

การเรียนรู้แบบเสริมแรงบนพื้นฐานของ CTM สำหรับการปรับปรุงการเลือก
โครงข่ายไร้สายแบบวิวิธพันธุ์



นายกิตติศักดิ์ กิตติเวทางค์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

CTM-based Reinforcement Learning for Improvement of Heterogeneous
Wireless Network Selection

Mr.Kittisak Kittiwaytang

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2010
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การเรียนรู้แบบเสริมแรงบนพื้นฐานของ CTM สำหรับการปรับปรุงการเลือกโครงข่ายไร้สายแบบวิธพันธุ

โดย

นายกิตติศักดิ์ กิตติเวทางค์


สาขาวิชา

วิศวกรรมไฟฟ้า

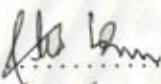
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ผู้ช่วยศาสตราจารย์ ดร.เชาวน์ดิศ อัสวกุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโท


..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศธีรฤกษ์)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(อาจารย์สุวิทย์ นาคพิระบุต)


..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.เชาวน์ดิศ อัสวกุล)


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


..... กรรมการภายนอกมหาวิทยาลัย
(ดร.จตุพร ชินรุ่งเรือง)

ศูนย์บริการพยาบาล
จุฬาลงกรณ์มหาวิทยาลัย

กิตติศักดิ์ กิตติเวทวงศ์ : การเรียนรู้แบบเสริมแรงบนพื้นฐานของ CTM สำหรับการปรับปรุงการเลือกโครงข่ายไร้สายแบบวิวิธพันธุ์. (CTM-based Reinforcement Learning for Improvement of Heterogeneous Wireless Network Selection)
 อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ. ดร. เซวาน์ดิศ อัสวกุล, 49 หน้า.


วิทยานิพนธ์ฉบับนี้ เสนอการพิจารณาสมรรถนะของการนำวิธีการเรียนรู้แบบเสริมแรง (reinforcement learning) มาประยุกต์ใช้เพื่อปรับปรุงกระบวนการตัดสินใจเลือกโครงข่ายเทคโนโลยีสื่อสารไร้สายที่เหมาะสมบนพื้นฐานของการจำลองการเคลื่อนที่แบบมณฑาทิมี่ชื่อว่า การจำลองการส่งผ่านเซลล์ (cell transmission model, CTM) ในโครงข่ายวิวิธพันธุ์ไร้สาย ในวิทยานิพนธ์นี้ได้เสนอผลการคำนวณเชิงตัวเลขบนโครงข่ายท้องถิ่น 1 มิติ โดยตั้งค่าให้ผู้ใช้มีตัวเลือกในการเชื่อมต่อกับโครงข่ายไร้สายแบบครอบคลุมพื้นที่ขนาดเล็กหรือโครงข่ายไมโครเซลล์ (micro cell network) และแบบครอบคลุมพื้นที่ขนาดใหญ่หรือโครงข่ายแมโครเซลล์ (macro cell network) ในแบบทิศทางตรงและแบบแอดฮ็อก (ad-hoc)


การจำลองแบบมณฑาทิมี่เสนอนี้กำหนดให้แหล่งกำเนิดกราฟฟิกจำลองมาจากการเคลื่อนที่แบบของไหลของยานพาหนะบนถนน ตรงกันข้ามกับแบบจำลองการเคลื่อนที่ของผู้ใช้แบบจุลภาค ซึ่งเป็นที่นิยมใช้ในงานวิจัยอดีตที่การวิเคราะห์จะซ่อนพฤติกรรมของผู้ใช้แต่ละคนไว้ ซึ่งส่งผลให้เกิดความซับซ้อนในการคำนวณ แบบจำลองดังกล่าวนี้มีความได้เปรียบในเรื่องความสามารถในการคำนวณเมื่อเปรียบเทียบกับแบบจำลองการเคลื่อนที่ของผู้ใช้แบบจุลภาค และสามารถอธิบายพฤติกรรมของผู้ใช้บนท้องถิ่นได้ผ่านทางสมการคณิตศาสตร์ แบบจำลองนี้จึงเหมาะสมกับการนำมาจำลองเพื่อหากลยุทธ์การเลือกโครงข่ายที่เหมาะสมแบบปรับตัวได้

การทดลองในวิทยานิพนธ์ที่เสนอกำหนดให้มีความแปรปรวนของระบบแตกต่างกัน เพื่อทดสอบการทำงานของวิธีการเรียนรู้แบบเสริมแรงในสภาพแวดล้อมที่แตกต่างกันออกไป ซึ่งผลการทดลองดังกล่าวนี้ แสดงให้เห็นว่าวิธีการเรียนรู้แบบเสริมแรงนั้นสามารถเลือกโครงข่ายที่เหมาะสมได้ในทุกผลการทดลองที่ได้นำเสนอ โดยไม่จำเป็นต้องใช้การคำนวณที่ซับซ้อนเหมือนกับอัลกอริทึมอื่นที่มีประสิทธิภาพใกล้เคียงกัน ทั้งยังสามารถปรับปรุงการเลือกโครงข่ายได้โดยอัตโนมัติหากสภาพแวดล้อมของระบบเปลี่ยนแปลงอย่างรวดเร็ว

ศูนย์วิทยทรัพยากร
 จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมไฟฟ้า
 สาขาวิชา วิศวกรรมไฟฟ้า
 ปีการศึกษา 2553

ลายมือชื่อนิสิต 

ลายมือชื่อที่ปรึกษาวิทยานิพนธ์หลัก 

5270656821 : MAJOR ELECTRICAL ENGINEERING

KEYWORDS: NETWORK SELECTION/ HETEROGENEOUS WIRELESS NETWORKS/
MACROSCOPIC MOBILITY MODEL.

KITTISAK KITTIWAYTANG : CTM-BASED REINFORCEMENT LEARNING
FOR IMPROVEMENT OF HETEROGENEOUS WIRELESS NETWORK SELEC-
TION . ADVISOR: ASST. PROF. CHAODIT ASWAKUL, Ph.D., 49 pp.

This thesis proposes the framework to find the optimal selection of heterogeneous wireless network. Reinforcement learning (RL) model has been used to find the best strategy to maximise the reward function expressed in terms of call blocking and call dropping probabilities. The reward-evaluation model is based on the well-established macroscopic cell transmission model (CTM). In this regard, numerical results have been given on a simple 1-dimensional road network settings with choices of moving users in making their connections via micro and macro cells in both direct and ad hoc modes.

With this approach, traffic sources have been modelled as the deterministic fluid flow of moving users travelling on their vehicles along a road. In contrast to the approaches with microscopic user mobility models often used in the past, the proposed framework has the advantage of computational efficiency and can be integrated well with RL in the herein developed optimisation framework.

The proposed framework has been considered RL-sensitivity in order to find the sensitivity of RL-parameters and evaluated in different scenarios which are the changes of bandwidth, stochastic incoming demands and unpredictable network problems. The results show that the RL algorithm can lead to the optimal solutions in all the tested scenarios with no large computational complexity as the other algorithms which have the efficiency nearly equal to RL. Furthermore, the RL algorithm can improve the network selection automatically even if the topology of the system has been changed immediately.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Department : .. Electrical Engineering
Field of study : .. Electrical Engineering
Academic year : .. 2010 ..

Student's signature
Advisor's signature



กิตติกรรมประกาศ

วิทยานิพนธ์ ฉบับนี้ สำเร็จ ลุล่วง ไป ได้ ด้วย ความ ช่วยเหลือ อย่าง ดี ยิ่ง จาก อาจารย์ ที่ปรึกษา วิทยานิพนธ์ ผศ.ดร.เชาวน์ดิศ อัครกุล ซึ่งได้ให้ความรู้และคำแนะนำอันมีค่ายิ่งต่อผู้วิจัย อีกทั้งตรวจทานงานวิทยานิพนธ์ฉบับนี้ด้วยดีเสมอมา ผู้วิจัยจึงขอกราบขอบพระคุณมา ณ ที่นี้

ขอขอบคุณกลุ่มวิจัยโครงข่าย (Network Research Group, NRG) ซึ่งดูแลโดย ผศ.ดร.เชาวน์ดิศ อัครกุล และ ผศ.ดร.ชัยเชษฐ์ สายวิจิตร ที่ได้จัดกิจกรรมเพื่อส่งเสริมการเรียนรู้และการทำงานของผู้วิจัย ให้มีประสิทธิภาพที่ดียิ่งขึ้น ทำให้งานวิทยานิพนธ์นี้สำเร็จได้อย่างสะดวกราบรื่น

ขอขอบคุณ โครงการ SP2 GE12 ของภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่สนับสนุนค่าเล่าเรียนและค่าใช้จ่ายรายเดือนตลอดการศึกษา

ขอขอบคุณ คุณปิติพงศ์ ชาญโลหะ ผู้คอยให้คำแนะนำ ข้อคิดเห็นต่าง ๆ อันเป็นประโยชน์ต่องาน วิทยานิพนธ์นี้ด้วยดีเสมอมา

ขอบคุณเพื่อน พี่น้อง นักวิจัย ทุกคน รวมถึงเจ้าหน้าที่ บุคลากรที่อยู่ในภาควิชาไฟฟ้า สาขา วิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ให้ความช่วยเหลือในเรื่องต่าง ๆ และเป็นกำลังใจที่ดียิ่ง ต่อผู้วิจัย

สุดท้ายนี้ขอขอบคุณครอบครัวของผู้วิจัย ซึ่งได้ให้การสนับสนุนและเป็นกำลังใจให้แก่ผู้วิจัยเสมอ มาจนสำเร็จการศึกษา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฌ
สารบัญภาพ	ญ
คำอธิบายสัญลักษณ์และคำย่อ	ฎ
1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิทยานิพนธ์	3
1.3 ขอบเขตวิทยานิพนธ์	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	4
1.5 ประมวลวิทยานิพนธ์	4
2 หลักการและทฤษฎี	5
2.1 หลักการเบื้องต้นของแบบจำลองการส่งผ่านเซลล์	6
2.2 ความสัมพันธ์พื้นฐานระหว่างความหนาแน่นและอัตราการไหล	7
2.3 พารามิเตอร์ของแบบจำลองการส่งผ่านเซลล์	7
2.3.1 ความสามารถในการส่ง	8
2.3.2 ความสามารถในการรับ	8
2.3.3 จำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ในการเชื่อมต่อ ของเซลล์แบบตามลำดับ	9
2.3.4 จำนวนยานพาหนะในเซลล์ตามกฎการอนุรักษ์การไหล	9
2.4 การเรียนรู้แบบเสริมแรง	10
2.4.1 องค์ประกอบของวิธีการเรียนรู้แบบเสริมแรง	11
2.4.2 วิธีการเรียนรู้แบบคิว	13
3 การตัดสินใจเลือกโครงข่ายโดยใช้วิธีการเรียนรู้แบบเสริมแรง	14
3.1 สมมติฐานของแบบจำลอง	14
3.2 แบบจำลองกราฟฟิกสำหรับกลยุทธ์การเลือกโครงข่าย	17
3.2.1 การคำนวณจำนวนผู้ใช้ที่ต้องการเชื่อมต่อกับโครงข่าย	18
3.2.2 การคำนวณจำนวนผู้ใช้ที่ถูกปฏิเสธการติดต่อกับโครงข่าย	18
3.2.3 การคำนวณแบนด์วิดท์ที่จัดสรรให้กับผู้ใช้	19
3.2.4 การคำนวณจำนวนผู้ใช้ที่สามารถติดต่อกับโครงข่ายได้สำเร็จ	19
3.2.5 การคำนวณจำนวนผู้ใช้ที่เสร็จสิ้นการติดต่อกับโครงข่าย	20
3.2.6 การคำนวณจำนวนผู้ใช้ที่อยู่ในสถานะทำงานในตอนท้ายของช่วงเวลา	21
3.2.7 การคำนวณแบนด์วิดท์ที่เหลืออยู่ก่อนช่วงเสร็จสิ้นการติดต่อกับโครงข่าย	21
3.3 การเรียนรู้แบบเสริมแรง	21

3.3.1	กำหนดนิยามของสถานะ	21
3.3.2	กำหนดนิยามของการกระทำ	22
3.3.3	วิธีการเรียนรู้แบบคิว	22
3.4	ตัวชี้วัดสมรรถนะของระบบ	23
3.4.1	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใช้โครงข่าย	23
3.4.2	ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัล (temporal difference, TD)	24
3.5	สรุป	24
4	ผลการทดสอบ	25
4.1	การเปรียบเทียบกับวิธีการหาผลเฉลยแบบตรงไปตรงมา (brute force search)	25
4.2	ผลกระทบที่เกิดจากการเปลี่ยนแปลงแบนด์วิดท์และความไม่แน่นอนของระบบจราจร	26
4.3	ผลการควบคุม การเลือก โครงข่าย ด้วยวิธีการเรียนรู้แบบเสริมแรง โดยมีการจำลองเหตุการณ์โครงข่ายล่ม	29
4.4	การทดสอบค่าความไว (sensitivity) ของวิธีการเรียนรู้แบบเสริมแรง	31
4.4.1	การทดลองปรับค่าอัตราการเรียนรู้ (α)	32
4.4.2	การทดลองปรับค่าอัตราลดทอนการเรียนรู้ (γ)	32
4.4.3	การทดลองปรับค่าอัตราสำรวจ (ϵ)	33
4.5	การเปรียบเทียบกับวิธีการจัดสรรแบนด์วิดท์โดยใช้กลยุทธ์ U	35
4.6	การเปรียบเทียบกับวิธีการจัดสรรแบนด์วิดท์แบบแบ่งน้ำหนักตามสัดส่วน (proportional fairness)	37
4.6.1	กรณีที่มีความจุของโครงข่ายแม่โครเซลล์สูง	37
4.6.2	กรณีที่มีความจุของโครงข่ายแม่โครเซลล์ต่ำ	39
4.6.3	กรณีที่มีการจำลองเหตุการณ์โครงข่ายล่ม	39
4.7	สรุป	42
5	บทสรุปและข้อเสนอแนะ	44
5.1	บทสรุป	44
5.2	ข้อเสนอแนะ	45
	รายการอ้างอิง	46
	ประวัติผู้เขียนวิทยานิพนธ์	49

สารบัญตาราง

ตารางที่ 2.1 ความแตกต่างของแบบจำลองจุลภาคและแบบจำลองมหภาค	5
---	---



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญญภาพ

รูปที่ 1.1	รูปแบบการแฮนด์ออฟทั้งสองประเภทในโครงข่ายไร้สาย	1
รูปที่ 2.1	การแบ่งถนนให้อยู่ในรูปเซลล์ของแบบจำลองการส่งผ่านเซลล์	6
รูปที่ 2.2	ความสัมพันธ์พื้นฐานระหว่างความหนาแน่นและอัตราการใช้	7
รูปที่ 2.3	การเชื่อมต่อเซลล์แบบตามลำดับของแบบจำลองการส่งผ่านเซลล์	8
รูปที่ 2.4	วงรอบของวิธีการเรียนรู้แบบเสริมแรง	10
รูปที่ 3.1	โครงข่ายไมโครเซลล์และแมโครเซลล์ในแบบจำลองการเชื่อมต่อเซลล์แบบตามลำดับของแบบจำลองการส่งผ่านเซลล์	14
รูปที่ 3.2	แผนภาพของตัวแปรขาเข้า ตัวแปรขาออก และตัวแปรสถานะในระบบ	16
รูปที่ 3.3	แบบจำลองลำดับการเกิดของเหตุการณ์ต่าง ๆ ในการเรียกเข้าใช้โครงข่ายไร้สายในแต่ละช่วงเวลา	17
รูปที่ 4.1	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ซึ่งลู่เข้าได้โดยใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์	26
รูปที่ 4.2	ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟซึ่งลู่เข้าได้โดยใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์	27
รูปที่ 4.3	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่โดยมีความจุแมโครเซลล์ต่ำและมีความไม่แน่นอนของระบบจราจร	28
รูปที่ 4.4	ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟโดยมีความจุแมโครเซลล์ต่ำและมีความไม่แน่นอนของระบบจราจร	28
รูปที่ 4.5	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่เมื่อมีการจำลองเหตุการณ์โครงข่ายล่ม	30
รูปที่ 4.6	ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟเมื่อมีการจำลองเหตุการณ์โครงข่ายล่ม	30
รูปที่ 4.7	ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรวมที่ค่า $\gamma = 0.001$ และ $\epsilon = 0.01$	32
รูปที่ 4.8	ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรวมที่ค่า $\alpha = 1$, และ $\epsilon = 0.01$	33
รูปที่ 4.9	ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรวมที่ค่า $\alpha = 1$ และ $\gamma = 0.65$	34
รูปที่ 4.10	ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่ค่า $\alpha = 1$ และ $\gamma = 0.65$	34
รูปที่ 4.11	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่จำนวนยานพาหนะต่าง ๆ	36
รูปที่ 4.12	ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่จำนวนยานพาหนะต่าง ๆ	36
รูปที่ 4.13	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่มีความจุแมโครเซลล์สูงเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน	38
รูปที่ 4.14	ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่มีความจุแมโครเซลล์สูงเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน	38
รูปที่ 4.15	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่มีความจุแมโครเซลล์ต่ำเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน	40
รูปที่ 4.16	ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่มีความจุแมโครเซลล์ต่ำเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน	40
รูปที่ 4.17	ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่เมื่อมีการจำลองเหตุการณ์โครงข่ายล่มเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน	41

รูปที่ 4.18 ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟเมื่อมีการจำลองเหตุการณ์โครงข่ายล่ม
 เทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน 42



ศูนย์วิทยทรัพยากร
 จุฬาลงกรณ์มหาวิทยาลัย

คำอธิบายสัญลักษณ์และคำย่อ

i	ดัชนีของเซลล์ของ CTM เป็นได้ตั้งแต่ $i = 1$ แทนเซลล์ที่ 1 ของ CTM จนถึง $i = N$ แทนเซลล์ที่ N ของ CTM
j	ดัชนีของสถานะของผู้ใช้ เป็นได้สองค่าดังนี้ 0 แทนสถานะของผู้ใช้ที่ไม่ได้ติดต่อกับโครงข่าย (สถานะไม่ทำงาน) และ 1 แทนสถานะของผู้ใช้ที่ติดต่อกับโครงข่าย (สถานะทำงาน)
k	ดัชนีของโครงข่ายไร้สาย เป็นได้ 3 ค่าดังนี้ 0 แทนโครงข่ายแมโครเซลล์, 1 แทนโครงข่ายไมโครเซลล์ที่ครอบคลุมเซลล์ที่ 1 ของ CTM และ N แทนโครงข่ายไมโครเซลล์ที่ครอบคลุมเซลล์ที่ N ของ CTM
$n_i(t)$	จำนวนรถยนต์ของเซลล์ที่ i ของ CTM ในช่วงเริ่มต้นช่วงเวลา t
$y_i(t)$	จำนวนรถยนต์ที่เคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ภายในช่วงเวลา t
$d_i^k(t)$	จำนวนผู้ใช้ที่อยู่ในสถานะทำงาน ในเซลล์ที่ i ของ CTM โดยติดต่อกับโครงข่าย k ในช่วงเวลา t ภายหลังจากที่มีผู้ใช้ในเซลล์ของ CTM เสร็จสิ้นการติดต่อโครงข่าย
w^k	น้ำหนักในการเลือกเข้าใช้โครงข่าย k
B_{max}^k	แบนด์วิทท์สูงสุดของโครงข่าย k
$BA_j^k(t)$	แบนด์วิทท์ของโครงข่าย k ที่จัดสรรไว้สำหรับผู้ใช้ในสถานะ j ในช่วงเวลา t
$BD^k(t)$	แบนด์วิทท์ของโครงข่าย k ที่เหลืออยู่ก่อนช่วงเสร็จสิ้นการติดต่อกับโครงข่ายในช่วงเวลา t
$r_{ij}(t)$	จำนวนผู้ใช้ที่อยู่ในสถานะ j ซึ่งต้องการเชื่อมต่อกับโครงข่ายต่าง ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t
$P_A(t)$	ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในช่วงเวลา t
$P_D(t)$	ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในแต่ละช่วงเวลา t
$A_{ij}^k(t)$	จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วสามารถติดต่อกับโครงข่าย k ได้สำเร็จในเซลล์ที่ i ของ CTM ในช่วงเวลา t
$D_i^k(t)$	จำนวนผู้ใช้ที่อยู่ในสถานะทำงานแล้วเปลี่ยนเป็นสถานะไม่ทำงานในเซลล์ที่ i ของ CTM โดยเสร็จสิ้นการติดต่อกับโครงข่าย k ในตอนท้ายช่วงเวลา t
$B_{ij}(t)$	จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วถูกปฏิเสธการติดต่อกับโครงข่ายต่าง ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t
$P_{B,j}$	ความน่าจะเป็นของผู้ใช้ในสถานะ j ที่ถูกปฏิเสธการติดต่อกับโครงข่าย

บทที่ 1

บทนำ

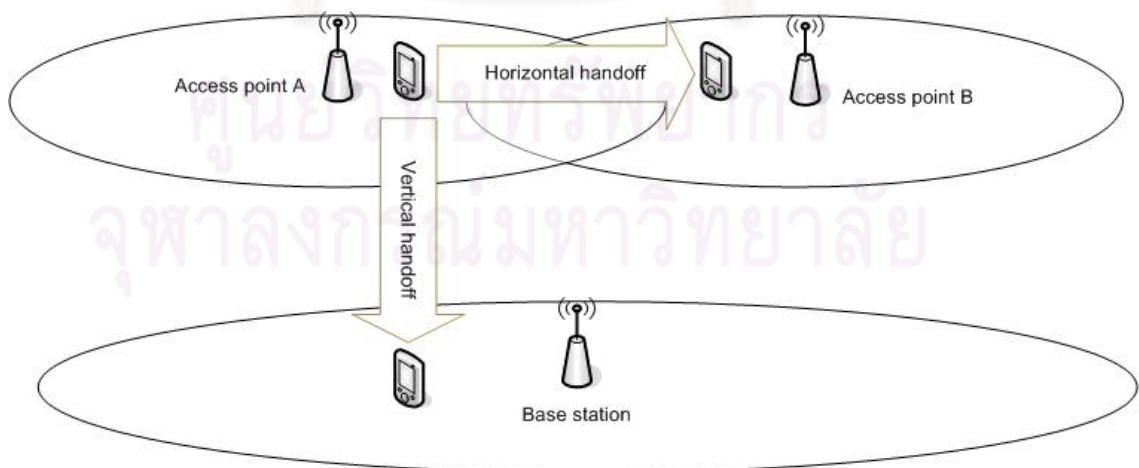
1.1 ความเป็นมาและความสำคัญของปัญหา

โครงข่ายเทคโนโลยีสื่อสารไร้สายในอนาคตจะถูกสร้างขึ้นให้สามารถตอบสนองความต้องการของผู้ใช้ทุกประเภท โดยมีการออกแบบและพัฒนาเทคโนโลยีสื่อสารไร้สายที่แตกต่างกันหลายชนิด เช่น โครงข่ายโทรศัพท์มือถือ (cellular networks), โครงข่ายไร้สายเขตมหานคร (IEEE 802.16-based wireless metropolitan area networks, WMANs), โครงข่ายไร้สายท้องถิ่น (IEEE 802.11-based wireless local area networks, WLANs) และโครงข่ายไร้สายส่วนบุคคล (IEEE 802.15-based wireless personal area networks, WPANs) เป็นต้น

เนื่องจากผู้ใช้แต่ละประเภทมีความต้องการที่จะส่งผ่านข้อมูลด้วยความเร็ว ขนาดของแบนด์วิดท์ และค่าใช้จ่ายในการส่งผ่านข้อมูลในแต่ละช่วงเวลามากน้อยแตกต่างกันตามสภาพแวดล้อมของระบบ จึงมีเทคโนโลยีสื่อสารไร้สายแตกต่างกันหลายประเภทซ้อนทับกันในพื้นที่หนึ่ง ๆ เรียกว่า "โครงข่ายวิวิธพันธุ์ไร้สาย" (heterogeneous wireless network)

อย่างไรก็ตาม ในปัจจุบันได้มีแนวคิดในการพัฒนาอุปกรณ์เพื่อรองรับเทคโนโลยีสื่อสารไร้สายได้หลาย ๆ เทคโนโลยี ทำให้ต้องมีการตัดสินใจเลือกว่าจะใช้ระบบใด (network selection) เมื่อต้องการเชื่อมต่อสัญญาณ โดยขึ้นอยู่กับกลยุทธ์ที่เลือกใช้

นอกจากนี้ยังมีการเปลี่ยนระบบในขณะที่เชื่อมต่อสัญญาณจากระบบที่ใช้อยู่ปัจจุบันไปเป็นอีกระบบหนึ่ง ซึ่งนิยามการเปลี่ยนระบบลักษณะนี้ว่า กระบวนการแฮนด์ออฟ (handoff) โดยกระบวนการแฮนด์ออฟนั้นสามารถแบ่งได้เป็น 2 ประเภท คือ การแฮนด์ออฟระหว่างโครงข่ายเทคโนโลยีเดียวกัน (horizontal handoff) กับ การแฮนด์ออฟระหว่างโครงข่ายต่างเทคโนโลยี (vertical handoff) ดังรูปที่ 1.1



รูปที่ 1.1: รูปแบบการแฮนด์ออฟทั้งสองประเภทในโครงข่ายไร้สาย

การแฮนด์ออฟระหว่างโครงข่ายเทคโนโลยีเดียวกัน คือการสับเปลี่ยนโครงข่ายของอุปกรณ์ไร้สาย

เมื่ออุปกรณ์เคลื่อนที่จากพื้นที่หนึ่งไปอีกพื้นที่หนึ่ง การแฮนด์ออฟระหว่างโครงข่ายต่างเทคโนโลยี คือ การสับเปลี่ยนโครงข่ายของอุปกรณ์ไร้สาย เมื่ออุปกรณ์อยู่ในพื้นที่ที่ถูกครอบคลุมด้วยเทคโนโลยีทั้งสองระบบ และอุปกรณ์ต้องการเปลี่ยนเทคโนโลยีที่ใช้ในปัจจุบัน ซึ่งการตัดสินใจว่าจะใช้เทคโนโลยีใดขึ้นอยู่กับพารามิเตอร์ที่ใช้ตัดสินใจ วิทยานิพนธ์นี้จะมุ่งพิจารณาการแฮนด์ออฟลักษณะนี้ โดยเลือกใช้พารามิเตอร์ที่เกี่ยวข้องกับค่าสถิติแบบจำลองการเคลื่อนที่ของผู้ใช้งาน

กระบวนการแฮนด์ออฟระหว่างโครงข่ายต่างเทคโนโลยีสามารถแบ่งได้เป็นสามขั้นตอน [1, 2] ประกอบด้วย การค้นหาระบบโครงข่าย (system discovery) การตัดสินใจแฮนด์ออฟระหว่างโครงข่ายต่างเทคโนโลยี (vertical handoff decision) และการดำเนินการแฮนด์ออฟระหว่างโครงข่ายต่างเทคโนโลยี (vertical handoff execution) ในช่วงที่อยู่ในขั้นตอนค้นหาระบบโครงข่าย อุปกรณ์สื่อสารไร้สายที่สามารถติดต่อได้หลายระบบจะตัดสินใจว่าระบบใดสามารถใช้งานได้ และพิจารณารูปแบบการบริการที่แต่ละระบบเสนอว่าควรใช้บริการใด ในช่วงการตัดสินใจแฮนด์ออฟระหว่างโครงข่าย อุปกรณ์จะตัดสินใจว่าควรเลือกใช้ระบบใดในกรณีที่มีหลายระบบสามารถใช้งานได้พร้อมกัน การตัดสินใจขึ้นอยู่กับพารามิเตอร์หลายอย่าง เช่น แบนด์วิดท์ที่สามารถใช้งานได้ ดีเลย์ จิตเตอร์ ค่าใช้จ่ายในการเข้าใช้ การใช้พลังงาน สถานะแบตเตอรี่ของอุปกรณ์ในขณะนั้น รวมไปถึงการตัดสินใจเลือกของผู้ใช้งาน ในช่วงการแฮนด์ออฟระหว่างโครงข่าย การเชื่อมต่อต้องเปลี่ยนเส้นทางการติดต่อในปัจจุบัน เป็นเส้นทางใหม่สำหรับระบบที่ตัดสินใจเลือกแฮนด์ออฟ ในขั้นตอนนี้รวมไปถึงกระบวนการลงทะเบียนและการยืนยันผู้ใช้งานด้วย

กระบวนการตัดสินใจว่าจะเลือกใช้เทคโนโลยีสื่อสารระบบใดขณะเริ่มการเชื่อมต่อโครงข่าย และการตัดสินใจเลือกระบบขณะแฮนด์ออฟระหว่างโครงข่ายต่างเทคโนโลยี เป็นปัจจัยสำคัญในแนวคิดการรวมเทคโนโลยีสื่อสารไร้สายต่าง ๆ บนอุปกรณ์สื่อสารพกพาเพียงอุปกรณ์เดียว ทำให้มีการทำมาตรฐานต่าง ๆ เกี่ยวกับแนวคิดนี้ในปัจจุบัน เช่น 3GPP [3] 3GPP2 [4] และ IEEE 802.21 [5] อย่างไรก็ตาม ในมาตรฐานเหล่านี้เปิดโอกาสให้ผู้ประยุกต์ใช้มาตรฐานเป็นผู้ตัดสินใจเองว่าจะเลือกใช้เทคโนโลยีใด

งานวิจัยที่ผ่านมาในอดีตได้มีงานวิจัยที่นำเสนอเกี่ยวกับอัลกอริทึมต่าง ๆ ที่ใช้ในการตัดสินใจเลือกใช้เทคโนโลยี ใน [6]-[10] ได้นำเสนอการแฮนด์ออฟระหว่างโครงข่ายเพื่อปรับปรุงคุณภาพของบริการ (quality of service) ด้วยวิธีต่าง ๆ ใน [11]-[13] ได้พิจารณาเกี่ยวกับการปรับเปลี่ยนวิธีการการแฮนด์ออฟระหว่างโครงข่ายให้มีความยืดหยุ่นมากขึ้นเพื่อลดปรากฏการณ์ปิงปอง (ping-pong effect) และใน [14] ได้พิจารณาการแฮนด์ออฟระหว่างโครงข่ายในระบบ 4G จากงานวิจัยต่าง ๆ ที่ได้ศึกษามาข้างต้นพบว่า งานวิจัยส่วนใหญ่มุ่งเน้นไปที่การปรับปรุงอัลกอริทึมเพื่อเพิ่มประสิทธิภาพในการแฮนด์ออฟ แต่งานวิจัยเหล่านั้นไม่ได้พิจารณาถึงการปรับตัวได้ของอัลกอริทึม มีผลทำให้อัลกอริทึมดังกล่าวไม่สามารถนำมาใช้ในทางปฏิบัติได้อย่างเหมาะสม กล่าวคือ ในทางปฏิบัติสภาพแวดล้อมของระบบมีการเปลี่ยนแปลงอยู่ตลอดเวลาและไม่สามารถทำนายการเปลี่ยนแปลงที่เกิดขึ้นได้ อัลกอริทึมที่สามารถนำมาประยุกต์ใช้ในทางปฏิบัติได้จึงจำเป็นต้องมีความยืดหยุ่นมากพอที่จะรองรับในทุกสภาพแวดล้อมที่เปลี่ยนแปลง

ในปัจจุบัน อัลกอริทึมหนึ่งที่ถูกนำมาประยุกต์ใช้อย่างแพร่หลายเพื่อรองรับความไม่แน่นอนของระบบคืออัลกอริทึมที่มีชื่อว่า การเรียนรู้แบบเสริมแรง [15] เป็นอัลกอริทึมที่เรียนรู้สภาพแวดล้อมของระบบด้วยตัวเอง ในการเรียนรู้แบบเสริมแรงนั้น กระบวนการทั้งหมดเกี่ยวข้องกับความสัมพันธ์ระหว่าง ตัวกระทำ (agent) และสภาพแวดล้อม (environment) โดยตัวกระทำจะพยายามเลือกการกระทำ (action) ใด ๆ ในชุดของการกระทำที่เป็นไปได้ทั้งหมดในสถานะ (state) ปัจจุบัน หลังจากนั้น การกระทำที่ถูกเลือก จะมีผลทำให้สภาพแวดล้อมเปลี่ยนแปลงไปและผู้เรียนก็จะได้รับผลรางวัล

(reward) ซึ่งขึ้นอยู่กับว่าการกระทำดังกล่าวมีผลให้สภาพแวดล้อมเปลี่ยนแปลงไปในทางใด

ใน [16]-[19] ได้นำอัลกอริทึมนี้มาประยุกต์ใช้ในการควบคุมการตัดสินใจเลือกเทคโนโลยีบนโครงข่ายวิวิธพันธุ์ไร้สาย อย่างไรก็ตาม งานวิจัยที่ได้กล่าวมาข้างต้นได้พิจารณาการเคลื่อนที่แบบจุลภาค (microscopic) ซึ่งมีข้อจำกัดหลายอย่างในการประยุกต์ใช้ในทางปฏิบัติ รวมไปถึงมีความซับซ้อนในการสร้างแบบจำลองและการคำนวณสูงมาก จึงทำให้ไม่เหมาะที่จะนำมาใช้ตัดสินใจเลือกโครงข่ายในทางปฏิบัติ ใน [20] ได้เสนอแบบจำลองการส่งผ่านเซลล์ (cell transmission model, CTM) [21] ซึ่งสามารถคำนวณจำนวนผู้ใช้ในถนนในช่วงเวลาใด ๆ ได้และมีการเปรียบเทียบความถูกต้องกับข้อมูลจริง [22] มาเป็นแบบจำลองการเคลื่อนที่ของผู้ใช้บนถนน ที่มีโครงข่ายไร้สายหลายประเภท เพื่อคำนวณสมรรถนะของระบบในกลยุทธ์การเลือกโครงข่ายต่าง ๆ กัน งานวิจัยดังกล่าวพิจารณากลยุทธ์ในการตัดสินใจเลือกโครงข่าย 2 กลยุทธ์ คือ กลยุทธ์เลือกโครงข่ายไมโครเซลล์ (กลยุทธ์ MI) และกลยุทธ์ผู้ใช้กำหนดเอง (กลยุทธ์ U) [20] โดยค่านำหนักที่ได้จากกลยุทธ์ทั้งสองกลยุทธ์จะเป็นค่าคงที่ ไม่ขึ้นกับเวลาและตำแหน่ง ซึ่งข้อจำกัดนี้มีผลทำให้ไม่สามารถนำกลยุทธ์ทั้งสองกลยุทธ์มาประยุกต์ใช้ในทางปฏิบัติได้ วิทยานิพนธ์ฉบับนี้ได้นำวิธีการเรียนรู้แบบเสริมแรง ซึ่งเป็นกลยุทธ์ที่มีการปรับเปลี่ยนค่านำหนักทุกช่วงเวลา มาประยุกต์ใช้เพื่อหากลยุทธ์ในการตัดสินใจเลือกโครงข่ายเทคโนโลยีสื่อสารไร้สายที่เหมาะสม โดยรายละเอียดของแบบจำลองการส่งผ่านเซลล์จะกล่าวโดยละเอียดในบทที่ 2

1.2 วัตถุประสงค์ของงานวิทยานิพนธ์

เพื่อพิจารณาสมรรถนะของการนำวิธีการเรียนรู้แบบเสริมแรง (reinforcement learning) มาประยุกต์ใช้เพื่อปรับปรุงกระบวนการตัดสินใจเลือกโครงข่ายเทคโนโลยีสื่อสารไร้สายที่เหมาะสมบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ในโครงข่ายวิวิธพันธุ์ไร้สายซึ่งประกอบไปด้วยโครงข่ายสองชนิดคือ แบบครอบคลุมพื้นที่ขนาดเล็ก (micro cell) เช่น IEEE 802.11 (Wi-Fi) เป็นต้น และแบบครอบคลุมพื้นที่ขนาดใหญ่ (macro cell) เช่น ระบบสื่อสารเซลลูลาร์ เป็นต้น

1.3 ขอบเขตวิทยานิพนธ์

1. พิจารณาโครงข่ายไร้สายแบบวิวิธพันธุ์ที่มีโครงข่ายสองชนิดในระบบคือโครงข่ายไมโครเซลล์ 1 โครงข่ายที่ครอบคลุมทั้งระบบและโครงข่ายไมโครเซลล์ 2 โครงข่าย โดยที่แต่ละโครงข่ายไมโครเซลล์ครอบคลุม 1 เซลล์ของ CTM
2. ประเมินสมรรถนะของระบบจากการคำนวณความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟในสมการรูปทั่วไปสำหรับโครงข่ายไร้สายแบบวิวิธพันธุ์ที่มีโครงข่ายสองชนิด
3. แบบจำลองการเคลื่อนที่ผู้ใช้แบบมหภาคที่ใช้คือแบบจำลอง CTM ที่จำลองถนนเส้นเดียวที่เชื่อมต่อเซลล์แบบตามลำดับ (cascading) โดยไม่คำนึงถึงประเภทของผู้ใช้ที่แตกต่างกัน
4. วิธีที่ใช้ในการเลือกโครงข่ายที่เหมาะสมคือ วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานแบบจำลองการส่งผ่านเซลล์ (CTM-based RL algorithm) โดยประเภทของการเรียนรู้แบบเสริมแรงที่ใช้คือ วิธีการเรียนรู้แบบคิว

5. ศึกษาค่าความไว (sensitivity) ของวิธีการเรียนรู้แบบเสริมแรงเมื่อมีการปรับค่าพารามิเตอร์ของวิธีการเรียนรู้แบบเสริมแรง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

สามารถนำวิธีการเรียนรู้แบบเสริมแรงมาประยุกต์ใช้เพื่อปรับปรุงกระบวนการตัดสินใจเลือกโครงข่ายเทคโนโลยีสื่อสารไร้สายที่เหมาะสมบนพื้นฐานของการจำลองการเคลื่อนที่แบบมหภาคในโครงข่ายวิวิธพันธุ์ไร้สายได้

1.5 ประมวลวิทยานิพนธ์

บทที่ 1 บทนำ: กล่าวถึง ลักษณะทั่วไปของโครงข่ายไร้สายแบบวิวิธพันธุ์ ปัญหาต่าง ๆ และการแก้ไขปัญหาที่เกิดขึ้นจากงานวิจัยในอดีต รวมไปถึงการเลือกใช้แบบจำลองที่เหมาะสมต่อการปรับปรุงสมรรถนะของโครงข่าย

บทที่ 2 หลักการและทฤษฎี: กล่าวถึง หลักการของแบบจำลองการส่งผ่านเซลล์ หลักการของวิธีการเรียนรู้แบบเสริมแรง และวิธีการคำนวณพารามิเตอร์ต่าง ๆ ที่เกี่ยวข้อง

บทที่ 3 การตัดสินใจเลือกโครงข่ายโดยใช้วิธีการเรียนรู้แบบเสริมแรง: กล่าวถึง การนำวิธีการเรียนรู้แบบเสริมแรงมาประยุกต์ใช้ในกระบวนการตัดสินใจเลือกโครงข่ายไร้สายแบบวิวิธพันธุ์บนแบบจำลองการส่งผ่านเซลล์

บทที่ 4 ผลการทดสอบ: แสดงผลการทดสอบการเลือกโครงข่ายไร้สายแบบวิวิธพันธุ์โดยใช้วิธีการเรียนรู้แบบเสริมแรง

บทที่ 5 บทสรุปและข้อเสนอแนะ: สรุปงานวิจัยทั้งหมดในวิทยานิพนธ์ฉบับนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

หลักการและทฤษฎี

ในการจำลองผู้ใช้งานจราจรบนท้องถนน สามารถแบ่งออกได้เป็น 2 ประเภท คือ 1) แบบจำลองจุลภาค (microscopic model) 2) แบบจำลองมหภาค (macroscopic model) ซึ่งมีลักษณะดังต่อไปนี้

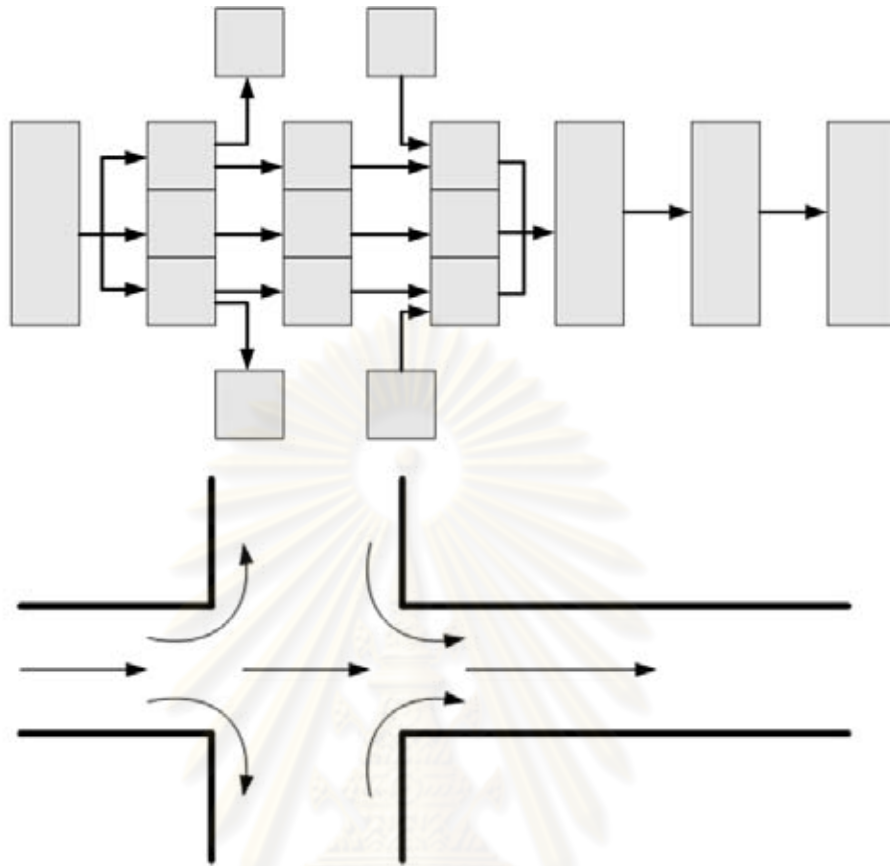
1. แบบจำลองจุลภาค จะพิจารณาผู้ใช้งานโครงข่ายจราจรแยกเป็นยานพาหนะแต่ละคัน ซึ่งผู้ขับที่มีการตัดสินใจแตกต่างกันไป และการเคลื่อนไหวของยานพาหนะมีความสัมพันธ์กัน ตัวอย่างแบบจำลองจุลภาค เช่น แบบจำลองการขับรถตามกัน (car following model) [26]
2. แบบจำลองมหภาค จะพิจารณาผู้ใช้งานโครงข่ายจราจรเป็นแบบองค์รวม กล่าวคือ จะพิจารณาเป็นกลุ่มของยานพาหนะ และกระแสนการเคลื่อนที่ของกลุ่มยานพาหนะนั้นแทน เช่น แบบจำลอง LWR [27, 28], แบบจำลองการส่งผ่านเซลล์ [21]

ตารางที่ 2.1: ความแตกต่างของแบบจำลองจุลภาคและแบบจำลองมหภาค

ข้อเปรียบเทียบ	แบบจำลองจุลภาค	แบบจำลองมหภาค
การหาค่าพารามิเตอร์ของแบบจำลอง	ยาก	ง่าย
ผลเฉลยทั่วไป	ไม่มี	มี
เวลาที่ใช้ในการประมวลผล	มาก	น้อย
ความถูกต้องในการจำลอง	มาก	น้อย
ความสามารถในการสร้างรูปทั่วไป	ทำได้ง่าย	ต้องอาศัยการทำงานเล็กน้อย

ตารางที่ 2.1 [22] แสดงความแตกต่างของแบบจำลองจุลภาคและแบบจำลองมหภาค เมื่อพิจารณาในด้านของเวลาที่ใช้ในการประมวลผล พบว่าแบบจำลองมหภาคใช้เวลาน้อยกว่าแบบจำลองจุลภาค ซึ่งเป็นปัจจัยสำคัญสำหรับการสร้างแบบจำลองการพิจารณาตัดสินใจเลือกโครงข่ายไร้สาย เนื่องจากการเคลื่อนที่ของยานพาหนะบนท้องถนนมีการเปลี่ยนแปลงตลอดเวลาอย่างรวดเร็ว ทำให้ในการตัดสินใจเลือกโครงข่ายของผู้ใช้แต่ละครั้งจำเป็นต้องพิจารณาจากสถานการณ์ในปัจจุบัน วิทยานิพนธ์นี้จึงได้นำเอาแบบจำลองการส่งผ่านเซลล์ซึ่งเป็นแบบจำลองมหภาคประเภทหนึ่ง มาเป็นเครื่องมือในการคำนวณจำนวนยานพาหนะบนถนนในแต่ละช่วงเพื่อใช้พิจารณาการตัดสินใจเลือกโครงข่าย

ในพื้นที่หนึ่ง ๆ มีโครงข่ายไร้สายอยู่ได้มากกว่าหนึ่งเทคโนโลยี แต่ละเทคโนโลยีจะมีคุณสมบัติและสถานะแวดล้อมที่แตกต่างกัน ซึ่งทำให้การเลือกใช้โครงข่ายไร้สายต่างเทคโนโลยีกันมีผลต่อสมรรถนะของระบบ การปรับปรุงสมรรถนะของระบบสื่อสารไร้สายจึงต้องมีการพิจารณาการตัดสินใจเลือกใช้โครงข่าย แต่เนื่องจากในทางปฏิบัติสภาพแวดล้อมของระบบมีการเปลี่ยนแปลงอยู่ตลอดเวลาและไม่สามารถทำนายการเปลี่ยนแปลงที่เกิดขึ้นได้ ในวิทยานิพนธ์ฉบับนี้จึงได้นำเสนอการนำวิธีการเรียนรู้แบบเสริมแรง ซึ่งเป็นอัลกอริทึมที่มีความยืดหยุ่นมากเพียงพอที่จะรองรับการเปลี่ยนแปลงของสภาพแวดล้อมมาเป็นกลยุทธ์ในการตัดสินใจเลือกใช้โครงข่าย ในบทนี้จะกล่าวถึงรายละเอียดของวิธีการเรียนรู้แบบเสริมแรงต่อไป



รูปที่ 2.1: การแบ่งถนนให้อยู่ในรูปเซลล์ของแบบจำลองการส่งผ่านเซลล์

2.1 หลักการเบื้องต้นของแบบจำลองการส่งผ่านเซลล์

แบบจำลองการส่งผ่านเซลล์พัฒนามาจากแบบจำลองที่เป็นที่รู้จักโดยทั่วไป คือ แบบจำลอง LWR ซึ่งเป็นแบบจำลองที่ใช้พื้นฐานกฎอนุรักษ์การไหล (flow conservation law) โดยแบบจำลองการส่งผ่านเซลล์ได้ปรับปรุงให้มีความซับซ้อนในการคำนวณน้อยกว่าแบบจำลอง LWR จึงทำให้สามารถนำไปประยุกต์ใช้และปรับปรุงต่อไปได้ง่ายกว่า

แบบจำลองการส่งผ่านเซลล์จะพิจารณาถนนเป็นส่วนย่อย ๆ โดยเรียกว่า เซลล์ (cell) หรือ เซลล์ของ CTM (CTM cell) ดังนั้นรูปแบบถนนจึงมีลักษณะเป็นเซลล์มาเชื่อมต่อกัน ดังแสดงตัวอย่างในรูปที่ 2.1 และมีการเปลี่ยนแปลงสถานะของระบบในแกนเวลาแบบวิยุต (discrete) ซึ่งแบ่งเป็น ช่องเวลา (time slot)

หลักการเบื้องต้นของการแบ่งถนนเป็นเซลล์ในแบบจำลองการส่งผ่านเซลล์ [21] มีดังนี้

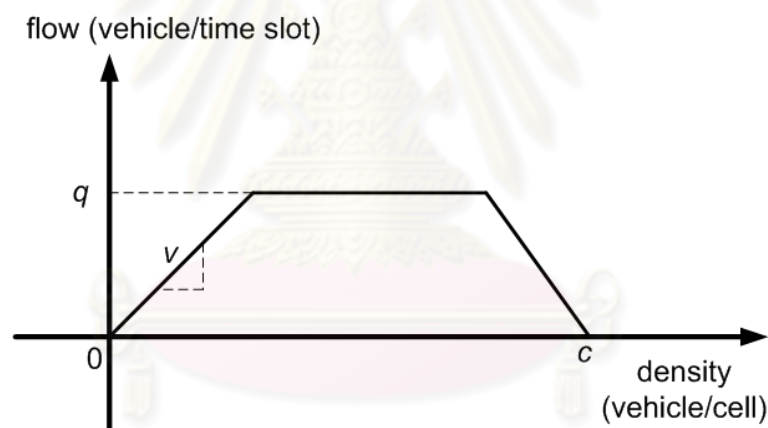
1. หากมีเซลล์ของ CTM 3 เซลล์อยู่ติดกันและยานพาหนะต้องการเคลื่อนที่ในทิศทางจากเซลล์ที่ 1 ไปยังเซลล์ที่ 3 แล้ว ยานพาหนะไม่สามารถเคลื่อนที่จากเซลล์ที่ 1 ไปถึงเซลล์ที่ 3 ภายในเวลา 1 ช่องเวลาได้ กล่าวคือ ความยาวของเซลล์จะต้องไม่สั้นเกินจนทำให้ยานพาหนะสามารถกระโดดข้ามผ่านเซลล์ได้ใน 1 ช่องเวลา ดังนั้น ความยาวของเซลล์ จึงคำนวณจาก ผลคูณระหว่างความเร็วเฉลี่ยเมื่อไม่มีการติดขัดของยานพาหนะ (free flow speed) กับ ระยะเวลา 1 ช่องเวลา

2. ระยะเวลา 1 ช่องเวลา พิจารณาตามอัตราการเปลี่ยนแปลงสภาพจราจร เช่น ในพื้นที่ที่มีการเปลี่ยนแปลงรอบเวลาสัญญาณไฟจราจรที่ค่อนข้างเร็ว ก็ควรกำหนดให้ค่า ระยะเวลา 1 ช่องเวลามีค่าน้อย ๆ เพื่อที่แบบจำลองจะได้เปลี่ยนแปลงทันตามการเปลี่ยนแปลงสภาพจราจรที่รวดเร็ว

การแบ่งถนนออกเป็นเซลล์นั้นในรูปทั่วไปแล้ว ความยาวเซลล์ของ CTM แต่ละเซลล์ไม่จำเป็นต้องมีค่าเท่ากัน [29] อย่างไรก็ตามสำหรับวิทยานิพนธ์นี้จะกำหนดให้เซลล์ในแบบจำลองการส่งผ่านเซลล์มีความยาวที่เท่ากันตามข้อกำหนดเบื้องต้นที่ได้มีนำเสนอไว้โดยผู้คิดค้นแบบจำลองนี้ [21] เมื่อสามารถกำหนดความยาวของเซลล์ได้ ขั้นตอนต่อไปคือการพิจารณาลักษณะการเชื่อมต่อของเซลล์

2.2 ความสัมพันธ์พื้นฐานระหว่างความหนาแน่นและอัตราการไหล

ก่อนจะพิจารณาลักษณะการเชื่อมต่อของเซลล์ ในเบื้องต้นจำเป็นต้องเข้าใจความสัมพันธ์พื้นฐาน (fundamental diagram) ระหว่างความหนาแน่น (density) และอัตราการไหล (flow), ความเร็วเมื่อไม่เกิดการติดขัด (free flow speed) (v) และ ความหนาแน่นของยานพาหนะบนถนน (density) ซึ่งเป็นพื้นฐานสำหรับสมการคำนวณค่าอัตราการไหล ดังรูปที่ 2.2 [25]



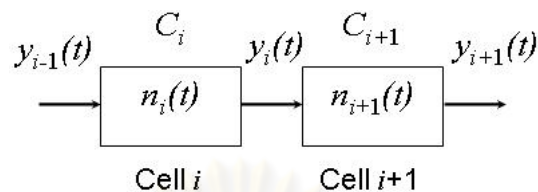
รูปที่ 2.2: ความสัมพันธ์พื้นฐานระหว่างความหนาแน่นและอัตราการไหล

จากรูปที่ 2.2 ในช่วงเริ่มต้นที่ความหนาแน่นของยานพาหนะบนถนนมีค่าน้อย ยานพาหนะจะเคลื่อนที่ด้วยความเร็วเมื่อไม่เกิดการติดขัด (v) (free flow speed) ทำให้ความสัมพันธ์ของอัตราการไหลและความหนาแน่นของยานพาหนะบนถนนเป็นแบบเชิงเส้นที่มีความชันเป็น v ในช่วงถัดมาอัตราการไหล q จะเริ่มเข้าสู่ช่วงอัตราการไหลคงที่ที่อัตราการไหลสูงสุด จากนั้นถ้าปริมาณยานพาหนะหนาแน่นขึ้นอีกจะเข้าสู่ช่วงสุดท้ายที่ยานพาหนะจะเริ่มเกิดการติดขัดทำให้อัตราการไหลจะลดลง จนเมื่อความหนาแน่นของยานพาหนะมีค่าเท่ากับความจุของถนน (c) อัตราการไหลจะมีค่าเท่ากับ 0 นั่นคือเกิดการติดขัดจนเคลื่อนที่ไม่ได้เลย

2.3 พารามิเตอร์ของแบบจำลองการส่งผ่านเซลล์

ในวิทยานิพนธ์ฉบับนี้พิจารณารูปแบบการเชื่อมต่อของแบบจำลองการส่งผ่านเซลล์เฉพาะการเชื่อมต่อแบบตามลำดับ (cascading) [22] โดยกำหนดให้ถนนช่วงที่พิจารณาประกอบด้วยเซลล์ของ

CTM ที่มีลำดับเรียงจากต้นทางไปปลายทางโดยเรียงตามลำดับเซลล์ที่ 1, 2, 3, ... ดังนั้นในดัชนีของเซลล์ที่เชื่อมต่อแบบตามลำดับในรูปที่ 2.3 จึงสามารถพิจารณาเซลล์ต้นทาง i ไปยังเซลล์ปลายทาง $i + 1$



รูปที่ 2.3: การเชื่อมต่อเซลล์แบบตามลำดับของแบบจำลองการส่งผ่านเซลล์

เป้าหมายในการใช้แบบจำลองการส่งผ่านเซลล์ในวิทยานิพนธ์ฉบับนี้ คือ การหาจำนวนยานพาหนะในเซลล์ของ CTM แต่ละเซลล์ในช่องเวลาต่าง ๆ ซึ่งแทนด้วยตัวแปร $n_i(t)$ โดยตัวแปรนี้หมายถึงจำนวนยานพาหนะที่เซลล์ i ในช่องเวลา t และการหาจำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ในช่องเวลาต่าง ๆ ซึ่งแทนด้วยตัวแปร $y_i(t)$ โดยตัวแปรนี้หมายถึงจำนวนยานพาหนะที่เคลื่อนออกจากเซลล์ต้นทาง i ไปยังเซลล์ปลายทาง $i + 1$ ในช่องเวลา t

ในการคำนวณหาตัวแปรข้างต้นทั้ง 2 ค่านี้ จำเป็นต้องทราบพารามิเตอร์พื้นฐาน 2 ค่า ได้แก่ ความสามารถในการส่ง (sending capability) ยานพาหนะของเซลล์ต้นทาง และความสามารถในการรับ (receiving capability) ยานพาหนะของเซลล์ปลายทาง ซึ่ง 2 ค่านี้เปรียบเทียบกับ อุปสงค์และอุปทานในการส่งรับยานพาหนะ

2.3.1 ความสามารถในการส่ง

ความสามารถในการส่ง คือ ความต้องการของเซลล์ต้นทางที่จะเคลื่อนยานพาหนะไปสู่เซลล์ปลายทาง โดยขึ้นอยู่กับจำนวนยานพาหนะทั้งหมดที่อยู่ในเซลล์ต้นทางในช่องเวลานั้น กับจำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ต้นทางใน 1 ช่องเวลา สามารถเขียนเป็นสูตรคำนวณได้ดังนี้

$$s_i(t) = \min\{n_i(t), q_i(t)\} \quad (2.1)$$

โดย

$s_i(t)$ คือ ความสามารถในการส่งของเซลล์ i ในช่องเวลา t

$q_i(t)$ คือ จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ i ไปยังเซลล์ปลายทางได้ ในช่องเวลา t

2.3.2 ความสามารถในการรับ

ความสามารถในการรับ คือ ความสามารถของเซลล์ปลายทางที่จะรองรับยานพาหนะที่เคลื่อนที่เข้ามาได้ โดยขึ้นอยู่กับจำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่เข้าสู่เซลล์ปลายทางใน 1 ช่องเวลา กับพื้นที่ว่างของเซลล์ปลายทางในช่องเวลานั้น สามารถเขียนเป็นสูตรคำนวณได้ดังนี้

$$r_i(t) = \min\{q_i(t), \delta_i(C_i - n_i(t))\} \quad (2.2)$$

โดย

$r_i(t)$ คือ ความสามารถในการรับของเซลล์ i ในช่วงเวลา t

δ_i คือ ค่าสัมประสิทธิ์การเพิ่มขึ้นของยานพาหนะของเซลล์ i

C_i คือ จำนวนยานพาหนะสูงสุดที่สามารถจุในเซลล์ i ได้

หลังจากคำนวณค่าความสามารถในการส่งของเซลล์และความสามารถในการรับของเซลล์ได้แล้ว ก็สามารถคำนวณค่าจำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ในช่องเวลาต่าง ๆ ได้ โดยการคำนวณจะแตกต่างกันไปตามลักษณะการเชื่อมต่อของเซลล์ ในวิทยานิพนธ์นี้จะพิจารณาเฉพาะการเชื่อมต่อของเซลล์แบบตามลำดับดังนี้

2.3.3 จำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ในการเชื่อมต่อของเซลล์แบบตามลำดับ

การคำนวณจำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ในการเชื่อมต่อของเซลล์แบบตามลำดับ จากรูปที่ 2.3 เริ่มต้นจากการคำนวณหาความสามารถในการส่งของเซลล์ต้นทางและความสามารถในการรับของเซลล์ปลายทาง จากนั้นจึงพิจารณาค่าที่น้อยที่สุดระหว่าง 2 ค่านี้ สามารถเขียนเป็นสูตรคำนวณได้ดังนี้

$$y_i(t) = \min\{s_i(t), r_{i+1}(t)\} \quad (2.3)$$

จากสมการ (2.1) (2.2) และ (2.3) สามารถยุบรวม เพื่อเป็นสูตรการคำนวณจำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ที่ $1, 2, \dots, N$ ของ CTM ในการเชื่อมต่อของเซลล์แบบตามลำดับได้ดังนี้

$$y_i(t) = \min\{n_i(t), q_i(t), \delta [C_{i+1} - n_{i+1}(t)]\} \quad \text{โดยที่ } i \in \{1, 2, \dots, N\} \quad (2.4)$$

หลังจากคำนวณหา จำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ในช่องเวลา t ได้แล้ว สามารถนำไปคำนวณหาจำนวนยานพาหนะที่อยู่ในเซลล์ของ CTM แต่ละเซลล์ในช่องเวลา $t + 1$ ตามกฎการอนุรักษ์การไหลในขั้นตอนต่อไป

2.3.4 จำนวนยานพาหนะในเซลล์ตามกฎการอนุรักษ์การไหล

กฎการอนุรักษ์การไหล เป็นสมการที่ใช้คำนวณความหนาแน่นของเซลล์ เมื่อคำนวณอัตราการไหลและทราบความหนาแน่นของเซลล์ในช่องเวลาก่อนหน้า

เมื่อประยุกต์เทียบกับจำนวนยานพาหนะจะได้ว่าอัตราการไหล คือ จำนวนยานพาหนะที่เคลื่อนระหว่างเซลล์ของ CTM แต่ละเซลล์ และความหนาแน่น คือ จำนวนยานพาหนะในเซลล์ ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$n_i(t + 1) = n_i(t) + y_{i-1}(t) - y_i(t) \quad \text{โดยที่ } i \in \{1, 2, \dots, N\} \quad (2.5)$$

จากการคำนวณด้วยแบบจำลองการส่งผ่านเซลล์ ทำให้สามารถทราบจำนวนยานพาหนะที่อยู่บนถนนในเวลาต่าง ๆ ได้ ยานพาหนะเหล่านี้ก็คือ ผู้ใช้งานในระบบโครงข่ายไร้สายแบบวีวีพพันซ์ ซึ่งจำนวนผู้ใช้ในเวลาต่าง ๆ จะถูกนำไปใช้คำนวณการตัดสินใจเลือกโครงข่ายไร้สายต่อไปในบทที่ 3

2.4 การเรียนรู้แบบเสริมแรง

การเรียนรู้แบบเสริมแรง [15] เป็นอีกสาขาย่อยหนึ่งของการเรียนรู้ของเครื่อง (machine learning) ซึ่งมีส่วนประกอบพื้นฐานดังนี้

1. ตัวกระทำ
2. การกระทำ
3. สภาพแวดล้อม

ซึ่งการทำงานของการเรียนรู้แบบเสริมแรงโดยพื้นฐานจะเป็นวงรอบดังรูปที่ 2.4



รูปที่ 2.4: วงรอบของวิธีการเรียนรู้แบบเสริมแรง

ในการเรียนรู้แบบเสริมแรงนั้น กระบวนการส่วนใหญ่จะเกี่ยวข้องกับความสัมพันธ์ระหว่างตัวกระทำและสภาพแวดล้อม โดยตัวกระทำจะพยายามเลือกการกระทำใด ๆ ในชุดของการกระทำที่เป็นไปได้ทั้งหมด ในสภาพแวดล้อมปัจจุบันหรือสถานะ หลังจากนั้นการกระทำที่ถูกเลือกจะมีผลทำให้สภาพแวดล้อมเปลี่ยนแปลงไปและตัวกระทำก็จะได้รับค่าผลรางวัล ซึ่งขึ้นอยู่กับว่าการกระทำดังกล่าวมีผลให้สภาพแวดล้อมเปลี่ยนแปลงไปในทางใด (ดีขึ้นหรือแย่ลง ถ้าดีขึ้นค่าผลรางวัลจะสูง แต่ถ้าแย่ลงค่าผลรางวัลจะต่ำ)

จากค่าผลรางวัลที่ได้รับ ตัวกระทำจะพยายามค้นหานโยบาย (policy) ในการเลือกการกระทำในสภาพแวดล้อมใด ๆ เพื่อให้ ค่าผลรางวัลในระยะยาวนั้นมีค่ามากที่สุด

สภาพแวดล้อมในการเรียนรู้แบบเสริมแรงนั้นมักจะอยู่ในรูปแบบของกระบวนการตัดสินใจมาร์คอฟ (markov decision process, MDP) และการวิเคราะห์พฤติกรรมของการเรียนรู้ประเภทนี้ส่วนใหญ่จะใช้เทคนิคของกำหนดการพลวัต (dynamic programming) และเพื่อให้ง่ายต่อการวิเคราะห์และการนำไปใช้งาน ในบางครั้งการเปลี่ยนแปลงของสภาพแวดล้อมและค่าผลรางวัลจะถูกกำหนดให้คงที่

การเรียนรู้แบบเสริมแรงนี้แตกต่างจากการเรียนรู้แบบมีผู้สอน (supervised learning) ตรงที่ตัวกระทำเองจะไม่ได้เรียนรู้จากชุดของข้อมูลตัวอย่าง (training set) แต่จะทำการโต้ตอบกับสภาพแวดล้อมที่ตัวกระทำกำลังทำงานอยู่โดยตรง ดังนั้นข้อมูลเดียวที่ตัวกระทำสามารถใช้ในการเรียนรู้ได้ก็คือค่าผลรางวัลที่ได้รับเมื่อมีการเลือกการกระทำใดการกระทำหนึ่ง ซึ่งการเรียนรู้ดังกล่าวเป็นลักษณะการเรียนรู้แบบลองผิดลองถูก (trial-and-error)

โดยส่วนใหญ่แล้วจะมีการนำการเรียนรู้แบบนี้ไปใช้ในงานที่เกี่ยวข้องกับการควบคุมหรือเกมส์ เช่น การควบคุมหุ่นยนต์ [15] หมากรุก [23] เป็นต้น

2.4.1 องค์ประกอบของวิธีการเรียนรู้แบบเสริมแรง

โดยทั่วไปแล้ววิธีการเรียนรู้แบบเสริมแรงจะประกอบด้วย 4 องค์ประกอบหลัก [15] ดังนี้

1. นโยบาย (policy) คือ ตัวกำหนดแนวทางของการเรียนรู้และการตัดสินใจในการเลือกการกระทำลำดับถัดไป เพื่อให้ได้ผลรางวัลเฉลี่ยในระยะยาวที่สูงที่สุด โดยกำหนดให้นโยบายแทนด้วย π ซึ่งได้มาจากการหาการกระทำที่ทำให้ได้ค่าสูงสุดของผลเฉลยรางวัลสะสมเพื่อนำไปใช้เป็นฟังก์ชันในการเลือกการกระทำดังสมการ

$$\pi(s) = \arg \max_a Q(s, a) \quad (2.6)$$

โดย

$\pi(s)$ คือ นโยบายการกระทำในแต่ละสถานะ s

$Q(s, a)$ คือ ค่าผลเฉลยรางวัลสะสมที่ได้จากการกระทำ a ในสถานะ s

2. ฟังก์ชันผลรางวัล (reward function) คือ ฟังก์ชันที่ใช้คำนวณผลรางวัลในระยะยาวของระบบที่ได้ดำเนินอยู่ในขณะนั้น โดยหากเป็นการตัดสินใจที่ถูกต้องจะได้ผลรางวัลในระดับที่สูง แต่หากเป็นการตัดสินใจที่ถูกต้องน้อยกว่าหรือเป็นการตัดสินใจที่ผิดจะได้รับผลรางวัลในระดับที่ต่ำลงมา ซึ่งผลรางวัลนี้จะเป็นตัวแสดงถึงความสามารถของตัวตัดสินใจว่าสามารถทำการตัดสินใจได้ดีหรือไม่ในช่วงเวลาในขณะนั้น จุดมุ่งหมายหลักของการแก้ปัญหาด้วยวิธีการเรียนรู้แบบเสริมแรงคือ ต้องการการเรียนรู้เพื่อให้ได้นโยบายในการตัดสินใจที่ส่งผลให้ได้ผลรางวัลในระยะยาวสูงที่สุด ซึ่งจะแสดงถึงการตัดสินใจที่ดีของวิธีการเรียนรู้แบบสนับสนุน หากพิจารณาผลรางวัลที่เกิดขึ้นภายหลังช่วงเวลา t ค่าผลรวมของผลรางวัลที่เกิดขึ้นตั้งแต่เวลา t จนถึงเวลา T จะสามารถคำนวณได้จากสมการ

$$R(t) = r(t+1) + r(t+2) + r(t+3) + \dots + r(T) \quad (2.7)$$

โดย

$R(t)$ คือ ผลรวมของผลรางวัลของช่วงเวลาตั้งแต่เวลา t จนถึงเวลา T

$r(t)$ คือ ผลรางวัลที่เกิดขึ้นในเวลา t

อนึ่ง ในหลายกรณีของการนำวิธีการเรียนรู้แบบเสริมแรงมาประยุกต์ใช้ กระบวนการที่เกิดขึ้นในระบบจะเกิดขึ้นเรื่อย ๆ ไม่มีช่วงเวลาจำกัด ($T = \infty$) การคำนวณผลรางวัลข้างต้นจึงจำเป็นต้องมีพารามิเตอร์ควบคุมผลรวมของผลรางวัลไม่ให้เกิดการลู่ออก สมการ (2.7) จึงถูกเขียนใหม่เป็น

$$R(t) = r(t+1) + \gamma r(t+2) + \gamma^2 r(t+3) + \dots = \sum_{k=0}^{\infty} \gamma^k r(t+k+1) \quad (2.8)$$

โดย

γ คือ อัตราการลดทอน (discount rate) โดยที่ $\gamma \in [0, 1)$

รายละเอียดของการนำวิธีการคำนวณผลรางวัลมาประยุกต์ใช้ในวิทยานิพนธ์ฉบับนี้จะกล่าวถึงต่อไปในบทที่ 3

3. ฟังก์ชันมูลค่า (value function) คือ ฟังก์ชันที่ใช้คำนวณหาผลรางวัลในระยะยาวที่คาดว่าจะได้รับหากเลือกการกระทำนั้น ๆ ภายใต้นโยบายเดียวกัน ซึ่งวิธีการเรียนรู้แบบเสริมแรง จะใช้ฟังก์ชันนี้ในการตัดสินใจเลือกการกระทำในลำดับถัดไป โดยถ้าให้นโยบาย π เป็นนโยบายที่ทำการเชื่อมโยงระหว่างสถานะ s และการกระทำ a ฟังก์ชันมูลค่าจะดำเนินไปภายใต้นโยบายตามสมการ

$$V^\pi(s) = E_\pi \{R(t)|s(t) = s\} \quad (2.9)$$

โดย

$V^\pi(s)$ คือ ฟังก์ชันมูลค่าที่สถานะ s ภายใต้นโยบาย π

$s(t)$ คือ สถานะของระบบในเวลา t

และค่าผลรางวัลเฉลี่ยที่ได้จากการกระทำ a ในสถานะ s ภายใต้นโยบาย π จะคำนวณได้จากสมการ

$$Q^\pi(s, a) = E_\pi \{R(t)|s(t) = s, a(t) = a\} \quad (2.10)$$

โดย

$Q^\pi(s, a)$ คือ ค่าผลเฉลี่ยรางวัลสะสมที่ได้จากการกระทำ a ในสถานะ s ภายใต้นโยบาย π

$a(t)$ คือ การกระทำที่ตัวกระทำได้ตัดสินใจเลือกกระทำในเวลา t

การนำวิธีการเรียนรู้แบบเสริมแรงมาใช้ในการแก้ปัญหา นั้น จะเป็นการหานโยบายที่จะทำให้เกิดผลรางวัลสะสมระยะยาวสูงที่สุด โดยจะทำการปรับนโยบายให้ดีขึ้นโดยตลอด ดังนั้นนโยบายใหม่จึงดีกว่าหรือเท่ากับนโยบายเดิม ซึ่งจะทำให้ได้นโยบายอย่างน้อยหนึ่งนโยบายที่ดีกว่าหรือเท่ากับนโยบายอื่น ๆ ซึ่งเรียกว่านโยบายที่เหมาะสม (optimal policy) และเมื่อให้ระบบดำเนินไปตามนโยบายนี้ จะได้ฟังก์ชันมูลค่าที่เหมาะสมจากสมการ

$$V^*(s) = \max_\pi V^\pi(s) \quad (2.11)$$

โดย

$V^*(s)$ คือ ฟังก์ชันมูลค่าที่สถานะ s ภายใต้นโยบายที่เหมาะสม

ในทำนองเดียวกัน ค่าผลเฉลี่ยรางวัลสะสมภายใต้นโยบายที่เหมาะสม ($Q^*(s, a)$) สามารถหาได้จากสมการ

$$Q^*(s, a) = \max_\pi Q^\pi(s, a) \quad (2.12)$$

โดย

$Q^*(s, a)$ คือ ค่าผลเฉลี่ยรางวัลสะสมที่ได้จากการกระทำ a ในสถานะ s ภายใต้นโยบายที่เหมาะสม

สำหรับในทุก ๆ คู่ (s, a) ของสถานะและการกระทำ (state-action pair) จะได้ความสัมพันธ์ของค่าผลเฉลี่ยรางวัลสะสมกับฟังก์ชันมูลค่าที่เหมาะสมจากสมการ

$$Q^*(s, a) = E \{r(t+1) + \gamma(V^*(s(t+1))) | s(t) = s, a(t) = a\} \quad (2.13)$$

4. แบบจำลองของสภาวะแวดล้อม (model of environment) คือ แบบจำลองของสิ่งแวดล้อมที่จะนำวิธีการเรียนรู้แบบสนับสนุนมาประยุกต์ใช้ ซึ่งต้องมีความสามารถที่จะแสดงพฤติกรรมได้เหมือนกับสภาวะแวดล้อมจริง ตัวอย่างเช่น แบบจำลองการเคลื่อนที่ของผู้ใช้แบบมหภาค (CTM) ที่ได้กล่าวไปในหัวข้อ 2.1 - 2.3 ซึ่งในวิทยานิพนธ์ฉบับนี้จะกล่าวถึงการนำวิธีการเรียนรู้แบบเสริมแรงมาประยุกต์ใช้กับแบบจำลองดังกล่าวนี้ในบทที่ 3

2.4.2 วิธีการเรียนรู้แบบคิว

วิธีการเรียนรู้แบบคิว (Q-Learning algorithm) [15] เป็นการเรียนรู้แบบเสริมแรงประเภทหนึ่ง que ที่เรียนรู้จากสภาพแวดล้อมโดยตรงและไม่มีการสะสมข้อมูล (memory storage) โดยผลเฉลี่ยรางวัลสะสมนั้นจะถูกปรับใหม่ในทุก ๆ วงรอบของการเรียนรู้ด้วยสมการ

$$Q(s(t), a(t)) \leftarrow Q(s(t), a(t)) + \alpha [R(t) + \gamma \max_a Q(s(t+1), a) - Q(s(t), a(t))] \quad (2.14)$$

โดย

$Q(s(t), a(t))$ คือ ค่าผลเฉลี่ยรางวัลสะสมที่ได้จากการกระทำ a ในสถานะ s ในเวลา t

$R(t)$ คือ ผลรางวัลที่เกิดขึ้นจากการกระทำใด ๆ ในเวลา t

α คือ อัตราการเรียนรู้ (learning rate) โดยที่ $\alpha \in (0, 1]$

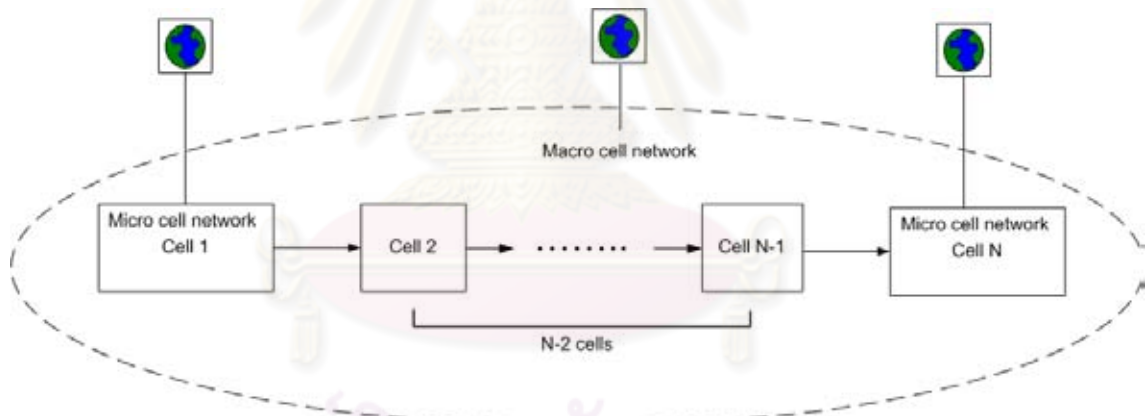
γ คือ อัตราการลดทอน (discount rate) โดยที่ $\gamma \in [0, 1)$

จากสมการ (2.14) เห็นได้ว่า ค่าผลเฉลี่ยรางวัลสะสมที่เหมาะสมนั้นจะถูกประมาณค่าได้โดยไม่ต้องมีนโยบายกำหนดการเลือกการกระทำ กล่าวคือ ค่าผลเฉลี่ยรางวัลสะสมนั้นสามารถปรับค่าจากสภาพแวดล้อมได้โดยตรง และสามารถเลือกการกระทำในครั้งต่อไปได้โดยการดึงข้อมูลจากค่าผลเฉลี่ยรางวัลสะสมโดยตรงเช่นกัน จากลักษณะดังกล่าวของวิธีการเรียนรู้แบบคิว ทำให้มีความซับซ้อนของกระบวนการน้อยมากเมื่อเทียบกับวิธีอื่น [15] ในวิทยานิพนธ์ฉบับนี้จึงเลือกใช้วิธีการเรียนรู้แบบคิวมาประยุกต์ในการควบคุมกระบวนการตัดสินใจเลือกโครงข่าย ซึ่งจะกล่าวถึงรายละเอียดของขั้นตอนการตัดสินใจเลือกโครงข่ายในบทที่ 3

บทที่ 3

การตัดสินใจเลือกโครงข่ายโดยใช้วิธีการเรียนรู้แบบเสริมแรง

ในวิทยานิพนธ์ฉบับนี้ จะพิจารณาสถานการณ์ที่ผู้ใช้บนท้องถนน ซึ่งต้องการเรียกใช้โครงข่ายไร้สายในตำแหน่งที่ต่างกัน และจำลองสภาพการจราจรโดยแบ่งถนนเป็นเซลล์ตามแนวคิดของแบบจำลองการส่งผ่านเซลล์เป็นจำนวน N เซลล์ของ CTM ดังรูปที่ 3.1 โครงข่ายในระบบแบ่งออกเป็น 2 ประเภท คือโครงข่ายไมโครเซลล์ โดย 1 เซลล์ของโครงข่ายไมโครเซลล์ ครอบคลุมพื้นที่ถนน 1 เซลล์ของ CTM และ 1 เซลล์ของโครงข่ายแมโครเซลล์ ครอบคลุมพื้นที่ทุกเซลล์ของ CTM ในระบบ สำหรับงานวิจัยนี้ได้กำหนดให้โครงข่ายไมโครเซลล์ของเซลล์ที่ 1 และเซลล์ที่ N ของ CTM เชื่อมต่อกับระบบอินเทอร์เน็ตได้โดยตรง และโครงข่ายแมโครเซลล์เชื่อมต่อกับระบบอินเทอร์เน็ต ดังนั้นผู้ใช้งานในแต่ละเซลล์ของ CTM จะสามารถเชื่อมต่อกับอินเทอร์เน็ตได้ 2 ทางคือ ผ่านโครงข่ายแมโครเซลล์และโครงข่ายไมโครเซลล์ สำหรับเซลล์ของ CTM ที่มีลักษณะพิเศษคือ เซลล์ของ CTM ที่ไม่มีโครงข่ายไมโครเซลล์เป็นของตัวเอง สามารถเชื่อมต่อกับอินเทอร์เน็ตได้ทางโครงข่ายแมโครเซลล์ และผ่านโหนดแอตท็อก คือ เชื่อมต่อผ่านระหว่างผู้ใช้ด้วยกันเองไปยังโครงข่ายไมโครเซลล์ของเซลล์ที่ 1 หรือเซลล์ที่ N ของ CTM



รูปที่ 3.1: โครงข่ายไมโครเซลล์และแมโครเซลล์ในแบบจำลองการเชื่อมต่อเซลล์แบบตามลำดับของแบบจำลองการส่งผ่านเซลล์

3.1 สมมติฐานของแบบจำลอง

เมื่อพิจารณาถึงระบบทั้งหมด จะสามารถแบ่งระบบออกเป็น 2 ระบบย่อย ได้แก่ ระบบการเคลื่อนที่ของผู้ใช้ ซึ่งในวิทยานิพนธ์นี้จะนำเอาแบบจำลองการเคลื่อนที่ CTM กรณีเชื่อมต่อแบบตามลำดับ มาใช้ และระบบของโครงข่ายไร้สายแบบวิธีพันธุ โดยมีสมมติฐานดังต่อไปนี้

1. ยานพาหนะ 1 คันในแบบจำลองการเคลื่อนที่ CTM เทียบเท่ากับผู้ใช้ 1 คนในระบบโครงข่ายไร้สายแบบวิธีพันธุ

2. โครงข่ายไร้สายมี 2 ประเภท คือ 1) โครงข่ายแมโครเซลล์มีพื้นที่ครอบคลุมผู้ใช้ทุกคนหรือเซลล์ของ CTM ทุกเซลล์ ในระบบจะมีโครงข่ายแมโครเซลล์เพียงโครงข่ายเดียวเท่านั้น 2) โครงข่ายไมโครเซลล์มีพื้นที่ครอบคลุมเท่ากับพื้นที่ของเซลล์ของ CTM โดยเซลล์ของ CTM แต่ละเซลล์จะมีจำนวนโครงข่ายไมโครเซลล์ครอบคลุมสูงสุดได้โครงข่ายเดียว อย่างไรก็ตามอาจมีเซลล์ของ CTM บางเซลล์ที่ไม่มีโครงข่ายไมโครเซลล์ครอบคลุม
3. ผู้ใช้ในระบบมีจุดมุ่งหมายในการออกสู่ระบบอินเทอร์เน็ต ไม่ใช่ต้องการติดต่อสื่อสารระดับเดียวกันภายในระบบ (peer-to-peer) จึงเสมือนว่าโครงข่ายแต่ละประเภทเป็นเกตเวย์ในการออกสู่อินเทอร์เน็ต จำนวนเกตเวย์จึงเท่ากับจำนวนโครงข่ายโดยโครงข่ายแต่ละประเภทจะมีขีดจำกัดสูงสุดในการรับผู้ใช้ที่เข้ามา
4. การเชื่อมต่อสู่ระบบอินเทอร์เน็ตเป็นการติดต่อในลักษณะกำหนดการเชื่อมต่อ (connection oriented) ซึ่งเป็นการจองทรัพยากรโครงข่ายตลอดจนกว่าจะสิ้นสุดการติดต่อสื่อสาร ไม่ใช่ลักษณะที่ไม่ต้องกำหนดการเชื่อมต่อ (connectionless) ที่ผู้ใช้ไม่จำเป็นต้องจองทรัพยากรโครงข่ายตลอดจนสิ้นสุดการติดต่อสื่อสาร
5. ผู้ใช้ในเซลล์ของ CTM ที่มีโครงข่ายไมโครเซลล์สามารถติดต่อสู่อินเทอร์เน็ตได้ 2 ทางคือ 1) ผ่านโครงข่ายแมโครเซลล์ 2) ผ่านโครงข่ายไมโครเซลล์ของตนเอง
6. ผู้ใช้ในเซลล์ของ CTM ที่ไม่มีโครงข่ายไมโครเซลล์สามารถติดต่อสู่อินเทอร์เน็ตได้ 2 ทางคือ 1) ผ่านโครงข่ายแมโครเซลล์ 2) ผ่านโครงข่ายไมโครเซลล์ของเซลล์ของ CTM ที่อยู่ข้างหน้า หรืออยู่ข้างหลังด้วยกระบวนการแอตช็อก

จากลักษณะของผู้ใช้และโครงข่ายไร้สายที่กล่าวมาข้างต้น สามารถนำเสนอเป็นแบบจำลองในกรณีทั่วไปได้โดยมองว่ามีเซลล์ของ CTM ทั้งหมด 2 ประเภท คือ เซลล์ที่มีโครงข่ายไมโครเซลล์ครอบคลุม กับเซลล์ที่ไม่มีโครงข่ายไมโครเซลล์

ในวิทยานิพนธ์ ฉบับนี้ จะ ทำ การ หา ค่า ของ จำนวน ผู้ใช้ ที่ ถูก ปฏิเสธ การ ติดต่อกับ โครงข่ายต่าง ๆ จากกระบวนการดังรูปที่ 3.2 เพื่อนำมาคำนวณค่านำหนักในการเลือกใช้โครงข่ายที่เหมาะสม โดยมีตัวแปรประเภทต่าง ๆ ดังนี้

ดัชนีระบุตำแหน่งของตัวแปร

i คือ ดัชนีของเซลล์ของ CTM เป็นได้ตั้งแต่ $i = 1$ แทนเซลล์ที่ 1 ของ CTM จนถึง $i = N$ แทนเซลล์ที่ N ของ CTM

j คือ ดัชนีของสถานะของผู้ใช้ เป็นได้สองค่าดังนี้ 0 แทนสถานะของผู้ใช้ที่ไม่ได้ติดต่อกับโครงข่าย (สถานะไม่ทำงาน) และ 1 แทนสถานะของผู้ใช้ที่ติดต่อกับโครงข่าย (สถานะทำงาน)

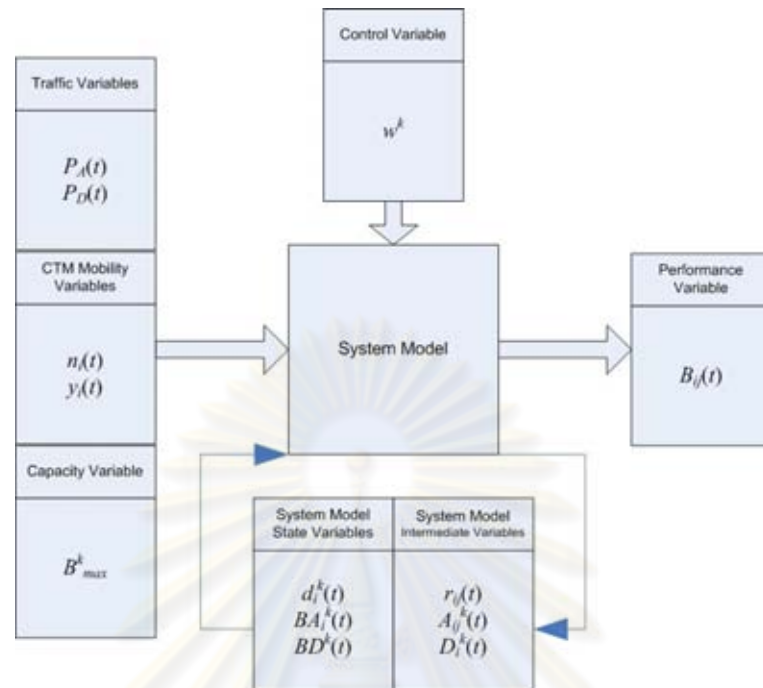
k คือ ดัชนีของโครงข่ายไร้สาย เป็นได้ 3 ค่าดังนี้ 0 แทนโครงข่ายแมโครเซลล์, 1 แทนโครงข่ายไมโครเซลล์ที่ครอบคลุมเซลล์ที่ 1 ของ CTM และ N แทนโครงข่ายไมโครเซลล์ที่ครอบคลุมเซลล์ที่ N ของ CTM

ตัวแปรขาเข้าของระบบ

$P_A(t)$ คือ ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในช่องเวลา t

$P_D(t)$ คือ ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในแต่ละช่องเวลา t

$n_i(t)$ คือ จำนวนรถยนต์ของเซลล์ที่ i ของ CTM ในช่วงเริ่มต้นช่องเวลา t



รูปที่ 3.2: แผนภาพของตัวแปรขาเข้า ตัวแปรขาออก และตัวแปรสถานะในระบบ

$y_i(t)$ คือ จำนวนรถยนต์ที่เคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ภายในช่วงเวลา t
 B_{max}^k คือ แบนด์วิดท์สูงสุดของโครงข่าย k

ตัวแปรสถานะของระบบ

$d_i^k(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะทำงาน ในเซลล์ที่ i ของ CTM โดยติดต่อกับโครงข่าย k ในช่วงเวลา t ภายหลังจากที่มีผู้ใช้ในเซลล์ของ CTM เสร็จสิ้นการติดต่อกับโครงข่าย

$BA_j^k(t)$ คือ แบนด์วิดท์ของโครงข่าย k ที่จัดสรรไว้สำหรับผู้ใช้ในสถานะ j ในช่วงเวลา t

$BD^k(t)$ คือ แบนด์วิดท์ของโครงข่าย k ที่เหลืออยู่ก่อนช่วงเสร็จสิ้นการติดต่อกับโครงข่าย ในช่วงเวลา t

$r_{ij}(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะ j ซึ่งต้องการเชื่อมต่อกับโครงข่ายต่าง ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t

$A_{ij}^k(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วสามารถติดต่อกับโครงข่าย k ได้สำเร็จในเซลล์ที่ i ของ CTM ในช่วงเวลา t

$D_i^k(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะทำงานแล้วเปลี่ยนเป็นสถานะไม่ทำงานในเซลล์ที่ i ของ CTM โดยเสร็จสิ้นการติดต่อกับโครงข่าย k ในตอนท้ายช่วงเวลา t

ตัวแปรควบคุมระบบ

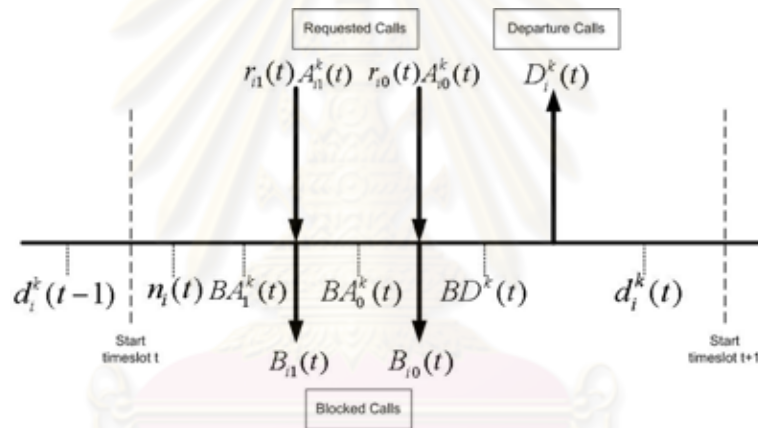
w^k คือ น้ำหนัก หรือสัดส่วน (proportion) ในการเลือกเข้าใช้โครงข่าย k ตัวแปรนี้เป็นตัวแปรที่ควบคุมการจัดสรรแบนด์วิดท์ให้กับผู้ใช้ที่ต้องการติดต่อกับโครงข่ายต่าง ๆ โดยแต่ละโครงข่ายจะมีการกำหนดเป็นค่าอัตราส่วนระหว่างจำนวนผู้ใช้ที่ติดต่อกับโครงข่ายนั้น ๆ ต่อจำนวนผู้ใช้ที่ต้องการติดต่อกับโครงข่ายทั้งหมด ในงานวิจัยนี้จะพิจารณาการปรับปรุงสมรรถนะของระบบโครงข่ายไร้สายโดยการปรับค่าสัดส่วนในการจัดสรรแบนด์วิดท์ เพื่อให้ได้ค่าสัดส่วนที่เหมาะสมที่สุด

ตัวแปรขาออกของระบบ

$B_{ij}(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วถูกปฏิเสธการติดต่อกับโครงข่ายต่าง ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t ตัวแปรนี้จะเป็นตัวชี้วัดสมรรถนะของระบบ โดยจำนวนของผู้ใช้ที่ถูกปฏิเสธการติดต่อกับโครงข่ายต่าง ๆ แสดงให้เห็นถึงประสิทธิภาพของการจัดสรรแบนด์วิดท์ของโครงข่าย เช่น ถ้าการจัดสรรแบนด์วิดท์ของโครงข่ายไม่ดีพอ จะมีจำนวนผู้ใช้ที่ถูกปฏิเสธการติดต่อกับโครงข่ายมาก เป็นต้น

3.2 แบบจำลองกราฟฟิกสำหรับกลยุทธ์การเลือกโครงข่าย

ในงานวิจัยนี้ได้พิจารณากระบวนการเรียกเข้าใช้โครงข่ายไร้สายที่เกิดขึ้นในเซลล์ i ของ CTM ในช่วงเวลา t ดังรูปที่ 3.3 ซึ่งกระบวนการเรียกเข้าใช้โครงข่ายไร้สายนี้จะมีลำดับความสำคัญ (priority) ในแต่ละเหตุการณ์ที่แตกต่างกัน โดยกระบวนการดังกล่าวจะดำเนินการเหตุการณ์ก่อนหน้าก่อนที่จะดำเนินการเหตุการณ์ถัดมา



รูปที่ 3.3: แบบจำลองลำดับการเกิดของเหตุการณ์ต่าง ๆ ในการเรียกเข้าใช้โครงข่ายไร้สายในแต่ละช่วงเวลา

แบบจำลองลำดับการเกิดของเหตุการณ์ต่าง ๆ ในการเรียกเข้าใช้โครงข่ายไร้สายในแต่ละช่วงเวลา หรือแบบจำลองกราฟฟิกของผู้ใช้ที่นำเสนอในวิทยานิพนธ์ฉบับนี้ ได้มีการดัดแปลงมาจากแบบจำลองกราฟฟิกของงานวิจัยในอดีต [24] โดยถอดตัวแปรที่ไม่จำเป็นออกเพื่อลดจำนวนสมการลง และนำตัวแปรหลาย ๆ ตัวแปรที่มีความหมายใกล้เคียงกันมานิยามใหม่ให้เหลือตัวแปรเพียงตัวแปรเดียว เพื่อให้ง่ายต่อการทำความเข้าใจและการศึกษาต่อ

ในทางปฏิบัติ การป้องกันไม่ให้เกิดความผิดพลาดในการแฮนด์ออฟจะมีความสำคัญมากกว่า การป้องกันไม่ให้เกิดการถูกปฏิเสธการเรียกเข้าใหม่ ในวิทยานิพนธ์ฉบับนี้จึงพิจารณาการจัดสรรแบนด์วิดท์เพียงวิธีเดียว คือจะจัดสรรแบนด์วิดท์ให้กับผู้ใช้ที่อยู่ในสถานะทำงานอยู่เดิมก่อนแล้วจึงจัดสรรแบนด์วิดท์ให้กับผู้ใช้ที่เดิมไม่ได้อยู่ในสถานะทำงานและต้องการขอเข้าใช้โครงข่ายในเวลานั้น ๆ

แบบจำลองกราฟฟิกที่ได้นำเสนอในวิทยานิพนธ์ฉบับนี้สามารถแยกการคำนวณเป็นขั้นตอนต่าง ๆ ดังนี้

3.2.1 การคำนวณจำนวนผู้ใช้ที่ต้องการเชื่อมต่อกับโครงข่าย

ในช่วงแรกของช่วงเวลาที่เราพิจารณามีผู้ใช้ในแต่ละสถานะต้องการเชื่อมต่อกับโครงข่ายต่าง ๆ ซึ่งจำนวนผู้ใช้ทั้งหมดที่ต้องการเชื่อมต่อกับโครงข่ายต่าง ๆ คำนวณได้จากสมการ

$$r_{ij}(t) = \sum_{\forall k \in \mathcal{K}(i)} \frac{d_{i-1}^k(t-1)}{n_{i-1}(t-1)} y_{i-1}(t-1), i \in \{1, 2, \dots, N\}, j = 1 \quad (3.1)$$

$$r_{ij}(t) = \left(n_i(t) - \sum_{\forall k \in \mathcal{K}(i)} \left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + A_{i1}^k(t) \right] \right) P_A(t), \quad i \in \{1, 2, \dots, N\}, j = 0 \quad (3.2)$$

โดย

$r_{ij}(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วต้องการเชื่อมต่อกับโครงข่ายต่าง ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t

$\mathcal{K}(i)$ คือเซตของโครงข่ายทุกโครงข่ายที่เป็นไปได้ในแต่ละเซลล์ที่ i ของ CTM

การคำนวณจำนวนผู้ใช้ที่ต้องการเชื่อมต่อกับโครงข่ายนั้นจะแยกเป็นสองกรณี คือ จำนวนผู้ใช้ที่อยู่ในสถานะทำงานอยู่เดิม ($j = 1$) ที่ต้องการเชื่อมต่อกับโครงข่ายต่าง ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t ซึ่งคำนวณได้จากจำนวนผู้ใช้ที่อยู่ในสถานะทำงานทั้งหมดที่เคลื่อนที่ออกจากเซลล์ที่ $i - 1$ ของ CTM ในช่วงเวลา $t - 1$ $\left(\frac{d_{i-1}^k(t-1)}{n_{i-1}(t-1)} y_{i-1}(t-1) \right)$ ทั้งนี้ การคำนวณจะตั้งอยู่

บนสมมุติฐานว่าผู้ใช้ในสถานะทำงานกระจายตัวอยู่ในกลุ่มยานพาหนะที่มีการเคลื่อนที่ในเซลล์ของ CTM อย่างสม่ำเสมอ ดังนั้นสัดส่วนของจำนวนผู้ใช้ในสถานะทำงานในเซลล์ของ CTM ต่อจำนวนผู้ใช้ทั้งหมดในเซลล์ของ CTM จึงเป็นค่าเดียวกับ สัดส่วนของจำนวนผู้ใช้ในสถานะทำงานที่เคลื่อนที่ออกจากเซลล์ของ CTM ต่อจำนวนผู้ใช้ทั้งหมดที่เคลื่อนที่ออกจากเซลล์ของ CTM

ใน ส่วน ของ จำนวน ผู้ใช้ ที่ อยู่ ใน สถานะ ไม่ ทำงาน ($j = 0$) แต่ ต้องการ เชื่อม ต่อ กับ โครงข่าย ต่าง ๆ ใน เซลล์ ที่ i ของ CTM ใน ช่วง เวลา t จะ คำนวณ ได้ จาก ผล ต่าง ระหว่าง จำนวน ผู้ใช้ ทั้งหมด ($n_i(t)$) กับ จำนวน ผู้ใช้ ที่ อยู่ ใน สถานะ ทำงาน อยู่ เดิม ก่อน แล้ว $\left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + A_{i1}^k(t) \right]$ คู่กับ ความ น่า จะ เป็น ที่ ผู้ ใช้ ที่ อยู่ ใน สถานะ ไม่ ทำงาน จะ เรียก เข้า ใหม่ มา ใน ระบบ ใน ช่วง เวลา t ($P_A(t)$) ซึ่ง จำนวน ผู้ใช้ ที่ อยู่ ใน สถานะ ทำงาน อยู่ เดิม มา จาก จำนวน ผู้ใช้ ที่ อยู่ ใน สถานะ ทำงาน ใน เซลล์ ที่ i ของ CTM ใน ช่วง เวลา $t - 1$ และ ไม่ได้ ออกจาก เซลล์ ที่ i ของ CTM ใน ตอน ท้าย ของ ช่วง เวลา $t - 1$ $\left(d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) \right)$ รวม กับ จำนวน ผู้ใช้ ที่ อยู่ ใน สถานะ ทำงาน ซึ่ง กำลัง เข้า สู่ เซลล์ ที่ i ของ CTM ใน ตอน ต้น ของ ช่วง เวลา t และสามารถ ติด ต่อ กับ โครงข่าย ได้ สำเร็จ ($A_{i1}^k(t)$) สำหรับ ทุก ๆ เซลล์ ที่ i ของ CTM

3.2.2 การคำนวณจำนวนผู้ใช้ที่ถูกปฏิเสธการติดต่อกับโครงข่าย

จำนวนผู้ใช้ที่ถูกปฏิเสธการติดต่อกับโครงข่ายใด ๆ สามารถคำนวณได้จาก

$$B_{ij}(t) = \sum_{\forall k \in \mathcal{K}(i)} \left[r_{ij}(t) w^k - A_{ij}^k(t) \right], i \in \{1, 2, \dots, N\} \quad (3.3)$$

โดย

$B_{ij}(t)$ คือ จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วถูกปฏิเสธการติดต่อกับโครงข่ายใด ๆ ในเซลล์ที่ i ของ CTM ในช่วงเวลา t

การคำนวณจำนวนผู้ใช้ที่ถูกปฏิเสธการติดต่อกับโครงข่ายนั้น จะคำนวณได้จากผลรวมของทุก ๆ โครงข่าย k ของผลต่างของจำนวนผู้ใช้ที่ต้องการเชื่อมต่อกับโครงข่าย ($r_{ij}(t)$) และถูกจัดสรรให้เชื่อมต่อกับโครงข่าย k ($r_{ij}(t)w^k$) กับจำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วสามารถติดต่อกับโครงข่าย k ได้สำเร็จในเซลล์ที่ i ของ CTM ในช่วงเวลา t ($A_{ij}^k(t)$)

3.2.3 การคำนวณแบนด์วิดท์ที่จัดสรรให้กับผู้ใช้

แบนด์วิดท์ที่จัดสรรให้ผู้ใช้ในแต่ละสถานะนั้นคำนวณได้จาก

$$BA_j^k(t) = B_{max}^k - \sum_{i=1}^N \left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) \right], k \in \{0, 1, N\}, j = 1 \quad (3.4)$$

$$BA_j^k(t) = B_{max}^k - \sum_{i=1}^N \left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + A_{i1}^k(t) \right], k \in \{0, 1, N\}, j = 0 \quad (3.5)$$

โดย

$BA_j^k(t)$ คือ แบนด์วิดท์ของโครงข่าย k ที่จัดสรรไว้สำหรับผู้ใช้ในสถานะ j ในช่วงเวลา t

ในการคำนวณแบนด์วิดท์ที่จัดสรรให้กับผู้ใช้นั้นจะต้องคำนึงถึงลำดับความสำคัญของผู้ใช้ด้วย ซึ่งผู้ใช้ในแต่ละสถานะจะมีลำดับความสำคัญที่ไม่เหมือนกัน โดยจะพิจารณาผู้ใช้ที่อยู่ในสถานะทำงาน ($j = 1$) ก่อน ซึ่งคำนวณได้จากผลต่างของแบนด์วิดท์สูงสุดของโครงข่าย k (B_{max}^k) กับจำนวนผู้ใช้ที่เชื่อมต่อกับโครงข่าย k ในเซลล์ที่ i ของ CTM ในช่วงเวลา $t-1$ และไม่ได้ออกจากเซลล์ที่ i ของ CTM ในตอนท้ายของช่วงเวลา $t-1$ ($d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1)$) สำหรับทุก ๆ โครงข่าย k หลังจากจัดสรรให้กับผู้ใช้ที่อยู่ในสถานะทำงานเสร็จสิ้นแล้ว จะนำแบนด์วิดท์ที่เหลือมาจัดสรรสำหรับผู้ใช้ที่อยู่ในสถานะไม่ทำงาน ($j = 0$) เป็นลำดับต่อมา

ในวิทยานิพนธ์นี้จะศึกษาเฉพาะกรณีที่แบนด์วิดท์ถูกแบ่งอย่างเด็ดขาดให้กราฟฟิกแต่ละประเภท ซึ่งได้แก่กราฟฟิกของผู้ใช้ที่อยู่ในสถานะทำงานและกราฟฟิกของผู้ใช้ที่เรียกเข้าใหม่ ไม่สามารถโยกย้ายแบนด์วิดท์ที่ใช้ไม่หมดให้กันและกันได้ เรียกกรณีเช่นนี้ว่า นโยบายไม่อนุรักษ์ (non-conserving policy) [24]

3.2.4 การคำนวณจำนวนผู้ใช้ที่สามารถติดต่อกับโครงข่ายได้สำเร็จ

จำนวนผู้ใช้ที่อยู่ในสถานะ j แล้วสามารถติดต่อกับโครงข่าย k ได้สำเร็จในเซลล์ที่ i ของ CTM ในช่วงเวลา t นั้นสามารถคำนวณได้จากจำนวนผู้ใช้ที่ต้องการและถูกจัดสรรให้เชื่อมต่อกับโครงข่าย k

$$A_{ij}^k(t) = \min \left(r_{ij}(t)w^k, BA_j^k(t) \right), i = k \in \{1, N\}, j \in \{0, 1\} \quad (3.6)$$

$$A_{ij}^k(t) = \min \left(r_{ij}(t)w^k, \left[BA_j^k(t) - A_{kj}^k(t) \right] \frac{r_{ij}(t)w^k}{\sum_{n=2}^{N-1} r_{nj}(t)w^k} \right),$$

$$i \in \{2, 3, \dots, N-1\}, k \in \{1, N\}, j \in \{0, 1\} \quad (3.7)$$

$$A_{ij}^k(t) = \min \left(r_{ij}(t)w^k, BA_j^k(t) \frac{r_{ij}(t)w^k}{\sum_{n=1}^N r_{nj}(t)w^k} \right), i \in \{1, 2, \dots, N\}, k = 0, j \in \{0, 1\} \quad (3.8)$$

โดยผู้ใช้ในแต่ละประเภทจะมีลำดับความสำคัญที่ไม่เหมือนกัน ขั้นตอนแรกจะพิจารณาผู้ใช้ที่ถูกจัดสรรให้ใช้โครงข่ายไมโครเซลล์ ($k \in \{1, N\}$) และอยู่ในเซลล์ที่ i ของ CTM ซึ่งเป็นเซลล์ที่มีโครงข่ายไมโครเซลล์เป็นของตัวเองก่อน ($i = k$) ซึ่งคำนวณได้จากค่าต่ำสุดระหว่างจำนวนผู้ใช้ในสถานะ j ที่ต้องการเชื่อมต่อกับโครงข่ายและถูกจัดสรรให้เชื่อมต่อกับโครงข่าย k ในเซลล์ที่ i ของ CTM ในช่องเวลา t และแบนด์วิดท์ของโครงข่าย k ที่จัดสรรไว้สำหรับผู้ใช้ในสถานะ j ในช่องเวลา t ดังสมการ (3.6) หลังจากจัดสรรให้กับผู้ใช้ที่อยู่ในเซลล์ที่ i ของ CTM ที่มีโครงข่ายไมโครเซลล์เป็นของตัวเองเสร็จสิ้นแล้ว จะนำแบนด์วิดท์ที่เหลือมาจัดสรรสำหรับผู้ใช้ที่อยู่ในเซลล์ที่ i ของ CTM ที่ไม่มีโครงข่ายไมโครเซลล์เป็นของตัวเอง แต่ถูกจัดสรรให้เชื่อมต่อกับโครงข่ายไมโครเซลล์ด้วยวิธีแอดฮ็อก และจะแบ่งสัดส่วนของแบนด์วิดท์ที่เหลือตามจำนวนของผู้ใช้ที่ต้องการเชื่อมต่อในแต่ละเซลล์ที่ i ของ CTM ที่ไม่มีโครงข่ายไมโครเซลล์เป็นของตัวเอง $\left(\frac{r_{ij}(t)w^k}{\sum_{n=2}^{N-1} r_{nj}(t)w^k} \right)$ ดังสมการ (3.7) สำหรับผู้ใช้ที่ถูก

จัดสรรให้ใช้โครงข่ายแมโครเซลล์ ($k = 0$) จะคำนวณแบบเดียวกับผู้ใช้ที่ถูกจัดสรรให้ใช้โครงข่ายไมโครเซลล์ และอยู่ในเซลล์ที่ i ของ CTM ที่มีโครงข่ายไมโครเซลล์เป็นของตัวเอง แต่จะแบ่งสัดส่วนของแบนด์วิดท์ที่เหลือตามจำนวนของผู้ใช้ที่ต้องการเชื่อมต่อในแต่ละเซลล์ที่ i ของ CTM ดังสมการ (3.8)

3.2.5 การคำนวณจำนวนผู้ใช้ที่เสร็จสิ้นการติดต่อกับโครงข่าย

หลังจากการคำนวณกระบวนการเรียกเข้าเสร็จสิ้น จะมีจำนวนผู้ใช้บางส่วนที่อยู่ในสถานะทำงานแล้วเปลี่ยนเป็นสถานะไม่ทำงานในเซลล์ที่ i ของ CTM โดยเสร็จสิ้นการติดต่อกับโครงข่าย k ในตอนท้ายของช่องเวลา t ซึ่งคำนวณได้จาก

$$D_i^k(t) = \left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + \sum_{j=0}^1 A_{ij}^k(t) \right] P_D(t),$$

$$i \in \{1, 2, \dots, N\}, k \in \{0, 1, N\} \quad (3.9)$$

ซึ่งก็คือจำนวนผู้ใช้ที่อยู่ในสถานะทำงาน ในเซลล์ที่ i ของ CTM โดยติดต่อกับโครงข่าย k ในช่องเวลา t ภายหลังจากกระบวนการจัดสรรแบนด์วิดท์เสร็จสิ้นแล้ว สมการนี้ตั้งอยู่บนสมมติฐานว่าผู้ใช้ที่เริ่มเข้าใช้งานโครงข่ายในช่องเวลา t สามารถเสร็จสิ้นบริการและออกจากโครงข่ายได้ในช่องเวลาเดียวกัน เทอม $\left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + \sum_{j=0}^1 A_{ij}^k(t) \right]$ คำนวณจากจำนวนผู้ใช้ที่เชื่อมต่อกับโครงข่าย k ในเซลล์ที่ i ของ CTM ในช่องเวลา $t-1$ และไม่ได้ออกจากเซลล์ที่ i ของ CTM

ในตอนท้ายของช่วงเวลา $t - 1$ รวมกับจำนวนผู้ใช้ทุก ๆ สถานะ j ที่สามารถติดต่อกับโครงข่าย k ได้สำเร็จในเซลล์ที่ i ของ CTM ในช่วงเวลา t คุณกับความน่าจะเป็นที่ผู้ใช้แต่ละคนที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในแต่ละช่วงเวลา t ($P_D(t)$)

3.2.6 การคำนวณจำนวนผู้ใช้ที่อยู่ในสถานะทำงานในตอนท้ายของช่วงเวลา

ในส่วนของจำนวนผู้ใช้ที่อยู่ในสถานะทำงาน ในเซลล์ที่ i ของ CTM โดยติดต่อกับโครงข่าย k ในช่วงเวลา t ภายหลังจากที่มีผู้ใช้ในเซลล์ของ CTM เสร็จสิ้นการติดต่อโครงข่าย คำนวณได้จาก

$$d_i^k(t) = \left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + \sum_{j=0}^1 A_{ij}^k(t) \right] (1 - P_D(t)),$$

$$i \in \{1, 2, \dots, N\}, k \in \{0, 1, N\} \quad (3.10)$$

ซึ่งก็คือจำนวนผู้ใช้ที่อยู่ในสถานะทำงาน ในเซลล์ที่ i ของ CTM โดยติดต่อกับโครงข่าย k ในช่วงเวลา t ภายหลังจากกระบวนการจัดสรรแบนด์วิธเสร็จสิ้นแล้ว คุณกับความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานยังคงติดต่อกับโครงข่ายในแต่ละช่วงเวลา t

3.2.7 การคำนวณแบนด์วิธที่เหลืออยู่ก่อนช่วงเสร็จสิ้นการติดต่อกับโครงข่าย

ส่วนแบนด์วิธของโครงข่าย k ที่เหลืออยู่ก่อนช่วงเสร็จสิ้นการติดต่อกับโครงข่าย ในช่วงเวลา t นั้นสามารถคำนวณได้จาก

$$BD^k(t) = B_{max}^k - \sum_{i=1}^N \left[d_i^k(t-1) - \frac{d_i^k(t-1)}{n_i(t-1)} y_i(t-1) + \sum_{j=0}^1 A_{ij}^k(t) \right], k \in \{0, 1, N\}$$

$$(3.11)$$

ซึ่งก็คือผลต่างของแบนด์วิธสูงสุดของโครงข่าย k กับจำนวนผู้ใช้ที่อยู่ในสถานะทำงาน ในเซลล์ที่ i ของ CTM โดยติดต่อกับโครงข่าย k ในช่วงเวลา t ภายหลังจากกระบวนการจัดสรรแบนด์วิธเสร็จสิ้นแล้ว ในวิทยานิพนธ์ฉบับนี้จะนำค่าดังกล่าวมาใช้ในกระบวนการเรียนรู้แบบเสริมแรงต่อไป

3.3 การเรียนรู้แบบเสริมแรง

ในวิทยานิพนธ์ฉบับนี้จะนำวิธีการเรียนรู้แบบเสริมแรงมาควบคุมค่าน้ำหนักในการเลือกเข้าใช้โครงข่าย ซึ่งมีขั้นตอนของการควบคุมที่เสนอดังนี้

3.3.1 กำหนดนิยามของสถานะ

จาก (3.11) ในงานวิจัยนี้กำหนดให้นิยามของสถานะเป็นดังสมการ

$$\tilde{BD}^k(t) = \left[\frac{BD^k(t)}{B_{max}^k} L \right] + I (BD^k(t) = 0), \quad (3.12)$$

โดยที่ $\tilde{B}D^k(t)$ คือค่าลำดับชั้นของแบนด์วิดท์ของโครงข่าย k ที่เหลืออยู่ก่อนช่วงเสร็จสิ้นการติดต่อกับโครงข่าย ในช่วงเวลา t และ L คือจำนวนของลำดับชั้นทั้งหมด

นำ (3.12) มานิยามสถานะได้ดังนี้

$$S := \{ \tilde{s} = [\tilde{B}D^0(t), \tilde{B}D^1(t), \tilde{B}D^N(t)] \} \quad (3.13)$$

3.3.2 กำหนดนิยามของการกระทำ

ในทุก ๆ ช่วงเวลา t กำหนดให้ $a(t) \in \{0, 1, \dots, \eta\}$ โดยที่ η คือค่าน้ำหนักที่มากที่สุด ซึ่งจะหาค่าน้ำหนักในการเลือกเข้าใช้โครงข่าย k ได้จาก

$$w^k = \frac{a(t)}{\eta}, \quad (3.14)$$

โดยที่ $\sum_k w^k = 1$

3.3.3 วิธีการเรียนรู้แบบคิว

เป้าหมายของกระบวนการเรียนรู้แบบเสริมแรงในวิทยานิพนธ์ฉบับนี้คือการควบคุมพารามิเตอร์ w^k ด้วยการปรับเปลี่ยนการกระทำ $a(t)$ เพื่อให้ค่าผลรางวัลเฉลี่ยที่สูงที่สุดในวิทยานิพนธ์ฉบับนี้ ได้เลือกวิธีการเรียนรู้แบบคิว ซึ่งเป็นการเรียนรู้แบบเสริมแรงประเภทหนึ่งที่เรียนรู้จากสภาพแวดล้อมโดยตรงและไม่มี การสะสมข้อมูล (memory storage) จึงมีความซับซ้อนของกระบวนการน้อยที่สุด [15] โดยความซับซ้อนของการคำนวณต่อหนึ่งวงรอบในแบบจำลองที่นำเสนอ นี้ จะอยู่ที่ $O(N\eta^{(\kappa-1)})$ เท่านั้น โดยที่ κ คือจำนวนเซลล์ของโครงข่าย ในวิทยานิพนธ์ฉบับนี้จึงใช้วิธีดังกล่าวมาประยุกต์ในการควบคุมกระบวนการตัดสินใจเลือกโครงข่าย

กำหนดให้

$Q(\tilde{s}, a)$	ค่าผลเฉลี่ยรางวัลสะสมที่ได้จากการกระทำ a เมื่อระบบอยู่ในเวกเตอร์สถานะ \tilde{s}
α	อัตราการเรียนรู้
γ	อัตราการลดทอนการเรียนรู้
$R(t)$	ค่าผลรางวัลที่ช่วงเวลา t

การควบคุมกระบวนการตัดสินใจเลือกโครงข่ายเทคโนโลยีสื่อสารไร้สายที่เหมาะสมโดยใช้วิธีการเรียนรู้แบบคิว มีขั้นตอนของกระบวนการดังนี้

1. ตั้งค่าเริ่มต้นของค่าผลรางวัลเฉลี่ย ซึ่งในงานวิจัยนี้จะให้ค่าเริ่มต้นของทุก ๆ เวกเตอร์สถานะ \tilde{s} และการกระทำ a มีค่าผลรางวัลเฉลี่ยเท่ากับ 0

$$Q(\tilde{s}, a) = 0, \forall(\tilde{s}, a) \quad (3.15)$$

ในทางปฏิบัติ ค่าผลรางวัลเฉลี่ยสามารถตั้งค่าเริ่มต้นเป็นเท่าไรก็ได้ [15] แต่เพื่อง่ายต่อการพิจารณาการทำงานของวิธีการเรียนรู้แบบเสริมแรง ในวิทยานิพนธ์ฉบับนี้จึงตั้งค่าเริ่มต้นการทดสอบเท่ากับ 0

- นำค่าของน้ำหนัก ในการ เลือก เข้า ใช้ โคร่งข่าย จาก การ เลือก การ กระทำ a เพื่อ ควบคุม กระบวนการตัดสินใจของโคร่งข่ายโดยใช้วิธีการเลือกแบบเอพชีรอนกรีดี (ϵ -greedy) ซึ่งเป็นการเลือกแบบสุ่มโดยการเลือกการกระทำ $a = \arg \max_{a'} Q(\tilde{s}, a')$ ด้วยความน่าจะเป็นเท่ากับ $1 - \epsilon + \frac{\epsilon}{\eta+1}$ ส่วนทุก ๆ การกระทำ $a \in \{0, 1, \dots, \eta\} - \{\arg \max_{a'} Q(\tilde{s}, a')\}$ จะถูกเลือกด้วยความน่าจะเป็นของแต่ละการกระทำเท่ากับ $\frac{\epsilon}{\eta+1}$
- นำการกระทำที่ได้จากขั้นตอนที่ 2 มาแทนค่าใน (3.14) และคำนวณหาค่าผลรางวัล $R(t)$ จาก

$$R(t) = - \sum_{i=1}^N [\beta B_{i0}(t) + (1 - \beta) B_{i1}(t)], \quad (3.16)$$

โดยที่ β คือค่าคงที่สำหรับปรับความสำคัญของค่าผลรางวัล

- นำค่าผลรางวัลที่ได้จาก (3.14) มาปรับค่าผลรางวัลเฉลี่ยใหม่โดย

$$Q(\tilde{s}, a) \leftarrow Q(\tilde{s}, a) + \alpha \left[R(t) + \gamma \max_{a'} Q(\tilde{s}', a') - Q(\tilde{s}, a) \right], \quad (3.17)$$

โดยที่ $Q(\tilde{s}', a')$ คือค่าผลรางวัลเฉลี่ยที่ได้จากการกระทำ a' ในเวกเตอร์สถานะถัดไป \tilde{s}' ซึ่งก็คือค่าผลรางวัลเฉลี่ยจะถูกปรับใหม่ด้วยการปรับค่าผลรางวัลเฉลี่ยที่สถานะและการกระทำนั้น ๆ โดยจะปรับให้ถูกปรับใหม่ในช่วง (0,1] เท่านั้น ซึ่งมีอยู่ 3 พจน์ย่อย คือค่าผลรางวัลปัจจุบัน ค่าผลรางวัลอนาคตที่ถูกปรับให้ค่าผลรางวัลรวมไม่ลู่ออก และค่าผลรางวัลในอดีต

- ปรับสถานะใหม่ $\tilde{s} \leftarrow \tilde{s}'$ โดยสถานะถัดไปจะเปลี่ยนเป็นสถานะปัจจุบันในการคำนวณค่าผลรางวัลเฉลี่ยครั้งถัดไป
- กลับไปทำขั้นตอนที่ 2 ใหม่จนกระทั่งสิ้นสุดการทดสอบ

3.4 ตัวชี้วัดสมรรถนะของระบบ

ในวิทยานิพนธ์ฉบับนี้จะพิจารณาสมรรถนะของระบบด้วยการวัดค่าพารามิเตอร์ต่าง ๆ ดังนี้

3.4.1 ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใช้โคร่งข่าย

ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใช้โคร่งข่ายจะแบ่งการพิจารณาเป็นสองตัวแปรย่อย คือความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่ช่วงเวลา t ($P_{B,0}(t)$) และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่ช่วงเวลา t ($P_{B,1}(t)$) ซึ่งตัวแปรดังกล่าวเป็นตัวชี้วัดที่สามารถนำมาวิเคราะห์สมรรถนะของระบบได้ โดยคำนวณได้จากสมการ

$$P_{B,j}(t) = \frac{\sum_i B_{ij}(t)}{\sum_i r_{ij}(t)} \quad (3.18)$$

3.4.2 ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัล (temporal difference, TD)

ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลเป็นตัวชี้วัดสมรรถนะของระบบเพื่อให้ง่ายต่อการเปรียบเทียบประสิทธิภาพของวิธีการเรียนรู้แบบเสริมแรง ซึ่งตัวแปรดังกล่าวเป็นตัวแปรบ่งชี้การลู่เข้าของการเรียนรู้ กล่าวคือ ถ้าค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลมีค่าใกล้ศูนย์มาก แสดงว่าระบบนั้นได้เรียนรู้จนกระทั่งสามารถเลือกการกระทำที่ใกล้เคียงกับการกระทำที่เหมาะสม (optimal action) มาก โดยตัวแปรนี้คำนวณได้จากสมการ

$$TD = R(t) + \gamma \max_{a'} Q(s', a') - Q(s, a) \quad (3.19)$$

3.5 สรุป

ในบทนี้ได้เสนอขั้นตอนการคำนวณตัวแปรต่าง ๆ ที่เกี่ยวข้องกับการใช้งานโครงข่าย โดยนำเอาพารามิเตอร์จำนวนยานพาหนะในระบบที่คำนวณจากแบบจำลองการส่งผ่านเซลล์มาประยุกต์ใช้ ทำให้สามารถทราบจำนวนผู้ใช้งานโครงข่ายในช่วงเวลาต่าง ๆ ได้ และสามารถคำนวณตัวชี้วัดสมรรถนะของกลยุทธ์แต่ละประเภทในสถานการณ์ต่าง ๆ ในขั้นตอนต่อไปได้ ซึ่งจะแสดงผลการทดสอบต่อไปในบทที่ 4

บทที่ 4

ผลการทดสอบ

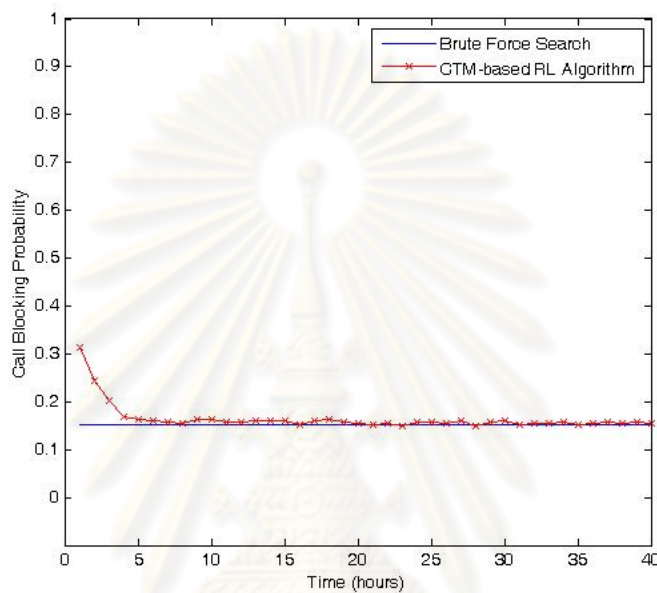
ในวิทยานิพนธ์ฉบับนี้ จะ ทดลอง นำ วิธีการ เรียนรู้ แบบ เสริม แรง มา เป็น อัล กอ ริ ทึม ใน การ ควบคุมกระบวนการตัดสินใจเลือกโครงข่ายบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ โดยพิจารณาสมรรถนะของระบบว่ามีความสัมพันธ์อย่างไรกับพารามิเตอร์ที่เกี่ยวข้องต่าง ๆ ทั้งนี้ การทดลองทุก ๆ การทดลองที่เสนอ กำหนดให้จำนวนเซลล์ของ CTM เท่ากับ 5 ($N = 5$) ค่าสัมประสิทธิ์การเพิ่มขึ้นของยานพาหนะของทุกเซลล์ที่ i เท่ากับ 0.8 ($\delta_i = 0.8, \forall i$) ทุกเซลล์ที่ i ของ CTM สามารถจุยานพาหนะได้สูงสุดเท่ากับ 100 คัน ($C_i = 100, \forall i$) และความยาวของหนึ่งช่องเวลามีค่าเท่ากับ 5 วินาที [20]

4.1 การ เปรียบเทียบ กับ วิธีการ หา ผล เฉลย แบบ ตรงไปตรงมา (brute force search)

ก่อนที่จะนำแบบจำลองโครงข่ายมาทดสอบวิธีการเรียนรู้แบบเสริมแรง ในหัวข้อนี้จะทำการเปรียบเทียบการแบ่งน้ำหนักในการเลือกเข้าใช้โครงข่าย โดยหาทุกอัตราส่วนที่เป็นไปได้ของการแบ่งน้ำหนักหรือหาผลเฉลยแบบตรงไปตรงมา [20] และนำมาเปรียบเทียบกับวิธีการเลือกเข้าใช้โครงข่ายด้วยการนำวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ (CTM-based RL) มาใช้ในการเลือก กำหนดให้ $y_0(t)$ คือ จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในช่องเวลา t ในการทดลองนี้จะกำหนดให้พารามิเตอร์ต่าง ๆ มีค่าดังนี้

- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 28,800 ช่องเวลา (40 ชั่วโมง) ซึ่งเป็นช่วงเวลาที่เพียงพอต่อการสังเกตการลู่เข้า (convergence) ของระบบ
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t เท่ากับ 50 คันต่อช่องเวลา ($y_0(t) = 50, \forall t$) [20]
- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่องเวลา t เท่ากับ 60 คันต่อช่องเวลา ($q_i(t) = 60, \forall i \forall t$) [20]
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 100 ช่องสัญญาณ ($B_{max}^0 = 100$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่องเวลา t เท่ากับ 0.5 ($P_A(t) = 0.5, \forall t$) [20]
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่องเวลา t เท่ากับ 0.5 ($P_D(t) = 0.5, \forall t$) [20]
- $\alpha = 0.01, \gamma = 0.0005, \epsilon = 0.01$ และ $\beta = 0.09$ (การปรับค่า β ให้มีค่าน้อย ๆ นั้นเพื่อต้องการให้ระบบให้ความสำคัญกับผู้ใช้ในสถานะทำงานมากกว่า)

จากค่าพารามิเตอร์ดังกล่าว สามารถนำมาหาค่าน้ำหนักที่ทำให้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟมีค่าน้อยที่สุดได้ [20] เท่ากับ 0.1521 และ 0 ตามลำดับ ซึ่งการทดลองนี้จะนำค่าที่ได้จากการหาผลเฉลยแบบตรงไปตรงมาเปรียบเทียบกับ การควบคุมการเลือกโครงข่ายโดยใช้วิธีการเรียนรู้แบบเสริมแรง จากค่าพารามิเตอร์ที่ได้กล่าวมาข้างต้น ทำให้ได้ผลการทดลองดังรูปที่ 4.1 และรูปที่ 4.2 ตามลำดับ



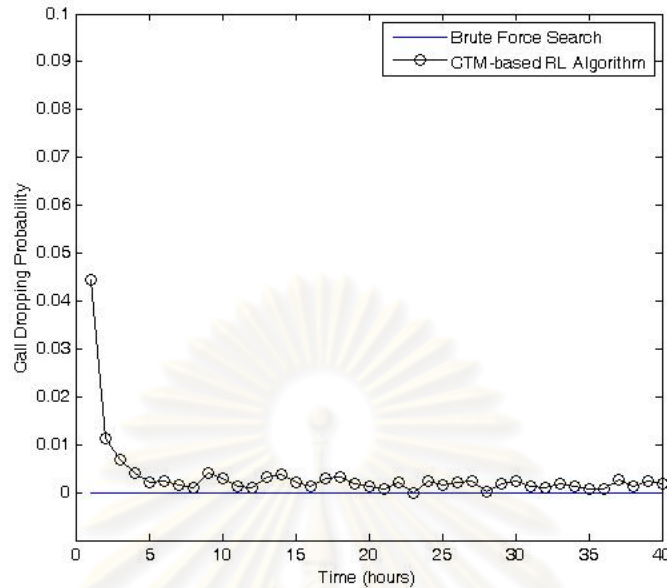
รูปที่ 4.1: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ซึ่งลู่เข้าได้โดยใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์

จากผลการทดลองข้างต้นแสดงให้เห็นว่า วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์สามารถหาค่าน้ำหนักในการเลือกใช้โครงข่ายที่เหมาะสมได้

4.2 ผลกระทบที่เกิดจากการเปลี่ยนแปลงแบนด์วิดท์และความไม่แน่นอนของระบบจราจร

ในการทดลองนี้จะเริ่มต้นทดสอบการนำวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ มาประยุกต์ใช้ในการแบ่งน้ำหนักการเลือกโครงข่ายที่จำลองขึ้นด้วยแบบจำลองของ CTM โดยจะลดความจุของแมคโครเซลล์ลงและเพิ่มความไม่แน่นอนของระบบจราจร ซึ่งมีค่าพารามิเตอร์ต่าง ๆ ดังนี้

- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 28,800 ช่องเวลา (40 ชั่วโมง)
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซองแบบวิยุตที่มีค่าพารามิเตอร์ = 60 คันต่อช่องเวลา ($y_0(t) = \text{Poisson}(60), \forall t$)

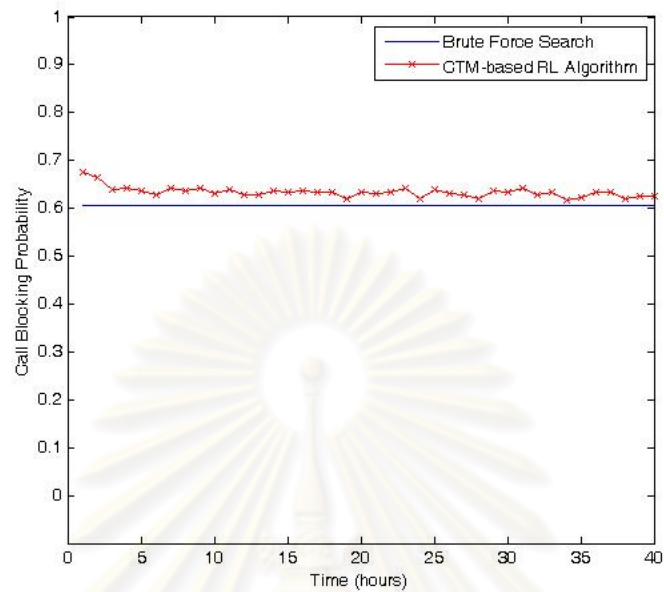


รูปที่ 4.2: ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟซึ่งลู่เข้าได้โดยใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์

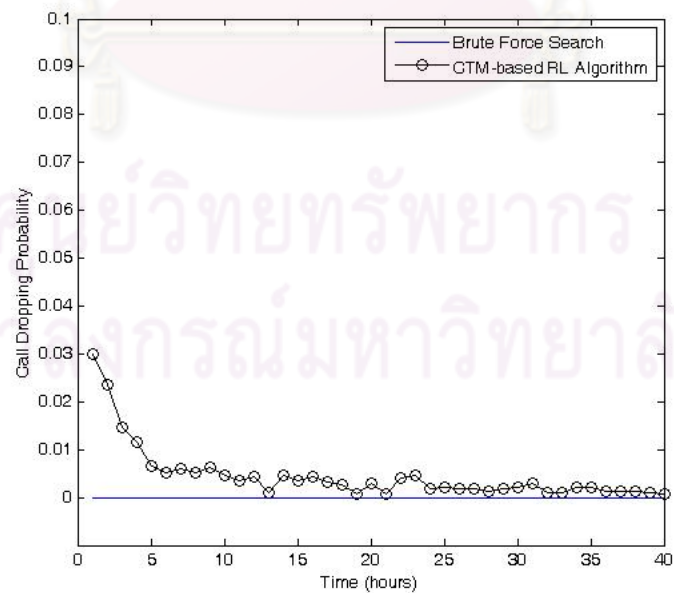
- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่วงเวลา t คำนวณจากการสุ่มปัวซองแบบวิฤตที่มีค่าพารามิเตอร์ = 60 ต่อช่วงเวลา ($q_i(t) = \text{Poisson}(60), \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแมโครเซลล์เท่ากับ 40 ช่องสัญญาณ ($B_{max}^0 = 40$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่วงเวลา t คำนวณจากการสุ่มยูนิฟอร์มแบบต่อเนื่องในช่วง 0 ถึง 1 ($P_A(t) = U(0, 1), \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่วงเวลา t คำนวณจากการสุ่มยูนิฟอร์มแบบต่อเนื่องในช่วง 0 ถึง 1 ($P_D(t) = U(0, 1), \forall t$)
- $\alpha = 0.01, \gamma = 0.0005, \epsilon = 0.01$ และ $\beta = 0.09$

จากค่าพารามิเตอร์ข้างต้น จะได้ค่าน้ำหนักที่ทำให้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟดังรูปที่ 4.3 และรูปที่ 4.4 ตามลำดับ

จากผลการทดลองแสดงให้เห็นว่า หลังจากใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์แล้ว ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่จะลดลงเพียงเล็กน้อยเมื่อเทียบกับช่วงต้นของการจำลอง และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟจะลดลงจนเหลือใกล้ ๆ ศูนย์หลังจากการเรียนรู้อยู่ในสภาวะคงที่ ซึ่งแสดงให้เห็นว่าวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์นั้นสามารถเลือกโครงข่ายที่เหมาะสมได้ ไม่ว่าจะมีการลดแบนด์วิดท์หรือเพิ่มความไม่แน่นอนของระบบบจรราก็ตาม



รูปที่ 4.3: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่โดยมีความจุแมคโครเซลล์ต่ำและมีความไม่แน่นอนของระบบจราจร



รูปที่ 4.4: ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟโดยมีความจุแมคโครเซลล์ต่ำและมีความไม่แน่นอนของระบบจราจร

ในวิทยานิพนธ์ฉบับนี้จะพิจารณาเฉพาะความสามารถในการปรับปรุงกระบวนการเลือกโครงข่ายของวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ที่ความแปรปรวนของสถานะแวดล้อมแตกต่างกันเท่านั้น ไม่มีการพิจารณาค่าความไว (sensitivity) ของตัวแปรในแบบจำลองที่เสนอ การทดลองนี้จึงไม่มีการทดสอบโดยการเปลี่ยนค่าเฉพาะตัวแปรใดตัวแปรหนึ่งและกำหนดให้ตัวแปรอื่นมีค่าคงที่เพื่อวิเคราะห์ผลจากการเปลี่ยนค่าของตัวแปรดังกล่าว

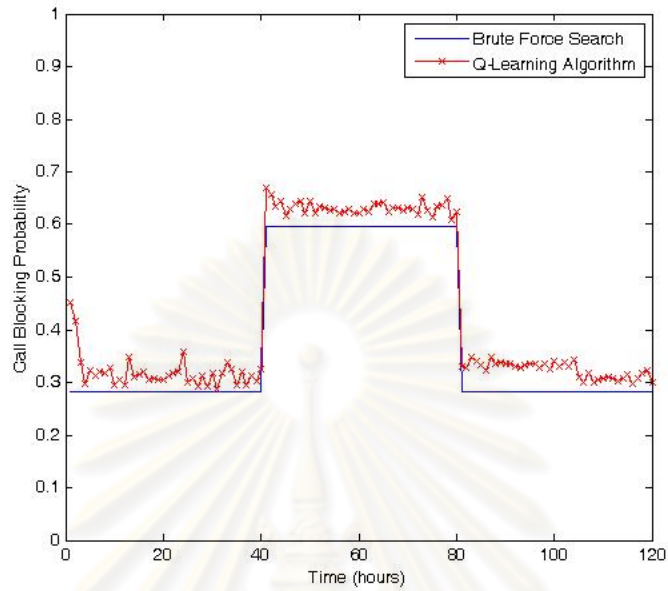
4.3 ผลการควบคุมการเลือกโครงข่ายด้วยวิธีการเรียนรู้แบบเสริมแรงโดยมีการจำลองเหตุการณ์โครงข่ายล้ม

ในการทดลองนี้จะทดสอบการนำวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ มาประยุกต์ใช้ในการแบ่งน้ำหนักการเลือกโครงข่ายโดยมีเหตุการณ์ที่ทดสอบการแก้ไขปัญหาเฉพาะหน้าของวิธีดังกล่าว ซึ่งการทดลองนี้ได้กำหนดค่าพารามิเตอร์ต่าง ๆ ดังนี้

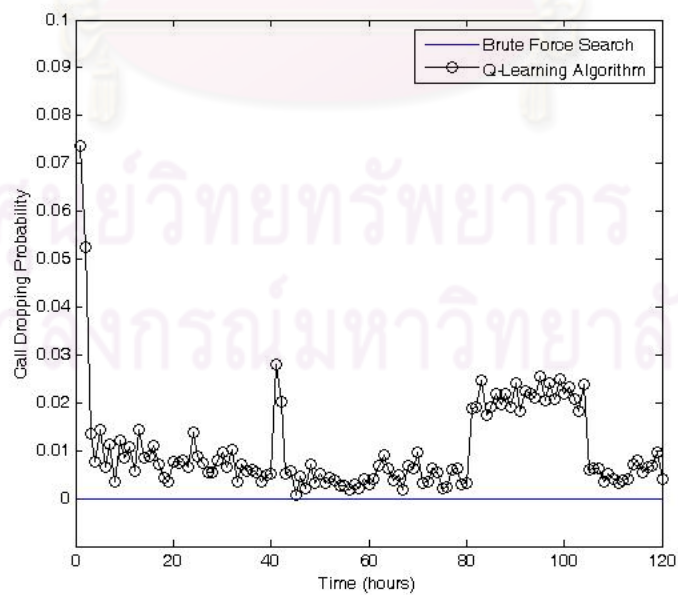
- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 86,400 ช่องเวลา (120 ชั่วโมง)
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซงแบบวิยุตที่มีค่าพารามิเตอร์ = 60 คันต่อช่องเวลา ($y_0(t) = \text{Poisson}(60), \forall t$)
- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซงแบบวิยุตที่มีค่าพารามิเตอร์ = 60 คันต่อช่องเวลา ($q_i(t) = \text{Poisson}(60), \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 100 ช่องสัญญาณ ในช่วงเวลาที่ 1 - 40 และ ชั่วโมงที่ 81 - 120 หรือในช่วงเวลาที่ 1 - 28,800 และช่องเวลาที่ 57,601 - 86,400 ($B_{max}^0(t) = 100, t \in [1, 28, 800] \cup [57, 601, 86, 400]$)
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 40 ช่องสัญญาณ ในช่วงเวลาที่ 41 - 80 หรือในช่วงเวลาที่ 28,801 - 57,600 ($B_{max}^0(t) = 40, t \in [28, 801, 57, 600]$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่องเวลา t คำนวณจากการสุ่มยูนิฟอร์มแบบต่อเนื่องในช่วง 0 ถึง 1 ($P_A(t) = \mathcal{U}(0, 1), \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่องเวลา t คำนวณจากการสุ่มยูนิฟอร์มแบบต่อเนื่องในช่วง 0 ถึง 1 ($P_D(t) = \mathcal{U}(0, 1), \forall t$)
- $\alpha = 0.01, \gamma = 0.0005, \epsilon = 0.05$ และ $\beta = 0.09$ (ปรับค่า ϵ ให้มีค่ามากกว่าเดิมเพื่อเพิ่มอัตราการสำรวจเนื่องจากการเปลี่ยนแปลงค่าแบนด์วิดท์ของโครงข่ายในช่วงเวลาที่ 41 และ 81)

จากค่าพารามิเตอร์ข้างต้น จะได้ค่าน้ำหนักที่ทำให้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟตั้งรูปที่ 4.5 และรูปที่ 4.6 ตามลำดับ

จากผลการทดลองแสดงให้เห็นว่าวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์นั้นสามารถเลือกโครงข่ายที่เหมาะสมได้แม้ว่าจะเกิดเหตุการณ์โครงข่ายล้มก็ตาม โดย



รูปที่ 4.5: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่เมื่อมีการจำลองเหตุการณ์โครงข่ายล่ม



รูปที่ 4.6: ความน่าจะเป็นของการปฏิเสธการเสนต์ออฟเมื่อมีการจำลองเหตุการณ์โครงข่ายล่ม

เมื่อเปรียบเทียบกับผลการทดลองในหัวข้อที่ 4.1 (การเปรียบเทียบกับวิธีการหาผลเฉลยแบบตรงไปตรงมา) และผลการทดลองในหัวข้อที่ 4.2 (ผลกระทบที่เกิดจากการเปลี่ยนแปลงแบนด์วิดท์และความไม่แน่นอนของระบบจราจร) นั้นจะเห็นได้ว่าช่วงที่โครงข่ายล่ม ผลจะไม่แตกต่างกับผลที่เกิดจากผลการทดลองในหัวข้อที่ 4.2 มากนัก และในช่วงที่โครงข่ายไม่มีการล่ม ผลการทดลองนั้นไม่แตกต่างกับผลที่เกิดจากผลการทดลองในหัวข้อที่ 4.1 เช่นกัน ถึงแม้ว่าจำนวนสถานะจะเพิ่มขึ้นในช่วงที่ 81 (ช่วงเวลา 57,601) ซึ่งมีผลทำให้นิยามของสถานะเปลี่ยนไปตามสมการที่ (3.12) และ (3.13) วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ยังสามารถเลือกโครงข่ายที่เหมาะสมได้หลังจากผ่านไปช่วงเวลาหนึ่ง แสดงให้เห็นว่าถ้ามีการเปลี่ยนแปลงแบนด์วิดท์อย่างกะทันหัน วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์นั้นยืดหยุ่นมากพอที่จะเลือกโครงข่ายที่เหมาะสมได้โดยการเรียนรู้ใหม่นั้นเอง

อย่างไรก็ตาม ผลการทดลองนี้แสดงให้เห็นว่าบางช่วงเวลา (ชั่วโมงที่ 81 - 100) วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ยังไม่สามารถเลือกค่าน้ำหนักที่เหมาะสมได้เนื่องจากประสิทธิภาพของวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ยังมีการปรับค่าพารามิเตอร์ที่ไม่เหมาะสม ซึ่งในวิทยานิพนธ์ฉบับนี้จะมีการทดลองปรับค่าพารามิเตอร์ในหัวข้อถัดไป

4.4 การทดสอบค่าความไว (sensitivity) ของวิธีการเรียนรู้แบบเสริมแรง

ในการทดลองนี้จะทำการทดสอบประสิทธิภาพของวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ โดยการทดลองปรับค่าพารามิเตอร์ต่าง ๆ เพื่อให้ได้ค่าพารามิเตอร์ที่เหมาะสมในแต่ละสภาพแวดล้อมของระบบ ในวิทยานิพนธ์ฉบับนี้จะนำเสนอตัวอย่างการทดลองปรับค่าพารามิเตอร์ของวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ โดยปรับค่าพารามิเตอร์ที่ใช้ในการทดสอบความไว 3 ค่า ได้แก่ ค่าอัตราการเรียนรู้ (α) ค่าอัตราการลดทอนการเรียนรู้ (γ) และค่าอัตราการสำรวจ (ϵ) โดยตลอดการทดลองจะกำหนดค่าพารามิเตอร์ต่าง ๆ ของระบบดังนี้

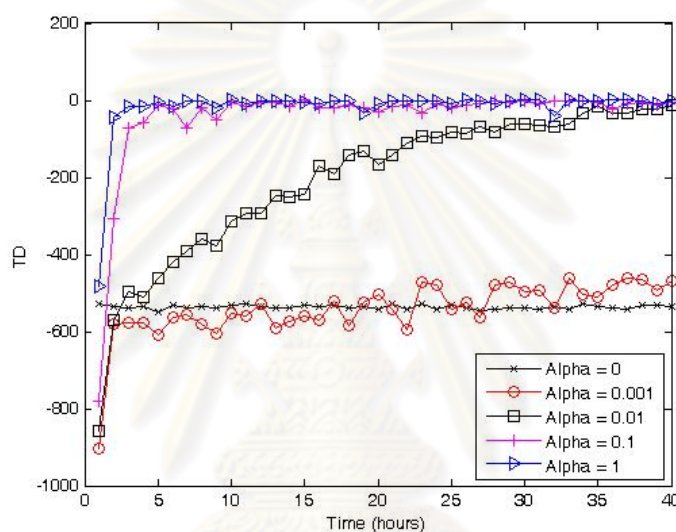
- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 28,800 ช่องเวลา (40 ชั่วโมง)
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซองแบบวิยุตที่มีค่าพารามิเตอร์ = 60 คั่นต่อช่องเวลา ($y_0(t) = \text{Poisson}(60), \forall t$)
- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่องเวลา t เท่ากับ 60 คั่นต่อช่องเวลา ($q_i(t) = 60, \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแมโครเซลล์เท่ากับ 40 ช่องสัญญาณ ($B_{max}^0 = 40$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่องเวลา t เท่ากับ 0.5 ($P_A(t) = 0.5, \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่องเวลา t เท่ากับ 0.5 ($P_D(t) = 0.5, \forall t$)

- $\beta = 0.09$

การทดลองนี้จะใช้ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลเป็นตัวชี้วัดสมรรถนะของระบบ เพื่อให้ง่ายต่อการเปรียบเทียบ

4.4.1 การทดลองปรับค่าอัตราการเรียนรู้ (α)

กำหนดให้ $\gamma = 0.001$ และ $\epsilon = 0.01$ จะได้ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลที่ค่า $\alpha = 0, 0.001, 0.01, 0.1$ และ 1 ดังรูปที่ 4.7



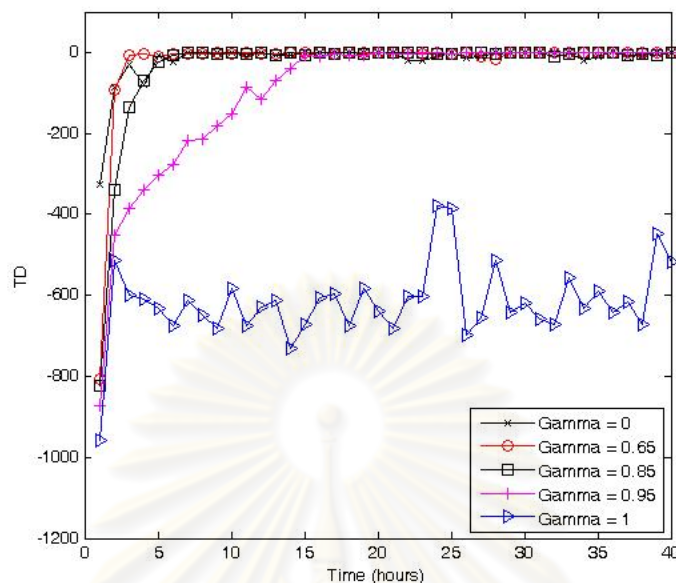
รูปที่ 4.7: ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลที่ค่า $\gamma = 0.001$ และ $\epsilon = 0.01$

จากผลการทดลองเห็นได้ว่า ค่าอัตราการเรียนรู้ที่เหมาะสมที่สุดในการทดลองนี้มีค่าเท่ากับ 1 ซึ่งถ้าพิจารณาสมการที่ (3.17) จะเป็นค่าอัตราการเรียนรู้ที่ทำให้การอัปเดตค่าผลรางวัลเฉลี่ยสะสมไม่ขึ้นกับค่าผลรางวัลเฉลี่ยสะสมในอดีต ทั้งนี้เนื่องจากระบบที่เสนอมีความแปรปรวนน้อยมาก จึงไม่มีความจำเป็นต้องใช้ค่าผลรางวัลเฉลี่ยสะสมในอดีตเพื่อพิจารณาค่าผลรางวัลของแต่ละการกระทำว่ามีผลรางวัลที่ผิดไปจากสถานการณ์ปัจจุบันอย่างไร โดยค่าอัตราการเรียนรู้ที่ได้จากการทดลองนี้จะนำมาใช้ในการทดลองปรับค่าอัตราการลดทอนการเรียนรู้ต่อไป

4.4.2 การทดลองปรับค่าอัตราการลดทอนการเรียนรู้ (γ)

กำหนดให้ $\alpha = 1$ และ $\epsilon = 0.01$ จะได้ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลที่ค่า $\gamma = 0, 0.65, 0.85, 0.95$ และ 1 ดังรูปที่ 4.8

จากผลการทดลองเห็นได้ว่า ค่าอัตราการลดทอนการเรียนรู้ที่เหมาะสมในการทดลองนี้จะอยู่ในช่วงระหว่าง 0 ถึง 0.65 ทั้งนี้ จากสมการที่ (3.17) เห็นได้ว่าการเพิ่มอัตราการลดทอนการเรียนรู้จะเป็นการเพิ่มความสำคัญให้กับการกระทำที่ได้จากค่าผลรางวัลเฉลี่ยสะสมในอดีต การกำหนดค่าอัตราการลดทอนการเรียนรู้สูงเกินไปจึงเป็นการตัดสินใจโดยพิจารณาข้อมูลเก่าเป็นหลักแทนที่จะพิจารณาข้อมูลปัจจุบัน และการพิจารณาข้อมูลเก่าจะเป็นการลดอัตราการเรียนรู้ลง ทำให้ระบบลู่เข้าสู่การ



รูปที่ 4.8: ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลที่ค่า $\alpha = 1$, และ $\epsilon = 0.01$

กระทำที่เหมาะสมได้ช้าลงด้วย ค่าอัตราการเรียนรู้ที่ได้จากการทดลองนี้จะนำมาใช้ในการทดลองปรับค่าอัตราการเรียนรู้ต่อไป

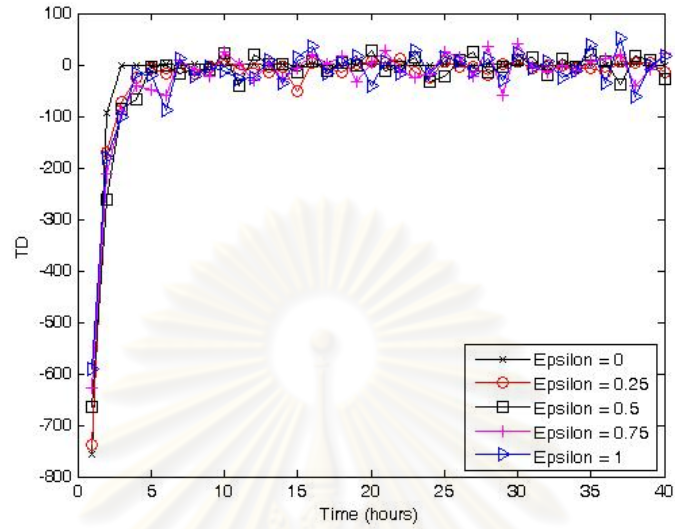
4.4.3 การทดลองปรับค่าอัตราการเรียนรู้ (ϵ)

กำหนดให้ $\alpha = 1$ และ $\gamma = 0.65$ จะได้ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลที่ค่า $\epsilon = 0, 0.25, 0.5, 0.75$ และ 1 ดังรูปที่ 4.9

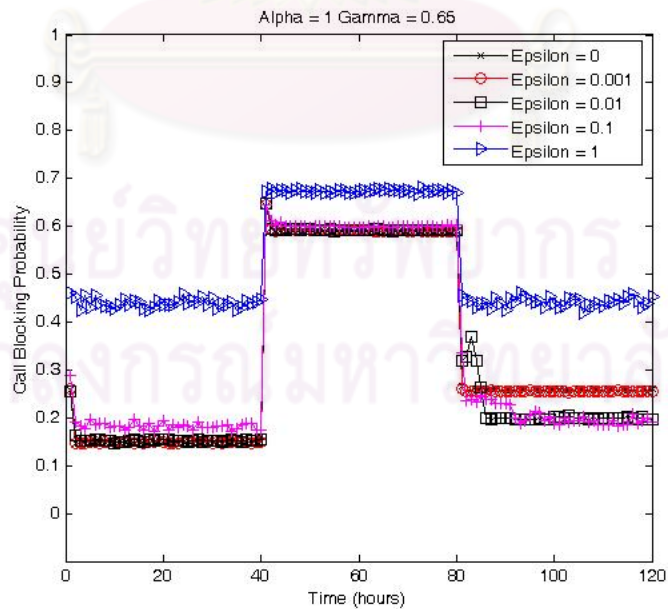
จากผลการทดลองเห็นได้ว่า ค่าอัตราการเรียนรู้ที่เหมาะสมที่สุดในการทดลองนี้มีค่าเท่ากับ 0 ทั้งนี้เนื่องจากค่าผลรางวัลในงานวิจัยนี้มีการปรับค่าทุก ๆ ช่วงเวลาและเป็นค่าที่ติดลบเสมอ (negative reward) และในงานวิจัยนี้ใช้วิธีการเลือกการกระทำแบบเอพซีรอนกรีดี ซึ่งเป็นการหาค่าสูงสุดของทุก ๆ การกระทำในแต่ละสถานะ ในช่วงต้นของการเรียนรู้จึงมีการทดลองเลือกการกระทำทุกการกระทำโดยอัตโนมัติ ทำให้มีการบันทึกค่าของทุก ๆ การกระทำในแต่ละสถานะโดยไม่จำเป็นต้องใช้ค่าอัตราการเรียนรู้ และเนื่องจากระบบที่เสนอมีความแปรปรวนน้อยมาก จึงไม่จำเป็นต้องเลือกการกระทำอื่นเพื่อหาการกระทำที่ให้ผลดีกว่าในช่วงเวลาต่อ ๆ ไป

ในขณะเดียวกัน หากแบบจำลองมีเหตุการณ์โครงข่ายล่มเหมือนการทดลองในหัวข้อที่ 4.3 (ผลการควบคุมการเลือกโครงข่ายด้วยวิธีการเรียนรู้แบบเสริมแรงโดยมีการจำลองเหตุการณ์โครงข่ายล่ม) โดยจะเพิ่มจำนวนช่วงเวลาทำการจำลองเป็น 86,400 ช่วงเวลา (120 ชั่วโมง) ซึ่งในการทดลองนี้จะใช้ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่เป็นตัวชี้วัดสมรรถนะของระบบแทน เนื่องจากค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลจะแสดงการเปลี่ยนแปลงที่เกิดจากเหตุการณ์โครงข่ายล่มไม่ชัดเจน จากพารามิเตอร์ที่กำหนดไว้ข้างต้นจะได้ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่ค่า $\epsilon = 0, 0.001, 0.01, 0.1$ และ 1 ดังรูปที่ 4.10

จากผลการทดลองเห็นได้ว่า เมื่อมีเหตุการณ์โครงข่ายล่ม ที่ค่าอัตราการเรียนรู้เท่ากับ 0 จะไม่ใช่ค่าอัตราการเรียนรู้ที่เหมาะสมที่สุด ค่าที่เหมาะสมจะเป็นค่าที่มากกว่า 0 เล็กน้อย ทั้งนี้เนื่องจากที่ค่าอัตรา



รูปที่ 4.9: ค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัลที่ค่า $\alpha = 1$ และ $\gamma = 0.65$



รูปที่ 4.10: ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่ค่า $\alpha = 1$ และ $\gamma = 0.65$

สำรวจเท่ากับ 0 จะทำให้ระบบไม่มีการทดลองเลือกการกระทำอื่น ซึ่งอาจทำให้ระบบลู่เข้าสู่ค่าที่ไม่เหมาะสม ในขณะที่เดียวกัน หากเลือกค่าอัตราสำรวจที่สูงเกินไป ก็จะมีผลทำให้ระบบไม่มีเสถียรภาพ เนื่องจากการทดลองเลือกการกระทำอื่นอยู่เรื่อย ๆ ตลอดเวลา

การทดลองในหัวข้อนี้เป็นตัวอย่างของการปรับค่าพารามิเตอร์ที่เหมาะสมในสภาพแวดล้อมของระบบที่ได้กำหนดมาข้างต้น ซึ่งในการทดลองอื่น ๆ ที่มีสภาพแวดล้อมของระบบแตกต่างกันอาจต้องออกแบบวิธีการปรับพารามิเตอร์เหมือนหรือแตกต่างกันออกไป โดยขึ้นอยู่กับความซับซ้อนของระบบและสมมติฐานของแบบจำลอง

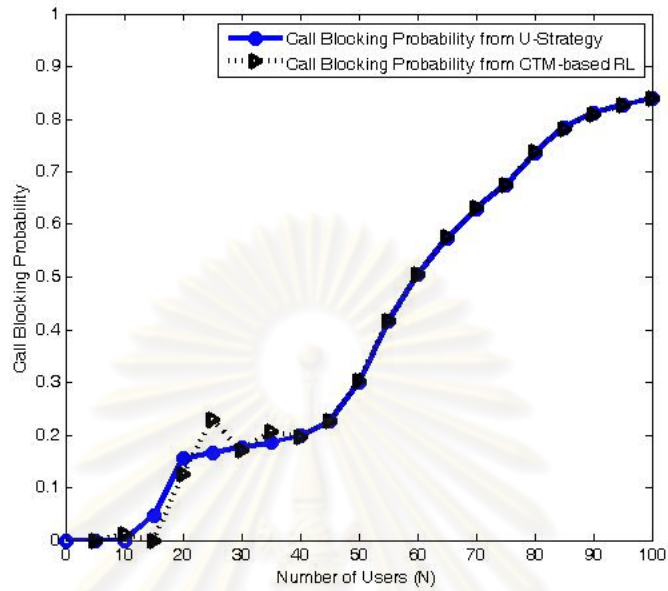
4.5 การเปรียบเทียบกับวิธีการจัดสรรแบนด์วิดท์โดยใช้กลยุทธ์ U

ในการทดลองนี้จะนำวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ที่ได้ทำการปรับค่าพารามิเตอร์ของวิธีการเรียนรู้แบบเสริมแรงจนได้ค่าพารามิเตอร์ที่เหมาะสมแล้ว มาเปรียบเทียบกับวิธีการจัดสรรแบนด์วิดท์โดยใช้กลยุทธ์ผู้ใช้กำหนดเอง (user preference, U) [24] ซึ่งเป็นการจัดสรรแบนด์วิดท์โดยกำหนดค่าน้ำหนักคงที่ตลอดการจำลอง โดยค่าน้ำหนักดังกล่าวจะแปรผันโดยตรงกับอัตราส่วนของแบนด์วิดท์สูงสุดของโครงข่าย การทดลองนี้ได้กำหนดค่าพารามิเตอร์ของระบบไว้ดังนี้ [24]

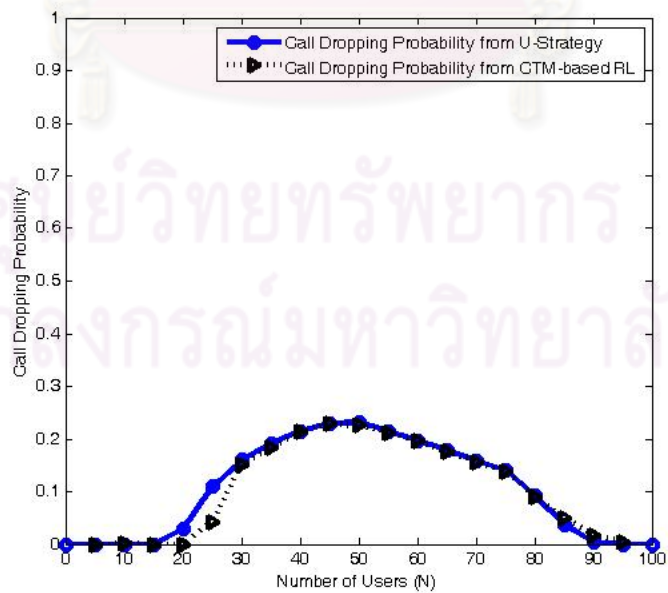
- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 100 ช่องเวลา (500 วินาที)
- ให้อัลกอริทึมของวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ได้เรียนรู้ระบบโดยการวนซ้ำ 39 รอบ และทดลองใช้จริงในรอบที่ 40
- จำนวนรถยนต์ทุก ๆ เซลล์ i ของ CTM ในทุกช่องเวลา t มีค่าคงที่ ($n_i(t) = \text{ค่าคงที่}, \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแมโครเซลล์เท่ากับ 5 ช่องสัญญาณ ($B_{max}^0 = 5$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ที่อยู่เซลล์ที่ 1 ของ CTM เท่ากับ 55 ช่องสัญญาณ ($B_{max}^1 = 55$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ที่อยู่เซลล์ที่ N ของ CTM เท่ากับ 5 ช่องสัญญาณ ($B_{max}^N = 5$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่องเวลา t เท่ากับ 0.1 ($P_A(t) = 0.1, \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่องเวลา t เท่ากับ 0.1 ($P_D(t) = 0.1, \forall t$)
- $\alpha = 0.89, \gamma = 0.001, \epsilon = 0.01$ และ $\beta = 0.4$

จากค่าพารามิเตอร์ข้างต้น เมื่อทดลองกับสภาพจราจรที่แตกต่างกัน โดยกำหนดจากจำนวนยานพาหนะคงที่ตั้งแต่ $n_i(t) = 0, 5, 10, \dots, 90, 95, 100$ จะได้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธแฮนด์ออฟดิงรูปที่ 4.11 และรูปที่ 4.12 ตามลำดับ

ในสภาพการจราจรที่มีจำนวนรถยนต์คงที่ จะส่งผลให้มีรูปแบบการเรียกเข้าใช้โครงข่ายคงที่ตลอดการจำลอง ทำให้ทั้งสองกลยุทธ์ที่ได้นำเสนอมานั้นให้ผลการทดลองออกมาใกล้เคียงกัน



รูปที่ 4.11: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่จำนวนยานพาหนะต่าง ๆ



รูปที่ 4.12: ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่จำนวนยานพาหนะต่าง ๆ

4.6 การเปรียบเทียบ กับ วิธีการ จัดสรร แบนด์วิดท์ แบบ แบ่ง น้ำหนัก ตาม สัดส่วน (proportional fairness)

ในการทดลองนี้จะนำวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ที่ได้ทำการปรับค่าพารามิเตอร์ของวิธีการเรียนรู้แบบเสริมแรงจนได้ค่าพารามิเตอร์ที่เหมาะสมแล้ว มาเปรียบเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน โดยอยู่บนพื้นฐานของวิธีการจัดสรรแบนด์วิดท์โดยใช้กลยุทธ์ผู้ใช้งานกำหนดเอง [24] ซึ่งวิธีดังกล่าวนี้จะแบ่งน้ำหนักเป็นสัดส่วนโดยตรงกับแบนด์วิดท์ที่เหลืออยู่ในแต่ละโครงข่าย โดยการทดลองนี้จะแบ่งการทดลองออกเป็น 3 การทดลองย่อย ๆ ซึ่งมีสภาพแวดล้อมที่แตกต่างกันออกไป

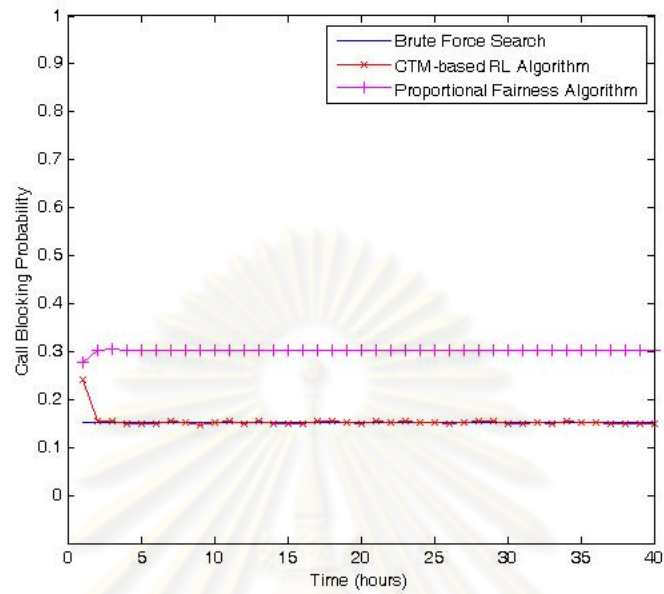
4.6.1 กรณีที่มีความจุของโครงข่ายแม่โครเซลล์สูง

ได้กำหนดพารามิเตอร์ต่าง ๆ ดังนี้

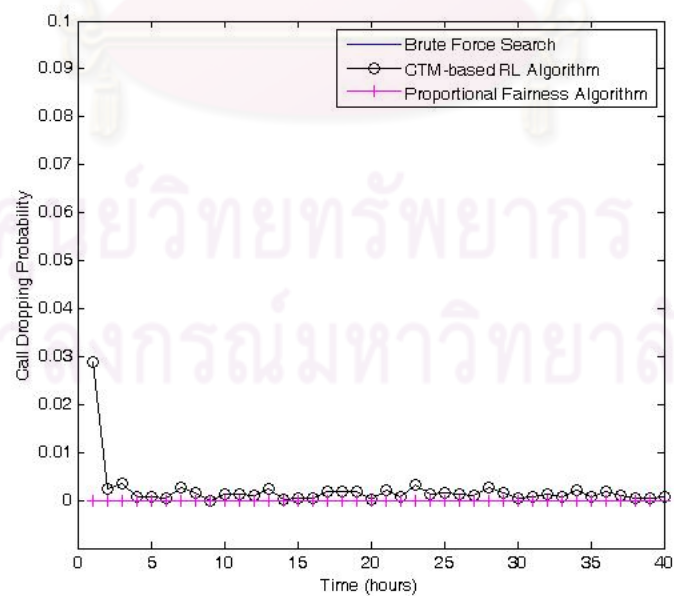
- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 28,800 ช่องเวลา (40 ชั่วโมง)
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซองแบบวิยุตที่มีค่าพารามิเตอร์ = 60 คั่นต่อช่องเวลา ($y_0(t) = \text{Poisson}(60), \forall t$)
- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่องเวลา t เท่ากับ 60 คั่นต่อช่องเวลา ($q_i(t) = 60, \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 100 ช่องสัญญาณ ($B_{max}^0 = 100$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่องเวลา t เท่ากับ 0.4 ($P_A(t) = 0.4, \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่องเวลา t เท่ากับ 0.1 ($P_D(t) = 0.1, \forall t$)
- $\alpha = 0.67, \gamma = 0.084, \epsilon = 0.01$ และ $\beta = 0.09$ (ค่าพารามิเตอร์เหล่านี้เป็นค่าพารามิเตอร์ที่ถูกปรับให้เหมาะสมแล้ว)

จากค่าพารามิเตอร์ข้างต้น จะได้ค่าน้ำหนักที่ทำให้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟดังรูปที่ 4.13 และรูปที่ 4.14 ตามลำดับ

จากผลการทดลองแสดงให้เห็นว่าวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ มีความสามารถในการควบคุมระบบให้อัตราการปฏิเสธการเรียกเข้าใหม่มีค่าน้อยกว่าวิธีแบ่งน้ำหนักตามสัดส่วน แต่ในส่วนของอัตราการปฏิเสธการแฮนด์ออฟ วิธีแบ่งน้ำหนักตามสัดส่วนนั้นทำได้ดีกว่า ทั้งนี้เนื่องจากในสถานะที่มีแบนด์วิดท์เหลือใช้ (undersaturated) การควบคุมการตัดสินใจเลือกใช้โครงข่ายตามสัดส่วนโดยตรงนั้น สามารถจัดสรรให้กับผู้ใช้ที่แฮนด์ออฟได้โดยไม่มีการถูกปฏิเสธแม้แต่คนเดียว เนื่องจากเป็นผู้ใช้ประเภทที่มีลำดับความสำคัญสูงกว่า ในขณะที่วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์พยายามทำให้ระบบมีความสมดุลมาก



รูปที่ 4.13: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่ความจุแม่โครเซลล์สูงเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน



รูปที่ 4.14: ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่ความจุแม่โครเซลล์สูงเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน

ที่สุด จึงพยายามลดอัตราการถูกปฏิเสธการเรียกเข้าให้ต่ำลงมากที่สุดและส่งผลให้เกิดความผิดพลาด (error) ในการลดอัตราการถูกปฏิเสธการแฮนด์ออฟ

4.6.2 กรณีที่มีความจุของโครงข่ายแม่โครเซลล์ต่ำ

ได้กำหนดพารามิเตอร์ต่าง ๆ ดังนี้

- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 28,800 ช่องเวลา (40 ชั่วโมง)
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซองแบบวิฤตที่มีค่าพารามิเตอร์ = 60 คันต่อช่องเวลา ($y_0(t) = \text{Poisson}(60), \forall t$)
- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่องเวลา t เท่ากับ 60 คันต่อช่องเวลา ($q_i(t) = 60, \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 40 ช่องสัญญาณ ($B_{max}^0 = 40$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่องเวลา t เท่ากับ 0.4 ($P_A(t) = 0.4, \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่องเวลา t เท่ากับ 0.1 ($P_D(t) = 0.1, \forall t$)
- $\alpha = 1, \gamma = 0, \epsilon = 0$ และ $\beta = 0.09$ (ค่าพารามิเตอร์เหล่านี้เป็นค่าพารามิเตอร์ที่ถูกปรับให้เหมาะสมแล้ว)

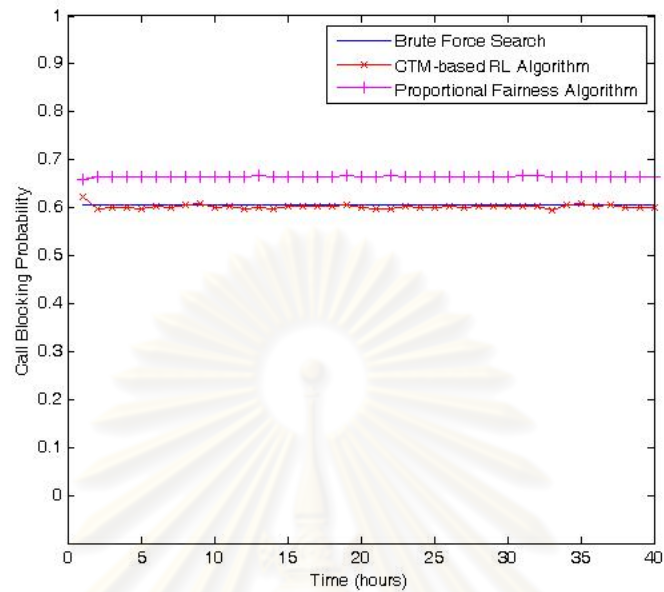
จากค่าพารามิเตอร์ข้างต้น จะได้ค่าน้ำหนักที่ทำให้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟดังรูปที่ 4.15 และรูปที่ 4.16 ตามลำดับ

จากผลการทดลองแสดงให้เห็นว่าวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ มีความสามารถในการควบคุมระบบให้อัตราการปฏิเสธการเรียกเข้าใหม่และอัตราการปฏิเสธการแฮนด์ออฟมีค่าน้อยกว่าวิธีแบ่งน้ำหนักตามสัดส่วน ทั้งนี้เนื่องจากวิธีแบ่งน้ำหนักตามสัดส่วนจะไม่คำนึงถึงพฤติกรรมของผู้ใช้ที่มีการเคลื่อนที่หรือการเรียกใช้อย่างไรในสภาวะที่มีแบนด์วิดท์ไม่เพียงพอกับความต้องการของกลุ่มผู้ใช้ (oversaturated) จึงเป็นข้อจำกัดของวิธีดังกล่าวที่ไม่สามารถจัดสรรแบนด์วิดท์ได้อย่างเหมาะสม

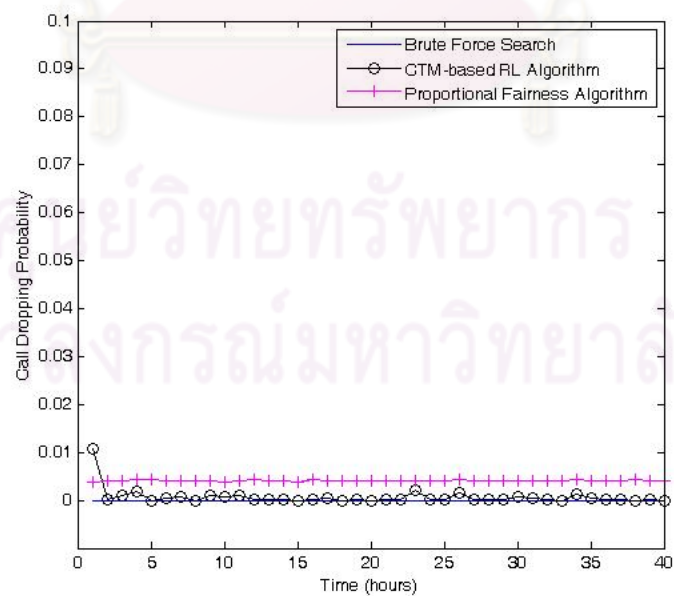
4.6.3 กรณีที่มีการจำลองเหตุการณ์โครงข่ายล่ม

ได้กำหนดพารามิเตอร์ต่าง ๆ ดังนี้

- จำนวนช่องเวลาที่ทำการจำลองเท่ากับ 86,400 ช่องเวลา (120 ชั่วโมง)
- จำนวนรถยนต์ที่เคลื่อนที่เข้าสู่เซลล์ที่ 1 ของ CTM ในทุกช่องเวลา t คำนวณจากการสุ่มปัวซองแบบวิฤตที่มีค่าพารามิเตอร์ = 60 คันต่อช่องเวลา ($y_0(t) = \text{Poisson}(60), \forall t$)



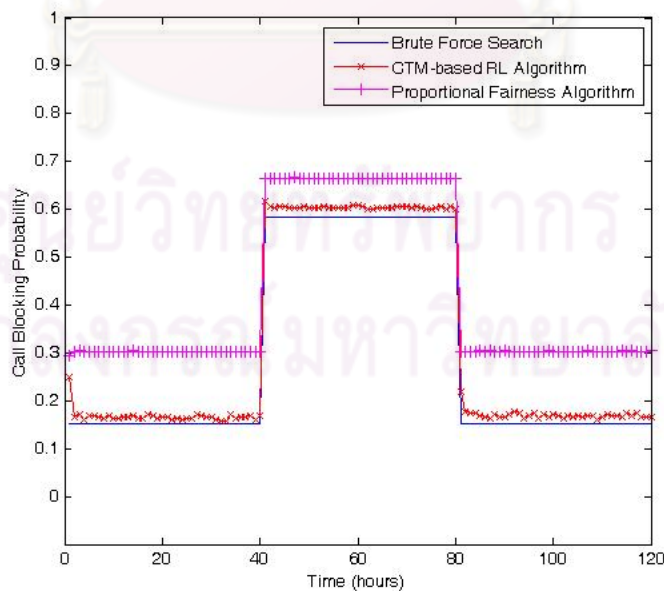
รูปที่ 4.15: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ที่ความจุแม่โครเซลล์ต่ำเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน



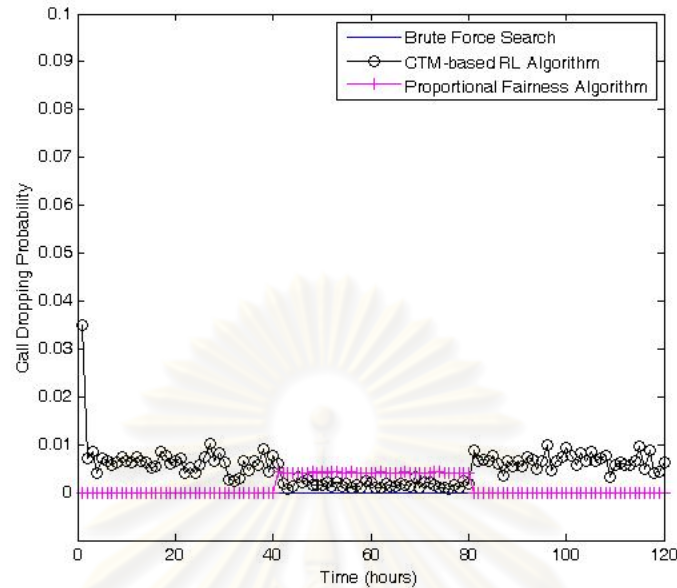
รูปที่ 4.16: ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟที่ความจุแม่โครเซลล์ต่ำเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน

- จำนวนยานพาหนะสูงสุดที่สามารถเคลื่อนที่ออกจากเซลล์ที่ i ของ CTM ได้ ในทุกช่วงเวลา t เท่ากับ 60 คันต่อช่วงเวลา ($q_i(t) = 60, \forall i \forall t$)
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 100 ช่องสัญญาณ ในช่วงเวลาที่ 1 - 40 และ ช่วงเวลาที่ 81 - 120 หรือในช่วงเวลาที่ 1 - 28,800 และช่วงเวลาที่ 57,601 - 86,400 ($B_{max}^0(t) = 100, t \in [1, 28, 800] \cup [57, 601, 86, 400]$)
- แบนด์วิดท์สูงสุดของโครงข่ายแม่โครเซลล์เท่ากับ 40 ช่องสัญญาณ ในช่วงเวลาที่ 41 - 80 หรือ ในช่วงเวลาที่ 28,801 - 57,600 ($B_{max}^0(t) = 40, t \in [28, 801, 57, 600]$)
- แบนด์วิดท์สูงสุดของโครงข่ายไมโครเซลล์ทุกโครงข่ายเท่ากับ 20 ช่องสัญญาณ ($B_{max}^1 = B_{max}^N = 20$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะไม่ทำงานจะเรียกเข้าใหม่มาในระบบในทุกช่วงเวลา t เท่ากับ 0.4 ($P_A(t) = 0.4, \forall t$)
- ความน่าจะเป็นที่ผู้ใช้ที่อยู่ในสถานะทำงานจะเสร็จสิ้นการติดต่อกับโครงข่ายในทุกช่วงเวลา t เท่ากับ 0.1 ($P_D(t) = 0.1, \forall t$)
- $\alpha = 1, \gamma = 0, \epsilon = 0.05$ และ $\beta = 0.09$ (ค่าพารามิเตอร์เหล่านี้เป็นค่าพารามิเตอร์ที่ถูกปรับให้เหมาะสมแล้ว)

จากค่าพารามิเตอร์ข้างต้น จะได้ค่าน้ำหนักที่ทำให้ค่าความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟตั้งรูปที่ 4.17 และรูปที่ 4.18 ตามลำดับ



รูปที่ 4.17: ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่เมื่อมีการจำลองเหตุการณ์โครงข่ายล่มเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน



รูปที่ 4.18: ความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟเมื่อมีการจำลองเหตุการณ์โครงข่ายล่มเทียบกับวิธีแบ่งน้ำหนักตามสัดส่วน

เมื่อเกิดเหตุการณ์โครงข่ายล่ม การควบคุมระบบเพื่อให้มีอัตราการปฏิเสธการเรียกเข้าใหม่และอัตราการปฏิเสธการแฮนด์ออฟมีค่าน้อยที่สุดนั้นยังสามารถทำได้ทั้งวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ และวิธีการแบ่งน้ำหนักตามสัดส่วน โดยผลการทดลองที่ได้จะเหมือนการทดลองกรณีที่มีความจุของโครงข่ายแม่โครเซลล์ต่ำในช่วงที่มีเหตุการณ์โครงข่ายล่ม และเหมือนการทดลองกรณีที่มีความจุของโครงข่ายแม่โครเซลล์สูงในช่วงปกติ

4.7 สรุป

จากผลการทดลองทั้งหมด เห็นได้ว่าวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ สามารถเลือกโครงข่ายที่เหมาะสมได้ในทุกผลการทดลองที่ได้นำเสนอ ทั้งยังเป็นอัลกอริทึมที่มีการปรับข้อมูลให้เข้ากับระบบในปัจจุบันอยู่ตลอดเวลา จึงไม่ต้องการใช้หน่วยความจำในการสะสมข้อมูล ซึ่งมีผลทำให้เกิดความซับซ้อนของระบบน้อย โดยความซับซ้อนของการคำนวณต่อหนึ่งวงรอบจะอยู่ที่ $O(N\eta^{(\kappa-1)})$ เท่านั้น โดยที่ κ คือจำนวนเซลล์ของโครงข่าย และสามารถนำมาประยุกต์ใช้ได้ในทางปฏิบัติเนื่องจากเป็นอัลกอริทึมที่มีความยืดหยุ่นมากพอเพื่อจะรองรับสภาพแวดล้อมที่มีการเปลี่ยนแปลงตลอดเวลา อย่างไรก็ตาม ในสถานะที่มีแบนด์วิดท์เหลือใช้ วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์พยายามทำให้ระบบมีความสมดุลมากที่สุด จึงพยายามลดอัตราการปฏิเสธการเรียกเข้าให้ต่ำลงมากที่สุดและส่งผลให้เกิดความผิดพลาด (error) ในการลดอัตราการปฏิเสธการแฮนด์ออฟ ส่งผลให้สมรรถนะในการแฮนด์ออฟของการใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ น้อยกว่าสมรรถนะในการแฮนด์ออฟของการใช้วิธีแบ่งน้ำหนักตามสัดส่วน

วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ที่ได้นำเสนอใน

วิทยานิพนธ์ฉบับนี้ ได้ปรับให้พิจารณาผู้ใช้ที่ต้องการแฮนด์ออฟก่อนที่จะพิจารณาผู้ใช้ที่ต้องการเรียกเข้าใหม่ การทดลองทั้งหมดจึงปรับให้การปฏิเสธการแฮนด์ออฟมีผลต่อค่าผลรางวัลมากกว่าการปฏิเสธการเรียกเข้าใหม่เพื่อให้สอดคล้องกับการตัดสินใจเลือกโครงข่ายในทางปฏิบัติ



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

วิทยานิพนธ์ฉบับนี้ได้ศึกษาระบบการแฮนด์ออฟและการตัดสินใจเลือกโครงข่ายไร้สาย โดยสภาพแวดล้อมที่พิจารณาจะเป็นผู้ใช้ที่อยู่บนท้องถนนที่สามารถเคลื่อนที่ได้ 1 ทิศทาง และประกอบไปด้วยโครงข่ายสื่อสารไร้สาย 2 ประเภท คือ โครงข่ายไมโครเซลล์ และโครงข่ายแมโครเซลล์

วิทยานิพนธ์ฉบับนี้ได้ศึกษาวิธีการเรียนรู้แบบเสริมแรง ซึ่งเป็น อัลกอริทึมที่มีความยืดหยุ่นมากพอที่จะรองรับสภาพแวดล้อมที่ไม่แน่นอนของโครงข่ายเทคโนโลยีไร้สาย เนื่องจากในทางปฏิบัติสภาพแวดล้อมของโครงข่ายเทคโนโลยีไร้สายจะมีการเปลี่ยนแปลงตลอดเวลาและไม่สามารถทำนายการเปลี่ยนแปลงที่เกิดขึ้นในอนาคตได้

วิทยานิพนธ์ฉบับนี้ได้เสนอการนำวิธีการเรียนรู้แบบเสริมแรงมาประยุกต์ใช้ในการปรับปรุกระบวนการตัดสินใจเลือกโครงข่ายที่เหมาะสมบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ ซึ่งเป็นแบบจำลองที่สามารถคำนวณจำนวนผู้ใช้ในถนนในช่วงเวลาใด ๆ ได้และมีการเปรียบเทียบความถูกต้องกับข้อมูลจริง [22] โดยวิทยานิพนธ์ฉบับนี้ได้ นำแบบจำลองดังกล่าวมาเป็นแบบจำลองการเคลื่อนที่ของผู้ใช้บนถนน ที่เป็นโครงข่ายไร้สายแบบวิวิธพันธุ์ เพื่อคำนวณสมรรถนะของระบบในกลยุทธ์การเลือกโครงข่ายต่าง ๆ กัน นอกจากนี้ การทดลองในวิทยานิพนธ์ที่เสนอยังกำหนดให้มีการเปลี่ยนแปลงของระบบแตกต่างกัน เพื่อทดสอบการทำงานของวิธีการเรียนรู้แบบเสริมแรงในสภาพแวดล้อมที่แตกต่างกันออกไป ซึ่งผลการทดลองดังกล่าวนี้ แสดงให้เห็นว่าวิธีการเรียนรู้แบบเสริมแรงนั้นสามารถเลือกโครงข่ายที่เหมาะสมได้ และสามารถปรับปรุการเลือกโครงข่ายได้โดยอัตโนมัติหากสภาพแวดล้อมของระบบเปลี่ยนไปอย่างกะทันหัน

วิทยานิพนธ์ฉบับนี้ได้พิจารณาสมรรถนะของการนำวิธีการเรียนรู้แบบเสริมแรงมาประยุกต์ใช้ในการตัดสินใจเลือกโครงข่ายไร้สายแบบวิวิธพันธุ์ โดยการใช้ตัวชี้วัดสมรรถนะของระบบ ซึ่งมีทั้งหมด 2 ตัวชี้วัด คือ ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใช้โครงข่าย (ความน่าจะเป็นของการปฏิเสธการเรียกเข้าใหม่ และความน่าจะเป็นของการปฏิเสธการแฮนด์ออฟ) กับค่าความแตกต่างในแต่ละช่วงเวลาของค่าผลรางวัล เมื่อคำนวณค่าตัวชี้วัดสมรรถนะของระบบได้แล้ว จะนำไปวิเคราะห์สมรรถนะของระบบเมื่อใช้วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ เปรียบเทียบกับวิธีอื่นซึ่งเป็นวิธีที่งานวิจัยในอดีต [24] ได้ใช้เป็นกลยุทธ์ในการตัดสินใจเลือกโครงข่าย

จากการพิจารณาสมรรถนะของระบบเมื่อนำวิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์มาประยุกต์ใช้ พบว่าวิธีดังกล่าวสามารถควบคุมการตัดสินใจเลือกโครงข่ายจนลดการปฏิเสธการเรียกเข้าใหม่ และการปฏิเสธการแฮนด์ออฟให้น้อยลงได้อย่างเหมาะสม อย่างไรก็ตาม หากเป็นกรณีที่แต่ละโครงข่ายมีความจุสูงมาก วิธีการเรียนรู้แบบเสริมแรงบนพื้นฐานของแบบจำลองการส่งผ่านเซลล์ จะทำได้ไม่ดีเท่าวิธีการแบ่งน้ำหนักตามสัดส่วนในส่วนของควบคุมอัตราการใช้การปฏิเสธการแฮนด์ออฟ

5.2 ข้อเสนอแนะ

หัวข้อที่ควรศึกษาและวิจัยต่อไปในอนาคตคือ

1. การเพิ่มทิศทางการเคลื่อนที่ เพิ่มระบบสัญญาณไฟจราจร และเพิ่มรายละเอียดของแบบจำลอง ในวิทยานิพนธ์ฉบับนี้พิจารณาท้องถนนที่สามารถเคลื่อนที่ได้ 1 ทิศทาง และไม่มีระบบสัญญาณไฟจราจร โดยในทางปฏิบัติ การเพิ่มทิศทางการเคลื่อนที่ในแบบจำลอง การเพิ่มระบบสัญญาณไฟจราจรในแบบจำลอง หรือการพิจารณาประเภทของยานพาหนะในแบบจำลอง จะทำให้งานวิจัยนี้มีแบบจำลองที่ใกล้เคียงกับโครงข่ายในทางปฏิบัติมากยิ่งขึ้น
2. การเพิ่มประเภทและจำนวนของโครงข่ายในระบบ
แบบจำลองที่เสนอ ในวิทยานิพนธ์ฉบับนี้ ประกอบไปด้วยโครงข่ายสื่อสารไร้สายอยู่เพียง 2 ประเภท คือ โครงข่ายไมโครเซลล์ และโครงข่ายแมโครเซลล์
ในสถานที่จริง เช่น ในกรุงเทพมหานครและปริมณฑล มีจำนวนโครงข่ายสื่อสารไร้สายที่ใช้ทำงานได้จริงอยู่หลายประเภท ในพื้นที่หนึ่ง ๆ อาจมีโครงข่ายไร้สายประเภทเดียวกันและมีโครงข่ายต่างประเภทซ้อนทับกันอยู่หลายโครงข่าย แบบจำลองที่พัฒนาต่อไปในอนาคต จึงควรเพิ่มประเภทและจำนวนของโครงข่ายสื่อสารไร้สาย เพื่อสามารถนำแบบจำลองมาเปรียบเทียบกับสภาพแวดล้อมในทางปฏิบัติจริงได้ดียิ่งขึ้น
3. การพิจารณาความจุของการแอตท็อก
ในวิทยานิพนธ์ฉบับนี้ กำหนดให้ผู้ใช้ที่ไม่สามารถเชื่อมต่อกับโครงข่ายได้โดยตรง สามารถเชื่อมต่อกับโครงข่ายได้ด้วยวิธีแอตท็อก ซึ่งในทางปฏิบัติ การแอตท็อกของอุปกรณ์สื่อสารจะสามารถแอตท็อกได้ด้วยจำนวนจำกัด ดังนั้น ถึงแม้โครงข่ายที่ต้องการเชื่อมต่อจะมีช่องสัญญาณว่างอยู่ หากไม่มีผู้ใช้ที่อยู่ในพื้นที่ครอบคลุมของโครงข่ายที่สามารถแอตท็อกได้ ผู้ใช้ที่อยู่นอกพื้นที่ครอบคลุมของโครงข่ายจะถูกปฏิเสธการเรียกเข้าใช้โครงข่ายและไม่สามารถเชื่อมต่อกับโครงข่ายนั้น ๆ ได้ ดังนั้น แบบจำลองที่พัฒนาต่อไปในอนาคต จึงควรพิจารณาความจุของการแอตท็อก เพื่อให้แบบจำลองมีความน่าเชื่อถือมากขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

- [1] McNair, J. and Zhu, F., "Vertical Handoffs in Fourth-generation Multinetwork Environment," *IEEE Wireless Communications*, 11, 3, 2004.
- [2] Chen, W., Liu, J. and Huang, H., "An Adaptive Scheme for Vertical Handoff in Wireless Overlay Networks," in *Proc. of ICPADS04*, 2004.
- [3] 3rd Generation Partnership Project (3GPP) [Online], <http://www.3gpp.org>
- [4] 3rd Generation Partnership Project 2 (3GPP2) [Online], <http://www.3gpp2.org>
- [5] IEEE 802.21 Media Independent Handover Working Group [Online], <http://www.ieee802.org/21>
- [6] Murray, K. and Pesch, D., "Call Admission and Handover in Heterogeneous Wireless Networks," *IEEE Internet Computing*, 11, 2, 2007: 44-52.
- [7] Lee, S., Sriram, K., Kim, K., Kim, Y. and Golmie, N., "Vertical Handoff Decision Algorithms for Providing Optimized Performance in Heterogeneous Wireless Networks," *IEEE Transactions on Vehicular Technology*, 58, 2, 2009: 865-881.
- [8] Wu, Y., Yang, K., Zhao, L. and Cheng, X., "Congestion-aware Proactive Vertical Handoff Algorithm in Heterogeneous Wireless Networks," *IET Communications*, 3, 7, 2009: 1103-1114.
- [9] Wu, Y., Yang, K. and Chen, H.-H., "An Adaptive Routing Protocol for Converged Ad Hoc and Cellular Networks," *IEEE/KICS J. Commun. Netw.*, 8, 4, 2006: 422-431.
- [10] Ali, R. B. and Pierre, S., "On the Impact of Mobility and Soft Vertical Handoff on Voice Admission Control in Loosely Coupled 3G/WLAN Networks," *IEEE Communications Letters*, 13, 5, 2009: 303-305.
- [11] Chang, B., Chen, J., Hsieh, C. and Liang, Y., "Markov Decision Process-based Adaptive Vertical Handoff with RSS Prediction in Heterogeneous Wireless Networks," *Wireless Communication and Networking Conference*, 2009: 1-6.
- [12] Singhrova, A. and Prakash, N., "Adaptive Vertical Handoff Decision Algorithm for Wireless Heterogeneous Networks," *High Performance Computing and Communications*, 2009: 476-481.
- [13] Hang, Y., Hsu, C., Gao, F. and Chen, H., "Performance of Adaptive Vertical Handoff in Heterogeneous Networks of WLAN and WCDMA Systems," *INC, IMS and IDC International Joint Conference*, 2009: 2103-2107.

- [14] Marques, L. B. S. and Motoyama, S., “Vertical Handoff Algorithms with QoS Guarantee for 4G Networks,” *IEEE Latin-American Communications Conference*, 2009: 1-6.
- [15] Sutton, R. S. and Barto, A. G., *Reinforcement Learning: An Introduction*. MA: The MIT Press, 1998.
- [16] Niyato, D. and Hossain, E., “Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach,” *IEEE Transactions on Vehicular Technology*, 58, 4, 2009: 2008-2017.
- [17] Xue, Y., Lin, Y., Feng, Z., Cai, H. and Chi, C., “Autonomic Joint Session Scheduling Strategies for Heterogeneous Wireless Networks,” *Wireless Communications and Networking Conference*, 2008: 2045-2050.
- [18] Feng, Z., Tan, L., Li, W. and Gulliver, T. A., “Reinforcement learning based Dynamic Network Self-optimization for heterogeneous networks,” *IEEE Pacific Rim Conference*, 2009: 319-324.
- [19] Mignanti, S., Di Giorgio, A. and Suraci, V., “A Model Based RL Admission Control Algorithm for Next Generation Networks,” *International Conference Networks*, 2009: 191-196.
- [20] Nitiwong, K., Komolkiti, P., and Aswakul, C.. CTM-based evaluation for herogenous wireless network selection strategy. Electrical Engineering/Electronics Computer Telecommunications and Information Technology International Conference (ECTI-CON), 2009.
- [21] Daganzo, C. F., “The Cell Transmission Model, part II: Network Traffic,” *Transportation Research*, 29b, 2, 1995: 79-93.
- [22] Tuerprasert, K., and Aswakul, C. Multiclass cell transmission model for heterogeneous mobility in general topology of road network. Journal of Intelligent Transportation Systems, 14, 2, 2010: 1-15.
- [23] Marivate, V.N., and Marwala, T. Social Learning methods in board game agents. Computational Intelligence and Games (CIG '08), 2008.
- [24] คณิน นิติวงศ์, *แบบจำลองการประเมินสมรรถนะกลยุทธ์การเลือกโครงข่ายไร้สายแบบวิวิธพันธุ์ซึ่งมีการเคลื่อนที่ของผู้ใช้แบบมหภาค*. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2009.
- [25] Daganzo, C. F. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research, 28, 4PartB, 1994: 269-287.

- [26] Yang, Q., and Koutsopoulos, H. N. A microscopic traffic simulator for evaluation of dynamic traffic management systems. Transportation Research Part C, 4, 1996: 1-3.
- [27] Lighthill, M. J., and Whitham, G.B. On kinematic waves: II. a theory of traffic flow on long crowded roads. Proceedings of the Royal Society Series A, 229, 1955: 317-345.
- [28] Richards, P. J. Shock waves on the highway. Operations Research 4, 1, 1956: 42-51.
- [29] Munoz, L., Sun, X., Horowitz, R., and Alvarez, L. Traffic density estimation with cell transmission model. Proceedings of the American Control Conference 5, 2003: 3750-3755.



ประวัติผู้เขียนวิทยานิพนธ์

กิตติศักดิ์ กิตติเวทางค์ เกิดเมื่อวันที่ 18 กรกฎาคม พ.ศ. 2529 ที่กรุงเทพมหานคร สำเร็จการศึกษาชั้นมัธยมศึกษาจากโรงเรียนกรุงเทพคริสเตียนวิทยาลัยในปี พ.ศ. 2547 จากนั้นได้เข้าศึกษาต่อที่คณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย จนสำเร็จหลักสูตรวิศวกรรมศาสตรบัณฑิตในปี พ.ศ. 2551 จากนั้นได้เข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต ณ จุฬาลงกรณ์มหาวิทยาลัย จนสำเร็จการศึกษาในปี พ.ศ. 2553

บทความทางวิชาการจากวิทยานิพนธ์

[1] K. Kittiyatang, P. Chanloha and C. Aswakul. CTM-based Reinforcement Learning Strategy for Optimal Heterogeneous Wireless Network Selection. Second International conference on Computational Intelligence, Modelling and Simulation (CIMSIm) 2010.



ศูนย์วิทยพัชการ
จุฬาลงกรณ์มหาวิทยาลัย