคีมากลางนี้จะต้องแก้ปัญหาเรื่องความหมายได้ด้วย

การสร้างสคีมากลางโดยผ่านกระบวนการบูรณาการโดยทั่วไปนั้นจะไม่ได้พิจารณาถึงเนื้อหาด้านความหมายของแหล่งข้อมูลสารสนเทศที่มีอยู่ในระบบก่อนจะเป็นการ

พิจารณาถึงความสมบูรณ์ (Completeness) และความถูกต้อง (Correctness) เพียงอย่างเดียว ซึ่งจะแตกต่างจากการสร้างสคีมากลางในระดับชั้นสหภาพนี้ที่จะถูกสร้างขึ้นจากการพิจารณาด้านความหมายเป็นหลัก (Semantic Essence) ซึ่งจะต้องมีเงื่อนไขและเนื้อหาสอดคล้องกลับสคีมาของแหล่งข้อมูลสารสนเทศที่นำเข้ามาบูรณาการด้วย

การใช้สคีมากลางนั้นก็เพื่อความชัดเจนในการแก้ไขปัญหาและรับมือกับความหลากหลายของสคีมาจากแหล่งข้อมูลที่หลากหลาย โดยจะต้องพิจารณาทั้งปัญหาในการบูรณาการด้านสคีมาและปัญหาในกระบวนการคิวรี่ด้วยเช่นกัน ในการที่จะแก้ไขปัญหาความหลากหลายด้านความหมายนั้น ระบบสหภาพสารสนเทศจำเป็นจะต้องทราบถึงรายละเอียดเมตาดาต้าประเภทตรรกะทั้งหมดของสคีมากลาง สคีมาของแหล่งสารสนเทศแต่ละแหล่ง คิวรี่ต่างๆ เป็นต้น ซึ่งเมตาดาต้าประเภทนี้อาจจะมาจากกฎหรือข้อบังคับที่ผู้ออกแบบระบบได้สร้างเอาไว้ เป็นต้น

ในการบูรณาการแหล่งข้อมูลสารสนเทศหลายแหล่งเข้าด้วยกันนั้น คงเป็นเรื่องยากที่จะทำให้ผู้ใช้แต่ละคนเข้าใจถึงสคีมาของแหล่งข้อมูลสารสนเทศแต่ละแหล่งข้อมูลได้ ยิ่งไปกว่านั้นถ้าหากแหล่งข้อมูลบางแหล่งมีการปรับเปลี่ยนสคีมาค่อนข้างบ่อย สคีมากลางจึงมีบทบาทสำคัญในการบูรณาการแหล่งข้อมูลที่มีสคีมาแตกต่างกัน เพื่อทำให้ผู้ใช้ทุกคนสามารถเข้าถึงข้อมูลสารสนเทศได้อย่างเป็นหนึ่งเดียวกัน โดยที่ผู้ใช้ไม่ต้องทราบรายละเอียดเกี่ยวกับสคีมาของแหล่งข้อมูลสารสนเทศทั้งหมด รูปแบบของระบบสหภาพแบบรัดกุมนั้นแสดงในรูปที่ 2.16



รูปที่ 2.16 ระบบสหภาพแบบรัดกุมในระบบสหภาพสารสนเทศ

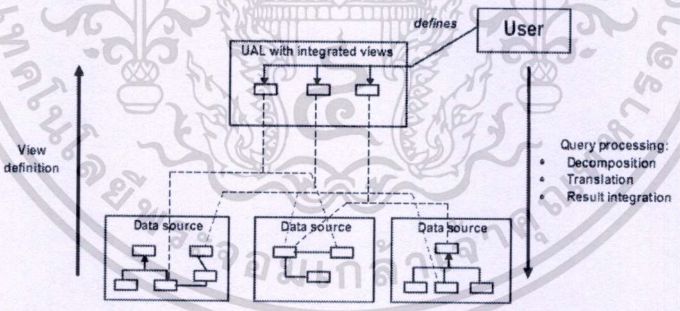
2.8.2.2 การออกแบบระบบสหภาพแบบอิสระ (Loose Federation)

ระบบสหภาพแบบอิสระนั้นได้นำเสนอภาษากลางที่ใช้ในการคิวรี่ (Multi-Database Query Language, MDBQL) [32] แทน ซึ่งถอดแบบมาจากภาษาคิวรี่ของแหล่งข้อมูลสารสนเทศต่าง ๆ และซ่อนความหลากหลายทางด้านเทคนิคและภาษาที่ใช้คิวรี่ที่แตกต่างกันจากผู้ใช้ได้อีกด้วย ซึ่งจะต่างจากระบบสหภาพแบบรัดกุมที่ได้นำเสนอสคีมากลางในการเข้าถึงข้อมูล ดังนั้นผู้ใช้ทุกคนในระบบสหภาพแบบอิสระจะต้องรับผิดชอบในเรื่องความหลากหลายทางด้าน

ตรรกะของแหล่งข้อมูลสารสนเทศด้วยตนเอง กล่าวคือ ผู้ใช้จะต้องทราบและเข้าใจถึงความหลากหลายของภาษาคิวรี่ในแหล่งข้อมูลต่างๆ, ความสัมพันธ์ต่างๆ หรือโครงสร้างต่าง ๆ ของแหล่งข้อมูลและเพื่อแก้ไขปัญหาเรื่องความหลากหลายทางด้านโครงสร้างนั้น MDBQL อาจจะต้องมีการปรับเปลี่ยนอีลิเมนต์บางส่วนของสคีมาให้เสมือนว่าตัวมันเป็นข้อมูลส่วนหนึ่ง ซึ่งเป็นสิ่งที่ไม่สามารถทำได้ใน SQL

ระบบสหภาพแบบอิสระนั้นจะสามารถสร้างขึ้นได้ก็ต่อเมื่อแหล่งข้อมูลสารสนเทศเหล่านั้นอนุญาตให้ระบบสหภาพสารสนเทศสามารถเข้าถึงข้อมูลของตนได้เท่านั้น ถ้าหากว่าแหล่งข้อมูลสารสนเทศเหล่านั้นไม่ได้อนุญาตให้เข้าถึงได้ หรือมีการจำกัดการเข้าถึงข้อมูลระบบสหภาพแบบอิสระนี้ก็จะไม่สามารทำงานได้ โดยในระบบสหภาพแบบรัดกุมนั้นสามารถแก้ไขปัญหาส่วนนี้ได้ แต่ในสหภาพแบบอิสระนั้นแก้ปัญหาในส่วนนี้ได้ยากกว่ามาก เนื่องระบบสหภาพประเภทนี้จะต้องออกแบบให้ผู้ใช้แต่ละคนมีความอิสระในการใช้คำสั่งคิวรี่เพื่อเข้าถึงแหล่งข้อมูลสารสนเทศที่มีอยู่ได้โดยตรงอีกด้วย

เพื่อเป็นการบรรเทาความยากลำบากในการใช้งานระบบสารสนเทศแบบบูรณาการระบบที่ใช้ MDBQL มักจะใช้การบูรณาการด้านมุมมอง (Integrating Views) เข้ามาช่วย หมายความว่าผู้ใช้สามารถกำหนดมุมมองบนแหล่งข้อมูลสารสนเทศ และอนุญาตให้ผู้ใช้ท่านอื่น ๆ ใช้งานได้ด้วย ซึ่งผู้ใช้แต่ละท่านอาจจะใช้มุมมองเหล่านี้ร่วมกันได้ โดยตัวอย่างของระบบสหภาพแบบอิสระนั้นได้แสดงในรูปที่ 2.17



รูปที่ 2.17 ระบบสหภาพแบบอิสระในระบบสหภาพสารสนเทศ

อาจกล่าวโดยสรุปได้ว่า สิ่งที่แตกต่างกันระหว่างสหภาพแบบรัดกุมและสหภาพแบบอิสระนั้น คือ ในสหภาพแบบรัดกุมนั้นมีการใช้สร้างสคีมากลางขึ้นมาใหม่เพื่อความเป็นหนึ่งเดียวกันของข้อมูลสารสนเทศ แต่ในสหภาพแบบอิสระนั้นสคีมาของแหล่งข้อมูลสารสนเทศต่าง ๆ จะถูกผู้ใช้มองเห็นอยู่ตามปรกติ กล่าวคือ ผู้ใช้สามารถมองเห็นข้อมูลบนแหล่งข้อมูลสารสนเทศได้อย่างอิสระไม่ได้ถูกจำกัดมุมมองแค่สคีมากลางเท่านั้น นอกจากนั้นผู้ใช้ยังสามารถกำหนดมุมมองเพิ่มเติมขึ้นใหม่เพื่อให้เข้าถึงข้อมูลเหล่านั้นได้ด้วยตนเองอีกด้วย

2.8.3 แนวทางการออกแบบสคีมากลาง

วิธีการออกแบบสคีมากลางสำหรับระบบสหภาพแบบรัดกุมนั้นแบ่งออกเป็นสองวิธีด้วยกันดังนี้คือ การออกแบบจากบนลงล่าง (Top-Down) เป็นการเริ่มต้นการออกแบบสคีมากลางโดยอาศัยความต้องการของผู้ใช้ก่อนและหลังจากนั้นจึงทำการเชื่อมต่อแหล่งข้อมูลสารสนเทศเข้ามาในระบบที่ได้ออกแบบไว้รองรับเป็นที่เรียบร้อยแล้ว การออกแบบจากล่างขึ้นบน (Bottom-up) เริ่มต้นด้วยการบูรณาการตามความต้องการของแหล่งข้อมูลสารสนเทศ โดยจะอธิบายแนวทางในการออกแบบทั้งสองวิธีอย่างละเอียดดังต่อไปนี้

2.8.3.1 การออกแบบจากบนลงล่าง (Top-Down)

การออกแบบสคีมากลางด้วยวิธีบนลงล่างนั้น ผู้สร้างจะต้องพิจารณาถึงต้องการโดยรวมของข้อมูลสารสนเทศเสียก่อน ยกตัวอย่างเช่น ในบริษัทแห่งหนึ่งต้องการที่จะเสนอบริการที่สามารถค้นหาราคาหนังสือที่ถูกที่สุดจากแหล่งข้อมูลทางอินเทอร์เน็ตหลายๆ แห่งให้กับลูกค้าของตน หรือระบบช่วยเหลือการตัดสินใจ มีความต้องการที่จะบูรณาข้อมูลลูกค้าเพียงอย่างเดียวที่กระจายกันอยู่ในฐานข้อมูลในแผนกต่างๆ เป็นต้น ซึ่งกรณีต่างๆเหล่านี้จะพบว่า การออกแบบสคีมากลางนั้นไม่ได้นำสคีมาของแหล่งข้อมูลสารสนเทศต่างๆ เข้ามาร่วมพิจารณาเลย เนื่องจากการออกแบบในลักษณะนี้เป็นการออกแบบซึ่งเกิดขึ้นจากความต้องการของระบบสารสนเทศเป็นหลัก โดยในการออกแบบสคีมากลางนี้จะมีข้อบังคับสำหรับ การบูรณาการสคีมาอยู่ 4 ประเด็นด้วยกันคือ ความสมบูรณ์ (Completeness) ความถูกต้อง (Correctness) ความเข้าใจ (Understandability) และความกะทัดรัด (Minimality) ประเด็นแรก ในการออกแบบสคีมากลางตามความต้องการในระบบสารสนเทศนั้น บางครั้งต้องการเพียงแค่ข้อมูลลูกค้าเท่านั้น ไม่ได้ต้องข้อมูลสารสนเทศทั้งหมด จึงไม่มีความจำเป็นต้องออกแบบสคีมากลางให้มีความเบ็ดเสร็จตามข้อบังคับข้อแรก ประเด็นที่สอง ไม่มีความจำเป็นที่จะต้องแสดงข้อมูลในระดับสคีมากลางตรงตามความเป็นจริงที่อยู่ในแหล่งข้อมูลสารสนเทศทุกอย่าง เนื่องจากในบางครั้งความต้องการของระบบสารสนเทศเพียงเป็นแค่เพียงความต้องการในการคำนวณหรือสรุปข้อมูลเท่านั้น จึงไม่จำเป็นที่จะต้องเป็นไปตามข้อบังคับข้อที่สอง ในบางตัวอย่างนั้นอาจจะมีการรวบรวมกลุ่มลูกค้าตามฐานเงินเดือนที่ได้รับและไม่ต้องการให้เก็บเป็นรายได้ที่ได้รับโดยตรง สคีมากลางที่ออกแบบโดยวิธีนี้นั้นอาจจะสร้างมาจากการบูรณาการข้อมูลสารสนเทศตามวิธีทั่วๆไปที่มี หรืออาจจะมาจากกระบวนวิเคราะห์ข้อมูลอย่างเป็นทางการก็ได้ (Formal Analysis Process) หรืออาจจะสร้างจากแนวทางที่แนะนำไปในหัวข้อที่ 2.8.2.1 ก็ได้ ซึ่งการบูรณาการแหล่งข้อมูลสารสนเทศด้วยวิธีการทั่วไปนั้นจะไม่ได้สนใจความหมายข้อมูลแต่อย่างใด

เนื่องจากการออกแบบสคีมากลางด้วยวิธีการนี้เป็นการออกแบบตามความต้องการของระบบสารสนเทศรวมหรือกระบวนการวิเคราะห์ข้อมูลอย่างเป็นทางการ ทำให้มีข้อได้เปรียบหลายอย่าง เช่น ในสถานการณ์ที่แหล่งข้อมูลสารสนเทศบางแหล่งมีการเปลี่ยนแปลง

โครงสร้างสคีมาของตนค่อนข้างบ่อย หรือมีการนำเข้าแหล่งข้อมูลสารสนเทศบ่อยครั้ง หรือบางระบบสารสนเทศนั้นไม่สามารถใช้กระบวนการบูรณาการระบบสารสนเทศได้หรือมีค่าใช้จ่ายค่อนข้างสูง หรือความต้องการสารสนเทศของระบบเปลี่ยนไป [33] เป็นต้น

2.8.3.1 การออกแบบจากล่างขึ้นบน (Bottom-Up)

ในการออกแบบสคีมากลางด้วยวิธีการนี้นั้นจะเริ่มต้นจากความต้องการที่จะเข้าถึงอย่างบูรณาการไปยังข้อมูลสารสนเทศที่กำหนดเอาไว้ ซึ่งโดยทั่วไปแล้วนั้นผู้ออกแบบจะต้องทราบถึงรายละเอียดการใช้งานของระบบสารสนเทศของผู้ใช้ทั้งหมด เพื่อที่จะนำมาสร้างแอปพลิเคชั่นกลาง โดยความแตกต่างระหว่างการออกแบบทั้งสองวิธีก็คือ การออกแบบวิธีนี้นั้นจะเป็นการสร้างการบูรณาการการเข้าถึงแหล่งข้อมูลสารสนเทศทั้งหมดที่มีในฐานข้อมูล ในขณะที่การออกแบบบนลงล่างนั้นจะเป็นการเข้าถึงข้อมูลเพียงบางส่วนโดยยึดตามความต้องการของระบบสารสนเทศเป็นหลักและไม่จำเป็นต้องเข้าถึงข้อมูลทั้งหมดจากทุกแหล่งข้อมูลที่มีในระบบ วิธีการออกแบบจากล่างขึ้นบนนี้สคีมาที่ได้รับการออกแบบจะต้องสามารถรับประกันได้ว่ามีความสมบูรณ์และความถูกต้องของข้อมูลอยู่ด้วย และสคีมาที่ได้รับการบูรณาการนี้อาจจะมีการใช้เทคนิคการบูรณาการอย่างเป็นทางการเข้ามาร่วมด้วยในการออกแบบวิธีนี้ถ้าหากว่าแหล่งข้อมูลสารสนเทศแหล่งหนึ่งมีการเปลี่ยนแปลงสคีมาไปจากเดิมอาจจะนำไปสู่การบูรณาการใหม่อีกครั้งหนึ่ง

บทที่ 3
การวิเคราะห์และออกแบบ

ปัจจุบันมีมัลแวร์จำนวนมากที่แพร่ระบาดอยู่ในอินเตอร์เน็ตซึ่งนับวันจะมีจำนวนเพิ่มมากขึ้น ผู้สร้างมัลแวร์เหล่านี้มักจะคิดค้นเทคนิคการทำงานของมัลแวร์ที่แปลกใหม่ขึ้นมาอยู่เสมอ ดังนั้นการศึกษาหาความรู้เกี่ยวกับมัลแวร์จากหนังสืออย่างเดียวจะทำให้ผู้ใช้ได้ข้อมูลที่ไม่ทันสมัย จึงจำเป็นต้องศึกษาข้อมูลมัลแวร์ผ่านทางเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์ที่ซึ่งเชื่อถือได้ และมีการปรับปรุงข้อมูลให้ทันสมัยอยู่เสมอ เวปไซต์แหล่งข้อมูลคำอธิบายมัลแวร์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์นั้นมักจะกระจายกันอยู่บนอินเตอร์เน็ต ตลอดจนแหล่งข้อมูลมัลแวร์แต่ละแหล่งนั้นจะเรียกมัลแวร์ตัวเดียวกันด้วยชื่อที่แตกต่างกันออกไป ซึ่งนับเป็นอุปสรรคอย่างหนึ่งที่ส่งผลต่อผู้ใช้ในการศึกษาค้นคว้าข้อมูลเกี่ยวกับมัลแวร์

เพื่อเป็นการแก้ปัญหาดังกล่าววิจัยชิ้นนี้จึงได้ออกแบบระบบฐานความรู้มัลแวร์ที่บูรณาการคำอธิบายมัลแวร์จากแหล่งข้อมูลต่าง ๆ ที่มีอยู่บนอินเตอร์เน็ตมาเก็บไว้ภายใต้มาตรฐานเดียวกัน โดยมีโปรแกรมที่ทำหน้าที่รวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ (Malware Description Gathering Engine) มาเก็บไว้ในฐานข้อมูลที่ออกแบบไว้โดยอัตโนมัติอีกด้วย และเพื่อเป็นการแก้ไขปัญหาเรื่องชื่อเรียกที่ซ้ำซ้อนกันของมัลแวร์ งานวิจัยนี้จึงเสนอวิธีการกำหนดหมายเลขประจำตัวร่วมมัลแวร์ (Malware Common ID) เพื่อระบุว่ามัลแวร์ตัวใดบ้างที่เป็นตัวเดียวกันถึงแม้ว่ามัลแวร์เหล่านั้นจะถูกเรียกด้วยชื่อที่ต่างกันไปในแต่ละแหล่งข้อมูลก็ตาม ซึ่งถ้าหากว่ามีแหล่งข้อมูลคำอธิบายมัลแวร์แหล่งใหม่เกิดขึ้นบนอินเตอร์เน็ต ระบบก็อนุญาตให้มีการเพิ่มเติมแหล่งข้อมูลใหม่นั้นลงในระบบฐานความรู้มัลแวร์ได้ด้วยระบบลงทะเบียนแหล่งข้อมูลคำอธิบายมัลแวร์ (Information Source Profile Registering) ซึ่งระบบฐานความรู้มัลแวร์จะช่วยให้ผู้ใช้สามารถเข้ามาอ่านและศึกษาข้อมูลเกี่ยวกับมัลแวร์ต่าง ๆ ได้อย่างง่ายดาย รวดเร็ว และรอบด้าน พร้อมทั้งเปิดโอกาสให้ผู้ที่มีความเชี่ยวชาญด้านมัลแวร์สามารถแบ่งปันความรู้หรือประสบการณ์ในการแก้ปัญหาที่เกิดขึ้นจากมัลแวร์กับผู้ใช้ทั่วไปผ่านทางเวปไซต์ที่ได้ออกแบบไว้ และยังอนุญาตให้ผู้ใช้ที่ลงทะเบียนเป็นสมาชิกสามารถลงคะแนนโหวตให้กับคำอธิบายมัลแวร์ที่อ่านแล้วเห็นว่ามีประโยชน์หรือมีเนื้อหาค่อนข้างครอบคลุมและละเอียดได้ อีกทั้งผู้ใช้ยังสามารถแสดงความเห็น คำแนะนำ หรือวิธีการแก้ไขที่สามารถนำไปใช้แล้วเกิดประโยชน์ได้จริง ผ่านการโพสต์ข้อความในเวปไซต์ได้อีกด้วย

3.1 ปัญหาที่พบในงานวิจัย

3.1.1 ปัญหาความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneous)

- ปัญหาที่เกิดจากการใช้ชื่อแอทริบิวท์ที่แตกต่างกัน

โดยปรกติแล้วเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์นั้นจะมีทีมงานวิจัยมัลแวร์และพัฒนาเวปไซต์ของตนเอง ดังนั้นการกำหนดชื่อมัลแวร์และชื่อแอทริบิวท์ที่ไม่ได้มีการกำหนดมาตรฐานกลางเอาไว้ จึงมีผลให้ข้อมูลที่นำมาแสดงให้ผู้ใช้อ่านทางเวปไซต์จึงมีจำนวนแอทริบิวท์หรือชื่อแอทริบิวท์ที่แตกต่างกันออกไป ชื่อแอทริบิวท์ดังกล่าวนั้นแต่ละแหล่งข้อมูลจะใช้คำศัพท์เทคนิค (Technical Terms) ที่กำหนดขึ้นเองในการตั้งชื่อ ยกตัวอย่างเช่น การอ้างถึงประเภทของมัลแวร์ “Type” นั้น เวปไซต์ Kaspersky ใช้คำศัพท์เทคนิคตั้งชื่อแอทริบิวท์ว่า “Behavior” ในขณะที่เวปไซต์ Trend Micro Incorporated คำว่า “Malware Type” เป็นต้น ซึ่งพบว่าแม้การใช้คำศัพท์เทคนิคที่ต่างกันในการให้ข้อมูลมัลแวร์ของแหล่งข้อมูลต่าง ๆ นั้น อาจจะสร้างความสับสนให้แก่ผู้ใช้ได้เนื่องจากการใช้ศัพท์เทคนิคและคำจำกัดความที่แตกต่างกัน

จากปัญหาดังกล่าวนี้เมื่อมีการบูรณาการแหล่งข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มารวมไว้ภายในแหล่งข้อมูลเดียวกันแล้วอาจจะเกิดการบันทึกข้อมูลลงในตำแหน่งที่ไม่ตรงกับความหมายที่แท้จริงก็ได้ เพื่อเป็นการแก้ไขปัญหานี้ งานวิจัยนี้จึงได้ออกแบบเมตาดาต้าสำหรับเปรียบเทียบคำศัพท์ (Mapping Metadata) ซึ่งทำหน้าที่เปรียบเทียบคำศัพท์เทคนิคของแอทริบิวท์ที่แตกต่างกันไปในแต่ละเวปไซต์กับคำศัพท์สามัญที่ได้ถูกออกแบบไว้ เพื่อให้ระบบสามารถนำข้อมูลคำอธิบายมัลแวร์มาจัดเก็บลงในฐานความรู้มัลแวร์ให้ตรงกับความหมายที่แท้จริง โดยจะสร้างเป็นตารางเปรียบเทียบคำศัพท์เทคนิค (Technical Terms to General Terms Mapping Table) ดังแสดงในตารางที่ 3.1

ตารางที่ 3.1 เปรียบเทียบคำศัพท์เทคนิค

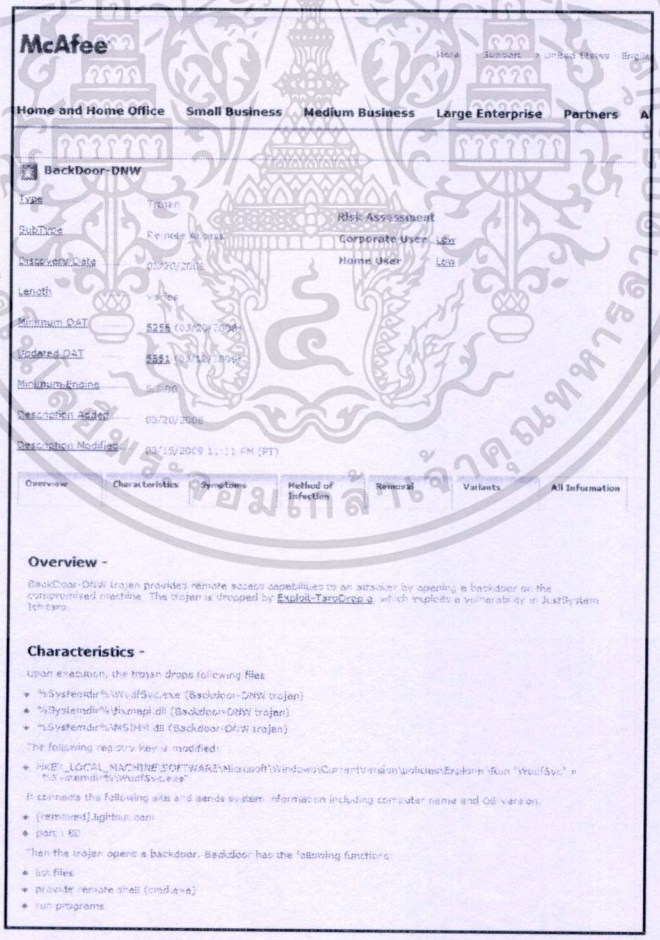
Technical term	Source	General term
Behavior	Kaspersky	Type
Type	Mcafee	Type
Malware type	Trend Micro Incorporated	Type
Type	Symantec Corporation	Type
Aliases	Kaspersky	Alias Name
Aliases	Mcafee	Alias Name
Aliases	Trend Micro Incorporated	Alias Name
Also Known As	Symantec Corporation	Alias Name
-	Kaspersky	Operation System

ตารางที่ 3.1 เปรียบเทียบคำศัพท์เทคนิค (ต่อ)

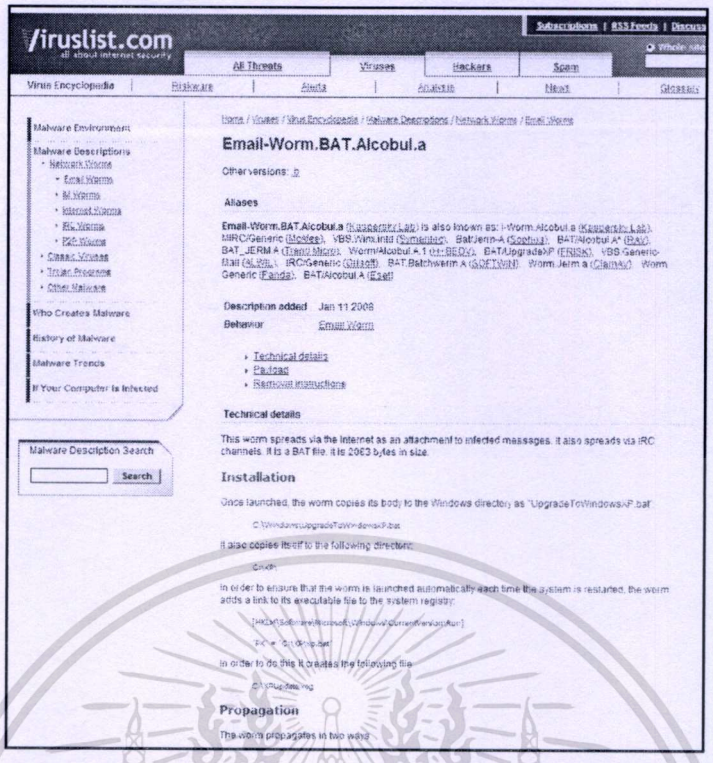
Technical term	Source	General term
-	Mcafee	Operation System
Platform	Trend Micro Incorporated	Operation System
Systems Affected	Symantec Corporation	Operation System
Systems Affected	Symantec Corporation	Operation System

- ปัญหาที่เกิดจากโครงสร้างของเวปไซต์ที่แตกต่างกัน

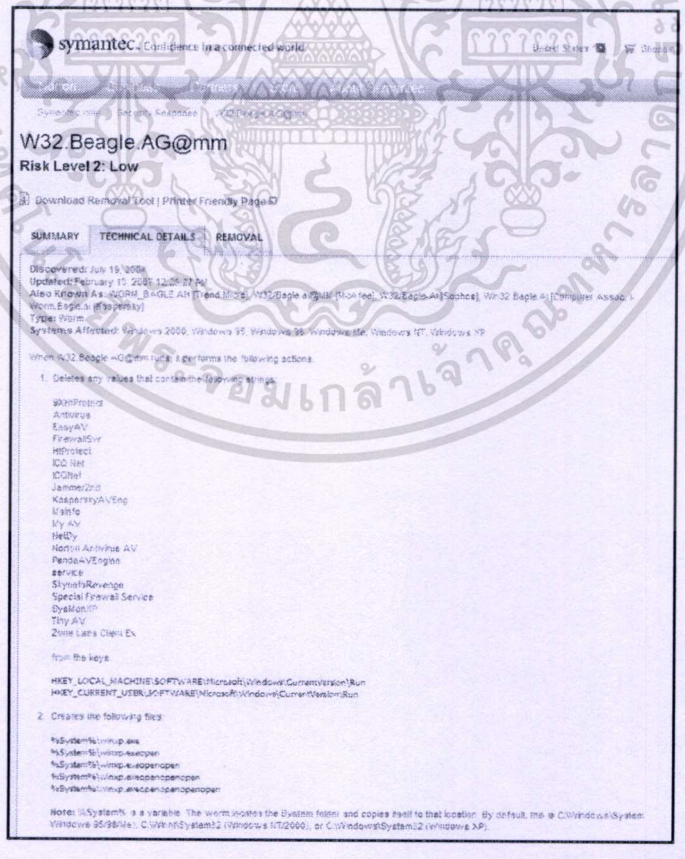
ในเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์แต่ละแห่งนั้น มักจะมีการนำเสนอข้อมูลคำอธิบายมัลแวร์ให้กับผู้ใช้ด้วยรูปแบบที่แตกต่างกันไป ด้วยสาเหตุนี้จึงทำให้โครงสร้างของเวปไซต์แต่ละแหล่งข้อมูลนั้นมีลักษณะที่ต่างกัน โดยตัวอย่างเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์ แสดงในรูปที่ 3.1 รูปที่ 3.2 และ รูปที่ 3.3



รูปที่ 3.1 ตัวอย่างคำอธิบายมัลแวร์จากเวปไซต์ Mcafee



รูปที่ 3.2 ตัวอย่างคำอธิบายมัลแวร์จากเวปไซต์ Kaspersky



รูปที่ 3.3 ตัวอย่างคำอธิบายมัลแวร์จากเวปไซต์ Symantec

นอกจากรูปแบบเวปไซต์ที่แตกต่างกันแล้ว โครงสร้างของเวปเพจในรูปแบบของภาษา HTML ในแต่ละเวปไซต์ก็แตกต่างกันไปด้วย กล่าวคือ ตำแหน่งจุดเริ่มต้นและจุดสิ้นสุดของข้อมูลแต่ละชุดนั้นแตกต่างกันออกไปในแต่ละเวปไซต์ จึงนับเป็นอุปสรรคสำคัญอย่างหนึ่งในการที่จะนำข้อมูลในเวปไซต์เหล่านี้มารวมกันไว้ในฐานข้อมูลมัลแวร์เดียวกันได้

เพื่อเป็นการแก้ไขปัญหาดังกล่าวนี้งานวิจัยนี้จึงได้ออกแบบเมตาดาต้าที่ใช้อธิบายรูปแบบการแสดงผล (Presentation Metadata) คำอธิบายมัลแวร์ที่แตกต่างกันออกไปในแต่ละเวปไซต์ ซึ่งจะทำหน้าที่เก็บข้อมูลโครงสร้างของเวปเพจในรูปแบบของภาษา HTML ของแต่ละเวปไซต์ โดยจะมีการเก็บจุดเริ่มต้น (Start point) และจุดสิ้นสุด (Stop point) ของชุดข้อมูลแต่ละส่วนที่มีในเวปไซต์ ซึ่งคำอธิบายข้อมูลที่เกี่ยวกับโครงสร้างภาษา HTML ของแต่ละเวปไซต์นั้นจะถูกบันทึกลงในตาราง ISProfile ซึ่งจะกล่าวถึงในส่วนของการออกแบบฐานข้อมูลต่อไป

3.1.2 ปัญหาความหลากหลายทางด้านความหมาย (Semantic Heterogeneous)

- ปัญหาด้านชื่อเรียกและความซ้ำซ้อนของมัลแวร์

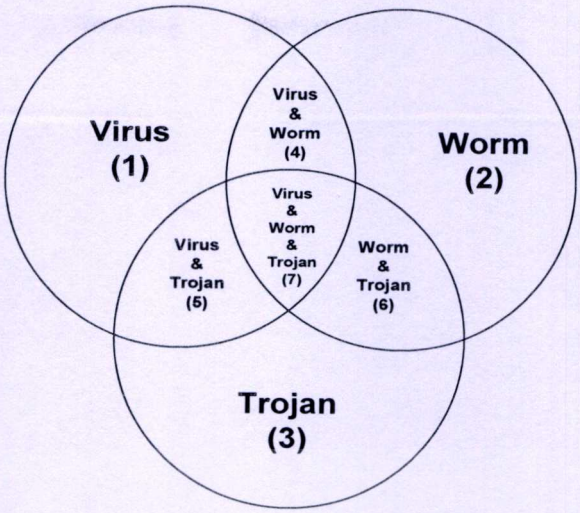
โดยปรกติแล้วมัลแวร์ใด ๆ เมื่อถูกค้นพบโดยผู้ผลิตโปรแกรมกำจัดมัลแวร์แล้ว ผู้ค้นพบมักจะตั้งชื่อมัลแวร์ดังกล่าวขึ้นมาเอง ซึ่งวิธีการนี้มักทำให้เกิดปัญหาในการอ้างอิงถึงมัลแวร์แต่ละตัว เนื่องจากมัลแวร์ตัวเดียวกันอาจถูกเรียกด้วยชื่อที่แตกต่างกันไปในแต่ละแหล่งข้อมูลได้ ดังนั้นในคำอธิบายมัลแวร์แต่ละตัวมักจะมีแอทริบิวท์ที่เป็นชื่อเรียก (Alias name) ของมัลแวร์ เพื่อเป็นการอ้างอิงถึงข้อมูลของมัลแวร์ดังกล่าวในแหล่งข้อมูลอื่น ๆ ยกตัวอย่างเช่น มัลแวร์ชื่อ “PWS-Gamania.gen.a” เป็นชื่อที่ใช้กันในเวปไซต์ Mcafee แต่เวปไซต์ Kaspersky จะเรียกว่า “Trojan.Win32.Vaklik.bkh” ส่วนเวปไซต์ Symantec จะเรียกว่า “W32.Gammima.AG” เป็นต้น จากปัญหาเหล่านี้จึงต้องมีวิธีการแยกแยะและจัดกลุ่มข้อมูลของมัลแวร์ตัวเดียวกันเอาไว้ด้วยกัน เพื่อให้ผู้ใช้ทราบว่ามัลแวร์ตามชื่อเรียกเหล่านั้นที่จริงแล้วเป็นมัลแวร์ตัวเดียวกัน

เพื่อเป็นการแก้ปัญหาดังกล่าวผู้เขียนจึงออกแบบหมายเลขประจำตัวมัลแวร์ (Malware Common ID: MCID) ขึ้นมาใช้เพื่ออ้างอิงมัลแวร์ที่มีหลายชื่อภายใต้หมายเลขประจำตัวมัลแวร์เดียวกัน กล่าวคือ ก่อนจะเพิ่มข้อมูลคำอธิบายมัลแวร์ลงในฐานข้อมูล ระบบจะตรวจสอบดูก่อนว่ามัลแวร์ตัวใหม่นี้มีชื่อเรียกตรงกับมัลแวร์ที่อยู่ในฐานข้อมูลก่อนหน้านั้นหรือไม่ ถ้ามีก็จะทำการกำหนดให้ใช้หมายเลขประจำตัวมัลแวร์เดียวกับมัลแวร์ที่มีอยู่แล้วในฐานข้อมูล ซึ่งวิธีการทำงานของระบบจะกล่าวไว้โดยละเอียดในส่วนของการออกแบบระบบ

- ปัญหาการตัดสินคุณสมบัติของมัลแวร์ที่แตกต่างกัน

โดยปรกติแล้วคำอธิบายมัลแวร์ที่มาจากแต่ละแหล่งข้อมูลนั้นจะมีการบอกถึงประเภทของมัลแวร์มาด้วยเสมอ ซึ่งบางแหล่งข้อมูลจะแสดงชื่อประเภทของมัลแวร์แตกต่างกันออกไป ถึงแม้ว่าจะเป็นมัลแวร์ประเภทเดียวกันก็ตาม เช่น แหล่งข้อมูลคำอธิบายมัลแวร์บางแหล่งเลือกที่จะใช้ชื่อคำว่า “ไวรัส” ในการเรียกมัลแวร์ประเภทต่าง ๆ เนื่องจากผู้ใช้คอมพิวเตอร์ส่วนใหญ่จะรู้จัก “ไวรัส” มากกว่าหนอนคอมพิวเตอร์และโทรจัน และมักจะเรียกมัลแวร์ที่พบในเครื่องคอมพิวเตอร์ของตนว่า “ไวรัส” ทำให้คำว่า “ไวรัส” กลายเป็นสัญลักษณ์แทนมัลแวร์ที่มีอยู่ทั่วไป วิธีนี้ไม่สามารถสร้างความรู้ความเข้าใจเกี่ยวกับมัลแวร์ที่ถูกต้องให้กับผู้ใช้ได้ เนื่องจากผู้ใช้จะเข้าใจว่ามัลแวร์ที่มีอยู่นั้นมีแต่ประเภทที่เป็นไวรัสและไม่ทราบถึงข้อแตกต่างระหว่างมัลแวร์ประเภทอื่น ๆ

ในปัจจุบันนี้มีการพัฒนาความสามารถของมัลแวร์ให้มีความหลากหลายมากขึ้น ซึ่งแตกต่างจากในอดีตที่มัลแวร์แต่ละตัวจะมีคุณสมบัติได้แค่หนึ่งประเภทเท่านั้น โดยมัลแวร์ที่มีคุณสมบัติประเภทเดียวนั้นจะแทนด้วยเซตของบริเวณที่ 1, 2 และ 3 ในเวนน์ไดอะแกรมในรูปที่ 3.4 ตัวอย่างของมัลแวร์ที่มีคุณสมบัติประเภทเดียวได้แก่ Jerusalem, Michelangelo, Stone ตามลำดับ เซตของบริเวณที่ 4, 5, 6 จะแสดงถึงมัลแวร์ที่มีคุณสมบัติสองประเภท ยกตัวอย่างเช่น W32/Zaflen.a นั้นสามารถแพร่เชื้อกระจายตนเองไปยังโฮสต์ไฟล์ในเครื่องและแพร่กระจายตนเองไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ได้ ส่วน W32/Conficker.worm.gen.b สามารถขโมยบัญชีและรหัสผ่านอีเมล์ของผู้ใช้แล้วส่งข้อมูลผ่านอินเตอร์เน็ตกลับไปให้ผู้สร้างมัลแวร์ และเซตของบริเวณที่ 7 จะแสดงถึงมัลแวร์ที่มีคุณสมบัติสามประเภท ยกตัวอย่างเช่น W32/Cekar สามารถขโมยบัญชีและรหัสผ่านอีเมล์ของผู้ใช้แล้วส่งข้อมูลผ่านทางอินเตอร์กลับไปให้ผู้สร้างมัลแวร์และแพร่กระจายไปยังโฮสต์ไฟล์ต่าง ๆ ในเครื่องของผู้ใช้ได้อีกด้วย ซึ่งคำอธิบายมัลแวร์จากบางแหล่งข้อมูลมักจะกล่าวถึงประเภทของมัลแวร์เพียงประเภทเดียวเท่านั้น ทั้งที่ในความเป็นจริงแล้วมัลแวร์หนึ่งตัวนั้นมีได้หลายคุณสมบัติ ทำให้ผู้ใช้มีความเข้าใจที่คลาดเคลื่อนจากความเป็นจริงได้



รูปที่ 3.4 แผนภาพแสดงมัลแวร์แต่ละประเภท

เพื่อเป็นการแก้ไขปัญหาดังกล่าวควรจะต้องมีโปรแกรมที่ทำหน้าที่ตัดสินคุณสมบัติของมัลแวร์ภายใต้มาตรฐานเดียวกันเพื่อให้ผู้ใช้เข้าใจถึงคุณสมบัติของมัลแวร์ได้อย่างตรงกัน โดยรายละเอียดของโปรแกรมวิเคราะห์คุณสมบัติของมัลแวร์นั้นจะกล่าวถึงในหัวข้อการออกแบบ

- ปัญหาการประเมินภัยคุกคามจากมัลแวร์ที่แตกต่างกัน

โดยปรกติแล้วแหล่งข้อมูลคำอธิบายมัลแวร์แต่ละแหล่งนั้นจะใช้มาตรฐานในการวิเคราะห์ภัยคุกคามที่เกิดจากมัลแวร์ด้วยมาตรฐานที่แตกต่างกันออกไป เพื่อเป็นการแก้ไขปัญหาดังกล่าว งานวิจัยนี้จึงออกแบบโปรแกรมวิเคราะห์ภัยคุกคามที่เกิดจากมัลแวร์ ซึ่งจะทำหน้าที่ประเมินระดับความเสียหายที่เกิดขึ้นจากมัลแวร์ (Damage Level) ระดับการแพร่กระจายของมัลแวร์ (Distribution Level) สถานการแพรกระจายของมัลแวร์ (Distribution Status) และระดับความเสี่ยงที่มีผลต่อผู้ใช้ (Risk Level) ซึ่งรายละเอียดจะกล่าวถึงในหัวข้อการออกแบบระบบ

3.2 การออกแบบฐานข้อมูล

เนื่องจากแหล่งข้อมูลคำอธิบายมัลแวร์ที่มีอยู่บนเวปไซต์ต่าง ๆ นั้นมีการออกแบบโครงสร้างฐานข้อมูลอย่างอิสระ จึงทำให้เกิดปัญหาในการบูรณาการแหล่งข้อมูลเหล่านั้นไว้ในระบบฐานความรู้มัลแวร์เดียวกัน

เพื่อเป็นการแก้ไขปัญหาดังกล่าวนี้ เราได้นำหลักการออกแบบสคีมากลางด้วยวิธีบนลงล่างเข้ามาช่วยแก้ปัญหา กล่าวคือ แนวทางการออกแบบลักษณะนี้จะเป็นการวิเคราะห์ตามความต้องการของระบบสารสนเทศที่มีความต้องการแตกต่างกันออกไป ซึ่งในงานวิจัยนี้เราได้ทำการ

วิเคราะห์พฤติกรรมและการทำงานของมัลแวร์ทั้ง 3 ประเภท คือ ไวรัส หนอนคอมพิวเตอร์ และโทรจัน และยังสามารถแบ่งประเภทของมัลแวร์ชนิดต่าง ๆ ลงไปตามลักษณะการทำงานของมัลแวร์แต่ละชนิด เช่น ไวรัสสามารถแบ่งออกได้ตามโฮสต์ไฟล์ที่มันอาศัยอยู่ หนอนคอมพิวเตอร์สามารถแบ่งออกตามประเภทของการแพร่กระจาย และโทรจันนั้นสามารถแบ่งออกตามจุดประสงค์ในการสร้างความเสียหาย เราได้นำข้อมูลที่ได้วิเคราะห์จากบทที่ 2 มาร่วมออกแบบโครงสร้างฐานข้อมูลด้วยเช่นกัน และยังได้ทำการศึกษาคำอธิบายมัลแวร์จากแหล่งข้อมูลต่าง ๆ ซึ่งประกอบด้วย NOD32 [34] Trend Micro [35] F-Secure [36] Sophos [37] Symantec [38] Mcafee [39] และ Kaspersky [40] เราพบว่าแม้คำอธิบายเหล่านั้นจะมีรูปแบบการแสดงผลหรือจำนวนแอทริบิวท์ที่แตกต่างกันไป แต่ก็ยังคงมีบางส่วนที่เหมือนกันอยู่เช่นกัน โดยเราสามารถจำแนกข้อมูลคำอธิบายมัลแวร์จากแหล่งข้อมูลต่าง ๆ ออกเป็น 6 ส่วนได้แก่

- **วันที่ (Date)**

  ในข้อมูลของคำอธิบายมัลแวร์จะต้องมีวันที่ซึ่งจะเป็นตัวบอกถึงวันที่มีความสำคัญเกี่ยวกับมัลแวร์ เช่น วันที่ค้นพบมัลแวร์ (Discovery date) วันที่คำอธิบายได้ออกเผยแพร่เป็นครั้งแรก (First Published date) วันที่คำอธิบายได้ถูกแก้ไขล่าสุด (Modified date) เป็น

- **สถานะ (Status)**

  โดยปรกติแล้วเมื่อนักวิจัยแต่ของผู้ผลิตโปรแกรมกำจัดมัลแวร์แต่ละแห่งนั้น ได้วิเคราะห์ตัวอย่างมัลแวร์เสร็จแล้ว จะมีการประเมินมัลแวร์ตัวนั้นในหลาย ๆ ด้าน ไม่ว่าจะเป็นด้านความเสียหายที่ส่งผลกระทบต่อผู้ใช้ ความรวดเร็วในการแพร่กระจาย ความเสี่ยงที่มีต่อผู้ใช้ เป็นต้น ซึ่งค่าสถานะที่ได้เหล่านี้จะช่วยให้ผู้ใช้สามารถประเมินสถานการณ์ได้ว่า มัลแวร์ตัวดังกล่าวนั้นมีการแพร่ระบาดอยู่จริงหรือไม่ มีระดับการสร้างความเสียหายสูงหรือไม่ เพื่อเป็นการระวังและป้องกันตนเองจากมัลแวร์นั้น

- **รายละเอียดทางเทคนิค (Technical Detail)**

  หมายถึงข้อมูลทางเทคนิคที่ทางเวปไซต์ผู้ผลิตโปรแกรมกำจัดมัลแวร์อธิบายถึงรายละเอียดของมัลแวร์เพื่อให้ผู้ใช้มีความเข้าใจเกี่ยวกับมัลแวร์ที่ดีขึ้น ยกตัวอย่างเช่น

  - **ประเภทของมัลแวร์ (Type)** เป็นข้อมูลที่บอกผู้ใช้ถึงประเภทของมัลแวร์ว่ามัลแวร์ตัวดังกล่าวนั้นจัดอยู่ในประเภท ไวรัส หนอนคอมพิวเตอร์ หรือโทรจัน
  - **ระบบปฏิบัติการ (Operation System)** เป็นชนิดของระบบปฏิบัติการที่มัลแวร์สามารถทำงานได้ เช่น ระบบปฏิบัติการวินโดว์ (Windows), ระบบปฏิบัติการดอส (DOS), ระบบปฏิบัติการแมคอินทอช (Macintosh) เป็นต้น

- **ชื่อเรียกมัลแวร์ (Alias Name)** เป็นชื่อของมัลแวร์ที่ถูกเรียกโดยแหล่งคำอธิบายมัลแวร์แหล่งอื่น ยกตัวอย่างเช่น มัลแวร์ชื่อ "PWS-Gamania.gen.a" ถูกค้นพบโดยเวปไซต์ของ Mcafee แต่เวปไซต์ของ Kaspersky จะเรียกมัลแวร์ดังกล่าวว่า "Trojan.Win32.Vaklik.bkh" และเวปไซต์ Symantec จะเรียกมัลแวร์ดังกล่าวว่า"W32.Gammima.AG" เป็นต้น
- **รุ่นของโปรแกรมที่สามารถกำจัดมัลแวร์ได้ (Version of Anti-Malware)** โดยปรกติแล้วในเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์นั้น มักจะบอกรุ่นของโปรแกรมกำจัดมัลแวร์ขั้นต่ำสุดที่สามารถกำจัดมัลแวร์ได้ เพื่อให้ผู้ใช้ทราบถึงรุ่นของโปรแกรมซึ่งเพียงพอต่อการนำมาใช้กำจัดมัลแวร์ออกจากเครื่องคอมพิวเตอร์ได้

- **พฤติกรรมและการทำงานของมัลแวร์ (Routines)**

ส่วนนี้นับได้ว่าเป็นส่วนที่มีความสำคัญส่วนหนึ่งในคำอธิบายของมัลแวร์ ดังที่กล่าวไปแล้วในบทที่ 2 เราสามารถแยกพฤติกรรมและการทำงานของมัลแวร์ออกได้เป็น 3 ส่วนด้วยกัน กล่าวคือมัลแวร์จะต้องมีส่วนการติดตั้ง (Installation routine) ซึ่งจะติดตั้งตนเองลงในเครื่องของผู้ใช้ จากนั้นมัลแวร์จะแพร่กระจายไปยังไฟล์ต่าง ๆ หรือเครื่องคอมพิวเตอร์เครื่องอื่นก็แล้วแต่ที่มัลแวร์นั้นได้ถูกออกแบบมา (Infection and Distribution routine) และส่วนที่สร้างความเสียหายให้กับผู้ใช้ (Damage routine)

- **วิธีการแก้ไขปัญหาจากมัลแวร์ (Solution)**

เป็นส่วนที่แต่ละแหล่งข้อมูลแนะนำหรืออธิบายวิธีการแก้ไขหรือกำจัดมัลแวร์ออกไปจากเครื่องคอมพิวเตอร์ของผู้ใช้ ซึ่งอาจจะเป็นการแนะนำให้ผู้ใช้ได้ทำการอัพเดทโปรแกรมให้เป็นเวอร์ชั่นปัจจุบันซึ่งจะสามารถกำจัดมัลแวร์ดังกล่าวออกไปได้ หรือเป็นการสอนให้ผู้ใช้สามารถกำจัดมัลแวร์ในเครื่องของผู้ใช้ด้วยตนเอง เป็นต้น

- **รายละเอียดของมัลแวร์ (Malware Details)**

ส่วนสุดท้ายนี้เป็นส่วนที่เพิ่มขึ้นมาเพื่อแก้ไขปัญหาความหลากหลายทางด้านโครงสร้างตามที่กล่าวไปในหัวข้อที่ 3.1.1 กล่าวคือ ในกรณีที่แหล่งข้อมูลมัลแวร์ที่นำเข้ามาในระบบนั้นมีจำนวนแอทริบิวท์มากกว่าระบบฐานความรู้มัลแวร์ ก็สามารถที่จะนำข้อมูลในแอทริบิวท์นั้นบันทึกลงในแอทริบิวท์ในส่วนนี้ของระบบได้ ดังนั้นแม้ว่าแหล่งข้อมูลสารสนเทศใหม่ที่นำเข้ามาในระบบนั้นจะมีจำนวนแอทริบิวท์ที่ไม่เท่ากัน แต่ก็สามารถที่จะเก็บบันทึกข้อมูลของแหล่งข้อมูลนั้น ๆ ลงในระบบฐานความรู้มัลแวร์ได้อย่างครบถ้วนซึ่งเป็นประโยชน์ต่อการศึกษาข้อมูลกับผู้ใช้

3.2.1 การออกแบบชื่อศัพท์คำสามัญ (General Terms)

จากปัญหาที่เกิดจากการใช้ชื่อแอทริบิวท์ที่แตกต่างกันในแต่ละแหล่งข้อมูลนั้น อาจจะทำให้ข้อมูลที่บันทึกลงในระบบฐานความรู้มัลแวร์นั้นไม่ตรงกับความหมายที่แท้จริง เนื่องจากแต่ละเวปไซต์นั้นใช้ชื่อแอทริบิวท์ของตนเอง ดังนั้นเพื่อเป็นการแก้ไขปัญหาและให้การสื่อความหมายเป็นไปได้อย่างชัดเจนยิ่งขึ้น เราจึงได้ออกแบบคำศัพท์สามัญ (General Terms) ขึ้นมาเพื่อให้ผู้ใช้สามารถอ่านแล้วเข้าใจถึงความหมายของคำศัพท์นั้นได้ในทันที ซึ่งคำศัพท์สามัญที่สื่อความหมายเหล่านี้จะถูกนำไปตั้งชื่อแอทริบิวท์ในระบบฐานความรู้มัลแวร์ โดยตารางที่ 3.2 จะแสดงคำศัพท์สามัญที่เราได้ทำการออกแบบไว้

ตารางที่ 3.2 คำศัพท์สามัญ (General Terms)

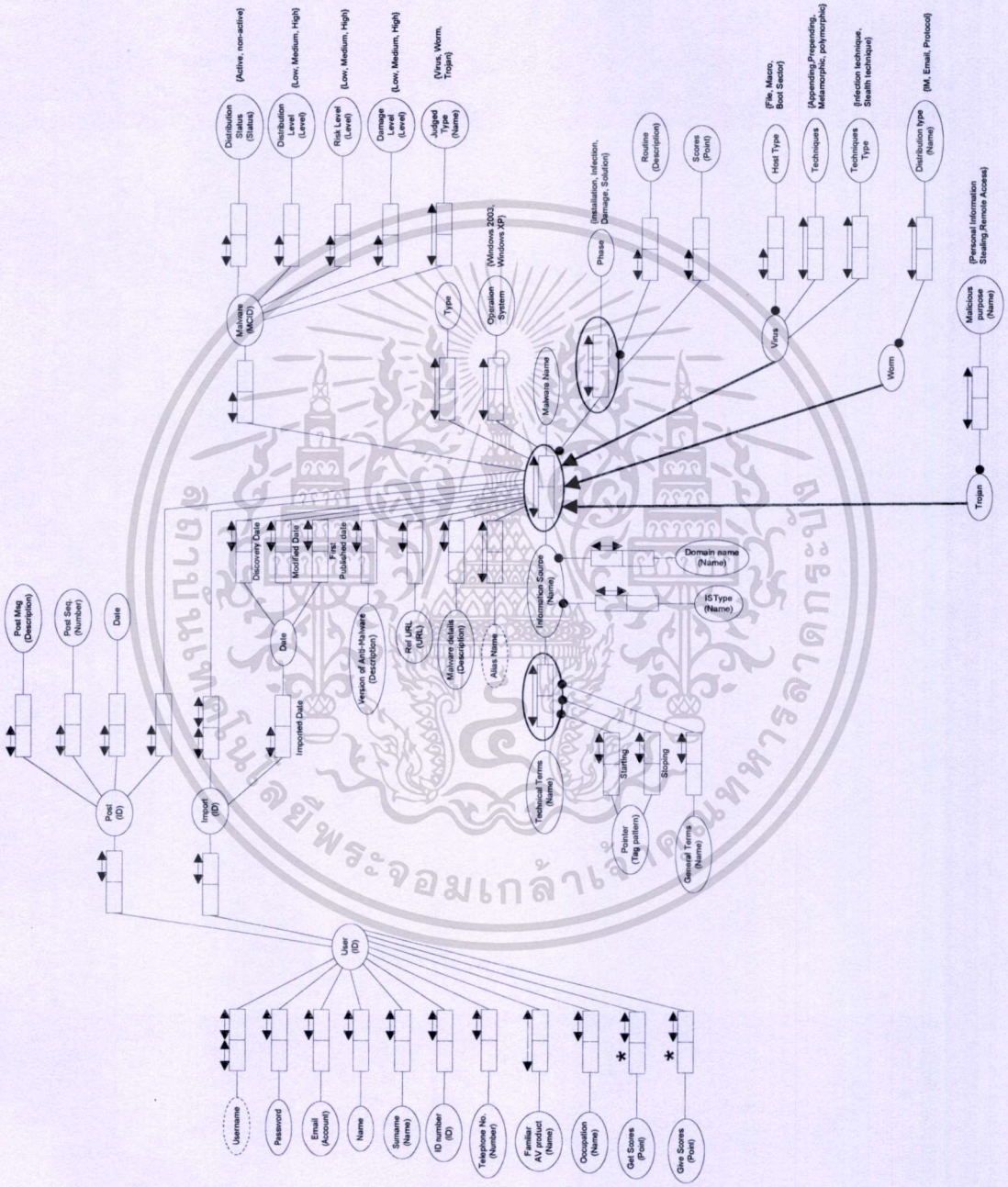
ชื่อคำสามัญ (General Terms)	**ความหมาย**	**ตัวอย่าง**
Name	ชื่อมัลแวร์	Trojan-Downloader.JS.Small.fi
Source	แหล่งข้อมูลคำอธิบายมัลแวร์	Mcafee, Symantec, Kaspersky
Discovery date	วันที่ค้นพบมัลแวร์	2007-07-04
First Published date	วันแรกที่คำอธิบายได้ถูกเผยแพร่	2008-10-29
Modified date	วันที่คำอธิบายได้ถูกแก้ไขล่าสุด	2008-11-29
Distribution Status	สถานะการแพร่ระบาดในปัจจุบัน	Active, Non-Active
Distribution Level	ระดับการแพร่กระจาย	Low, Medium, High
Damage Level	ระดับความเสียหาย	Low, Medium, High
Risk Level	ระดับความเสี่ยง	Low, Medium, High
Type	ประเภทของมัลแวร์	Virus, Worm, Trojan
Operation System	ระบบปฏิบัติการ	Windows, DOS, Macintosh
Alias Name	ชื่อเรียกหรือนามแฝงของมัลแวร์	Blood_Sugar.416
Version of Anti-Malware	รุ่นของโปรแกรมที่สามารถกำจัดมัลแวร์ได้	5287 (05/02/2008)

ตารางที่ 3.2 ตารางคำศัพท์สามัญ (General Terms) ต่อ

ชื่อคำสามัญ (General Terms)	**ความหมาย**	**ตัวอย่าง**
Installation routine	รูทีนการติดตั้ง	The Trojan copies its executable file to the Windows system directory: …
Infection and Distribution routine	รูทีนการแพร่กระจาย	When infecting a computer, the worm launches an HTTP server on a random TCP port. This is then used to load the worm's executable file to other computers. …
Damage routine	รูทีนการสร้างความเสียหาย	When launching, the worm injects its code into the address space of one of the "svchost.exe" system processes. This code is responsible for the worm's malicious payload:
Solution	วิธีการแก้ไข	Delete the system registry key shown below:: [HKLM\SYSTEM\CurrentControl Set\Services\netsvcs]
Malware details	รายละเอียดของมัลแวร์	This is a virus detection. Viruses are programs that self-replicate recursively, meaning …

3.2.2 แบบจำลองไนแอม (NIAM Model)

หลังจากที่ได้ทำการวิเคราะห์ข้อมูลมัลแวร์จากแหล่งข้อมูลต่าง ๆแล้วเราสามารถออกแบบโครงสร้างฐานข้อมูลโดยใช้แบบจำลองไนแอมได้ดังนี้



รูปที่ 3.4 แผนภาพไนแอม

3.2.1 ตารางข้อมูล

ตารางที่ 3.3 ข้อมูลผู้ใช้ (ชื่อตาราง User)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Username	Varchar(10)	PK	ชื่อผู้ใช้	Somchai
Password	Varchar(32)		รหัสผ่าน	p@ssw0rd
Email	Varchar(50)		อีเมล์	abc@hotmail.com
Name	Varchar(50)		ชื่อ	Somchai
Surname	Varchar(50)		นามสกุล	Sudjai
ID number	Varchar(15)		เลขบัตรประชาชน	012345678912345
Telephone No.	Varchar(20)		หมายเลขโทรศัพท์	012-483-4879
Occupation	Varchar(50)		อาชีพ	Engineer, student
Get Scores	Int(5)		คะแนนที่ได้รับ	15
Give Scores	Int(5)		คะแนนที่ให้	5

ตารางที่ 3.4 ข้อมูลโปรแกรมกำจัดมัลแวร์ที่ใช้เป็นประจำ (ชื่อตาราง AV Product)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Username	Varchar(10)	PK	ชื่อผู้ใช้	Somchai
Familiar AV product	Varchar(50)	FK	โปรแกรม Anti-malware ที่ใช้เป็นประจำ	Symantec, Mcafee

ตารางที่ 3.5 รายละเอียดข้อมูลการโพสต์ (ชื่อตาราง Post_Detail)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Post ID	Int(5)	PK	หมายเลขการโพสต์	12345
Username	Varchar(10)		ชื่อผู้ใช้	Somchai
Post Seq	Int(5)		ลำดับที่โพสต์	9
Post Msg	Text		ข้อความที่โพสต์	"You can manually remove this worm.."
Date	Date		เวลาที่โพสต์	2008-11-30
Information Source	Varchar(50)		ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky

ตารางที่ 3.5 ตารางรายละเอียดข้อมูลการโพสต์ (ชื่อตาราง Post_Detail) (ต่อ)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Malware Name	Varchar(50)		ชื่อมัลแวร์	Trojan-Downloader..

ตารางที่ 3.6 รายละเอียดการนำเข้าข้อมูลมัลแวร์ (ชื่อตาราง Import_Detail)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Import ID	Int(5)	PK	หมายเลขการโพสต์	12345
Information Source	Varchar(50)	FK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar(50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Username	Varchar(10)		ชื่อผู้ใช้	Somchai
Date	Date		วันที่ที่นำเข้าข้อมูล	2008-11-30

ตารางที่ 3.7 ข้อมูลลงทะเบียนของแหล่งข้อมูล (Information Source Registered data) (ชื่อตาราง ISProfile)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Technical Terms	Varchar(50)	FK	ชื่อคำศัพท์เทคนิค	Behavirous, Also Known as, Payload.
General Terms	Varchar(50)		ชื่อคำศัพท์สามัญ	Name, Alias Name
Start Pointer	Varchar(255)		แท็ค HTML ที่บอกตำแหน่งจุดเริ่มต้นของชุดข้อมูล	<td nowrap class enc…" /td>..
Stop Pointer	Varchar(255)		แท็ค HTML ที่บอกตำแหน่งจุดสิ้นสุดของชุดข้อมูล	<table class= "enc_tbl" border= "0"…

ตารางที่ 3.10 รายละเอียดประเภทของมัลแวร์ (ชื่อตาราง Malware_Type)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar(50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Type	Varchar(50)		ประเภทของมัลแวร์	Trojan, Virus, Worm

ตารางที่ 3.11 รายละเอียดระบบปฏิบัติการ (ชื่อตาราง Malware_OS)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar(50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Operation System	Varchar(50)		ระบบปฏิบัติการ	Windows Server 2003, Windows XP.

ตารางที่ 3.12 รายละเอียดชื่อเรียกมัลแวร์ (ชื่อตาราง Malware_Alias_name)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar(50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Alias Name	Varchar(50)		ชื่อเรียกมัลแวร์	Blood_Sugar.416

ตารางที่ 3.13 รายละเอียดรูทีนการทำงานของมัลแวร์ (ชื่อตาราง Malware_Malware_routine)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar(50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Phase	Varchar(50)		ชื่อเรียกมัลแวร์	Installation, Damage,…
Routine	Text		รูทีนการทำงานของมัลแวร์	The Trojan copies its executable file to the

ตารางที่ 3.13 รายละเอียดรูทีนการทำงานของมัลแวร์ (ชื่อตาราง Malware_Malware_routine) (ต่อ)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Scores	Int(5)		คะแนนที่ของส่วนรูทีนที่ได้รับการโหวต	25

ตารางที่ 3.14 รายละเอียดของมัลแวร์ร่วม (ชื่อตาราง MCID_details)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
MCID	Int(5)	PK	หมายเลขประจำตัวร่วมมัลแวร์	143
Distribution Status	Varchar(10)		สถานะการแพร่กระจาย	Active, non-active
Distribution Level	Varchar(10)		ระดับการแพร่กระจาย	Low, Medium, High
Risk Level	Varchar(10)		ระดับความเสี่ยงที่มีผลต่อผู้ใช้	Low, Medium, High
Damage Level	Varchar(10)		ระดับความเสียหาย	Low, Medium, High
Information Source	Varchar(50)		ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar(50)		ชื่อมัลแวร์	Trojan-Downloader..

ตารางที่ 3.15 รายละเอียดของประเภทของมัลแวร์ที่ถูกประเมินแล้ว (ชื่อตาราง Judged Type)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
MCID	Int(5)	PK	หมายเลขของมัลแวร์	143
Judged Type	Varchart (10)	FK	ประเภทของมัลแวร์ที่ได้รับการประเมินแล้ว	Virus, worm, Trojan

ตารางที่ 3.16 รายละเอียดโฮสต์ของไวรัส (Virus_Host_details)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar (50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Host Type	Varchar (50)		โฮสต์ของไวรัส	File, Boot sector,…

ตารางที่ 3.17 รายละเอียดเทคนิคของไวรัส (Virus_Technique_details)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar (50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Techniques	Varchar (50)	FK	เทคนิคของไวรัส	Appending, Prepending,..

ตารางที่ 3.18 รายละเอียดประเภทของเทคนิคไวรัส (Virus_Technique_Type_details)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar (50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Techniques Type	Varchar (50)	FK	ประเภทเทคนิคของไวรัส	Infection technique, Stealth technique.

ตารางที่ 3.19 รายละเอียดประเภทการแพร่กระจายของหนอนคอมพิวเตอร์ (Worm_Distribution_details)

ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar (50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Distribution Type	Varchar (50)	FK	ประเภทการแพร่กระจายของหนอนคอมพิวเตอร์	IM, Email, Network, Removable storages device.

ตารางที่ 3.20 รายละเอียดจุดประสงค์ที่มุ่งร้ายของโทรจัน (Trojan_Malicious_Purpose_details)

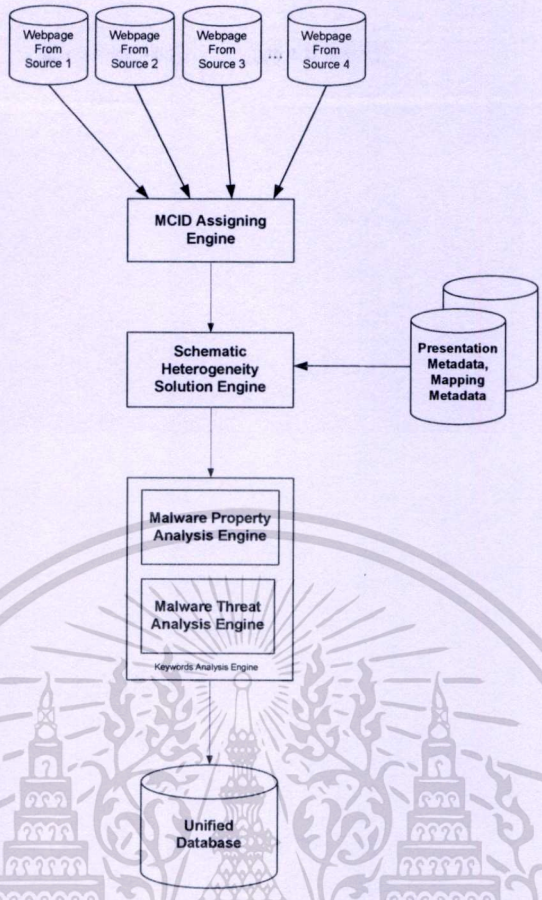
ชื่อ	ชนิดข้อมูล	คีย์	ความหมาย	ตัวอย่าง
Information Source	Varchar(50)	PK	ชื่อแหล่งข้อมูล	Symantec, Mcafee, Kaspersky
Malware Name	Varchar (50)	FK	ชื่อมัลแวร์	Trojan-Downloader..
Malicious purpose	Varchar (50)	FK	จุดประสงค์ที่มุ่งร้ายของโทรจัน	Personal information stealing, remote access,..

3.3 การออกแบบการทำงานของระบบ

3.3.1 โปรแกรมรวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ (Malware Description Gathering Engine)

งานวิจัยนี้ได้ออกแบบโปรแกรมรวบรวมข้อมูลคำอธิบายมัลแวร์ที่ทำหน้าที่รวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาเก็บลงในฐานความรู้มัลแวร์โดยอัตโนมัติ หลังจากป้อน URL ของหน้าเวปไซต์คำอธิบายมัลแวร์ ซึ่งโปรแกรมจะทำหน้าที่นำข้อมูลคำอธิบายแอทริบิวท์ของมัลแวร์มาใช้ในการวิเคราะห์ เพื่อแยกแยะข้อมูลและบันทึกลงในฐานความรู้มัลแวร์ได้อย่างถูกต้องตามความหมายอย่างอัตโนมัติ

โปรแกรมรวบรวมข้อมูลคำอธิบายมัลแวร์นั้นประกอบไปด้วยโปรแกรมย่อย 4 ส่วน คือ หนึ่ง โปรแกรมกำหนดหมายเลขประจำตัวร่วมมัลแวร์ (MCID Assigning Engine) ซึ่งทำหน้าที่ตรวจสอบถึงความซ้ำซ้อนของคำอธิบายมัลแวร์ที่เข้ามาในระบบ ค้นหาข้อมูลของมัลแวร์เดียวกันในต่างเวปไซต์ และกำหนดหมายเลขประจำตัวร่วมมัลแวร์ สอง โปรแกรมแก้ไขปัญหาด้านความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneous Solution Engine) ซึ่งทำหน้าที่แก้ปัญหาความหลากหลายที่เกิดขึ้นจากการบูรณาการข้อมูล สาม โปรแกรมวิเคราะห์คีย์เวิร์ดในคำอธิบายมัลแวร์ (Malware Properties Analysis Engine) ทำหน้าที่วิเคราะห์คีย์เวิร์ดที่มีในคำอธิบายมัลแวร์เพื่อนำมาจำแนกประเภทของมัลแวร์ และสุดท้ายคือ โปรแกรมประเมินภัยคุกคามที่เกิดจากมัลแวร์ (Malware Threat Analysis Engine) โดยภาพรวมของการทำงานของระบบรวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ นั้นแสดงในรูปที่ 3.5



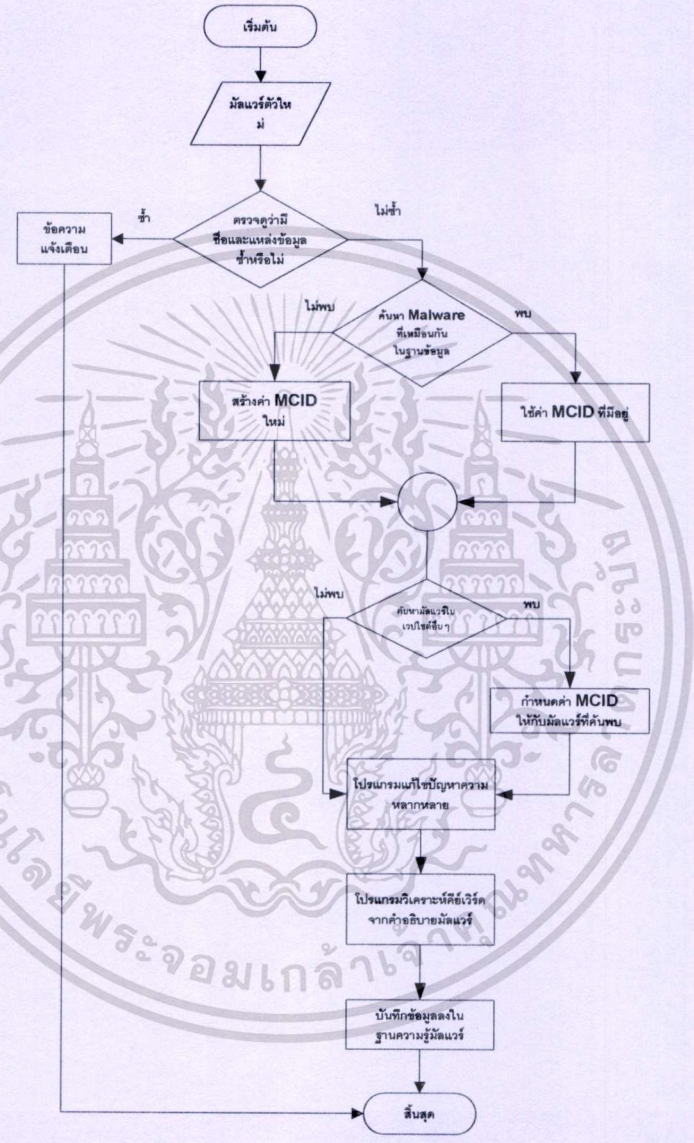
รูปที่ 3.5 ระบบรวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปเพจต่าง ๆ

- โปรแกรมกำหนดหมายเลขประจำตัวร่วมมัลแวร์ (MCID Assigning Engine)

ก่อนที่มัลแวร์ตัวใหม่จะเข้าสู่ระบบฐานความรู้มัลแวร์นั้น จะต้องมีการตรวจสอบถึงความซ้ำซ้อนของมัลแวร์เสียก่อน โดยโปรแกรมกำหนดหมายเลขประจำตัวร่วมมัลแวร์นี้จะตรวจสอบดูว่าในฐานความรู้มัลแวร์นั้นมีมัลแวร์ตัวดังกล่าวแล้วหรือไม่ ถ้ามีแล้วก็จะแจ้งข้อความเตือนถึงการพบมัลแวร์ที่ซ้ำซ้อนกัน แต่ถ้าไม่มีโปรแกรมจะทำการตรวจสอบต่อไปว่าชื่อหลักและชื่อเรียกของมัลแวร์ตัวใหม่นั้นตรงกับมัลแวร์ที่มีอยู่แล้วในฐานความรู้มัลแวร์หรือไม่ ถ้าหากว่าพบในฐานความรู้มัลแวร์แสดงว่าเป็นมัลแวร์เดียวกันเพียงแต่มีชื่อเรียกต่างกันและอาจจะมีข้อมูลอื่น ๆ เพิ่มเติมโปรแกรมก็จะนำค่าหมายเลขประจำตัวร่วมมัลแวร์จากมัลแวร์เดิมที่พบในฐานความรู้มัลแวร์มากำหนดให้กับมัลแวร์ตัวใหม่ เพื่อผู้ใช้จะได้ทราบว่ามัลแวร์ทั้งสองตัวนี้เป็นมัลแวร์ตัวเดียวกันแต่มีชื่อเรียกที่ต่างกัน แต่ถ้าหากว่าไม่พบข้อมูลของมัลแวร์นั้นในฐานความรู้มัลแวร์ โปรแกรมจะกำหนดค่า MCID ค่าใหม่ให้แก่มัลแวร์ตัวใหม่ซึ่งจะต้องเป็นค่าที่ไม่ซ้ำกับ MCID ที่มีอยู่แล้วในฐานข้อมูล

ต่อจากนั้นโปรแกรมจะนำชื่อมัลแวร์ไปค้นหาในเวปไซต์ที่ได้ลงทะเบียนไว้ในฐานข้อมูล เพื่อค้นหาว่าชื่อของมัลแวร์ตัวใหม่นั้นตรงกับชื่อเรียกของมัลแวร์ในแหล่งข้อมูล

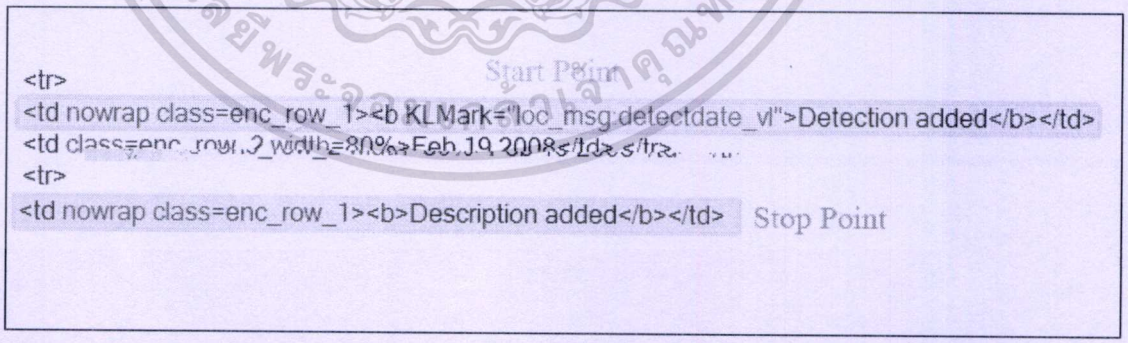
มัลแวร์ใดบ้าง โดยถ้าโปรแกรมพบว่ามีเวปไซต์อื่นที่มีชื่อเรียกตรงกับชื่อของมัลแวร์ตัวดังกล่าว ระบบก็จะทำการกำหนดค่า MCID ให้กับมัลแวร์ทั้งหมดที่พบด้วยค่า MCID เดียวกัน แต่ถ้าไม่พบ ระบบก็จะเข้าสู่กระบวนการต่อไป โดยขั้นตอนการทำงานของระบบกำหนดหมายเลขประจำตัวร่วมมัลแวร์นั้นได้แสดงไว้ในรูปที่ 3.6



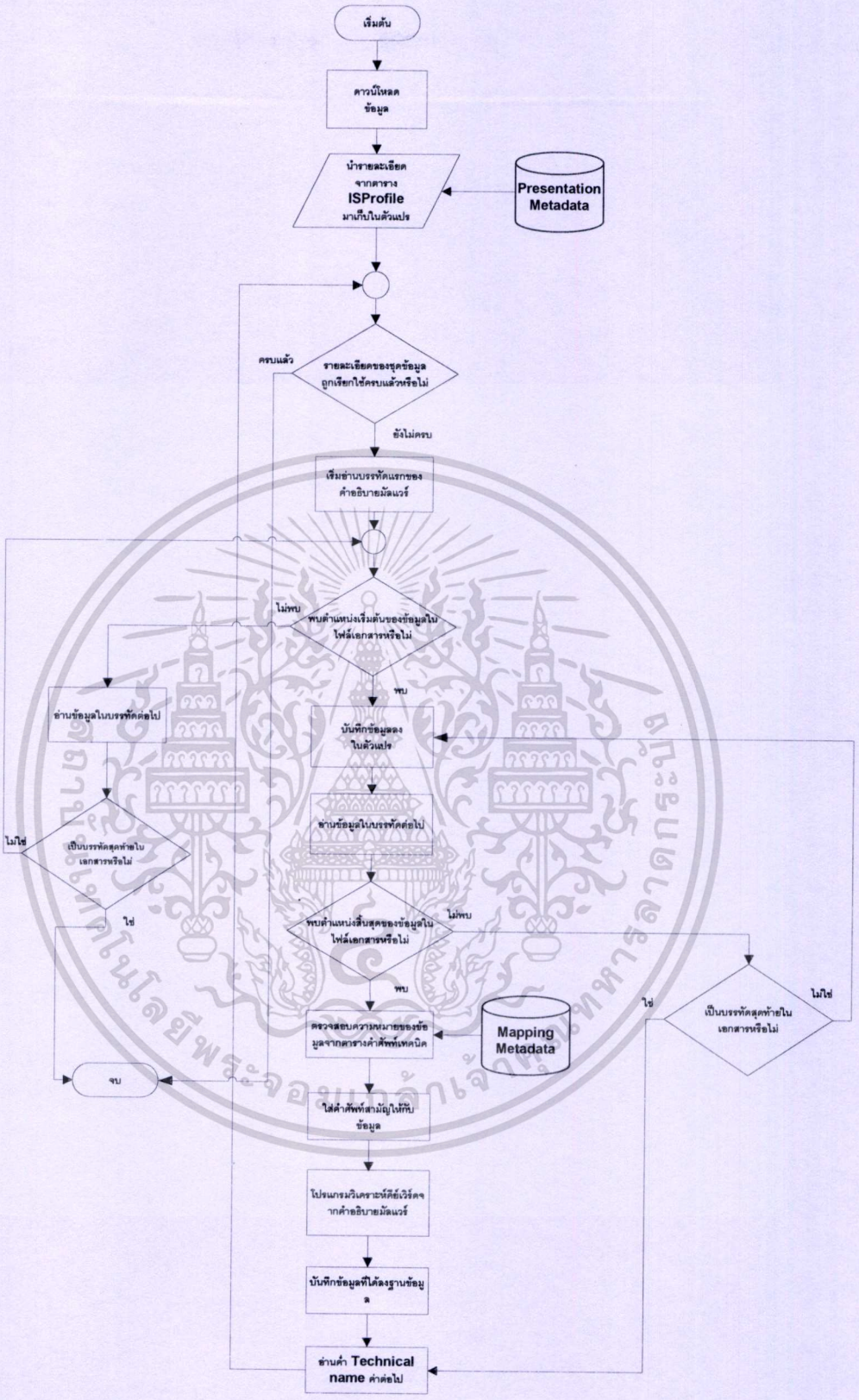
รูปที่ 3.6 อธิบายการทำงานโปรแกรมกำหนดหมายเลขประจำตัวร่วมมัลแวร์

- โปรแกรมที่แก้ไขปัญหาความหลากหลายทางด้านโครงสร้าง (Schematic Heterogeneous Solution Engine)

หลังจากที่คำอธิบายมัลแวร์ได้ผ่านกระบวนการตรวจสอบในโปรแกรมกำหนดหมายเลขประจำตัวร่วมมัลแวร์ไปแล้วนั้น โปรแกรมแก้ไขปัญหาด้านความหลากหลายจะทำการดาวน์โหลดคำอธิบายมัลแวร์จากเวปไซต์มาเก็บเป็นไฟล์เอกสาร HTML ซึ่งจะประกอบด้วยข้อมูลคำอธิบายมัลแวร์ที่บรรจุอยู่ภายในเอกสาร HTML รวมถึงโค้ดของภาษา HTML ดังแสดงในรูปที่ 3.7 จากรูปจะเห็นว่าข้อมูลที่ต้องการนั้นจะอยู่รวมกับโค้ดของภาษา HTML ซึ่งจะมีจุดเริ่มต้น (Start Point) และจุดสิ้นสุด (Stop Point) เพื่อบอกถึงตำแหน่งที่เริ่มต้นของข้อมูลและตำแหน่งที่สิ้นสุดของข้อมูลตามลำดับ โดยการจะแยกแต่ละชุดข้อมูลในคำอธิบายมัลแวร์ได้นั้น เราจะต้องใช้เมตาดาต้าที่ได้ออกแบบไว้แล้วเข้ามาช่วย กล่าวคือ หลังจากที่ดาวน์โหลดคำอธิบายมัลแวร์มาแล้วนั้น โปรแกรมจะนำรายละเอียดจุดเริ่มต้นและจุดสิ้นสุดของชุดข้อมูลแต่ละชุดออกมาจากตาราง ISProfile ต่อจากนั้นโปรแกรมจะเริ่มค้นหาข้อมูลแต่ละชุด เมื่อได้ข้อมูลดังกล่าวแล้ว โปรแกรมจะนำเมตาดาต้าด้านความหมายในตารางเปรียบเทียบคำศัพท์เทคนิคมาตรวจสอบดูถึงคำศัพท์สามัญของข้อมูลดังกล่าว เพื่อให้ทราบถึงความหมายที่แท้จริงของข้อมูล ต่อจากนั้นจะโปรแกรมจะบันทึกคำศัพท์สามัญซึ่งบอกถึงความหมายของข้อมูลลงในตัวแปร เพื่อเป็นการบอกว่าข้อมูลที่ได้นั้นมีความหมายที่แท้จริงเช่นไร เพื่อที่จะได้เขียนข้อมูลลงในฐานความรู้มัลแวร์ได้อย่างถูกต้องตามความหมายที่แท้จริง ต่อจากนั้นเมื่อได้ข้อมูลแต่ละชุดข้อมูลครบแล้ว โปรแกรมจะนำข้อมูลแต่ละชุดที่ได้ไปวิเคราะห์คีย์เวิร์ดต่อไปโดยขั้นตอนการทำงานของโปรแกรมแก้ไขความหลากหลายนั้นได้แสดงไว้ในรูปที่ 3.8



รูปที่ 3.7 การกำหนดตำแหน่งเริ่มต้นและสิ้นสุดโดยใช้แท็ก HTML



รูปที่ 3.8 โปรแกรมแก้ไขปัญหาความหลากหลายทางด้านโครงสร้าง

- โปรแกรมวิเคราะห์คุณสมบัติของมัลแวร์ (Malware Properties Analysis Engine)

จากการที่เราได้วิเคราะห์พฤติกรรมความแตกต่างของมัลแวร์แต่ละประเภทในบทที่ 2 มาแล้วนั้น เราพบว่าคำอธิบายมัลแวร์แต่ละแหล่งข้อมูลนั้นมีการอธิบายการทำงานของมัลแวร์ประเภทต่าง ๆ ด้วยกลุ่มคีย์เวิร์ดที่เหมือนกัน ยกตัวอย่างเช่น ในคำอธิบายมัลแวร์ของไวรัสนั้นมักจะระบุถึงวิธีการแพร่เชื้อไวรัสไปยังโฮสต์ไฟล์เป้าหมายเสมอ เช่น Append, Prepend หรือ Infect เป็นต้น คำอธิบายมัลแวร์ของหนอนคอมพิวเตอร์นั้นก็มักจะมีคำศัพท์ที่บ่งบอกถึงการแพร่กระจายตนเองเอาไว้เหมือนกัน เช่น Distribute, Spread หรือ Propagate เป็นต้น และนำคำอธิบายมัลแวร์ของโทรจันนั้นก็จะมีคำศัพท์ที่บ่งบอกถึงการทำงานของโทรจันเอาไว้ เช่น Password stealing, Key logging หรือ Backdoor Trojan เป็นต้น

ดังนั้นในการตัดสินคุณสมบัติของมัลแวร์แต่ละตัวนั้น จึงเป็นการคำนวณหาคีย์เวิร์ดของไวรัส หนอนคอมพิวเตอร์ และโทรจัน ที่ปรากฏอยู่ในคำอธิบายมัลแวร์ เราสามารถตัดสินคุณสมบัติของมัลแวร์แต่ละตัวได้จากสูตรการคำนวณในการวิเคราะห์คุณสมบัติมัลแวร์ซึ่งจะกล่าวในหัวข้อถัดไป

○ **สูตรคำนวณในการวิเคราะห์คุณสมบัติของมัลแวร์**

จากหลักการแนวคิดในการตัดสินคุณสมบัติของมัลแวร์ตามที่ได้กล่าวไปแล้วนั้น สามารถคำนวณจำนวนคำศัพท์ที่พบในคำอธิบายมัลแวร์แต่ละประเภทออกมาในรูปของเปอร์เซ็นต์ ซึ่งสูตรการคำนวณดังกล่าวนี้แสดงด้วยสมการต่อไปนี้

$$\text{Virus}(\%) = \frac{\text{Sum}(\text{Keywords}_{\text{Virus}})}{\text{Total}} \tag{3.1}$$

$$\text{Worm}(\%) = \frac{\text{Sum}(\text{Keywords}_{\text{Worm}})}{\text{Total}} \tag{3.2}$$

$$\text{Trojan}(\%) = \frac{\text{Sum}(\text{Keywords}_{\text{Trojan}})}{\text{Total}} \tag{3.3}$$

$$\text{Total} = \text{Sum}(\text{Keywords}_{\text{Virus}}) + \text{Sum}(\text{Keywords}_{\text{Worm}}) + \text{Sum}(\text{Keywords}_{\text{Trojan}}) \tag{3.4}$$

$$\text{Sum}(\text{Keywords}) = \sum(\text{Keywords}_i) \tag{3.5}$$

เมื่อ Keywords_i เป็นคีย์เวิร์ดคำที่ i ที่พบในคำอธิบายมัลแวร์

$\text{Sum}(\text{Keywords}_{\text{Virus}})$ เป็นผลรวมของคีย์เวิร์ดประเภทไวรัสที่พบในคำอธิบายมัลแวร์

Sum ($Keywords_{Worm}$) เป็นผลรวมของคีย์เวิร์ดประเภทหนอนคอมพิวเตอร์ที่พบในคำอธิบายมัลแวร์

Sum ($Keywords_{Trojan}$) เป็นผลรวมของคีย์เวิร์ดประเภทโทรจันที่พบในคำอธิบายมัลแวร์

Virus (%) เป็นอัตราส่วนจำนวนคีย์เวิร์ดของไวรัสที่พบในคำอธิบายมัลแวร์ เมื่อเปรียบเทียบกับจำนวนคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ เป็นเปอร์เซ็นต์

Worm (%) เป็นอัตราส่วนจำนวนคีย์เวิร์ดของหนอนคอมพิวเตอร์ที่พบในคำอธิบายมัลแวร์ โดยเปรียบเทียบกับคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ เป็นเปอร์เซ็นต์

Trojan (%) เป็นอัตราส่วนจำนวนคีย์เวิร์ดของโทรจันที่พบในคำอธิบายมัลแวร์ โดยเปรียบเทียบกับคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ เป็นเปอร์เซ็นต์

"…Infected files will increase in size by 1,701 bytes. Evil is not able to recognize when it has previously infected a file, so it may reinfect .COM files several times. Each infection will result in another 1,701 bytes of viral code being appended to the file. Like PhoenixD, Evil will infect files when they are opened for any reason, in addition to when they are executed. The simple act of copying a .COM file will result in both the source and target .COM files being infected. Systems infected with the Evil virus will experience problems with executing CHKDSK.COM. …"

รูปที่ 3.9 ตัวอย่างคีย์เวิร์ดไวรัสที่พบในคำอธิบายมัลแวร์

"…This worm is dropped by W32/Autorun.worm.zu.dr. This worm spreads by copying itself to network shares and to removable devices, along with an "Autorun.inf". This description is for a worm that is capable of spreading through removable devices and network shares. The characteristics of this worm in regards to file names, folders created etc. …"

รูปที่ 3.10 ตัวอย่างคีย์เวิร์ดหนอนคอมพิวเตอร์ที่พบในคำคำอธิบายมัลแวร์

"..Password Stealers may steal data from the hard drive. This data might include: CD Keys for various games credit card details your local username/password It may also log keystrokes for login details for banking applications, for example while Internet Explorer is open and connected to specific websites. As it is trivial for the malware author to modify the Password Stealer to transmit data to a different website or web address, McAfee write detection routines for these Trojans which as a general rule do not include these strings in the detection routine.."

รูปที่ 3.11 ตัวอย่างคีย์เวิร์ดโทรจันที่พบในคำอธิบายของมัลแวร์

ในการตัดสินคุณสมบัติของมัลแวร์นั้นเราจะค้นหาคีย์เวิร์ดของมัลแวร์ประเภทต่าง ๆ ที่ปรากฏอยู่ในคำอธิบายมัลแวร์ซึ่งแสดงในรูปที่ 3.10 – 3.12 จากนั้น จะทำการคำนวณจำนวนคีย์เวิร์ดที่พบออกมาเป็นเปอร์เซ็นต์โดยเทียบคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ การคำนวณคีย์เวิร์ดนั้นเราจะไม่นำคำศัพท์ที่มีความหมายซ้ำกันมาคำนวณอีกเป็นครั้งที่สอง ยกตัวอย่างเช่น รูปที่ 3.10 คำว่า Infect, Infection, Infected ที่ปรากฏในเอกสารนั้นจะถูกโปรแกรมนับเป็นคีย์เวิร์ดเพียงหนึ่งคีย์เวิร์ดเท่านั้นแม้ว่าจะพบคีย์เวิร์ดเหล่านี้อีกเป็นจำนวนมากก็ตาม

จากรูปที่ 3.10, 3.11 และ 3.12 เราสามารถคำนวณจำนวนของคีย์เวิร์ดที่พบในคำอธิบายมัลแวร์ได้ สมมุติว่าในคำอธิบายมัลแวร์มีคีย์เวิร์ดของโทรจันจำนวน 13 คีย์เวิร์ด มีคีย์เวิร์ดของหนอนคอมพิวเตอร์จำนวน 4 คีย์เวิร์ด และมีคีย์เวิร์ดของไวรัสจำนวน 1 คีย์เวิร์ด โปรแกรมจะนำคีย์เวิร์ดที่พบไปคำนวณในสมการที่ (3.1) – (3.3) ซึ่งเราจะได้ผลลัพธ์ดังต่อไปนี้ คีย์เวิร์ดของไวรัสถูกพบในคำอธิบายมัลแวร์คิดเป็น 5.56% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ คีย์เวิร์ดของหนอนคอมพิวเตอร์ถูกพบในคำอธิบายมัลแวร์คิดเป็น 22.22% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์และคีย์เวิร์ดของโทรจันถูกพบในคำอธิบายมัลแวร์คิดเป็น 72.22% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์

ต่อจากนั้นโปรแกรมจะนำผลลัพธ์ดังกล่าวนี้ไปคำนวณหาค่า Diff A และ Diff B ซึ่งเป็นค่าผลต่างของเปอร์เซ็นต์คีย์เวิร์ดของชนิดมัลแวร์ที่พบมากที่สุดและรองลงมา และค่าผลต่างของเปอร์เซ็นต์คีย์เวิร์ดของชนิดมัลแวร์ที่พบมาเป็นอันดับสองและอันดับสุดท้าย ตามลำดับ ดังแสดงในสมการที่ (3.6) และ (3.7) โดยค่า Diff A และ Diff B นั้นเท่ากับ 50% และ 16.66% ตามลำดับ จากนั้นโปรแกรมจะนำค่า Diff A และ Diff B ที่ได้จากการคำนวณนี้ไปพิจารณาในเงื่อนไขของสมการ (3.8) – (3.10) เพื่อตัดสินว่ามัลแวร์ตัวดังกล่าวมีคุณสมบัติประเภทใด

$$\text{Diff A} = \text{Max (\%)} - \text{Mid (\%)} \quad (3.6)$$

$$\text{Diff B} = \text{Mid (\%)} - \text{Min (\%)} \quad (3.7)$$

เมื่อ Max (%) เป็นเปอร์เซ็นต์ของคีย์เวิร์ดที่พบมากเป็นอันดับที่หนึ่งในคำอธิบายมัลแวร์
Mid (%) เป็นเปอร์เซ็นต์ของคีย์เวิร์ดที่พบมากเป็นอันดับที่สองในคำอธิบายมัลแวร์
Min (%) เป็นเปอร์เซ็นต์ของคีย์เวิร์ดที่พบมากเป็นอันดับที่สามในคำอธิบายมัลแวร์
Diff A เป็นค่าผลต่างของเปอร์เซ็นต์คียเวิร์ดที่พบในคำอธิบายมัลแวร์ของคีย์เวิร์ดที่พบมากเป็นอันดับที่หนึ่งและอันดับที่สองในคำอธิบายมัลแวร์ มีหน่วยเป็นเปอร์เซ็นต์
Diff B เป็นค่าผลต่างของเปอร์เซ็นต์คียเวิร์ดที่พบในคำอธิบายมัลแวร์ของคีย์เวิร์ดที่พบมากเป็นอันดับที่สองและอันดับที่สามในคำอธิบายมัลแวร์ มีหน่วยเป็นเปอร์เซ็นต์

จากการทดลองวิเคราะห์คำอธิบายมัลแวร์ทั้งหมด 1,000 คำอธิบาย โดยแบ่งเป็นคำอธิบายประเภทไวรัสจำนวน 400 คำอธิบาย หนอนคอมพิวเตอร์ 300 คำอธิบาย และโทรจัน 300 คำอธิบาย เราทดลองแทนค่าเทรชโฮลด์ค่าต่าง ๆ เพื่อตรวจสอบว่าเทรชโฮลด์ค่าใดที่สามารถตัดสินคุณสมบัติของมัลแวร์ได้แม่นยำที่สุด ในกรณีการตัดสินมัลแวร์ที่มีคุณสมบัติประเภทเดียวนั้น เราพบว่าค่าเทรชโฮลด์เท่ากับ 30% เป็นค่าที่สามารถตัดสินมัลแวร์ที่มีคุณสมบัติประเภทเดียวได้ถูกต้องคิดเป็น 92% ของจำนวนมัลแวร์ทั้งหมด เราเรียกค่าเทรชโฮลด์นี้ว่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียว ($\text{Threshold}_{\text{Single}}$) และค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียวนี้สามารถนำมาใช้ร่วมกับค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติสองประเภท ($\text{Threshold}_{\text{Double}}$) ในการตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติสองประเภทและมัลแวร์ที่มีคุณสมบัติสามประเภท ซึ่งเราเลือกใช้ค่าเทรชโฮลด์เท่ากับ 15% ในการตัดสิน กล่าวคือ มัลแวร์ใดที่มีค่า Diff A ที่คำนวณได้ต่ำกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียวซึ่งมีค่าเท่ากับ 30% และค่า Diff B ที่ได้จากการคำนวณมีค่ามากกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติสองประเภทซึ่งมีค่าเท่ากับ 15% จะถูกตัดสินว่าเป็นมัลแวร์ที่มีคุณสมบัติสองประเภท และมัลแวร์ใดที่มีค่า Diff A ที่คำนวณได้ต่ำกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียวซึ่งมีค่าเท่ากับ 30% และค่า Diff B ที่ได้จากการคำนวณมีค่าน้อยกว่าหรือเท่ากับค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติสองประเภทซึ่งมีค่าเท่ากับ 15% จะถูกตัดสินว่าเป็นมัลแวร์ที่มีคุณสมบัติสามประเภท จากผลการทดลองนี้เราจึงได้นำค่าเทรชโฮลด์ทั้งสองค่ามาใช้ในการตัดสินคุณสมบัติของมัลแวร์ประเภทต่าง ๆ ดังต่อไปนี้

- การตัดสินมัลแวร์ที่มีคุณสมบัติประเภทเดียว (Single Property)

หลังจากที่ได้ผลลัพธ์จากสมการที่ (3.6) และ (3.7) แล้วนั้น เราจะนำผลลัพธ์มาตรวจสอบกับเงื่อนไขในสมการที่ (3.8) เพื่อหาคุณสมบัติของมัลแวร์ ยกตัวอย่างเช่น สมมุติว่าผลลัพธ์จากการคำนวณในสมการที่ (3.1) – (3.3) มีค่าเป็นดังนี้ คีย์เวิร์ดของไวรัสถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 61% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ คีย์เวิร์ดของหนอนคอมพิวเตอร์ถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 22% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ และคีย์เวิร์ดของโทรจันถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 17% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ ค่า Diff A ที่ได้จากการคำนวณมีค่าเท่ากับ 39% เมื่อพิจารณาในสมการที่ (3.8) พบว่าค่า Diff A มีค่ามากกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียวซึ่งมีค่าเท่ากับ 30% ซึ่งตรงกับเงื่อนไขที่กำหนดไว้ ดังนั้นมัลแวร์ตัวดังกล่าวจึงเป็นมัลแวร์ที่มีคุณสมบัติประเภทเดียวและเป็นประเภทไวรัส เพราะมีการพบคีย์เวิร์ดของไวรัสมากเป็นอันดับหนึ่งเมื่อเทียบกับคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์

$$M \in \{M_S\} \text{ If Diff A} > \text{Threshold}_{\text{Single}} \qquad (3.8)$$

เมื่อ M เป็นมัลแวร์ตัวที่กำลังพิจารณา

M_S เป็นมัลแวร์ที่มีคุณสมบัติประเภทเดียว

- การตัดสินมัลแวร์ที่มีคุณสมบัติสองประเภท (Double properties)

ในกรณีนี้จะใช้วิธีการพิจารณาคล้ายกับการตัดสินมัลแวร์ที่มีคุณสมบัติประเภทเดียว แต่แทนที่จะใช้สมการ (3.8) ในการตัดสินคุณสมบัติของมัลแวร์ เราจะใช้สมการที่ (3.9) ในการตัดสินคุณสมบัติของมัลแวร์แทน ยกตัวอย่างเช่น สมมุติว่าผลลัพธ์จากการคำนวณในสมการที่ (3.1) – (3.3) เป็นดังนี้ คีย์เวิร์ดของไวรัสถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 7% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ คีย์เวิร์ดของหนอนคอมพิวเตอร์ถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 60% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ และคีย์เวิร์ดของโทรจันถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 33% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ โดยค่า Diff A และ Diff B ที่ได้จากการคำนวณนั้นมีค่าเท่ากับ 27% และ 26% ตามลำดับ เมื่อพิจารณาในสมการที่ (3.9) พบว่าค่า Diff A มีค่าน้อยกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียวซึ่งมีค่าเท่ากับ 30% และค่า Diff B มีค่ามากกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติสองประเภท ซึ่งตรงกับเงื่อนไขที่กำหนดไว้ ดังนั้นมัลแวร์ตัวดังกล่าวจึงเป็นมัลแวร์ที่มีคุณสมบัติสองประเภท มัลแวร์ตัวดังกล่าวนั้นมีคุณสมบัติเป็นทั้งหนอนคอมพิวเตอร์และโทรจัน เนื่องจากมีการพบคีย์เวิร์ดของหนอนคอมพิวเตอร์มากเป็นอันดับหนึ่ง

เมื่อเทียบกับคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ และมีการพบคีย์เวิร์ดของโทรจันมากเป็นอันดับสองเมื่อเทียบกับคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์

$$M \in \{M_D\} \text{ If } (\text{Diff A} \leq \text{Threshold}_{\text{Single}}) \text{ AND } (\text{Diff B} > \text{Threshold}_{\text{Double}}) \qquad (3.9)$$

เมื่อ M เป็นมัลแวร์ตัวที่กำลังพิจารณา

M_D เป็นมัลแวร์ที่มีคุณสมบัติสองประเภท

- การตัดสินมัลแวร์ที่มีคุณสมบัติสามประเภท (Triple properties)

ในกรณีนี้จะเป็นการพิจารณาว่ามัลแวร์นั้นมีคุณสมบัติสามประเภทหรือไม่ โดยใช้สมการที่ (3.10) ในการตัดสินมัลแวร์ที่คุณสมบัติสามประเภท ยกตัวอย่างเช่น สมมุติว่าผลลัพธ์จากการคำนวณในสมการที่ (3.1) – (3.3) เป็นดังนี้ คีย์เวิร์ดของหนอนคอมพิวเตอร์ถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 43% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ คีย์เวิร์ดของโทรจันถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 32% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ และคีย์เวิร์ดของไวรัสถูกพบในเอกสารคำอธิบายมัลแวร์คิดเป็น 25% ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ ค่า Diff A และ Diff B ที่ได้จากการคำนวณมีค่าเท่ากับ 11% และ 7% ตามลำดับ จากนั้นนำค่าที่ได้มาพิจารณาเงื่อนไขในสมการที่ (3.10) พบว่าค่า Diff A มีค่าน้อยกว่าค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติประเภทเดียวซึ่งมีค่าเท่ากับ 30% และค่า Diff B มีค่าน้อยกว่าหรือเท่ากับค่าเทรชโฮลด์ที่ใช้ตัดสินแยกกลุ่มมัลแวร์ที่มีคุณสมบัติสองประเภท ซึ่งตรงกับเงื่อนไขที่กำหนดไว้ ดังนั้นจึงสรุปได้ว่า มัลแวร์ตัวดังกล่าวมีคุณสมบัติของไวรัส หนอนคอมพิวเตอร์ และโทรจันรวมอยู่ในตัวเดียวกัน

$$M \in \{M_T\} \text{ If } (\text{Diff A} \leq \text{Threshold}_{\text{Single}}) \text{ AND } (\text{Diff B} \leq \text{Threshold}_{\text{Double}}) \qquad (3.10)$$

เมื่อ M เป็นมัลแวร์ตัวที่กำลังพิจารณา

M_T เป็นมัลแวร์ที่มีคุณสมบัติสามประเภท

- โปรแกรมวิเคราะห์ภัยคุกคามที่เกิดจากมัลแวร์ (Malware Threat Analysis Engine)

โปรแกรมวิเคราะห์ภัยคุกคามที่เกิดจากมัลแวร์นั้นจะทำหน้าที่วิเคราะห์ระดับความเสียหายที่เกิดขึ้นจากมัลแวร์ ระดับการแพร่กระจาย สถานะการแพร่กระจาย และอัตราเสี่ยงที่ส่งผลต่อผู้ใช้ โดยมีรายละเอียดต่าง ๆ ดังนี้

○ **ระดับของความเสียหายที่เกิดจากมัลแวร์ (Damage Level)**

จากทฤษฎีในบทที่ 2 สามารถวิเคราะห์ระดับของความเสียหายที่เกิดขึ้นจากมัลแวร์ออกเป็น 3 ระดับดังต่อไปนี้

- ระดับสูง (High) เป็นความเสียหายที่ส่งผลต่อผู้ใช้เป็นอย่างมาก โดยความเสียหายดังกล่าวเกิดจากการที่มัลแวร์อาจจะใช้คำสั่งฟอร์แมตเครื่องคอมพิวเตอร์ของผู้ใช้ ทำการลบไฟล์เอกสารของผู้ใช้ออกจากเครื่องคอมพิวเตอร์ ขโมยบัญชีผู้ใช้และรหัสผ่านของอีเมล์ ขโมยบัญชีผู้ใช้และรหัสผ่านสำหรับทำธุรกรรมทางการเงินของผู้ใช้ ทำการควบคุมเครื่องคอมพิวเตอร์ของผู้ใช้จากระยะไกลเพื่อใช้เครื่องคอมพิวเตอร์ของผู้ใช้ในการสร้างผลประโยชน์ให้กับคนสร้างมัลแวร์ โดยตัวอย่างของคีย์เวิร์ดเหล่านี้ได้แก่ Format C:\, Deletes All Files, Delete data, Password Stealing เป็นต้น

- ระดับปานกลาง (Medium) ความเสียหายในส่วนนี้เกิดขึ้นจากการที่มัลแวร์ได้ทำการปรับแก้ไฟล์ของระบบปฏิบัติการของเครื่องคอมพิวเตอร์ มัลแวร์อาจจะทำการซ่อนคำสั่งหรือเมนูบางอย่างของระบบหรือโปรแกรม ทำให้ผู้ใช้ไม่สามารถใช้งานโปรแกรมหรือเมนูได้ตามปรกติ มัลแวร์รบกวนการทำงานของผู้ใช้โดยการเปลี่ยนแปลงหรือทำให้อินเตอร์เฟสบางอย่างไม่สามารถทำงานได้ตามปรกติ เช่น ทำให้เมาส์ของผู้ใช้ไม่สามารถคลิ๊กได้ ทำให้แป้นพิมพ์ของผู้ใช้ไม่สามารถใช้งานได้ตามปรกติ เป็นต้น ซึ่งความเสียหายในส่วนนี้มักจะเป็นการสร้างอุปสรรคในการทำงานให้กับผู้ใช้ โดยตัวอย่างของคีย์เวิร์ดเหล่านี้ได้แก่ Registry modified, Hide Folder option, Disable mouse or keyboards เป็นต้น

- ระดับต่ำ (Low) ความเสียหายในส่วนนี้แม้ไม่ได้เกิดจากความตั้งใจของมัลแวร์โดยตรง แต่การที่มัลแวร์อาศัยอยู่บนเครื่องคอมพิวเตอร์ของผู้ใช้นั้นนับเป็นการใช้ทรัพยากรของเครื่องคอมพิวเตอร์ไปโดยเปล่าประโยชน์ เช่นมีการใช้หน่วยความจำของเครื่อง มีการใช้พื้นที่ดิสก์ และมีการใช้หน่วยประมวลผลกลางอีกด้วยซึ่งความเสียหายในส่วนนี้แม้จะไม่ได้มีความร้ายแรงมากแต่ก็นับว่าเป็นความเสียหายที่เกิดขึ้นจากการที่มีมัลแวร์อาศัยอยู่บนเครื่องของผู้ใช้ด้วยเช่นกัน

○ **ระดับการแพร่กระจายของมัลแวร์ (Distribution Level)**

จากการศึกษาพฤติกรรมการทำงานของมัลแวร์ในบทที่ 2 นั้นพบว่า ไวรัส หนอนคอมพิวเตอร์ และโทรจัน มีระดับการแพร่กระจายที่แตกต่างกันออกไป ไวรัสและโทรจันนั้นมีระดับการแพร่กระจายในระดับต่ำ เนื่องจากตัวของมันเองไม่ได้ถูกออกแบบมาให้แพร่กระจายไปนอกเครื่องคอมพิวเตอร์ การที่จะแพร่กระจายไวรัสและโทรจัน

ออกไปนอกเครื่องคอมพิวเตอร์ได้นั้นจำเป็นที่จะต้องมีผู้ใช้คอมพิวเตอร์เองที่เป็นพาหะหรือเป็นตัวกลางในการแพร่กระจายไวรัสและโทรจันออกไปด้วยตนเอง ซึ่งจะแตกต่างจากหนอนคอมพิวเตอร์ที่ได้ถูกโปรแกรมมาตั้งแต่แรกแล้ว โดยสามารถแบ่งระดับการแพร่กระจายของหนอนคอมพิวเตอร์ไปยังเครื่องอื่น ๆ ได้เป็น 3 ระดับ ดังต่อไปนี้

- ระดับสูง (High) มัลแวร์ที่มีการแพร่กระจายในระดับสูงนั้นมักจะเป็นมัลแวร์ในกลุ่มของหนอนคอมพิวเตอร์ที่ใช้อีเมล์เป็นช่องทางการแพร่กระจาย เนื่องจากหนอนคอมพิวเตอร์ประเภทที่แพร่กระจายผ่านทางอีเมล์นั้นจะทำการค้นหารายชื่อของผู้ใช้ท่านอื่น ๆ จากสมุดรายชื่อในเครื่องคอมพิวเตอร์ที่ตกเป็นเหยื่อ แล้วส่งสำเนารหัสโปรแกรมต่อไปยังที่อยู่ที่ปรากฏในสมุดรายชื่อดังกล่าว จากการทำงานเช่นนี้จึงมักจะทำให้หนอนคอมพิวเตอร์ประเภทนี้สามารถแพร่กระจายตนเองไปยังคอมพิวเตอร์เครื่องอื่น ๆ ได้อย่างรวดเร็วมาก อีกช่องทางหนึ่งที่หนอนคอมพิวเตอร์ใช้ในการแพร่กระจายได้อย่างรวดเร็วก็คือการใช้โปรโตคอล UDP, TCP เป็นต้น เนื่องจากหนอนคอมพิวเตอร์ที่ใช้วิธีการแพร่กระจายในลักษณะนี้จะสร้างแพ็คเก็จขนาดเล็กขึ้นมาแล้วส่งออกไปในระบบเครือข่ายหรืออินเตอร์เน็ต ซึ่งเมื่อแพ็คเก็จดังกล่าวนี้ไปถึงเครื่องผู้ใช้แล้วแพ็คเก็จดังกล่าวจะทำงานทันทีที่เข้าถึงเครื่องผู้ใช้ โดยไม่ต้องอาศัยผู้ใช้คอมพิวเตอร์ในการเปิดหรือเรียกไฟล์หนอนคอมพิวเตอร์ตัวนั้นขึ้นมาทำงานทั้งสิ้น โดยตัวอย่างของคีย์เวิร์ดเหล่านี้ได้แก่ Mass mailing, E-mail worm, TCP worm เป็นต้น

- ระดับปานกลาง (Medium) นอกเหนือจากการใช้อีเมล์และโปรโตรคอลเพื่อเป็นช่องทางในการแพร่กระจายของหนอนคอมพิวเตอร์ดังที่กล่าวมาแล้วนั้น ยังมีช่องทางการสื่อสารที่หนอนคอมพิวเตอร์สามารถแพร่กระจายตนเองไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ได้เช่นกัน ยกตัวอย่างเช่น หนอนคอมพิวเตอร์ที่อาศัยโปรแกรมสนทนาซึ่งเป็นที่นิยมเป็นตัวกลางในการแพร่กระจายไปยังคอมพิวเตอร์เครื่องอื่น ๆ เช่น โปรแกรม MSN Messenger, Yahoo Messenger, ICQ, IRC เป็นต้น หรือหนอนคอมพิวเตอร์บางตัวอาจจะอาศัยโปรแกรมแลกเปลี่ยนไฟล์ซึ่งใช้ระบบเพียร์ทูเพียร์ในการแพร่กระจายตนเองไปยังเครื่องอื่น ๆ เช่น โปรแกรม Bitcomet, Limewire เป็นต้น แม้ว่าหนอนคอมพิวเตอร์จะใช้วิธีการแพร่กระจายตนเองโดยอาศัยโปรแกรมสนทนาออนไลน์หรือโปรแกรมแลกเปลี่ยนไฟล์แบบเพียร์ทูเพียร์ตามที่ได้กล่าวมาแล้วนั้นพบว่าหนอนคอมพิวเตอร์เหล่านี้ยังมีความสามารถในการแพร่กระจายค่อนข้างต่ำเมื่อเทียบกับหนอนคอมพิวเตอร์ที่ใช้อีเมล์เป็นตัวกลาง กล่าวคือ หนอนคอมพิวเตอร์ประเภทนี้จะแพร่กระจายตนเองได้ก็ต่อเมื่อผู้ใช้ปลายทางเปิดโปรแกรมสนทนาออนไลน์หรือโปรแกรมแลกเปลี่ยนไฟล์เอาไว้เท่านั้น ซึ่งถ้าหากว่าผู้ใช้ไม่ได้เปิดโปรแกรมเอาไว้หนอนคอมพิวเตอร์จะไม่สามารถส่งไฟล์ไปยังเครื่องผู้รับทางโปรแกรมสนทนาออนไลน์ได้เลย ซึ่งจากเหตุนี้เองที่ทำให้หนอนที่ใช้อีเมล์เป็นช่องทางในการแพร่กระจายสามารถ

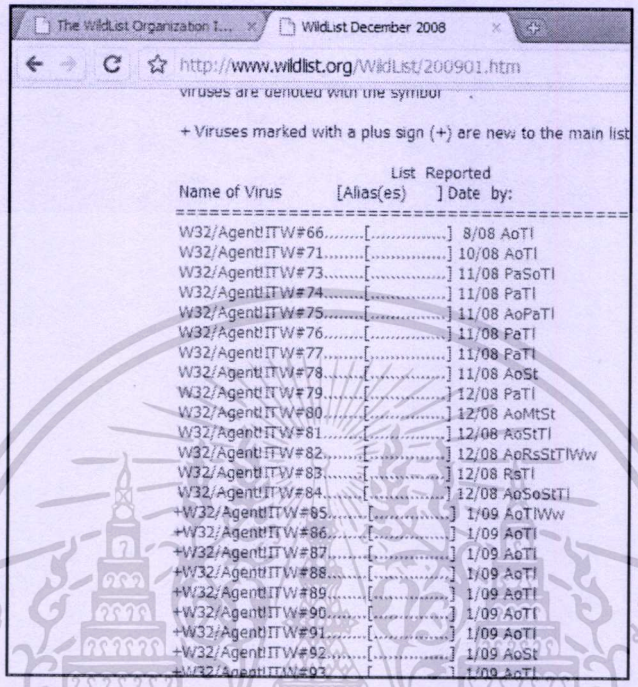
แพร่กระจายไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ได้มากกว่าหนอนคอมพิวเตอร์ประเภทนี้ เนื่องจากว่าผู้ใช้ที่ได้รับอีเมล์นั้นจะเปิดอ่านไฟล์ที่มีหนอนคอมพิวเตอร์อยู่ในเวลาใดก็ได้ โดยตัวอย่างของคีย์เวิร์ดเหล่านี้ได้แก่ MSN worm, P2P worm, Yahoo Messenger เป็นต้น

- ระดับต่ำ (Low) หนอนคอมพิวเตอร์บางตัวนั้นแพร่กระจายไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ โดยอาศัยอุปกรณ์เก็บข้อมูลแบบพกพาเป็นตัวกลาง ซึ่งหนอนคอมพิวเตอร์ชนิดนี้จะสร้างไฟล์ autorun.inf ขึ้นมาในอุปกรณ์เก็บข้อมูลแบบพกพา โดยไฟล์ autorun.inf นี้จะทำหน้าที่เรียกไฟล์ของหนอนคอมพิวเตอร์ที่อยู่ในอุปกรณ์ให้ทำงานโดยอัตโนมัติทุกครั้งที่ผู้ใช้นำอุปกรณ์ดังกล่าวนี้เชื่อมต่อเข้ากับเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ซึ่งเป็นสาเหตุให้หนอนคอมพิวเตอร์ประเภทนี้แพร่กระจายไปยังคอมพิวเตอร์เครื่องอื่นได้ แม้ว่าหนอนคอมพิวเตอร์ประเภทนี้จะแพร่กระจายไปยังเครื่องคอมพิวเตอร์เครื่องอื่นได้ทันทีที่ผู้ใช้ได้เชื่อมต่ออุปกรณ์เก็บข้อมูลแบบพกพาเข้ากับเครื่องคอมพิวเตอร์ แต่ระดับการแพร่กระจายของหนอนคอมพิวเตอร์ชนิดนี้ยังอยู่ในระดับต่ำเนื่องจากหนอนคอมพิวเตอร์ชนิดนี้ไม่สามารถแพร่กระจายตนเองผ่านเครือข่ายอินเตอร์เน็ตได้ด้วยตนเองยังคงต้องอาศัยผู้ใช้ในการเชื่อมต่ออุปกรณ์ที่มีหนอนคอมพิวเตอร์อาศัยอยู่ไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ด้วยตัวของผู้ใช้เองทีละเครื่อง ทำให้จำนวนการแพร่ระบาดของหนอนคอมพิวเตอร์ประเภทนี้ไม่มากนักเมื่อเทียบกับหนอนคอมพิวเตอร์ที่อาศัยอีเมล์และโปรแกรมสนทนาออนไลน์ในการแพร่กระจายซึ่งสามารถแพร่กระจายตนเองได้โดยอัตโนมัติผ่านเครือข่ายอินเตอร์เน็ตและสามารถแพร่กระจายไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ได้เป็นจำนวนมาก โดยตัวอย่างของคีย์เวิร์ดเหล่านี้ได้แก่ autorun, autorun.inf, flash drive worm เป็นต้น

○ **สถานะการแพร่กระจายของมัลแวร์ (Distribution Status)**

องค์กรไวด์ลิสต์ (Wildlist Organization) [41] นั้นก่อตั้งขึ้นจากแนวความคิดของ Joe Wells ผู้เชี่ยวชาญด้านโปรแกรมแอนตี้มัลแวร์ ซึ่งได้รวบรวมรายงานการแจ้งพบมัลแวร์ที่ระบาดอยู่บนอินเตอร์เน็ต เขาและกลุ่มนักวิจัยด้านมัลแวร์ท่านอื่น ๆได้สร้างรายชื่อของมัลแวร์ที่มีการแจ้งว่ามีการแพร่ระบาดอยู่บนอินเตอร์เน็ตจริง มาเผยแพร่เป็นครั้งแรกในปี ค.ศ. 1993 โดยรายชื่อมัลแวร์ที่มีการระบาดอยู่บนอินเตอร์เน็ตนั้นได้รับการเรียกว่า รายชื่อมัลแวร์ที่พบว่ามีสถานะการแพร่กระจายอยู่ในปัจจุบัน (Wildlist) โดยรายชื่อมัลแวร์เหล่านี้จะมีการเผยแพร่ทางเวปไซต์ขององค์กรไวด์ลิสต์ในทุกวันที่ 15 ของเดือน โดยชื่อของมัลแวร์ที่ปรากฏในรายชื่อมัลแวร์ที่พบว่ามีสถานะการแพร่กระจายอยู่ในปัจจุบันนั้นบ่งบอกว่ามัลแวร์ตัวดังกล่าวมีการแพร่กระจายอยู่บนอินเตอร์เน็ตจริงในช่วงเวลาหนึ่งเดือนที่ผ่านมา

งานวิจัยนี้ได้นำรายชื่อมัลแวร์ที่พบว่ามีสถานะการแพร่กระจายอยู่ในปัจจุบันมาช่วยในการวิเคราะห์ระดับความเสี่ยงที่มีผลต่อผู้ใช้ โดยตัวอย่างรายชื่อมัลแวร์ที่พบว่ามีสถานะการแพร่กระจายอยู่ในปัจจุบันนั้นแสดงในรูปที่ 3.12



รูปที่ 3.12 ตัวอย่างรายชื่อมัลแวร์ที่พบว่ามีสถานะการแพร่กระจายอยู่ในปัจจุบัน

- หลักการคำนวณระดับความเสียหายที่เกิดจากมัลแวร์และระดับการแพร่กระจายของมัลแวร์

ในส่วนการคำนวณระดับความเสียหายที่เกิดจากมัลแวร์และระดับการแพร่กระจายของมัลแวร์นั้น โปรแกรมจะค้นหาคีย์เวิร์ดที่บ่งบอกถึงความเสียหายที่เกิดจากมัลแวร์และการแพร่กระจายของมัลแวร์ซึ่งปรากฏอยู่ในคำอธิบายมัลแวร์และคำนวณจำนวนคีย์เวิร์ดที่พบออกมาเป็นเปอร์เซ็นต์โดยเทียบกับจำนวนของคีย์เวิร์ดทั้งหมดที่ปรากฏในคำอธิบายมัลแวร์ โดยหลักการคำนวณนั้นจะคล้ายกับการจำแนกประเภทของมัลแวร์ซึ่งได้กล่าวมาแล้วในหัวข้อก่อนหน้านี้ เมื่อโปรแกรมได้คำนวณจำนวนคีย์เวิร์ดความเสียหายที่เกิดจากมัลแวร์และการแพร่กระจายของมัลแวร์ที่ปรากฏในคำอธิบายมัลแวร์ออกมาเป็นเปอร์เซ็นต์แล้ว โปรแกรมจะนำค่าที่ได้จากการคำนวนนั้นมาวิเคราะห์ต่อไปโดยพิจารณาจากตารางที่ 3.21 ซึ่งเป็นตารางที่ประเมินความเสียหายในระดับสูงสุดที่มัลแวร์สามารถสร้างความเสียหายให้กับผู้ใช้ได้และการแพร่กระจายของมัลแวร์ในระดับสูงสุดที่มัลแวร์สามารถแพร่กระจายตนเองไปยังเครื่องคอมพิวเตอร์เครื่องอื่น ๆ ได้ เพื่อหาค่าระดับความเสียหายที่เกิดจากมัลแวร์และระดับการแพร่กระจายของมัลแวร์ยกตัวอย่างเช่น สมมุติว่าในการคำนวณจำนวนคีย์เวิร์ดที่บ่งบอกถึงความเสียหายที่เกิดจากมัลแวร์

นั้นพบว่า มีคีย์เวิร์ดของความเสียหายที่เกิดจากมัลแวร์ในระดับต่ำปรากฏในคำอธิบายมัลแวร์คิดเป็น 65.56 % ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ มีคีย์เวิร์ดของความเสียหายที่เกิดจากมัลแวร์ในระดับสูงปรากฏในคำอธิบายมัลแวร์คิดเป็น 24.44 % ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ และมีคีย์เวิร์ดของความเสียหายที่เกิดจากมัลแวร์ในระดับกลางปรากฏในคำอธิบายมัลแวร์คิดเป็น 10 % ของคีย์เวิร์ดทั้งหมดที่พบในคำอธิบายมัลแวร์ โปรแกรมจะนำผลลัพธ์ที่ได้จากการคำนวณไปวิเคราะห์โดยพิจารณาจากตารางที่ 3.21 ซึ่งในตัวอย่างนี้จะพบว่ามีการค้นพบคีย์เวิร์ดความเสียหายที่เกิดจากมัลแวร์ในระดับสูง มีการค้นพบคีย์เวิร์ดความเสียหายที่เกิดจากมัลแวร์ในระดับกลางและมีการค้นพบคีย์เวิร์ดความเสียหายที่เกิดจากมัลแวร์ในระดับต่ำในคำอธิบายมัลแวร์ โดยโปรแกรมจะประเมินระดับความเสียหายที่เกิดจากมัลแวร์ให้อยู่ในระดับสูง แม้ว่าในคำอธิบายมัลแวร์นั้นจะมีจำนวนคีย์เวิร์ดความเสียหายที่เกิดจากมัลแวร์ในระดับต่ำซึ่งมีจำนวนมากที่สุดเมื่อเทียบกับคีย์เวิร์ดของความเสียหายที่เกิดจากมัลแวร์ทั้งหมด

เนื่องจากเมื่อมีคีย์เวิร์ดความเสียหายที่เกิดจากมัลแวร์ในระดับสูงปรากฏในคำอธิบายมัลแวร์นั้นหมายถึงมัลแวร์ตัวดังกล่าวอาจจะสร้างความเสียหายในระดับสูงกับผู้ใช้ได้ ซึ่งถือว่าเป็นอันตรายต่อผู้ใช้เป็นอย่างมาก ซึ่งการประเมินค่าระดับความเสียหายที่เกิดจากมัลแวร์และระดับการแพร่กระจายของมัลแวร์ในตารางที่ 3.21 จะทำให้ผู้ใช้เพิ่มความระมัดระวังตนเองไม่ให้เครื่องคอมพิวเตอร์ของตนตกเป็นเหยื่อของมัลแวร์ตัวดังกล่าวซึ่งอาจจะสร้างความเสียหายให้กับผู้ใช้ได้เป็นอย่างมาก

ตารางที่ 3.21 เปรียบเทียบค่าระดับความเสียหายที่เกิดจากมัลแวร์และระดับการแพร่กระจายของมัลแวร์

สูง	กลาง	ต่ำ	ผลลัพธ์
✓	✓	✓	สูง
✓	✓	✗	สูง
✓	✗	✓	สูง
✓	✗	✗	สูง
✗	✓	✓	ปานกลาง
✗	✓	✗	ปานกลาง
✗	✗	✓	ต่ำ

-ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้ (Risk Level)

ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้นั้นเป็นค่าที่ได้การประเมินระดับภัยคุกคามทั้งหมดที่เกิดจากมัลแวร์ ซึ่งประกอบด้วยระดับความเสียหายที่เกิดจากมัลแวร์ ระดับการแพร่กระจายของมัลแวร์และสถานะการแพร่กระจายของมัลแวร์ว่าส่งผลต่อผู้ใช้คอมพิวเตอร์ในระดับใด

กล่าวคือ ค่าระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้นั้นเป็นค่าที่บอกผู้ใช้คอมพิวเตอร์ทั่วไปว่ามีโอกาสที่จะได้รับผลกระทบจากมัลแวร์ในระดับใด โดยหลักเกณฑ์ในการประเมินระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้นั้น จะพิจารณาจากปัจจัยทั้งสามค่าด้วยกัน คือ ระดับความเสียหายที่เกิดจากมัลแวร์ ระดับการแพร่กระจายของมัลแวร์และสถานะการแพร่กระจายของมัลแวร์ ซึ่งแสดงในตารางที่ 3.22โดยจะพิจารณาค่าที่ส่งผลกระทบต่อผู้ใช้มากที่สุด ซึ่งก็คือค่าระดับความเสียหายที่เกิดจากมัลแวร์เนื่องจากค่าดังกล่าวนี้นับว่าส่งผลกระทบต่อผู้ใช้มากที่สุด ต่อจากนั้นจึงจะพิจารณาค่าสถานะการแพร่กระจายของมัลแวร์และระดับการแพร่กระจายตามลำดับ กล่าวคือ ถ้าพบว่ามัลแวร์ตัวหนึ่งมีความสามารถในการสร้างความเสียหายในระดับสูง มีความสามารถในการแพร่กระจายในระดับสูง แต่พบว่าไม่ได้มีการแพร่ระบาดจริงในปัจจุบันนั้น ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้ก็จะมีค่าต่ำลงไปด้วยเนื่องจากมัลแวร์ที่ไม่ได้มีการแพร่ระบาดอยู่จริงในปัจจุบันย่อมมีโอกาสส่งผลกระทบต่อผู้ใช้ได้น้อยกว่ามัลแวร์ที่มีการแพร่ระบาดอยู่จริงในปัจจุบันเป็นต้น

ตารางที่ 3.22 ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้

ระดับความเสียหายที่เกิดจากมัลแวร์	ระดับการแพร่กระจายของมัลแวร์	สถานะการแพร่กระจายของมัลแวร์	ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้
สูง	สูง	มีการแพร่ระบาด	สูง
สูง	สูง	ไม่มีการแพร่ระบาด	ต่ำ
สูง	กลาง	มีการแพร่ระบาด	สูง
สูง	กลาง	ไม่มีการแพร่ระบาด	ต่ำ
สูง	ต่ำ	มีการแพร่ระบาด	กลาง
สูง	ต่ำ	ไม่มีการแพร่ระบาด	ต่ำ
กลาง	สูง	มีการแพร่ระบาด	กลาง
กลาง	สูง	ไม่มีการแพร่ระบาด	ต่ำ
กลาง	กลาง	มีการแพร่ระบาด	กลาง
กลาง	กลาง	ไม่มีการแพร่ระบาด	ต่ำ
กลาง	ต่ำ	มีการแพร่ระบาด	ต่ำ
กลาง	ต่ำ	ไม่มีการแพร่ระบาด	ต่ำ
ต่ำ	สูง	มีการแพร่ระบาด	ต่ำ
ต่ำ	สูง	ไม่มีการแพร่ระบาด	ต่ำ

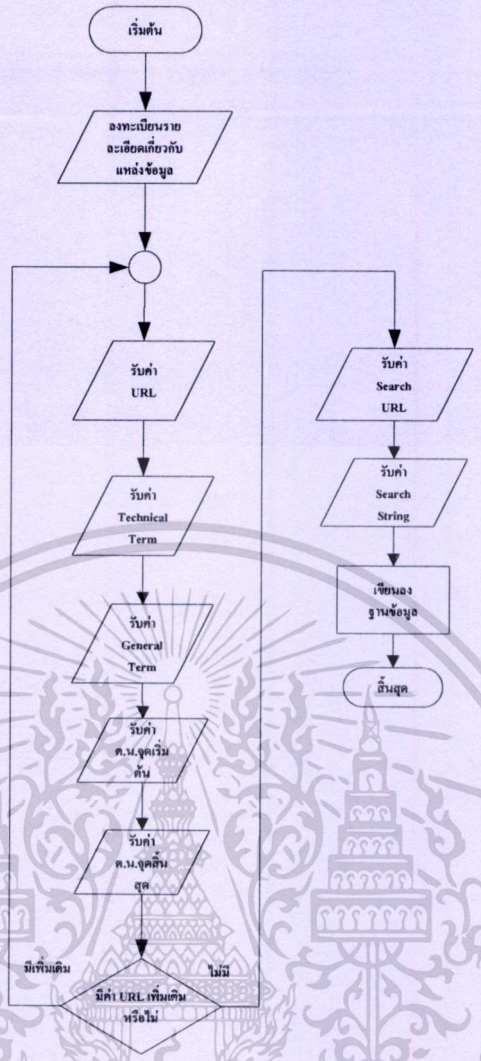
ตารางที่ 3.22 ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้ (ต่อ)

ระดับความเสียหายที่เกิดจากมัลแวร์	ระดับการแพร่กระจายของมัลแวร์	สถานะการแพร่กระจายของมัลแวร์	ระดับความเสี่ยงที่มีผลกระทบต่อผู้ใช้
ต่ำ	กลาง	มีการแพร่ระบาด	ต่ำ
ต่ำ	กลาง	ไม่มีการแพร่ระบาด	ต่ำ
ต่ำ	ต่ำ	มีการแพร่ระบาด	ต่ำ
ต่ำ	ต่ำ	ไม่มีการแพร่ระบาด	ต่ำ

จากการวิเคราะห์ระดับภัยคุกคามที่เกิดจากมัลแวร์นั้น จะมีส่วนช่วยให้ผู้ใช้ทราบสถานการณ์หรือผลกระทบที่เกิดจากมัลแวร์ในปัจจุบันได้และยังสามารถป้องกันตนเองจากการได้รับผลกระทบจากมัลแวร์ตัวดังกล่าวได้อีกด้วย

3.3.2 ระบบลงทะเบียนแหล่งข้อมูลคำอธิบายมัลแวร์ (Information Source Profile Registering)

ในการศึกษาคำอธิบายมัลแวร์สิ่งที่สำคัญที่สุดก็คือแหล่งข้อมูลที่สามารถเชื่อถือได้ ซึ่งแหล่งข้อมูลคำอธิบายมัลแวร์ในปัจจุบันนั้นมีด้วยกันหลากหลายแหล่ง ซึ่งแต่ละแหล่งข้อมูลนั้นจะมีลักษณะการเขียนข้อมูลที่แตกต่างกันออกไป บางแหล่งข้อมูลวิเคราะห์พฤติกรรมการแพร่กระจายของมัลแวร์ไว้ค่อนข้างละเอียด บางแหล่งข้อมูลมีรายละเอียดวิธีการแก้ไขปัญหาที่เกิดจากมัลแวร์ไว้ค่อนข้างมากและสามารถนำไปใช้งานได้จริง จึงเห็นได้ว่ายิ่งถ้ามีแหล่งข้อมูลคำอธิบายมัลแวร์เป็นจำนวนมากก็ยิ่งจะทำให้ผู้ใช้ที่เข้ามาอ่านได้รับความรู้ที่กว้างขวางและมีประโยชน์ยิ่งขึ้น ผู้เขียนจึงออกแบบ ระบบลงทะเบียนแหล่งข้อมูลคำอธิบายมัลแวร์ ซึ่งสามารถเพิ่มแหล่งข้อมูลคำอธิบายมัลแวร์จากแหล่งข้อมูลใหม่ ๆ เข้ามาในระบบได้หรือถ้าหากแหล่งคำอธิบายมัลแวร์ที่มีอยู่ได้มีการเปลี่ยนแปลงรูปแบบข้อมูลไปจากเดิมระบบก็สามารถที่จะแก้ไขโครงสร้างของคำอธิบายมัลแวร์ที่มีอยู่เดิมได้เช่นกัน โดยวิธีการทำงานของระบบได้แสดงไว้ในรูปที่ 3.13



รูปที่ 3.13 อธิบายการทำงานของระบบลงทะเบียนคำอธิบายมัลแวร์

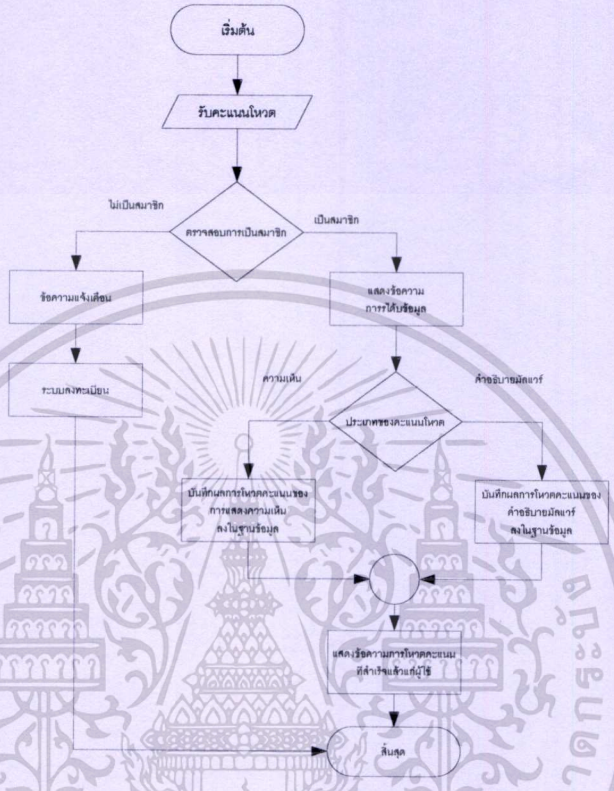
3.3.3 ระบบสมาชิกกลุ่มผู้ที่สนใจด้านมัลแวร์

เป็นระบบที่เปิดให้ผู้ใช้ที่สนใจด้านมัลแวร์สามารถสมัครเป็นสมาชิกของกลุ่มได้ผ่านทางเวปไซต์ โดยการลงทะเบียนนี้ผู้สมัครจะต้องกรอกข้อมูลรายละเอียดต่าง ๆ ที่ทางระบบกำหนดไว้ หลักจากที่ผู้ใช้ได้เป็นสมาชิกของกลุ่มแล้ว ผู้ใช้มีสิทธิในการร่วมการโหวตคะแนนให้กับคำอธิบายที่ตนเองเห็นว่ามีประโยชน์และสามารถนำไปใช้ได้จริง ซึ่งคะแนนดังกล่าวนี้จะถูกเก็บไว้ในฐานข้อมูล โดยในครั้งต่อไปถ้าผู้ใช้ท่านอื่นเข้ามาอ่านคำอธิบายมัลแวร์ข้อมูลคำอธิบายที่มีคะแนนโหวตสูงจะถูกแสดงให้อยู่เป็นลำดับแรก ซึ่งข้อมูลคำอธิบายมัลแวร์ที่มีคะแนนต่ำกว่าก็จะแสดงให้เห็นในลำดับต่อไป

- ระบบโหวตคะแนน

ในการทำงานของระบบโหวตคะแนนนั้น ระบบจะทำการตรวจสอบค่าที่ได้รับจากผู้ใช้ก่อนว่าเป็นสมาชิกของทางเวปไซต์หรือไม่ ถ้าไม่ใช่ระบบจะนำผู้ใช้ไปสู่หน้าลงทะเบียนสมาชิกเพื่อสมัครเป็นสมาชิกต่อไป ถ้าเป็นสมาชิกอยู่ก่อนแล้วระบบจะตรวจสอบดูว่าผู้ใช้ทำการ

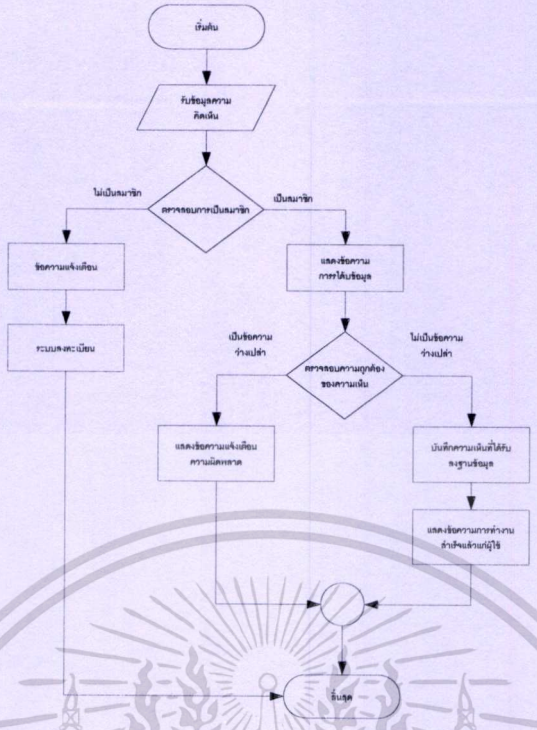
โหวตคะแนนให้กับความเห็นหรือคำอธิบายมัลแวร์และจัดเก็บรายละเอียดที่จำเป็นลงในฐานข้อมูลและแสดงข้อความแจ้งเตือนแก่ผู้ใช้ถึงการโหวตคะแนนของผู้ใช้นั้นได้ถูกบันทึกลงในฐานข้อมูลของระบบเรียบร้อยแล้ว



รูปที่ 3.14 อธิบายระบบโหวตคะแนนให้กับคำอธิบายมัลแวร์และความเห็น

- ระบบแสดงความเห็น

เมื่อผู้ใช้ทำการแสดงความคิดเห็นผ่านทางเวปไซต์แล้ว ระบบจะทำการตรวจสอบก่อนว่าผู้ใช้ที่แสดงความคิดเห็นกับทางระบบนั้นเป็นสมาชิกหรือไม่ ถ้าไม่ได้เป็นสมาชิกระบบจะแสดงข้อความแจ้งเตือนและนำผู้ใช้ไปสู่หน้าเวปไซต์การลงทะเบียนสมาชิกของระบบ แต่หากว่าผู้ใช้เป็นสมาชิก ระบบจะตรวจสอบต่อไปว่าข้อความเห็นที่สมาชิกโพสต์นั้นเป็นข้อความว่างเปล่าหรือไม่ ถ้าเป็นความว่างเปล่าระบบจะแสดงข้อความแจ้งเตือนความผิดพลาดให้กับสมาชิก แต่ถ้าข้อความเห็นที่โพสต์ไว้มีข้อมูลอยู่จริง ระบบจะนำข้อความดังกล่าวบันทึกลงฐานข้อมูลต่อไป และแสดงข้อความแจ้งการทำงานที่สำเร็จแก่ผู้ใช้ต่อไป



รูปที่ 3.15 อธิบายระบบการแสดงความเห็นของผู้ใช้

- ระบบผู้ดูแลระบบ

ผู้ดูแลระบบจะทำหน้าที่คอยตรวจสอบข้อมูลคำอธิบายมัลแวร์ที่มีการนำเข้ามาเก็บลงในฐานข้อมูลมัลแวร์ พร้อมกับทำหน้าที่คอยตรวจสอบความเห็นที่ผู้ใช้ได้แสดงไปในคำอธิบายมัลแวร์ต่าง ๆ และทำหน้าที่ตรวจสอบบันทึกการทำงานของโปรแกรมนำเข้าคำอธิบายมัลแวร์ด้วย โดยจะแบ่งการทำงานของผู้ดูแลระบบออกเป็นสองส่วนด้วยกัน

○ **ส่วนดูแลบริหารจัดการเกี่ยวกับผู้ใช้ (Manage user)**

ผู้ดูแลระบบสามารถเรียกดูประวัติการใช้งานของผู้ใช้แต่ละรายได้ โดยระบบจะแสดงให้เห็นถึงรายละเอียดต่าง ๆ ที่ผู้ใช้ได้บันทึกไว้ตอนลงทะเบียน เช่น ชื่อ นามสกุล ที่อยู่ เบอร์โทรศัพท์ อีเมล์ ที่สามารถติดต่อได้ เป็นต้นและผู้ดูแลระบบยังสามารถลบผู้ใช้ออกจากระบบสมาชิกได้หากมีความจำเป็น นอกจากนั้นยังมีการแสดงประวัติการเข้ามามีส่วนร่วมกับทางเวปไซต์อีกด้วย กล่าวคือ จะมีการบันทึกลงฐานข้อมูลว่า ผู้ใช้แต่ละคนนั้นมีการโหวตให้คะแนนกับคำอธิบายใดไปจำนวนกี่ครั้ง มีการโพสต์ข้อความแสดงความเห็นหรือไม่ มีจำนวนกี่ครั้ง ได้รับคะแนนโหวตกี่คะแนน เป็นต้น ซึ่งรายละเอียดเหล่านี้จะช่วยให้ผู้ดูแลระบบสามารถประเมินการมีส่วนร่วมของผู้ใช้กับทางเวปไซต์ได้และอาจจะมีการเลื่อนระดับให้กับผู้ใช้ทั่วไปสามารถดูแลระบบและบริหารจัดการเวปไซต์ด้วยเช่นกัน

○ **ส่วนดูแลบันทึกการทำงานของโปรแกรม (Log file monitor)**

เนื่องจากระบบนำเข้าข้อมูลคำอธิบายมัลแวร์นั้นจำเป็นต้องดึงคำอธิบายมัลแวร์จากแหล่งข้อมูลต่าง ๆ บนอินเตอร์เน็ตเป็นจำนวนมาก จึงจำเป็นต้องมีบันทึกการทำงานของโปรแกรมเพื่อให้ผู้ดูแลระบบสามารถตรวจสอบรายละเอียดการนำเข้าคำอธิบายแต่ละตัวได้ และในกรณีที่การนำเข้าข้อมูลมีปัญหาหรือเกิดความผิดพลาดขึ้น ผู้ดูแลระบบสามารถเข้ามาตรวจสอบดูบันทึกของโปรแกรมได้ว่ามีการนำเข้าคำอธิบายมัลแวร์ในเวลาใด มาจาก URL ใด ใครเป็นคนนำเข้า URL ดังกล่าว เป็นต้น

- ระบบผู้ใช้

เป็นระบบที่ผู้ใช้สามารถเข้ามาตรวจสอบดูประวัติของตนเองในการมีส่วนร่วมกับทางเวปไวต์ได้ โดยจะแบ่งเป็นสองส่วนด้วยกัน

○ **รายละเอียดของผู้ใช้ (User profile)**

ในส่วนนี้ผู้ใช้สามารถแก้ไขข้อมูลส่วนตัวได้ ยกตัวอย่างเช่น ชื่อ นามสกุล เบอร์โทรศัพท์ อีเมล์ เป็นต้น

○ **รายละเอียดในการมีส่วนร่วมกับเวปไซต์ (User history)**

ในส่วนนี้ผู้ใช้สามารถดูประวัติของตนเองว่ามีการโพสต์แสดงความเห็นไปแล้วกี่ความเห็น ตนเองได้รับคะแนนโหวตจากผู้ใช้คนอื่นแล้วกี่คะแนนและได้โหวตคะแนนให้กับคำอธิบายมัลแวร์หรือความเห็นไปแล้วกี่คะแนน มีการนำเข้าคำอธิบายมัลแวร์แล้วจำนวนเท่าไหร่ เป็นต้น

บทที่ 4

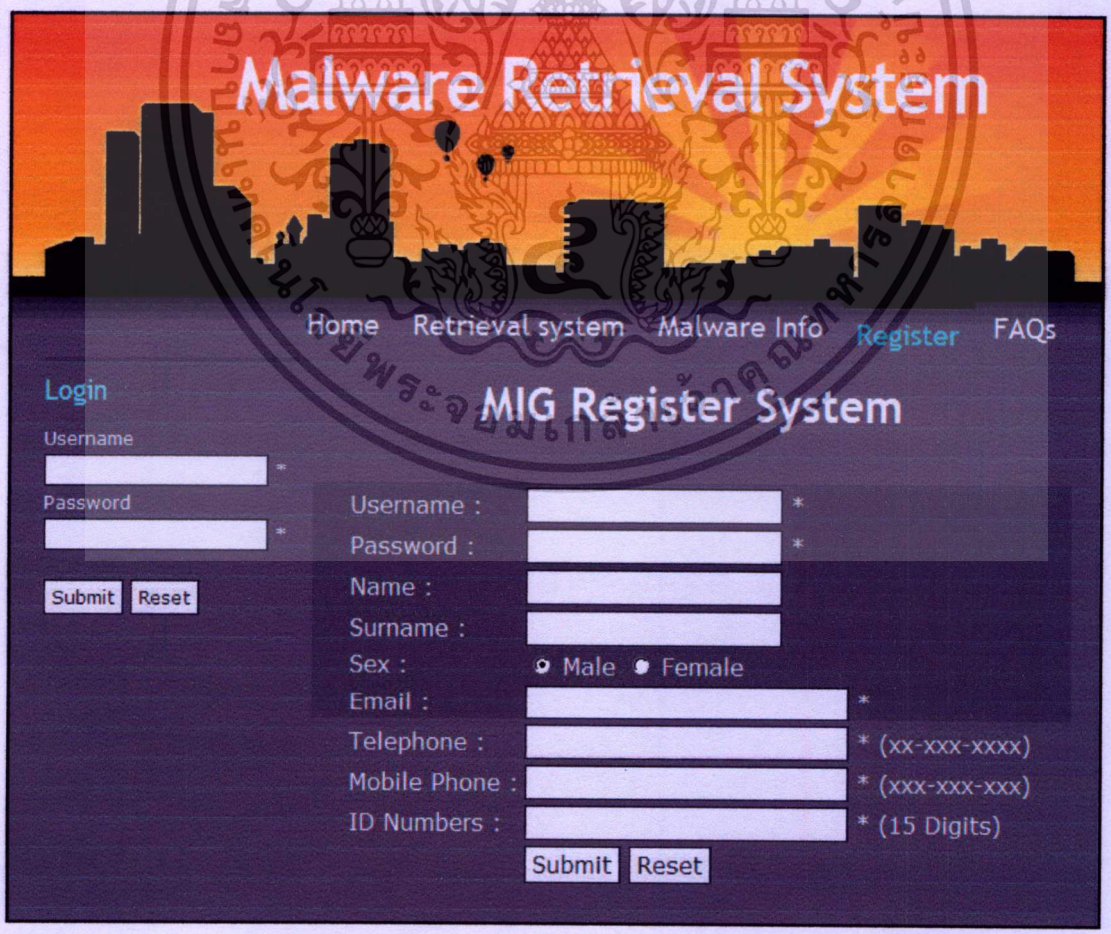
ผลการทดลอง

4.1 การทดลองและผลการทดลองของระบบลงทะเบียน

ระบบลงทะเบียนในเวปไซต์จะมีด้วยกันสองประเภท ประเภทแรกจะเป็นระบบลงทะเบียนสมาชิกผู้ที่สนใจด้านมัลแวร์ และประเภทที่สองจะเป็นการลงทะเบียนเพิ่มแหล่งข้อมูลคำอธิบายมัลแวร์ใหม่เข้ามาในระบบ

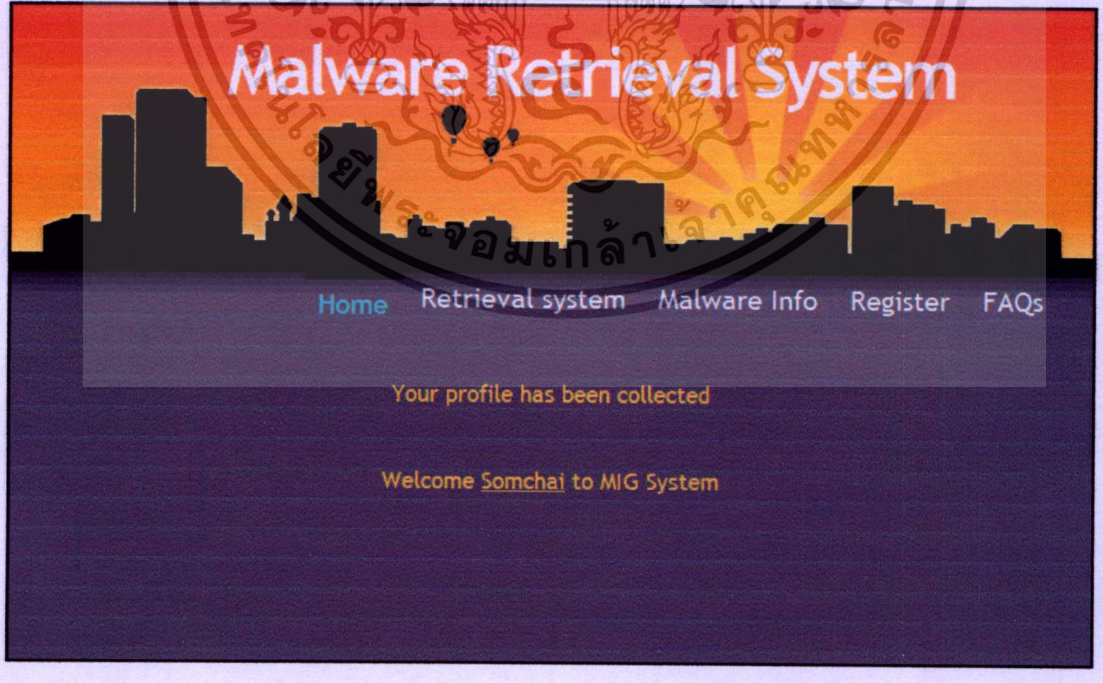
4.1.1 การลงทะเบียนสมัครสมาชิกกลุ่มผู้สนใจด้านมัลแวร์

ก่อนที่ผู้ใช้จะสามารถมีส่วนร่วมกับทางเวปไซต์ได้นั้นจำเป็นต้องลงทะเบียนสมัครเป็นสมาชิกของกลุ่มผู้สนใจด้านมัลแวร์ก่อน โดยจะต้องกรอกข้อมูลที่จำเป็นในการตรวจสอบลงในแบบฟอร์มตามรูปที่ 4.1

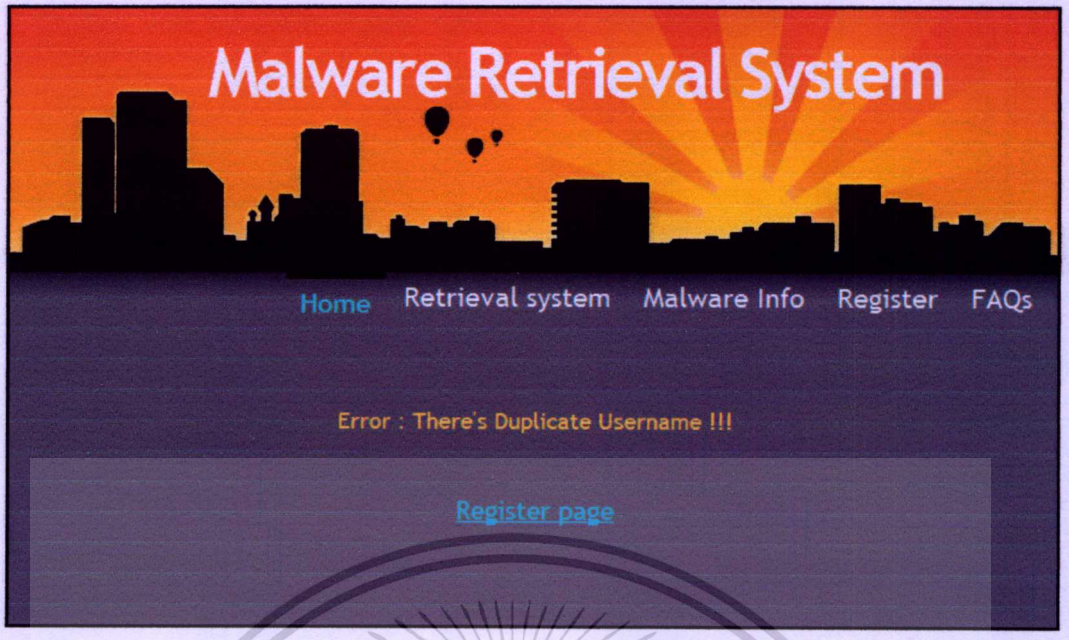


รูปที่ 4.1 หน้าเวปไซต์สมัครสมาชิกกลุ่มผู้สนใจด้านมัลแวร์

รูปที่ 4.2 แบบฟอร์มสำหรับสมัครสมาชิกกลุ่มผู้สนใจด้านมัลแวร์



รูปที่ 4.3 หน้าเวปเพจแสดงข้อความเมื่อผู้ใช้กรอกข้อมูลถูกต้อง



รูปที่ 4.4 หน้าเวปเพจแสดงข้อความเมื่อผู้ใช้กรอกข้อมูลไม่ถูกต้อง

รูปที่ 4.5 หน้าเวปเพจที่ผู้ใช้ล็อกอินเข้ามาในระบบด้วยชื่อที่ลงทะเบียนแล้ว

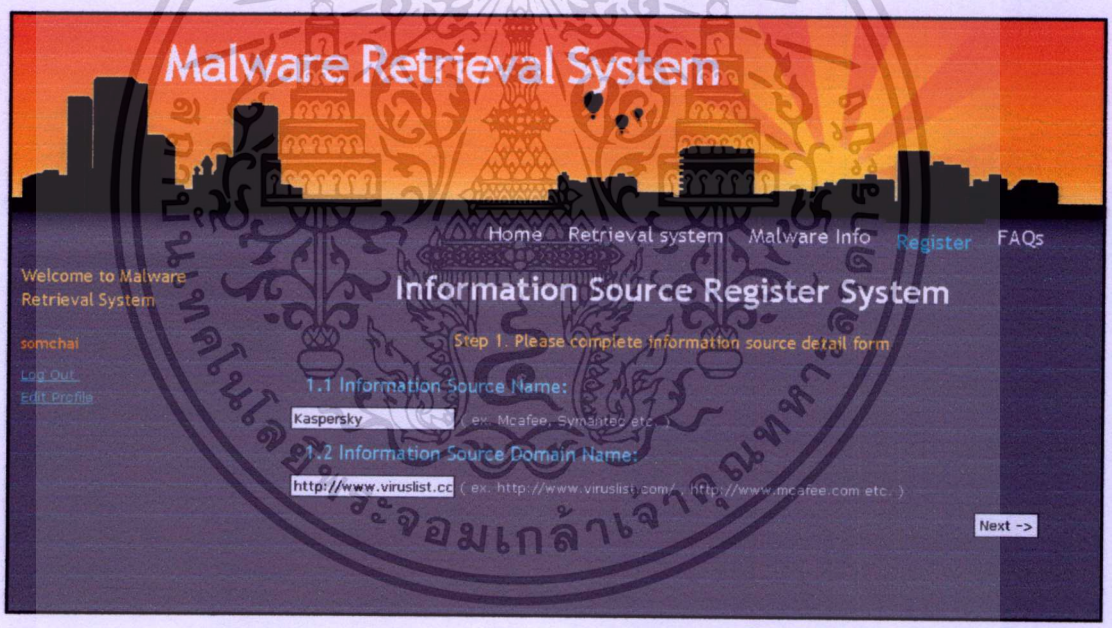
4.1.2 การลงทะเบียนเพิ่มแหล่งข้อมูลคำอธิบายมัลแวร์

ระบบนี้อนุญาตให้ผู้ใช้ที่เป็นสมาชิกสามารถเพิ่มแหล่งข้อมูลมัลแวร์แหล่งใหม่เข้ามาในฐานข้อมูลได้ โดยระบบลงทะเบียนจะทำการเก็บข้อมูลรายละเอียดของเวปไซต์ยกตัวอย่างเช่น ชื่อของแหล่งข้อมูล ตำแหน่งจุดเริ่มต้นและจุดสิ้นสุดของข้อมูลแต่ละชุด คำศัพท์เทคนิค เป็นต้น ระบบลงทะเบียนเพิ่มแหล่งข้อมูลคำอธิบายมัลแวร์จะประกอบด้วยขั้นตอนต่าง ๆ ดังนี้

○ **ขั้นตอนที่ 1** เป็นขั้นตอนที่ผู้ใช้ต้องใส่รายละเอียดเกี่ยวกับแหล่งข้อมูลลงในระบบซึ่งรายละเอียดนั้นมีดังนี้

- ชื่อของแหล่งข้อมูล (Information Source name) ผู้ใช้ต้องใส่ชื่อของแหล่งข้อมูลใหม่ที่จะนำมาเพิ่มเติมในฐานข้อมูลลงในช่องรับข้อมูลที่ 1.1 ดังแสดงในรูป 4.6

- โดเมนเนมของแหล่งข้อมูล (Information Source Domain Name) ผู้ใช้ต้องใส่โดเมนเนมเวปไซต์ของแหล่งข้อมูลคำอธิบายมัลแวร์



รูปที่ 4.6 ขั้นตอนการเพิ่มแหล่งข้อมูลมัลแวร์ในเวปไซต์ขั้นตอนแรก

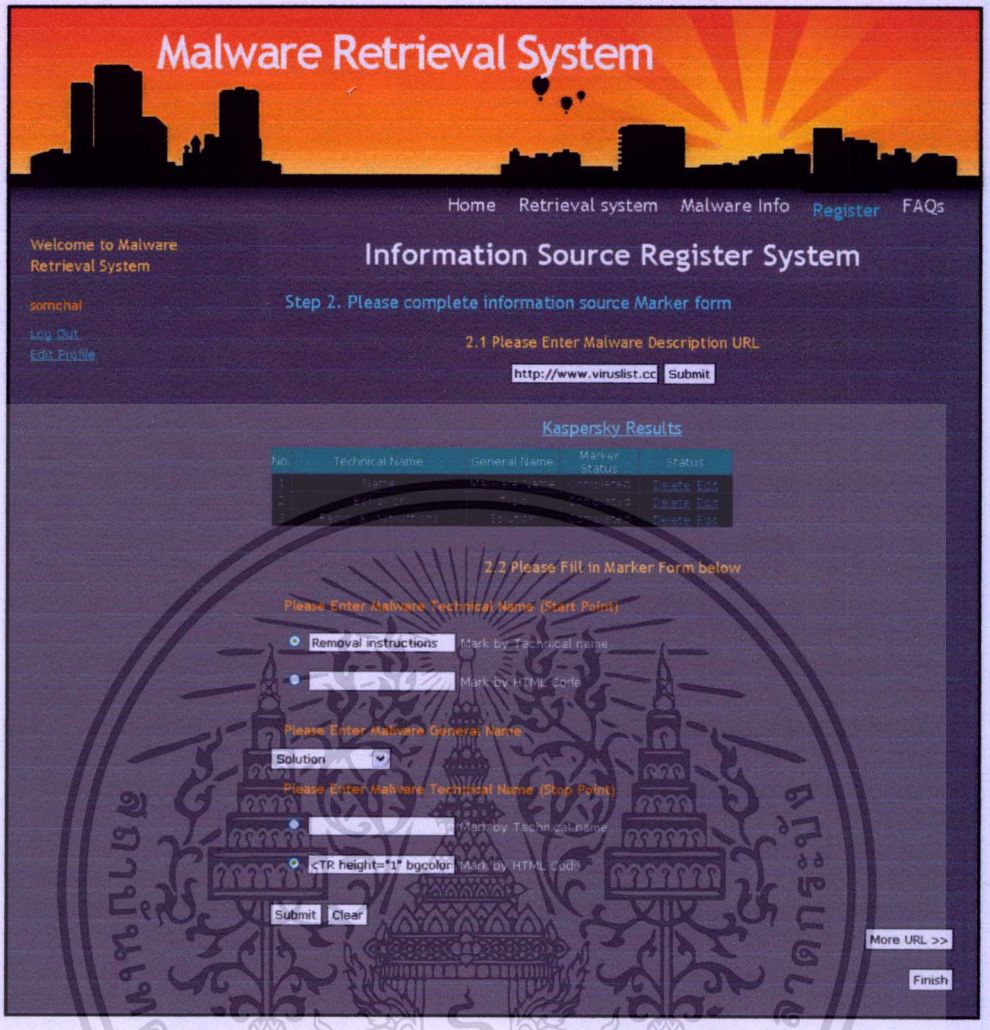
○ **ขั้นตอนที่ 2** ขั้นตอนนี้นับได้ว่ามีส่วนสำคัญที่สุดในการลงทะเบียนเพิ่มแหล่งข้อมูล เนื่องจากว่าผู้ใช้จะต้องกรอกข้อมูลรายละเอียดและตำแหน่งที่เป็นจุดเริ่มต้นและจุดสิ้นสุดของแต่ละชุดข้อมูล เพื่อให้โปรแกรมสามารถนำรวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ของแหล่งข้อมูลดังกล่าวมาเก็บไว้ในฐานข้อมูลได้ โดยรายละเอียดดังกล่าวจะมีดังนี้

- URL คำอธิบายมัลแวร์ (Malware Description URL) ในหัวข้อ 2.1 จากรูปที่ 4.7 นี้ผู้ใช้ต้องใส่ URL คำอธิบายมัลแวร์ของแหล่งข้อมูลดังกล่าวซึ่งจะต้องเป็นคำอธิบายข้อมูลที่มีอยู่จริง จากนั้นระบบจะเปิดหน้าต่างเวปเบราว์เซอร์ของคำอธิบายข้อมูลมัลแวร์ใหม่ขึ้นมา เพื่อให้ผู้ใช้สามารถนำข้อมูลที่ต้องการใส่ในแบบฟอร์มต่อไปได้

- ใส่ข้อมูลตัวกำหนดขอบเขตข้อมูล ในหัวข้อ 2.2 จากรูปที่ 4.7 นี้ผู้ใช้จะต้องใส่ข้อมูลที่จำเป็นด้วยกัน 3 อย่างคือ

- **คำศัพท์เทคนิค (Technical Name)** เป็นคำศัพท์เทคนิคที่ใช้ในการตั้งชื่อแอทริบิวต์ของแต่ละแหล่งข้อมูลที่แตกต่างกันออกไป ซึ่งคำศัพท์เทคนิคที่ปรากฏในเวปไซต์คำอธิบายมัลแวร์นั้นคือตำแหน่งที่เป็นตัวกำหนดจุดเริ่มต้นและจุดสิ้นสุดของข้อมูลแต่ละชุด โดยในรูปที่ 4.7 นั้น ผู้ใช้สามารถเลือกกำหนดขอบเขตข้อมูลได้โดยการใส่คำศัพท์เทคนิคเพียงหรืออาจจะใช้โค้ดของภาษา HTML เป็นตัวกำหนดขอบเขตข้อมูลก็ได้
- **คำศัพท์สามัญ (General Name)** เป็นคำศัพท์สามัญที่เราได้ออกแบบไว้ในบทที่ 3 ซึ่งเป็นคำที่สามารถสื่อถึงความหมายของข้อมูลได้ดี โดยในหัวข้อนี้ระบบจะสนับสนุนการใช้งานของผู้ใช้โดยการแสดงผลข้อมูลด้วย กล่องแสดงรายการ (List box) ซึ่งจะเป็นคำศัพท์สามัญที่มีอยู่ในฐานข้อมูล เพื่อให้ผู้ใช้สามารถเลือกคำศัพท์เหล่านี้ได้โดยง่าย

โดยในหัวข้อที่ 2.2 จากรูปที่ 4.7 นี้ผู้ใช้จะต้องกรอกข้อมูลตัวกำหนดขอบเขตข้อมูลคำศัพท์เทคนิค และคำศัพท์สามัญ ที่มีอยู่ในแหล่งข้อมูลใหม่ที่ต้องการจะเพิ่มลงในฐานข้อมูลให้ครบทุกตัว ซึ่งระบบสามารถรองรับการใส่ตัวกำหนดขอบเขตข้อมูลซึ่งเป็นชื่อคำศัพท์เทคนิคโดยตรงหรือแท็คภาษา HTML จากนั้นรายละเอียดของผลที่บันทึกลงในฐานข้อมูลจะแสดงให้เห็นในตาราง Results ซึ่งผู้ใช้สามารถแก้ไขหรือลบรายการข้อมูลได้ด้วยตนเอง ในกรณีที่แหล่งข้อมูลคำอธิบายมัลแวร์นั้นใช้ URL มากกว่าหนึ่งค่าในการแสดงคำอธิบายมัลแวร์ ผู้ใช้สามารถคลิ๊กที่ปุ่ม “More URL>>” เพื่อเข้าสู่หน้าเวปแพจของขั้นตอนที่สองใหม่อีกครั้งหนึ่ง เพื่อกรอกข้อมูลในหัวข้อที่ 2.1 ในรูปที่ 4.7 ด้วยค่า URL ซึ่งเพิ่มขึ้นใหม่ จากนั้นจึงกรอกข้อมูล 2.2 ในรูปที่ 4.7 ต่อไป เมื่อผู้ใช้ได้กรอกข้อมูลครบทุก URL ที่ใช้แสดงค่าคำอธิบายมัลแวร์แล้ว สามารถกดปุ่ม “Finish” เพื่อเข้าสู่ขั้นตอนต่อไป

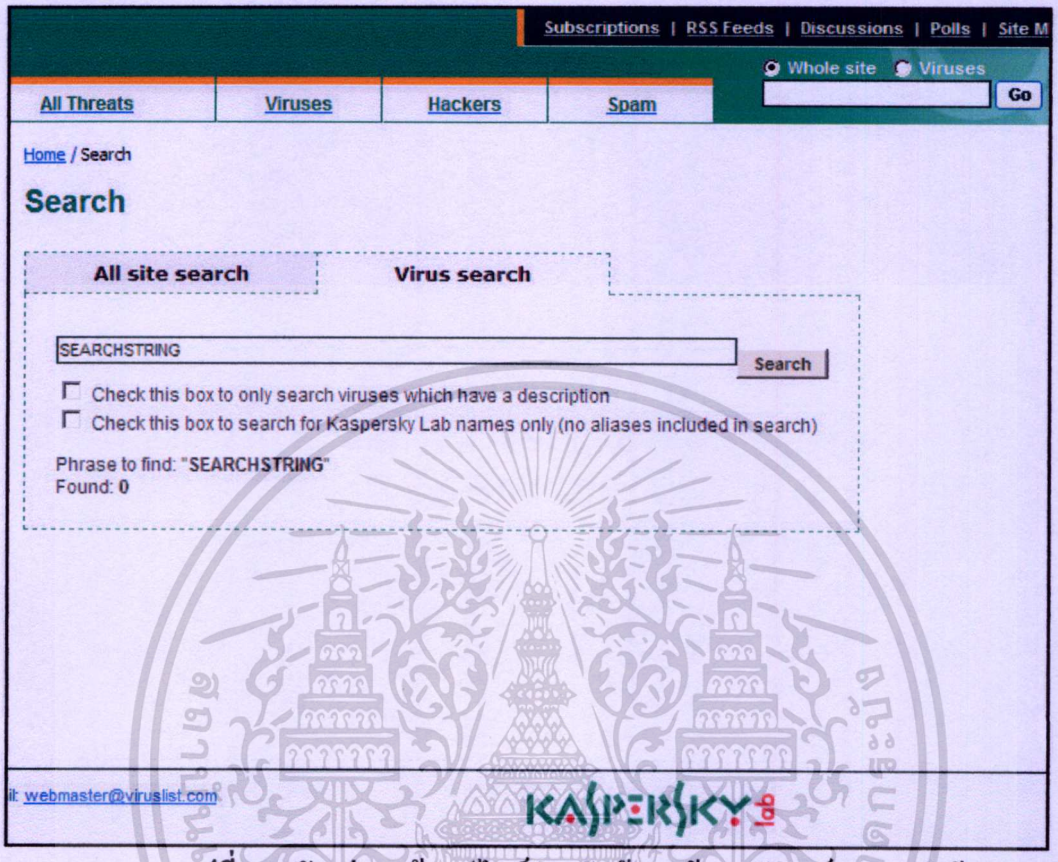


รูปที่ 4.7 ขั้นตอนการเพิ่มแหล่งข้อมูลมัลแวร์ในเวปไซต์ขั้นตอนที่สอง

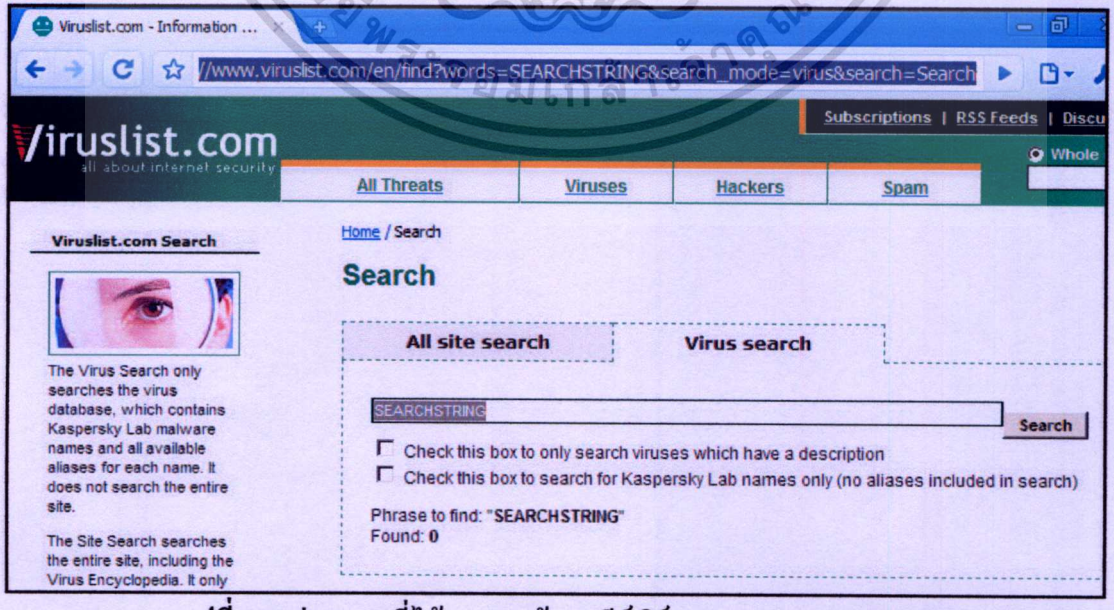
○ **ขั้นตอนที่ 3** ขั้นตอนนี้เป็นการลงทะเบียนสำหรับการค้นหาข้อมูลมัลแวร์จากในเวปไซต์ของแหล่งข้อมูล กล่าวคือ ในการทำงานของโปรแกรมรวบรวมคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ นั้นจะต้องนำชื่อหรือชื่อเรียกของมัลแวร์ไปค้นหาในเวปไซต์ของแหล่งข้อมูลต่าง ๆ เสมอ เพื่อให้ได้ข้อมูลคำอธิบายมัลแวร์จากแหล่งข้อมูลที่หลากหลายมากยิ่งขึ้น โดยมีขั้นตอนการลงทะเบียนดังนี้

ในตอนเริ่มแรกนั้นให้ผู้ใช้เปิดเวปเพจสำหรับค้นหาข้อมูลมัลแวร์ของแหล่งข้อมูลที่กำลังลงทะเบียนขึ้นมา จากนั้นให้ผู้ใช้พิมพ์คำว่า “SEARCHSTRING” ช่องค้นหาข้อมูลดังแสดงในรูปที่ 4.8 เมื่อค้นหาข้อมูลเสร็จแล้วให้ผู้ใช้คัดลอก URL ของเวปเพจนั้นซึ่งจะอยู่ในช่องแอดเดรสบาร์ของโปรแกรมบราวเซอร์ซึ่งแสดงในรูปที่ 4.9 มาใส่ลงในหัวข้อ 3.1 ตามรูปที่ 4.10 ต่อจากนั้นให้ผู้ใช้กดปุ่ม “Finish” เพื่อเป็น

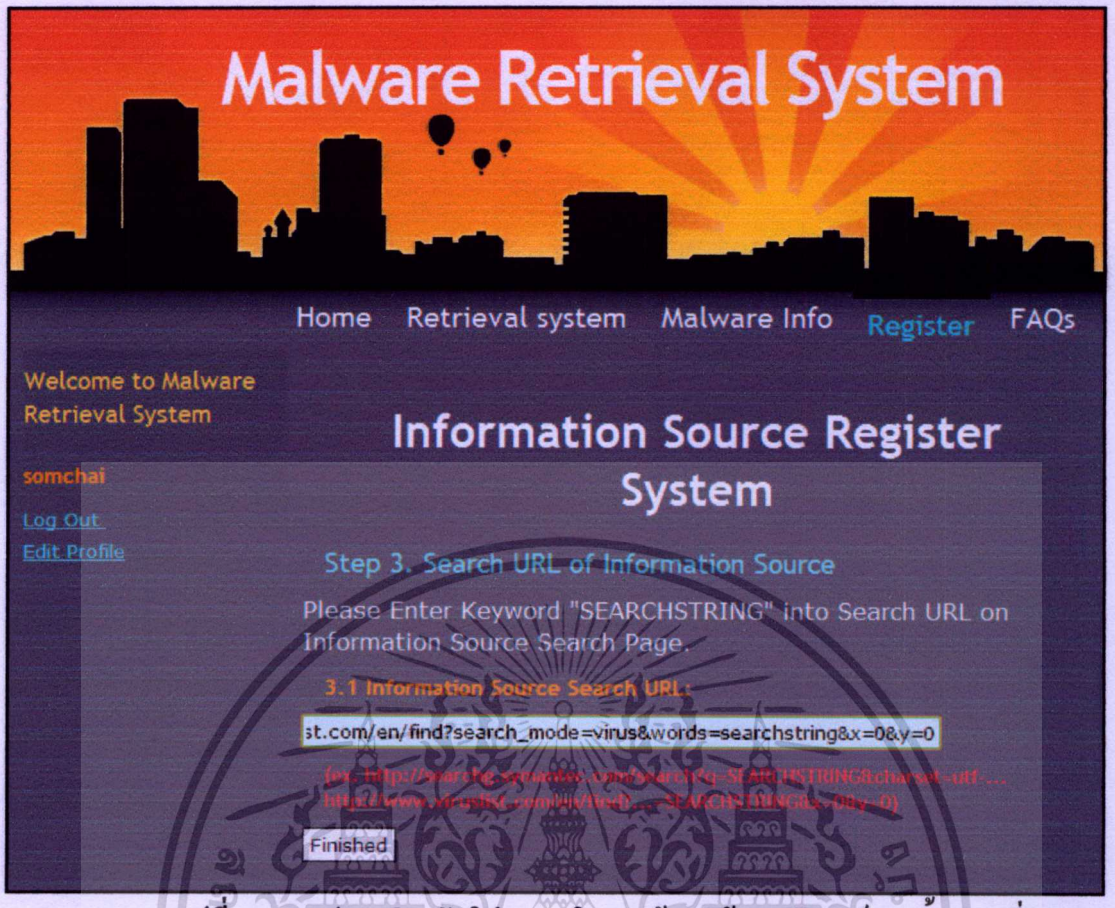
การจบการทำงานในขั้นตอนที่ 3 ซึ่งระบบจะทำการบันทึกรายละเอียดต่าง ๆ ที่ผู้ใช้ได้ลงทะเบียนเอาไว้ในฐานข้อมูลต่อไป



รูปที่ 4.8 ตัวอย่างหน้าเวปไซต์สำหรับค้นหาข้อมูลมัลแวร์ของแหล่งข้อมูลคำอธิบายมัลแวร์



รูปที่ 4.9 ค่า URL ที่ได้จากการค้นหาคีย์เวิร์ด “SEARCHSTRING” จะถูกแสดงไว้ด้านบนในช่องแอดเดรสบาร์

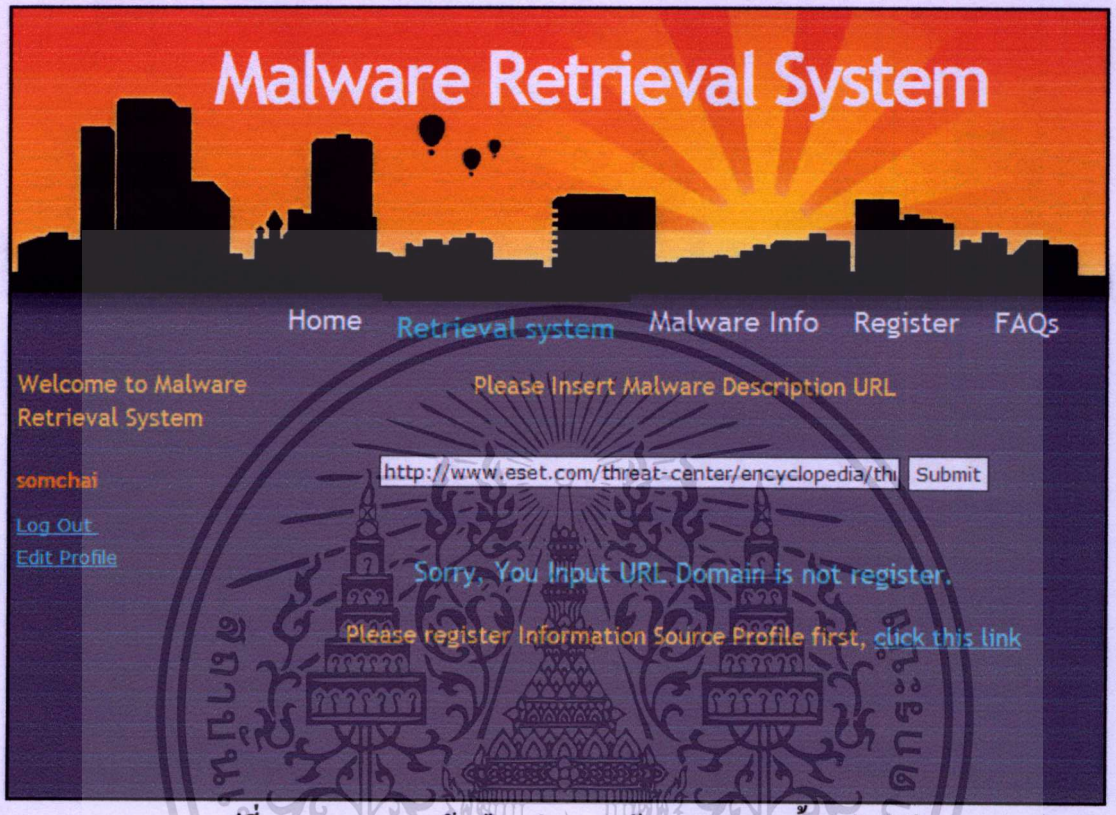


รูปที่ 4.10 เวปเพจสำหรับใส่ URL ในการค้นหาข้อมูลมัลแวร์ของขั้นตอนที่สาม

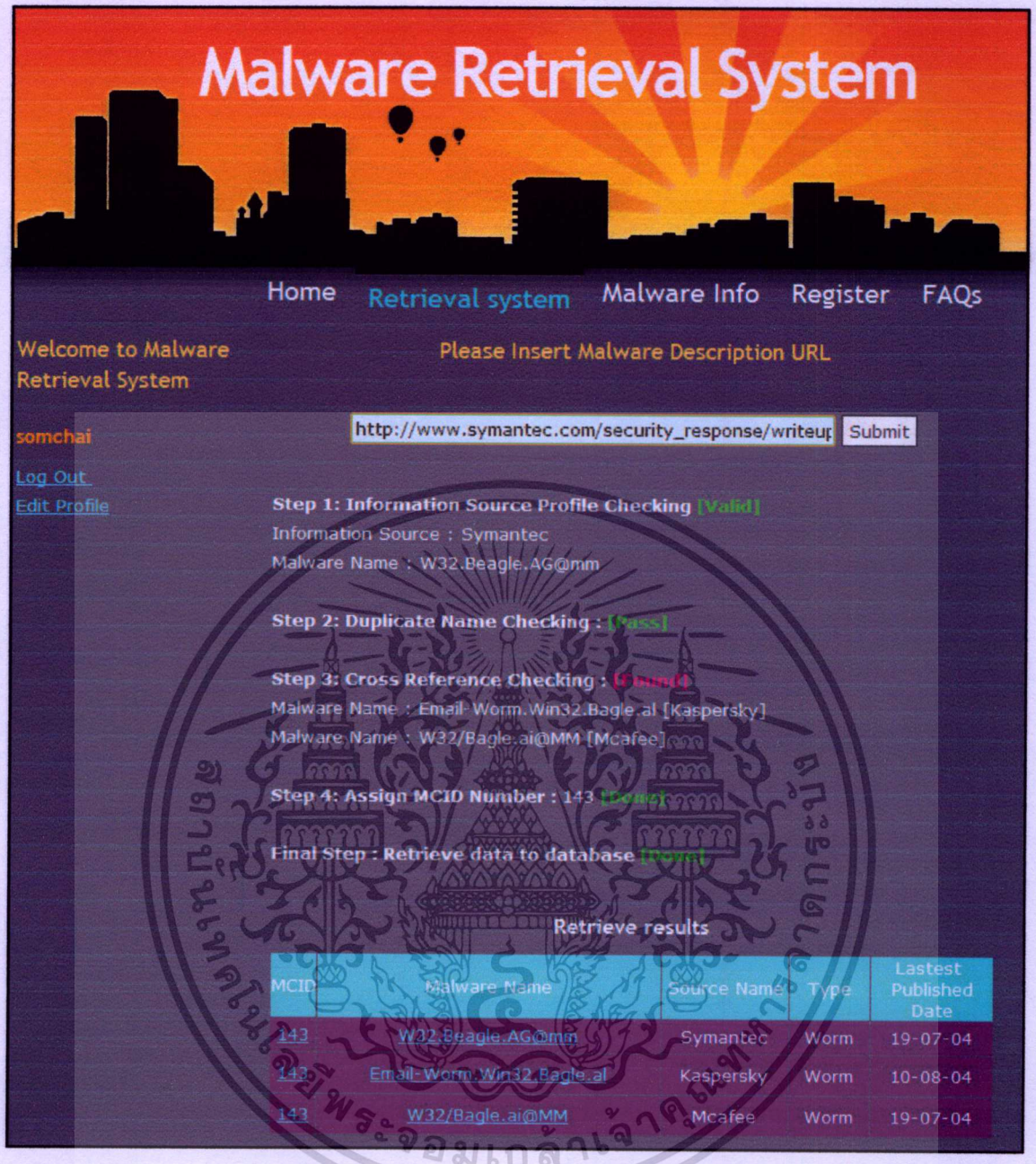
4.2 การทดลองและผลการทดลองโปรแกรมรวบรวมข้อมูลคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ

ในระบบสืบค้นข้อมูลมัลแวร์นี้ผู้ใช้สามารถนำเข้าคำอธิบายมัลแวร์จากเวปไซต์ของผู้ผลิตโปรแกรมกำจัดมัลแวร์แหล่งต่าง ๆ ได้ โดยที่ผู้ใช้แค่ทำการคัดลอก URL คำอธิบายของมัลแวร์ตัวดังกล่าวมาไว้ในช่องใส่ URL จากนั้นกดปุ่ม submit ซึ่งแสดงในรูปที่ 4.10 โปรแกรมก็จะเริ่มทำการตรวจสอบในขั้นตอนที่ 1 โดยระบบจะทำการตรวจสอบดูว่าโดเมนของ URL ค่าดังกล่าวนั้นได้มีในข้อมูลลงทะเบียนของแหล่งข้อมูลแล้วหรือไม่ ถ้ายังไม่มีระบบจะแจ้งเตือนผู้ใช้ให้ลงทะเบียนรายละเอียดแหล่งข้อมูลเสียก่อนดังแสดงในรูปที่ 4.10 แต่ถ้าโดเมนของเวปไซต์ดังกล่าวนั้นได้ลงทะเบียนกับระบบไว้ก่อนแล้วก็จะทำการประมวลผลขั้นที่ 2 ต่อไปโดยจะตรวจสอบดูว่าชื่อมัลแวร์และแหล่งข้อมูลที่จะนำเข้ามาใหม่นั้นมีอยู่ในฐานข้อมูลแล้วหรือยัง ถ้ายังโปรแกรมจะดำเนินการขั้นที่ 3 คือ นำชื่อและชื่อเรียกของมัลแวร์ตัวดังกล่าวไปตรวจสอบในเวปไซต์อื่น ๆ ที่ได้ลงทะเบียนกับทางระบบไว้ก่อนแล้ว เมื่อพบว่ามีข้อมูลมัลแวร์ที่เป็นตัวเดียวกันแต่ถูกเรียกด้วยชื่อที่ต่างกันแล้วนั้น โปรแกรมจะกำหนดหมายเลข MCID ใหม่ให้กับข้อมูลคำอธิบายมัลแวร์เหล่านั้น เพื่อให้ผู้ใช้สามารถอ้างถึงมัลแวร์ที่มีชื่อที่แตกต่างกันแต่เป็นมัลแวร์ตัวเดียวกันได้

ภายใต้ MCID เดียวกัน ต่อจากนั้นโปรแกรมจะวิเคราะห์คีย์เวิร์ดในคำอธิบายมัลแวร์เหล่านี้และบันทึกข้อมูลของมัลแวร์ลงในฐานข้อมูลซึ่งผู้ใช้สามารถเรียกดูคำอธิบายมัลแวร์เหล่านี้ได้จากระบบการแสดงผล ซึ่งจะกล่าวถึงในหัวข้อต่อไป โดยผลลัพธ์ที่ได้จากขั้นตอนการทำงานต่าง ๆ ของจะแสดงในรูปที่ 4.12



รูปที่ 4.11 ระบบจะแจ้งเตือนว่าแหล่งข้อมูลดังกล่าวนั้นยังไม่ได้ลงทะเบียน

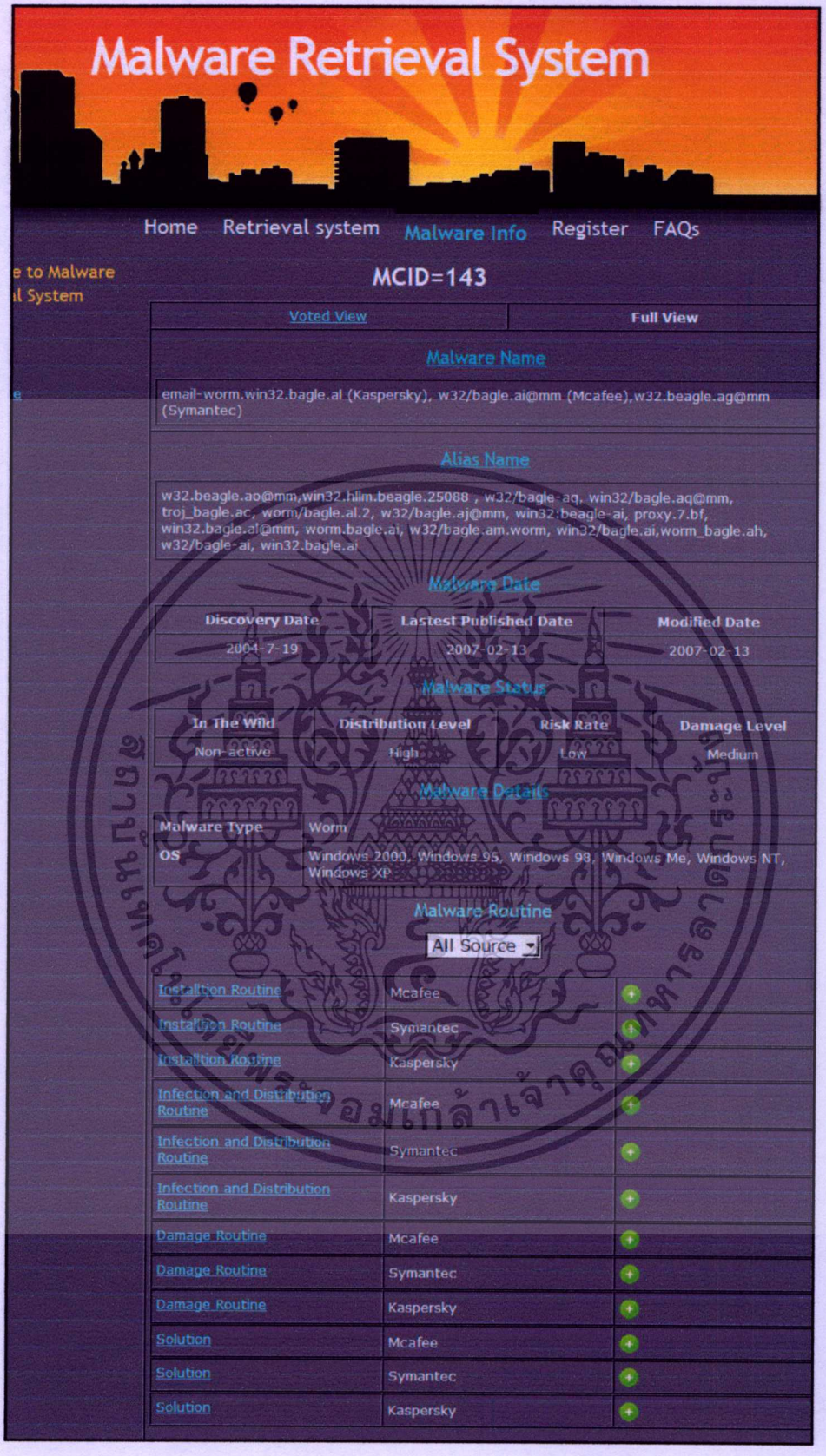


รูปที่ 4.12 ผลการทดลองการนำเข้าคำอธิบายมัลแวร์จากค่ายต่าง ๆ

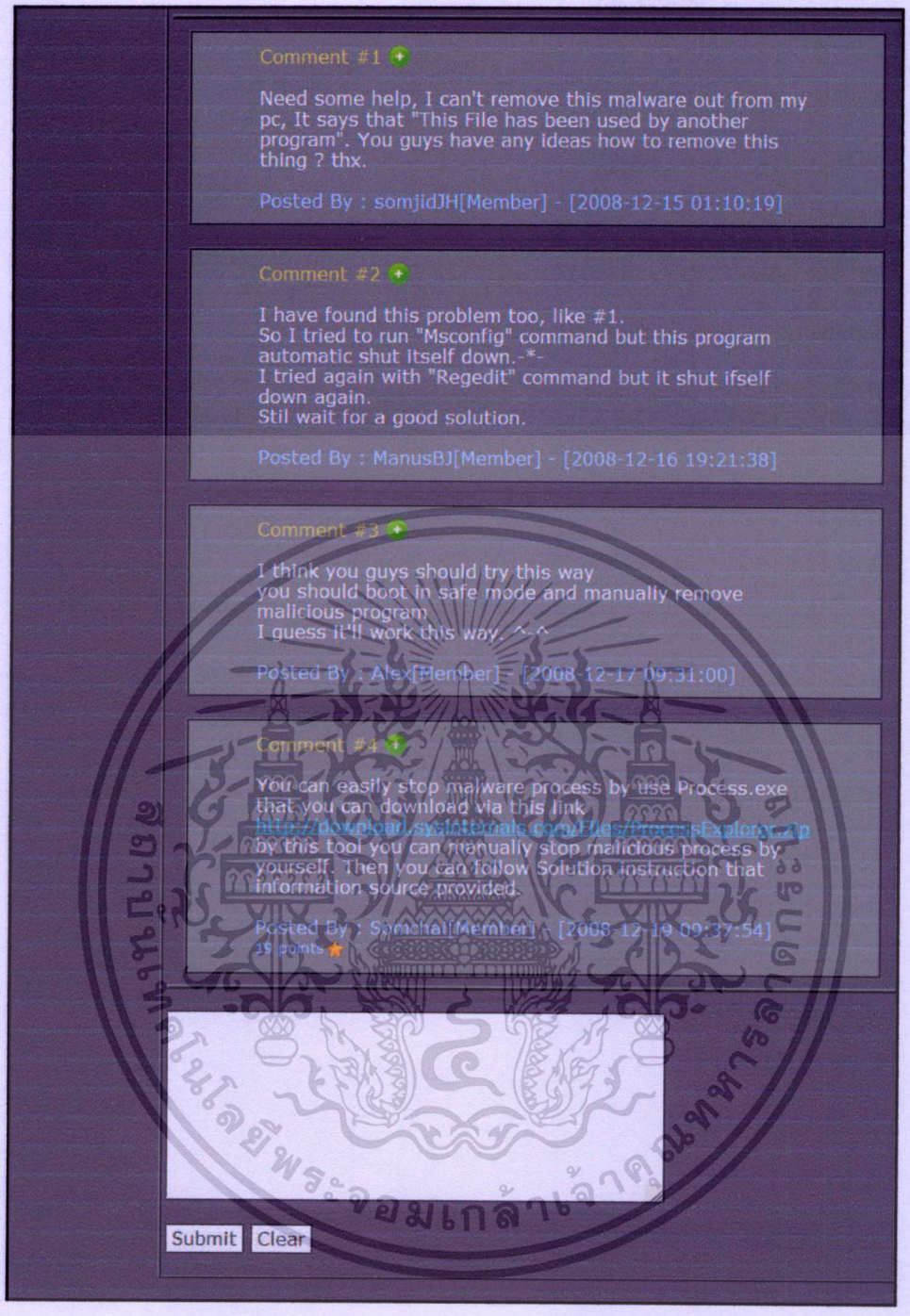
4.3 ระบบการแสดงผล (Results Viewer system)

เพื่อเป็นการอำนวยความสะดวกแก่ผู้ใช้ในการเรียกดูข้อมูลคำอธิบายมัลแวร์ เราได้นำเสนอการแสดงผลให้มีสองชนิดคือ การแสดงผลอย่างเต็มรูปแบบ (Full View) และการแสดงผลตามคะแนนโหวต (Vote View)

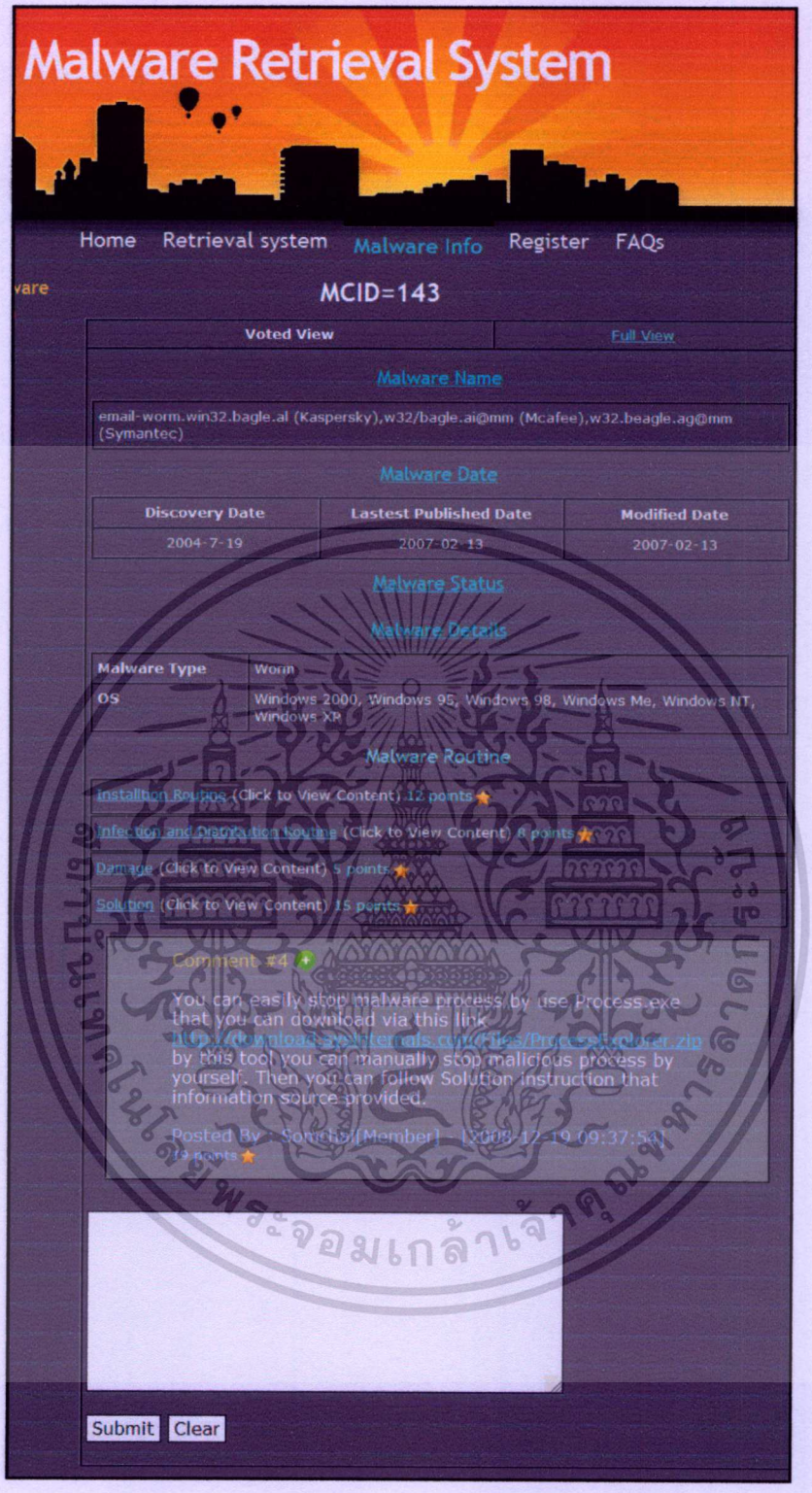
- **การแสดงผลอย่างเต็มรูปแบบ (Full view)** จะแบ่งรายละเอียดของข้อมูลที่จะนำมาแสดงออกเป็นสองส่วนด้วยกัน กล่าวคือ ส่วนแรกจะเป็นข้อมูลรายละเอียดทั้งหมดของมัลแวร์ซึ่งเป็นผลลัพธ์ที่ระบบได้ทำการนำเข้าคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาเก็บลงในฐานข้อมูล ซึ่งรายละเอียดข้อมูลดังกล่าวนี้จะประกอบด้วยข้อมูล 6 ประเภท ได้แก่ วันที่, สถานะต่าง ๆ, รายละเอียดทางเทคนิค, พฤติกรรมและการทำงานของมัลแวร์ และวิธีการแก้ไขปัญหาจากมัลแวร์ ดังแสดงในรูปที่ 4.13 – 4.14 ซึ่งผู้ใช้สามารถกดเรียกที่หัวข้อในการอ่านข้อมูลได้และสามารถกดที่หัวข้อเพื่อซ่อนข้อมูลดังกล่าวเมื่ออ่านเสร็จแล้ว ซึ่งจะช่วยให้ผู้ใช้ไม่ต้องอ่านข้อมูลที่มีจำนวนมากพร้อม ๆ กัน ส่วนที่สองจะเป็นการแสดงข้อมูลการแสดงความเห็นเกี่ยวกับมัลแวร์จากผู้ใช้ท่านต่าง ๆ ที่เคยโพสต์เอาไว้ ซึ่งข้อมูลส่วนนี้นับว่ามีประโยชน์มาก เนื่องจากข้อมูลในส่วนนี้เป็นข้อมูลที่ได้มาจากประสบการณ์ใช้งานจริง ซึ่งผู้ใช้ที่ได้อ่านข้อความเห็นเหล่านี้สามารถนำไปใช้แก้ไขปัญหาได้จริง
- **การแสดงผลตามคำแนนโหวต (Voted view)** เป็นการแสดงผลลัพธ์ข้อมูลคำอธิบายมัลแวร์และความเห็นที่ได้รับการโหวตแล้วจากผู้ใช้ โดยระบบจะแสดงคำอธิบายและความเห็นที่ได้รับคะแนนโหวตสูงขึ้นมาเป็นอันดับแรก เพื่อให้ผู้ใช้สามารถประหยัดเวลาในการอ่านได้ เนื่องจากคำอธิบายมัลแวร์ที่ได้รับการโหวตให้คะแนนจะเป็นคำอธิบายที่ครอบคลุม เข้าใจได้ง่ายและมีรายละเอียดครอบถ้วน ตลอดจนความเห็นที่มีประโยชน์ต่าง ๆ ที่ได้รับการโหวตให้คะแนนก็จะนำมาแสดงเป็นอันดับแรกเช่นกัน ดังนั้นระบบนี้จึงช่วยให้ผู้ใช้สามารถอ่านคำอธิบายมัลแวร์และความเห็นที่มีประโยชน์ได้ในระยะเวลาอันสั้น เพื่อให้ผู้ใช้สามารถอ่านข้อมูลที่มีประโยชน์มากที่สุด ดังแสดงในรูปที่ 4.15



รูปที่ 4.13 การแสดงข้อมูลอย่างเต็มรูปแบบ



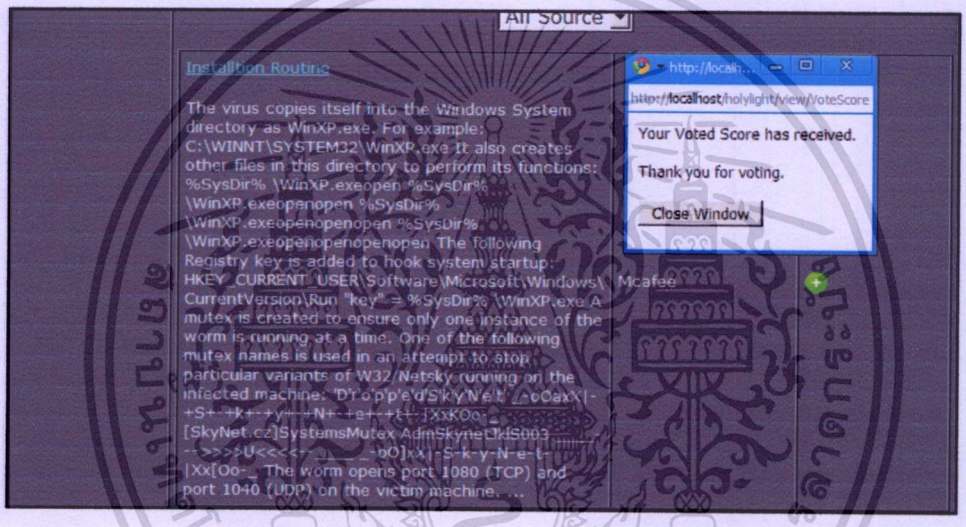
รูปที่ 4.14 การแสดงข้อมูลอย่างเต็มรูปแบบ (ต่อ)



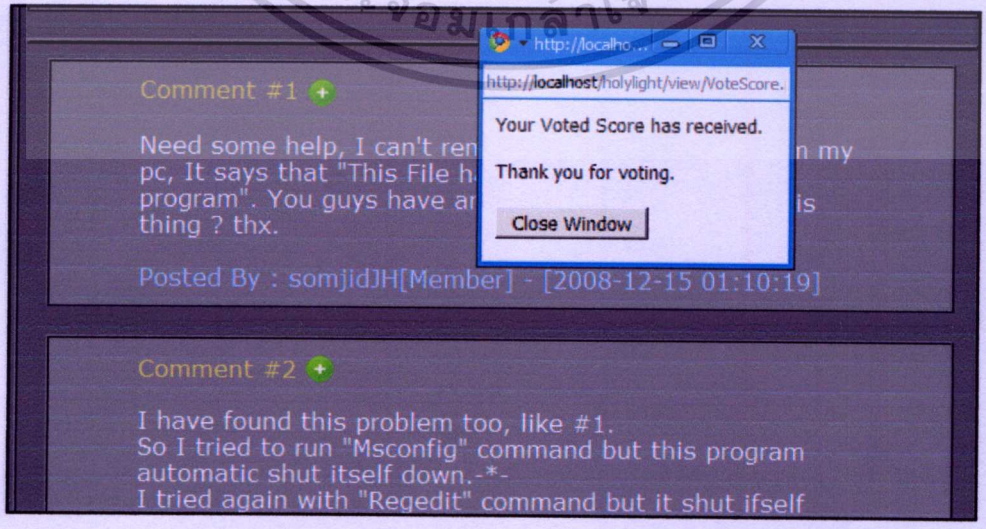
รูปที่ 4.15 ตัวอย่างคำอธิบายที่แสดงผลที่ได้รับการโหวตแล้ว

4.4 การทดลองและผลการทดลองระบบโหวตคะแนน (Malware Voting System)

เนื่องจากแต่ละแหล่งข้อมูลนั้นมีการอธิบายมัลแวร์ตัวเดียวกันด้วยรายละเอียดที่แตกต่างกันไป บางแหล่งข้อมูลให้รายละเอียดเกี่ยวกับวิธีแก้ไขค่อนข้างละเอียด บางแหล่งข้อมูลให้รายละเอียดเกี่ยวกับข้อมูลทางเทคนิคของมัลแวร์ที่ชัดเจน ในขณะที่บางค่ายไม่ได้อธิบายรายละเอียดในการติดตั้งหรือการแพร่กระจายเอาไว้ เราจึงเปิดโอกาสให้ผู้ใช้ที่เข้ามาอ่านสามารถโหวตให้คะแนนกับคำอธิบายมัลแวร์และข้อความแสดงความคิดเห็นต่าง ๆ ที่ครอบคลุม เข้าใจได้ง่ายและมีรายละเอียดครอบถ้วน ซึ่งคะแนนที่ได้จากการโหวตนี้จะถูกบันทึกลงในฐานข้อมูลเพื่อนำไปประมวลผลในการแสดงข้อมูลในหน้าผลลัพธ์แบบ Voted view ต่อไป โดยตัวอย่างการโหวตนั้นแสดงในรูปที่ 4.16



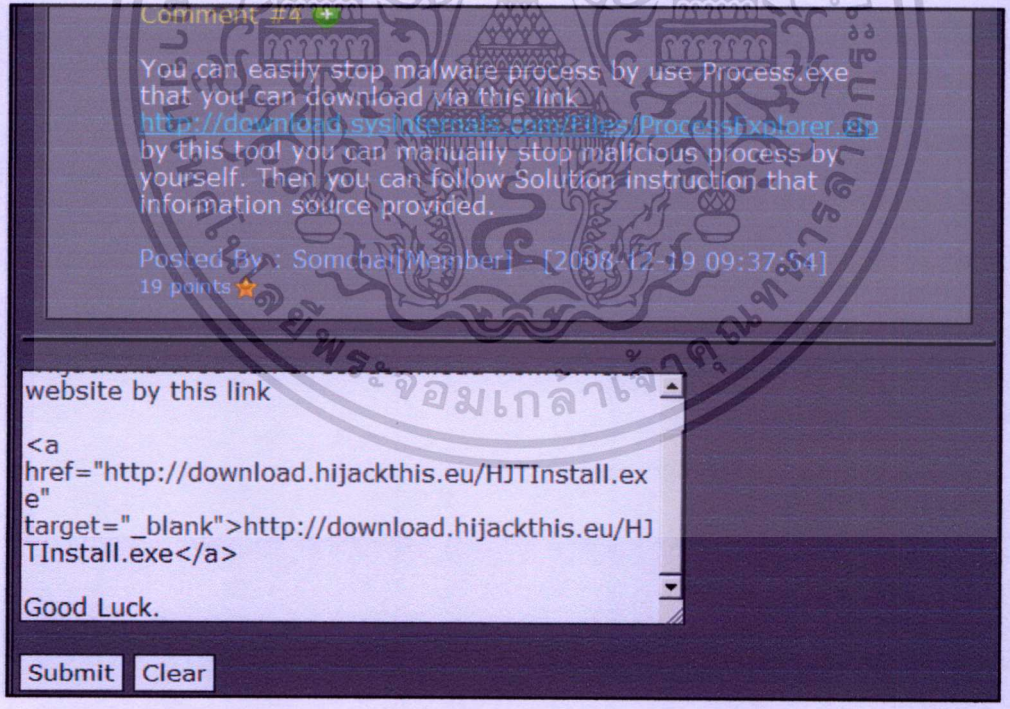
รูปที่ 4.16 ผลลัพธ์จากการโหวตคะแนนให้กับคำอธิบายมัลแวร์



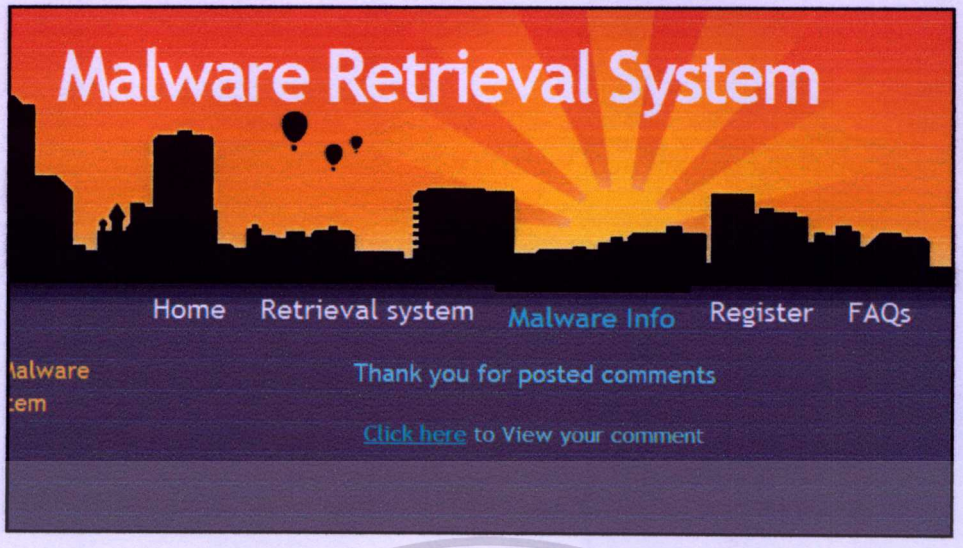
รูปที่ 4.17 ผลลัพธ์จากการโหวตคะแนนให้กับข้อความแสดงความเห็น

4.5 การทดลองและผลการทดลองระบบแสดงความคิดเห็น (Malware Comment System)

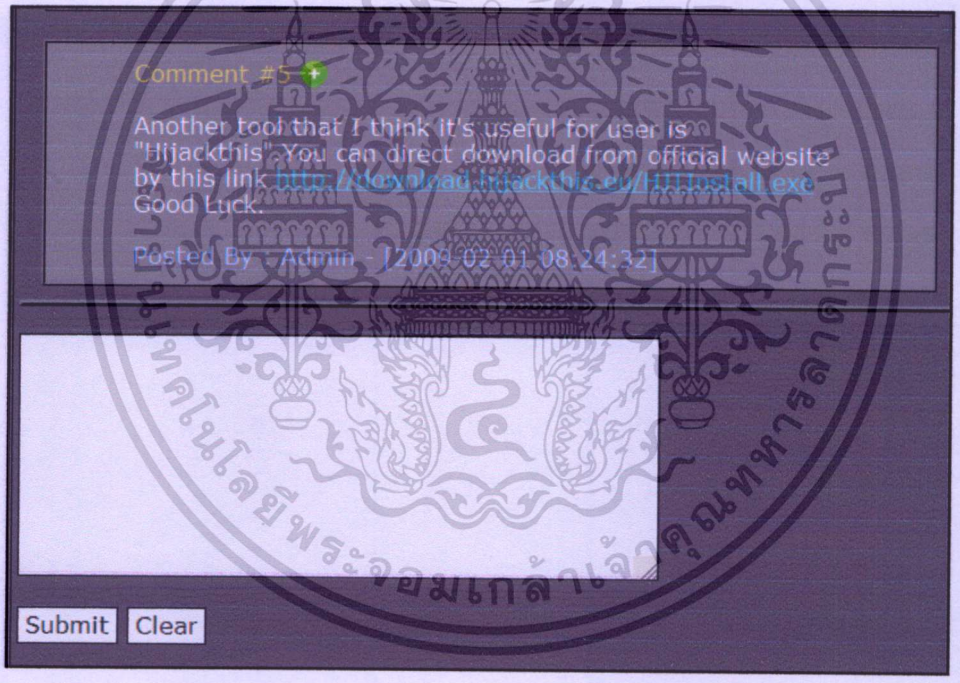
นอกเหนือจากข้อมูลคำอธิบายมัลแวร์ที่มีประโยชน์จากแหล่งข้อมูลต่าง ๆ แล้วนั้น สิ่งที่สำคัญอีกสิ่งหนึ่งก็คือความเห็นหรือคำแนะนำจากผู้ที่มีประสบการณ์เกี่ยวกับมัลแวร์ ซึ่งจะมีส่วนช่วยเหลือผู้ใช้ได้เป็นอย่างมาก เนื่องจากความเห็นเหล่านี้เป็นสิ่งที่ผู้ใช้แต่ละท่านได้ประสบมากับตนเอง จึงสามารถนำมาอธิบายให้ผู้ใช้ท่านอื่นทราบได้อย่างละเอียด ตลอดจนสามารถนำไปใช้งานหรือแก้ไขปัญหาได้จริง ซึ่งในงานวิจัยนี้มองเห็นถึงประโยชน์ของการร่วมกันแสดงความคิดเห็นนี้ จึงอนุญาตให้สมาชิกที่ลงทะเบียนกับทางระบบนั้นสามารถแสดงความเห็นหรือข้อชี้แนะเกี่ยวกับมัลแวร์ตัวดังกล่าวนั้นได้ ไม่ว่าจะเป็นการอธิบายรายละเอียดบางส่วนเพิ่มเติม หรือเป็นการแนะนำวิธีการแก้ไขปัญหาที่เกิดจากมัลแวร์ตัวดังกล่าว หรือการกำจัดมัลแวร์ออกจากเครื่องของผู้ใช้ เป็นต้น ในการโพสต์ข้อความแสดงความเห็นนั้น ผู้ใช้สามารถพิมพ์ข้อความที่ต้องการจะโพสต์ลงใช้ช่องพื้นที่ว่างในรูปที่ 4.18 จากนั้นกดปุ่ม Submit เพิ่มโพสต์ข้อความลงในระบบ ต่อจากนั้นระบบจะนำข้อความดังกล่าวบันทึกลงในฐานข้อมูลและแสดงข้อความตอบรับการโพสของผู้ใช้ดังแสดงในรูปที่ 4.19 ความคิดเห็นใหม่ที่ผู้ใช้ได้โพสจะแสดงในรูปที่ 4.20



รูปที่ 4.18 ตัวอย่างการโพสต์ข้อความแสดงความคิดเห็น



รูปที่ 4.19 เมื่อผู้ใช้ได้ทำการโพสต์ข้อความแสดงความคิดเห็นแล้ว



รูปที่ 4.20 ความคิดเห็นที่ผู้ใช้โพสต์เข้ามาใหม่

บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอระบบฐานความรู้มัลแวร์ซึ่งเป็นการบูรณาการคำอธิบายมัลแวร์จากแหล่งข้อมูลที่หลากหลายมารวบรวมไว้ภายใต้มาตรฐานเดียวกัน โดยใช้ตารางเปรียบเทียบคำศัพท์เทคนิคและเมตาเดต้าในการแก้ไขปัญหาเรื่องความหลากหลายทางด้านโครงสร้าง (Schematic heterogeneity) และยังนำเสนอวิธีการแก้ไขปัญหาที่เกิดจากความหลากหลายด้านความหมาย โดยการกำหนดหมายเลขประจำตัวร่วมมัลแวร์ (Malware Common ID) ซึ่งมีส่วนช่วยในการรวบรวมมัลแวร์ที่เป็นตัวเดียวแต่มีเรียกแตกต่างกันไปตามแหล่งข้อมูลให้อยู่ภายใต้ฐานความรู้มัลแวร์เดียวกันและมีเลขประจำตัวร่วมมัลแวร์หมายเลขเดียวกันรวมถึงการจำแนกประเภทมัลแวร์และการประเมินระดับภัยคุกคามที่เกิดจากมัลแวร์ซึ่งสามารถบันทึกข้อมูลประเภทของมัลแวร์และระดับภัยคุกคามที่เกิดจากเกิดจากมัลแวร์ที่มาจากแหล่งข้อมูลที่แตกต่างกันลงในฐานความรู้มัลแวร์ภายใต้มาตรฐานเดียวกันได้ พร้อมทั้งเปิดโอกาสให้ผู้ที่มีความเชี่ยวชาญด้านมัลแวร์สามารถแบ่งปันความรู้หรือประสบการณ์ในการแก้ปัญหาที่เกิดขึ้นจากมัลแวร์กับผู้ใช้ทั่วไปผ่านทางเวปไซต์ที่เราได้ออกแบบไว้ และยังอนุญาตให้ผู้ใช้ที่ได้ลงทะเบียนเป็นสมาชิกกับทางเวปไซต์สามารถลงคะแนนโหวตให้กับคำอธิบายมัลแวร์ที่ตนเองอ่านแล้วเห็นว่ามีประโยชน์หรือมีเนื้อหาค่อนข้างครอบคลุมและละเอียดได้ด้วยตนเอง นอกจากนี้ระบบฐานความรู้มัลแวร์ยังมีเครื่องมือสำหรับช่วยรวบรวมคำอธิบายมัลแวร์จากเวปไซต์ต่าง ๆ มาเก็บลงในฐานข้อมูลอีกด้วย

บรรณานุกรม

[1] Saudi Madihah and Jomhari Nazean **"Knowledge Structure on Virus for User Education",** Faculty of Science & Technology, Islamic University College of Malaysia, 2006.

[2] M. Helenius, **"A system to support the analysis of antivirus products' virus detection capabilities"**, PhD Dissertation, Department of Computer and Information Sciences, University of Tampere, 2002.

[3] สัญญา คล่องในวัย, "**ก้าวทันมัลแวร์กับการใช้อุปกรณ์ IT อย่างปลอดภัย**" [Online]. Available: http://pclab.nectec.or.th/Documents/Support/Article/Malware.pdf

[4] Kaspersky Lab Industry, "Classic Viruses." [Online]. Available: http://www.viruslist.com/en/virusesdescribed?chapter=152540474

[5] J. Aycock, "**Computer Viruses and Malware**", University of Calgary, Springer Science Business Media, Canada 2006.

[6] Symantec. "**What is the difference between viruses, worms, and Trojans?**". [Online]. Avaliable:http://service1.symantec.com/SUPPORT/nav.nsf/pfdocs/1999041209131106?Open&docid=1999041209131106&nsf=nav.nsf&view=docid

[7] Kaspersky Lab Industry, "**Macro Viruses.**", [Online]. Available: http://www.viruslist.com/en/virusesdescribed?chapter=152540474#macro

[8] DOE-CIRC. "**I-023: Macro Virus Update ((WM.CAP, XM.Laroux, WM.Concept, WM.Wazzu, WM.NPAD))**". [Online]. Available: http://ciac.llnl.gov/ciac/bulletins/i-023.shtml

[9] P. Szor, "**The art of computer virus research and defense**", Addison Wesley Professional, 2005.

[10] S. Klongnaivai and Thanawit Chewaprapanon, "**Polymorphic and Metamorphic Viruses: The formidable adversaries**" [Online]. Available http://www.thaicert.nectec.or.th/paper/virus/polymorphic.php

[11] T. Yetiser, "**Polymorphic Viruses: Implementation, Detection and Protection**", VDS Advanced Research Group, Jan, 1993. [Online]. Available http://vx.netlux.org/lib/ayt01.html

[12] M. Karresand, "**Separating Trojan Horses, Viruses, and Worms – A proposed taxonomy of software weapons**". In Proceedings of the 2003 IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY June 2003.

[13] ThaiCERT. **"W32.Mydoom.BB@MM."** [Online], Available: http://thaicert.nectec.or.th/advisory/alert/mydoom_bb.php#reference

[14] ThaiCERT. **"W32.MSN.Worm."** [Online], Available: http://thaicert.nectec.or.th/advisory/alert/msnworm.php

[15] ThaiCERT. **"W32.IRCBot.AJY."** [Online], Avaliable: http://thaicert.nectec.or.th/advisory/alert/ircbot.php

[16] N. Weaver, V Paxson, S Staniford, R Cunningham. **"A taxonomy of computer worms"**. In Proceedings of the 2003 ACM workshop on Rapid malcode, 2003.

[17] ThaiCERT. **"VBS.Godzilla."** [Online], Available: http://thaicert.nectec.or.th/paper/virus/godzilla.pdf

[18] ThaiCERT. **"VBS.Solow."** [Online], Available: http://thaicert.nectec.or.th/paper/virus/rundll64.pdf

[19] Kaspersky LabIndustry, **"Trojan programs"** [Online]. Available: http://www.viruslist.com/en/virusesdescribed?chapter=152540521

[20] ดวงกลม ทรัพย์พิทยากร, **"สปายแวร์และวิธีป้องกัน"** [Online], Available: http://www.thaicert.org/paper/spyware/IntroToSpyware.pdf

[21] G. M. Nijssen and Terry Halpin, **"Conceptual Schema and Relational Database Design"**, (Prentice Hall, Sydney:1989).

[22] Terry Halpin Microsoft Corporation, "**Object-Role Modeling: ORM/NIAM**" [Online], Available: http://www.orm.net/pdf/springer.pdf

[23] Wikipedia **"Information integration"** [Online], Available: http://en.wikipedia.org/wiki/Information_integration

[24] งามนิจ อาจอินทร์, ศรัญญา กัลย์จาฤก **"ระบบบูรณาการสารสนเทศเพื่อแก้ปัญหาความหลากหลายของข้อมูล"** ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

[25] Ngamnij Arch-int, Peraphon Sophatsathit, **"A Reference Architecture for Integrating Heterogeneous Information Sources Using XML and Agent model"** Advanced Virtual and Intelligent Computing (AVIC) Research Center.

[26] S Busse, RD Kutsche, U Leser and H Weber **"Fedearted Information Systems: Concepts, Terminology and Architectures"** Technische Universität Berlin, 1999.

[27] The Dublin Core Metadata Initiative. [Online], Available: http://dublincore.org/

[28] MARC Standard. [Online], Available: http://www.loc.gov/marc/

[29] ID3. [Online], Avaialbe: http://www.id3.org/

[30] C. Batini, M. Lenzerini, S.B. Navathe, **"A Comparative Analysis of Methodologies fo Database Schema Integration"**, ACM Computing Surveys, Vol. 18, No. 4, pp. 323-364, Dec1986.

[31] I. Schmitt, **"Schemaintegration für den Entwurf föderierter Datenbanken"**, Inx Verlag, Sankt, Augustin, 1998.

[32] W. Litwin, L. Mark, N. Roussoupoulos, **"Interoperability of Multiple Autonomous Databases"**, ACM Computing Surveys, Vol. 22, No. 3, pp. 267-293, Sep. 1990.

[33] U. Leser, **"Maintenance and Mediation in Federated Databases"**, 8th Workshop on Information Technology and Systems, Helsinki, Finland, TR-19, University of Jyvaeskylae pp.187-196, 1998.

[34] NOD32 Antivirus Software. [Online], Available: http://www.nod32th.com/

[35] Trend Micro Incorporated. [Online], Available: http://us.trendmicro.com/us/home/

[36] F-Secure Corporation. [Online], Available: http://www.f-secure.com

[37] Sophos Plc. [Online], Available: http://www.sophos.com

[38] Symantec Corporation. [Online], Available: http://www.symantec.com/index.jsp

[39] McAfee Corporate. [Online], Available: http://www.mcafee.com

[40] Kaspersky Lab Industry. [Online] Available: http://www.viruslist.com

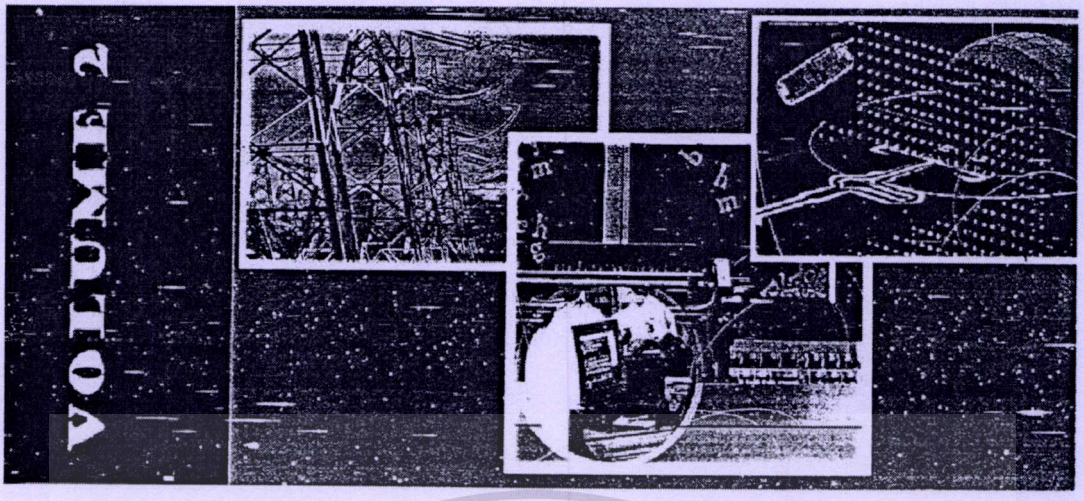
[41] The WildList Organization International. [Online] Available: http://www.wildlist.org/

ภาคผนวก

ภาคผนวก ก.

ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่

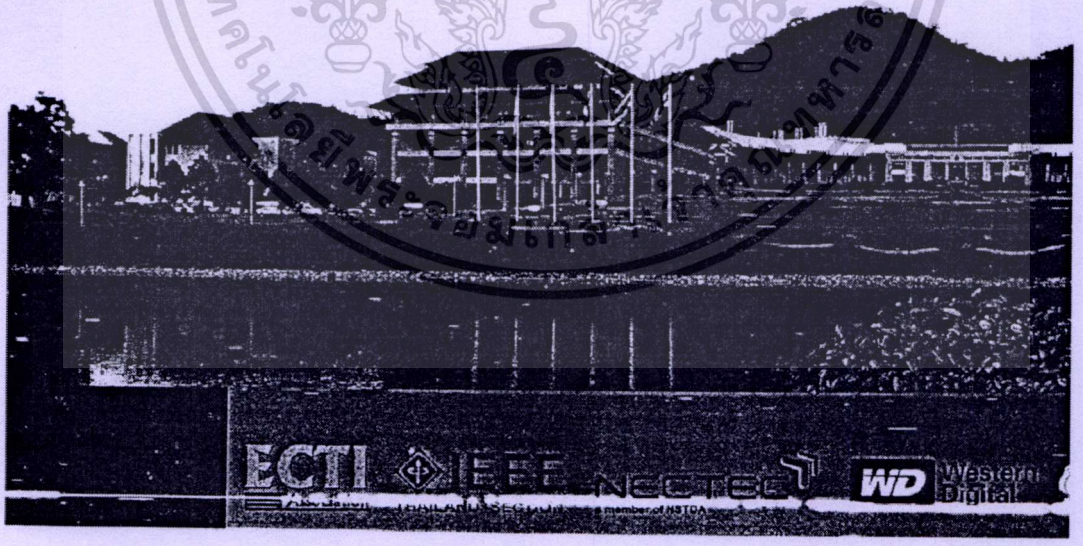
1. Chittanond Pheunphiphop and Mayuree Lertwatechakul, ***"Computer Viruses Metadata"*** ECTI-CON 2007, pp. 1252-1255, Mae Fah Luang University,Chiang Rai ,Thailand, May 9-12, 2007.
2. Chittanond Pheunphiphop and Mayuree Lertwatechakul, ***"Unified Malware Descriptions Platform and a Retrieval Tool"*** JCSSE 2009, Phuket, Thailand, May 13-15, 2009.



ECTI-CON 2007

Mae Fah Luang University, Chiang Rai, Thailand
May 9-12, 2007

VOLUME 2
- Communication Systems
- Signal Processing
- Computer and Information



Computer Viruses Metadata

Chittanond Pheunphiphop, Mayuree Lertwatechakul
Department of Information Engineering
Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang
Bangkok, Thailand 10520
Email: s47061137@yahoo.com, klmayure@kmitl.ac.th

***Abstract*-As to enable a variety of the Internet users to easily achieve and share virus description, a convenient platform and well defined general terms are required. In this paper, various virus techniques, behaviors and damages have been analyzed to design the virus metaschema. We defined a set of general terms to represent virus description words for several conflict terms. This could help people to easily understand and make it comfortable to exchange to the public.**

I. INTRODUCTION

At the moment, the outbreak of virus has spread increasingly than ever while virus information source or description of each virus limited only in the websites of antivirus companies or organizations. Description of each virus has been provided as individual without distinct data exchange or sharing with other company or organization.

In this paper, analytical study has been done on technique and behavior of the viruses in order to design the structure of database by using NIAM model which is easy to be applied by illustrated presentation. This metadata presentation is provided with simple description and definition for easy access and meant to publicize this knowledge to anyone interests in virus, for better understanding and gain more viruses data exchange among people, communities & organization.

II. THE STUDY OF VIRUSES

Computer virus [1] is a type of computer program designed to rapidly replicate itself into targeted computer's files without user notice or permission. But virus unable to replicate itself across the other computer except that the user manages himself as a carrier, for instance, virus can be distributed only when infected user attaches file or document via email. Exchanging diskette, CD-ROM and Flash drive with virus infected are the other ways that allow virus to spread into other computers. That's it, the virus keep replicate itself over and over again.

Computer virus can be divided into two categories. Based on infected environment or infected methods [2]. In this paper, we have designed the concept virus metadata which is mainly concern on infected environment type.

Most viruses can be found in one of the following environments:

- **File**
- **Boot sector**
- **Macro command**
- **Script host**

A. Virus behavior

Among the various types of virus we confront with, there are always common traits which can be divided into three phases as following:

Installation phase: The phase that virus writer intends to install virus code into user's computer.

Infection phase: The phase after virus installation, virus will spread itself into targeted file as it designed by writer.

Damaged phase: This phase is meant to cause damage to user directly from virus or its side effect. The damage also includes the effect that occurred during the infection phase.

B. Virus technique

We can categorize virus technique as following [3, 4, 5]:

- **Stealth:** Stealth virus can conceal their presence from antivirus program. It can also inhibit in computer without user's awareness.
- **Polymorphic or Encrypted:** This type of virus got many appearances in order to hide their signature by scrambling virus body to avoid a detection of antivirus programs. Moreover its encrypted virus body always be changed.
- **Metamorphic:** This technique is similar to polymorphic but it has additional part which calls "Mutation Engine" which can change both of decryption routine and encryption virus body.
- **Memory Resident:** Viruses use this technique to load itself in to computer memory and to infect the next loaded program file.
- **Multipartite Viruses:** This virus infects both in file and boot sector. It can also spread itself across infected program file to boot sector.
- **Overwriting Viruses:** This virus type replaces itself to a target file.
- **Companion Viruses:** This virus is similar to overwriting virus, moreover it rename the original file to the new one. For instance PROGRAM.EXE to PROGRAM.EXD, when computer run PROGRAM.EXE virus code will activate first and execute original afterwards.
- **Prepending Viruses:** These viruses write itself to the beginning of the file.

- **Appending Viruses**: These viruses write itself to the end of the file.

C. Damages of viruses

In this paper, we define damage as an action caused by virus means harm that occurs to user no matter what it came from virus writer's intention or side effect. Damages occurred to user can categorized as following:

- **Interface damaged:** Annoying user's interface such as display monitor, mouse, speaker etc. For example, virus distorts displayed font on user monitor.
- **Data damaged:** Damage to user's documents or files by erasing or modifying user's data, for example, some virus searched text in user's document and replaced it with the virus text.
- **System file damaged:** Damage to operational system, boot sector and CMOS, for example, when virus erases or modifies system file of the operational system or boot sector of hard disk or CMOS. So that user could not run computer as usual.
- **Accessible damaged:** Virus may changes or create password access mode that's used to access a file or document. So user cannot open document or file.
- **Program damaged:** Damage to programs, virus would erase or hide significant menu, sometimes virus make program unable to run normally.
- **Security damaged:** Damage to security system, virus would erase antivirus program files or modify security menu to allow virus intrusion.
- **Overall data damaged:** Damage to all storage data, virus would erase or destroy all data in the computer, for example, virus may using command such as Format C:/, DEL *.* etc.
- **Performance damaged:** Damage which is caused by existence of virus in computer memory which mean computer has wasted its resource. This category includes computer halt, shutdown or restart itself unexpected.

III. NIAM CONCEPTUAL SCHEMA MODEL

After we have analyzed the detail of viruses' behaviors, viruses' techniques and viruses' damages from various information sources. We could design a computer virus conceptual schema as shown in Figure 1. The schema was designed on purpose that virus description has to cover all important data that useful for user and represent it in the way that is easy to understand.

According to our design, virus description has been grouped into two groups, first group is mandatory data to user such as virus name, OS, technique of virus, infected environment of virus, virus's routine, damage of virus and virus infected file type. The second group consists of supplementary data such as alias name of virus, in the wild, risk rate, reporter's name, reporting date and virus's message.

NIAM conceptual schema in Figure 1. could be mapped by optimal algorithm to relational schema as shown in Figure 2.

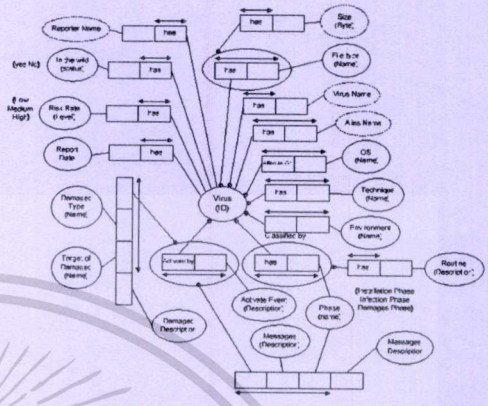


Figure 1. NIAM conceptual schema model of computer virus

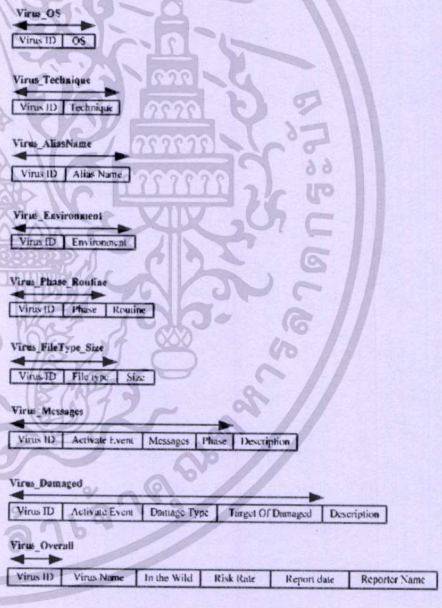


Figure 2. Corresponding relational computer virus database schemas mapped from Figure 1.

TABLE 1

Computer virus metadata

Entity \ Label name	Description	Example
Virus (ID)	Identity of virus	ID0001
Virus Name	Name of virus	Jerusalem
Alias Name	Virus alias name	Israel
OS (Name)	Operation System	Windows XP
Technique (Name)	Virus technique	Stealth
Environment (Name)	Type of virus	File virus
Phase (Name)	Phase of virus procedure	Installation
Activate Event (Description)	Event that activate virus.	Friday 13th, July 18th.
Messages (Description)	Messages that virus display to user.	"!Hola yo soy Ruperto y……"
Damaged Type (Name)	Virus damaged type.	Interface Damaged
Target of Damaged (Name)	Target of virus damage.	Monitor, Keyboards
Damaged description	Description of virus damaged.	"The systems slow down…"
Messages Description	Messages description.	"The messa…"
Report date	Date of reporting	13/10/2006
Risk Rate (Level)	Risk rate of virus.	Low.
In the wild (Status)	Show that virus is in the wild or not.	Yes, No.
Routine (Description)	Virus routine.	"Virus infec…"
File Type (Name)	Target file that virus intend to infect.	.COM .EXE
Size (Byte)	Size of virus.	1,080 bytes.
Reporter name (Name)	Name of description reporter	"MR. Smith", "MR. Neo".

Table 1 is virus metadata that describes entities name, descriptions and examples of corresponding label of NIAM schema computer viruses.

IV. EXPERIMENT

As to prove the consistency of the designed schema, we tried to populate existing data of viruses from various information sources as following:

- F-Secure Corporation [6]
- Sophos Plc [7]
- Trend Micro Incorporated [8]
- Symantec Corporation [9]

Because of limitation in this paper, we could show only some parts of some tables that are populated by existing data of Jerusalem virus. Table 2-9 is created in order to show that each virus information sources is concerning on different set of virus description attributes.

TABLE 2

Example of data in Virus_Overall table from F-Secure Corporation

Virus_ID	Virus_Name	Risk_Rate	In_the_wild	
ID0001	Jerusalem	-	-	

("-" = Not mention to.)

TABLE 3

Example of data in Virus_Overall table from Sophos Plc

Virus_ID	Virus_Name	Risk_Rate	In_the_wild	
ID0001	Jerusalem	-	-	

("-" = Not mention to.)

TABLE 4

Example of data in Virus_Overall table from Trend Micro Incorporated

Virus_ID	Virus_Name	Risk_Rate	In_the_wild	
ID0001	Jerusalem	Low	Yes	

TABLE 5

Example of data in Virus_Overall table from Symantec Corporation

Virus_ID	Virus_Name	Risk_Rate	In_the_wild	
ID0001	Jerusalem.1808	Low	Yes	

(Jerusalem.1808 in Symantec Corporation is similar to Jerusalem on other virus information source.)

TABLE 6

Example of data in Virus_Phase_Routine table from F-Secure Corporation

Virus_ID	Phase	Routine
ID0001	Installation	-
ID0001	Infection	Virus infects .exe over and over.
ID0001	Damaged	When virus is activates on Friday 13 th, it deleting programs that run on that day. Thirty minutes after an infected program is run virus will slows the computer down and make a part of the screen scroll up two lines. Virus continues infects the same file over and over in order to wastes computer resource.

("-" = Not mention to.)

TABLE 7

Example of data in Virus_Phase_Routine table from Sophos Plc

Virus_ID	Phase	Routine
ID0001	Installation	-
ID0001	Infection	Virus infected exe file over and over again and increased they size. But .com files it infected only once
ID0001	Damaged	After infected thirty minutes, virus scrolled screen up two lines, then it slows the system down. On Friday 13th it deletes every program run.

("-" = Not mention to.)

TABLE 8

Example of data in Virus_Phase_Routine table from Trend Micro

Virus_ID	Phase	Routine
ID0001	Installation	When user run infected file virus load itself in memory
ID0001	Infection	Virus stay in memory then infect program files as they are accessed
ID0001	Damaged	Virus increased infected files in order to wastes computer resource. Thirty minutes after an infected program is run virus will slows the computer down and will delete any program started on Friday the 13th of any year.

TABLE 9

Example of data in Virus_Phase_Routine table from Symantec Corporation

Virus_ID	Phase	Routine
ID0001	Installation	-
ID0001	Infection	-
ID0001	Damaged	Virus active every Friday 13th, it deletes any running program. Thirty minutes after first deletion, the computer slows down and the screen scrolls up two lines.

("-" = Not mention to.)

Each information sources got their technical terms that might be incommon and this may difficult for users to understand. As to solve this problem, we have to choose the most popular terms or defined more meaningful names to be used as general virus technical terms. The well defined general terms will be stored in Technical to general terms mapping table as shown Table 10.

TABLE 10

Technical to general terms mapping table

Entity\Company	F	S	T	Sy
OS	Platform	Affected Operating Systems	Platform	Systems Affected
Damaged	-	-	-	Damage
File Type	-	-	-	Target Of Infection
Alias Name	Alias	Aliases	Aliases	Also Known As
Reporter name	-	-	-	Writeup By
In the wild	-	-	In the Wild	Wild
Risk rate	-	-	Overall Risk Rating	Risk Level

F= F-Secure Corporation, S= Sophos Plc, T= Trend Micro Incorporated, Sy= Symantec Corporation

V. RESULT

After we have populated existing data of Jerusalem virus from various information sources into virus metadata tables. We found that each virus information source describes viruses in arbitrary way, some provides sufficient details and some doesn't. Table 11 is created to show what kind of mandatory virus description data that we found in each virus information source.

Table 11 shows that Trend Micro Incorporated reveals virus description covers almost all mandatory data except for OS information. Whereas Symantec Corporation is less mention on mandatory data which is required on each virus description such as virus's technique, infected environment of virus and routine process.

TABLE 11

Mandatory virus description table

Entity\Company	F	S	T	Sy	A
Virus Name	Yes	Yes	Yes	Yes	Yes
OS	No	No	No	Yes	Yes
Technique	Yes	No	Yes	No	Yes
Environment	No	No	Yes	No	Yes
Damaged description	Yes	Yes	Yes	Yes	Yes
Routine*	No	No	Yes	No	Yes
File type	Yes	Yes	Yes	Yes	Yes

F= F-Secure Corporation, S= Sophos Plc, T= Trend Micro Incorporated, Sy= Symantec Corporation, A= Our metadata, Routine* = Description has to covers all phase of the routine process.

VI. CONCLUSION

In this paper, we present metadata for computer viruses that and the well defined general virus technical terms for easily understand. The content of technical to general terms mapping table is intended to solve the conflict terms problems of various information sources. Furthermore this metadata can be used as exchangeable format among interested people or viruses' community.

REFERENCES

[1] Sanya Klongnaivai, "Malware on IT equipment", http://pclab.nectec.or.th/Documents/Support/Article/Malware.pdf.

[2] Kaspersky LabIndustry, "Classic Viruses.", http://www.viruslist.com/en/virusesdescribed?chapter=152540474

[3] Subramanya, S.R., Lakshminarasimhan, N., "Computer viruses", in *Potentials IEEE*, Volume 20, Issue 4, Oct-Nov 2001, pp. 16 – 19.

[4] Sanya Klongnaivai and Thanawit Chewaprapanon, "Polymorphic and Metamorphic Viruses: The formidable adversaries", http://www.thaicert.nectec.or.th/paper/virus/polymorphic.php.

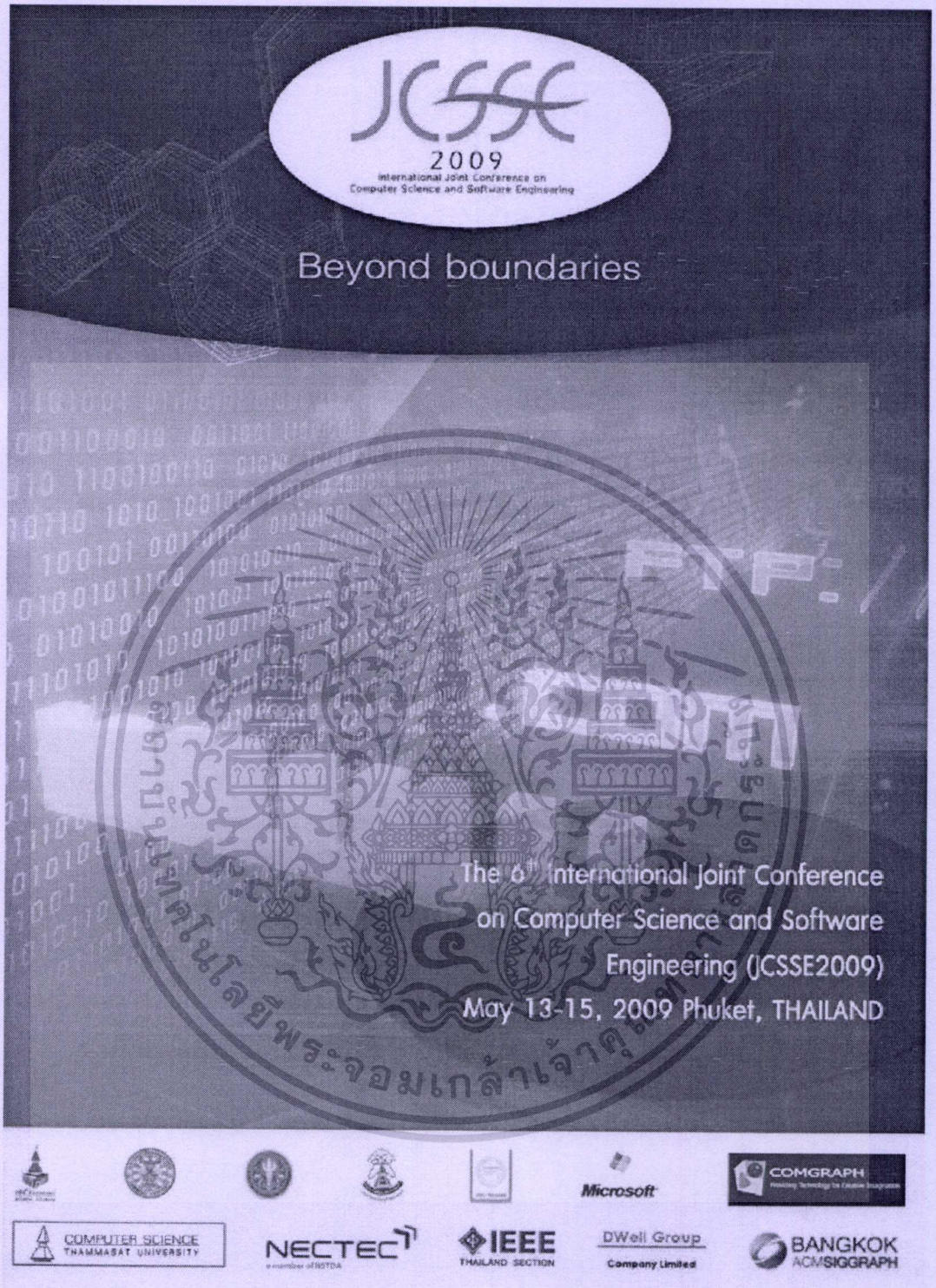
[5] Peter Szor, "The art of computer virus research and defense", Addison Wesley Professional, 2005.

[6] F-Secure Corporation, http://www.f-secure.com.

[7] Sophos Plc., http://www.sophos.com.

[8] Trend Micro Incorporated, http://www.trendmicro.com.

[9] Symantec Corporation, http://www.symantec.com.



JCSSE2009, May 13-15, 2009 (Vol. 1)

Unified Malware Descriptions Platform and a Retrieval Tool

Chittanond Pheunphiphop and Mayuree Lertwatechakul
Department of Information Engineering
Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang.
Bangkok, Thailand
Email: S47061137@yahoo.com, klmayure@kmitl.ac.th

Abstract

Nowadays, malware problem becomes major issue in the Internet. As to provide knowledge about malware to people many websites or anti-virus vendors provide malware description through their websites. But each of them has their individual technical terms which may make the information become non-integrity to people's sense. In order to solve the problem, we have designed malware metadata which covers almost all malware descriptions from various popular sources. We also provide tool to retrieve, analyze and store malware descriptions as to present it publicly in a unified pattern.

Key Words: Malware, Classification, Unified Descriptions Platform

1. Introduction

In order to protect people from malwares, users should to have knowledge about malwares as to understand how they propagate and infect into their computers. Generally, malwares are classified from their properties into viruses, worms and Trojan horses. Presently, malware descriptions can be easily found on the Internet which is provided by many information sources. Since malware information provided by various information sources can cause user to be confused about their technical terms which are defined arbitrarily by each sources.

Usually, when a new malware is discovered by anti-malware vendors, they would give it a name which may not be synchronized to the names given by another vendors. So that it could be difficult to identify which malwares are identical.

As to solve this problem, we have been studied and analyzed malwares' properties and its behaviors that cause affecting damages. We have designed a malware metadata in order to provide users a unified information which covers all important entities that required for malware descriptions from various sources. The malware metadata contains extra information that present a malware common name for each malware to solve the malware naming conflict. Besides, we also developed a tool for malware interesting group to retrieve malware description from websites and collect them into the malware database.

2. The study of malware

Malware [1] is a name abbreviated from malicious software which means software with malicious intentions. We can categorize malwares by their property characteristics into three types which consists of viruses, worms, and Trojan horses as shown in details below.

Computer viruses [2] are computer programs designed to inhabit in executable files. They could not being stand alone. They are automatically replicate themselves to any target files when user runs the infected files.

Anyway, computer viruses could not propagate themselves across the computers except for when a user act himself as a carrier by exchanging media such as CD-ROM, email or removable data storage which contain viruses.

Computer viruses can be divided by type of hosted file, boot sector and macro that they inhabit in as the following:

- File type: Computer viruses that infect in executable files such as .EXE, .COM and etc. Addition with script files type such as .BAT, .VBS and etc.
- Boot sector type: Some viruses do not infect in executable files, they infect only within a specific place at the boot sector or boot record on disks.
- Macro type: Mean viruses that are attached in Macro command persisted in user's document such as Macro virus that we have found in MS Office document.

Worms [3] are quite similar to computer viruses with their self-replicate property. They can automatically propagate themselves to other computers through removable storage devices and network. The different between worms and viruses is that worms exist as a stand-alone file, whereas viruses need some host files to inhabit in. Worms can be categorized by their propagation characteristics as the following:

- Email type: This kind of worms copies and propagates itself to the other computers through email by using address book data on victims' computers. This is a concealed process which may be happened without user's notice.
- Instant messaging type: Instant messaging worm propagates itself through online communication channel by using conversation programs such as Internet Chat Relay (IRC), MSN messenger and Yahoo messenger. This type of worm copies and sends itself to other user who is having online conversation with victim's user at that moment, whereas some worms would send an URL which led users to the specified malicious website.
- Network type: Network worm propagates to other computer within the local network or the Internet through the existing communication protocols, such as TCP, UDP, P2P and etc. In some cases, it could place its copies into frequently used network resources and to be waiting for another user to download.
- Removable storage device type: This worm is frequently found on user's storage devices such as removable hard disk, USB flash drive, memory stick and etc. They would modify files on users' devices to automatically execute the worm file whenever users plug his/her device into computer.

Trcjan horses [4] are programs that designed for the specific malicious functions. Generally, Trcjans' behaviors are different from viruses and worms because they could not infect to the other host files or self-propagation to other computers. Trcjan can be stratified by their malevolent functions as following:

- Personal Information Stealing type: Some Trojans were designed to steal users' private information, for example email accounts, bank accounts and their access passwords.
- Remote Access type: Trcjans of this type would remotely access into users' computers in order to do some malicious tasks such as redirect user's browser to a specified website as to raise click-counter on the target websites. Some Trcjans induce victim machines to behave like DoS (Denial of Service) as to attack some target websites.

A. Malware Behaviors

After we have studied malwares, we can divide their behavior into three phases as following:

- Installation phase: Malwares install themselves on user's machines by using several techniques such as hooking on computers' memory, injecting themselves into systems' registry. During this phase, malwares prepare themselves to propagate to other machine or other file.
- Propagation & Infection phase: After malwares installed themselves on users' machines, they would begin to spread themselves from one place to another. Viruses infect executable files whereas worms spread themselves across machines. In contrast, Trcjans do not spread to the other files or machines, excepting for the case that an user accidentally carries them to the other machines.
- Damaged phase: Within this phase, malwares may take some actions that cause damage to users' machines directly likes deleting some files, stealing some information or creating some huge garbage files.

B. Damages caused by Malwares

We can define term 'damages' as the effects that according to malwares' behaviors on the victims' machines, no matter what have they directly done or just their side effects. In this paper, we include inhabitation of malwares as a kind of damage too. Because they can waste machine resources such as disk space, memory space and cause CPU to do more work. Thus, by using this concept damages can be divided into:

- Data damages: The damage that occurred to users' files or documents. Malwares would erase or modify the data or files and this may make the users unable to use or to recover them.
- Private information stealing: Users' private information such as bank accounts, email accounts and their access passwords may be stolen to be used by some criminalities.
- System file damages: Malwares would erase or configure the system files which make the system out of order or do something to support malwares' functions.

- Program damages: This damage would be happened to users' programs, malwares would modify or erase program files which make the programs unable to run.
- Network resource damages: The damages that occurred to network resources especially for the network bandwidth capacity that would be aggressively consumed by DoS or worm.

3. Unified malware description platform

In order to provide a unified malware description platform, we have to design a database that can completely store various malware data from multiple sources. Besides, some special set of tables must be provided to solve the malware naming conflict problems that may confuse the users. Such kind of the information is still useful for containing of the specific malware technical terms which describe malwares behavior, malware property and type of malware. Moreover, we have developed a tool that can be helpful to gather and analyze malware information to be store in the malware database. Finally, the malware data that resides in the database could be easily retrieved and to be represented in a meaningful and unified platform.

A. Malware Metadata

Previously, we have studied on viruses and then designed a computer viruses metadata as presented in [6]. The concept and idea of that work such as technical term mapping table, is also applicable for worms and Trcjans, so that we extended the metadata to cover most types of malware. The new metadata model is composed of malware type, malware detail, first published date, modified date and MCID (Malware Common ID) as shown in Figure 1.

According to design the malware metadata, we can divide the malware information into two groups. First is the necessary data that required on every description such as malware name, malware routines (that describe processes of malware), malware property (which are likely to be either viruses or worms or Trcjans), malware OS (which tell users that what Operation System that could be affected from malware).

The second group is supportive data that helps user to know more about malwares, for example "Distribution Status" [5] (which tell users the malware is still spreading), "damage level" hmeans level of damage that malware can caused to user's machine), discovery date, "risk rate" (means a level of opportunity for user to be infected a malware)

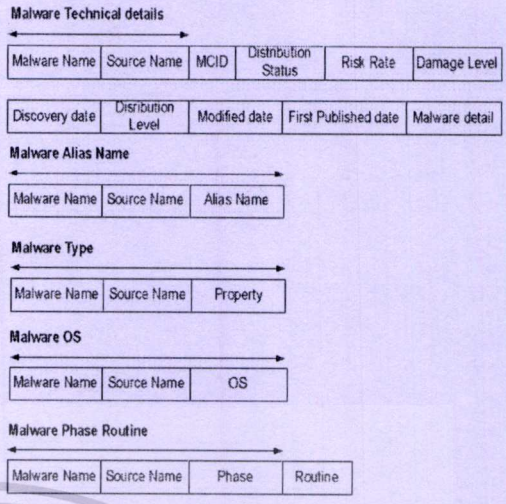


Figure 1. Malware metadata

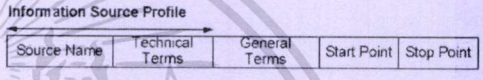


Figure 2. Information source profile metadata

Furthermore, we have designed a malware common grouping table to overcome malware conflict naming problem. A MCID number can cross reference the information of a malware record to the other records of the same malware that may has different names defined by the other malware information sources.

4. Malware retrieval tool

Besides of the malware metadata that we have described above, we also designed Information Source (IS) profile table to store profile metadata of each source. This metadata consist of information source name, technical terms (which may be defined differently by each sources), general terms (mean the most proper and meaningful words selected for describing malware property and behavior as referred in [6]), start point and end point (which are the beginning and the ending points which shows the boundary of the specific contents on each sources that may be given as HTML tags) as shown in Table 1.

Normally, if someone wants to gather information that available in the Internet and store it into a designed database, she/he may download the page, copy and paste it into the database. The process is tedious and hard work. As to ease the process, we

have developed a tool that helps user to retrieve malware content from a specified source through the Internet automatically. Before retrieving malware information from any sources, a user has to do a registration process for each source by giving the retrieval tool about the URLs of the source and marking boundary of the malwares' description contents separately by its property, behavior, damages and so on. The block diagram of the malware retrieval tool engine is shown in Figure 3.

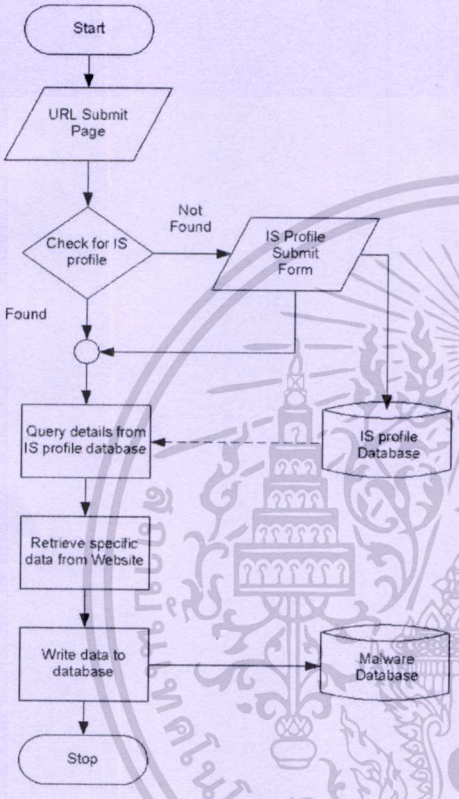


Figure 3. Block diagram of the engine

A. Registration phase

To retrieve malware data from the Internet, user has to give the required information of the new preferred source such as information source name (may be an anti-malware developing company) into IS profile. User also required to mark on a set of html documents that represents a malware description shown on the website. The marked text and html tags indicate boundary of the content and relate them onto corresponding data fields in the malware tables, the example of data that resides in IS profile table is shown in Table 1.

B. Retrieval phase

After a user filled a malware source profile data with the preferred website information, the user could ask the retrieval tool as to download and analyze malware information component from the specified URLs of any malware descriptions of that site. The tool would identify information source from the URL, download the html documents and then analyze the malware information to be filled into the malware database by referring to the corresponding mapping information that stored in the information source profile table through the registration process.

Table 1. Example of data in IS profile

IS_Name	Technical term	General Term	Start point	Stop point
Symantec	Name	Malware Name	<meta http-equiv="Conten. ..	<meta http-equiv= "descr
Symantec	Also known as	Alias Name	Also Known As: ...	Type: Viru ...

5. Experiment

As to prove the proposed unified malware metadata structure, we have retrieved malware contents from various information sources and stored it into our designed database. Because of the limitation of paper length, we could show only malware descriptions from four well known sources Kaspersky Lab [7], McAfee, Inc [8], Symantec Corporation [9].

Results of this experiment were done by using the malware description retrieval tool. We just filled in the forms the required necessary information of each source only one time. The information is IS name, technical terms, general terms, start point, and stop point of each part of content. This process was done by using the user friendly interface thought webpages that ease user to complete the form.

6. Result

After we populated data to our designed database, we found that the malware common grouping table can help users to cross reference malware descriptions among various information sources as

presented in Table 2. This is because of each source may describe malware content in an arbitrary way. Since some sources describe malware with information that cover all necessary data but some sources may not. So there were some null values in the malware metadata table

Lack of some information may not help users to understand the interesting malware. To solve this problem, we may need some volunteers to gather malware description and to vote for the most clarify description part from every sources. The most clarify content will be used to represent instead of the null value from the users' specified source as an output example shown in Figure 4.

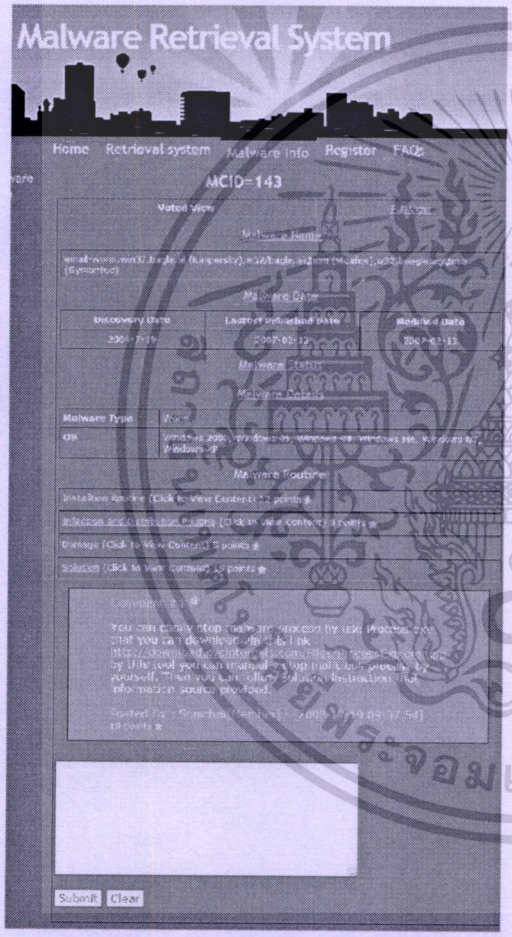


Figure 4. Interface of Malware Retrieval System

Table 2. Malware Common Grouping Table

M_Name	IS_Name	MCID	
Email-worm.win32.bagle.al	Kaspersky	143	
W32/bagle.ai@mm	McAfee	143	
W32.beagle.ag@mm	Symantec	143	

7. Conclusion

We present retrieval and analysis tool that assists malware interesting group in order to automatically retrieve malware description from sources' URLs which available on the Internet. In this paper, we have solved malware conflict naming problem by designing of the malware common grouping table. Information of the table allows user to cross reference malwares' description of many sources. Furthermore, we have provided facilities for malware experts to share their knowledge to public. The experts may use malware description retrieval tool to retrieve, analyze and store malware description from various websites as to present it publicly in a unified pattern.

8. Referrences

[1] John Aycock, "Computer Viruses and Malware," Springer, 2006.

[2] Kaspersky Lab, Industry, "Classic Viruses. ," http://www.viruslist.com/en/virusesdescribed?chapter=152540474.

[3] N Weaver, V Paxson, S Staniford, R Cunningham, "A taxonomy of computer worms," in Proceedings of the 2003 ACM workshop on Rapid Malcode, October 27, 2003.

[4] Kaspersky Lab, Industry, "Trojan Programs," http://www.viruslist.com/en/virusesdescribed?chapter=152540521.

[5] The WildList Organization International, http://www.wildlist.org.

[6] Chittanond Pheunphiphop, "Computer Viruses Metadata," in ECTI-CON 2007, Volume 2, 2007, pp.1252-1255.

[7] Kaspersky Lab, Industry http://www.kaspersky.com.

[8] McAfee, Inc., http://www.mcafee.com.

[9] Symantec Corporation, http://www.symantec.com.