

Reinforcement Learning for BESS SARSA

5 June 2025

Part of the Scada Automation AI Initiative

วัตถุประสงค์

- ใ้ด้นี้เป็นการจำลองระบบการจัดการพลังงานด้วย แบตเตอรี่เก็บพลังงาน (Battery Energy Storage System - BESS) โดยใช้เทคนิค **Reinforcement Learning แบบ SARSA (State-Action-Reward-State-Action)** เพื่อ
- 1 ลดภาระโหลดจากกริด (Peak Shaving)
- 2 ลดค่าไฟฟ้า โดยเลือกเวลาในการชาร์จและปล่อยพลังงานจากแบตเตอรี่อย่างเหมาะสม แบ่งเป็น ชาร์จ, ดิสชาร์จ และ ไม่ทำอะไร



จุดเด่นเชิงวิศวกรรม

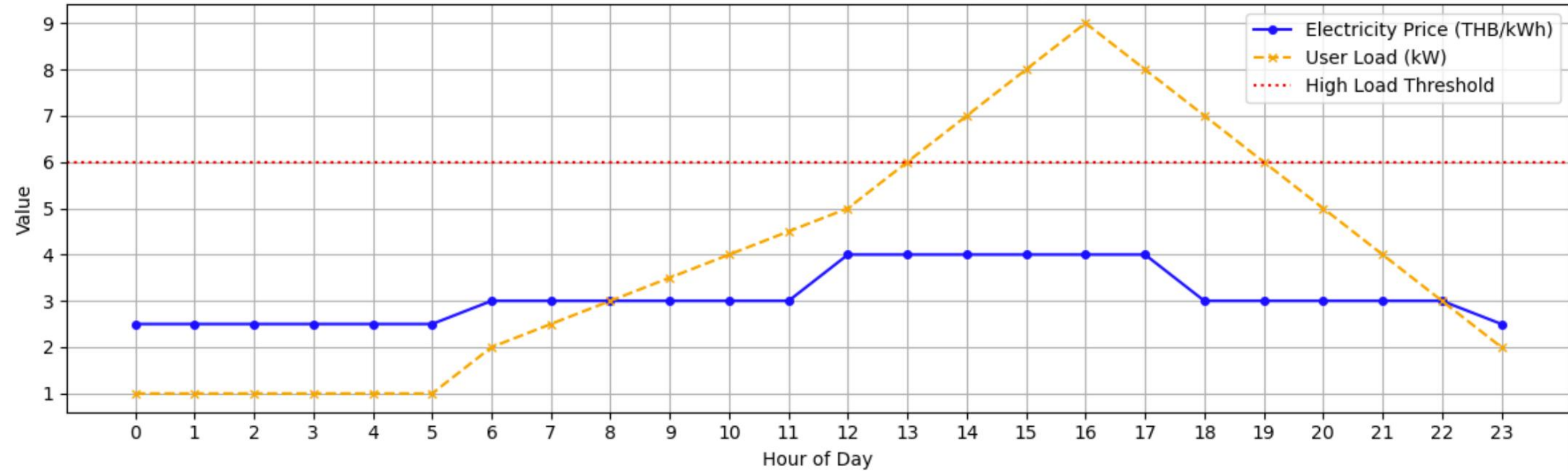
- ใช้ **SARSA** ซึ่งเป็น on-policy learning → ตอบสนองต่อ environment จริง
- มีการให้ **penalty** แบบมีน้ำหนัก ตามสถานการณ์
- ระบบถูกออกแบบมาให้ **ใช้งานง่ายและอธิบายได้ทางวิศวกรรมพลังงาน**

หัวข้อ	Q-learning	SARSA (State–Action–Reward–State–Action)
ประเภทของการเรียนรู้	Off-policy (เรียนจากนโยบายที่ "ดีที่สุด")	On-policy (เรียนจากนโยบายที่ "ใช้อยู่ตอนนั้น")
การอัปเดตค่า Q	ใช้ค่าที่ ดีที่สุด ของ action ถัดไป	ใช้ action ถัดไปที่ เลือกจริงๆ ตามนโยบาย
สูตรอัปเดต	$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma * \max_{a'} Q(s', a') - Q(s, a)]$	$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma * Q(s', a') - Q(s, a)]$
การเรียนรู้จากการทดลอง	มองโลกในแง่ดี – สมมติว่าตัดสินใจดีที่สุดในอนาคต	เรียนรู้จากการกระทำจริง – อาจไม่ใช่ดีที่สุด
การทำงานใน สภาพแวดล้อมสุ่ม (stochastic)	อาจเสี่ยงกว่า เพราะไม่ได้ดูการกระทำจริง	ปลอดภัยกว่า เพราะตามพฤติกรรมที่ใช้จริง
การนำไปใช้ (policy)	อัปเดตค่านโยบาย “greedy” ทันทีแม้จะไม่เคยใช้	อัปเดตจากนโยบายที่ “ใช้จริง” ขณะเรียนรู้
ข้อดี	เรียนรู้เร็ว ถ้าสภาพแวดล้อมไม่สุ่มมาก	ปลอดภัยกว่าในสถานการณ์ที่มีความไม่แน่นอนสูง
ข้อเสีย	อาจ overestimate Q ถ้าสภาพแวดล้อมไม่แน่นอน	ช้ากว่า Q-learning ในการหานโยบายดีที่สุด
ตัวอย่างแนวคิด	"ถ้าฉันจะทำให้ดีที่สุด ฉันควรทำอะไร?"	"ฉันเลือกทำแบบนี้ ฉันจะเรียนจากผลของมัน"

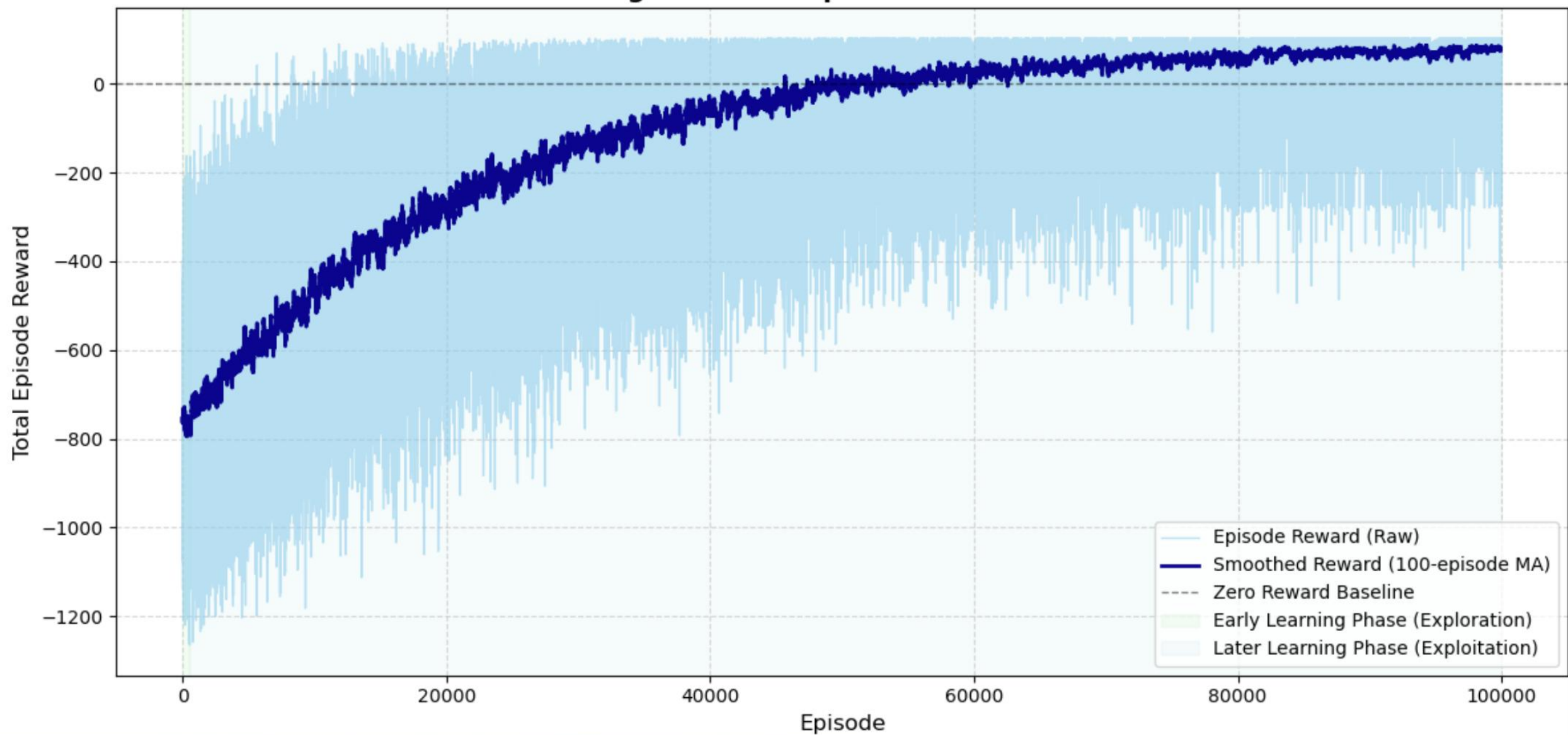
✨ สรุปง่าย ๆ:

- **Q-learning** มองไปข้างหน้าแบบ มองโลกในแง่ดี (คาดหวังว่าจะเลือก action ที่ดีที่สุด)
- **SARSA** เรียนรู้จาก สิ่งที่ทำจริง ในตอนนั้น (รวมความไม่แน่นอนด้วย)
- ถ้าทำระบบที่มี **ความไม่แน่นอนสูงหรือ noisy** เช่น พลังงานหรือโหลดผันผวนมาก — **SARSA** อาจให้ผลที่เสถียรกว่า
- แต่ถ้าระบบ **ค่อนข้างแน่นอนและต้องการเรียนรู้เร็ว** — **Q-learning** ก็เป็นตัวเลือกที่ดี

Electricity Price & User Load Profile (Hourly)

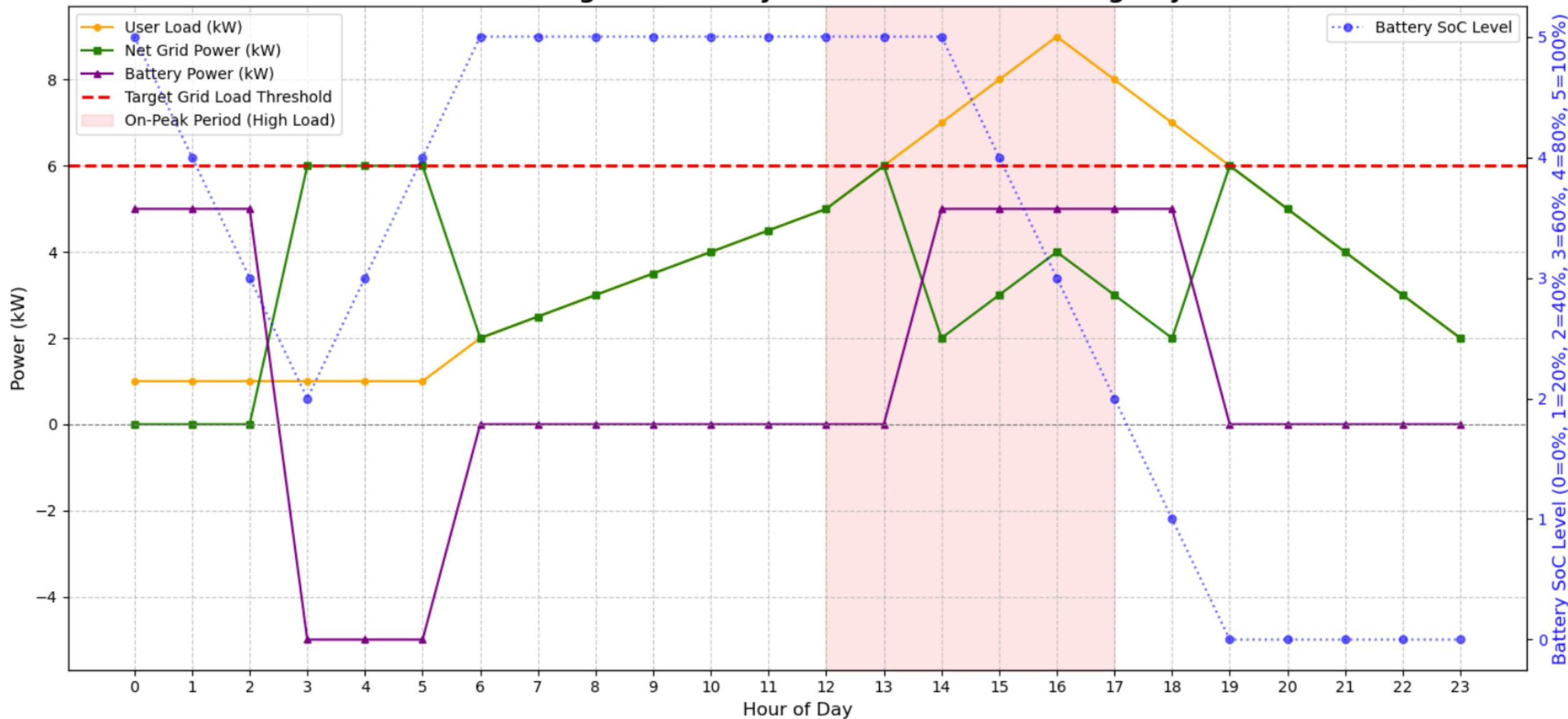


SARSA Progress: Total Episode Rewards Over Time



Q-Learning

Power Flow Throughout the Day with Strict Peak Shaving Objective



SARSA

Power Flow Throughout the Day with Strict Peak Shaving Objective (SARSA)

