บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

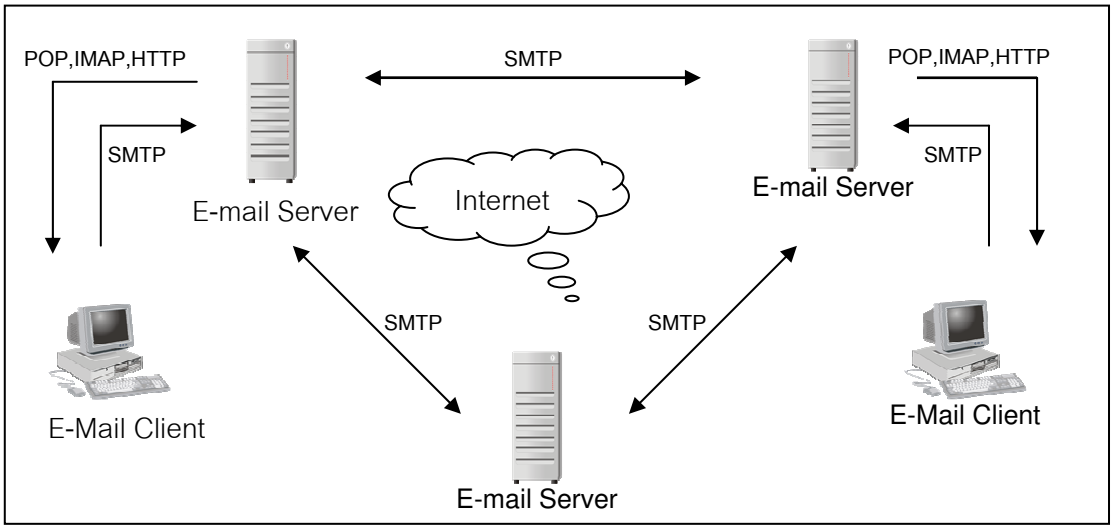
บทนี้จะกล่าวถึง ทฤษฎีและงานวิจัยที่เกี่ยวข้องในวิทยานิพนธ์ฉบับนี้ ทฤษฎีที่นำมาใช้ในการทำวิจัย ประกอบไปด้วย ทฤษฎีการทำงานของระบบจดหมายอิเล็กทรอนิกส์ ทฤษฎีการทำงานของบอตเน็ต งานวิจัยที่เกี่ยวข้องกับจดหมายอิเล็กทรอนิกส์สแปมและบอตเน็ต และกรอบแนวคิดการวิจัย

2.1 ทฤษฎีการทำงานของระบบจดหมายอิเล็กทรอนิกส์

ระบบการทำงานของจดหมายอิเล็กทรอนิกส์ (Electronic Mail) มีรูปแบบการสื่อสารดังนี้

2.1.1 ระบบการทำงานของจดหมายอิเล็กทรอนิกส์

ภาพที่ 2.1 แสดงการรับ-ส่งจดหมายอิเล็กทรอนิกส์ผ่านระบบอินเตอร์เน็ต



จากภาพที่ 2.1 ประกอบด้วย เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ (Electronic Mail Serve, E-mail Server) เครื่องลูกข่าย (Electronic Mail Client) และ โปรโตคอล (Protocol) ที่ใช้ในการรับส่งจดหมายอิเล็กทรอนิกส์

กระบวนการส่งจดหมายอิเล็กทรอนิกส์ มีขั้นตอนดังนี้

1) กระบวนการส่งจดหมายอิเล็กทรอนิกส์ (Sending) เริ่มจากผู้ใช้งานต้องการส่งข้อความผ่านโปรแกรมส่งจดหมายอิเล็กทรอนิกส์ในเครื่องลูกข่าย เช่น Outlook Express, Microsoft Outlook ฯลฯ โดยที่เครื่องลูกข่าย ทำการส่งข้อมูลผ่านโปรโตคอล SMTP (Simple Mail Transfer Protocol) ทำหน้าที่ส่งข้อมูลจดหมายอิเล็กทรอนิกส์ส่งต่อให้เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ เมื่อได้ข้อมูลจากเครื่องลูกข่ายเรียบร้อยแล้ว เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์จะทำการส่งจดหมายอิเล็กทรอนิกส์ไปตามที่อยู่ของผู้รับ (E-mail Address) ผ่านการส่งข้อมูลให้เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของฝ่ายผู้รับโดยผ่าน โปรโตคอล SMTP และโปโตคอล TCP ผ่านพอร์ต 25

2) กระบวนการรับจดหมายอิเล็กทรอนิกส์ (Receiving) เมื่อข้อมูลจดหมายอิเล็กทรอนิกส์ถูกส่งไปถึงเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของผู้รับแล้ว ผู้รับจะติดต่อเพื่อรับข้อมูลดังกล่าวโดยร้องขอไปเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของผู้รับ ผ่านระบบตรวจสอบชื่อผู้ใช้งาน (Login) เข้าไปที่เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของตนเอง โดยติดต่อรับข้อมูลผ่านโปรโตคอล POP3, IMAP และใช้โปรแกรมอ่านจดหมายอิเล็กทรอนิกส์อ่านจดหมายที่อยู่ในเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์หรือรับจดหมายอิเล็กทรอนิกส์ผ่านโปรโตคอล เช่น POP, IMAP ในการสื่อสารกับเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ โปรโตคอลที่ใช้คือ POP (Post Office Protocol) เป็นโปรโตคอลที่ใช้อ่านจดหมายในตู้จดหมาย ปัจจุบันใช้ Version 3 (POP3)

ขั้นตอนที่ 1 การทำงานของระบบ POP3 เริ่มตั้งแต่ที่ผู้ใช้งานต้องการดึงข้อมูลจากเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของตัวเอง โดยอาศัยการเชื่อมต่อผ่านโปรโตคอล TCP พอร์ต 110 โดยมีการทำงาน 3 ขั้นตอนด้วยกัน ขั้นตอนแรก ผู้ใช้งานต้องทำการเข้าใช้งานระบบ โดยทำการใส่ ชื่อผู้ใช้งาน (User Name) และ รหัสผ่าน (Password) เพื่อตรวจสอบสิทธิ์การใช้งาน

ขั้นตอนที่ 2 ระบบจะทำการดึงข้อมูลจากเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ ขั้นตอนนี้สามารถระบุได้ว่าจดหมายฉบับไหนที่ต้องการรับและไม่ต้องการรับจากเครื่องแม่ข่ายฯ

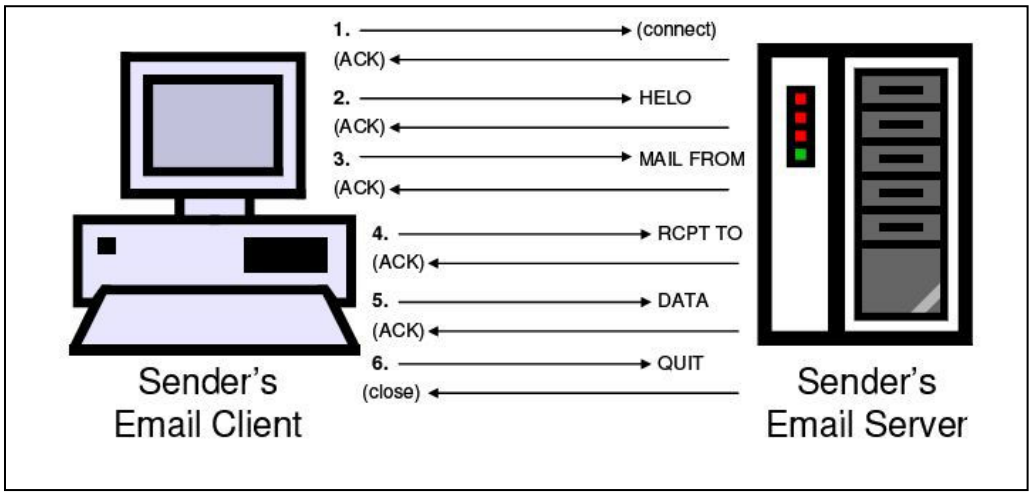
ขั้นตอนที่ 3 การสิ้นสุดการเชื่อมต่อ โดยเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์จะทำการลบจดหมายอิเล็กทรอนิกส์ของผู้ใช้คนนั้นออกจากตู้จดหมายของผู้ใช้และทำการสิ้นสุดการเชื่อมต่อ ทั้งสามขั้นตอนดังกล่าวอยู่ในช่วงการเชื่อมต่อ TCP พอร์ต 110

IMAP (Internet Message Access Protocol) เป็นโปรโตคอลที่ถูกพัฒนาให้มีประสิทธิภาพมากกว่า POP3 โดยปัจจุบันล่าสุดเป็น Version 4 แก้ไขที่ 1 (IMAP4rev1) โดยให้ผู้ใช้สามารถจัดการตู้จดหมายอิเล็กทรอนิกส์ ของตัวเองได้ที่เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ เช่น สามารถสร้าง Folder เพื่อจัดการเก็บจดหมายอิเล็กทรอนิกส์ สามารถย้ายจดหมายอิเล็กทรอนิกส์ไปมาระหว่าง แฟ้มข้อมูลของตัวเองได้ IMAP ยังอนุญาตให้ผู้ใช้สามารถ Download เฉพาะบางส่วนของจดหมายอิเล็กทรอนิกส์ ผ่านการทำงานของ IMAP ทางพอร์ท TCP 143 (จตุชัย แพงจันทร์ และคณะ, 2546)

2.1.2 **พื้นฐานการทำงานของโปรโตคอล** SMTP

หลักการทำงานของโปรโตคอล SMTP เป็นโปรโตคอลพื้นฐานที่ใช้ในการส่งจดหมายอิเล็กทรอนิกส์ มีหลักการทำงานดังนี้

ภาพที่ 2.2 แสดงการติดต่อสื่อสารผ่านโปรโตคอล SMTP



ที่มา: Weinberg, 2005

การติดต่อสื่อสารระหว่างผู้ส่ง และเครื่องแม่ข่ายที่ทำหน้าที่ส่งจดหมายอิเล็กทรอนิกส์ (Email Server) จะทำงานผ่านโปรโตคอลที่ชื่อ SMTP มี 6 ขั้นตอนดังนี้

1) สร้างการเชื่อมต่อโดยเครื่องลูกข่ายติดต่อไปยังเครื่องแม่ข่าย เมื่อเครื่องแม่ข่ายรับข้อมูลแล้วจะตอบกลับเป็นสัญญาณ Acknowledge (ACK)

2) เครื่องลูกข่ายส่งคำสั่ง HELLO เพื่อระบุว่าเป็นเครื่องคอมพิวเตอร์เครื่องใดเป็นผู้ส่ง เมื่อเครื่องแม่ข่าย รับข้อมูลแล้วจะตอบกลับเป็นสัญญาณ Acknowledge (ACK)

3) เครื่องลูกข่ายส่งคำสั่ง MAIL FROM เพื่อระบุว่าเป็นเครื่องคอมพิวเตอร์เครื่องใดเป็นผู้ส่ง เมื่อ เครื่องแม่ข่าย รับข้อมูลแล้วจะตอบกลับเป็นสัญญาณ Acknowledge (ACK)

4) เครื่องลูกข่ายส่งคำสั่ง RCPT TO เพื่อระบุว่าใครเป็นผู้รับ เมื่อเครื่องแม่ข่าย รับข้อมูลแล้วจะตอบกลับเป็นสัญญาณ Acknowledge (ACK)

5) เครื่องลูกข่ายส่งคำสั่ง DATA มาพร้อมกับข้อความในจดหมาย เมื่อเครื่องแม่ข่ายรับข้อมูลแล้วจะตอบกลับเป็นสัญญาณ Acknowledge (ACK)

6) เครื่องลูกข่ายส่งคำสั่ง QUIT เมื่อเครื่องแม่ข่ายรับข้อมูลแล้วจะตอบกลับเป็นสัญญาณยกเลิกการติดต่อ Close

2.1.3 **จดหมายอิเล็กทรอนิกส์สแปม**

จุดมุ่งหมายที่สำคัญของ จดหมายอิเล็กทรอนิกส์สแปมคือ ต้องการส่งจดหมายอิเล็กทรอนิกส์ให้ได้จำนวนมากที่สุด ดังนั้นงานที่นักพัฒนาจดหมายอิเล็กทรอนิกส์สแปม (Spammer) จะต้องทำเป็นอันดับแรกก็คือ การหารายชื่อบัญชีจดหมายอิเล็กทรอนิกส์ (E-mail Address) ให้ได้มากที่สุดเท่าที่จะมากได้ โดยใช้วิธีที่หลากหลาย อาทิเช่น การรวบรวมรายชื่อบัญชีจดหมายอิเล็กทรอนิกส์ จากนักพัฒนาสแปมด้วยกันเอง หรือจากหน่วยงานที่มีการซื้อขายรายชื่อบัญชีจดหมายอิเล็กทรอนิกส์ (List Merchants) การดำเนินการแบบ DHA (Directory Harvest Attacks) เพื่อทำการกวาดเอารายชื่อบัญชีจดหมายอิเล็กทรอนิกส์ที่เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของหน่วยงานซึ่งอาศัยช่องทางของ SMTP (Simple Mail Transfer Protocol) และเครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ที่ไม่ได้ทำการปิดการทำงานของระบบส่งต่อข้ามโดเมน (Relay) จุดประสงค์ของการปิดระบบ Relay เพื่อเป็นการระบุให้ผู้ส่งต้องมีรายชื่อบัญชีจดหมายอิเล็กทรอนิกส์อยู่ภายในโดเมน (Domain) ที่เป็นสมาชิกเท่านั้น บุคคลภายนอกโดเมน เช่น ผู้ส่งใช้เครื่องแม่ข่ายจดหมายอิเล็กทรอนิกส์ของ Yahoo จะไม่สามารถใช้งานเครื่องแม่ข่ายรายชื่อบัญชีจดหมายอิเล็กทรอนิกส์ของอีกบริษัทหนึ่งเป็นทางผ่านเพื่อส่งจดหมายอิเล็กทรอนิกส์ไปหายังบุคคลอื่นที่ไม่ได้อยู่ภายในโดเมนของเครื่องแม่ข่ายอิเล็กทรอนิกส์ได้ ซึ่งเป็นผลให้ระบบของโดเมน

นั้นถูกลักลอบขโมยข้อมูลบัญชีรายชื่อจดหมายอิเล็กทรอนิกส์ได้ (เลอศักดิ์ ลิ้มวิวัฒน์กุล, 2005) การใช้หุ่นยนต์สแปมบอต (Spambot) เพื่อทำการค้นหารายชื่อบัญชีจดหมายอิเล็กทรอนิกส์จาก Website ต่างๆ หลักการทำงานคล้ายกับตัวค้นหาของ Web Crawler ของ Search Engine จาก newsgroups, special-interest group (SIG) postings, และ chat-room conversation (Spambot, Wikipedia) สำหรับผลลัพธ์ที่ได้จากการหาบัญชีรายชื่อจดหมายอิเล็กทรอนิกส์ ผลลัพธ์ที่ได้คือ บัญชีรายชื่อจดหมายอิเล็กทรอนิกส์ที่มีตัวตนของผู้รับอยู่จริงรวมถึงมีรายการที่มีบัญชีรายชื่อจดหมายอิเล็กทรอนิกส์ที่ไม่มีผู้รับปลายทางที่ถูกต้องรวมอยู่ด้วย

2.2 ทฤษฎีการทำงานของบอตเน็ต

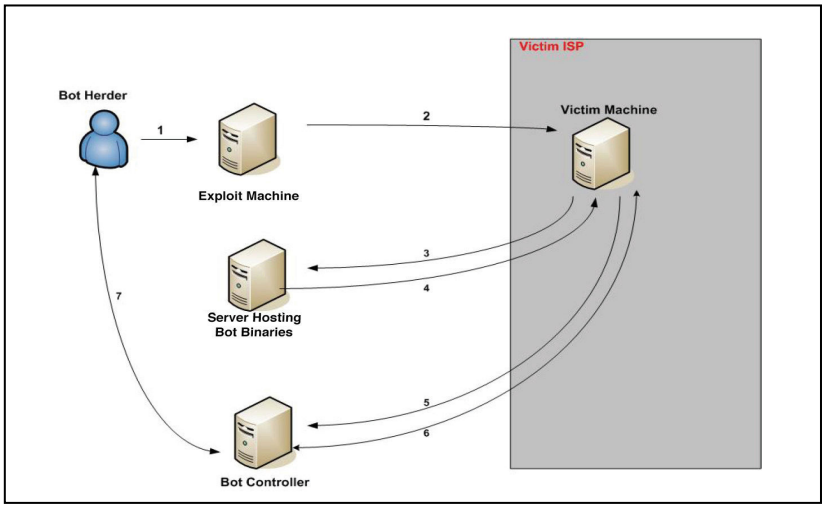
ในปัจจุบันการส่งจดหมายอิเล็กทรอนิกส์สแปมมีรูปแบบที่หลากหลายมากขึ้น รูปแบบหนึ่งที่ถูกนำมาพัฒนาและนำมาใช้ในการส่งจดหมายอิเล็กทรอนิกส์สแปม คือ การนำเอาหลักการของบอตเน็ต Botnet (roBOT NETwork) บางครั้งถูกเรียกว่า กองทัพซอมบี้ (Zombie PCs) เข้ามาเป็นเครื่องมือในการส่งจดหมายอิเล็กทรอนิกส์สแปม เครื่องที่มีโอกาสโดนโจมตีส่วนใหญ่ได้แก่ เครื่องคอมพิวเตอร์ที่อยู่ตามบ้านพักอาศัย โรงเรียน หน่วยงานรัฐบาล รวมถึงภาคธุรกิจ ซึ่งส่วนใหญ่จะทำงานบนระบบปฏิบัติการวินโดส์ (Windows) และเครื่องดังกล่าวมักมีการติดต่อกับระบบอินเตอร์เน็ตตลอดเวลา ซึ่งถ้าไม่มีการดูแลรักษาในด้านความปลอดภัยที่ดีพอก็มีโอกาสที่จะได้รับผลกระทบและถูกควบคุมจากผู้ไม่หวังดีเพื่อประโยชน์ในการส่งอิเล็กทรอนิกส์สแปม

จุดมุ่งหมายการทำงานของบอตเน็ต มีด้วยกันหลายวัตถุประสงค์ เช่น ใช้เป็นเครื่องมือในการส่งจดหมายอิเล็กทรอนิกส์สแปม ประมาณการว่ากว่า 70 เปอร์เซ็นต์ ของจดหมายอิเล็กทรอนิกส์สแปม ถูกส่งออกมาจากเครื่องที่เป็นบอตเน็ต และใช้ในการ Key Logger หรือ การคอยดักจับรหัส Password ข้อมูลบัตรเครดิต เลขที่บัญชีธนาคาร ประเด็นปัญหาที่สำคัญอีกประการหนึ่งคือ เมื่อกลุ่มของบอตมีจำนวนมากเพิ่มขึ้นจนกลายเป็นบอตเน็ต ซึ่งมีเครื่องที่เป็นบอตมากกว่า 1 ล้านเครื่องด้วยความสามารถขั้นพื้นฐานของการสื่อสารระดับ 128 Kbps. ของอุปกรณ์ MODEM ธรรมดา และพิจารณาการทำงานร่วมกันของเครื่องที่เป็นบอต ก็มีพลังในการสื่อสารถึง 128 Gigabits ซึ่งมีพลังงานมากพอที่จะให้บริษัท 500 บริษัทหยุดชะงักการให้บริการเนื่องจากถูกโจมตีในลักษณะเพื่อให้เกิดการหยุดการให้บริการแบบกระจาย (Distributed Denial of Service (DDoS) Attack) ถ้าเกิดการรวมตัวในหลายๆ กลุ่มบอตเน็ตก็จะสร้างความเสียหายในระดับประเทศรวมไปถึงหลายๆ ประเทศ (Baylor, Brown, 2006)

2.2.1 ลักษณะการแพร่กระจายของบอตเน็ต

ลักษณะการทำงานของบอตเน็ต เมื่อเครื่องคอมพิวเตอร์กลายเป็นบอตเน็ต (Infected) จากเครื่องบอตเน็ตอื่นที่อยู่ในเครือข่าย โดยอาศัยการแพร่กระจายของมัลแวร์ (Malware) เช่น ไวรัสคอมพิวเตอร์ (Computer Virus) เวิร์มหรือหนอนอินเตอร์เน็ต (Internet Worms) ม้าโทจัน (Trojan Horse) และสปายแวร์ (Spyware) ที่มีอยู่มากมายในระบบอินเตอร์เน็ต โดยมีวัตถุประสงค์ของการทำงานที่แตกต่างกันไป มัลแวร์เหล่านี้ถือว่าเป็นสื่อ (Media) ที่จะนำมาใช้ส่งรหัสโปรแกรม (Program Code) ที่ถูกสร้างจากผู้บุกรุก (Hacker, Spammer) เมื่อเครื่องของผู้ใช้งานกลายเป็นสมาชิกของกลุ่มบอตเน็ตแล้ว เครื่องดังกล่าวจะได้รับการติดต่อและสั่งงานจากศูนย์การสั่งการ (Bot Controller) ซึ่งจะทำการติดตั้งรหัสโปรแกรมที่สามารถทำงาน (Execute) ได้ และสามารถติดต่อสื่อสารไปยังเครื่องคอมพิวเตอร์อื่นๆ ที่อยู่ในกลุ่มของเครื่องที่เป็นบอตเน็ตซึ่งมีมากกว่า 100,000 เครื่องทั่วโลก ในปัจจุบันบอตเน็ตถูกพบว่ามันมีการทำงานที่ผสมผสานกันกับการทำงานของรูปแบบมัลแวร์ประเภทอื่นๆ เช่น สามารถแพร่ขยายตัวเองได้เหมือนกับหนอนอินเตอร์เน็ต สามารถซ่อนตัวเองจากการตรวจพบคล้ายกับไวรัสคอมพิวเตอร์ และมีรูปแบบการโจมตีโดยอาศัย Toolkits (Cooke, 2005) การแพร่กระจายของบอตเน็ตจะใช้เครื่องคอมพิวเตอร์ที่เป็นเครื่องควบคุมหลักเรียกว่า Bot Controller เป็นเครื่องหลักในการสั่งงานเครื่องที่ถูกโจมตี ตามภาพที่ 2.3

ภาพที่ 2.3 แสดงการแพร่กระจายของบอตเน็ต



ที่มา: Baylor, Brown, 2006

จากภาพที่ 2.3 มีขั้นตอนการทำงานรายละเอียดดังนี้

ขั้นตอนที่ 1 ผู้บุกรุก (Bot Header, Bot Master) จะทำการใส่ข้อมูลโปรแกรมเข้าไปที่เครื่องที่ขาดการป้องกันที่ดีพอ (Exploit Machine) ซึ่งเครื่องดังกล่าวสถานะอาจจะเป็นเครื่องที่ยังไม่เคยเป็นบอตมาก่อนหรือเป็นเครื่องที่กลายเป็นบอตเรียบร้อยแล้ว บอตส่วนใหญ่จะใช้ File Sharing และ Port RPC ในการแพร่กระจาย เพื่อให้แน่ใจว่าเครื่องที่ถูกโจมตีนี้ ได้รับการตั้งค่าเริ่มต้นที่เหมาะสมเพียงพอที่จะสามารถติดต่อไปที่เครื่อง Bot Controller ได้ในอนาคต

ขั้นตอนที่ 2 Exploit Machine ทำการ Scan ตรวจหาเครื่องที่ไม่มีระบบป้องกันภัยที่ดีเพื่อทำให้เครื่องเป้าหมายเครื่องนี้ (Victim Machine) กลายเป็นเครื่องที่มีคุณสมบัติที่จะเป็นบอตได้ในอนาคต (Exploit Machine)

ขั้นตอนที่ 3 และ ขั้นที่ 4 Victim Machine จะถูกสั่งให้ Download โปรแกรม Binary Code จากเครื่องคอมพิวเตอร์เครื่องแม่ข่ายอื่นๆ ซึ่งปรกติแล้วจะมาจากพวก Web Site หรือ FTP Server

ขั้นตอนที่ 5 Binary Code ที่ถูกติดตั้งในขั้นตอนที่ 3 และ 4 จะทำการสั่งให้ทำงานบน Victim Machine เพื่อเปลี่ยนให้เครื่องนี้กลายเป็นบอตที่สมบูรณ์ และเมื่อเครื่องนี้กลายเป็นบอตแล้วมันจะทำการติดต่อไปยังเครื่อง Bot Controller เพื่อคอยรับคำสั่งจากเครื่อง Bot Controller และคอยรายงานผลลัพธ์ต่างๆ ให้ Bot Controller ทราบ

ขั้นตอนที่ 6 Bot Controller ทำการออกคำสั่งไปที่เครื่อง Victim Machine (กลายเป็นบอตแล้ว) ซึ่งคำสั่งที่ถูกส่งออกไป อาจจะเป็นคำสั่งให้เครื่อง Victim Machine ทำการติดตั้งโปรแกรมใหม่ๆ ที่ถูกพัฒนาขึ้น เช่น โปรแกรมสำหรับสั่งให้ขโมยเลขที่บัญชี ติดตั้ง Spyware สั่งให้โจมตีเครื่องอื่นๆ และทำการส่งจดหมายอิเล็กทรอนิกส์สแปม

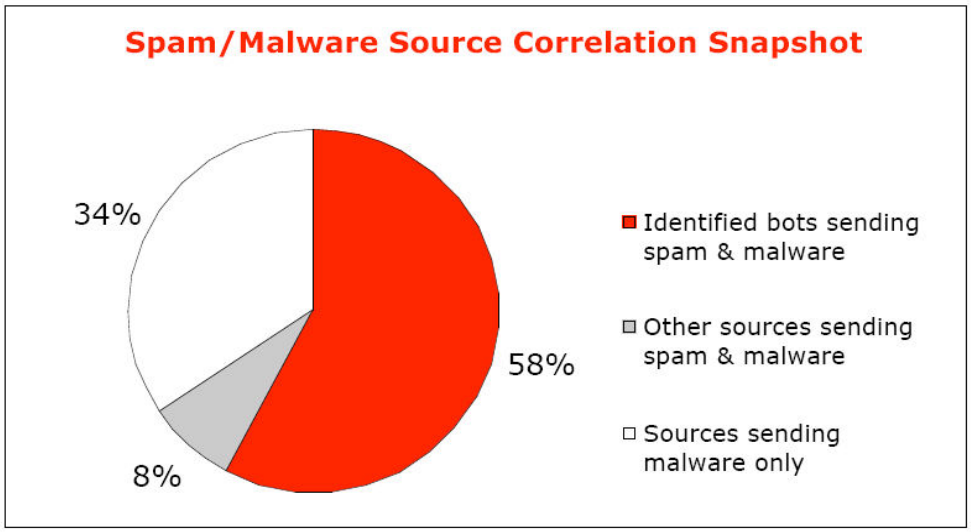
ขั้นตอนที่ 7 Bot Header คือบุคคลที่คอยควบคุมเครื่องบอตทั้งหมด โดยการส่งคำสั่งต่างๆ ผ่านเครื่อง Bot Controller

2.3 งานวิจัยที่เกี่ยวข้องกับจดหมายอิเล็กทรอนิกส์สแปมและบอตเน็ต

ปัญหาจดหมายอิเล็กทรอนิกส์สแปม ได้เป็นปัญหาที่ทั่วโลกกำลังประสบอยู่ เนื่องด้วยปัญหาจดหมายอิเล็กทรอนิกส์สแปมเป็นปัญหาที่มีความซับซ้อน และปัจจุบันได้พัฒนาวิธีการส่งจดหมายอิเล็กทรอนิกส์สแปมโดยนำ บอตเน็ต (Botnet) เข้ามาเป็นตัวแทนในการส่งจดหมายอิเล็กทรอนิกส์สแปม (Lanelli, Hackworth, 2006) ผลลัพธ์ของการนำบอตเน็ตมาใช้งาน

คือ ปริมาณจดหมายอิเล็กทรอนิกส์สแปมจำนวนมากถูกส่งออกมาจากเครื่องคอมพิวเตอร์หลายๆ เครื่องพร้อมกัน และบางกรณีจะส่งผลให้ Bandwidth ถูกใช้ไปอย่างมากทำให้การทำงานของระบบโดยรวมช้าลงและอาจส่งผลให้ระบบหยุดชะงักได้ถ้าเกิดปัญหาที่รุนแรงมาก

ภาพที่ 2.4 แสดงอัตราส่วนการส่งจดหมายอิเล็กทรอนิกส์สแปมจากแหล่งต่างๆ



ที่มา: Alt-N Technology, 2007

จากภาพที่ 2.4 จะเห็นได้ว่า กว่า 58 เปอร์เซ็นต์ ของจดหมายอิเล็กทรอนิกส์สแปมเกิดจากเครื่องที่เป็นบอตเน็ต

งานวิจัยในการตรวจสอบจดหมายอิเล็กทรอนิกส์สแปม (Hershkop, 2006) และ (Weinberg, 2005) แบ่งประเภทงานวิจัยแยกตามวิธีการที่ใช้ในการตรวจสอบ เช่น การใช้การตรวจสอบคำโดยพิจารณาจากเนื้อหาภายในจดหมายอิเล็กทรอนิกส์ (Content Based Filtering) การตั้งกฎในการรับจดหมายอิเล็กทรอนิกส์ (Rule-Based Scoring System) ทฤษฎีของ Bayesian Filters ที่เรียนรู้เปรียบเทียบกับจดหมายอิเล็กทรอนิกส์ก่อนหน้านั้น เช่น การใช้ Black List IP การใช้ Reverse DNS เป็นต้น โดยส่วนใหญ่จำเป็นต้องเปิดดูรายละเอียดข้อมูลในจดหมายอิเล็กทรอนิกส์แล้วนำมาวิเคราะห์เพื่อระบุว่าเป็นจดหมายอิเล็กทรอนิกส์สแปมหรือไม่ มีงานวิจัยบางงานวิจัย ได้นำเสนอการตรวจสอบ ในระดับเครือข่าย (Network Level) ของจดหมายอิเล็กทรอนิกส์สแปมกับเครื่องบอตเน็ต ซึ่งแสดงให้เห็นว่ากลุ่มของจดหมายอิเล็กทรอนิกส์สแปม ที่

ถูกส่งออกมาส่วนใหญ่จะอยู่ในกลุ่มของเครื่องที่เป็นบอตเน็ต (Ramachandran, Feamster, 2006)

งานวิจัยหลายๆ ฉบับใช้วิธีการตรวจสอบพฤติกรรมการติดต่อสื่อสารของบอตเน็ต เพื่อค้นหาเครื่องคอมพิวเตอร์ที่ผู้ควบคุมบอตใช้ในการติดต่อกับบอตเน็ตช่องทางหนึ่งที่นิยมคือ IRC (Internet Relay Chat) ผ่านทาง IRC Port หมายเลข 6667 แต่ปัญหาของผู้ควบคุมบอตเน็ตพบมาก คือ ช่องทางดังกล่าวอาจถูกปิดจากผู้ดูแลระบบ (James R. Binkley, Suresh Singh) ทำให้ผู้พัฒนาบอตต้องหาวิธีรูปแบบอื่นๆ ในการติดต่อ เช่นใช้วิธีการติดต่อกันโดยตรงโดยไม่ผ่านระบบ IRC เป็นต้น สำหรับงานวิจัยที่ผู้วิจัยได้นำเข้ามาร่วมในการออกแบบนี้ ได้แก่ งานวิจัยของ Choi โดยได้ทำการศึกษากลุ่มของกิจกรรมของเครื่องคอมพิวเตอร์ที่อยู่ในเครือข่าย (Group Activities) ที่มีพฤติกรรมในการร้องขอการตรวจสอบ Domain Name จากระบบ DNS และแสดงให้เห็นว่ากลุ่มเครื่องคอมพิวเตอร์ใดที่มีการร้องขอให้ตรวจสอบ Domain Name เดียวกันในช่วงเวลาเดียวกัน Domain Name นั้นเป็น Domain Name ที่บอตเน็ตใช้ในการติดต่อสื่อสาร และยังสามารถสอบกลับได้ว่า IP Address ใดบ้างที่ร้องขอเพื่อติดต่อกับ Domain Name นี้ รายละเอียดอัลกอริธึมของ Choi แสดงดังภาพที่ 2.5 และภาพที่ 2.6

ภาพที่ 2.5 อัลกอริธึมในการตรวจสอบบอตเน็ตจากการร้องขอผ่านระบบ DNS

Detect-BotDNS-Query (A_r)

1 FOR k = 1 to n
2 IF (A_{r1} => DN_k) is equal to (A_r => DN_k)
3 similarity(A_{r1} => $IPList_k$, A_{r2} => $IPList_k$)
4 S = computed similarity
5 IF $S > a$, a = Similarity threshold
6 DN_k is dotnet domain name
7 ELSE IF S = - 1 THEN insert(BL, DN_k) BL = blacklist
8 ELSE insert(W, DN_k)
9 ENDIF

End of Detect-BotDNS-Query

ที่มา: Choi H, et al, 2007

จากภาพที่ 2.5 สามารถอธิบาย รายละเอียดการออกแบบอัลกอริธึมได้ดังนี้

1) ระบบจะทำการวนรอบตั้ง 1 ครั้งถึง N ครั้ง

2) ถ้า Domain Name ณ เวลาที่1 (A) เป็นชื่อ Domain Name เดียวกันกับ Domain Name ที่ 2 (B) ณ เวลาที่ 2 จะทำการคำนวณหาคา Similarity

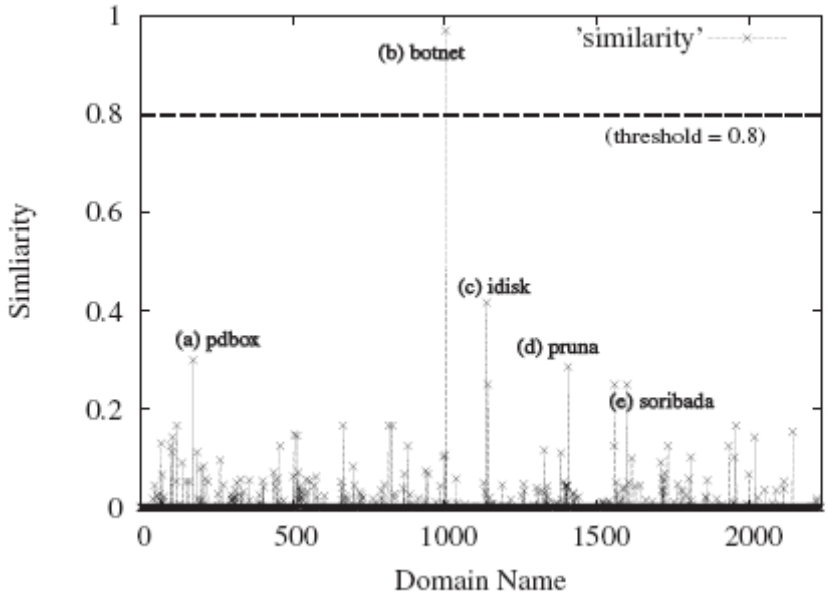
สูตร Similarity = 0.5 * (C/A + C/B) โดยที่ (A และ B <> 0)

3) ทำการเปรียบเทียบค่า Similarity ที่ได้จากการคำนวณในข้อ 2 กับ Threshold ถ้ามากกว่าค่า Threshold ซึ่งในงานวิจัยของ Choi ระบุค่า Threshold เป็น 0.8 ดังนั้นถ้า Domain Name ใดมีค่า Semilarity เกินกว่า 0.8 จะถือว่าเป็น Domain Name ที่บอตเน็ตต้องการจะติดต่อด้วย

4) ถ้าค่า Similarity ที่ได้เป็น -1 จะให้ถือว่าอยู่ในกลุ่ม Black List

5) ถ้าค่า Similarity มากกว่า -1 และต่ำกว่า 0.8 จะเป็น Domain ที่ไม่มีปัญหาให้ใส่ค่าไว้ใน White List

ภาพที่ 2.6 แสดงภาพ Domain Name ที่มีค่า Similarity สูงๆ



ที่มา: Choi H, et al, 2007

จากภาพที่ 2.6 แสดงให้เห็นว่า Domain ทั้งหมด 2,300 Domain Name มีเพียง Domain Name เดียวเท่านั้นที่เป็น Botnet คือ (b) botnet ซึ่งมีค่า Similarity สูงกว่า Threshold ที่ตั้งไว้คือ 0.8

ในปัจจุบันการศึกษาระบบป้องกันจดหมายอิเล็กทรอนิกส์สแปม (Anti-Spam E-mail) โดยส่วนใหญ่แล้วจะใช้วิธีการเปิดอ่านข้อมูลเนื้อหาของจดหมายอิเล็กทรอนิกส์ (Content based) วัตถุประสงค์เพื่อ การกรองจดหมายอิเล็กทรอนิกส์สแปมออกจากระบบจดหมายอิเล็กทรอนิกส์ ทั้งนี้งานวิจัยนี้ได้นำเสนอแนวทางที่แตกต่างออกไป ผู้วิจัยไม่ได้ศึกษาเพื่อสร้างระบบการกรองจดหมายอิเล็กทรอนิกส์สแปม ผู้วิจัยมีวัตถุประสงค์เพื่อลดอัตราการส่งจดหมายอิเล็กทรอนิกส์สแปมที่ออกมาจากเครื่องบอตเน็ต โดยนำแนวคิดเบื้องต้นในเรื่องของการลดอัตราการแพร่กระจายของหนอนอินเตอร์เน็ต (Internet Worms) ของ Chen และ Tang ซึ่งพัฒนาสถาปัตยกรรม DAW (Distributed Anti-worms Architecture) เพื่อใช้ในการลดอัตราการแพร่กระจายของหนอนอินเตอร์เน็ต (Slowing Down Internet Worms) โดยการวิเคราะห์อัตราการส่งที่ผิดพลาด (Failure Rate) ที่เกิดขึ้นเนื่องจากหนอนอินเตอร์เน็ตส่ง SYN Package ไปยัง IP Address แบบสุ่ม วัตถุประสงค์เพื่อกระจายตัวเอง (Random IP Address) บางครั้งเป็น IP Address ที่ไม่มีอยู่จริงในระบบอินเตอร์เน็ต ดังนั้นค่าที่ตอบกลับจาก TCP/IP จึงเป็นอัตราการส่งที่ผิดพลาดปริมาณสูง หลังจากที่ระบบตรวจสอบพบ ระบบจะทำการลดอัตราการแพร่กระจายโดยวิธีจำกัด Bandwidth ที่ทำการส่งข้อมูลของ IP Address นั้นให้น้อยลงจนกระทั่งหยุดการให้บริการส่งข้อมูลออกจากเครื่องนั้น (Block SYN Package) จากแนวความคิดดังกล่าว ผู้วิจัยได้เกิดแนวความคิดในลักษณะเดียวกันนี้โดยนำมาปรับใช้กับการลดอัตราการส่งจดหมายอิเล็กทรอนิกส์สแปมจากเครื่องที่เป็นบอตเน็ตและทำการส่งจดหมายอิเล็กทรอนิกส์สแปม โดยนำเสนอระบบ Detecting and Slowing Down Spam E-mail System (DSDSE) ระบบ DSDSE มีวัตถุประสงค์ในการลดอัตราการส่งจดหมายอิเล็กทรอนิกส์สแปมที่ถูกส่งออกจากเครื่องคอมพิวเตอร์ที่เป็นบอตเน็ต โดยใช้เทคนิคการตรวจสอบบอตเน็ตจากพฤติกรรมการสอบถามข้อมูล Domain Name ผ่านระบบ Domain Name System (DNS) และเทคนิคการตรวจสอบความผิดปรกติในการส่งจดหมายอิเล็กทรอนิกส์ เมื่อนำเอาข้อมูลทั้งสองอย่างมาวิเคราะห์ผลลัพธ์ที่ได้คือ เครื่องคอมพิวเตอร์ที่มีแนวโน้มที่จะเป็นบอตเน็ตและกำลังส่งจดหมายอิเล็กทรอนิกส์สแปม ในท้ายที่สุดระบบจะพยายามลดขนาด Bandwidth ของเครื่องดังกล่าวเพื่อลดอัตราการส่งจดหมายอิเล็กทรอนิกส์สแปม โดยไม่ได้มุ่งเน้นที่ในการยกเลิกการส่งจดหมายอิเล็กทรอนิกส์สแปม (Reject Spam E-mail) ผลลัพธ์ที่ได้สามารถทำให้ภาพรวมการใช้งาน Bandwidth ในระบบเครือข่ายถูก

นำมาใช้อย่างมีประสิทธิภาพจากระบบ DAW ของ Chen และ Tang นั้นเน้นการลดอัตราการเติบโตของหนอนอินเตอร์เน็ต โดยในงานวิจัยนี้ดำเนินการกับจดหมายอิเล็กทรอนิกส์สแปม ไม่ได้ลดอัตราการเติบโตของจดหมายอิเล็กทรอนิกส์สแปม ดังนั้นประเด็นนี้ผู้วิจัยจึงปรับกระบวนการไปในแนวทางผลสัมฤทธิ์ในแง่ของ Bandwidth ที่ได้กลับคืนมาจากการลดอัตราการส่งจดหมายอิเล็กทรอนิกส์สแปม