ตัวแบบทางคณิตศาสตร์ในการเลือกตำแหน่งที่ตั้งคลังสินค้าและศูนย์กระจายสินค้า สำหรับบริษัท ไปรษณีย์ไทยดิสทริบิวชั่น จำกัด

Mathematical Modeling for Location Selection of Warehouse and Distribution Centers for Thailand Post Distribution Co., Ltd.

สิวนีย์ ปงลังกา[1*] และ วลัยลักษณ์ อัตธีรวงศ์[2]

[1,2]ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

[*]E-mail: Siwanee2985@gmail.com

บทคัดย่อ

การตั้งศูนย์กระจายสินค้าใกล้กับแหล่งลูกค้าปลายทางจะช่วยให้การขนส่งสินค้าไปยังลูกค้ามีความสะดวกรวดเร็วมากขึ้น แต่อาจทำให้ต้นทุนการขนส่งโดยรวมเพิ่มสูงขึ้นหากมีจำนวนศูนย์กระจายสินค้ามากเกินไป ปัจจุบันบริษัท ไปรษณีย์ไทยดิสทริบิวชั่น จำกัด เป็นผู้ให้บริการขนส่งสินค้ากลุ่มยาและเวชภัณฑ์ไปยังลูกค้าในจังหวัดต่างๆ ทั่วประเทศ โดยมีคลังสินค้า 1 แห่งตั้งอยู่ที่จังหวัดสมุทรปราการและศูนย์กระจายสินค้าอีก 9 แห่งกระจายอยู่ในภูมิภาคต่างๆ ที่ผ่านมาทางบริษัทยังไม่เคยวิเคราะห์เกี่ยวกับตำแหน่งที่ตั้งคลังสินค้าและจำนวนศูนย์กระจายสินค้าที่เหมาะสม ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อหาตำแหน่งที่ตั้งคลังสินค้าและศูนย์กระจายสินค้าที่เหมาะสมให้กับทางบริษัทในการที่จะขนส่งสินค้าไปยังลูกค้าปลายทางด้วยต้นทุนการขนส่งโดยรวมที่ต่ำที่สุดโดยใช้ตัวแบบทางคณิตศาสตร์และประมวลผลด้วยโปรแกรม LINGO ผลการวิจัยสรุปว่า ถ้าบริษัทมีคลังสินค้า 1 แห่งที่จังหวัดสมุทรปราการและมีศูนย์กระจายสินค้าทั้งหมด 12 แห่ง จะสามารถลดต้นทุนการขนส่งได้ 1,205,427 บาทต่อเดือน คิดเป็นร้อยละ 5.52

คำสำคัญ: ปัญหาการเลือกตำแหน่งที่ตั้ง, คลังสินค้าและศูนย์กระจายสินค้า, ตัวแบบทางคณิตศาสตร์

Abstract

Construction Distribution Center proximity to customer will help product transportation to customers more convenient and faster. But it may increase total transportation cost if we have too many distribution centers. Currently Thailand Post Distribution Co., Ltd. is the pharmaceutical product transporter to customers in various provinces around the country. They have a warehouse in Samut Prakan province and nine distribution centers in various regions. In the past, they have never analysis about a warehouse location and optimal number of distribution center. So the objective of this research is to select location of warehouse and distribution center for this company to ship products to customers with the lowest transportation costs using mathematical model and processing with LINGO. The result shows that if this company has a warehouse in Samut Prakan province and distribution center in twelve provinces. It can reduce the transportation cost by 1,205,427 baht per month or 5.52%

Keywords: Location Problem, Warehouse and Distribution Center, Mathematical Model

เอกสารอ้างอิง

สำนักงานสถิติแห่งชาติ. (2560). **สรุปผลที่สำคัญ สำรวจการมีการใช้เทคโนโลยีสารสนเทศและการสื่อสารในครัวเรือน พ.ศ. 2560**. ค้นเมื่อ 31 พฤษภาคม 2562 จาก http://www.nso.go.th/sites/2014/สำรวจ/เทคโนโลยีสารสนเทศ/เทคโนโลยีในครัวเรือน.aspx

ศูนย์ประสานการรักษาความมั่นคงปลอดภัยระบบคอมพิวเตอร์ประเทศไทย (ไทยเซิร์ต). (2562). **สถิติภัยคุกคามประจำปี พ.ศ. 2562**. ค้นเมื่อ 31 พฤษภาคม 2562 จากhttps://www.thaicert.or.th/statistics/statistics.html

MindPhp v4.0 (2561) **Honeypots (ฮันนีพอต) ระบบข้อมูลไฟร์วอล** ค้นหาเมื่อ 15 กันยายน 2562 https://www.mindphp.com/%E0%B8%9A%E0%B8%97%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1/244-security/5742-honeypots-security.html

Quora. (2560).**What is a state of Markov chain**. ค้นเมื่อ 31 พฤษภาคม 2562 https://www.quora.com/What-is-a-state-of-Markov-chain

Ka Ching ChanC. T. LenardTerence M MillsTerence M Mills. (2012). **An Introduction to Markov Chains**. ค้นเมื่อ 31 พฤษภาคม 2562 จาก https://www.researchgate.net/publication/258927967_An_Introduction_to_Markov_Chains

Juang, B. H., and Rabiner, l. R. (1986).**An introduction to hidden Markov models**. ค้นเมื่อ 31 พฤษภาคม 2562 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.957.202&rep=rep1&type=pdf

Eugine Kang. (2560). **Hidden Markov Models.** ค้นเมื่อ 31 พฤษภาคม 2562 จาก https://medium.com/@kangeugine/hidden-markov-model-7681c22f5b9

Sanjay Dorairaj . (2561). **Hidden Markov Models Simplified**. ค้นเมื่อ 31 พฤษภาคม 2562 จาก https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab

Mark Stamp. (2018). **A Revealing Introduction to Hidden Markov Models**. ค้นเมื่อ 31 พฤษภาคม 2562 จาก https://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf

รูปภาพที่ 7 Emission Matrix ของแบบจำลอง Hidden Markov Model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00000	0.00000	0.00000	0.00000	0.00007	0.00000	0.00097	0.00003	0.00049	0.00000	0.00003	1.00411	0.00000	0.00000	0.00003	0.00003	0.00000	0.00003	0.00003	0.00003	0.00000	0.00000	0.00000
2	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	0.00039	0.00001	0.00013	0.00000	0.00001	1.00172	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
3	0.00000	0.00000	0.00000	0.00000	0.33302	0.00000	0.00048	0.00002	0.00017	0.00000	0.00002	0.66915	0.00000	0.00000	0.00002	0.00002	0.00000	0.00002	0.00002	0.00002	0.00000	0.00000	0.00000
4	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00013	0.00000	0.00004	0.00000	0.00000	1.00056	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
5	0.00000	0.00000	0.00000	0.00000	0.00013	0.00000	0.00193	0.00007	0.00067	0.00000	0.00007	1.00851	0.00000	0.00000	0.00007	0.00007	0.00000	0.00007	0.00007	0.00007	0.00000	0.00000	0.00000
6	0.00000	0.00000	0.00000	0.00000	0.00005	0.00000	0.00072	0.00002	0.00025	0.00000	0.00002	1.00320	0.00000	0.00000	0.00002	0.00002	0.00000	0.00002	0.00002	0.00002	0.00000	0.00000	0.00000
7	0.00000	0.00000	0.00000	0.00000	0.00013	0.00000	0.00193	0.00007	0.00067	0.00000	0.00007	1.00851	0.00000	0.00000	0.00007	0.00007	0.00000	0.00007	0.00007	0.00007	0.00000	0.00000	0.00000
8	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000	0.00029	0.00001	0.99961	0.00000	0.00001	0.00176	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
9	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	0.00040	0.00001	0.00014	0.00000	0.00001	1.00176	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
10	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	0.00038	0.00001	0.00013	0.00000	0.00001	1.00169	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
11	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000	0.00027	0.00001	0.00009	0.00000	0.00001	1.00117	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
12	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.95655	0.04348	0.00004	0.00000	0.00000	0.00066	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
13	0.00000	0.00000	0.00000	0.00000	0.00005	0.00000	0.00073	0.00003	0.00032	0.00000	0.00003	1.00313	0.00000	0.00000	0.00003	0.00003	0.00000	0.00003	0.00003	0.00003	0.00000	0.00000	0.00000
14	0.00000	0.00000	0.00000	0.00000	0.00007	0.00000	0.00097	0.00003	0.00033	0.00000	0.00003	0.00503	0.00000	0.00000	0.00003	0.33311	0.00000	0.33311	0.33311	0.00003	0.00000	0.00000	0.00000
15	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00013	0.00000	0.00005	0.00000	0.00000	1.00058	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
16	0.00000	0.00000	0.00000	0.00000	0.00020	0.00000	0.00289	0.00010	0.00100	0.00000	0.99778	0.01509	0.00000	0.00000	0.00010	0.00010	0.00000	0.00010	0.00010	0.00010	0.00000	0.00000	0.00000
17	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00145	0.00005	0.00050	0.00000	0.00005	0.00755	0.00000	0.00000	0.49947	0.00005	0.00000	0.00005	0.00005	0.49947	0.00000	0.00000	0.00000
18	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	0.00042	0.00001	0.00014	0.00000	0.00001	1.00183	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
19	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00013	0.00000	0.00004	0.00000	0.00000	1.00056	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
20	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	1.00009	0.00001	0.00014	0.00000	0.00001	0.00216	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
21	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000	0.00027	0.00001	0.00009	0.00000	0.00001	1.00120	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
22	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	0.00040	0.00001	0.00014	0.00000	0.00001	1.00176	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
23	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000	0.00042	0.00001	0.00014	0.00000	0.00001	1.00185	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00001	0.00000	0.00000	0.00000

รูปภาพที่ 6 Transition Matrix ของแบบจำลอง Hidden Markov Model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00007	0.00003	0.00005	0.00003	0.00003	0.00007	0.00003	0.00005	0.00003	0.00003	0.00003	0.00005	0.99903	0.00003	0.00010	0.00003	0.00007	0.00003	0.00005	0.00003	0.00003	0.00003	0.00003
2	0.00001	0.00001	0.00001	0.00001	0.00004	0.00001	0.00001	0.00001	0.00001	0.04474	0.01122	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.31720	0.00001	0.00001	0.01110	0.00001	0.61548
3	0.00002	0.00002	0.00003	0.00002	0.00002	0.00005	0.00002	0.83302	0.00002	0.00002	0.00002	0.00005	0.00005	0.00002	0.00005	0.00002	0.16648	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
4	0.00000	0.32320	0.00000	0.00000	0.00001	0.00000	0.00001	0.00000	0.31697	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.04348	0.00000	0.00002	0.00000	0.00000	0.00000	0.31619	0.00006
5	0.66555	0.00007	0.00008	0.00007	0.33293	0.00009	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00014	0.00007	0.00007	0.00007	0.00007	0.00008	0.00007	0.00007	0.00007	0.00007	0.00007
6	0.00003	0.00003	0.00004	0.00003	0.00003	0.00007	0.00003	0.99936	0.00003	0.00003	0.00003	0.00005	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003
7	0.00009	0.00007	0.00007	0.00007	0.66519	0.00007	0.33334	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007	0.00007
8	0.09981	0.00001	0.60012	0.00001	0.00001	0.00010	0.00001	0.00001	0.00001	0.00001	0.04651	0.00001	0.09992	0.00001	0.00001	0.00001	0.00001	0.00001	0.10000	0.00001	0.05337	0.00001	0.00001
9	0.00001	0.00001	0.00001	0.00001	0.00004	0.00001	0.00001	0.00001	0.00001	0.47100	0.01260	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.31589	0.00001	0.00001	0.01245	0.00001	0.18779
10	0.03273	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.08212	0.00001	0.00001	0.00001	0.02090	0.00001	0.00001	0.00001	0.00001	0.77504	0.08897	0.00001	0.00001
11	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00002	0.00001	0.99978	0.00001	0.00001	0.00001	0.00002	0.00001	0.00001	0.00001	0.00001
12	0.00000	0.00002	0.00000	0.99987	0.00000	0.00000	0.00001	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000
13	0.00003	0.00003	0.00006	0.00003	0.00003	0.99927	0.00003	0.00006	0.00003	0.00003	0.00003	0.00005	0.00007	0.00003	0.00003	0.00003	0.00004	0.00003	0.00005	0.00003	0.00003	0.00003	0.00003
14	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.49948	0.00005	0.00005	0.00005	0.00005	0.00005	0.49947	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005
15	0.00000	0.00000	0.00001	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.99988	0.00000	0.00000	0.00000	0.00000
16	0.00010	0.00010	0.00010	0.00012	0.00014	0.00010	0.99774	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010
17	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.99890	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005	0.00005
18	0.08637	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.23321	0.00001	0.00001	0.00001	0.02761	0.00001	0.00001	0.00001	0.00001	0.13335	0.51920	0.00001	0.00001
19	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.99989	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
20	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.72773	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.27197	0.00001	0.00001
21	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00005	0.00001	0.99973	0.00001	0.00001	0.00001	0.00004	0.00001	0.00001	0.00001	0.00001
22	0.00001	0.00001	0.00001	0.00001	0.00004	0.00001	0.00001	0.00001	0.00001	0.52170	0.01276	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.31770	0.00001	0.00001	0.01265	0.00001	0.13492
23	0.02149	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.40623	0.00001	0.00001	0.00001	0.01642	0.00001	0.00001	0.00001	0.00001	0.02981	0.52579	0.00001	0.00001

รูปภาพที่ 5 Transition Matrix ของแบบจำลอง Markov Chain Model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.008850	0.000000	0.000000	0.000000	0.000000	0.000000	0.008850	0.982301	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.053814	0.766953	0.179233	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.097420	0.000000	0.000122	0.000791	0.882926	0.000000	0.001339	0.010770	0.000000	0.000000	0.005233	0.000061	0.000000	0.000000	0.000548	0.000000	0.000669	0.000000	0.000122
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000866	0.000067	0.000050	0.000000	0.000025	0.998959	0.000000	0.000000	0.000025	0.000000	0.000000	0.000000	0.000008	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000496	0.002835	0.000567	0.000000	0.002268	0.993763	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000071	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.227815	0.000000	0.010140	0.000799	0.023143	0.000000	0.000949	0.729813	0.000000	0.000000	0.006310	0.000000	0.000000	0.000000	0.000999	0.000000	0.000000	0.000000	0.000033
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.005087	0.000000	0.000000	0.000000	0.000424	0.000000	0.000000	0.994065	0.000000	0.000000	0.000424	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000007	0.000000	0.000000	0.001589	0.000000	0.174392	0.020451	0.052932	0.000000	0.006719	0.733210	0.000000	0.000000	0.010363	0.000012	0.000000	0.000000	0.000175	0.000000	0.000112	0.000000	0.000037
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.001340	0.007257	0.000000	0.000000	0.991403	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.002949	0.000000	0.001201	0.000000	0.833352	0.000655	0.000437	0.010593	0.000000	0.000000	0.148083	0.000000	0.000000	0.000000	0.002730	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.465421	0.024299	0.510280	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.007790	0.000000	0.000000	0.000216	0.000000	0.000000	0.000000	0.007934	0.000000	0.000000	0.000505	0.000000	0.000000	0.001731	0.637911	0.000000	0.000000	0.343912	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.321590	0.000377	0.000000	0.000000	0.000000	0.006037	0.000000	0.000000	0.000849	0.000000	0.000000	0.001603	0.005706	0.023533	0.409451	0.230853	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.539474	0.000000	0.000000	0.000000	0.032895	0.000000	0.427632	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.159091	0.000000	0.000000	0.000000	0.005682	0.000000	0.000000	0.000000	0.835227

ตารางที่ 1 สรุปผลประเมินความแม่นยำ

Model	Data Evaluation (ร้อยละ)
แบบจำลอง Markov Chain (MC)	79.29
แบบจำลอง Hidden Markov Model (HMM)	90.40

จากตางรางที่ 2 ข้างต้น จะเห็นได้ว่าค่าความแม่นยำของแบบจำลอง Markov Chain (MC) มีค่าความแม่นยำร้อยละ 79.29 และค่าความแม่นยำของแบบจำลอง Markov Chain (MC) มีค่าร้อยละ 90.40 ซึ่งการวัดผลความแม่นยำนั้นทำโดยคำนวณจากการหาสถานะถัดไปโดยใช้โมเดล MC, HMM มาเปรียบเทียบกับข้อมูลที่ใช้ในการทดสอบ หากนำมาเปรียบเทียบความแม่นยำสำหรับการการศึกษาพฤติกรรมที่ผิดปกติจากข้อมูลที่ได้มาจากหน่วยงานราชการด้านความมั่นคงแห่งนี้ จะเห็นได้ว่าแบบจำลองที่เหมาะกับการตรวจจับพฤติกรรมที่ผิดปกติของเครื่องคอมพิวเตอร์แม่ข่ายมากที่สุด คือ แบบจำลอง Hidden Markov Model (HMM) จะเหมาะสมในการตรวจจับและป้องกันพฤติกรรมที่ผิดปกติในการเข้าถึงเครื่องคอมพิวเตอร์แม่ข่ายมากที่สุด

2. แบบจำลอง Hidden Markov Model (HMM)

จากการทดลอบด้วยแบบจำลอง Hidden Markov Model อาจแบ่ง States ได้ทั้งหมด 23 States จึงดำเนินการกำหนดพารามิเตอร์ที่เป็น States (N) ได้ทั้งหมดจำนวน 23 ตัว โดยเป็น Observation (M) ทั้งหมดจำนวน 23 ตัว โดยกำหนดขนาดของ Observation Sequence (T) มีค่าเท่ากับ 199 เนื่องจากข้อมูลที่ได้จากขั้นตอนการเตรียมข้อมูลได้จำนวน Sequence ที่เป็นการโจมตี มีจำนวน Sequence 199 เป็นลำดับเยอะที่สุด ผู้วิจัยจึงเลือกใช้ Observation Sequence ที่มีขนาด 199 จะได้

mod = hmm.HMM(N, M, T)

เมื่อได้ค่า mod แล้ว จะนำค่า mod มาประมวลผลเพื่อหา Transition Matrix เพื่อหาความน่าจะเป็นในการเปลี่ยน Emission Matrix ของแบบจำลอง Hidden Markov Model ซึ่งได้ Transition Matrix และ Emission Matrix ของแบบจำลอง Hidden Markov Model โดยมีรายละเอียดตามรูปภาพที่ 6 และ 7 ตามลำดับ

เมื่อได้ Transition Matrix และ Emission Matrix แล้ว หลังจากนั้นก็นำมาหาความน่าจะเป็น จาก Observation Prior ในการเปลี่ยนเป็น States เดิม โดยการคำนวณนั้นจะใช้ค่าที่เป็น Maximum likelihood ดังนี้

19,17,15,8,11,8,4,8,4,8,11,8,11,8,11,11,11,8,11,11,11,6,11,11,11,11,11
11,6,11,11,11,11,11,11,6,11,11,11,11,11,11,6,11,11,11,11,11,11,6,11,11
11,11,11,11,6,11,11,11,11,11,11,6,11,11,11,11,11,11,6,11,11,11,11,11,11
6,11,11,11,11,11,11,6,11,11,11,11,11,11,6,11,11,11,11,11,6,11,11,11,11
11,11,6,11,11,11,11,11,11,6,11,11,11,11,11,11,6,11,11,11,11,11,11,8,11
6,11,11,11,6,11,11,11,6,11,11,11,6,11,11,11,6,11,11,11,6,11,11,11,6,11
11,11,6,11,11,11,6,11,11,11,6,11,11,11,6,11,11,11,6,11,11,11,6,11,11,11
6,11,11,11,7,11,10,11,11,11,11,11,11,11,8,11,11,8,11,14,18

โดยข้อมูลที่ทำการทดสอบถือว่าเป็นข้อมูลที่โดนโจมตีทั้งหมด จากขนาดของ Observation Sequence ทั้งหมด 199 ทำให้ต้องหา likelihood ที่สูงที่สุด (Maximum likelihood) ของ ในแต่ละ epochs โดยกำหนดค่าความต่างของ likelihood ในแต่ละ epochs น้อยกว่า 0.001

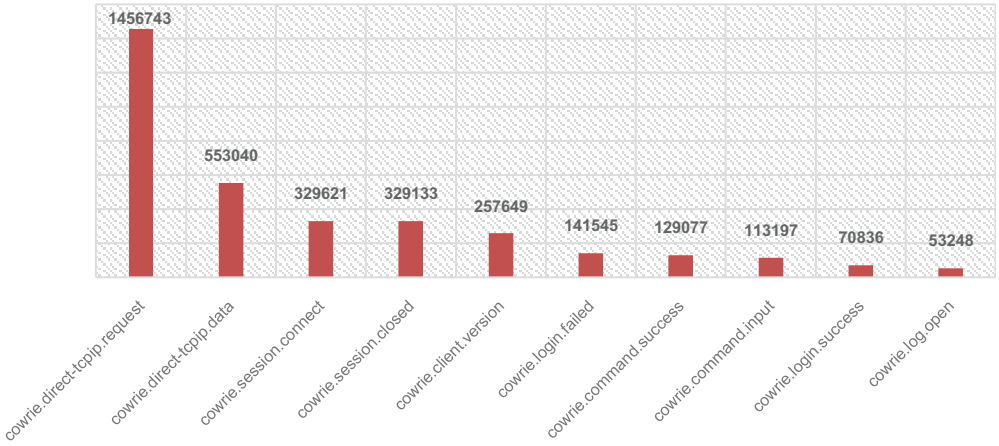
เมื่อได้ค่าความต่างที่น้อยกว่า 0.001 แล้วจึงหยุดทำการทดสอบ แล้วจะนำผลที่ได้มาประเมินความถูกต้องโดยวิธี เปรียบเทียบความแม่นยำระหว่างข้อมูลได้จาก Honeypots กับข้อมูลจากการใช้แบบจำลอง HMM ซึ่งได้ค่าความแม่นยำร้อยละ 90.40

5. สรุปผลการศึกษาวิจัย

การทดสอบนี้เป็นการวิจัยการตรวจจับพฤติกรรมที่ผิดปกติของเว็บแอพพลิเคชันจากการเรียนรู้ของเครื่องกลหรือ Machine learning โดยการนำข้อมูลที่ร้องขอใช้งานเว็บแอพพลิเคชันซึ่งมีรูปแบบ Honeypots ที่เป็นการหลอกให้ผู้ไม่ประสงค์ดีต่อเว็บแอพพลิเคชันมาโจมตี Honeypots แทนระบบจริง หลังจากนั้นก็จัดเก็บข้อมูลและคัดกรองข้อมูลเพื่อนำมาทำการวิจัยต่อไป

เมื่อได้ข้อมูลการร้องขอข้อมูลเว็บแอพพลิเคชันจาก Honeypots แล้วได้นำเข้าสู่กระบวนการคัดกรองและจัดกลุ่มข้อมูล เพื่อนำมาประมวลผลและและประมวลผลผ่านแบบจำลอง MC และHMM และได้นำไปประเมินความแม่นยำของทั้งสองแบบจำลอง สรุปได้ตามตารางที่ 1 รายละเอียดดังนี้

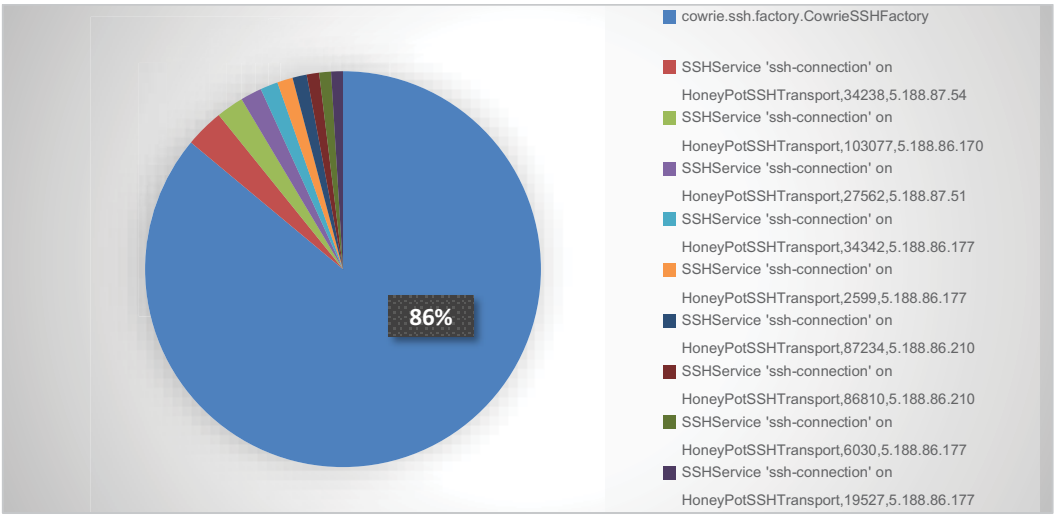
แผนภูมิที่ 6 Event Distribution



3. Highest Number of Attacks

Highest Number of Attacks คือ พฤติกรรมที่ผิดปกติแบ่งตามกลุ่มของไอพีแอดเดรสต้นทาง (Source IP) ที่เข้ามาโจมตีเครื่องคอมพิวเตอร์แม่ข่าย โดยพิจารณาจัดกลุ่มประเภทได้ดังนี้ 1) cowrie.ssh.factory.cowriesshfactory มีสัดส่วนมากที่สุด คิดเป็นร้อยละ 86 โดยมี SSHService ‘ssh-connection’ on HoneyPotSSHTransport,34238,5.188.87.54 มีมากเป็นลำดับต่อมา และ SSHService ‘ssh-connection’ on HoneyPotSSHTransport,103077,5.188.86.170 มีมากเป็นลำดับที่สาม ซึ่งพิจารณาได้จากแผนภูมิที่ 6 ดังนี้

แผนภูมิที่ 6 Highest Number of Attacks



4.2 ข้อมูลจากการทดสอบของแบบจำลอง

1. แบบจำลอง Markov Chain Model (MC)

แบบจำลอง Markov Chain หลักจากที่ผ่านขั้นตอนการเตรียมข้อมูลมาแล้ว จะได้ข้อมูลที่เป็นลักษณะ List ของ State ในแต่ Session ตัวอย่างดังนี้ หลักจากนั้นทำการคำนวณหา Transition Matrix เพื่อหาความน่าจะเป็นในการเปลี่ยนของแต่ละ State จากข้อมูลที่ผ่านการคัดแยกมาแล้ว มีข้อมูลทั้งหมด 1,058,090 แถว ผลที่ได้จากการคำนวณหา Transition Matrix ได้ผลดังนี้ โดยมีรายละเอียดตามรูปภาพที่ 5

เมื่อได้ Transition Matrix แล้วจึงนำมาประมวลผลแล้วนำผลลัทธ์ที่ได้มาประเมินความถูกต้องโดยวิธีเปรียบเทียบความแม่นยำระหว่างข้อมูลได้จาก Honeypots กับข้อมูลจากการใช้แบบจำลอง Markov Chain Model (MC) ซึ่งจะได้ค่าความแม่นยำร้อยละ 79.29

ขั้นตอนการวัดประสิทธิภาพ (Data Evaluation) เป็นขั้นตอนการประเมินความแม่นยำของแบบจำลอง ซึ่งจะนำมาผลการวัดประสิทธิภาพของแบบจำลอง Hidden Markov Model (HMM) และแบบจำลอง Markov Chain (MC) มาเปรียบเทียบประสิทธิภาพความแม่นยำ โดยวิธีนำข้อมูลที่ Model ทำการพยากรณ์ มาเทียบกับข้อมูลที่อยู่ใน State ถัดไป หากข้อมูลที่พยากรณ์ตรงกับข้อมูล State ถัดไปจะเป็นการพยากรณ์ที่ถูกต้อง

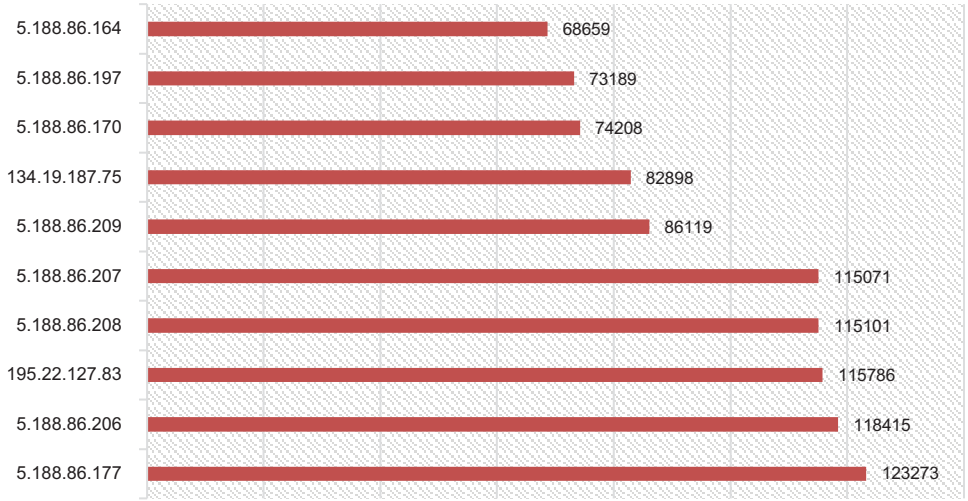
4. รายงานผลการวิจัย

4.1 ข้อมูลทั่วไปของรูปแบบของข้อมูลในการวิจัย

1. ไอพีแอดเดรสต้นทาง (Source IP)

ไอพีแอดเดรสต้นทาง คือ ไอพีแอดเดรสของต้นทางที่มีการ SSH เข้ามาทำการใช้งานระบบ ซึ่งมีการร้องของใช้งานเว็บแอพพลิเคชันจากไอพีแอดเดรสต้นทาง 5 ลำดับแรกดังนี้ ไอพีแอดเดรส 5.188.86.177 มากที่สุด จำนวน 123,273 ครั้ง ต่อมาไอพีแอดเดรส 5.188.86.206 จำนวน 118,415 ครั้ง ไอพีแอดเดรส 195.22.127.83 จำนวน 115,786 ครั้ง ไอพีแอดเดรส 5.188.86.208 จำนวน 115,101 ครั้ง และ ไอพีแอดเดรส 5.188.86.207 จำนวน 115,071 ครั้ง ตามลำดับ โดยมีรายละเอียดตามแผนภูมิที่ 5 ดังนี้

แผนภูมิที่ 5 ไอพีแอดเดรสต้นทาง (Source IP)



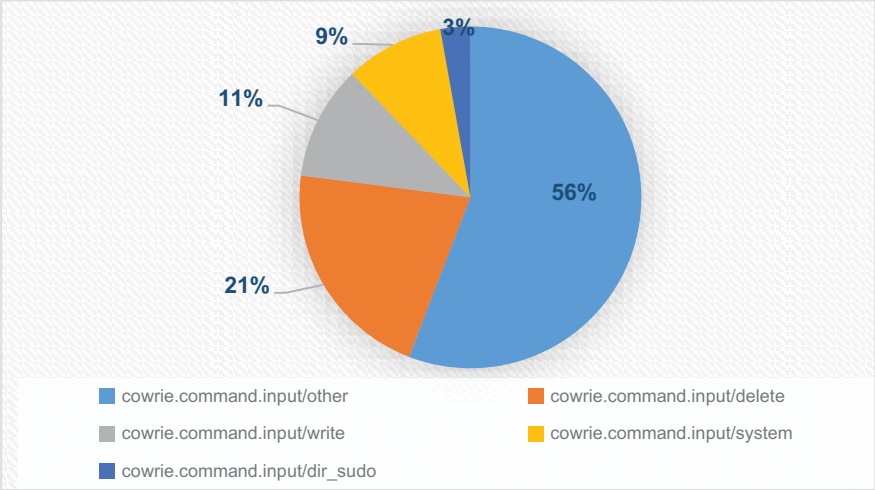
2. Event Distribution

Event Distribution คือ ประเภทกลุ่มข้อมูลที่แบ่งตามประเภทการใช้งานที่ Honeypots เก็บไว้ แบ่งประเภทเหตุการณ์ต่าง ๆ โดยจัดลำดับประเภทเข้ามาใช้งานมากที่สุด เรียกจากมากที่สุดไปหาน้อยที่สุด

ซึ่งแยกได้เป็นประเภท cowrie.direct-tcpip.request มีจำนวนมากที่สุด จำนวน 1,456,743 ครั้ง ต่อมาเป็นการร้องขอใช้งานประเภท cowrie.direct-tcpip.data เป็นลำดับต่อมา จำนวน 553,040 ครั้ง และประเภท cowrie.session.connect จำนวน 329,621 ครั้ง ซึ่งพิจารณารายละเอียดอื่น ๆ ได้จากแผนภูมิที่ 6 ดังนี้

ข่าย ผู้วิจัยจึงดำเนินการวิเคราะห์และประมวลผลข้อมูลประเภท 'cowrie.command.input โดยได้ประเภทข้อมูลเพิ่มเติม ตามแผนภูมิที่ 4 ดังนี้

แผนภูมิที่ 4 ข้อมูลประเภท cowrie.command.input



จากแผนภูมิที่ 4 ข้างต้น เมื่อนำข้อมูลกลุ่มประเภท cowrie.command.input มาวิเคราะห์เพื่อแยกประเภทของอินพุท จะแบ่งแยกประเภทของกลุ่มของข้อมูลได้อีก 5 ประเภท รายละเอียดดังนี้

1. 'cowrie.command.input/other' คือคำสั่งทั่วไปที่ไม่มีผลกระทบต่อการทำงานของระบบจัดกลุ่มได้มากที่สุด คิดเป็นร้อยละ 56

2. 'cowrie.command.input/delete' คือคำสั่งลบไฟล์ออกจากระบบ คิดเป็นร้อยละ 21

3. 'cowrie.command.input/write' คือคำสั่งเขียนไฟล์ คิดเป็นร้อยละ 11

4. 'cowrie.command.input/system' คือคำสั่งที่เป็นการทำงานของ OS คิดเป็นร้อยละ 9

5. 'cowrie.command.input/dir_sudo' คือคำสั่งที่กระทำโดย Super User คิดเป็นร้อยละ 3

เมื่อนำข้อมูล 'cowrie.command.input ไปวิเคราะห์และประมวลผล จะได้ประเภทข้อมูลเพิ่มอีก 5 ประเภท รวมแล้ว มีประเภทของกลุ่มของข้อมูลที่เข้ามาใช้งานเครื่องคอมพิวเตอร์แม่ข่ายทั้งสิ้น 23 ประเภทของกลุ่มของข้อมูล

3.2 ขั้นตอนการประมวลผล (Data Processing)

ขั้นตอนการประมวลผล (Data Processing) เป็นขั้นตอนที่นำข้อมูลที่ผ่านการเตรียมข้อมูลมาแล้วมาเข้ารูปแบบจำลองที่เลือกมาเพื่อให้เหมาะสมกับการวิจัย ซึ่งผู้วิจัยคัดเลือกแบบจำลองมา 2 แบบจำลองคือ แบบจำลอง Markov Chain (MC) และ แบบจำลอง Hidden Markov Model (HMM) โดยมี รายละเอียดดังนี้

1. แบบจำลองมาร์คอฟ (Markov Chain หรือ MC)

ในขั้นตอนการประมวลผลโดยใช้แบบจำลอง Markov Chain หลังจากที่ผ่านขั้นตอนการเตรียมข้อมูลมาแล้ว จะได้ข้อมูลที่เป็นลักษณะ List ของ State ในแต่ Session หลังจากนั้นทำการคำนวณหา Transition Matrix เพื่อหาความน่าจะเป็นในการเปลี่ยนของแต่ละ State จากข้อมูลที่ผ่านการคัดกรองและแบ่งกลุ่มมาแล้ว

2. แบบจำลองมาร์คอฟซ่อนเร้น (Hidden Markov Model หรือ HMM)

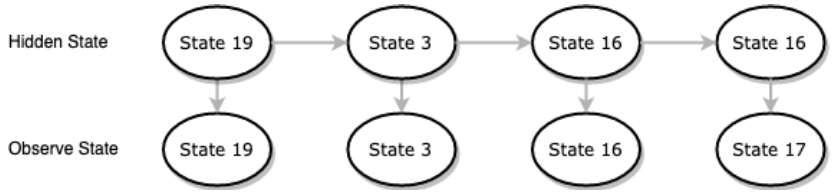
ขั้นตอนการประมวลผลในเบื้องต้นจะมีหลักการคล้ายแบบจำลองมาร์คอฟ คือเมื่อได้ข้อมูลที่เป็นลักษณะ List ของ State ในแต่ Session แล้วจะคำนวณหา Transition Matrix และ Emission Matrix เพื่อหาความน่าจะเป็นในการเปลี่ยนของแต่ละ State โดยใช้ข้อมูลเหล่านี้เป็นตัวแทนของ Observe State เพื่อหา Hidden State ที่มีความน่าจะเป็นมากที่สุด

3.3 ขั้นตอนการวัดประสิทธิภาพ (Data Evaluation)

แนวความคิดของแบบจำลอง Markov Chain รวมเข้าข้อมูลของสถานะเป็น Probabilistic Function เกิดเป็น “แบบจำลองมาร์คอฟซ่อนเร้น” (Hidden Markov Model หรือ HMM) ซึ่งคุณสมบัติของแบบจำลองนี้คือ ไม่จำเป็นต้องทราบสถานะที่เกิดขึ้นในกระบวนการก็ได้

ดังนั้น สรุปได้ว่าแบบจำลอง Hidden Markov Model เป็นโมเดลที่ใช้ความน่าจะเป็นมาคำนวณร่วมกับเซตของสถานะที่ซ่อนอยู่ (Hidden State) ที่ได้รับจากชุดของสถานะที่สังเกตได้ (Observe State) ดังนั้นเมื่อทราบความน่าจะเป็นของสถานะที่ซ่อนอยู่และสถานะที่สังเกตได้ จะกำหนดลำดับดีที่สุดที่จะเป็นไปได้ โดยลำดับที่มีความน่าจะเป็นสูงสุดและเลือกลำดับนั้นเป็นลำดับที่ดีที่สุดของสถานะที่ซ่อนอยู่ถัดไป

รูปภาพที่ 4 HMM State



จากรูปภาพที่ 4 หากพิจารณาสถานะด้านล่างซึ่งสถานะของตัวแปรที่ซ่อนอยู่ (Hidden State) ซึ่งสถานะดังกล่าวจะเป็นสถานะที่ไว้สำหรับทำนายลำดับของสถานะต่างๆ ที่มาจากกลุ่มข้อมูลตัวอย่าง สถานะตัวแปรที่สังเกตได้ (Observe State) คือสถานะของข้อมูลที่นำเข้าไปทดสอบผลการทำนายสถานะถัดไปว่าควรจะเป็นอะไรจากตัวอย่าง รูปภาพที่ 4 สถานะ 19,3,16,17 เป็นตัวอย่างของสถานะที่ส่งเข้าไปทดสอบ โดยมีลูกศรแสดงการเปลี่ยนจากสถานะที่ซ่อนอยู่ไปยังสถานะที่ซ่อนอยู่อื่น (Transition) หรือจากสถานะที่ซ่อนอยู่ไปยังตัวแปรที่สังเกตได้ (Emission) จะเห็นได้ว่าความน่าจะเป็นที่จะเกิดแต่ละสถานะขึ้นอยู่กับสถานะก่อนหน้า และสถานะที่ซ่อนอยู่ไปยังตัวแปรที่สังเกตได้เท่านั้น โดยไม่จำเป็นต้องอยู่ในสถานะก่อนหน้าอื่น ๆ เป็นต้น

3. ขั้นตอนวิธีการดำเนินการ

3.1 ขั้นตอนการเตรียมข้อมูล (Data Preparation)

ขั้นตอนการเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนแรกเพื่อจัดเตรียมข้อมูลที่ได้รับมาในรูปแบบ Honeypots ที่มีความหลากหลายของข้อมูล นำมาสกัดให้ได้ข้อมูลการเข้าใช้งานเครื่องคอมพิวเตอร์แม่ข่ายที่นำมาใช้ในการวิจัย โดยมีขั้นตอนการเตรียมข้อมูลดังนี้

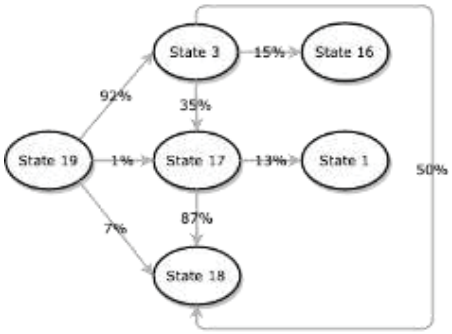
1. คัดแยกข้อมูลที่มีการร้องขอผ่าน HTTP (Hypertext Transfer Protocol) ยกตัวอย่างเช่น TCP data และ TCP request ออก เนื่องจากการวิจัยนี้เป็นการวิเคราะห์การเข้าใช้งานเครื่องคอมพิวเตอร์แม่ข่าย ซึ่งจะพิจารณาเฉพาะการร้องของในรูปแบบ SSH (Secure Shell) เท่านั้น

2. จัดกลุ่มของข้อมูลตามประเภทของข้อมูลตามเข้ามาใช้งานเครื่องคอมเพิวเตอร์แม่ข่ายได้ทั้งหมด 18 ประเภท รายละเอียดดังนี้

'cowrie.client.size', 'cowrie.client.var', 'cowrie.client.version', 'cowrie.command.failed', 'cowrie.command.input', 'cowrie.command.success', 'cowrie.direct-tcpip.data', 'cowrie.direct-tcpip.request', 'cowrie.log.closed', 'cowrie.log.open', 'cowrie.login.failed', 'cowrie.login.success', 'cowrie.session.closed', 'cowrie.session.connect', 'cowrie.session.file_download', ‘cowrie.session.file_upload', 'cowrie.session.input’, 'cowrie.client.fingerprint'

เมื่อจัดกลุ่มประเภทของข้อมูลที่ ทำการ SSH เข้ามาใช้งานเครื่องได้ 15 ประเภทแล้ว พบว่าข้อมูลประเภท 'cowrie.command.input ไม่อาจบ่งชี้ได้ว่าเป็นการเข้ามาใช้ทำงานคำสั่งอะไรในการเข้ามาใช้งานเครื่องคอมพิวเตอร์แม่

รูปภาพที่ 3 Markov Chain (MC)



หากพิจารณาจากรูปภาพที่ 3 เป็นตัวอย่างแบบจำลอง Markov Chain ของสถานะการเข้ามาใช้งานที่มี 3 สถานะคือ cowrie.session.connect, cowrie.client.version, cowrie.login.success, cowrie.session.closed, cowrie.login.failed, cowrie.client.size โดยตัวเลขที่อยู่ในเส้นบ่งบอกถึงความน่าจะเป็นในการเปลี่ยนสถานะ โดยกำหนดให้แต่ละสถานะเป็นดังนี้

State 19: cowrie.session.connect

State 3: cowrie.client.version

State 17: cowrie.login.success

State 18: cowrie.session.closed

State 16: cowrie.login.failed

State 1: cowrie.client.size

กำหนดค่าความน่าจะเป็นของการเปลี่ยนสถานะได้ ซึ่งแทนด้วย Matrix A

$$A = \{a_{ij}\} = \begin{bmatrix} 0.0 & 0.92 & 0.1 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.35 & 0.5 & 0.15 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.87 & 0.0 & 0.13 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

สมมติให้สถานะการเข้าใช้งานในหนึ่ง Session มีสถานะการใช้งานเป็นลำดับ ซึ่งหากจะคาดการณ์ความน่าจะเป็นที่สถานะอีก 4 สถานะข้างหน้าจะเป็นอย่างไร จะพิจารณาสถานะการใช้งานเป็น “S19-S3-S17-S18” มีค่าเท่าไหร่ ซึ่งจะกำหนดให้ลำดับข้อมูลการเปลี่ยนสถานะเหล่านี้แทนด้วย O โดยที่ O = {S19, S3, S17, S18 } ซึ่งตรงกับสถานะที่ t = 1, 2, 3, 4 จะได้ค่าความน่าจะเป็นที่สถานะจะเป็นไปตาม O คือ

$$P(O|Model) = P[S_{19}, S_3, S_{17}, S_{18} | Model]$$

$$= P[S_3].P[S_3| S_{19}].P[S_{17}| S_3].P[S_{18}| S_{17}]$$

$$= 1.(0.92)(0.35)(0.87)$$

$$= 2.801 \times 10^{-4}$$

2.4 แบบจำลองมาร์คอฟซ่อนเร้น (Hidden Markov Model หรือ HMM)

หากพิจารณาแบบจำลอง Markov Chain จะเกิดขึ้นจากการเปลี่ยนแปลงสถานะ ทำให้แบบจำลอง Markov Chain มีข้อจำกัดในการนำไปใช้ เนื่องจากต้องทราบเหตุการณ์ต่อไปก่อนจึงจะทำนายสถานะต่อไปได้ ดังนั้นหากนำ

การให้ระบบคอมพิวเตอร์เรียนรู้โดยใช้ข้อมูลที่มีอยู่ซึ่งจะแตกต่างจากระบบคอมพิวเตอร์ธรรมดาๆ ตรงที่ Machine learning จะมีข้อมูลเป็นข้อมูล (Data) และมีเอาต์พุตเข้าไปเพื่อทำนายอนาคต ของข้อมูลที่อินพุตเข้าไปประมวลผลจะมีเอาต์พุตออกมาเป็นตามที่ทำ Machine learning แบ่งออกเป็น 3 ประเภท ดังนี้

1. Supervised Learning ระบบคอมพิวเตอร์เรียนรู้โดยมีข้อมูลมาสอน โดยโปรแกรมจำแนกข้อมูลได้โดยจาก Training Data เพื่อให้เรียนรู้จากสิ่งที่ทำนายออกมา

2. Unsupervised Learning ระบบคอมพิวเตอร์จะเรียนรู้ได้ด้วยตัวเอง โดยจะนำข้อมูลที่มีอยู่หาความสัมพันธ์ข้อมูลเอง

3. Reinforcement learning ระบบคอมพิวเตอร์เรียนรู้ตามสภาพแวดล้อมโดยนำข้อมูลเปลี่ยนตามการคำนวณ

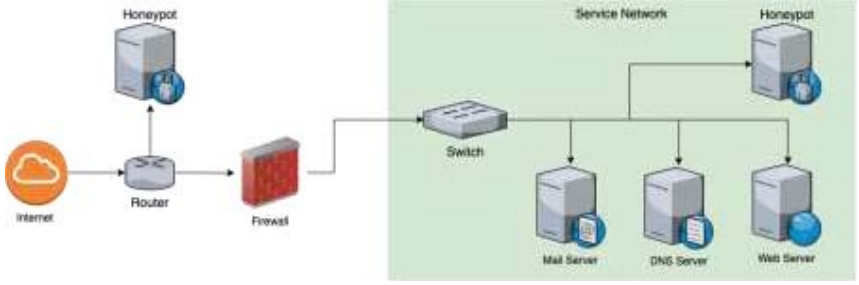
2.2 รูปแบบข้อมูลแบบ Honeypots

ปัจจุบันการทำงานบนเครือข่ายคอมพิวเตอร์มีมากมายหลากหลายรูปแบบ ยิ่งเมื่อการใช้งานบนเครือข่ายอินเทอร์เน็ตเป็นที่นิยมมากขึ้น ส่งผลให้มีการดำเนินธุรกรรมต่าง ๆ ผ่านเครือข่ายอินเทอร์เน็ตมากเช่นกัน อาจจะส่งผลให้เกิดมีช่องโหว่ หรือเกิดความเสี่ยงต่อการใช้งาน รวมถึงระบบเครือข่ายได้ นอกจากนี้อาจจะส่งผลให้เกิดการโจรกรรมข้อมูลต่าง ๆ ตลอดเวลา เช่น การแอบดักจับข้อมูล เพื่อนำไปใช้สำหรับผลประโยชน์ของตนเอง หรือต่อองค์กรนั้น ๆ ด้วยเหตุผลดังกล่าว จึงมีการพัฒนาการดักจับพฤติกรรมที่ประสงค์ร้ายกับเครื่องคอมพิวเตอร์แม่ข่าย โดยทำการล่อลวงผู้ที่ประสงค์ร้ายต่อเครื่องคอมพิวเตอร์แม่ข่ายให้เข้ามาใช้งานและเก็บข้อมูลเหล่านี้เอาไว้ ในรูปแบบแบบ Honeypots

Honeypots (ฮันนีพอต) คือกับดักที่ติดตั้งเพื่อดักจับ เบี่ยงเบน หรือบางครั้งอาจตอบโต้การพยายามใช้ระบบสารสนเทศโดยไม่ได้รับอนุญาต เป็นระบบข้อมูลไฟร์วอลล์ หรือเครื่องคอมพิวเตอร์แม่ข่ายฐานข้อมูลไฟล์ต่าง ๆ เหมือนระบบทั่วไป มีวัตถุประสงค์เพื่อเลียนแบบระบบที่ผู้ประสงค์ร้ายต้องการเจาะ และล่อลวงให้ผู้ประสงค์ร้ายเข้าไปใช้งานเครื่องคอมพิวเตอร์แม่ข่ายนั้นๆแต่จำกัดไม่ให้ผู้ประสงค์ร้ายเหล่านั้นเข้าถึงเครือข่ายได้ทั้งหมด ดังนั้นการติดตั้ง Honeypots จึงมักติดตั้งภายในไฟร์วอลล์ (Firewall) เพื่อให้ควบคุมและจัดการการเข้าถึงเครือข่ายภายใน และจำกัดการส่งข้อมูลออกภายนอกเครือข่ายได้

โดยทั่วไปมักจะนำ Honeypots มาใช้งานเพื่อศึกษาวิจัยพฤติกรรมของผู้โจมตีระบบต่าง ๆ เพื่อลดความเสี่ยงจากการถูกโจมตี เพื่อให้โจมตีที่ Honeypots แทนระบบจริง นอกจากนี้ Honeypots ยังช่วยตรวจสอบและแจ้งเตือนการบุกรุกระบบ Honeypots จะติดตั้งไว้ในเครือข่ายเดียวกับระบบจริงเพื่อเข้าถึงและรู้ระบบการทำงานได้ เพื่อช่วยให้การตรวจสอบได้ง่ายขึ้น

รูปภาพที่ 2 : Honeypots Deployment

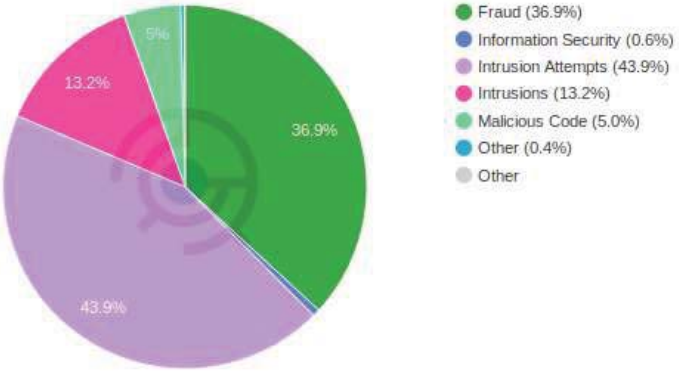


ที่มา: Honeypots (ฮันนีพอต) ระบบข้อมูลไฟร์วอล

2.3 แบบจำลองมาร์คอฟ (Markov Chain หรือ MC)

แบบจำลองมาร์คอฟ (Markov Chain หรือ MC) เป็นการพิจารณาระบบที่อธิบายถึงช่วงเวลาหนึ่งของกลุ่มสถานะที่แน่นอนจำนวน N สถานะ คือ S_1 ถึง S_N ที่แสดงให้เห็นดังรูปภาพที่ 3 โดยกำหนดให้ N = 6 และค่า a_{ij} เป็นค่าความน่าจะเป็นในการเปลี่ยนสถานะหนึ่งไปยังอีก สถานะหนึ่ง (โดยที่ i เป็นสถานะต้น และ j เป็นสถานะปลาย)

แผนภูมิที่ 3 : จำแนกตามประเภทภัยคุกคามทางอินเทอร์เน็ต



ที่มา: ThaiCERT (2562)

ด้วยเหตุผลดังกล่าว จะเห็นว่าการโจมตีทางไซเบอร์มีแนวโน้มเพิ่มมากขึ้น ดังนั้น ผู้วิจัยจึงสนใจศึกษา **"การตรวจจับพฤติกรรมที่ผิดปกติต่อเครื่องคอมพิวเตอร์แม่ข่าย โดยใช้แบบจำลองมาร์คอฟ และแบบจำลองมาร์คอฟซ่อนเร้น"** โดยใช้เทคโนโลยีการเรียนรู้ของเครื่องกล (Machine learning) เข้ามาประยุกต์เพื่อตรวจจับและหาวิธีป้องกันพฤติกรรมผิดปกติที่ประสงค์ร้ายเหล่านั้น ก่อนจะเกิดเหตุรุนแรงต่อไป

1.2 วัตถุประสงค์การวิจัย

1. เพื่อสร้างแบบจำลอง (Model) การวิเคราะห์รูปแบบพฤติกรรมที่ประสงค์ร้ายต่อการใช้งานเครื่องคอมพิวเตอร์แม่ข่าย โดยใช้วิธีการเรียนรู้ของเครื่องกล

2. เพื่อวิเคราะห์แบบจำลอง (Model) และเลือกแบบจำลองที่เหมาะสม ที่มีประสิทธิภาพในการดักจับพฤติกรรมที่ประสงค์ร้ายกับเครื่องคอมพิวเตอร์แม่ข่ายในการวิจัยครั้งนี้มากที่สุด

3. เพื่อทดสอบประสิทธิภาพของแบบจำลองพฤติกรรมที่ประสงค์ร้ายต่อการใช้งานเครื่องคอมพิวเตอร์แม่ข่าย

1.3 ขอบเขตของการศึกษา

1. ข้อมูลที่นำมาวิเคราะห์ในการวิจัยนี้ได้รับความอนุเคราะห์จากหน่วยงานราชการด้านความมั่นคงแห่งหนึ่งในประเทศไทย ซึ่งเป็นข้อมูลแบบประเภทข้อความ (Log file) ในรูปแบบ Honeypots

2. ข้อมูลในรูปแบบ Honeypots ที่นำมาวิเคราะห์ในการวิจัยนี้ เก็บข้อมูลระหว่างวันที่ 1 ธันวาคม 2560 ถึง 31 ธันวาคม 2561

3. ใช้แบบจำลองมาร์คอฟ และแบบจำลองมาร์คอฟซ่อนเร้นเข้ามาวิเคราะห์และประมวลผล เพื่อหาแบบจำลอง (Model) ที่เหมาะสมสำหรับข้อมูลในการวิจัยมากที่สุด

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สร้างแบบจำลอง (Model) การวิเคราะห์รูปแบบพฤติกรรมที่ผิดปกติต่อการใช้งานเครื่องคอมพิวเตอร์แม่ข่าย

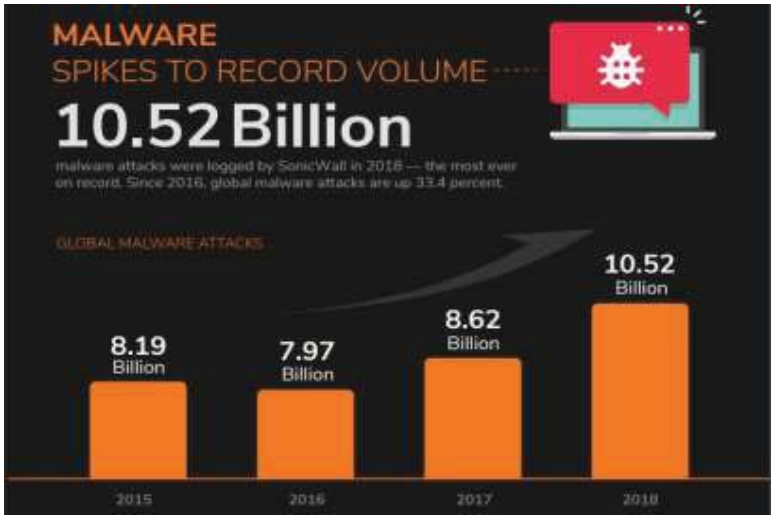
2. วิเคราะห์แบบจำลอง (Model) รวมถึงเลือกแบบจำลองที่เหมาะสมและมีประสิทธิภาพในการดักจับพฤติกรรมที่ผิดปกติกับเครื่องคอมพิวเตอร์แม่ข่าย ในการวิจัยครั้งนี้มากที่สุด

3. ทดสอบประสิทธิภาพของแบบจำลองรูปแบบพฤติกรรมที่ผิดปกติต่อการใช้งานเครื่องคอมพิวเตอร์แม่ข่าย

2. แนวคิด ทฤษฎี

2.1 การเรียนรู้ของเครื่องกล (Machine Learning)

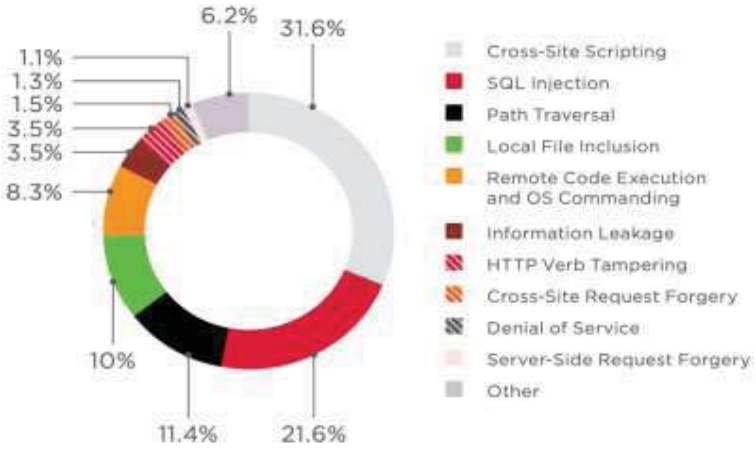
รูปภาพที่ 1 สถิติการถูกโจมตีจากทั่วโลก



ที่มา : SonicWall (2019)

จะเห็นได้ว่ามีการโจมตีจากผู้ประสงค์ร้ายเพิ่มขึ้น โดยมีแนวโน้มเพิ่มมากขึ้นทุกปี หากพิจารณาประเภทการโจมตีเว็บแอพพลิเคชันที่ผู้ประสงค์ร้ายนิยมใช้มากที่สุด 10 อันดับ จาก Positive Technologies ในปี 2018 พบว่ามีการโจมตีแบบ Cross-Site Script มากที่สุดคิดเป็นร้อยละ 36.1 อันดับสอง คือ SQL injection คิดเป็นร้อยละ 21.6 อันดับสามคือ Path Traversal ร้อยละ 11.4 อันดับที่สี่ คือ Local File inclusion คิดเป็นร้อยละ 10 อันดับที่ 5 คือ Remote Code Execution and OS commanding คิดเป็นร้อยละ 8.3 โดยมีการโจมตีแบบ Information Leakage, HTTP Verb Tampering, Cross-Site Request Forgery, Denial of Service และ Server-Side Request Forgery ตามลำดับ ดังนี้

แผนภูมิที่ 2 10 อันดับการโจมตีมากที่สุดทั่วโลก



ที่มา: Positive Technologies (2019)

สำหรับประเทศไทย Thaicert หรือ ศูนย์ประสานการรักษาความมั่นคงปลอดภัยระบบคอมพิวเตอร์ประเทศไทย (ไทยเซิร์ต) โดยกระทรวงวิทยาศาสตร์และเทคโนโลยี ระบุว่า พ.ศ. 2561 มีการแจ้งเหตุจากภัยคุกคามด้านอินเทอร์เน็ตของประเทศไทยสูงถึง 2,520 ครั้ง (เฉพาะที่มีการแจ้งภัยคุกคาม) หากพิจารณาแยกตามประเภทภัยคุกคาม พบว่าเป็นภัยคุกคามด้าน Intrusion Attempts มากที่สุดถึง 1,102 ครั้ง หรือคิดเป็นร้อยละ 43.9 รองลงมาคือ Fraud จำนวน 929 (ร้อยละ 36.9) และ Intrusions จำนวน 335 ครั้ง (ร้อยละ 13.2) ตามแผนภูมิที่ 3 ดังนี้

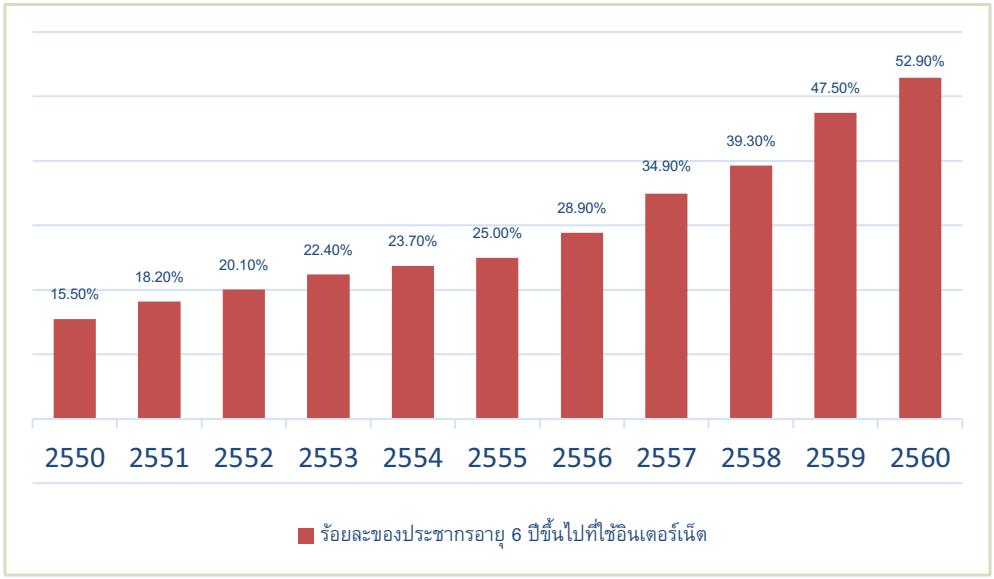
Keyword: Machine Learning, detecting, anomaly behavior, Markov Model, Hidden Markov Model

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการพัฒนาเครือข่ายอินเทอร์เน็ตได้ขยายพื้นที่ครอบคลุมได้ไกลและเป็นวงกว้างมากขึ้น ในทางตรงกันข้ามค่าบริการในการใช้งานอินเทอร์เน็ตกลับมีราคาถูกลงกว่าเดิมมาก การเข้าถึงอินเทอร์เน็ตจึงทำได้ง่ายและสะดวกมากขึ้น ส่งผลให้คนไทยใช้งานอินเทอร์เน็ตมากขึ้นในการดำเนินชีวิต ตั้งแต่การทำธุรกิจ การสื่อสาร การทำธุรกรรมด้านการเงินการธนาคาร การซื้อขายออนไลน์ การเรียนการสอน หรือการใช้อินเทอร์เน็ตเพื่อความบันเทิง เช่น ดูหนัง ฟังเพลง เล่นเกมส์ ซึ่งล้วนทำให้ชีวิตเราง่ายขึ้นเพียงปลายนิ้วคลิก ข้อมูลดังกล่าวสอดคล้องกับข้อมูลการใช้อินเทอร์เน็ตในประเทศไทย พ.ศ. 2550 – 2560 ของสำนักงานสถิติแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ซึ่งพบว่า พ.ศ. 2550 มีจำนวนผู้ใช้อินเทอร์เน็ตร้อยละ 15.50 และปี พ.ศ. 2560 มีผู้ใช้อินเทอร์เน็ตเพิ่มสูงขึ้นเป็นร้อยละ 52.90 นั่นคือภายใน 10 ปี อัตราการใช้งานอินเทอร์เน็ตเพิ่มขึ้นถึง 3.4 เท่า และมีแนวโน้มว่าจะมีผู้ใช้งานอินเทอร์เน็ตเพิ่มขึ้นตามแผนภูมิที่ 1 ดังนี้

แผนภูมิที่ 1 จำนวนผู้ใช้อินเทอร์เน็ตทั่วราชอาณาจักร พ.ศ. 2550 – 2560



ที่มา : สำนักงานสถิติแห่งชาติ (2560)

การใช้งานอินเทอร์เน็ตที่เพิ่มขึ้นอย่างรวดเร็ว ทำให้ธุรกิจเปลี่ยนรูปแบบการให้บริการ ผู้คนใช้งานบริการผ่านระบบเว็บแอพพลิเคชันมากขึ้น ส่งผลให้ผู้ประสงค์ร้ายปรับเปลี่ยนรูปแบบการก่อกวน บุกรุก โจมตี เพื่อสร้างความเสียหายต่อธุรกิจผ่านระบบเว็บแอพพลิเคชันมากขึ้นด้วยเช่นกัน

จากสถิติของ SonicWall Capture Threat Network ซึ่งมีเซ็นเซอร์มากกว่าหนึ่งล้านเครื่องทั่วโลก ในเดือน ที่บันทึกข้อมูลการโจมตีใน ค.ศ. 2018 พบว่าการถูกโจมตีจากผู้ประสงค์ร้ายมีแนวโน้มเพิ่มขึ้นทุกปี ตั้งแต่ ค.ศ. 2016 ที่มีการโจมตีจำนวน 7.97 พันล้านครั้ง ค.ศ. 2017 ถูกโจมตี 8.62 พันล้านครั้ง และ ค.ศ. 2018 ถูกโจมตี 10.52 พันล้านครั้ง และในช่วง ค.ศ. 2016 - 2018 มีการโจมตีเพิ่มขึ้นร้อยละ 33.4 รายละเอียดตามรูปภาพที่ 1

การตรวจจับพฤติกรรมที่ผิดปกติต่อเครื่องคอมพิวเตอร์แม่ข่าย โดยใช้แบบจำลองมาร์คอฟ และแบบจำลองมาร์คอฟซ่อนเร้น

Anomaly Behavior Detection on the Server by Markov Model and Hidden Markov Model

ณัฐ นักปราชญ์ [1*] และ ปราโมทย์ กั่วเจริญ[2]
[1,2] วิทยาศาสตร์มหาบัณฑิต (สาขาวิทยาการข้อมูล) คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
*E-mail: nukprach.n@gmail.com

บทคัดย่อ

ปัจจุบันเครือข่ายอินเทอร์เน็ตได้พัฒนาให้ขยายพื้นที่ครอบคลุมกว้างไกลขึ้น อีกทั้งค่าบริการก็มีราคาถูกลง ทาให้เข้าถึงเครือข่ายอินเทอร์เน็ตได้ง่ายและสะดวกขึ้น ส่งผลให้เกิดการโจมตีเครื่องคอมพิวเตอร์แม่ข่ายผ่านเครือข่ายอินเทอร์เน็ตมากขึ้น การประยุกต์ใช้เทคโนโลยีการเรียนรู้ของเครื่องกล (Machine learning) เพื่อตรวจจับพฤติกรรมการใช้งานเครื่องคอมพิวเตอร์แม่ข่าย อาจป้องกันการโจมตีจากผู้ประสงค์ร้ายได้ก่อนเกิดเหตุ [อันไม่พึงปรารถนา/ดุกคาม]

งานวิจัยนี้มีวัตถุประสงค์เพื่อ “ตรวจจับพฤติกรรมที่ผิดปกติต่อเครื่องคอมพิวเตอร์แม่ข่าย โดยใช้แบบจาลองมาร์คอฟและแบบจาลองมาร์คอฟซ่อนเร้น” โดยได้ความอนุเคราะห์ข้อมูลจากหน่วยงานราชการด้านความมั่นคงของไทย และมีขั้นตอนการดาเนินการดังนี้ 1) ขั้นตอนการเตรียมข้อมูล เพื่อคัดแยกและจัดกลุ่มข้อมูล 2) ขั้นตอนการประมวลผลโดยใช้แบบจาลองมาร์คอฟและแบบจาลองมาร์คอฟซ่อนเร้น เนื่องจากเหมาะสมกับรูปแบบของข้อมูลที่ใช้ในการศึกษา 3) ขั้นตอนการประเมินความถูกต้อง เพื่อประเมินประสิทธิภาพของแบบจาลองทั้งสอง

ผลการศึกษาสรุปว่า แบบจาลองมาร์คอฟและแบบจาลองมาร์คอฟซ่อนเร้น มีค่าความถูกต้องแม่นยาร้อยละ 79.29 และร้อยละ 90.40 ตามลาดับ หากเปรียบเทียบความแม่นยาสาหรับการศึกษาพฤติกรรมที่ผิดปกติจากข้อมูลที่ได้จากหน่วยงานราชการด้านความมั่นคง พบว่า แบบจาลองมาร์คอฟซ่อนเร้น เหมาะสมในการตรวจจับและป้องกันพฤติกรรมผิดปกติในการเข้าถึงเครื่องคอมพิวเตอร์แม่ข่ายมากที่สุด

คำสำคัญ: การเรียนรู้ของเครื่องกล, ตรวจจับ, พฤติกรรมที่ผิดปกติ, แบบจำลองมาร์คอฟ, แบบจำลองมาร์คอฟ ซ่อนเร้น

Abstract

The development of the internet network nowadays has expanded to offer wider coverage. The service fee of the internet network also becomes cheaper, providing easier and more readily network access. This leads to more server attacks via the internet network. Applying machine learning technology to detect anomaly behaviors in accessing servers could prevent such attacks by malicious hackers from undesirable or threatening outcomes.

This study aims to detect anomaly behaviors toward servers using Markov model and Hidden Markov model. The study process includes: 1) Data Preparation for data classification and clustering 2) Data Processing using Markov model and Hidden Markov model 3) Data Evaluation to evaluate the effectiveness of the models.

This study found that the detection accuracy using Markov model and Hidden Markov model is 79.29 percent and 90.40 percent, respectively. Therefore, we conclude that the most accurate method to detect anomaly behaviors toward server access is the Hidden Markov model.