CERIAS

The Center for Education and Research in Information Assurance and Security

Adversarial Booking Attack for Autonomous On-demand Mobility Services

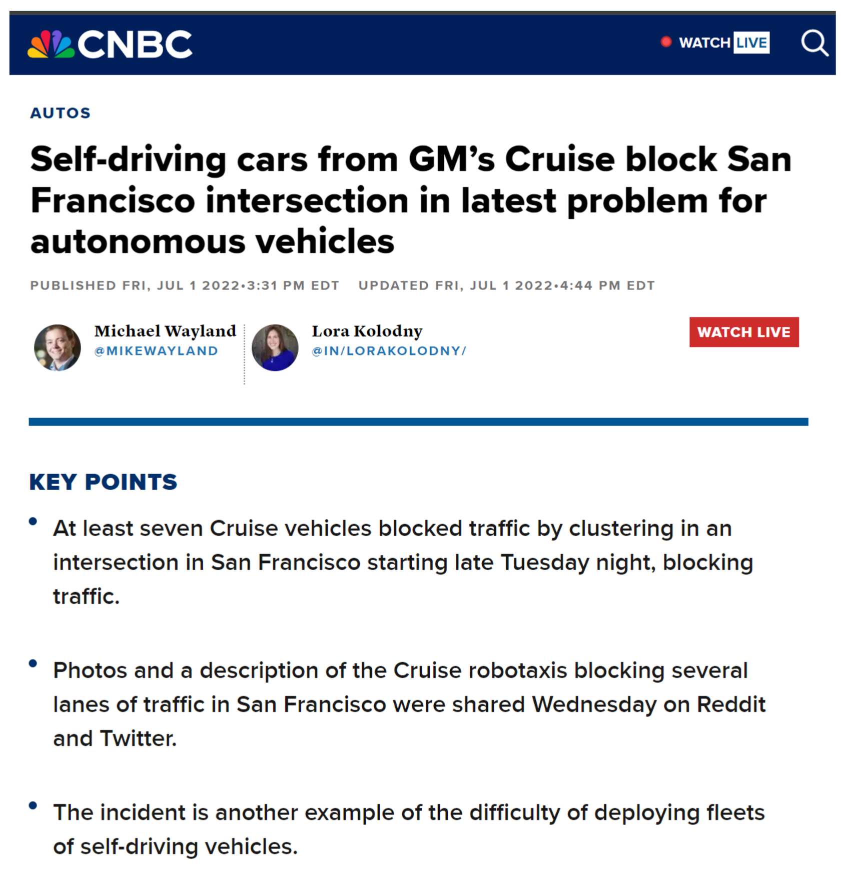
Zengxiang Lei [1], Satish V. Ukkusuri [1]

[1] Lyles School of Civil Engineering, Purdue University

Urban Mobility, Networks and Intelligence Lab

Background

- On-demand mobility services (OMS) provide online vehicle scheduling and routing instructions.
- More controls are expected to be added, e.g., autonomous driving, V2X coordination.
- Little is known about the vulnerabilities and associated risks of these controls. though there is real-world lessons from robo-taxis pilots.



CNBC WATCH LIVE

AUTOS

Self-driving cars from GM's Cruise block San Francisco intersection in latest problem for autonomous vehicles

PUBLISHED FRI, JUL 1 2022•3:31 PM EDT UPDATED FRI, JUL 1 2022•4:44 PM EDT

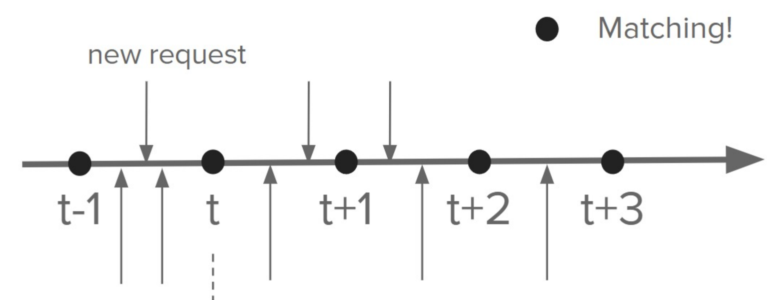
Michael Wayland @MIKEWAYLAND | Lora Kolodny @IN/LORAKOLODNY/ WATCH LIVE

KEY POINTS

- At least seven Cruise vehicles blocked traffic by clustering in an intersection in San Francisco starting late Tuesday night, blocking traffic.
- Photos and a description of the Cruise robotaxis blocking several lanes of traffic in San Francisco were shared Wednesday on Reddit and Twitter.
- The incident is another example of the difficulty of deploying fleets of self-driving vehicles.

Preliminaries

- We focus on a vehicle-passenger matching algorithm, the batch matching, used by major ride-hailing service providers such as Uber and Didi.
- Input: vehicle j and passenger i matching weights c_{ij} within certain time windows.
- Output: matched vehicle-passenger pairs represented by indicators x_{ij}.



$$\max \sum_{i=1}^{|V^t|} \sum_{j=1}^{|O^t|} c^0_{ij} x^t_{ij}$$

$$\text{s.t.} \sum_{j=1}^{|O^t|} x^t_{ij} \le 1, \forall i$$

$$\sum_{i=1}^{|V^t|} x^t_{ij} \le 1, \forall j$$

$$x^t_{ij} \in \{0,1\}, \forall i,j$$

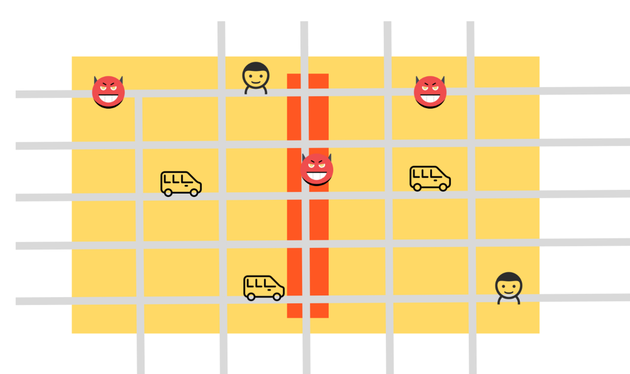
Adversarial booking attack

The threat model:

- The attacker controls K compromised accounts that can send requests with customizable coordinates.
- The attacker will send the requests then cancel it after 3 minutes, which is assumed to be the threshold for the service provider to collect cancelling fee.
- The attacker's objective is to disrupt the services by reducing the number of successfully matched passengers and inducing traffic to a congested area.
- The attacker knows: the matching time window, the coordinates of vacant vehicles, and a good approximator to the matching weights.
- The attacker may also know the coordinates of ongoing requests.

A bi-level optimization problem:

- Upper-level: decide the coordinates of fake requests (u_j, v_j) within a polygon.
- Lower-level: batching matching.
- Solve it by reducing to single level.



$$\min \sum_{i=1}^{|V^t|} \sum_{j=1}^{|O^t|} x^t_{ij} - \sum_{j=|O|^t+1}^{|O|^t+K} r_j$$

minimize the successfully matched real requests and maximize the trips through the target (orange) area

$$\max \sum_{i=1}^{|V^t|} \sum_{j=1}^{|O^t|+K} c^1_{ij} x^t_{ij}$$

$$\text{s.t.} \sum_{j=1}^{|O^t|+K} x^t_{ij} \le 1, \forall i$$

$$\sum_{i=1}^{|V^t|} x^t_{ij} \le 1, \forall j$$

$$x^t_{ij} \in \{0,1\}, \forall i,j$$

$$c^1_{ij} = \alpha|u_i - u_j| + \beta|v_i - v_j|, \forall i, \forall j$$

$$A_0 u_j + B_0 v_j + D_0 \le 0, \forall j \in \{|O|^t+1, \ldots, |O|^t+K\}$$

$$A_1 u_j + B_1 v_j + D_1 \le M(1-r_j), \forall j \in \{|O|^t+1, \ldots, |O|^t+K\}$$

$$r^t_j \in \{0,1\}, \forall j \in \{|O|^t+1, \ldots, |O|^t+K\}$$

Numerical experiment

Simulation framework:

- Routing engine (https://github.com/Project-OSRM/osrm-backend) + SUMO (https://sumo.dlr.de/docs/index.html).
- For each matching time window:

1. Information of vacant vehicles are collected from SUMO.
2. The attacker generates the fake requests based on the current vacant vehicles' coordinates and estimated request locations by solving the bi-level optimization problem.
3. The platform solves the batch matching problem with travel time/distance estimated by the routing engine, then updates the states of all requests and vehicles.

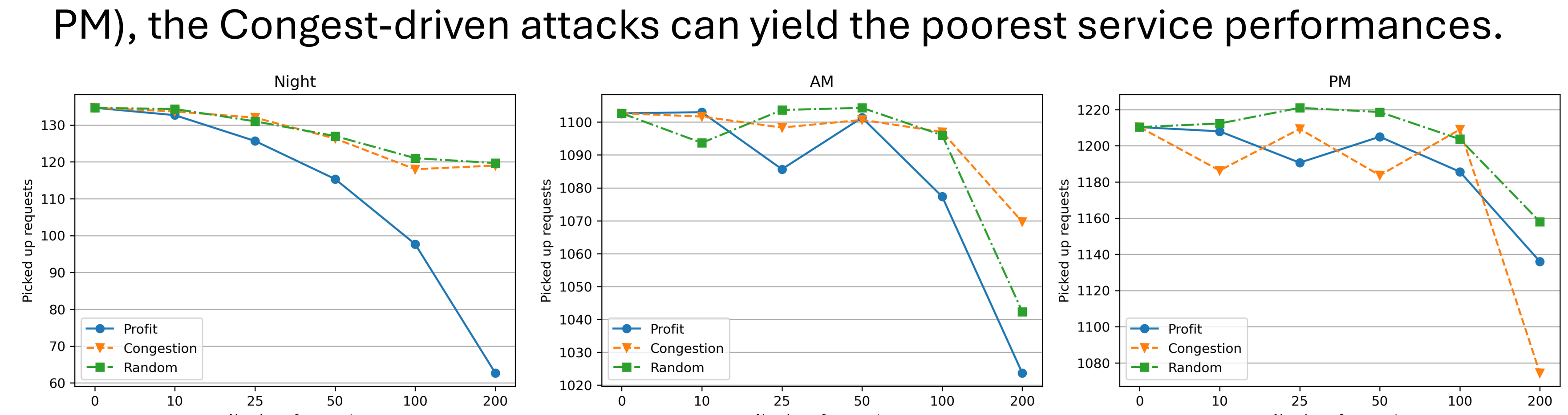
Experiment settings

- Three workdays (2023/4/17-2023/4/19) in NYC with three time periods (AM 7:30-9:00, PM 18:00-19:30, and Night 2:30-4:00) with 30 min attacks in the middle of 90 min simulation.
- Three attack strategies: **Random** generated attacks, **Profit**-driven attack by considering the first term, and **Congest**-driven attack by considering both.
- Defense of the platform:

1. One account can only send one trip request at a time.
2. Following the cancellation of a request, the account must wait for 5 minutes before submitting another one.

Key results

- The Profit-driven attack is the most effective in most cases.
- However, when the congestion effect becomes particularly noticeable (e.g., in PM), the Congest-driven attacks can yield the poorest service performances.



Summary

- We investigate a threat model that can exploit the vulnerabilities in passenger-vehicle matching.
- We develop a simulation framework to evaluate the attack's impact to OMS performances with the consideration of congestion.
- The results show that a limited number of compromised accounts can cause significant reduction in service performances, which suggest sharing real-time vacant vehicle locations would introduce significant risks.

PURDUE UNIVERSITY

CERIAS