Co-funded by the European Union

CHESS

Cyber-security Excellence Hub in Estonia and South Moravia

Trust Management in Digital Ecosystems

David Halasz and Barbora Buhnova

DevConf.cz 2023, Brno, June 17, 2023

Funded by the European Union under Grant Agreement No. 101087529. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

MUNI
FI

Trust Management in Digital Ecosystems

David Halasz and Barbora Buhnova, DevConf.cz 2023, **Brno**, June 17, 2023

Faculty of Informatics, Masaryk University, Czech Republic

- Masaryk University (MU)
  - Established in 1919
  - 2nd largest in Czechia
  - Over 30,000 students
- Faculty of Informatics, MU
  - Established in 1994
  - 1st faculty of comp. science
  - Over 2,000 students

MUNI
FI

DIGITALIZATION ADVANCEMENT

The Dual-Use Dilemma

Technology facilitates and speeds up activities around us = force multiplier

- Can be used for the good, as well as to cause harm
- E.g. it helps people to organize for the good, as well as for the bad

If we want to boost the good, **opening up to its enormous potential**, we need to simultaneously **boost the protection against the bad**

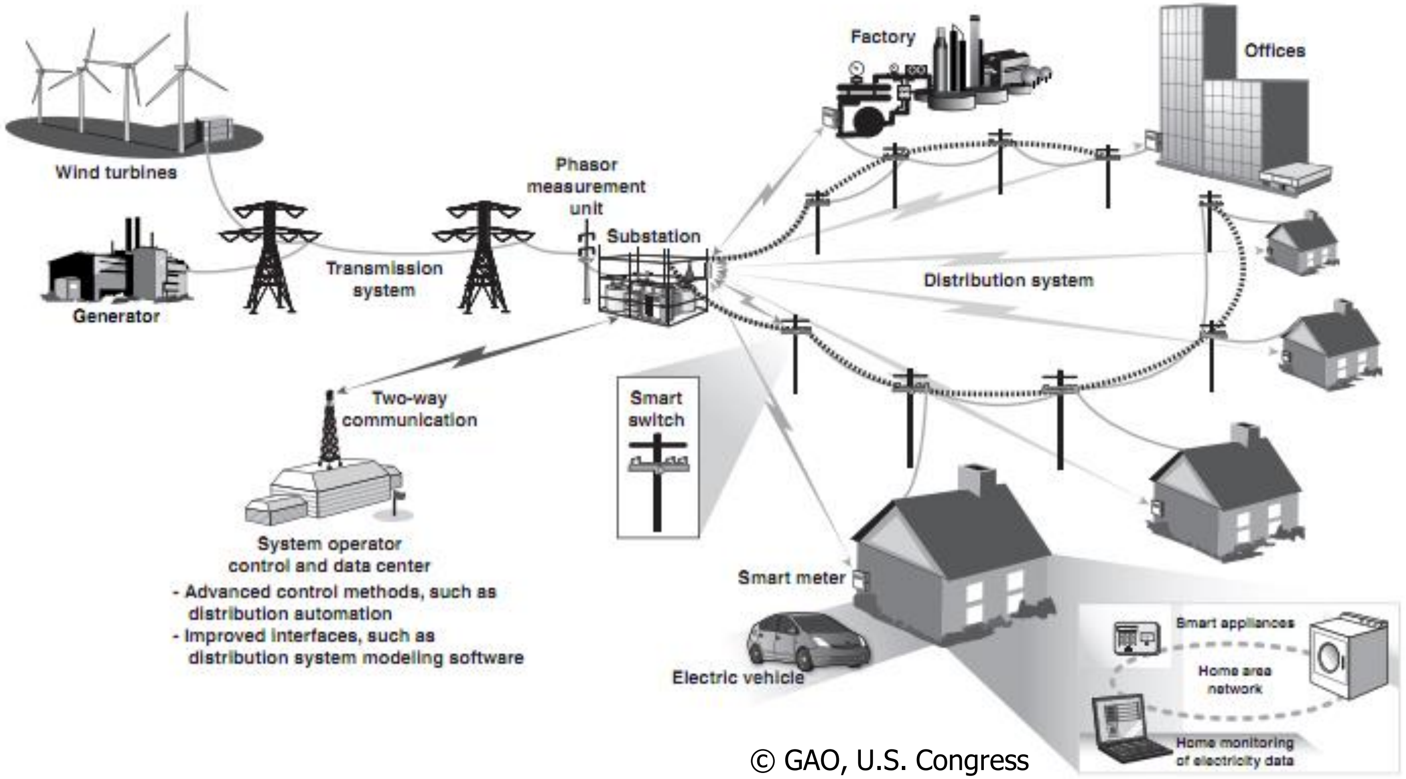
Context-related Challenges

- Hyperconnected world and business landscape, problem cascading, unpredictable impacts
- Uncertainty about the trustability of connected devices
- Highly distributed environment, entry points to secure, data inconsistency, unreliable sensors, partial failures
- Securing against threats that are not existing yet

Digitalization meets Critical Infrastructures

What makes these infrastructures critical?

- The cyber and physical space merged into one
- If we stayed all digital, not much would be in danger, but we go into remote control of everything


Wind turbines
Generator
Transmission system
Phasor measurement unit
Substation
Factory
Offices
Distribution system
Two-way communication
Smart switch
System operator control and data center
- Advanced control methods, such as distribution automation
- Improved interfaces, such as distribution system modeling software
Smart meter
Electric vehicle
Smart appliances
Home area network
Home monitoring of electricity data

© GAO, U.S. Congress

Trust Management in Digital Ecosystems

Image from IEEE CSS

Trust Management in Internet of Vehicles

Information Exchange in Coordinated Moves

Image from Parking Network

Collision Avoidance
with Misbehaving Vehicle

Information Exchange
in Coordinated Moves

Trust Management in Internet of Vehicles

Image from Parking Network

Trust Management in Internet of Vehicles

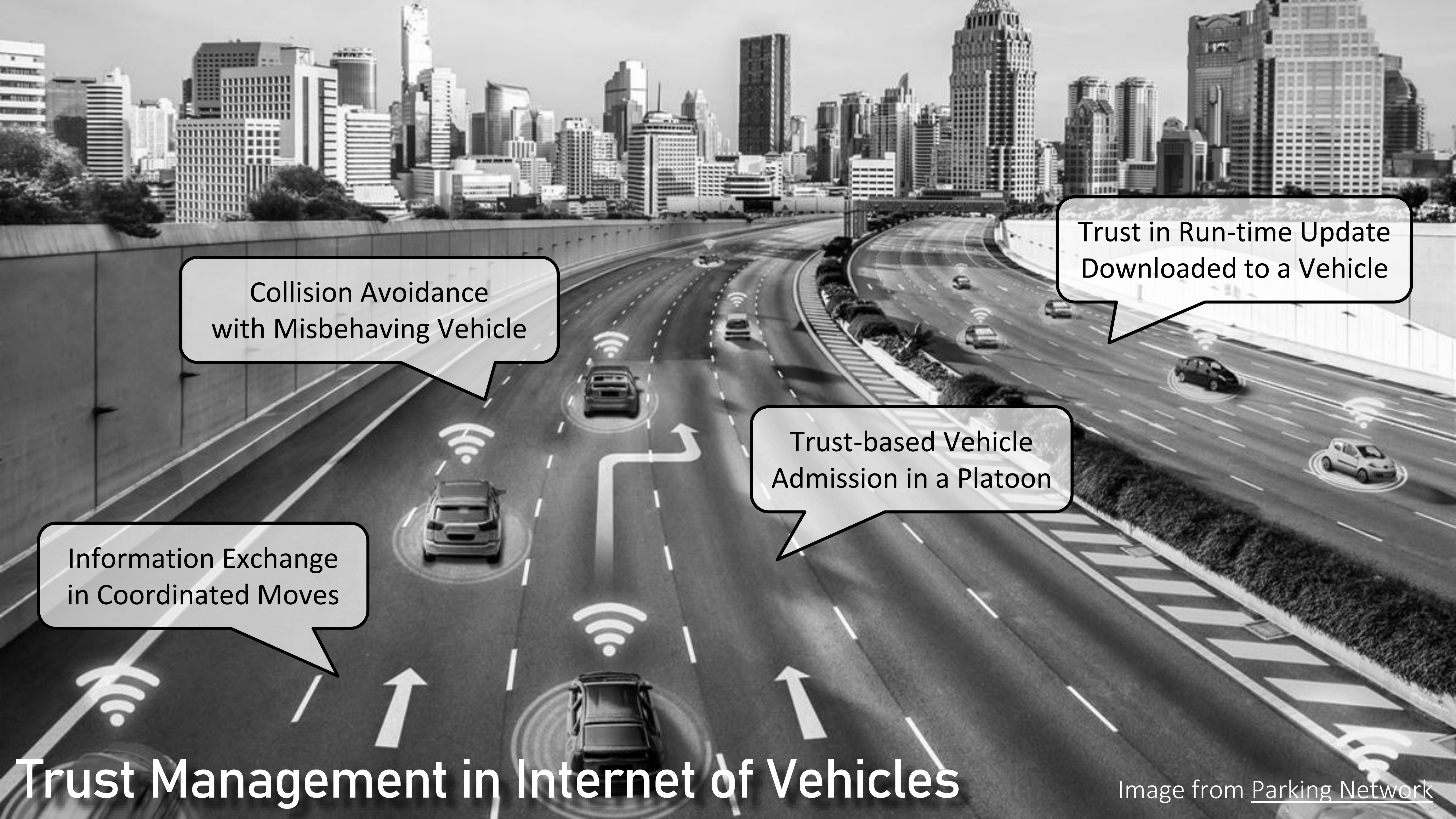
Collision Avoidance with Misbehaving Vehicle

Trust-based Vehicle Admission in a Platoon

Information Exchange in Coordinated Moves

Image from Parking Network

Trust Management in Internet of Vehicles



Collision Avoidance with Misbehaving Vehicle

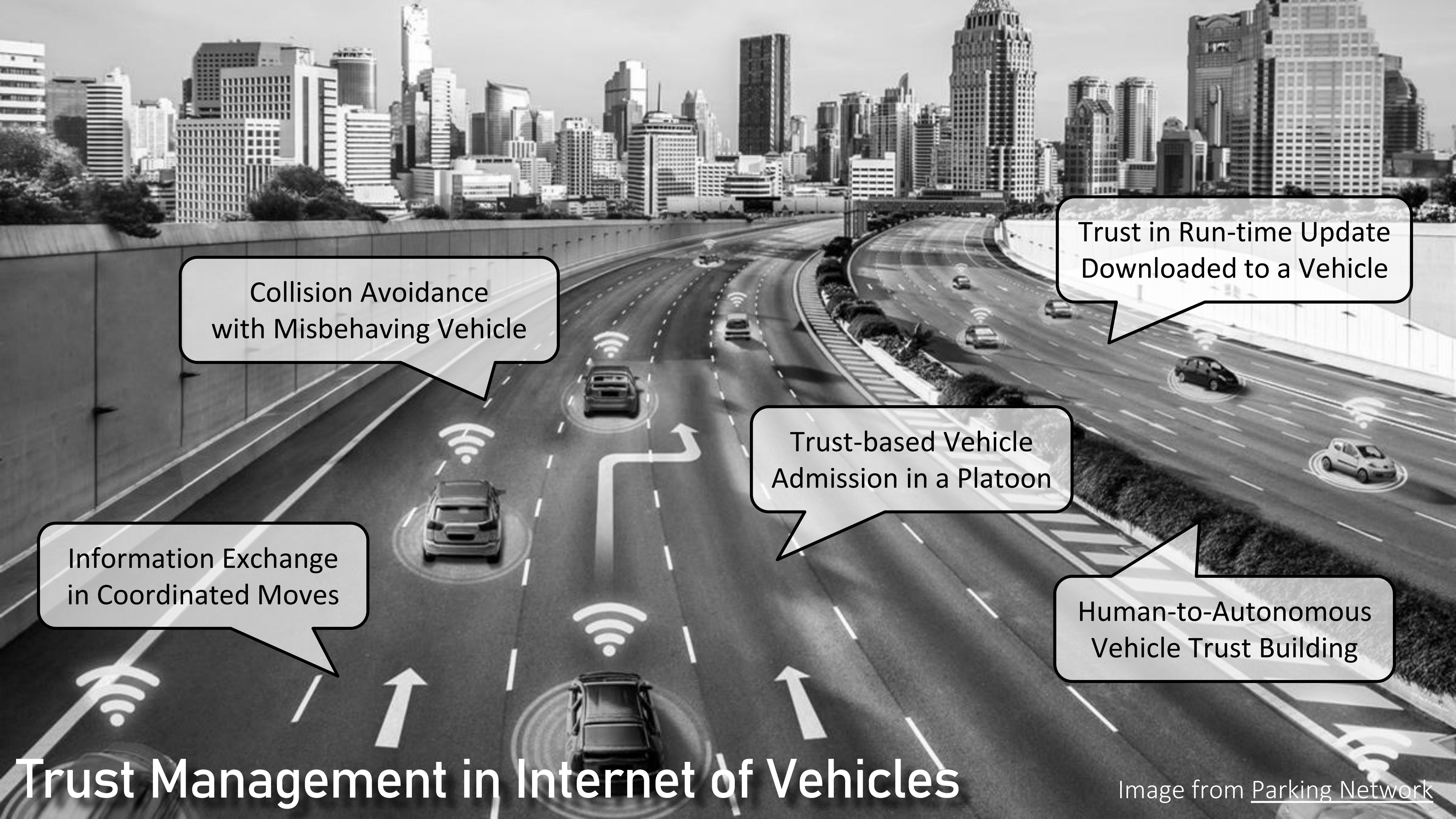
Trust in Run-time Update Downloaded to a Vehicle

Trust-based Vehicle Admission in a Platoon

Information Exchange in Coordinated Moves

Image from Parking Network

Trust Management in Internet of Vehicles



Collision Avoidance with Misbehaving Vehicle

Trust in Run-time Update Downloaded to a Vehicle

Trust-based Vehicle Admission in a Platoon

Information Exchange in Coordinated Moves

Human-to-Autonomous Vehicle Trust Building

Image from Parking Network

Trust Management in Internet of Vehicles

Image from WardsAuto

Trustworthiness does NOT guarantee Trust

- Approaches exist to ensure trustworthiness of the individual ecosystem components, via improving their **security**, **reliability**, **availability**, etc.
- Trust is difficult to get addressed via such solutions.
- Trust is a social psychological concept crucial for forming partnerships, it is conceptually a **belief** about a system that is **out of our control**.
- Although the system might **declare its trustworthiness**, this does not give a guarantee that it can be trusted.
- This is an effect of the fact that **malicious objects** can enter the ecosystem with the intention to disrupt the basic functionality of a network for malicious purposes.

Agents with Malicious Intentions

- Malicious objects can enter the ecosystem with **the intention to disrupt** the basic functionality of the ecosystem for malicious purposes.
- This can be done via **causing harm** directly or via **damaging the reputation** of good (well behaving) objects or by increasing the trustworthiness of misbehaving objects.

What if tech progress gets out of our control? Is tech ban a solution?

- Not really. A safe digital ecosystem therefore needs to be **equipped for dealing with the misbehaving objects** (which are capable of jeopardizing the ecosystem functionality) by restricting their services and prioritizing the trustworthy alternatives.

Trust Management in Internet of Vehicles

Image from Parking Network

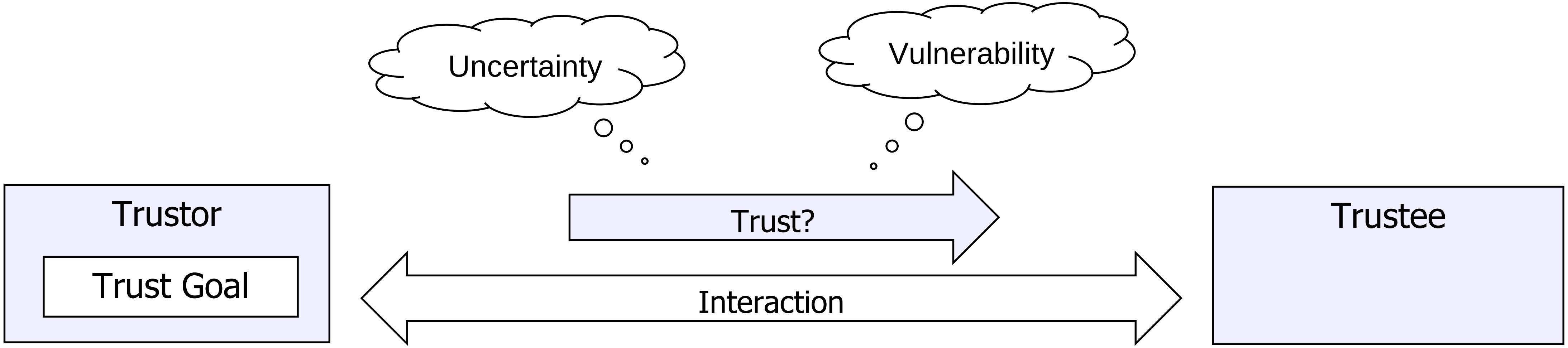
UNDERSTANDING TRUST

What is Trust?

- **Trust in Sociology:** Subjective probability that another party will perform an action that will not hurt my interest under uncertainty or ignorance.
- **Trust in Psychology:** Cognitive learning process obtained from social experiences based on the consequences of trusting behaviors.
- **Trust in Economics:** Expectation upon a risky action under uncertainty and ignorance based on the calculated incentives for the action.
- **Trust in Automation:** Attitude or belief that an agent will help achieve another agent’s goal in a situation characterized by uncertainty and vulnerability.

What is Trust?

Trust: *„the attitude or belief of an agent (trustor) to achieve a specific goal in interaction with another agent (trustee) under uncertainty and vulnerability.*"



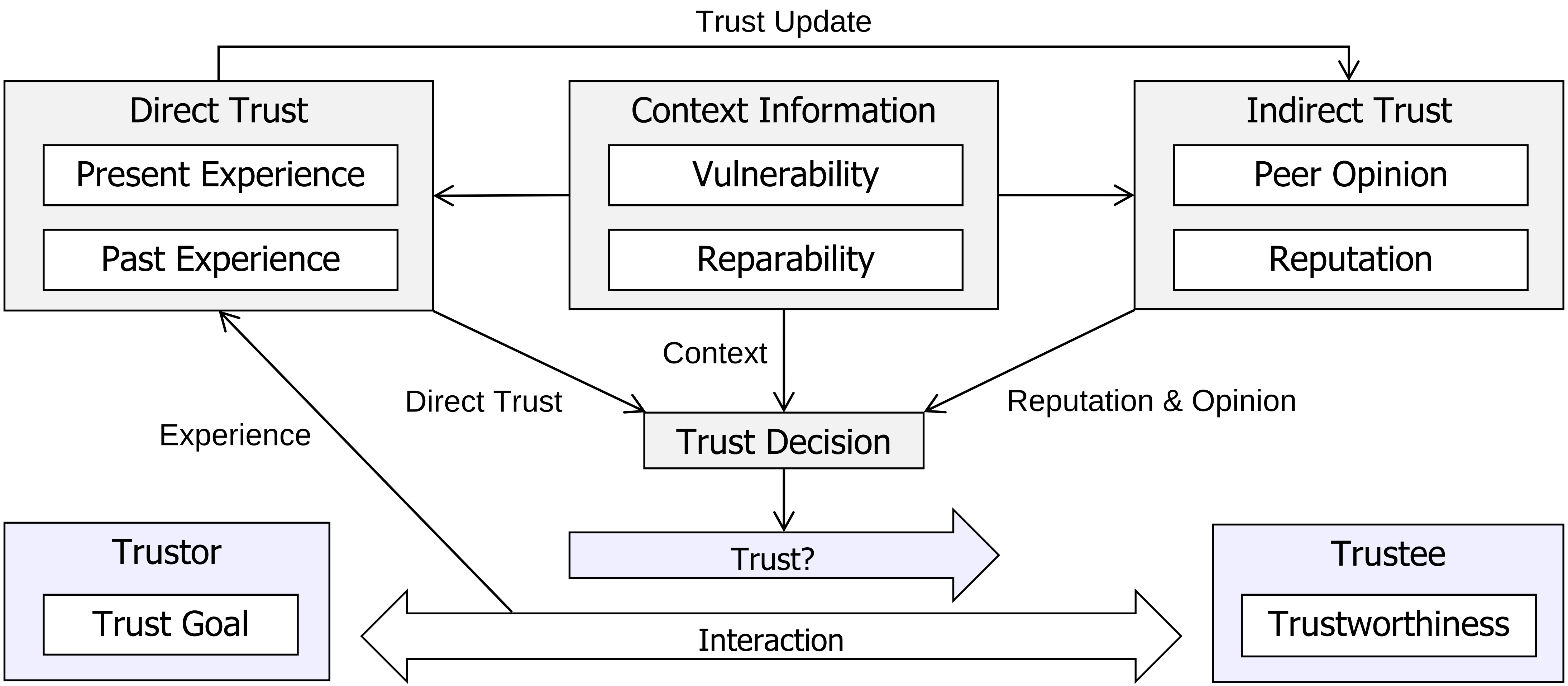
Characteristics of Trust

- **Subjective:** Trust is viewed using the centrality of an agent, wherein the trust is computed based on trustor's observation (i.e., direct trust) as well as the opinion (i.e., feedback or indirect trust) of the other agents.
- **Asymmetric:** Trust is an asymmetric property, i.e., if an agent A trusts another agent B, it does not guarantee that B also trusts A.
- **Transitive:** System agent A is more likely to develop trust towards an agent B if A trust agent C that trusts agent B.

Scope of Trust Evaluation

- **Local:** It represents the trust based on an agent-agent relationship, wherein an agent evaluates the trustworthiness of another agent using local information such as its current observation and past experience.
- **Global:** Global trust is based on the reputation of an agent within the ecosystem, wherein the reputation of each agent might be influenced by the local trust score of each of the other agents in the ecosystem.
- **Context-specific:** Trust of an agent towards another agent varies with context. A trust relation between the agents is usually dynamic and depends on multiple factors such as temporal factors or location.

Trust Management Components



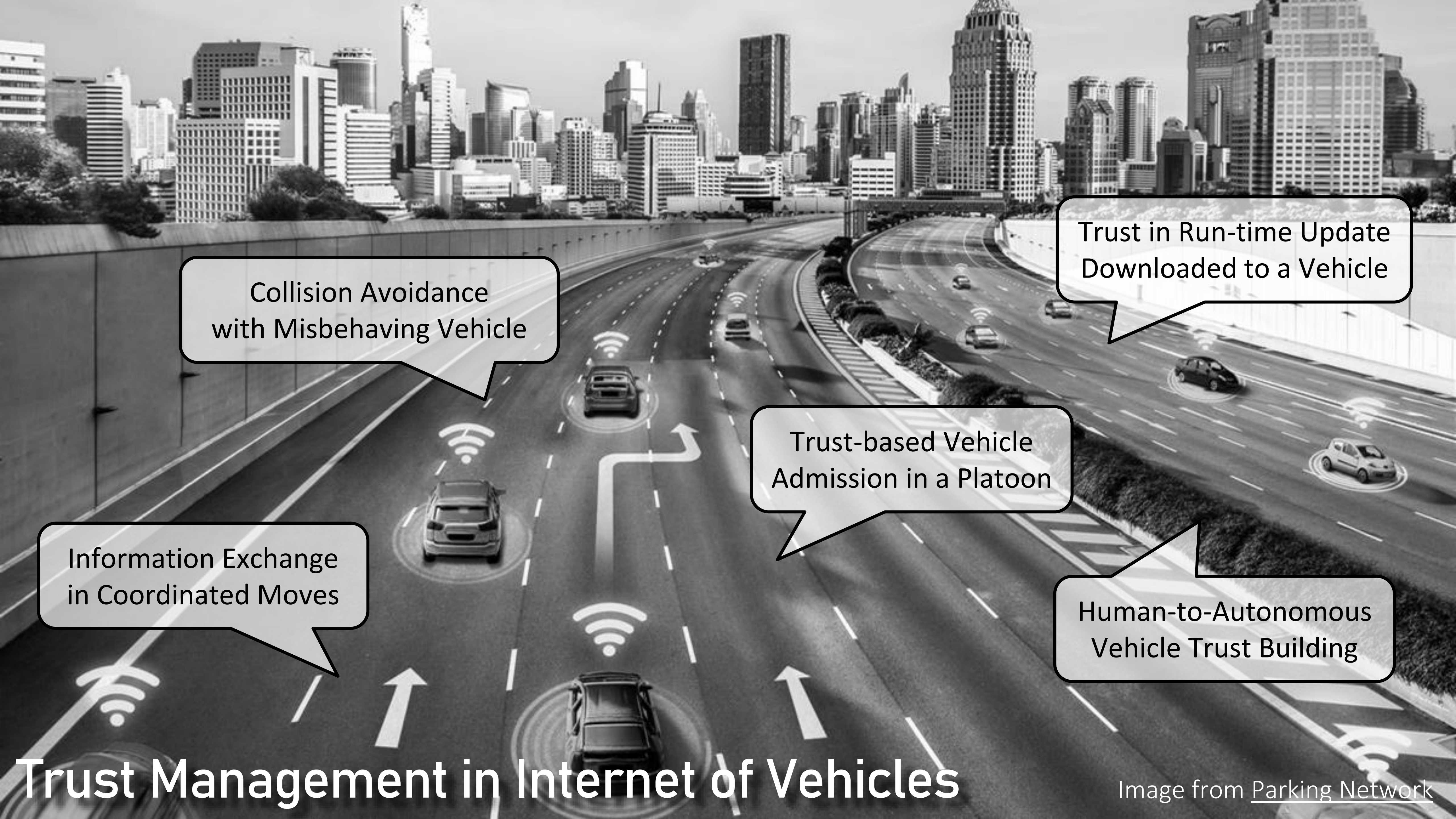
Direct-Trust Evaluation – Trust Metrics

Trust Metrics refer to the features that are chosen and combined in trust computation. These features can refer to:

- **QoS Metrics**, which represent the confidence that an agent is able to offer high quality of the delivered service, e.g. in terms of reliability, availability, security or accuracy.
- **Social Metrics**, which represent the social relationships among ecosystem agents, which can include integrity, benevolence, honesty, friendship, openness, altruism, unselfishness.

SOCIAL METRICS?
What do you mean?

Trust Management in Internet of Vehicles



Collision Avoidance with Misbehaving Vehicle

Trust in Run-time Update Downloaded to a Vehicle

Trust-based Vehicle Admission in a Platoon

Information Exchange in Coordinated Moves

Human-to-Autonomous Vehicle Trust Building

Image from Parking Network

Interesting Problems to Address

1. **Thing-to-Update Trust (update scenario):** A vehicle is downloading a black-box update at runtime. May I trust that update and give it access to my critical driving functions?
2. **Thing-to-Thing Trust (collision avoidance scenario):** Two vehicles approaching each other. May I trust the other vehicle that it does not intend to cause a crash?
3. **Trust-Based Adaptive Safety:** How shall I adapt my safety mechanisms to the level of trust? What if I misdudge trust (false postivies/negatives)?
4. **Trust Management and Governance:** What mechanisms (e.g., incentives, evidence collection, reparation) shall be in place to protect and govern trust values?

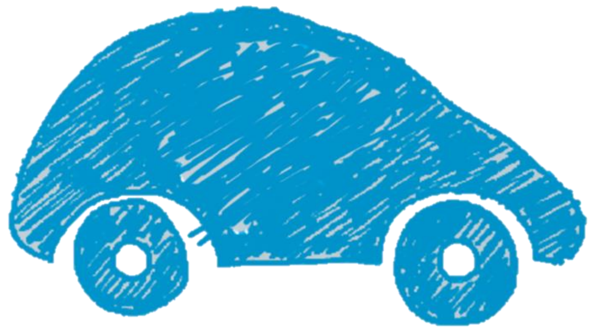
PROBLEM 1
Thing-to-Update Trust (update scenario)

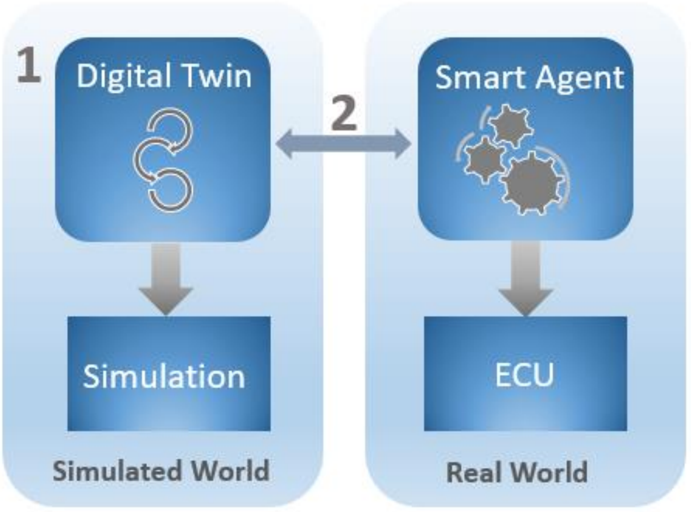
Building Trust in a SW Smart Agent

- Trust of a vehicle into an automated update downloaded to the car.
- DT of the black-box smart agent provided.
- DT run in a simulation environment to get predictive awareness of possible harmful effects.
- A fail-over behavior can be triggered for the system in the real world.

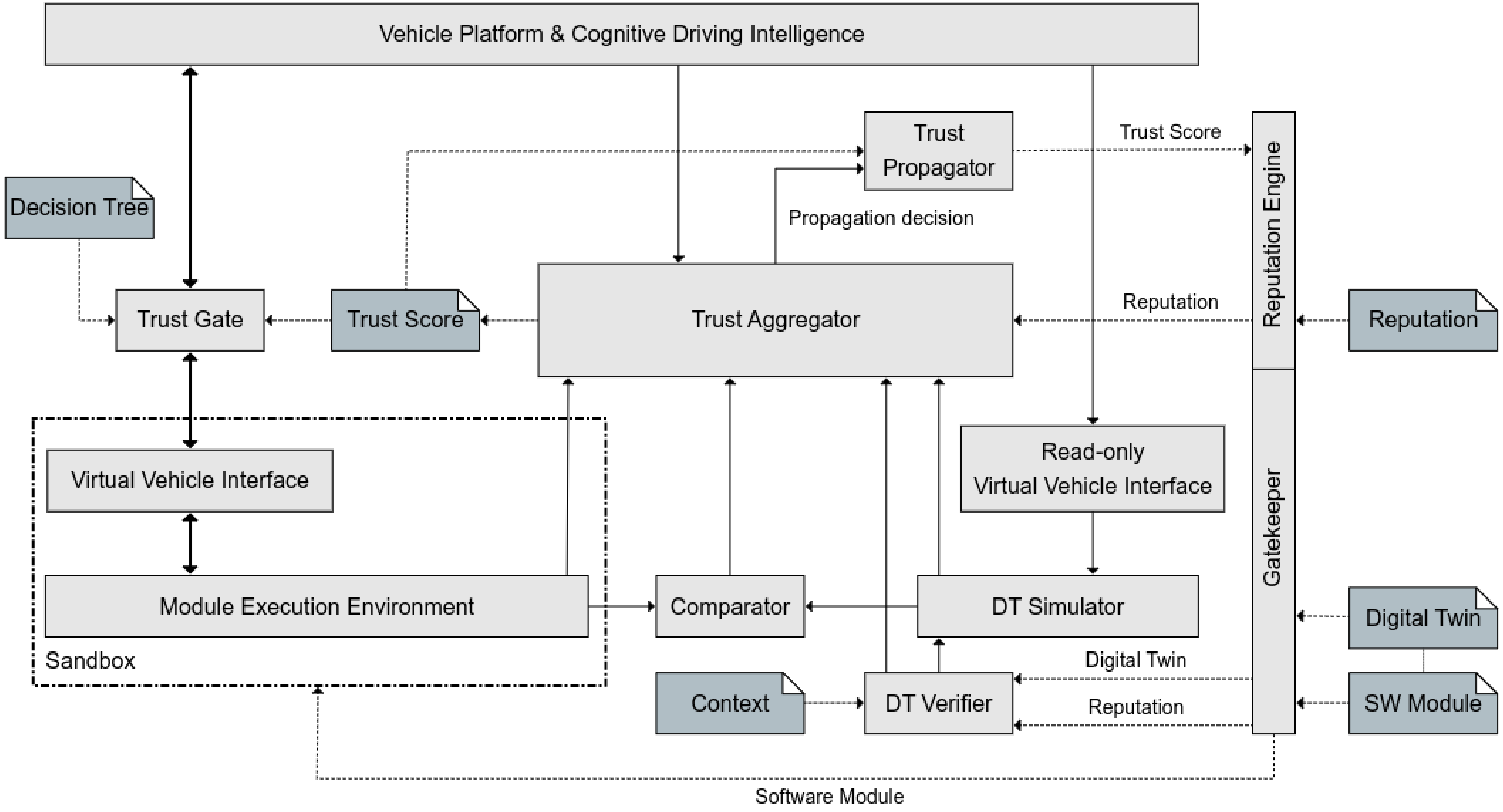
[Ref] Cioroaica, E., Kuhn, T., and Buhnova, B. (2019, May). (Do not) trust in ecosystems. In 2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER) (pp. 9-12). IEEE.

SW Smart Agent





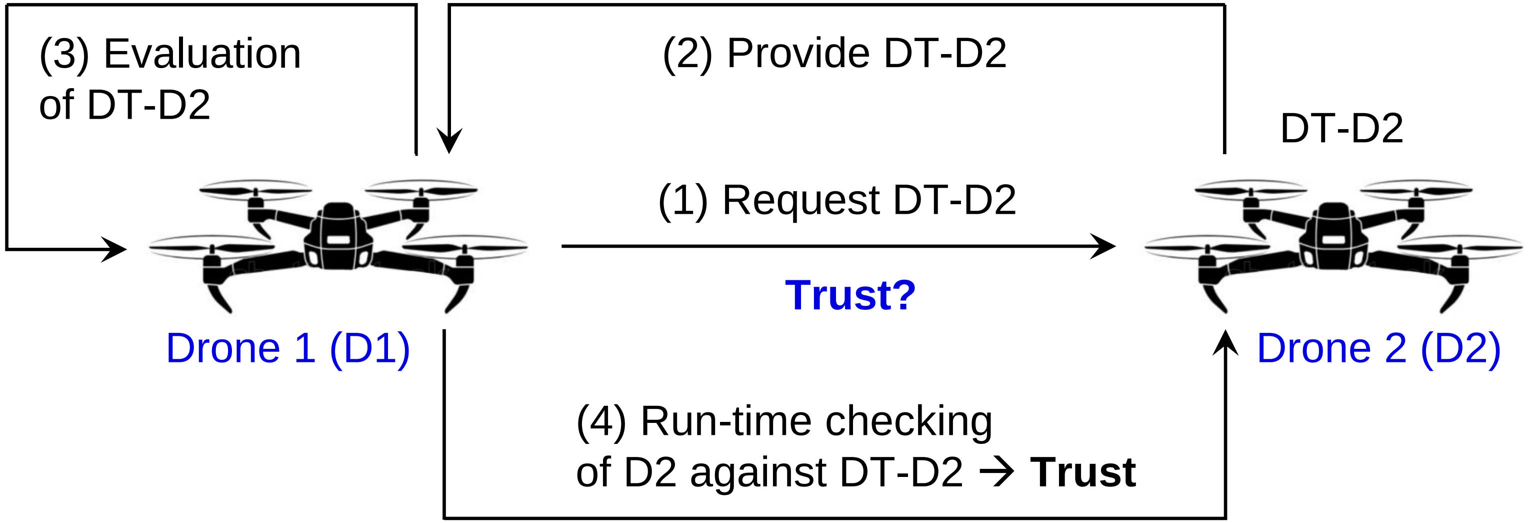
Conceptual Framework of Runtime Trust Evaluation



PROBLEM 2
Thing-to-Thing Trust (collision avoidance)

Trust Building via Predictive Simulation

- Consider **Drone 1** assessing its level of **trust in Drone 2**, as illustrated below.
- From the point of view of Drone 1, Drone 2 is **a black box** and **out of our control**, with **unknown intentions** (possibly malicious, hidden behind good behaviour).



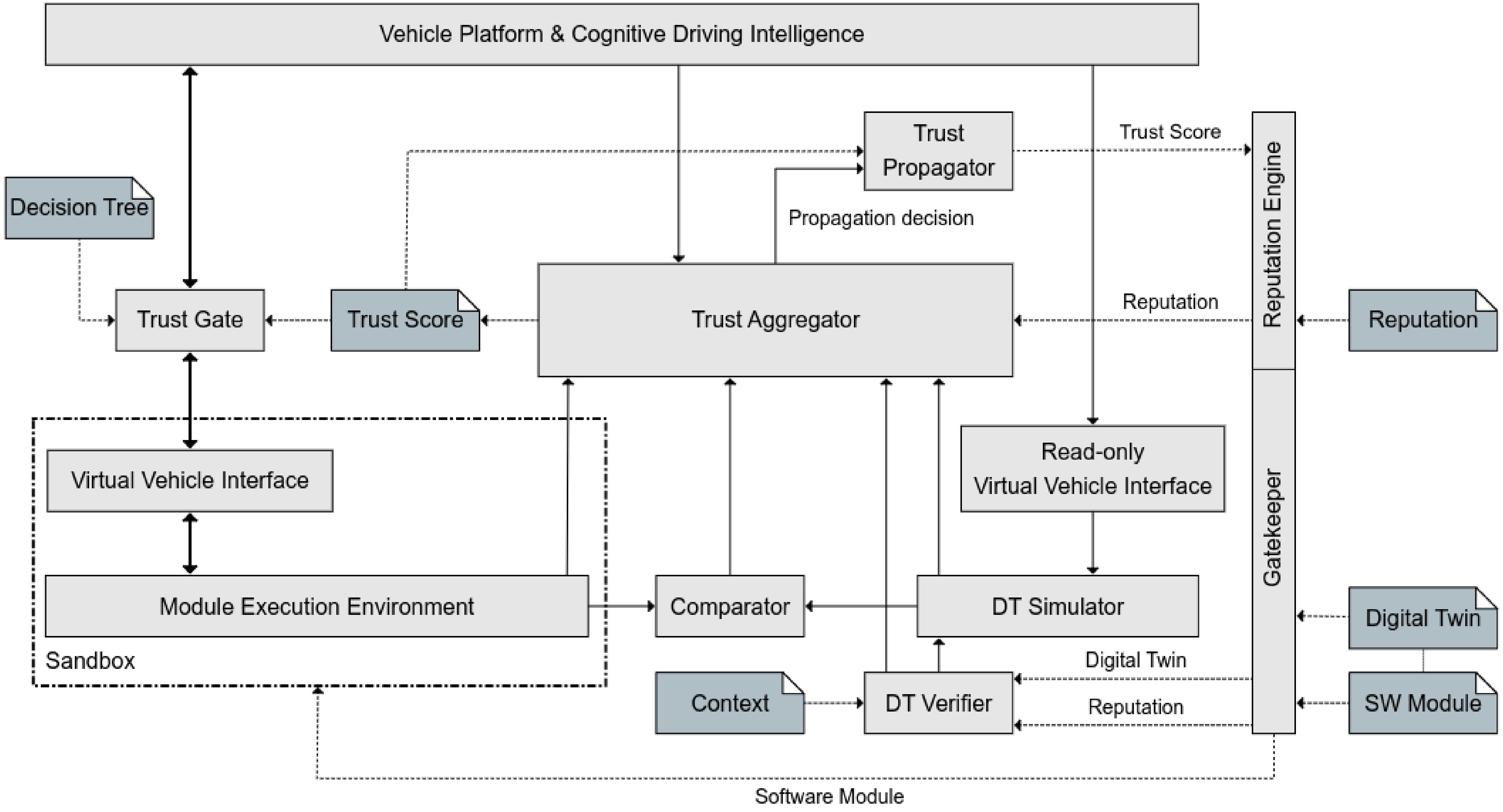
[Ref] Iqbal, D., and Buhnova, B. (2022). Model-based Approach for Building Trust in Autonomous Drones through Digital Twins. In IEEE SMC 2022 International Conference on Systems, Man, and Cybernetics (pp. 9-12). IEEE.

PROBLEM 3
Trust-Based Adaptive Safety

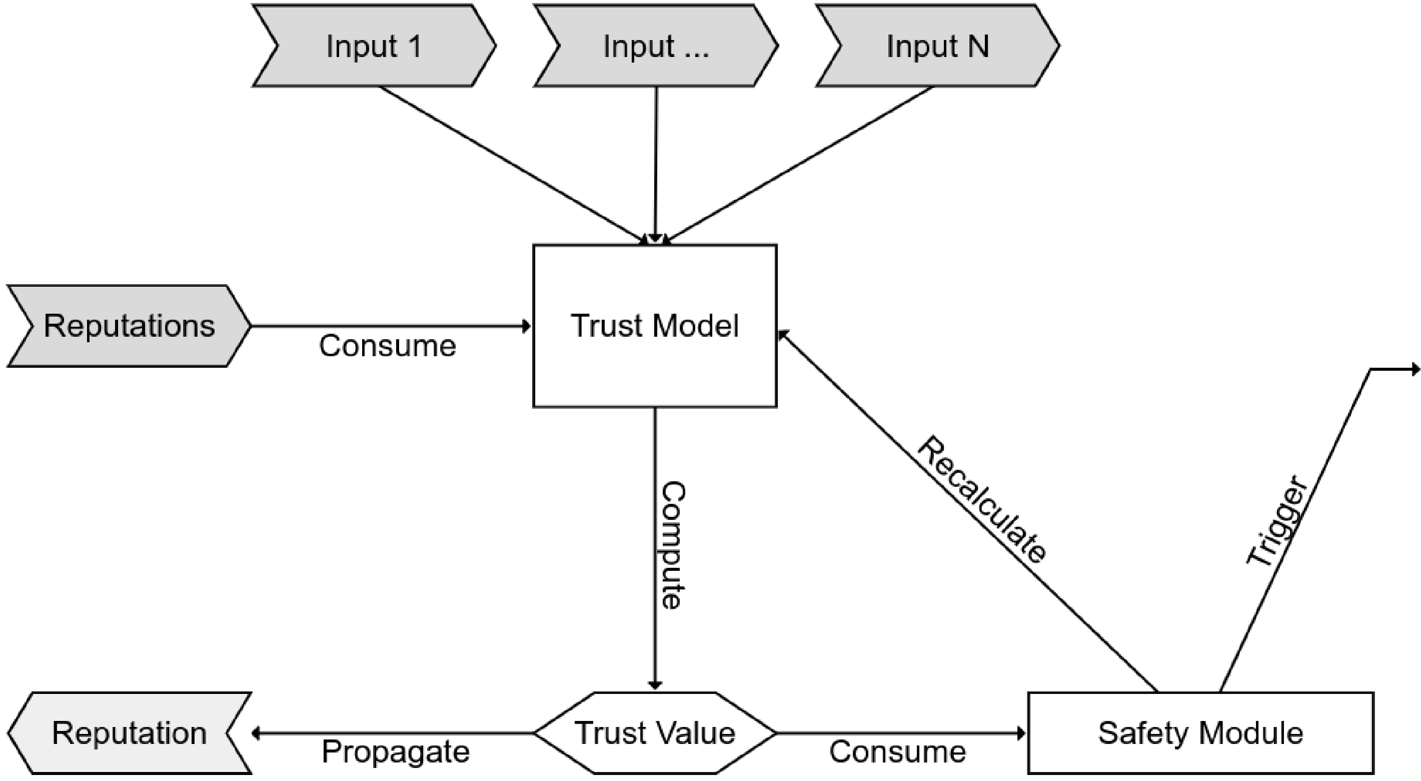
Safety Assurance in Face of Untrusted Agents

- **Run-time adaptive safety:** Evolution of safety mechanisms is needed to support dynamic and self-adaptive architectures of autonomous ecosystems.
- **Adaptation to the level of trust:** Responding to the level of trust among autonomous agents.
- **Safety supervision and control:** An agent that is reported as untrusted might fall under safeguarding of its trustworthy operation, with enabling/disabling its (safety) features.
- **False positives and negatives:** Need to accommodate for trust misjudgment with gradual safety mechanisms.
- **Technical feasibility:** The safeguarding mechanisms can be checked/downloaded on entry to the ecosystem (e.g., the city, highway).

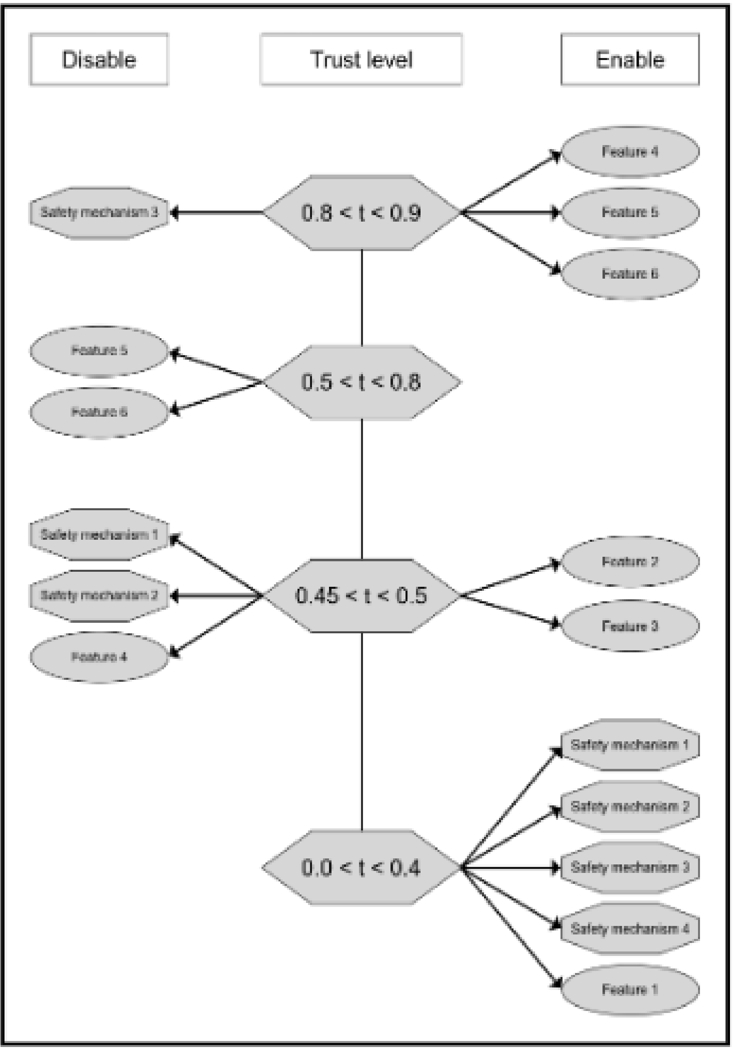
Sandboxing within our Conceptual Framework



Trust-Driven Adaptive Safety



Decision tree



PROBLEM 4
Trust Management and Governance

Trust Governance Mechanisms

- **Trust score** calculation, propagation, update
- **Incentives**, i.e., rewards and punishment mechanisms
- **Reparation** and redemption mechanisms
- **Evidence** collection
  - Pre-incident to predict somebody is attempting misbehaviour
  - Post-incident to either identify <u>the source of misbehaviour</u>, or to understand whether a <u>corrective action needs to be taken</u>

Further Mechanisms to Consider

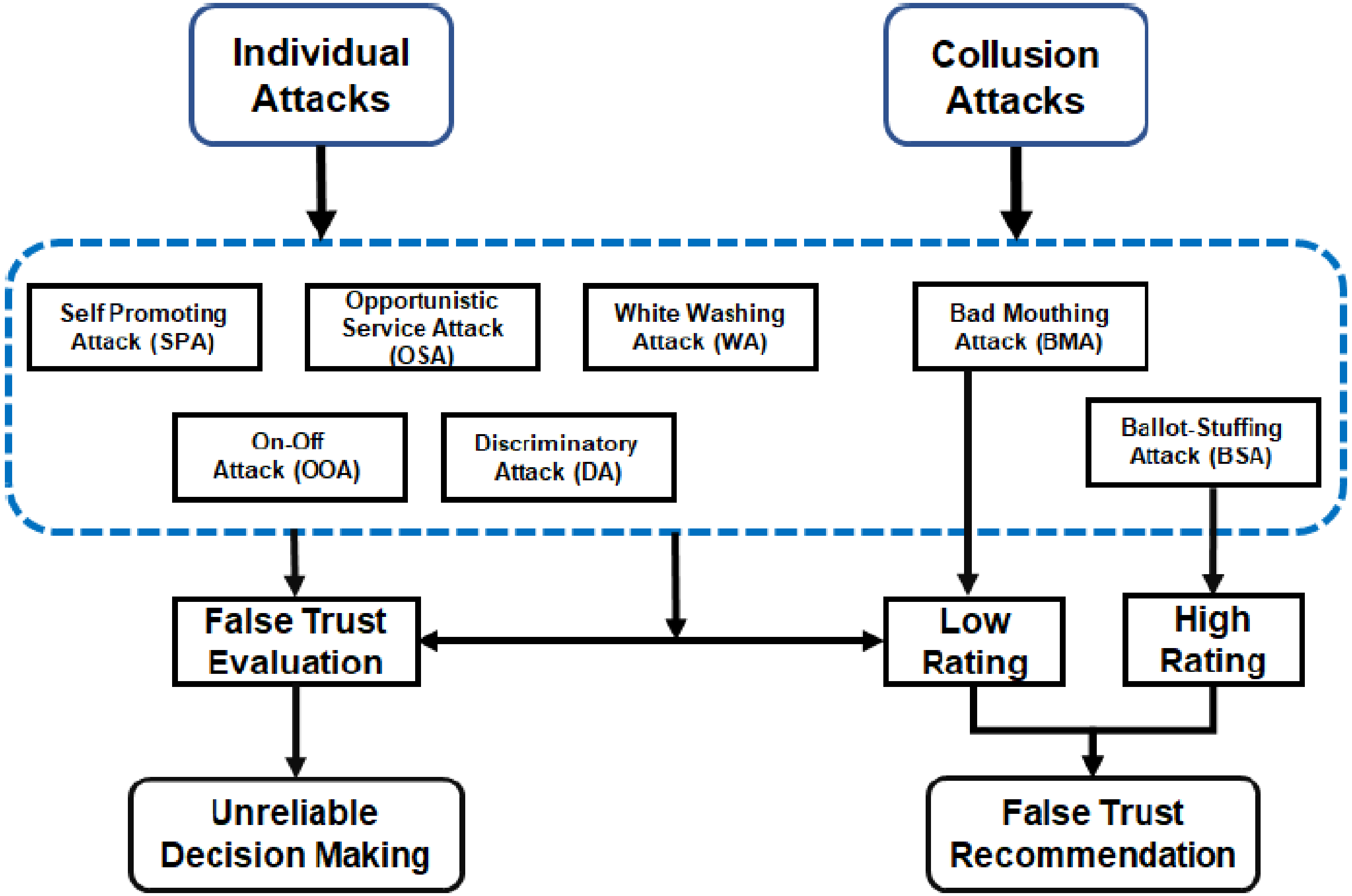
- **Default Trust Score of New Agents:** on which trust score shall a new agent start
- **Trust Erosion:** trust score is subject to decay in case of no or too few interactions
- **Building Trust in the Trustworthy:** employing explanation to give evidence of trustworthiness
- **Black Swan Blindness** and its other sources

WHAT ARE THE CHALLENGES OF TRUST-MANAGEMENT SYSTEMS?

Challenges of Trust Management in IoE

- **Situational Scope of Trust:** high dependence of trust building on the context of trustor
- **Subjectivity of Trust:** influence by the factors inherent to the trustor (e.g. in taking risks)
- **Default Trust Score of New Agents:** on which trust score shall a new agent start
- **Trust Erosion:** trust score is subject to decay in case of no or too few interactions
- **Detection of Hidden Malicious Intentions:** hard to detect, likely to make mistakes in detection
- **Safety Assurance in Face of Untrusted Agents:** an ingredient of immune-response capability
- **Building Trust in the Trustworthy:** employing explanation to give evidence of trustworthiness
- **High Degree of Dynamism and Uncertainty in IoE:** possibly with missing information that is needed to make a decision, leading to misjudgment and bias

Trust Attacks


Individual Attacks
Collusion Attacks
Self Promoting Attack (SPA)
Opportunistic Service Attack (OSA)
White Washing Attack (WA)
Bad Mouthing Attack (BMA)
On-Off Attack (OOA)
Discriminatory Attack (DA)
Ballot-Stuffing Attack (BSA)
False Trust Evaluation
Low Rating
High Rating
Unreliable Decision Making
False Trust Recommendation

Trust Attacks – Individual Attacks

Individual Attacks refer to the attacks launched by an individual agent, which can take form of:

- **Self-Promoting Attacks:** an agent promotes its significance by providing good recommendation for itself so as to be selected as a service provider, and then acts maliciously.
- **Whitewashing Attacks:** an agent exits and re-joins the ecosystem to recover its reputation and to wash-away its own bad reputation.
- **Discriminatory Attacks:** an agent explicitly attacks other agents that do not have common friends with it, i.e. it performs well for a particular service/agent and badly for some other services/agents.
- **Opportunistic Service Attacks:** an agent might offer a great service to improve its reputation when its reputation falls too low.
- **On-Off Attacks:** an agent provides good and bad services on and off (randomly) to avoid being labeled as a low-reputed agent.

Trust Related Attacks – Collusion-based Attacks

Collusion-based Attacks represent the attacks launched by a group of agents to either provide a high rating or low rating to a particular agent, such as:

- **Bad-Mouthing Attacks:** In this type of attack, a group of agents diminishes the reputation of a trustworthy agent within the ecosystem by providing bad recommendations about it.
- **Ballot-Stuffing Attacks:** In this type of attack, a group of agents boosts the reputation of bad agents within the ecosystem by providing good recommendations for them.

THANK YOU

Thank you for your attention

Barbora Buhnova, FI MU Brno
buhnova@fi.muni.cz
www.fi.muni.cz/~buhnova

David Halasz, FI MU & Red Hat
halasz@mail.muni.cz

FACULTAS ARTIS INFORMATICÆ • UNIVERSITAS MASARYKIANA •

CHESS
Cyber-security Excellence Hub in
Estonia and South Moravia

Co-funded by
the European Union