جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Resilient Computing and Cybersecurity Center

Why are we still far from safe and secure autonomous vehicles?

Paulo Esteves-Veríssimo

Research Fellow at LASIGE, ULISBOA (PT)
Former Professor at KAUST and Director of RC3 Center (SA)
pjverissimo@ciencias.ulisboa.pt , https://lasige.pt/
LASIGE, Universidade de Lisboa FCUL, Portugal

89th IFIP WG 10.4 Meeting, Kaunas (LT), 4-7th May, 2026

Is the *autonomous vehicles world* (cyber)-safe?

The past reveals clouds in the horizon of … the safety side … Are they gone?

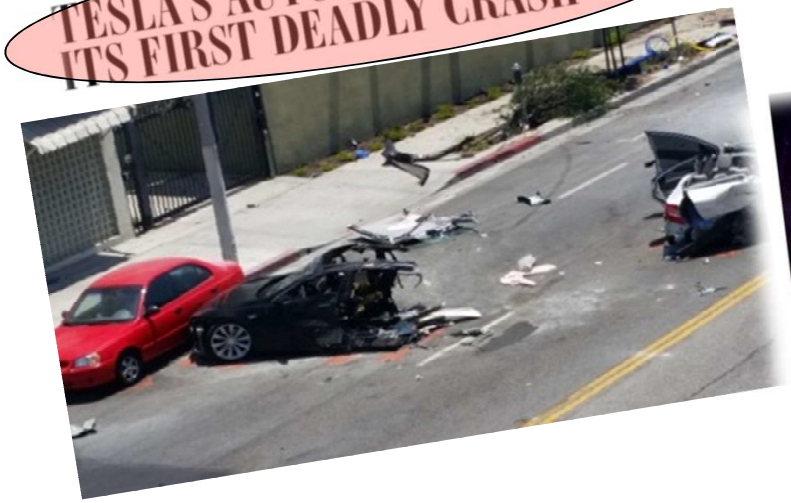
Toyota "Unintended Acceleration" Has Killed 89

A 2005 Toyota Prius, which was in an accident, is seen at a police station in Harrison, New York, Wednesday, March 10, 2010. The driver of the Toyota Prius told police that the car accelerated on its own, then lurched down a driveway, across a road and into a stone wall. (AP Photo/Seth Wenig) AP PHOTO/SETH WENIG

Unintended acceleration in Toyota vehicles may have been involved in the deaths of 89 people over the past decade, upgrading the number of deaths possibly linked to the massive recalls, the government said Tuesday.

TESLA FSD 10.4 tries to turn into incoming traffic, AGAIN!

TESLA'S AUTOPILOT HAS HAD ITS FIRST DEADLY CRASH




Is the *autonomous vehicles world at least* (cyber)-dependable in, say, “normal” days?

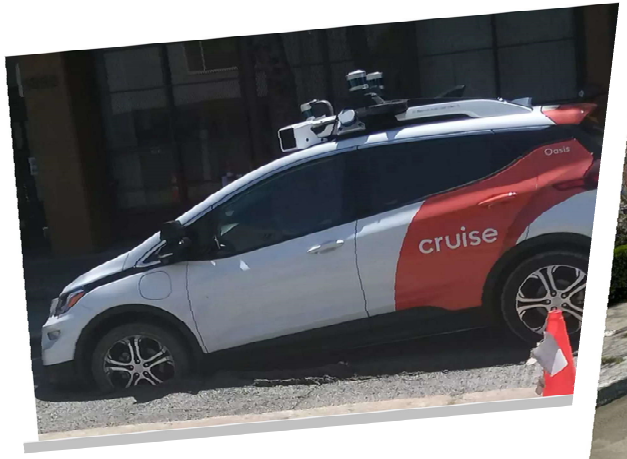
SO, what about "normal" case behaviour ? ...

SFGATE

Newsletters

Cruise vehicle gets stuck in wet concrete while driving in San Francisco

By Joshua Bote
Aug 15, 2023



r/shittyrobots • 2 yr. ago
by Jasaj4

Join

Delivery robot tries to walk across undried cement

CARSCOOPS

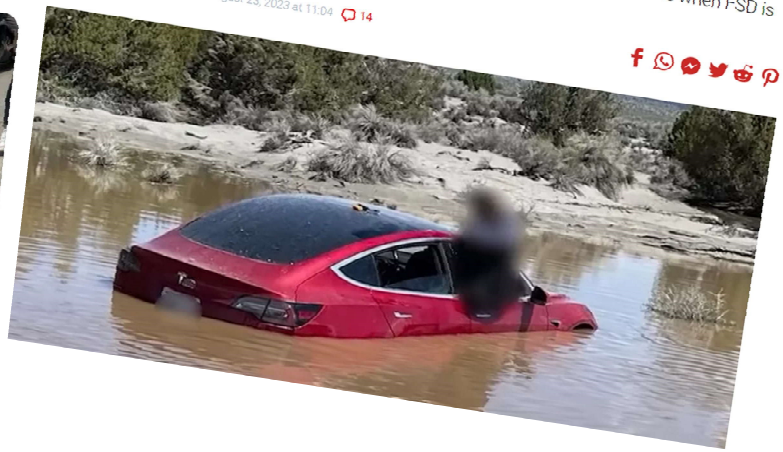
LATEST NEW CARS SCOOPS

VIDEO

Tesla Model 3 Driver Ignores FSD Limitations, Drives Through Flooded Road

The driver of the Tesla Model 3 seems to have forgotten that drivers are still responsible when FSD is

by Brad Anderson August 23, 2023 at 11:04 14



Is the *autonomous vehicles world* (cyber)-secure?

Security gap in Vehicle Systems

naked security by SOPHOS

Award-winning computer security news

SOPHOS.COM >

The Jeep hackers return to ditch a car going 60 mph

03 AUG 2016 · 5

Security threats, Vulnerability

Home About

Keen Security Lab Blog

2016-09-19

Car Hacking Research: Remote Attack Tesla Motors

By Keen Security Lab of Tencent

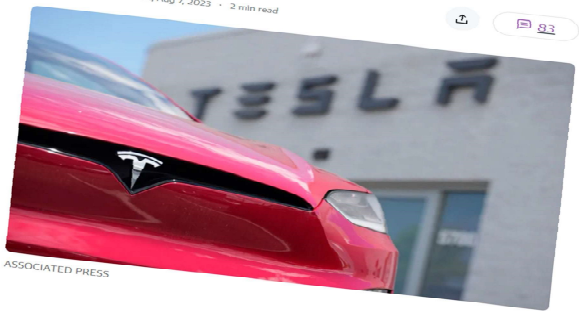
With several months of in-depth research on Tesla Cars, we have discovered multiple security vulnerabilities and successfully implemented remote, aka none physical contact, control on Tesla Model S in both Parking and Driving Mode. It is worth to note that we used an unmodified car with latest firmware to demonstrate the attack.

engadget Sections

Researchers reveal Tesla jailbreak that could unlock Full Self-Driving for free

The group found a hardware exploit they say would be hard for Tesla to mitigate.

Kris Holt
Contributing Reporter
Updated Mon, Aug 7, 2023 · 2 min read



ASSOCIATED PRESS

naked security by SOPHOS

Award-winning computer security news

SOPHOS.COM >

The Jeep hackers return to ditch a car going 60 mph

03 AUG 2016 · 5

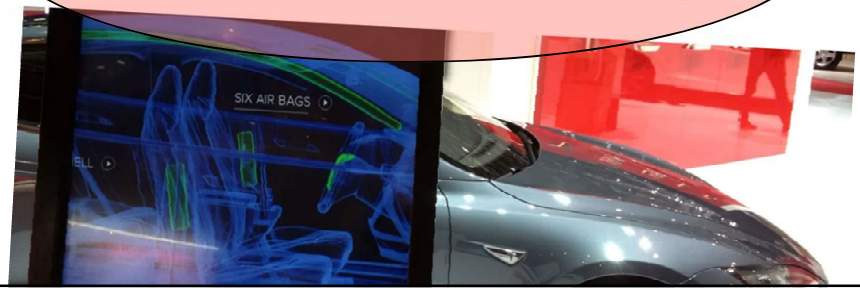
Security threats, Vulnerability

electrek Exclusives Autos Alt. Transport Autonomy Energy

AUGUST 27

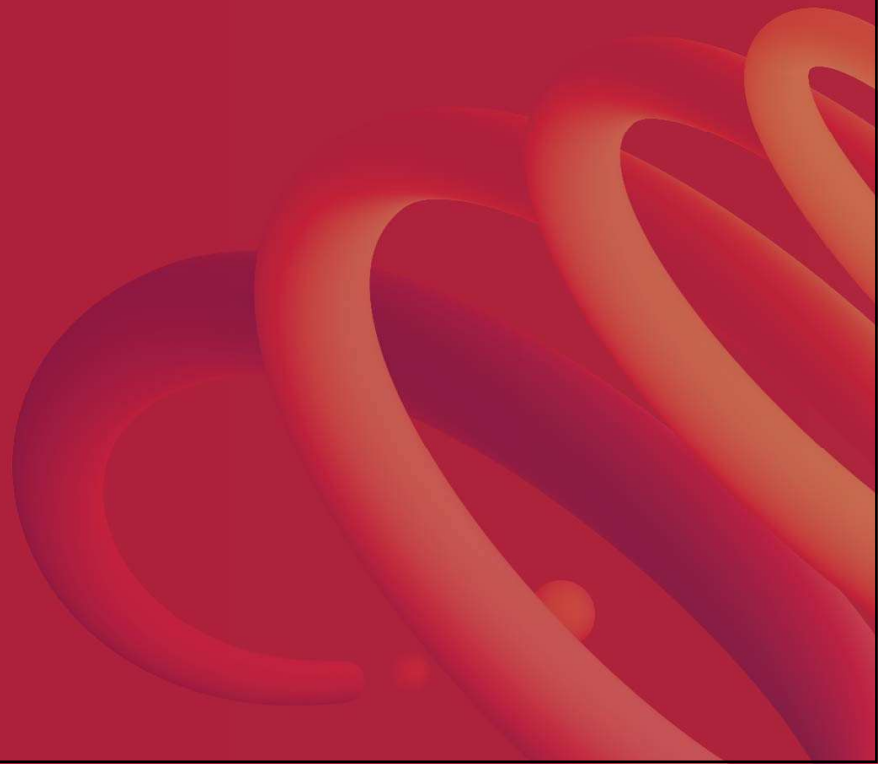
The Big Tesla Hack: A hacker gained control over the entire fleet, but fortunately he's a good guy

Fred Lambert - Aug. 27th 2020 3:29 pm ET @FredericLambert



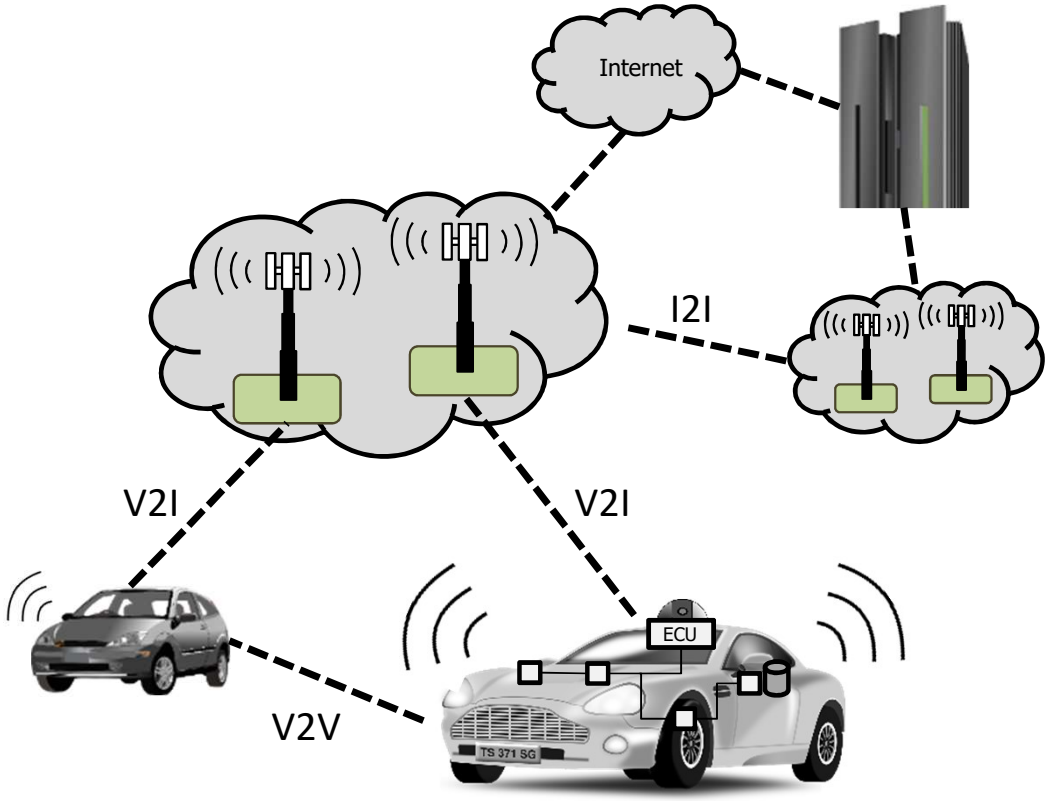
And the recent months are full of more evidence ...

So, what's wrong about the current autonomous vehicles ecosystem?

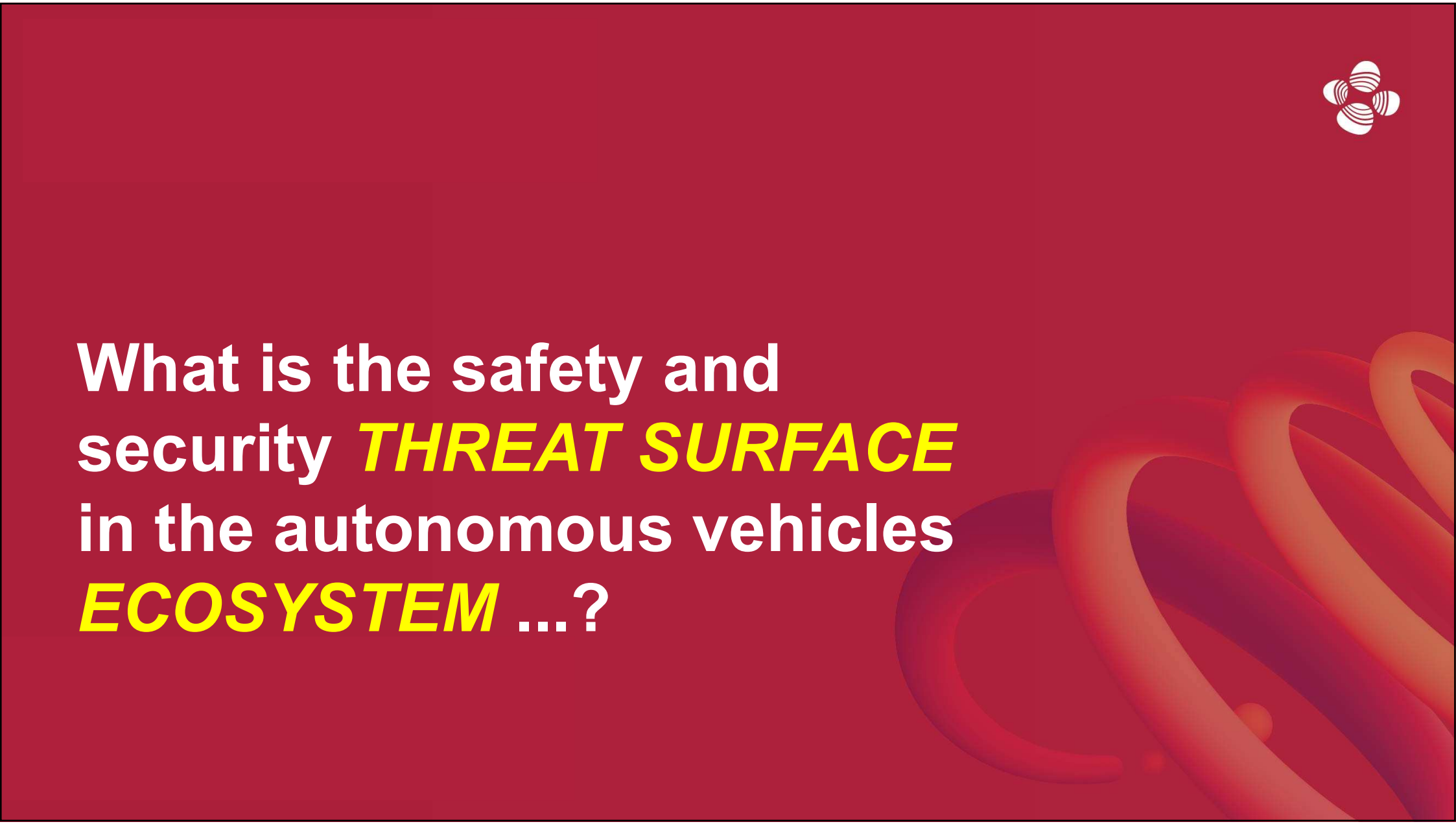


- *To start with, the very notion that there is an ecosystem is inexistent*
- *An analysis of the ecosystem as a critical infrastructure is missing*

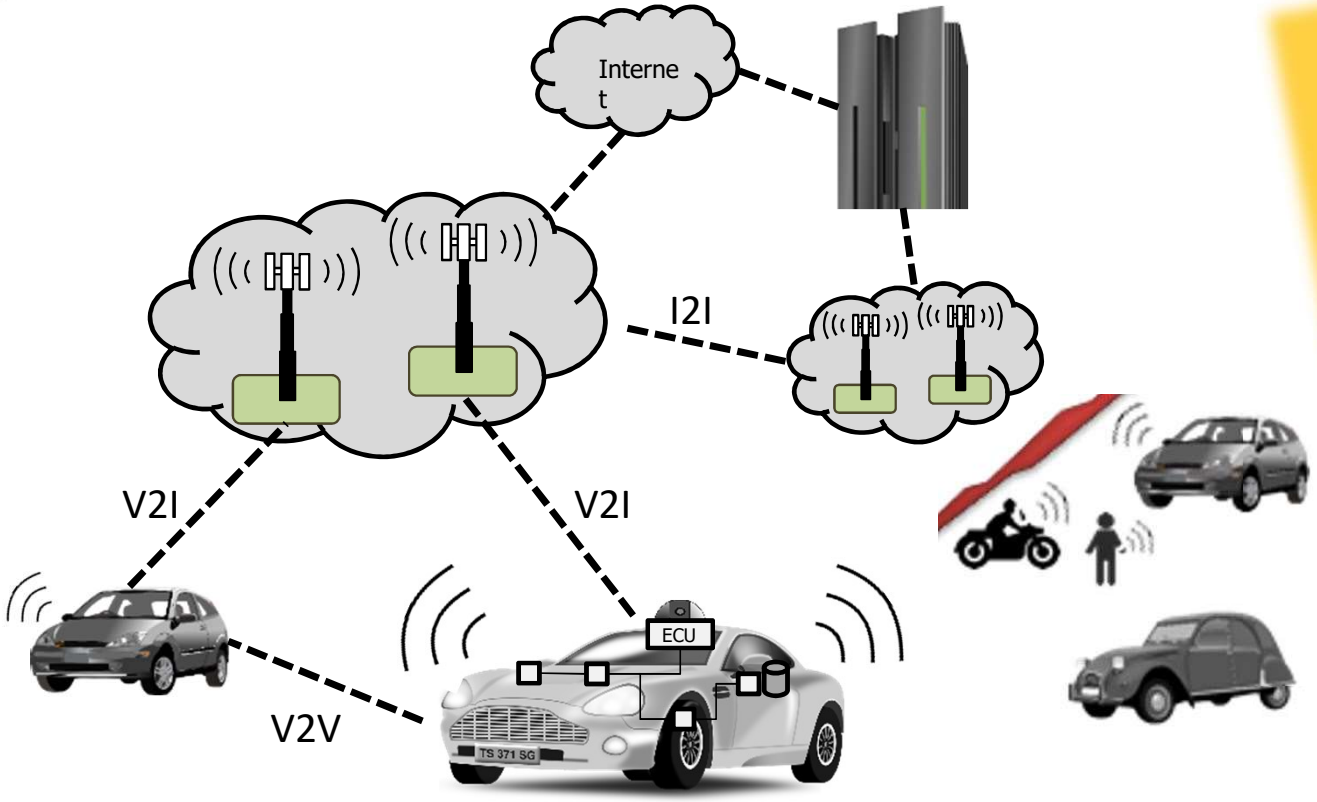
Autonomous Vehicle Ecosystem



Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. Lima, A; Rocha, F; Volp, M; Verissimo, P. in Proc's 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (2016, October) @CCS, Vienna-Austria


What is the safety and
security THREAT SURFACE
in the autonomous vehicles
ECOSYSTEM ...?

Autonomous Vehicle Ecosystem

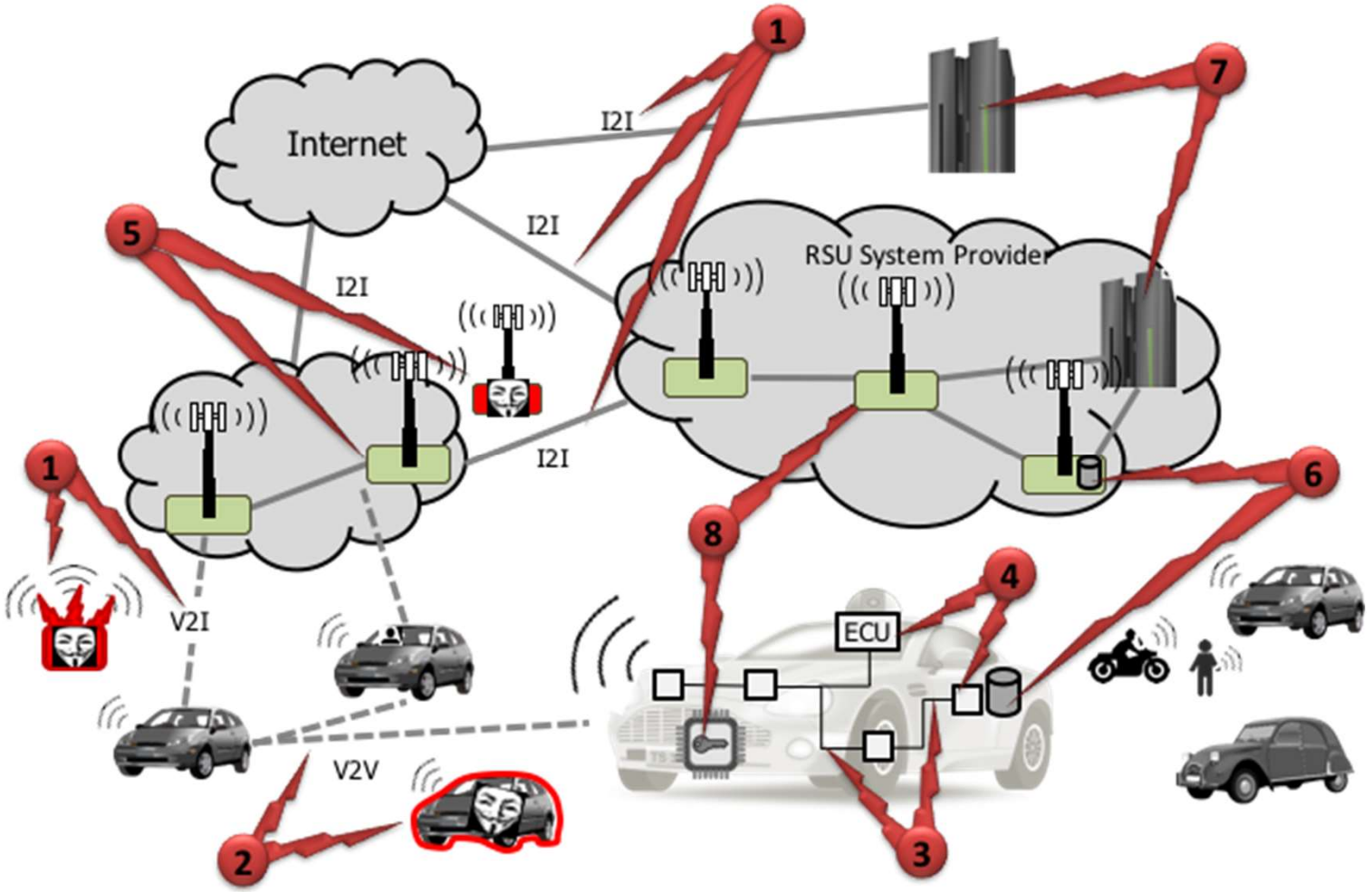


FIRST COMPREHENSIVE STUDY OF THE THREAT SURFACE AND SAFETY-SECURITY GAP OF AUTONOMOUS AND COOPERATIVE VEHICLE ECOSYSTEMS

2016

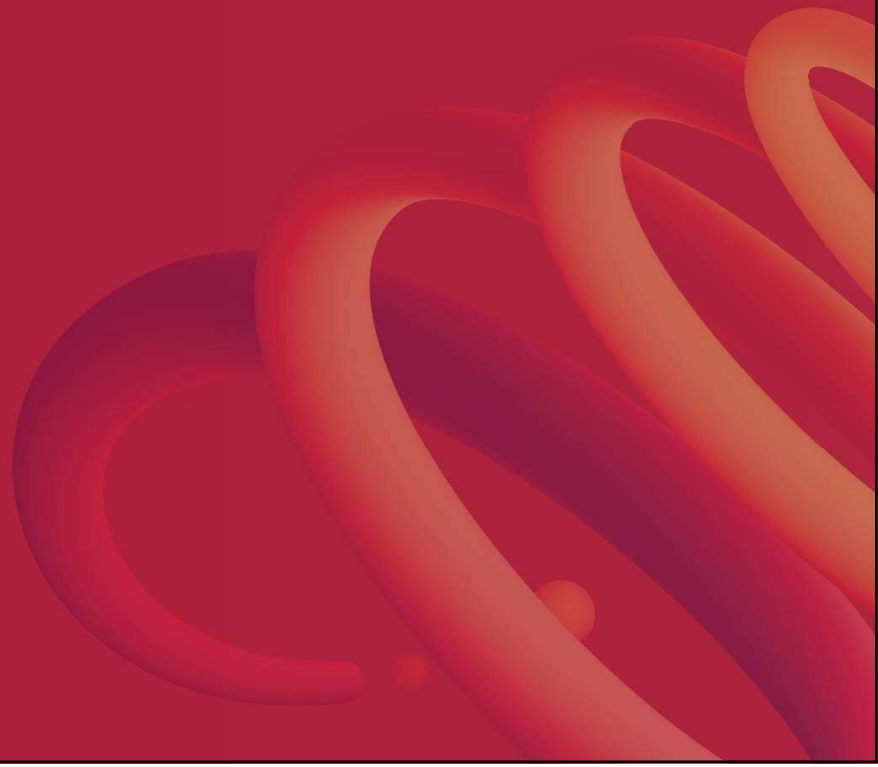
Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. Lima, A; Rocha, F; Volp, M; Verissimo, P. in Proc's 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (2016, October) @CCS, Vienna-Austria

Autonomous vehicle ecosystem
threat surface perhaps wider than many think



Threat Vectors

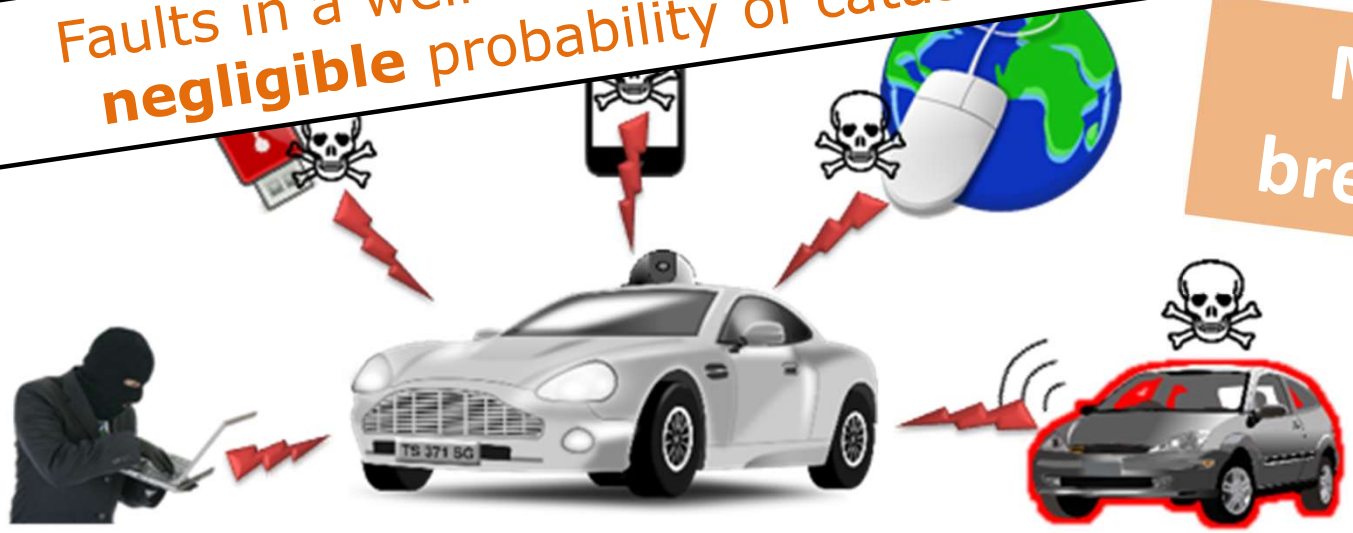
The
SAFETY GAP
in the autonomous
vehicles area ...



Safety gap in vehicle ecosystems

Faults in a well designed car ecosystem lead to an **infinitesimal and acceptable** probability of catastrophic failure;

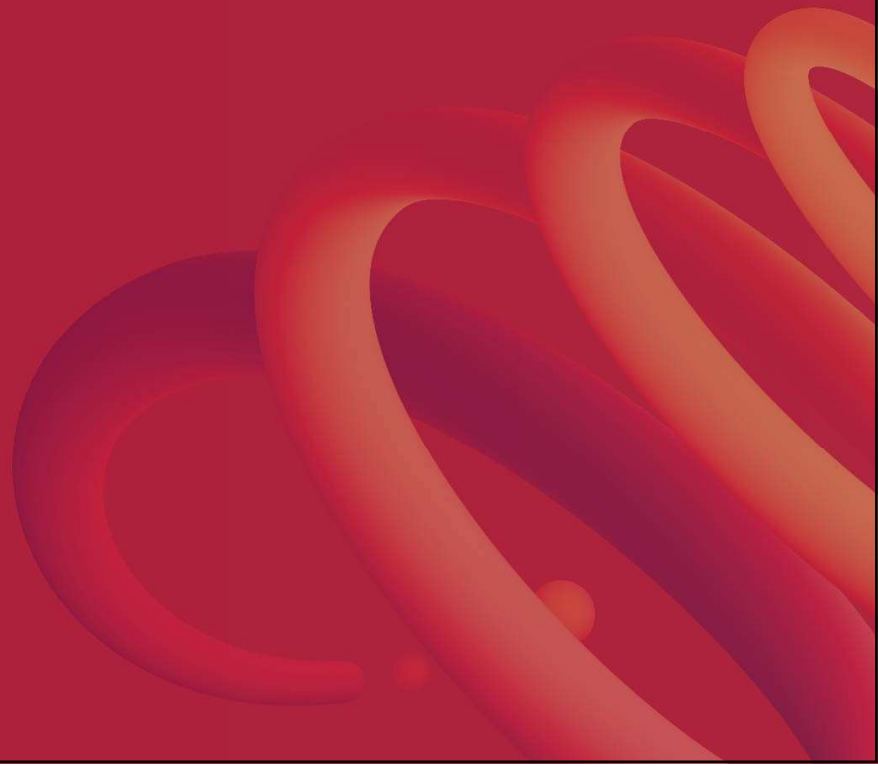
Faults in a well designed car may imply a **non-negligible** probability of catastrophic failure



Move fast break things?

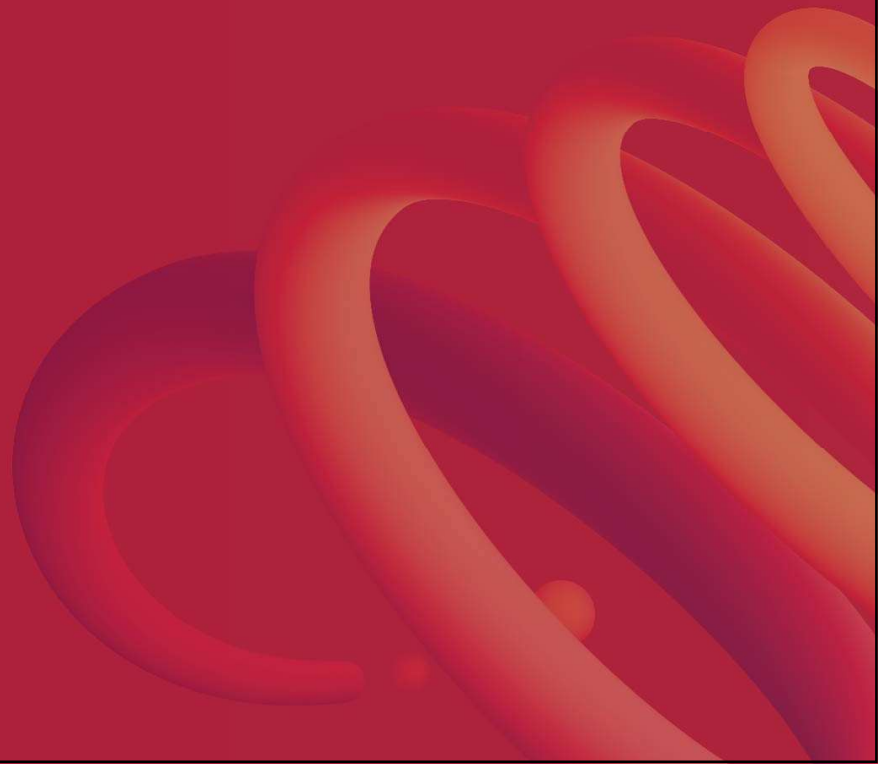
Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. Lima, A; Rocha, F; Volp, M; Verissimo, P. in Proc's 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (2016, October) @CCS, Vienna-Austria

So...
maybe those reported accidents ... were not really just bad luck?



But it can get worse:

The
SAFETY-SECURITY GAP
in the autonomous
vehicles area ...
... *(land, air, space)*



Safety-security gap in vehicle ecosystems

Faults in a well designed car ecosystem lead to an **infinitesimal and acceptable** probability of catastrophic failure;

Faults in a well designed car may imply a **non-negligible** probability of catastrophic failure

Vulnerabilities in a car ecosystem **will** lead, rather sooner than later, to catastrophic failures;

Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. Lima, A; Rocha, F; Volp, M; Verissimo, P. in Proc's 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (2016, October) @CCS, Vienna-Austria

How serious is that?

«IF IT AIN'T SECURE, IT AIN'T SAFE»

Safety-security gap in vehicle ecosystems

Faults in a well designed car ecosystem lead to an **infinitesimal and acceptable** probability of catastrophic failure;

Faults in a well designed car may imply a **non-negligible** probability of catastrophic failure

Vulnerabilities in a car ecosystem **will** lead, rather sooner than later, to catastrophic failures;

Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. Lima, A; Rocha, F; Volp, M; Verissimo, P. in Proc's 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (2016, October) @CCS, Vienna-Austria

But maybe the AI/ML driven world is a bit more secure or safer?

It can get really bad…
BAD as in 'blind'

https://www.reddit.com/r/ThatsInsane/comments/r3fxpi/tesla_radar_did_not_recognize_a_camel_cusing_an/?rdt=49822

Tesla vision did not recognize a camel, causing an accident in the UAE

LESSONS LEARNT?

Snake?
Dust Cloud? Bush?
Naaah, nothing ahead!

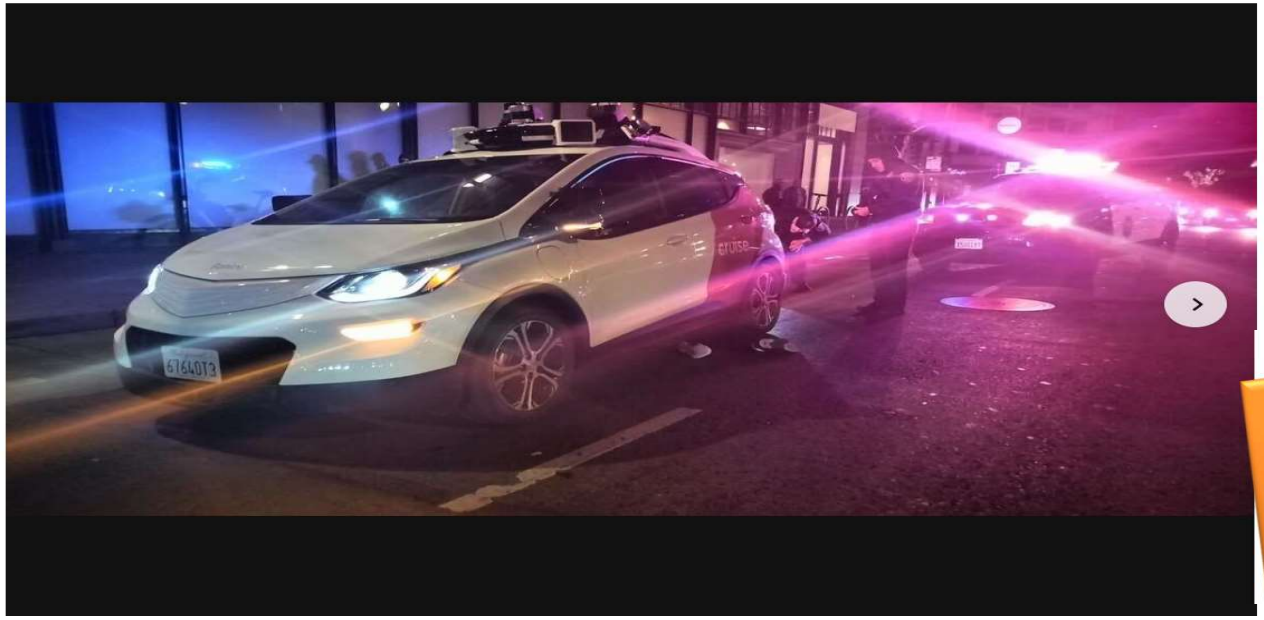
Cruise driverless car runs over woman and stops

BAY AREA // SAN FRANCISCO

Driver hits woman in S.F., then Cruise driverless car runs her over; photo shows victim trapped

Jordan Parker, Nora Mishanec

Oct. 2, 2023 | Updated: Oct. 3, 2023 3:52 p.m.



LESSONS LEARNT?

One of Uber's Self-Driving Cars Hit and Killed a Woman in Arizona

Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk

The automated car lacked "the capability to classify an object as a pedestrian unless that object was near a crosswalk," an NTSB report said.

DISASTERS

Dashcam video of deadly self-driving Uber crash released

By Nicole Darrah , · Fox News

Published March 22, 2018 12:45am EDT | Updated March 22, 2018 12:59am EDT

Dashcam catches the moment self-driving Uber hits pedestrian

LESSONS LEARNT?

The serious ecosystem security risks

Solutions? ...

COMPONENT-BASED, INDIVIDUALIZED

ATTACK PREVENTION, ACCESS CONTROL, FWALLS, ETC.

VULNERABILITY PREVENTION AND REMOVAL

HUMAN-STEERED AD-HOC MITIGATION

A long journey towards

RESILIENT AUTONOMOUS VEHICLE ECOSYSTEMS

Two+ decades of joint work with several colleagues at Navigators/LASIGE@ULISBOA, CRITIX@UNILU (and many others from Univ/R&D across EU).
More recently, A. Shoker and R. Yasmin at CybeResil@KAUST, M.Voelp CRITIX@UNILU, V. Rahli @U.BIRMINGHAM, J. Decouchant@U.DELFT

CORTEX Project Info *[2001-04]*

INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME

Project acronym: ***CORTEX***
Project full title:
CO-operating Real-time senTient objects: architecture and EXperimental evaluation

- Members:
  - Univ. Lisboa Fac. Of Sciences (PT) **(proj. coord.)**
  - Trinity College of Dublin (IR)
  - U. of Lancaster (UK)
  - U. of Ulm (DE)
- Duration:
  - 3 years, starting April 2001
- Budget:
  - 2 MEURO

(Philosophical) ambition:

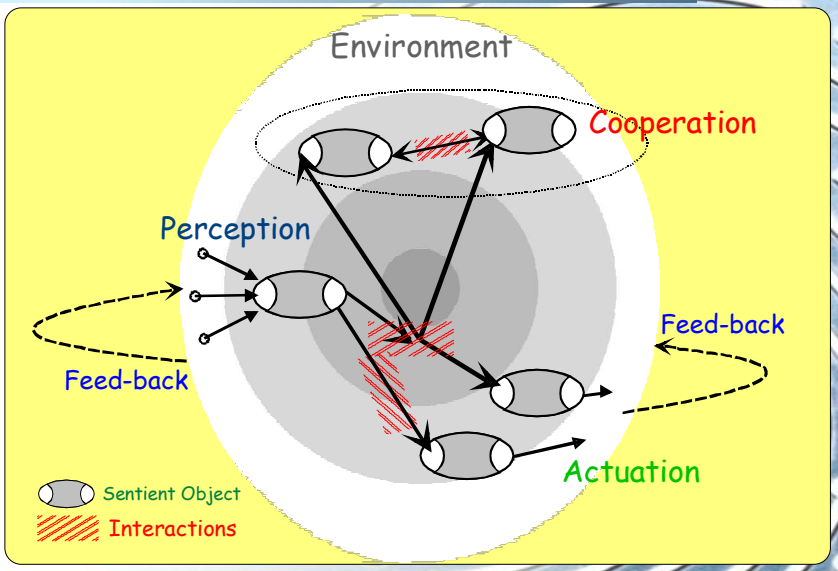
«Control the physics of event interleaving of autonomous object ecosystems, interacting in real time, reconciling the (needed) predictability with the (unavoidable) uncertainty of open environments»

Cortex project [2001]

'Sentient objects' interaction model

Abstract safe distributed real-time (DRT) autonomous control of free-running objects

- should support the needed **classes of R/T interactions**: environment-object, object-object
- **context awareness and sentience** of body and of environment;
- **generic and semantics-agnostic predicates for predictability and correctness in face of uncertainty**



[P. Veríssimo and A. Casimiro. The Timely Computing Base Model and Architecture. IEEE Tacs. on Computers, 2002]


KARYON PROJECT :
Kernel-Based
ARchitecture for safetY-
critical cONtrol

KARYON PROJECT: Kernel-Based ARchitecture for safetY-critical cONtrol

2011-2014

Academia & Research Institutes
SMEs and Industry

Proof-of-concept prototypes
Simulations

FACULDADE DE CIÊNCIAS UNIVERSIDADE DE LISBOA

gmv INNOVATING SOLUTIONS

EMBRAER

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

SP Science Partner

CHALMERS UNIVERSITY OF TECHNOLOGY

4S Group Technology for Sustainability

Avionics

UAS/Aircraft flight mission

Automotive

Adaptive cruise control
Coordinated lane change
Coordinated intersection crossing

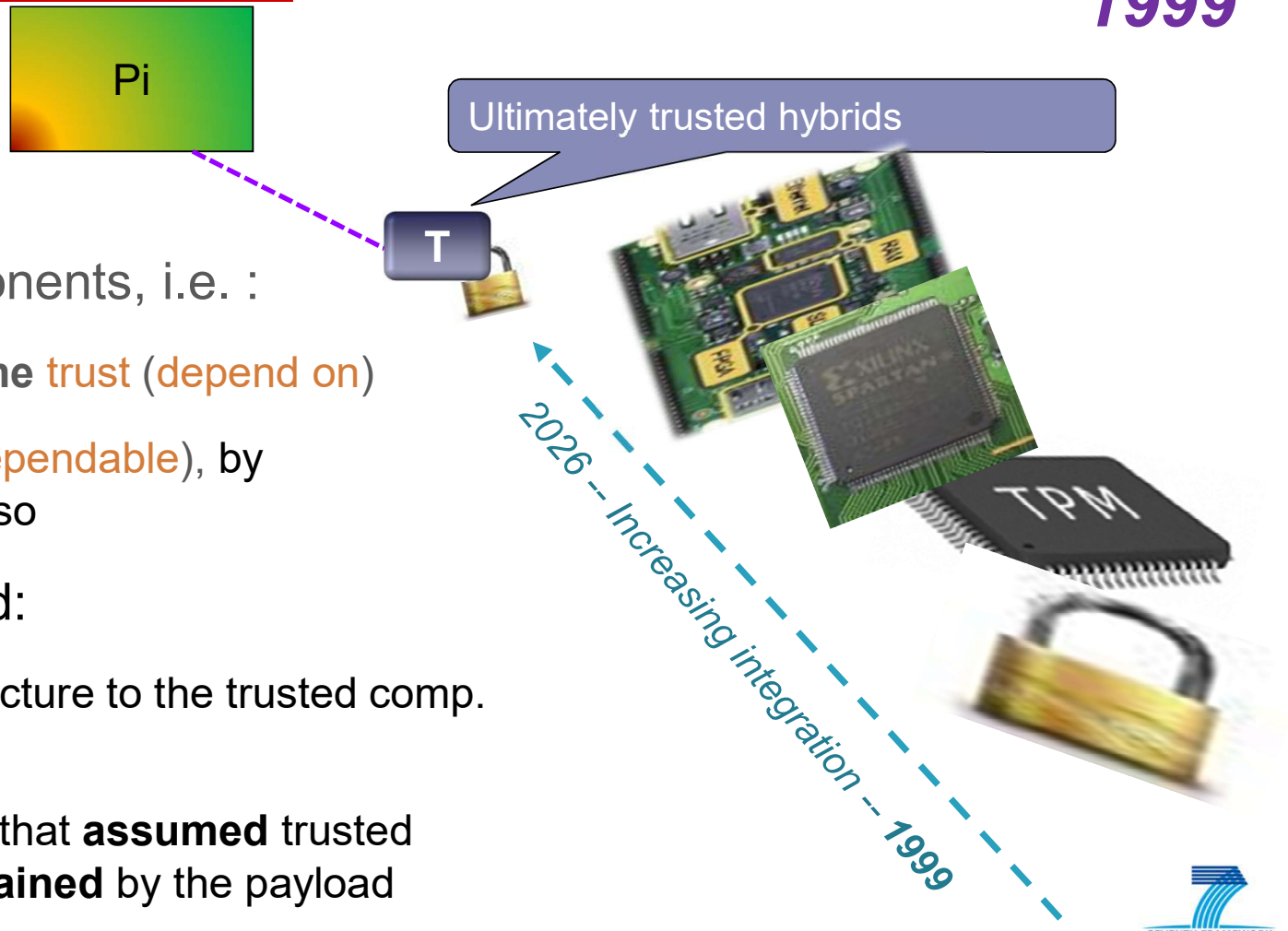
- Provide system solutions for predictable and safe coordination of smart vehicles that autonomously cooperate and interact in an open and inherently uncertain environment

33

Divide-and-conquer I: Architectural Hybridisation

Models and architectures Giving substance to assumptions

1999



- ***Trusted-Trustworthy*** components, i.e. :
  - components on which you **assume** trust (depend on)
  - because they are trustworthy (dependable), by **construction**, and verified to be so
- Not enough! We further need:
  - a ***hybrid architecture*** giving structure to the trusted comp. plus payload compound
  - a ***computational model*** proving that **assumed** trusted comp. properties **are indeed obtained** by the payload

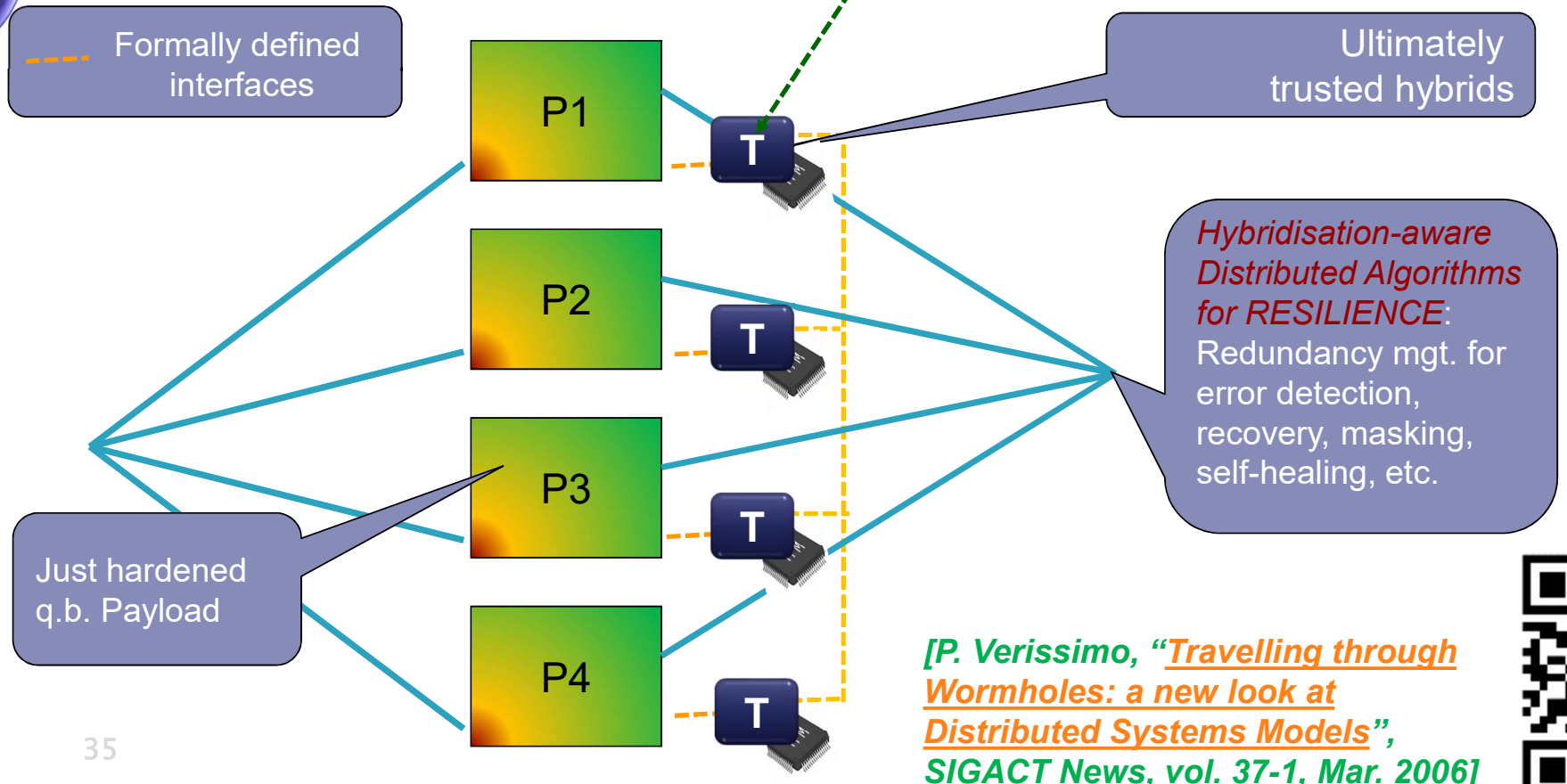
KARYON

SEVENTH FRAMEWORK PROGRAMME

Hybridisation-aware distributed algorithms, models, and architectures

Hybridisation-aware algorithms: models, architect., and control

Leveraging trusted-trustworthy components and TEE, with the right set of simple functions (failure detectors, monotonic counters, reliable timers or clocks, PRG, signatures, indelible logs, binary consensus)



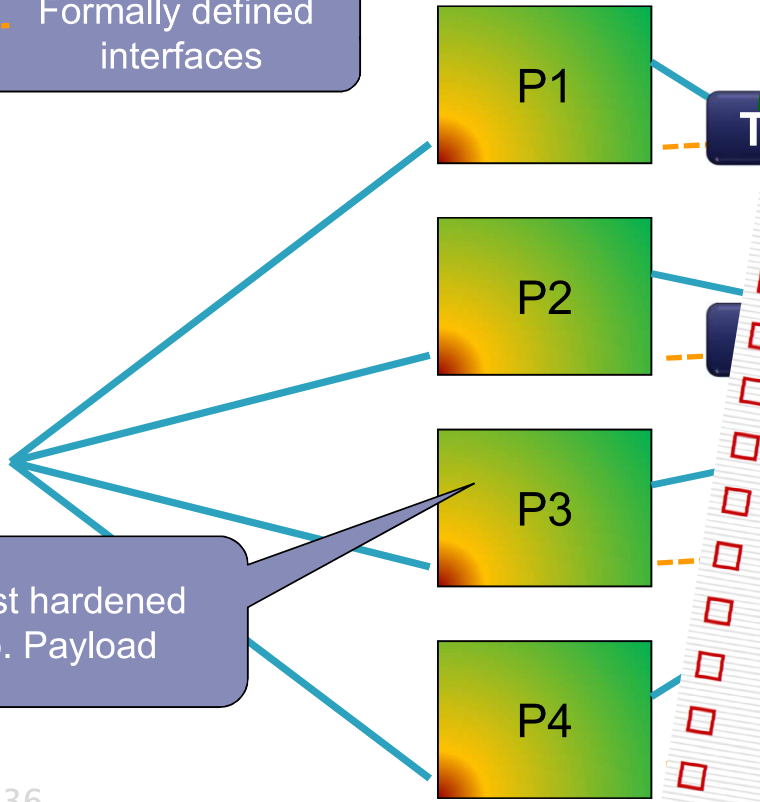
[P. Verissimo, "Travelling through Wormholes: a new look at Distributed Systems Models", SIGACT News, vol. 37-1, Mar. 2006]

Hybridisation-aware distributed algorithms, models, and architectures

Leveraging trusted-trustworthy components and TEE, with ... (failure detectors, monotonic counters, reliable timers or clocks, PRG, signa...

Hybridisation-aware algorithms: models, architect., and control

Formally defined interfaces



Just hardened q.b. Payload

Some example works on explicit or implicit hybrid models:

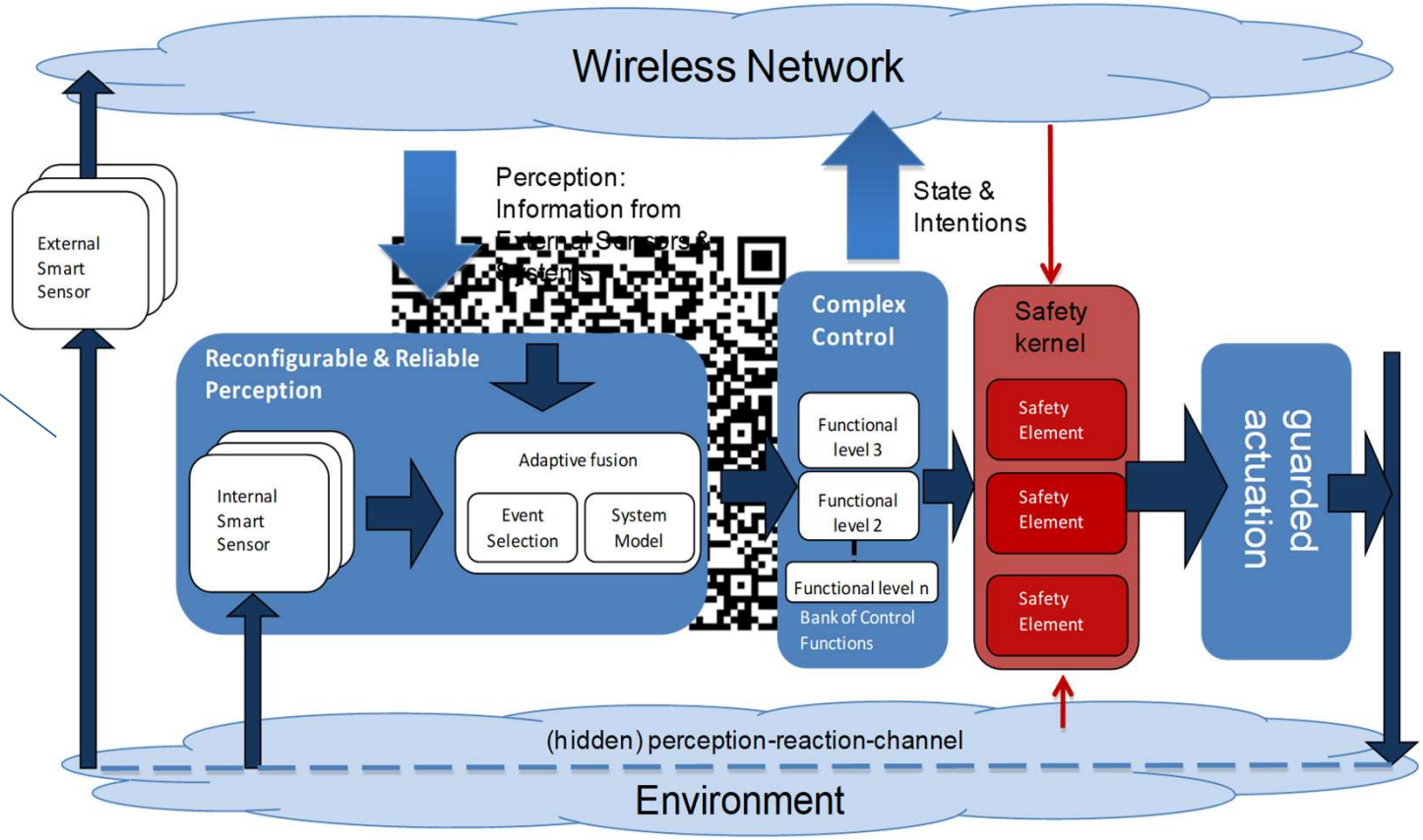
- Aguilera et al.
- Babay, Amir et al.
- Baldoni et al.
- Chun, Maniatis, Kubiatowicz et al.
- Clement, Junqueira, et al.
- Distler, Kapitza, Reiser et al.
- Guerraoui, Vukolic et al.
- Krishnamurthy, Sanders, Cukier
- Kapitza, Cachin et al.
- Levin, Douceur et al.
- Liu, Asokan et al.
- Malkhi et al.
- Nogueira R., Raynal et al.
- Ramasamy, Agbaria, Sanders
- Roeder, Schneider
- Vukolic et al.
- Weisberg, Dolev et al.

2006]

KARYON architectural view: proof of concept of *modularity* and *hybridisation* for safety

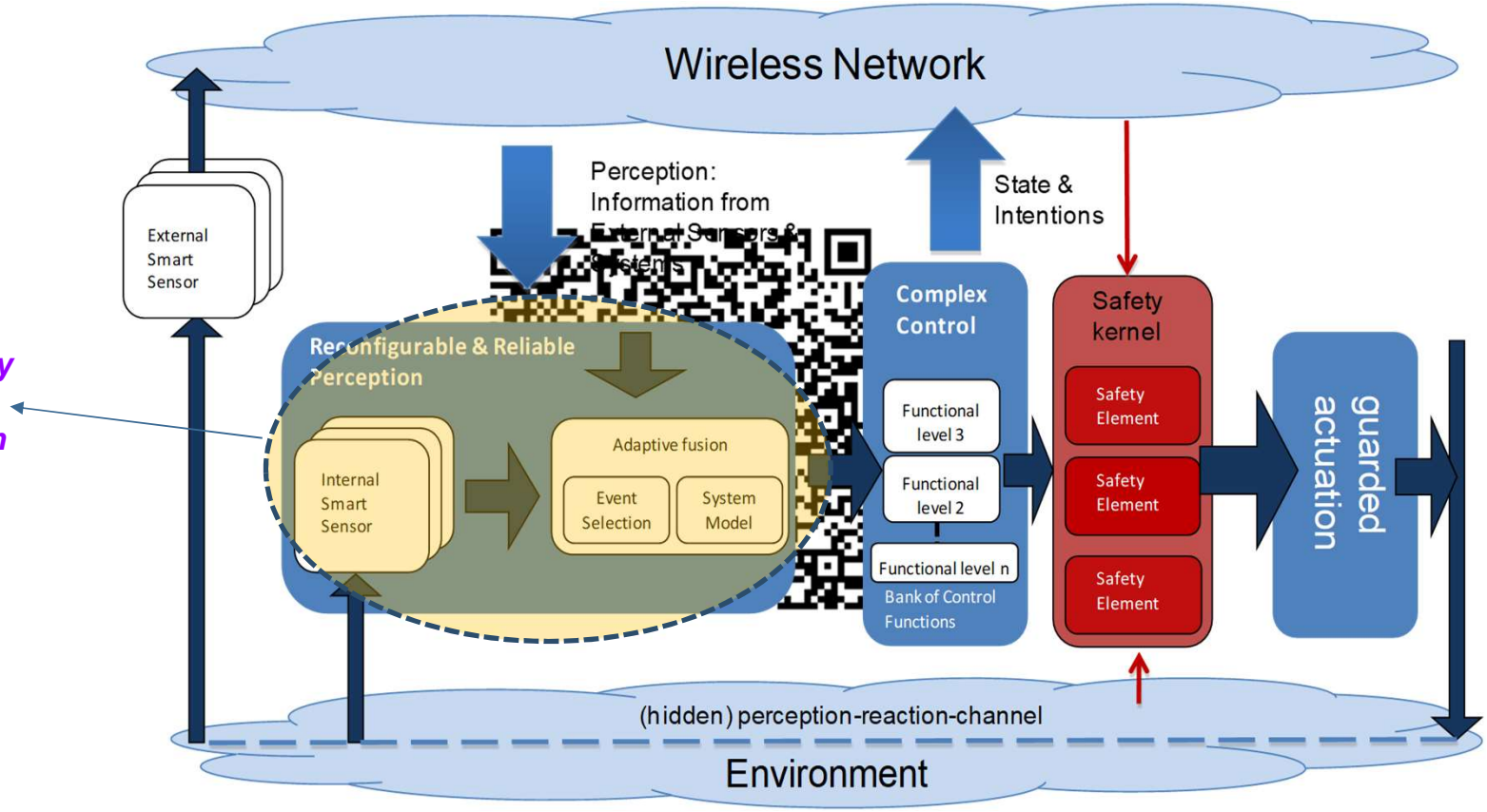
[2011-2014]

modularity and hybridisation for safety



A. Casimiro, J. Kaiser, E. Schiller, P. Costa, J. Parizi, R. Johansson, R. Librino, "The KARYON Project: Predictable and Safe Coordination in Cooperative Vehicular Systems", in 2nd Workshop on Open Resilient Human-aware CPS (WORCS'13), Jun. 2013.

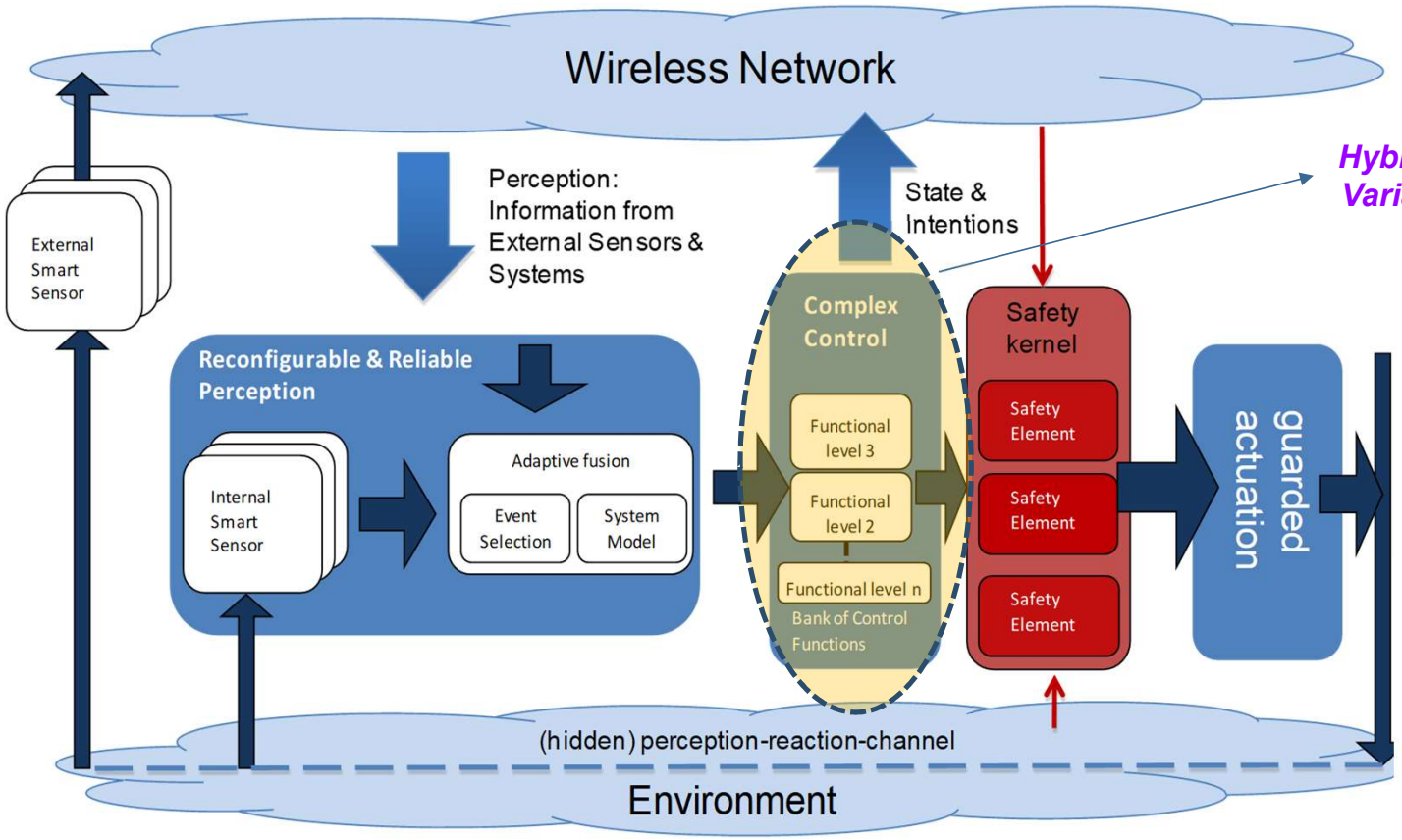
KARYON architectural view: proof of concept of modularity and hybridisation for safety



Sensor Diversity
Correlation
Adaptive fusion

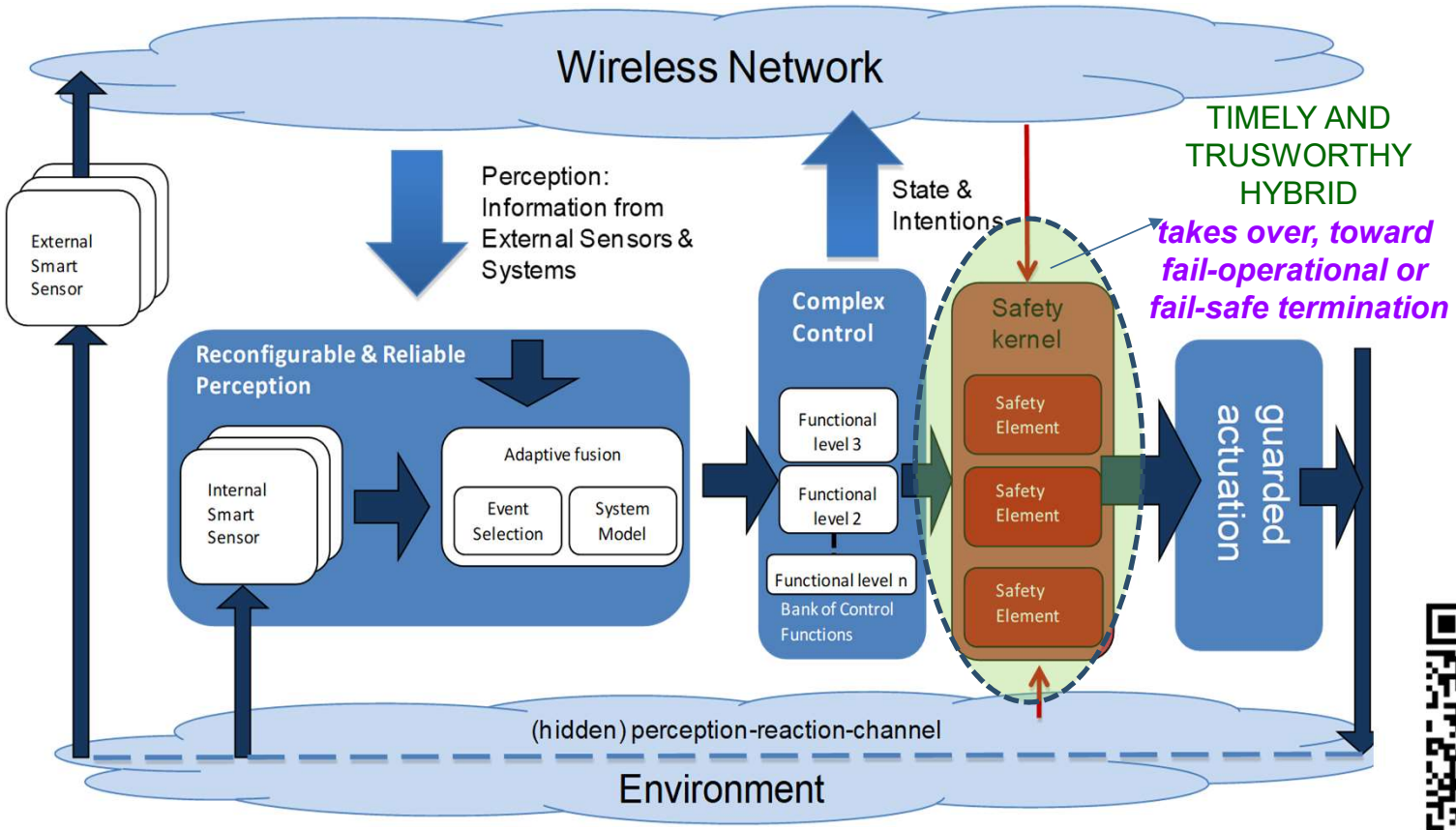
A. Casimiro, J. Kaiser, E. Schiller, P. Costa, J. Parizi, R. Johansson, R. Librino, "The KARYON Project: Predictable and Safe Coordination in Cooperative Vehicular Systems", in 2nd Workshop on Open Resilient Human-aware CPS (WORCS'13), Jun. 2013.

KARYON architectural view: proof of concept of modularity and hybridisation for safety



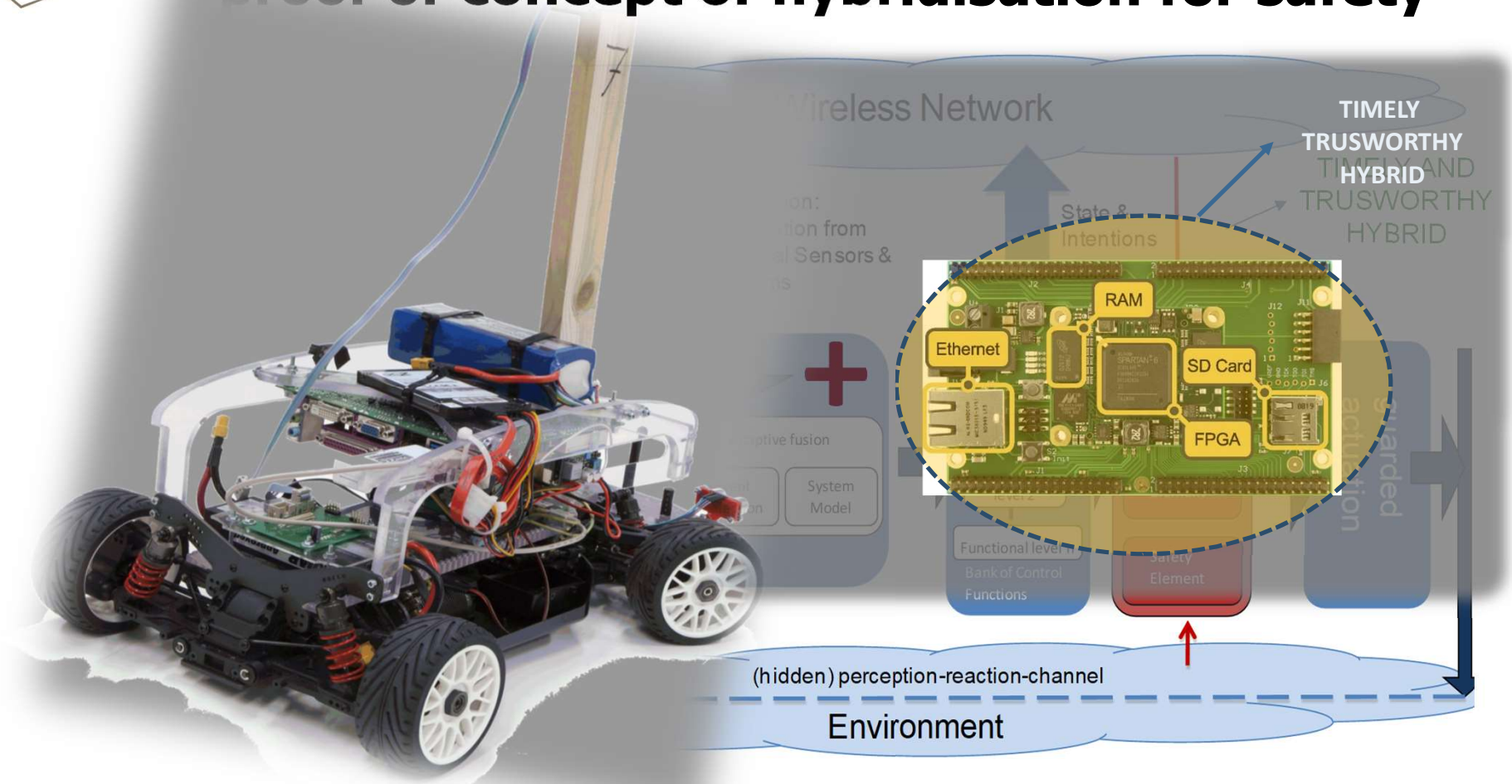
A. Casimiro, J. Kaiser, E. Schiller, P. Costa, J. Parizi, R. Johansson, R. Librino, "The KARYON Project: Predictable and Safe Coordination in Cooperative Vehicular Systems", in 2nd Workshop on Open Resilient Human-aware CPS (WORCS'13), Jun. 2013.

KARYON architectural view: proof of concept of modularity and hybridisation for safety



A. Casimiro, J. Kaiser, E. Schiller, P. Costa, J. Parizi, R. Johansson, R. Librino, "The KARYON Project: Predictable and Safe Coordination in Cooperative Vehicular Systems", in 2nd Workshop on Open Resilient Human-aware CPS (WORCS'13), Jun. 2013.

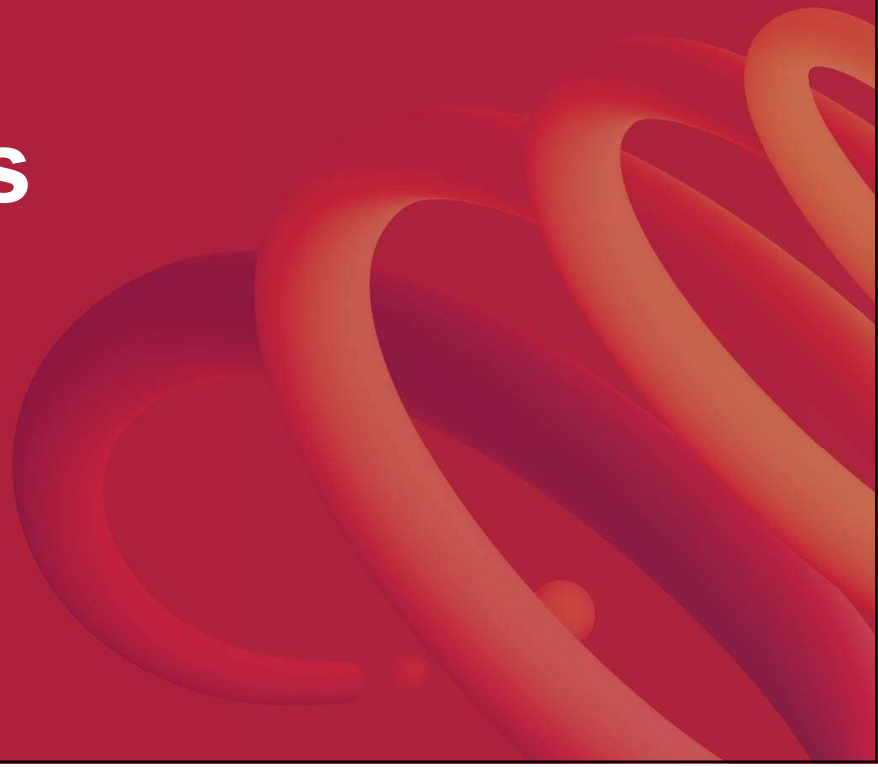
KARYON architectural view: proof of concept of hybridisation for safety



A. Casimiro, J. Kaiser, E. Schiller, P. Costa, J. Parizi, R. Johansson, R. Librino, "The KARYON Project: Predictable and Safe Coordination in Cooperative Vehicular Systems", in 2nd Workshop on Open Resilient Human-aware CPS (WORCS'13), Jun. 2013.

Intel Collaborative Research Institute (ICRI)

Collaborative Autonomous & Resilient Systems (CARS)



Intel Collaborative Research Institute for

Collaborative Autonomous & Resilient Systems *(CARS)*

securityandtrust.lu

CRITIX

https://www.icri-cars.org/

2017-2020

ICRI-CARS » Resilient Autonomy » Mission

- ICRI-CARS
- Mission
- Research Topics
- Principal Investigators
- TU Darmstadt
- Aalto University
- Ruhr-University Bochum
- Critix@ University of Luxembourg
- TU Wien
- Collaborations

Intel Collaborative Research Institute for Collaborative Autonomous & Resilient Systems (ICRI-CARS)

About Collaborative Autonomous and Resilient Systems (CARS)

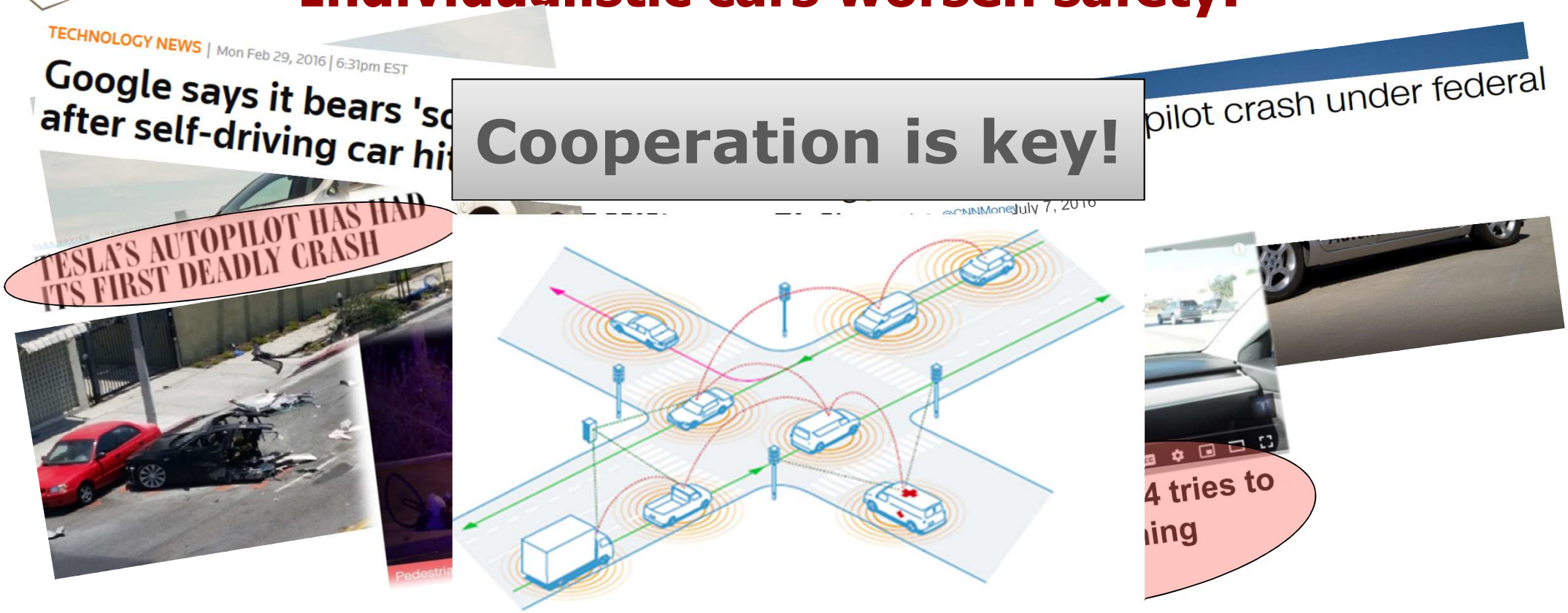
The mission of the ICRI-CARS is the study of security, privacy, and safety of autonomous systems that may collaborate with each other. Examples include drones, self-driving vehicles, or collaborative systems in industrial automation. CARS introduce a new paradigm to computing that is different from conventional systems in a very important way: they must learn, adapt, and evolve with minimal or no supervision. A fundamental question therefore, is what rules and principles should guide the evolution of CARS?

This raises security related questions in multiple research areas:

1. Trustworthy and Controllable Autonomy
2. Fair and Safe Collaboration Tolerating Failures and Attacks
3. Intelligent Security Strategies for Self-Defense and Self-Repair
4. Integration of Safety, Security, and Real-time Guarantees
5. Autonomous Systems, Ecosystem Scenarios, Requirements, Case Studies, and Validation
6. Advanced Platform Security for Long-term Autonomy

Ecosystem approach: Cooperation is key!

Individualistic cars worsen safety!



Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. Lima, A; Rocha, F; Volp, M; Verissimo, P. in Proc's 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (2016, October) @CCS, Vienna-Austria

Real-Time and Byzantine Resilient Digital Twins:
Beyond mere SCADA near-Real-Time Data Dissemination

Modularly build

PISTIS

PISTIS: Real-Time Byzantine Atomic Broadcast

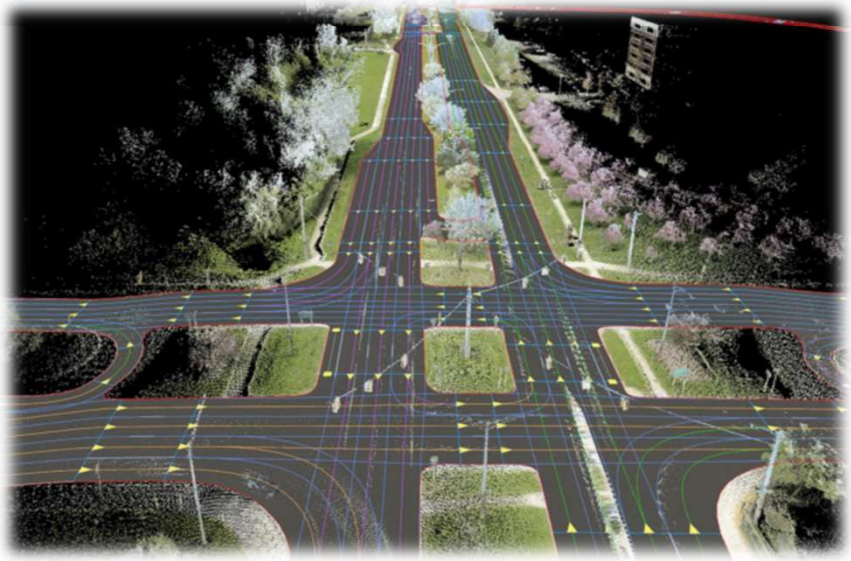
Real-Time Byzantine Consensus

RT-ByzCast

RT-ByzCast: Real-Time Byzantine-Resilient Reliable Broadcast

Poorly behaving network (unbounded probabilistic losses)

intel Research Institute for Collaborative Autonomous and Resilient Systems



Accurate Real-Time Digital Maps for Autonomous Driving

*D. Kozhaya, J. Decouchant and P. Esteves-Veríssimo, "RT-ByzCast: Real-Time Byzantine-Resilient Reliable Broadcast" , IEEE Transactions on Computers 2019, Core A**

*Kozhaya, D., Decouchant, J., Rahli, V., & Esteves-Verissimo, P. (2021). PISTIS: An Event-Triggered Real-time Byzantine Resilient Protocol Suite. IEEE TPDS. doi:10.1109/tpds.2021.3056718, Core A**

Real-Time and Byzantine Resilient Digital Twins:
Beyond mere SCADA near-Real-Time Data Dissemination

Modularly build

PISTIS: Real-Time Byzantine Atomic Broadcast

PISTIS

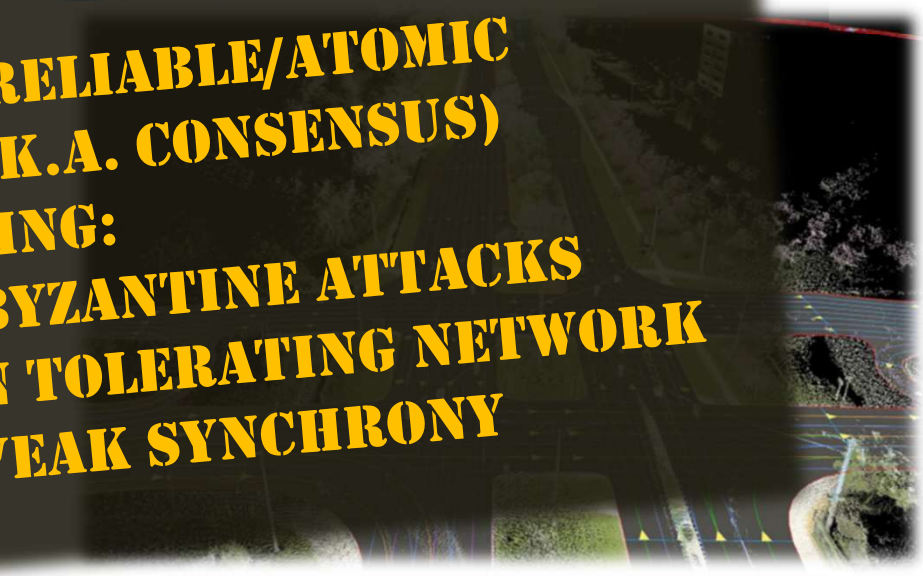
RT-ByzCast

Poorly behaving network
(unbounded probabilistic losses)

intel Research Institute for Collaborative Autonomous and Resilient Systems

WORLD-FIRST BYZANTINE RELIABLE/ATOMIC BROADCAST PROTOCOL (A.K.A. CONSENSUS) SIMULTANEOUSLY PROVIDING:

- **RESILIENCE AGAINST BYZANTINE ATTACKS**
- **REAL-TIME OPERATION TOLERATING NETWORK UNCERTAINTIES AND WEAK SYNCHRONY**



Accurate Real-Time Digital Maps for Autonomous Driving

*D. Kozhaya, J. Decouchant and P. Esteves-Veríssimo, "RT-ByzCast: Real-Time Byzantine-Resilient Reliable Broadcast" , IEEE Transactions on Computers 2019, Core A**

*Kozhaya, D., Decouchant, J., Rahli, V., & Esteves-Verissimo, P. (2021). PISTIS: An Event-Triggered Real-time Byzantine Resilient Protocol Suite. IEEE TPDS. doi:10.1109/tpds.2021.3056718, Core A**

Intrusion Resilience System (IRS)

Trustworthy Autonomous Vehicles Architecture (SAVVY)

KAUST In-house Projects

2021----

Towards sustainable security and safety

In AV control

Towards sustainable security and safety

(inspired by precursor projects Karyon (EU) and ICRI CARS (INTEL)

Resilient DRT autonomous control --- general driving

Collaboration among autonomous vehicles (V2V, V2I)

Fault and intrusion tolerant control in-vehicle by eliminating SPOFs, in particular at operating-system level

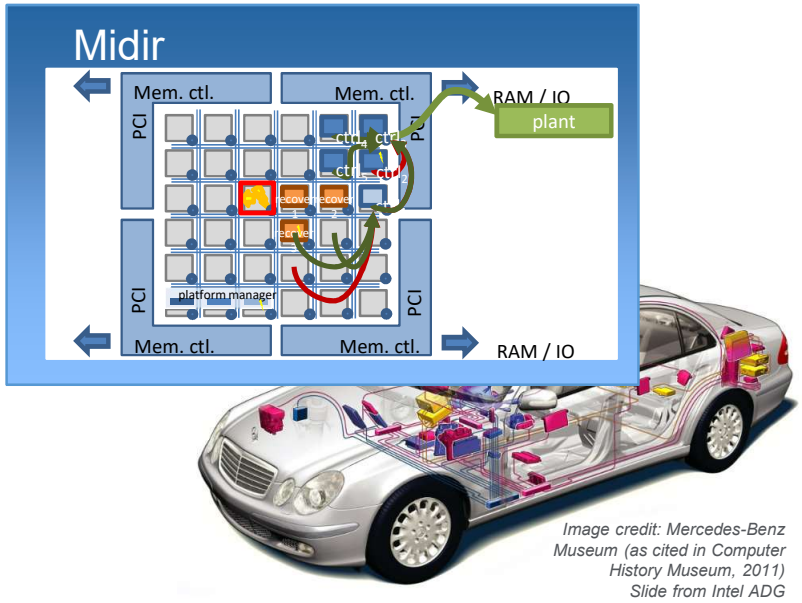


Image credit: Mercedes-Benz Museum (as cited in Computer History Museum, 2011)
Slide from Intel ADG

Intrusion Resilience System (IRS)

Towards sustainable security and safety

In AV control

Intrusion Resilience System (IRS)

The Concept: intrusion masking for real-time fault and intrusion tolerance (R/T FIT)

Our scope: ***In-vehicle systems as Distributed systems of ECUs***

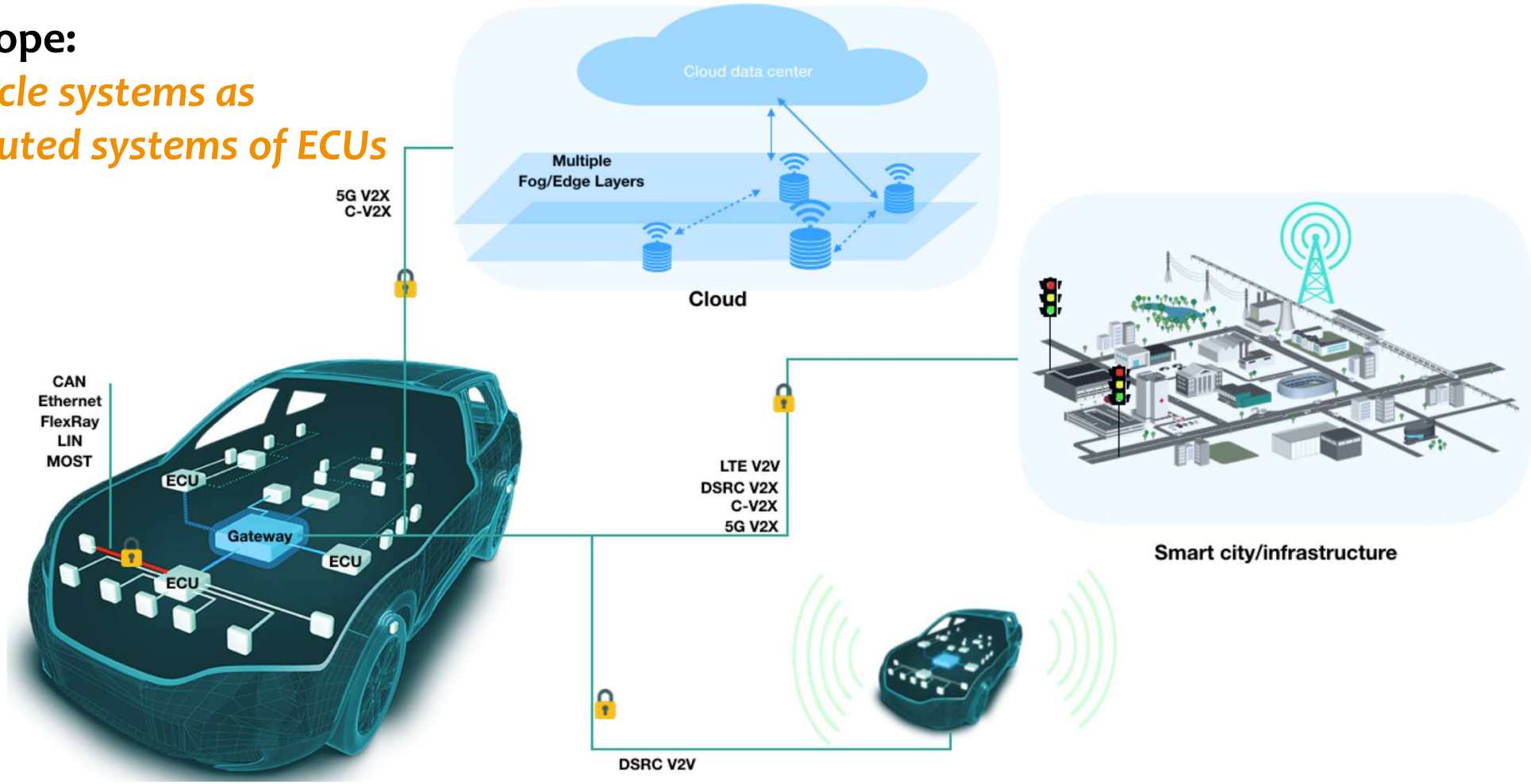
- IRS as a **distributed** service/middleware/library securing critical real-time in-car applications
- ***Distributed State Machines*** over a number of diverse ECUs

A. Shoker, V. Rahli, J. Decouchant and P. Esteves-Verissimo, "Intrusion Resilience Systems for Modern Vehicles," 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023

Automotive Ecosystem

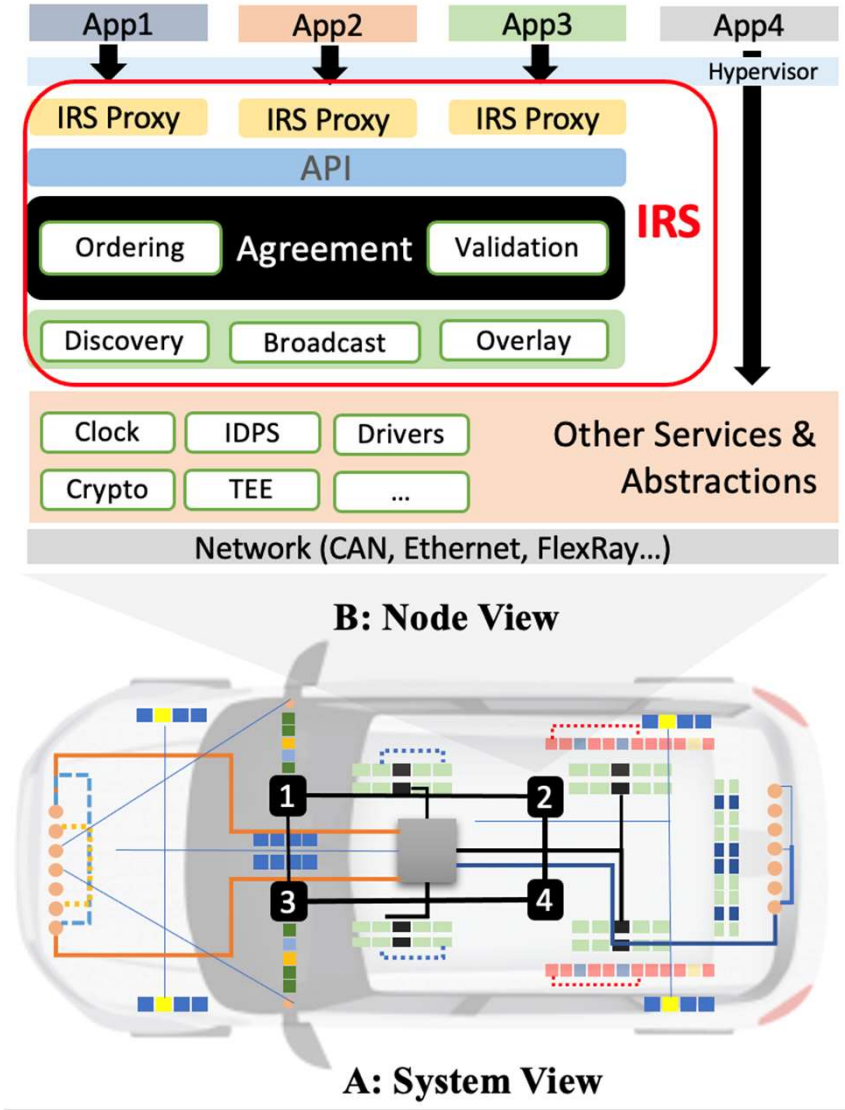
Our scope:

In-vehicle systems as
Distributed systems of ECUs



Intrusion Resilience System (IRS)
The Concept: intrusion masking

- IRS as a **distributed** service/middleware/library
- A critical application (process) is **fully replicated**
- Replicas form a ***Distributed State Machine*** over a number of ECUs
- Decisions are only made through **Byzantine agreement** (BA/BFT)
- Integrity of decisions is guaranteed despite intrusion faults of ***f* out of *N*** (3*f*+1/2*f*+1) replicas



The Path to Fault- and Intrusion-Resilient Manycore Systems on a Chip

The «YES WE CAN» paper

- ***distributed, parallelized, reconfigurable, heterogeneous…***
  - the very features that cause many of the imminent and emerging security and resilience challenges, can, through ...
- ***replication, hybridization, diversity, rejuvenation, adaptation,***
  - also open avenues for their cure through SoC architecting …

- This disruptive paper (@DSN2023 Disrupt track) suggests paths across the entire SoC hardware/software stack.
- **Modular FIT in modern cars offers a promising application domain**

Shoker, P. Esteves-Verissimo and M. Völp, "The Path to Fault- and Intrusion-Resilient Manycore Systems on a Chip," 53rd IEEE/IFIP DSN Int'l Conference, Disrupt Track (DSN-S), Porto, Portugal, 2023. doi: 10.1109/DSN-S58398.2023.00043.

Distributed Systems-on-a-Chip (DisSoC)

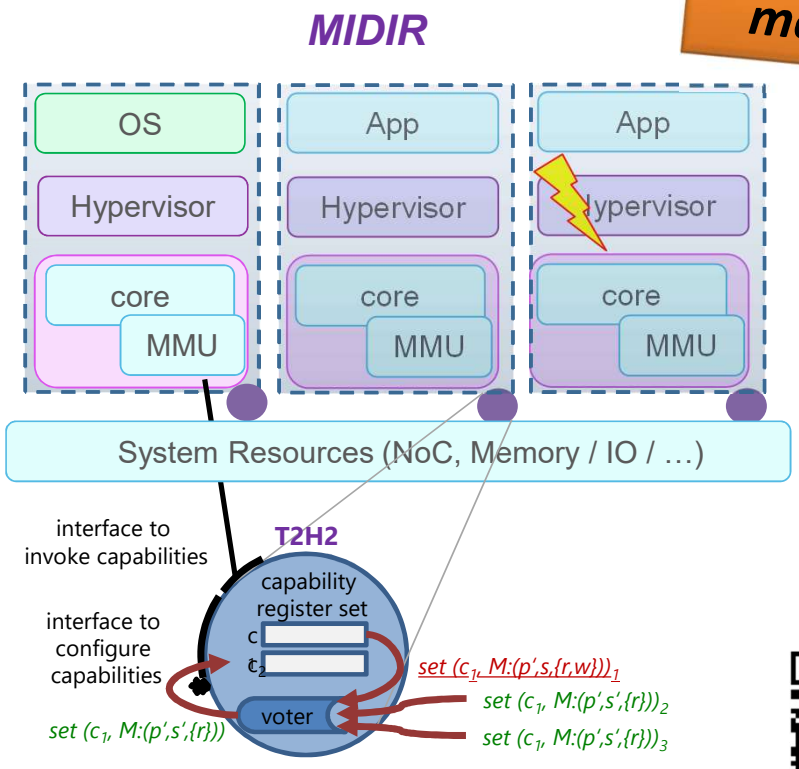
leveraging Ultra-resilient minimal roots-of-trust

[2016]

The Enabling mechanism

=> Patent applications

- Threats have been permeating all levels of architecture.
- And we are always one step "late":
  - we rely on high-level protection (Paxos, BFT,…)
  - threats haunt below (hyp, ME, hw)
  - lost battle: general 0-defect infeasible
- Leverage properties of manycore systems:
  - **distributed systems-on-a-chip (DisSoC)**
  - reinstantiate protection techniques at low enough level (detection, self-check, tolerance)



Behind the Last Line of Defense -- Surviving SoC Faults and Intrusions. Pinto Gouveia, Ines; Voelp, Marcus; Esteves-Verissimo, Paulo. arXiv preprint arXiv:2005.04096 (2020).
Computers & Security, Vol.123, 2022, https://doi.org/10.1016/j.cose.2022.102920.

Trustworthy Autonomous Vehicles Architecture (SAVVY)

Towards sustainable security and safety *In AV control*

Savvy: Trustworthy AI/ML powered Autonomous Vehicles Architecture

Homogeneous ML-based systems cannot give strong assurance and resilience guarantees

Revisit the current fundamentals of GPT based safety-critical AV architectures:

(i) finding a balance between **intelligence** **and** **trustworthiness**, considering *efficiency and functionality* brought in by AI/ML, while prioritizing indispensable *safety and security*;

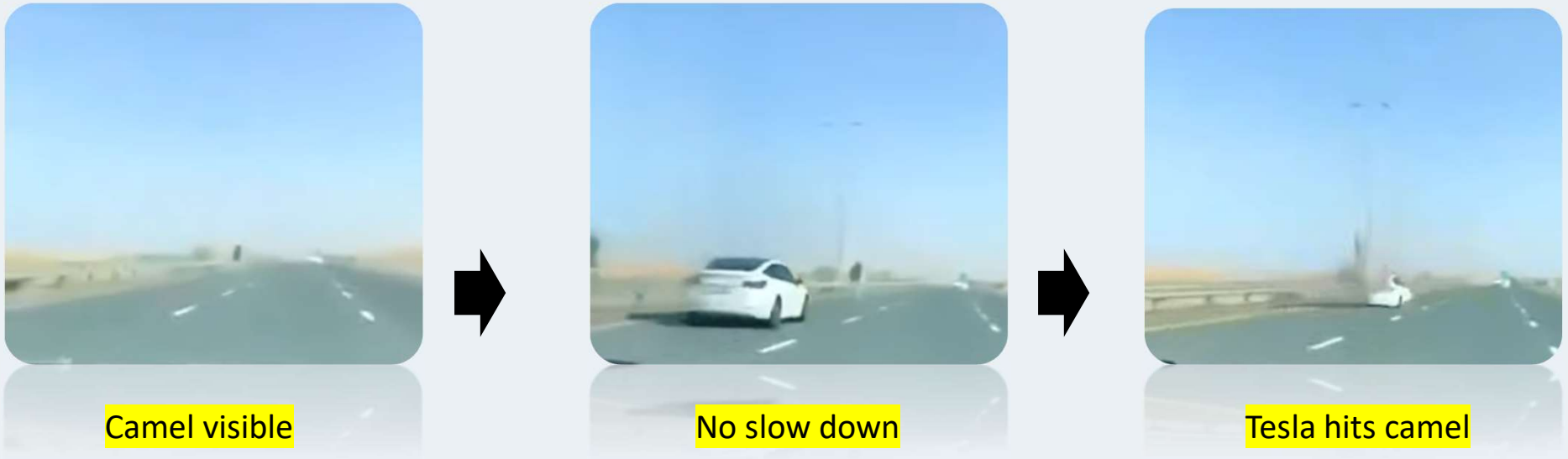
(ii) developing an advanced architecture reconciling the **stochastic** nature of AI/ML (***uncertainty***) with the **determinism** of driving control theory (***predictability***)

Ali Shoker, Rehana Yasmin & Paulo Esteves-Verissimo. RC3@KAUST. (Work in progress) . Symposium on Vehicle Security and Privacy (VehicleSec 2024) @NDSS Feb. 2024, San Diego, CA-US. arXiv https://doi.org/10.48550/arXiv.2402.14580

Autonomous Driving under attack

“Adversary”:

Inadequate or insufficient Machine Learning mechanisms!



Ever seen Tesla hit a Camel??

Predicates abstracting the main AI/ML-based AV failure syndromes

- **Issue 1**
  Confusion in Command and Control
  – *(ML model mapping of the controlled process and environment)*

- **Issue 2**
  Better-precise-than-timely (All-or-Nothing)
  – *(ML classification paradigm is timeliness-agnostic)*

Incident Analysis (NTSB & NHTSA)
Tesla, Volvo, GM Cruise, Honda Acura

Issue 1

Confusion in Command and Control

Vehicle has not made any slow-down or braking	• AD system could not make a decision • Late driver handover is being done
Features disabled, ignored sensor inputs	• No reliable system that oversees vehicle state • No reliable system to take over vs. waiting handover forever

Incident Analysis (NTSB & NHTSA)
Tesla, Volvo, GM Cruise, Honda Acura

Issue 1

Confusion in Command and Control

Vehicle has not made any slow-down or braking	• AD system could not make a decision • Driver handover is being done
Features disabled, broken or ignored sensors	• No reliable system that oversees vehicle state • No reliable system to take over vs waiting handover late or forever

Issue 2

ML classification oriented to Better-precise-than-timely (All-or-Nothing)

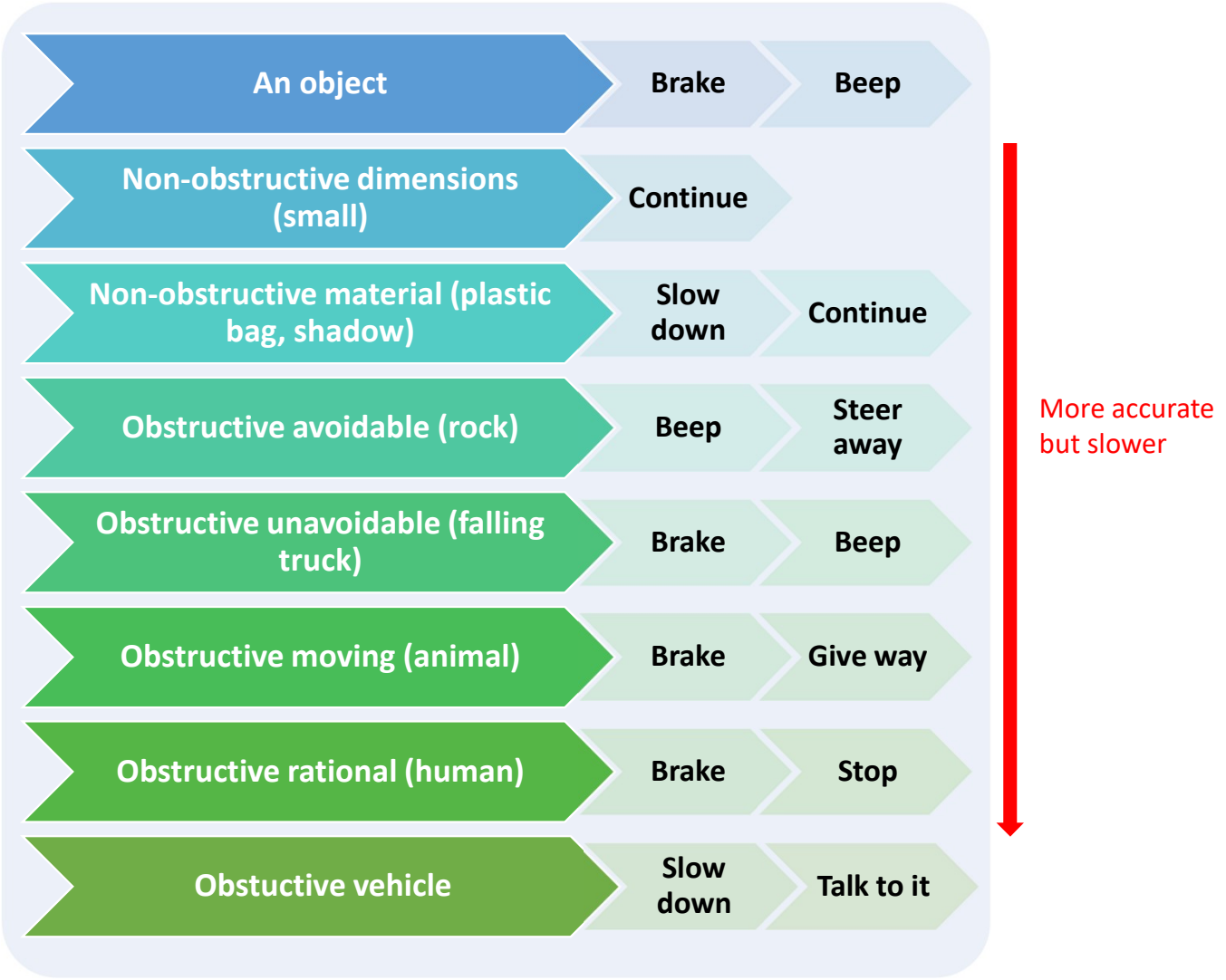
No mentioning to ”invalid” or ”indeterminate” or “not-converging” classification	• ML has not delivered early enough • ML failed to recognize an obstacle

Solution Hypothesis

Tune ML to infer a hierarchy of useful insights that are: **time-bounded; trade accuracy for delay**

Dynamic Neural Networks that allow for model deformation using depth and width adjustment (early exiting, skipping, pruning, etc.),
choosing the adequate protocol using Neural Architecture Search or parameter (Weights, Space, or Channel).

E.G.: Obstacle Avoidance Task



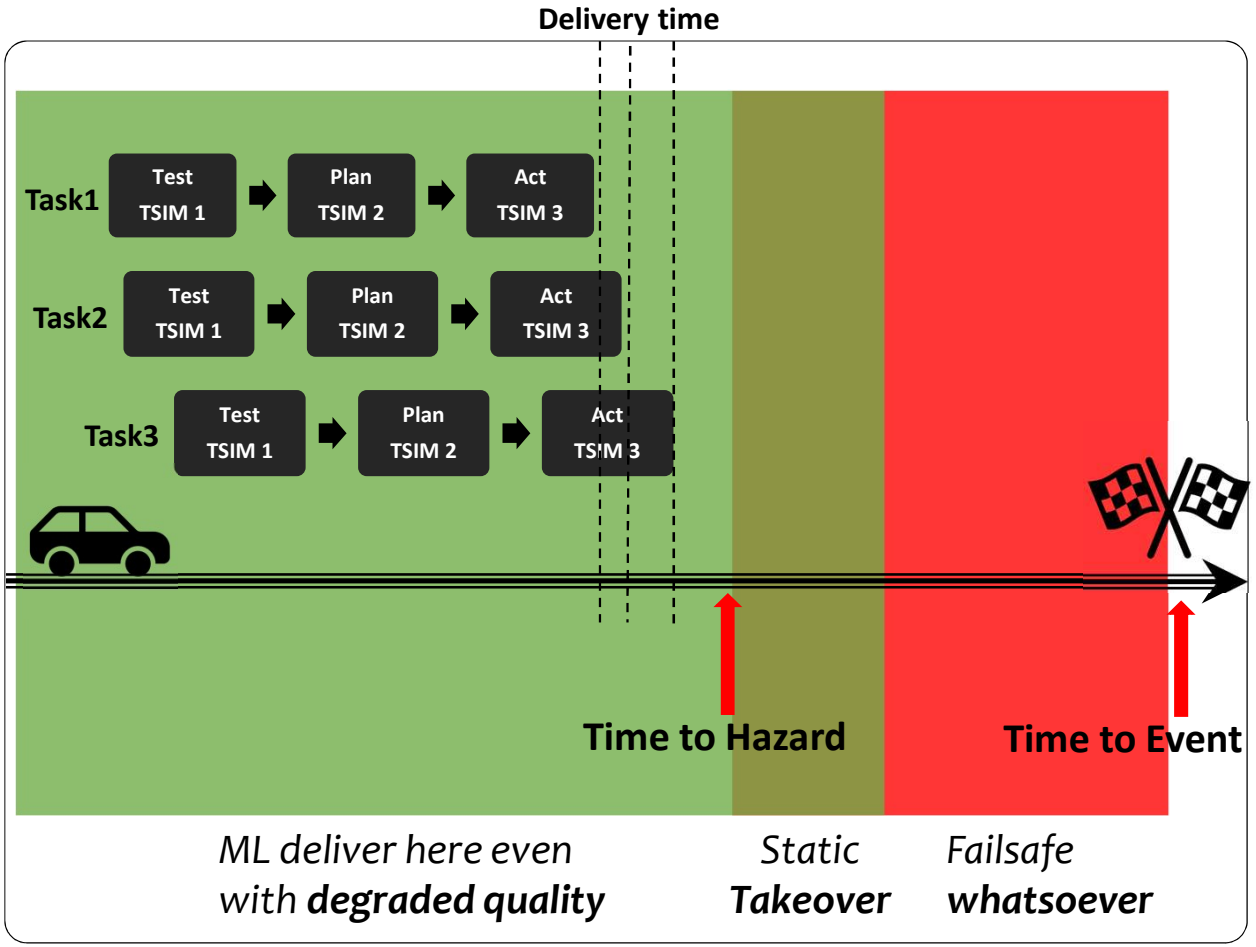
Savvy's approach

Issue
ML optimized for Better-precise-than-timely (All-or-Nothing)

Solution
calibrated ML for -Time-aware predictive quality degradation

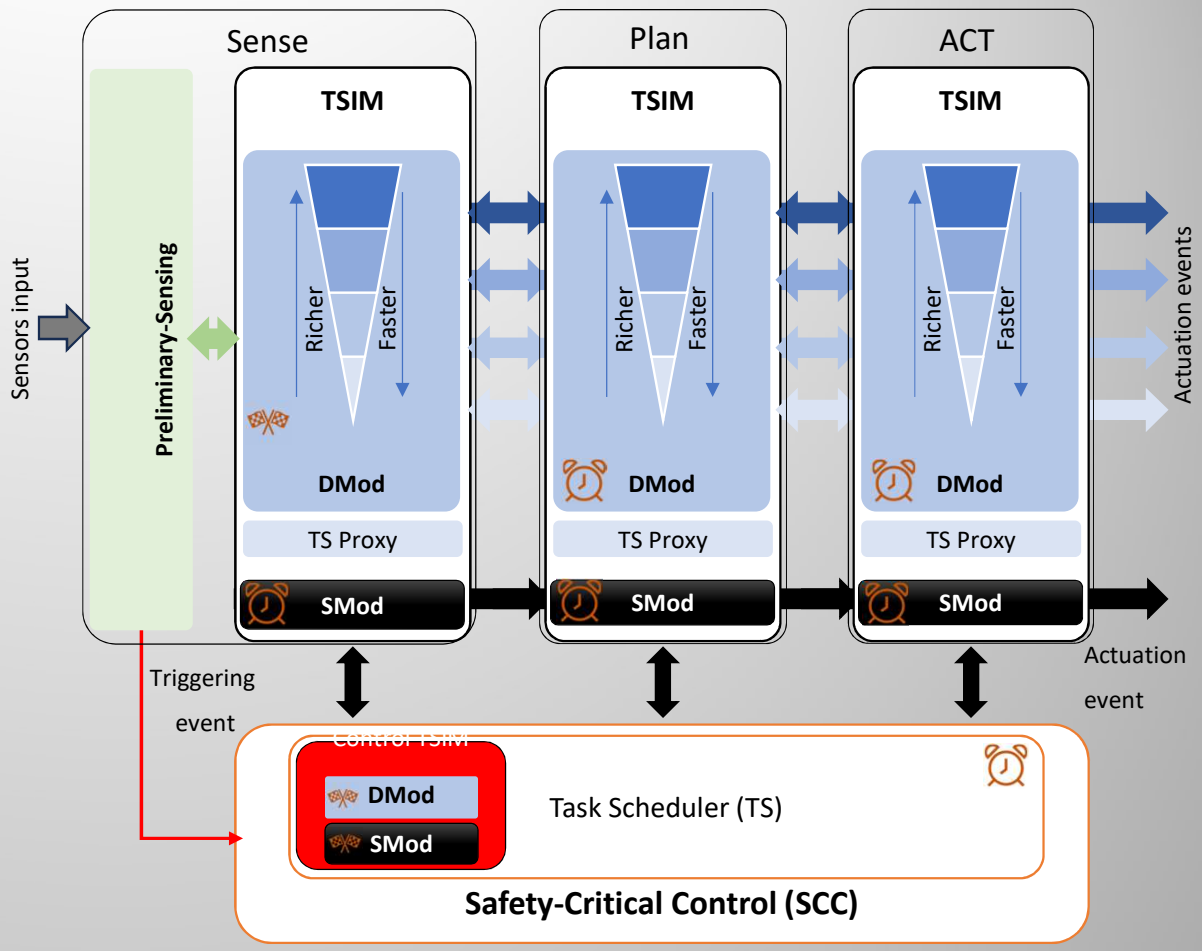
Issue
Confusion in Command and Control

Solution
Safety-critical Superv. Control System
Hybrid takes-over whatsoever



Savvy Architecture

- Preliminary Sensing
  - Detect an Event
  - Define Time-to-Event (T2E)
- Safety-Critical Control (SCC)
  - Define Time-to-Hazard (T2H)
  - Set T2E and T2H timers
  - Schedule Tasks over Time-Sensitive Intelligent Modules (TSIM)
- Timer T2H << T2E:
  - TSIM tunes ML model to deliver before T2H
- Timer T2H = T2E
  - Fail-operational: SCC takes over



Indispensable Key AV Design Principles

(whose value was confirmed by the consequences of their lack, over the talk)

- ***Fundamental principles:***
  - *Real-time context awareness and sentience of AV body and of environment;*
  - *Reconciling predictability with uncertainty of open environments*
- ***-- Modularity and hybridisation for efficient, safe and economic control architectures***
  - *Fault and intrusion tolerant control in-vehicle, eliminating SPOFs (by attacks or accidents)*
  - *Modular In-vehicle systems as Distributed systems of ECUs (leverage open ECU market)*
  - *ECUs from multicore chips as Distributed Systems-on-a-Chip (DisSoC) (for cost efficient implementation of functions above)*
  - *Trusted-trustworthy hardware-assisted Architectural Hybridisation (ultimately trusted minimal modules)*
- ***--- Algorithms and mechanisms predicates for predictability and correctness in face of uncertainty***
  - *Sensor Diversity, Correlation, Adaptive fusion, providing context awareness with acceptable coverage in open environments*
  - *Control SW Hybridisation layers of Variable complexity; these layers not converging or faulty, ultra-resilient hybrids (roots-of-trust) enforce timely and ultimately trusted take-over as last resort (not hand-over!), toward fail-op or fail-safe termination*
  - *Resilient R/T communication (V2V, V2I, V2NonSentientPlayers, e.g. humans), to enable Cooperation (no Individualistic cars!)*
  - *Protocols and architectures reconciling the stochastic nature of AI/ML with the determinism of driving control theory*

KEY TAKE-AWAYS
(or… a wake-up call to AV manufacturers):

- *Ecosystem mindset is indispensable*
- *Laws and regulations, "no Far-West"*
- *AV systems (AI/ML or other) cannot ignore distributed real-time systems and control theory, i.e. …. computer interpreted physics*
- *«If it ain't secure, it ain't safe» -- safety and security go together*
- *Reconciliation of uncertainty with predictability must be an inherent design predicate, not an after thought, a question of "training better"*
- *Modular and tech. neutral resilience solutions, from mechanical to cyber*

Search

Home My Network Jobs

Edit article View post

Credit legalbeagle.com

On the Quantitative PaperMetrics culture in research

Paulo Esteves-Veríssimo
Research Fellow at FCUL– Faculdade de Ciências da Uni. Lisboa (PT); Research Fellow at SnT– Uni. Luxembourg (LU); Former Professor and...

March 29, 2024

About a side subject discussed at the meeting (ref. a LI post made some time ago):

In my humble opinion, there are two sides to what's happening:

(1) the quantitative paper-metrics culture itself;
(2) its use to evaluate/compare researchers/ faculty across areas or subareas.

(1) is *inadequate*,
(2) is intellectually *dishonest*.

The end.

Thank you!

Paulo Esteves-Veríssimo