16. Cyber-Physical Security of Connected & Autonomous Vehicles

金力 Li Jin
li.jin@sjtu.edu.cn
上海交通大学密西根学院
Shanghai Jiao Tong University UM Joint Institute

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Outline

- Background
- Attacker-defender interaction
- Basics of game theory*

Background

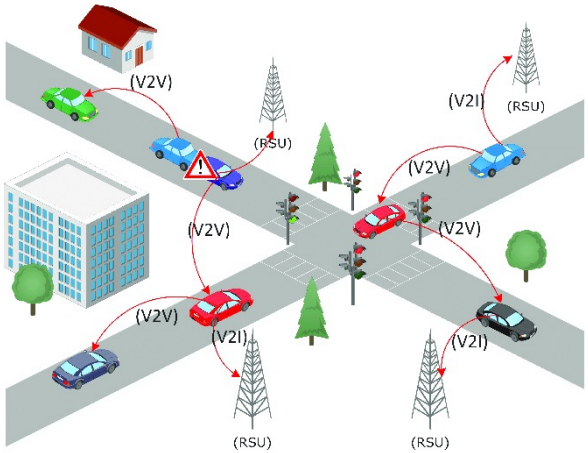
- Application layer
- Network layer
- System level
- Privacy leakage

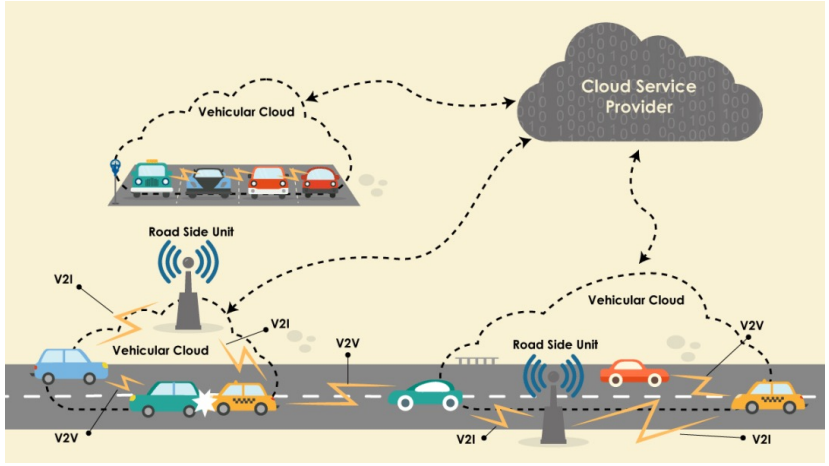
Challenge: cyber-physical security for CAVs

- Adversaries that maliciously compromise (disconnect or pollute) information flow
- Why people do this? Beyond the scope of this talk…
- But people can do and have done this…

Vulnerabilities

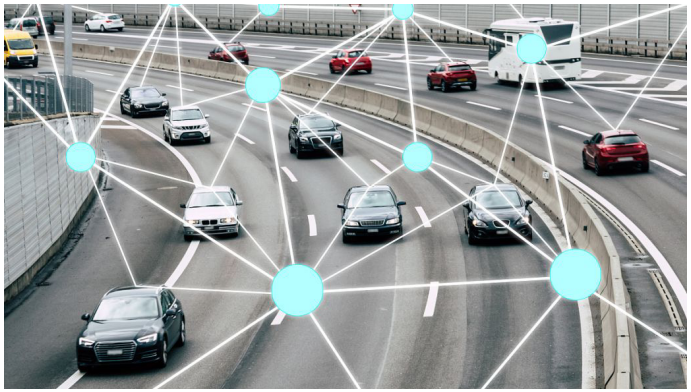
- To achieve automated cooperative driving, vehicles need to have access to each other's information.
- Such information enhances the ability of the autonomous vehicle to plan ahead and make better decisions to improve the overall safety and performance of the vehicle.
- One possible way of achieving inter-vehicle communication is through vehicular ad hoc networks (VANETs).





Vulnerabilities

- Although VANET technologies have matured over time, they are still riddled with security issues.
- Faults in software components can potentially lead to devastating effects for autonomous vehicles and other vehicles sharing the roadway.
- Thus, it is important to design the system to be robust, secure against malicious attacks, privacy preserving, and fault tolerant.



Vulnerabilities

- We group the security attacks on CAVs as application layer, network layer, system level, and privacy leakage attacks.
- All these attacks can potentially impact the string stability of the system, and compromise the safety and privacy of the passengers of the CAV stream.
- Such attacks can be launched by either an outsider or insider adversary.
- While leveraging state-of-the-art security architectures can potentially limit the capabilities of outsider attacks, there can still be disruptive insider attacks.

Vulnerabilities

- Application layer
- Network layer
- System level
- Privacy leakage

- Amoozadeh, M., Raghuramu, A., Chuah, C. N., Ghosal, D., Zhang, H. M., Rowe, J., & Levitt, K. (2015). Security vulnerabilities of connected vehicle streams and their impact on cooperative driving. *IEEE Communications Magazine*, *53*(6), 126-132.

Application layer

- Application layer attacks affect the functionality of a particular application such as CACC beaconing, or message exchange in the platoon management protocol.
- In a message falsification attack, the adversary starts listening to the wireless medium and, upon receiving each beacon, manipulates the content meaningfully and rebroadcasts it
- In a spoofing attack, the adversary impersonates another vehicle in the stream in order to inject fraudulent information into a specific vehicle.
- In a replay attack, the adversary receives and stores a beacon sent by a member of the stream and tries to replay it at a later time with malicious intent.

Application layer

- How to address application layer attacks?
- For outsiders (invader):
  - Prevent unauthorized outsiders from intruding
  - State-of-the-art security architectures employing a strong cryptographic system
  - Digital signatures provide data integrity for beacon messages and protect them from unauthorized change.
  - But many practical challenges involved in deployment, implementation, and standardization
- For insiders (spy):
  - Not very much can be done with cryptography & authentication.
  - Main solution: misbehavior/anomaly detection techniques
  - Multiple sources of data for cross verification
  - False negative and false positive rates.

Network layer

- Network layer attacks have the potential to affect the functioning of multiple user applications.
- (Distributed) denial-of-service (DoS/DDoS) attack
  - CAVs have a tamper-proof hardware security module (HSM), which is responsible for storing digital keys as well as performing all cryptographic operations, such as message signing/verification, encryption, and hashing. A botnet can target this limitation to overwhelm an autonomous vehicle and make its HSM unavailable.
  - Radio jamming to deliberately disrupt communications over small or wide geographic areas.

System-level attacks

- Another type of attack is tampering with vehicle hardware or software, which can be done by a malicious insider at the manufacturing level or by an outsider in an unattended vehicle (e.g., by replacing or altering certain vehicle sensors).
- Even if the communication channel is secure, and a state-of-the-art security architecture is deployed in the VANET if the onboard hardware/software is tampered with or faulty, the input information to the system will not be accurate.
- One possible solution is to use tamper-proof sensors.
- If a tamper-proof version is not available or too expensive to deploy on a large scale, the misbehavior detection techniques discussed below can be useful.

Privacy leakage attacks

- An eavesdropping attack extracts valuable information about the CAVs such as its trace by linking position data, and use it for her own benefit.
- Eavesdropping is a type of passive attack, and hence is difficult to detect, especially in broadcast wireless communication.
- However, it is possible to prevent the success of eavesdropping by using encryption to achieve data privacy or using anonymity techniques to achieve identity and location privacy.

A philosophical question

Why would people do such bad things?

Attacker-defender interaction

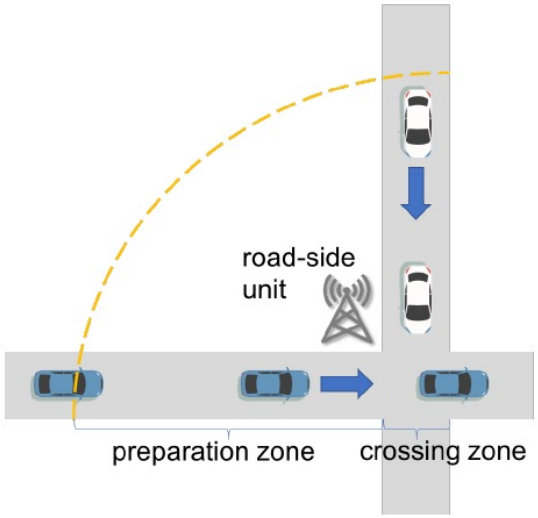
- Baseline setting
- Attacker model
- Defender model

Baseline setting

- CAV intersection, FCFS
- Standard trajectory planning algorithm (road-side)
  - Generates trajectory to go
- Standard trajectory tracking algorithm (road-side)
  - Generates target speed
- Standard car-following/collision avoidance system (onboard)
  - Generates engine torque

Problem formulation

- Two one-way routes without turning
- Radius of preparation zone = L
- Initial conditions $x_i^k(0), v_i^k(0) = \bar{v}$
  - $x_i^k(0) - x_{i+1}^k(0) \geq d$ (safe)
  - $\left|L - x_i^k(0)\right| + \left|L - x_j^{-k}(0)\right| \geq d$ (safe)
  - $\bar{v}$ = nominal speed



- Furthermore, we assume that in the absence of attacks, every vehicle can keep traveling at constant speed $\bar{v}$ and cross the intersection safely.
- Safely = satisfying the distance constraint

Research questions

- How to create an accident by falsification or spoofing?
- How many pieces of information have to be spoofed to create an accident?
- How much redundant information is needed to detect spoofing?
- To what extent can we recover spoofed information?

Attacker

- Attacking modes
  - DoS: erase certain pieces of information
  - Spoofing: modify certain pieces of information
  - Manipulation: modify instructions
- Budget
  - Probability of erasure/modification
  - # of erasures/modifications
- Objective
  - DoS: maximize efficiency loss
  - Spoofing: create accident

Defender

- Defending actions
  - Increase safety margin
  - Detect spoofed data
  - Recover erased/spoofed data
  - Activate backup sensing
  - Increase security level
- Budget
  - Probability of successful detection
  - # of nodes with increased security
  - # of backup sensing
- Objective
  - Avoid accident
  - Optimize efficiency

Attacking modes

- DoS: suppose the attacker can erase the information of a vehicles per time step.
  - Simultaneously erase $x_i^k(t)$ and $v_i^k(t)$
  - When state of vehicle i is erased, vehicle $i+1$ switches to ACC mode and relies on onboard sensing for vehicle following. This leads to an inter-vehicle spacing $d' > d$.
  - This leads to an increase in travel time
  - Attacker wants to maximizes this increase
  - Which vehicle to erase?

Attacking modes

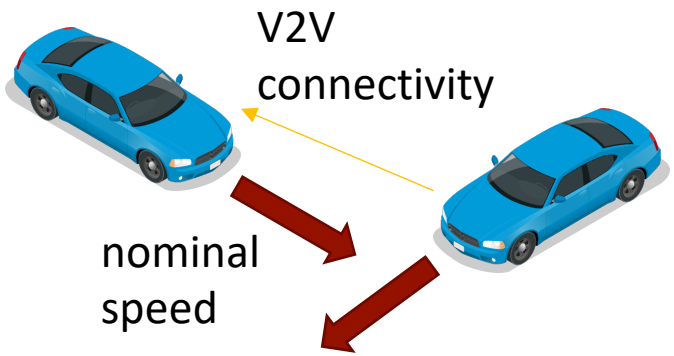
- Spoofing: create non-existent vehicle to cause "phantom" delay
  - Inject information $\xi(t)$ (position) and $v(t)$ (speed) of a phantom vehicle
  - This misleads the RSU to induce an unnecessary spacing between vehicles
  - Attacker wants to maximize the induced delay
- Falsification: modify state observation
  - Modify $x_i^k(t)$ and $v_i^k(t)$
  - Creates accidents
  - That is, drive the system to an unsafe state

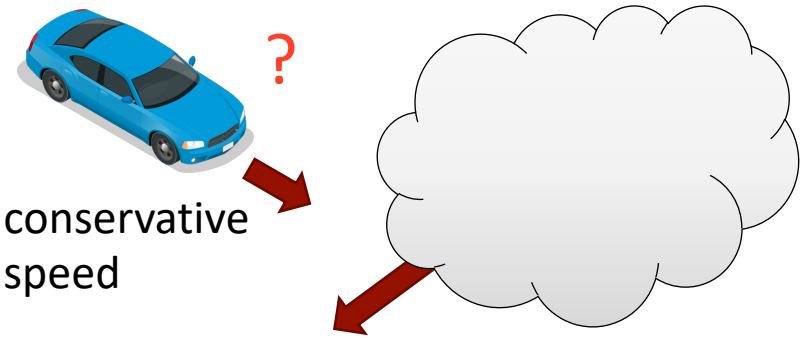
Attacker's decision-making problem

- Data:
  - CAV parameters
- Decision variable
  - Which piece information to attack
  - When to attack
- Constraint:
  - Budget
  - Technological
- Objective
  - Compromise safety
  - Compromise efficiency

Defending modes

- Proactive actions:
  - Increase security investment (e.g. hardware)
  - Improved encryption technology
  - Maintain safety margins
- Reactive actions:
  - Isolate detected malicious information
  - Switch to or activate redundant information sources
  - Intervene the trajectories of neighboring vehicles





Defender’s decision-making problem

- Data:
  - CAV parameters
  - Observed data
- Decision variable
  - Proactive actions
  - Reactive actions
- Constraint:
  - Budget
  - Performance
- Objective
  - Ensure safety
  - Improve efficiency

Question:

How about the attack and defense for a CACC-based vehicle platoon?

Attacker-defender interaction

- Let's summarize the setting...
- Two decision makers have conflicting objectives
- One side's decision will influence the gain/loss of the other side
- Both sides have some extent of expectation/belief of the behavior of the other side
- --> Game!

Why game theory?

Multiple decision makers + conflicting objectives

= game theory

Basics of game theory*

- Formulation
- Strategy
- Equilibrium

Game vs. optimization

Optimization

- A single decision maker
- A single objective function
- Optimal solution: no other solution can improve objective value

Game

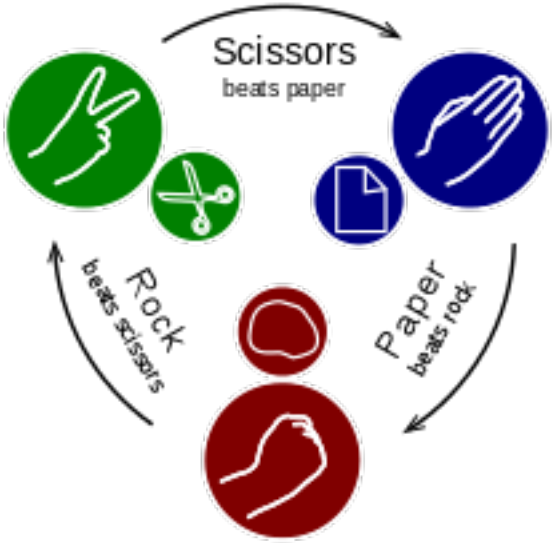
- Two (or multiple) decision makers (players)
- Two players have different (typically conflicting) objectives
- Equilibrium: neither player can improve its utility by unilaterally varying its action

Strategic-form game

- Set of players $i \in \mathcal{I} = \{1,2, \dots, I\}$
- Pure-strategy space S_i
- Space of pure strategies $S = \prod_{i=1}^{I} S_i$
- Payoff function $u_i: \prod_{j=1}^{I} S_j \to \mathbb{R}$
- Strategy profile $s = (s_1, s_2, \dots, s_I)$
- Zero-sum game: $\sum_{i=1}^{I} u_i(s) = 0$ for all s.
- Common knowledge: all players know the structure of the strategic form and other players' knowledge.

Example: Rock paper scissors

- Two players $i \in \{1,2\}$
- Pure strategy space $S_i = \{R, P, S\}$ for $i = 1,2$
- Space of pure strategies $S = \{R, P, S\}^2$
- Payoff function $u_i: \{R, P, S\}^2 \to \{-1,0,1\}$ #
- Strategy profile $s = (s_1, s_2)$
- Zero-sum game: $\sum_{i=1}^{I} u_i(s) = 0$ for all s
- Common knowledge: rules for the game; simultaneous move



Mixed strategy

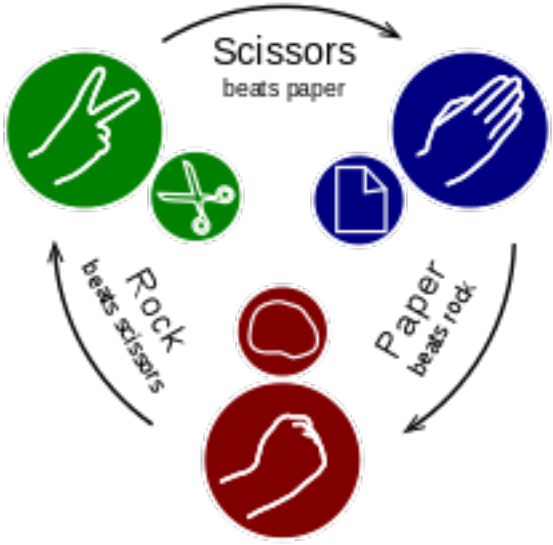
- A mixed strategy σ_i is a probability distribution over pure strategies.
- Each player's randomization is statistically independent of others'.
- Space of player i's mixed strategies = Σ_i
- Mixed strategy profile $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_I)$
- Space of mixed strategy profiles Σ
- Payoff $u_i(\sigma_i) = \sum_{s \in S_i} \left(\prod_{j=1}^{I} \sigma_j(S_j) \right) u_i(s)$

Example: Rock paper scissors

- Randomized action (mixed strategy)

$$\sigma_i = \{p_R^i, p_P^i, p_S^i\}$$
$$\sigma_i \geq 0, \|\sigma_i\|_1^1 = 1$$

- Player i's mixed strategy space $\Sigma_i =$? #
- Mixed strategy profile $\sigma = (\sigma_1, \sigma_2)$
- Space of mixed strategy profiles Σ
- Payoff $u_i(\sigma_i) = \sum_{s \in S_i} \left(\prod_{j=1}^{I} \sigma_j(S_j)\right) u_i(s)$



Dominated strategies

- Pure strategy s_i is strictly dominated for player i if there exists $\sigma_i' \in \Sigma_i$ such that

$$u_i(\sigma_i', s_{-i}) > u_i(s_i, s_{-i}) \text{ for all } s_{-i} \in S_{-i}$$

- The strategy s_i is weakly dominated if
  - $u_i(\sigma_i', s_{-i}) \geq u_i(s_i, s_{-i})$ for all $s_{-i} \in S_{-i}$
  - There exists $-i$ such that $u_i(\sigma_i', s_{-i}) > u_i(s_i, s_{-i})$
- Iterated (strict) dominance: process of elimination of dominated strategies
- Only a small portion of games can be solved by iterated dominance
- R-P-S: dominated strategy exists?

Nash equilibrium

- Provides solutions to general games
- A Nash equilibrium is a profile of strategies such that each player's strategy is an optimal response to the others'
- Mathematically, a mixed-strategy profile σ^* is a Nash equilibrium if for all players i

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(s_i, \sigma_{-i}^*) \text{ for all } s_i \in S_i$$

- Similarly, a pure-strategy Nash equilibrium is s^* such that for all players i

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) \text{ for all } s_i \in S_i$$

Nash equilibrium

- A Nash equilibrium is strict if each player has a unique best response to others'
- That is, s^* is a strict equilibrium if and only if for all i and for all $s_i \neq s_i^*$

$$u_i(s_i^*, s_{-i}^*) > u_i(s_i, s_{-i}^*)$$

- A strict equilibrium must be a pure-strategy equilibrium
- If one round of iterated dominance yields a unique strategy profile s^*, then s^* must be a unique Nash equilibrium
- Any Nash-equilibrium strategy profile puts weight only on strategies that are not strictly dominated

Nash equilibria

- Nash equilibria are consistent predictions of how game will be played in the sense that if all players predict a particular Nash equilibrium, then no player has an incentive to play differently.
- Many games have multiple Nash equilibria
- A Nash equilibrium is played if there exists some mechanism that leads all players to expect the same equilibrium
- Focal point: a priori preference for certain equilibria
- Risk dominance: Pareto-dominant equilibrium

Example: Rock paper scissors

- Does this game have a pure-strategy NE?

Existence of mixed-strategy equilibrium

- **Theorem**: Every finite strategic-form game has a mixed-strategy equilibrium.
- Remember that a pure-strategy equilibrium is an equilibrium in degenerate mixed strategies.
- The theorem does not assert the existence of an equilibrium with nondegenerate mixing.
- R-P-S: no pure-strategy equilibrium, but with mixed-strategy equilibrium #

Cournot game

- So far we have motivated the solution concepts of dominance, iterated dominance, and Nash equilibrium by supposing that
  - players predict opponents' play by introspection and deduction
  - common knowledge
- Cournot adjustment process: an alternative approach to introspection for explaining why Nash equilibria are played

Cournot game: definition

- Players take turns choosing their actions
- Each player's each action is a best response to the history of actions
- Suppose two players 1 and 2
- Player 1 moves first and chooses a_1^0
- Player 2 responds with $a_2^1 = r_2(a_1^0)$
- Player 1 then responds with $a_1^2 = r_1(r_2(a_1^0))$
- If $a_1^k \to a_1^*$ and $a_2^k \to a_2^*$, then $a_2^* = r_2(a_1^*)$ and $a_1^* = r_1(a_2^*)$
- (a_1^*, a_2^*) is a Nash equilibrium

Cournot game: convergence

- If $\left(a_1^k, a_2^k\right) \rightarrow \ (a_1^*, a_2^*)$ for all (a_1^0, a_2^0) within a neighborhood of (a_1^*, a_2^*), the equilibrium is asymptotically stable.
- If $\left(a_1^k, a_2^k\right) \rightarrow \ (a_1^*, a_2^*)$ for all (a_1^0, a_2^0) in the action space, the equilibrium is globally stable.

Existence of Nash equilibrium in infinite games

- **Theorem**: Consider a strategic-form game whose strategy space S_i are nonempty compact convex subsets of an Euclidean space. If the payoff functions u_i are continuous in s and quasi-concave in s_i, then there exists a pure-strategy Nash equilibrium.

Quasiconvex and quasiconcave

- A function $f: S \to \mathbb{R}$ defined on a convex subset S of a real vector space is quasiconvex if for all $x, y \in S$ and $\lambda \in [0,1]$ we have

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$$

- A function $f: S \to \mathbb{R}$ defined on a convex subset S of a real vector space is quasiconcave if for all $x, y \in S$ and $\lambda \in [0,1]$ we have

$$f(\lambda x + (1-\lambda)y) \geq \max\{f(x), f(y)\}$$

Existence of Nash equilibrium in infinite games

- **Theorem**: Consider a strategic-form game whose strategy spaces S_i are nonempty compact subsets of a metric space. If the payoff functions u_i are continuous, then there exists a Nash equilibrium in mixed strategies.
- If payoff functions are discontinuous, mixed strategies may not exist.

Security game between CAV attacker & defender

- Players
- Strategy spaces
- Payoffs
- Common knowledge
- Pure strategy
- Mixed strategy
- Equilibrium

Summary questions

- What types of attacks do CAVs face?
- For CAVs, how does cyber security influence physical safety?
- What are the actions and objectives for CAV attackers and defenders?

Next time

- Cyber-physical security of network systems.