D 6.1 – DEFINITION AND PLANNING OF QUANTIFICATION TRIALS

Authors	Wojciech Jaworski (RTC), Michał Niezgoda (RTC), Anna Olejniczak-Serowiec (RTC), Adam Dąbrowski (RTC), Aris Lalos (ISI), Christos Didachos (ISI), Apostolos Fournaris (ISI), Petros Kapsalas (PASEU), Gerosimos Arvanitis (UPAT), Pavlos Kosmides (CTL), Gianmarco Genchi (CRF)
Work Package	WP6 – Industry Driven Trial and Evaluation

Abstract

This report contains the output of Task 6.1 which lays the ground for the pilot evaluation studies. It consists of designing and planning the way the pilot studies will be organized, supported, and managed throughout the duration of the project. The deliverable is organized around two use case groups. First one Human-Robot Interaction in manufacturing environment includes following use cases: a design operation continuum evaluation and resilience and safety. Second group of connected and autonomous L3-L4 vehicles use cases consists of: human in the loop control, cybersecurity issues and cooperative awareness. The protocol includes a description of the high-level procedure to carry out the validation of the CPSoSaware system based also on the results of other ongoing tasks.

Funded by the Horizon 2020 Framework Programme
of the European Union

Deliverable Information

Work Package	WP6 Industry Driven Trial and Evaluation
Task	T6.1 Pilot trials specification and assessment protocol
Deliverable title	Definition and planning of quantification of trials
Dissemination Level	Public
Status	Final
Version Number	2.0
Due date	30/06/2020

Project Information

Project start and duration	1.01.2020-31.12.2022
Project Coordinator	Industrial Systems Institute, ATHENA Research and Innovation Center 26504, Rio-Patras, Greece
Partners	1. ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS (ISI) - Coordinator 2. FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA (I2CAT), 3. IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD (IBM ISRAEL 4. ATOS SPAIN SA (ATOS), 5. PANASONIC AUTOMOTIVE SYSTEMS EUROPE GMBH (PASEU) 6. EIGHT BELLS LTD (8BELLS) 7. UNIVERSITA DELLA SVIZZERA ITALIANA (USI), 8. TAMPEREEN KORKEAKOULUSAATIO SR (TAU) 9. UNIVERSITY OF PELOPONNESE (UoP) 10. CATALINK LIMITED (CATALINK) 11. ROBOTEC.AI SPOLKA Z OGRANICZONA ODPOWIEDZIALNOSCIA (RTC) 12. CENTRO RICERCHE FIAT SCPA (CRF) 13. PANEPISTIMIO PATRON (UPAT)
Website	www.cpsosaware.eu

Control Sheet

VERSION	DATE	SUMMARY OF CHANGES	AUTHOR
0.1	1/04/2020	Table of Content distributed to the Consortium	*Wojciech Jaworski* *Anna Olejniczak-Serowiec* *Michał Niezgoda* *Adam Dąbrowski*
1.0	17/06/2020	Final version for internal review	*Wojciech Jaworski* *Anna Olejniczak-Serowiec* *Michał Niezgoda* *Adam Dąbrowski* *Aris Lalos* *Christos Didachos,* *Apostolos Fournaris* *Petros Kapsalas* *Gerosimos Arvanitis* *Pavlos Kosmides* *Gianmarco Genchi*
1.1	24/06/2020	Reviewed version	*Georgios Keramidas* *Rodrigo Diaz*
2.0	30/06/2020	Final version after review	*Michał Niezgoda* *Wojciech Jaworski* *Anna Olejniczak-Serowiec*

	NAME
Prepared by	RTC
Reviewed by	UoP, ATOS
Authorised by	ISI

DATE	RECIPIENT
1/04/2020	Project Consortium
30/06/2020	European Commission

Table of contents

List of figures

List of tables

1 Introduction

1.1 Scope and objectives of industry driven trials

The goal of industry driven trials and evaluation is to validate and demonstrate the whole set of functionalities of the CPSoSaware framework. In order to enable early verification of basic features, the trials will consist of two iterations (preliminary and final trials), both including two key use case groups: connected car and human-robot collaboration.

The first use case is focused on connected semi-autonomous vehicles where we will perform trials focused on Human in the loop scenarios, like non predictable failures that may involve the human driver and how this affect the design operation continuum support of the CPSoSaware solution as well as human situational awareness enhancement when using the CPSoSaware architecture. We also use this use-case to access the cybersecurity mitigation strategies using the CPSoSaware architecture and its response to cyberattacks.

The second use case will be focused on HRC (Human-Robot Collaboration) in the manufacturing environment and will involve trails that challenge the MOOD CPSoSaware concept and trails on accidents/failures as well as cybersecurity attacks that challenge the collaborative control mechanism and the autonomic decentralized operation of the CPSoSaware solution as well as the design operation continuum support in the presence of cybersecurity attacks. The two use cases complement one another since they have different requirements and specificities (open spaces and moving CPS, close interconnection with humans versus closed space environment, static CPS and more relaxed interaction with humans).

1.2 Role of the assessment protocol in CPSoSaware project

The assessment protocol will include a detailed description of the procedure to carry out the validation of the CPSoSaware system. Protocol will focus on the following parts of validation process:

- inclusion and exclusion criteria,
- experimental conditions,
- data collection instruments,
- outcome measures.

All components of CPSoSaware framework must be assessed precisely. To ensure appropriate validation, the following parameters need to be considered:

- limits of detection/quantification,
- working range,
- precision (repeatability, intra, and inter-laboratory reproducibility).

1.3 Evaluated CPSoSaware pillars

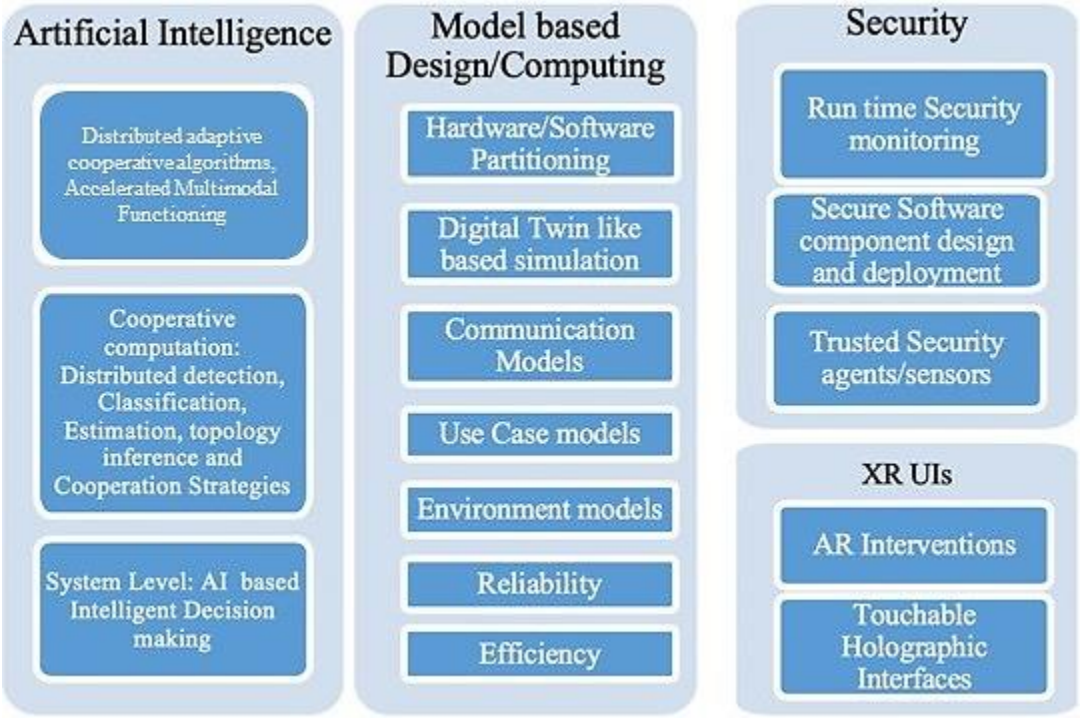

Figure 1. CPSoSaware pillars.

The CPSoSaware objectives are matched by the four CPSoSaware pillars that are presented in Figure 1:

- Cognitive AI pillar,
- Model based design and computing pillar,
- Security pillar,
- Extended reality (XR) pillar.

These pillars are manifested in the CPSoS design phase and continue to appear in the CPSoS operation, commissioning, and decommissioning phase thus, supporting the full CPSoS lifecycle. Each of the pillars is briefly described in the following section.

1.3.1 Artificial Intelligence (distribute, adaptive, cooperative algorithms, accelerated multimodal fusion)

This pillar considers multiple (heterogeneous) CP(H)S's that cooperate in multiple tasks (multi-device multi-tasking paradigm - MDMT) and provides scalable and distributed signal processing and learning (SPL) algorithms for the processing, and in-network fusion of heterogeneous data (e.g., GPS, Camera, Lidar as well as driver/user status related data) in order to tackle challenges related to: i) User State Monitoring, ii) Multimodal localization and scene analysis as well as iii) identification of abnormalities attributed to cyber failures or cyber-attacks. This mechanism also provides a set of incentives that induces cooperation between CPSs, even when all of them are assumed to act selfish in their own interest (i.e., by

relying on the property that CPSs return favours for each other). The aim is to let these CPSs cooperate in order to achieve a better overall performance and to support the design-operation continuum throughout the CPSoS lifecycle as compared to the case where the CPSs would operate on their own or where they would exchange raw data in an uncontrolled fashion. This bottom-up approach is highly novel, and forms the basis of several challenges that this pillar tackles, such as:

- Heterogeneity: each CPS has its own mode of operation, which results in, among others, (i) different SPL tasks within the network (see also next bullet), (ii) different CPS characteristics, and (iii) different sampling rates/ data resolution.
- CPS-specific interests: each CPS has its own task or 'interest', which possibly conflicts with other CPSs, e.g., a source may be desired for one CPS, but at the same time an interferer for another CPS. This makes the choice of the fusion rules at the different CPSs a highly non-trivial task and very different from the case where the CPS share the same interest.
- Avoiding selfish behaviour: CPS are expected to cooperate by exchanging useful data with each other. In the MDMT context, nodes are selfish; hence they need to have an incentive to cooperate. This selfish behaviour is mostly absent in traditional (top-down) CPSoS, where the MDMT paradigm holds.
- Distributed network/system topology inference/selection: CPSs have to be able to infer information about their place and role within the overall network/system topology. Since the topology is a network/system-wide characteristic, it will be non-trivial to infer such information in an adaptive distributed fashion.

This pillar provides a highly intelligent, yet generic, execution environment tailor-made for multiple CPSs where a CPS utility is quantified by all the other CPSs in the network. In addition, this utility also takes the network topology into account. The supported SPL functionalities are able to operate also in ad hoc networks and they can cope with the rapidly changing statistics of recorded signals/data and changes in the network topology. More importantly, they are able to infer information about the network topology and identify opportunities with the aim of improving a system wide performance. This will be achieved by defining a proper utility measure that can be computed in a distributed fashion, and that quantifies the 'importance' or 'usefulness' of each CPS with respect to the network/system-wide performance. This measure should include the usefulness and quality of the data that a CPS delivers, and at the same time the CPS position in the network/system. To achieve this goal, we will rely on a rating system where CPSs give a score to each of their neighbours. From these local utility scores, a network-wide utility score is evaluated, relying on the following principle: if CPS B assigns a high local score to its neighbour CPS A (B->A), and if CPS C assigns a high local score to its neighbour CPS B (C->B), then CPS C implicitly believes that CPS A should have a large score (C->(B)->A). This is akin to the spectral graph theory (SGT) concept of eigenvector centrality (EC). Note that EC is a centrality measure, hence it favours CPSs with central positions. Therefore, the major tasks executed by this submodule for achieving autonomic operation, higher reliability, and redeployment-commissioning/decommissioning of CPSs, are the following:

- Assignment of a local utility score to their neighbours (e.g., based on received data quality or their willingness to cooperate in a coalition).
- Execution of distributed SPL and AI approaches, inspired by eigenvector centrality, to compute a network-wide utility score for each CPS, based on the local utility scores.
- Identification of selfish CPSs which focus on manipulating the rating system (e.g. to get a higher individual rating), and deployment of proper measures to disallow this.

1.3.2 Model based design/computing (OpenCL optimization, simulation, hardware/software partitioning, CPS models, use case models, reliability, efficiency)

The Model Based Design of the CPSoSaware project is responsible for the modelling of the CPSoS functionality using high level modelling languages and open access tools in order to describe discrete, continuous and hybrid time models of CPS and cross CPSs processes as well as orchestration procedures of the CPSoS. The modelling process will also include a series of Requirement-KPIs in order to create a meta-model. Most importantly, our modelling methodology will include models that will capture the user behaviour and security functionality. The models will be used in order to produce an initial detailed description of the functional processing components of each CPS (including processes that can be executed concurrently or in specific order following a producer-consumer style). The output of our modelling layer will be quantified and be used in order to choose components from CPSoSaware software and hardware component libraries and create the actual runnables (either as software or hardware components) to be deployed in the CPSs of the system. The application description using the model-based design will be done using the OpenCL standard. The provided descriptions will then be optimized based on the CPSoSaware meta-model schemas that are produced using the requirement KPIs and their associated weighted values. Using the schemas, the OpenCL code for each CPS will be optimized in order to match the CPSoS KPIs.

The optimized OpenCL will then be inserted to code extraction tools (either for software or hardware) that in order to address the heterogeneity of each CPS use appropriate Hardware/Software IP code libraries (OpenCL kernels). The software executables can be optimally executed on the CPS single and multiple cores, GPUs or FPGA logic but also automatically generate hardware FPGA deployable bitstreams. Generated bitstreams can be directly mapped on the CPS/CPHS System-on-Chip (SoC), as long as the CPS/CPHS supports FPGA technology, thus providing the ability to change the CPS SoC hardware structure itself.

Our aim is to offer a practical approach that will take as a single source file (based on OpenCL) and perform automatically all the following steps:

- Hardware-software partitioning,
- Enhancing the software components with reliability and security means using an orchestrated source-to-source translation approach,
- Enhancing the hardware components with reliability and security aware structures by relying on HLS (High-level Synthesis techniques).

In this pillar, the CPSoSaware project will also deal with reliability and resilience issue, aiming to "harden" the HW/SW component libraries with fault tolerance and reliability techniques like in tandem execution of Software operations with multiple cores executing the same OpenCL kernel simultaneously by controlling the level of redundancy at the thread level (a voting mechanism will ensure the correct operation) and the degree of instructions re-executed (replayed) in a lock-step fashion at the instruction level, and the frequency of memory scrubbing operation. Also, in hardware components, controlling of ECC bits in memories and size of redundant/spare resources at the processor and memory/cache levels. Our reliability goals are to guaranty the real-time operation of the system in the presence of faults.

Finally, the Model based design pillar encapsulates a simulation environment that provides the integration and orchestration framework for various, different simulation subcomponents that simulate specific areas of the CPSoS architecture like the CPS/CPHS Hardware and Software processes, the CPS/CPHS

to CPS/CPHS and CPS/CPHS to System network environment and the CPS to Human interaction. The CPSoSaware simulation and orchestration will be able to extract simulation data from all the subcomponents and will provide a data control and collection environment for CPSoS specific simulators (e.g. use-case 1: Connected semi-autonomous car simulator provided by Panasonic Automotive/I2CAT and use case 2: Human-Robot collaboration in manufacturing environment simulator provided by CRF). Concepts like hardware-in-the-loop simulation and analysis will be explored and the simulation outcomes will be used in order to finetune the model-based design approach employed in the project.

1.3.3 Security (Run time security monitoring, secure hardware/software component design and deployment, trusted security agents/sensors)

The CPSoSaware pillar on Cybersecurity and Security by-Design/Trust is aimed at operating mostly proactively to detecting anomalies and cyber threats before they become an actual failure. Thus, in CPSoSaware solution we aim at a full-scale protection, horizontally and vertically, that involves detection, identification, response, and mitigation (before, during and after a cyberattack).

The CPSoSaware pillar spans on both, the system layer and the CPS layer of the CPSoSaware architecture and includes components that are designed to be secure following the security-by-design approach and are also collecting input to assess cybersecurity anomalies at run time. More specifically, the CPSs are security hardened to achieve a high security level and be infused with trust in their computation flow. Following this security-by-design approach we introduce in each CPS, appropriate security hardened agents/sensor that operate at runtime, collect security related events and sends them to the CPSoSaware Security run time Monitoring Mechanism (SRMM) for analysis in order to detect and respond to anomalies that maybe related to cyberthreats.

Security 'hardening" starts from modelling of appropriate security functions (security/cryptography primitives e.g., Secure communication, message integrity, authenticity) and then mapping their functionality at specific hardware and software components within each CPS (following the optimization approach using the CPSoSaware modelling pillar). In this process, we apply secure design principles in the HW/SW security primitive components but also on any HW/SW component of the CPSoSaware modelling libraries that handles sensitive information (this knowledge is gained from the appropriate modelling of CPS functionality).

Furthermore, the concept of security-by-design is been infused throughout the CPoSaware modelling approach so that the optimization process considers specific design rules that are known to introduce vulnerabilities (e.g., Code patterns that introduce buffer overflows or Return Oriented Programming attacks). Among the security requirements that will be included at design time there will be side channel resistance of the security HW/SW components in the CPSoSaware Modelling libraries. Design methodologies and best practice for achieving robustness from physical attacks will be put in place and the possibility of automatically apply them will be considered and integrated into the design flow when possible.

The above designed components are used in the CPS layer and are meant to support the information collection for the CPSoSaware Security Runtime Monitoring mechanism (SRMM) which constitutes the main security sentinel of the CPSoS at operation phase. SRMM spans vertically in the CPSoS architecture and includes the installed security monitoring agents in the CPS layer that log events happening during each CPS operation. The agents will potentially be able to perform lightweight data analytics in order to process events and extract/generate alerts when some possible abnormal behaviour

is observed locally. The CPSoSaware SRMM tool that operates at the CPSoS system level will be able collect alerts from all the CPSs' agents and access the status of the CPSoS in terms of security (possible abnormal behaviours that indicate cyberthreats or actual cyber-attacks). It categorizes security issues according to their criticality level and may forwards such information to the CPSoSaware system decision support system (CSAIE) that may trigger a redesign process and provide suggestions on the recommended components that need to be included in the new design and possibly changes the requirement KPIs.

1.3.4 XR UIs (AR interventions)

The major objective of the extended reality technologies is to provide highly innovative and interactive tools empowering users, including operators of the CPSoS system and end users interacting with the individual connected CPHS's, to create easily accessible, sustainable, and interactive interfaces by:

- tracking human physical responses to different emergency situations and dynamically adapting the visual, auditory, and tactile information rendered in the real world, ensuring that will not exceed the user ability to handle it;
- increasing significantly quality of experience via touchable holographic interfaces, and augmented gesture control with natural tactile feedback in the mid-air using an array of ultrasonic emitters.

Haptic perception is an important modality in reinforcing the presence and interactivity of virtual or remote targets, however providing haptic feedback in dynamic environments remains a very challenging task. In recent works, the issue of haptic tele-presence and tele-control systems has been addressed by improving and integrating real-time haptic rendering of unstructured spatial data and collaborative interaction with a remote end-user. Contact with physical objects are directly estimated from streaming point-cloud data (without surface reconstruction), and haptic guidance cues that restrict or promote the users' motion are provided by both predefined triggers and gesture-based input from a helper. The haptic device delivering these cues range from wearables to grounded force-feedback controllers. Very recently, a new breed of haptic devices utilizing focused ultrasound can deliver tactile cues to users' hands from a proximal distance and without the need for any additional wearables of or hand-held controllers therefore increasing the overall usability, applicability and accessibility of these touchless haptic controllers. In a critical setting (e.g., automotive case) this enabling technology has shown impressive potential for improving user (e.g., driver) safety by reducing visual distraction. In CPSoSaware we will go way beyond this feasibility study and use case by adding high-resolution, high-fidelity, and highly informative haptics to virtual objects and functions, that complement touchable holographic interfaces, and augment gesture control with natural vibrotactile feedback to increase the feeling of the immersive experience, through the use of mid-air ultrasonic feedback provided by UH development kits.

In both cases (Connected Cars and HRC Manufacturing), using different augmented reality tools (e.g. AR Glasses, mobile devices, and marker less tracking), the user will receive

- information streams regarding the task under way or a machine involved, improving focus on some crucial information/elements,
- personalized reminders regarding other parallel or scheduled tasks significantly improving response time,
- notifications and visual aids highlighting imminent dangers or accident related factors.

Multi-modal projections, that visually describe the significant environment parameters will act as guidelines through situations with high task load in challenging and multi-tasking environments. These awareness enhancement tools will go beyond current state of the art approaches introducing a real-time personalized AR -based visual aid for CPHS users and/or CPSoS operators.

2 Pilot quantification phase

2.1 Concept of pilot quantification phase

Pilot quantification phase is the first iteration of Industry Driven Trials and Evaluations being the main subject of Work Package 6. It aims to validate and demonstrate the whole set of functionalities of the CPSoSaware framework in 2 inspection cases: connected car and human-robot interaction. Despite the differences in analysed scenarios, some similarities can be found in terms of collected data and metric used for systems assessment. Important aspects of CPSoSaware valid for both use-cases are:

- Utilization of sensor data for machine perception,
- Secure communication of Systems in System of Systems,
- Simulation supported validation of use cases.

Key ideas leading to proper CPSoS components performance involve human-oriented design and safe human-robot interaction (HRI), which regardless of the context (automotive or manufacturing) is the most important factor in building trust towards automatized systems. Figure 2 provides the overview of methods for depicting safety in HRI.

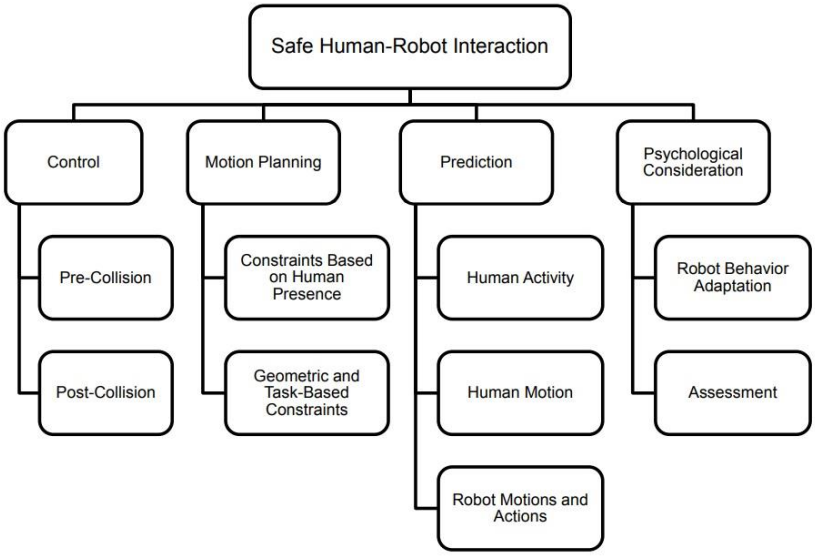

Figure 2. Diagram depicting the major methods of providing safety in HRI [source: Lasota, Fong, & Shah, 2014]

2.2 Human-Robot Interaction in Manufacturing Environment

CRF will provide a work-cell environment (Figure 3) where there is an assembly line with robotic entities that are supported by human operators. Initially, in the quantification phase, a small number of

human operators will be recruited and participate in small HRC activities inside the test site work-cell in a small time period (few months). There will be a division of the experimental group in subgroups in order to provide advice on specific aspects of the HRC operations performed on-site. A collection of experimental data will be acquired over time, both from the human behaviors and health status in the presence of robots as well as the human to robot on-site interactions. The goal of this activity would be to collect enough data in order to structure a HRC operational model with enough information for the quantification of the CPSoSaware user models and risk prediction framework.

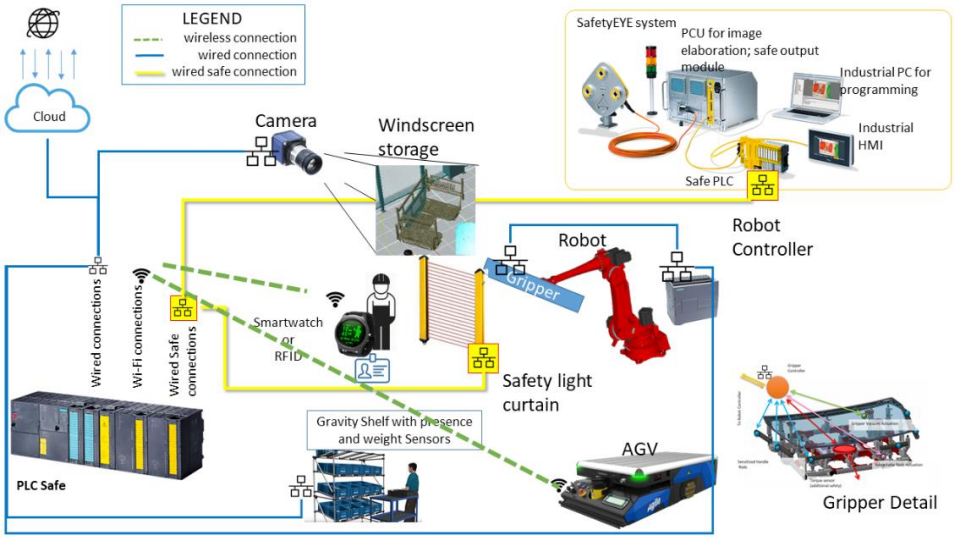

Figure 3. CRF work-cell environment.

2.2.1 A design operation continuum evaluation

2.2.1.1 Detailed use case concept

In this use case, we consider the inclusion of a new component in the assembly line of an automotive manufacturing factory in which a specific vehicle model is developed through a HRC approach. Typically, a specific component is assembled on the vehicle in a collaborative workplace using collaborative robots. For example, the rear-view mirror is assembled on a windshield which is held and manipulated by a high payload collaborative robot. In the normal operation cycle the operator checks from the HMI the specific model of the rear-view mirror to be assembled on that specific model, he picks it up from a side line logistic kit which is prepared in advance by an automated warehouse management specifically by a movable robot equipped with a flexible gripper and a 3D camera for the pick an place of the parts. After picking the rear-view mirror from the logistics he performs the assembly according to known assembly rules which are provided to the operator (Figure 4). The operation is performed daily a specific number of times.

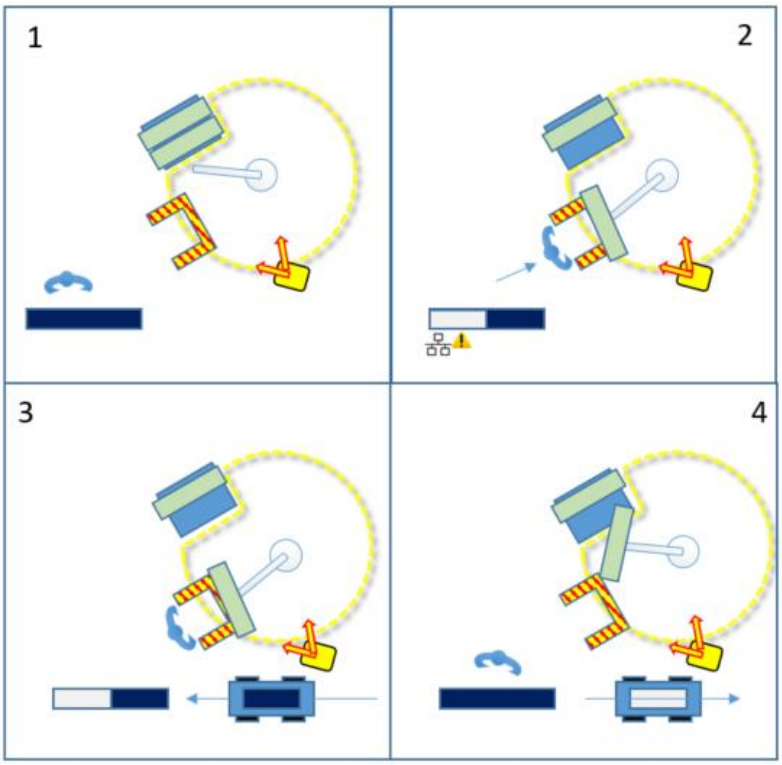

Figure 4. Normal operation cycle.

The assembly line work cell is locally managed by an intelligent software which dialogs with the MES (Manufacturing Execution Software) to update the number of parts produced, notify serial numbers of parts whose traceability is requested and, of course, manages and controls the equipment in the cell. The human operator is defined inside the software as a dynamic model in order to have an updated model of the two actors in the Human Robot Collaboration. The overall cell is a system defined inside the line "system" which is nested into the more complex factory system. The parts' drawings are defined both as CAD models and as part numbers in an ERP (Enterprise resource planning) warehouse management which is connected to other complex systems such as the Central Purchasing, after market and so on.

We consider also that the operator's HMI includes augmented reality component that are used to provide him with feedback on the assembly process. Finally, HRC work-cell will collaborate with a layer of simulated ERP type of systems (e.g. logistics, purchasing etc). The operators will be equipped with wearable devices that will allow us to monitor their health and send them instructions. Wearable devices with AR have started to become ubiquitous in the industry and new applications have been created to satisfy the challenging needs of this area. These devices are used directly into workers' workflows and they can superimpose virtual 3D objects within the physical environment. Their use perfectly fits into a CPSoS environment, providing mobile-friendly solutions, touch input methods, location services, and more. Users can include this new technology into several different types of existing solutions, functionalities, and workflows. Workers can always have access to appropriate data and in any location, they could be. More specifically, wearable devices provide the ability to deliver data to workers while leaving their hands free to work on current tasks. Workers can be assisted by experts remotely and can participate in AR-based training programs. It allows an experienced or expert worker to mentor another worker who is in the field.

This kind of remote collaboration between in-house experts and field workers can be assisted by solutions based on shared extended reality. AR devices are also able to provide on-time information to the workers, through the wearable device, including diagrams, text, images, checklists, manuals, videos, and 3D virtual objects. These info deliveries are directly in the worker's line of sight. Extended reality glasses and headsets are also capable of interacting with devices within the Internet of Things, allowing for collaboration with other linked devices too. This functionality also involves monitoring in real-time and constantly providing relevant alerts, warnings, and updates regarding the significance of the situation

2.2.1.2 Validation procedure

CRF will perform appropriate surveys, with operators in manufacturing plants. Needs and requirements will be prioritised and, on this basis, the target use cases of the CPSoSaware system will be elaborated, under a prioritization perspective aligned to the one of the requirements. The surveys will be conducted with users in Italy (CRF). Users will be engaged in interviews, based on well-defined criteria and upon written informed consent.

2.2.1.3 Data acquisition

Data for this use case evaluation will be collected during actual use of the cell with specific functional tests performed by multiple operators in the cell equipped with the different sensors. The data that need to be collected are following:

- Data from the cell's safety components:
  - SafetyEYE system: Safety Zone Violation (TCP/IP strings)
  - Safety light curtain: Safety Zone Violation (TCP/IP strings)
- Smartwatch
  - ECG
- PLC
  - Cell status
- Other cell components
  - Camera
  - Gravity Shelf
  - Robot Controller

In addition to collected data, very important part of this scenario validation will be performed with questionnaires. A widely used scale for assessing subjectively experienced workload can be used:

- NASA-TLX (Hart & Staveland, 1988) - subjective, multidimensional assessment tool developed by the Human Performance Group at NASA that rates perceived workload, that can be used in order to assess the HRI quality. The questionnaire will be administered to the participants twice: before and after the trial, to compare in repeated measures design, how the CPSoS components influenced the perceived workload level.

In addition, a dedicated questionnaire will be developed to assess subjective perception design operation continuum. The questionnaire will be administered to CRF workers participating in the trial. The development of such questionnaire requires several steps (see Figure 5):

- Phase I – collecting the functionalities to be performed/improved by the CPSoS component from its authors.
- Phase II – composition of a questionnaire in which each of the functionality is addressed with one item. The respondent is asked to relate to each item on a 5-point Likert scale.
- Phase III – pilot study on a sample of 15 CRF workers – potential trial participants – where they are asked to comment whether the items are easily understood, and how do they understand each of them.
- Phase IV – questionnaire improvements targeted at unambiguous understanding.

Figure 5. Overview if the design operation continuum related questionnaire development process.

2.2.1.4 Outcome measures

The goal of this use case evaluation is to measure two aspects of manufacturing process: productivity and worker satisfaction. It is crucial for human-robot cooperation to provide benefit in both aspects, so meaningful metrics have to be defined and analyzed all together. Metrics that are going to be used to precisely assess advantages of using CPSoSaware platform in manufacturing environment are following:

- Productivity related metrics:
  - Time Cycle – the duration of a single cycle (expressed in seconds). The expected cycle time with robots engaged is shorter than with automation disengaged.
  - Effectiveness – metric being combination of performance, quality, and availability, describing overall effectiveness of production process.
- Worker satisfaction related metrics:
  - Quality – worker satisfaction in HRI, thrust, and stress levels as subjectively experienced and expressed in questionnaires.

2.2.2 Resilience and safety

2.2.2.1 Detailed use case concept

In an industrial manufacturing environment, cybersecurity, safety, and robustness/resilience are very important. The human worker's health and the robotic equipment must be protected against unexpected errors/failures and malicious attacks that will cause an industrial accident. In this use case scenario, we focus on the aspect of cybersecurity, resilience and safety in the work-cell even when some unexpected failure event occurs (Figure 6).

The use case involves the use of all CPSs that are included in the provided work-cell i.e., the surveillance cameras, the robot and the operator HMI device. The protection of the manufacturing environment and monitoring of correct system functionality against cyberattacks is a critical element of this scenario. The response of the system must focus on a) protecting the environment and b) providing

information of the attack or error in order to refine the process and model of CPSoSaware of the system. The information will be used in two different ways: on the one hand to inform cybersecurity experts of the system about the alarm or event and on the other hand to provide information about how the system behaved according to the fulfilment of the cybersecurity requirements identified at design time.

The wearable devices (smartwatch) can be connected to each other facilitating the communication between the workers. The peer-to-peer communication can provide a convenient real-time solution for any case of urgency. The auxiliary feature is brought with AR thus it is covered by the continuously expanding range of smartphones that support AR technologies.

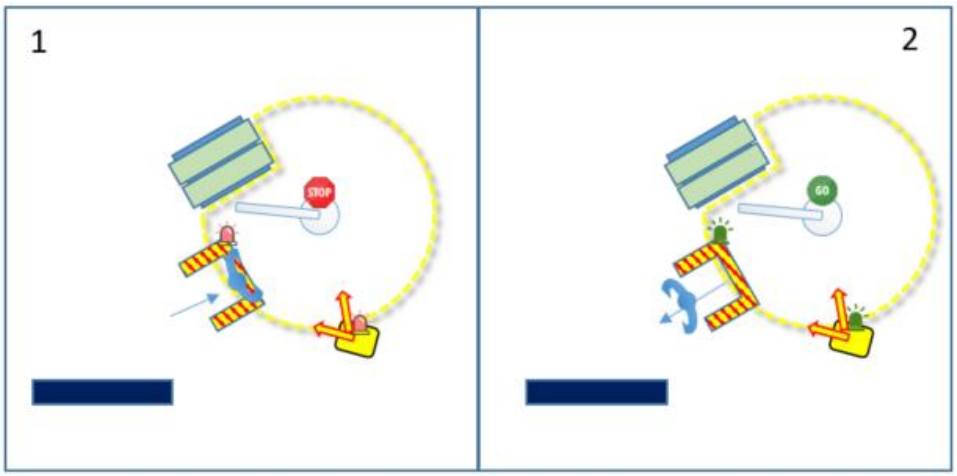

Figure 6. Work-cell unexpected failure.

2.2.2.2 *Validation procedure*

CRF will perform appropriate surveys, with operators in manufacturing plants. Needs and requirements will be prioritised and, on this basis, the target use cases of the CPSoSaware system will be elaborated, under a prioritization perspective aligned to the one of the requirements. The surveys will be conducted with users in Italy (CRF). Users will be engaged in interviews, based on well-defined criteria and upon written informed consent.

Furthermore, a special focus will be given in validating CPSoSaware system regarding cybersecurity attacks as described below:

- *Providing means to intervene to the Human operator*: in order to enhance the situational awareness of the Human operator, CPSoSaware system will provide ways for the Human operator to intervene using his HMI device when needed. In order to validate this, Human operators will have to answer to specific questionnaires that will be distributed.
- *Identify emergency and communicate problem using Vision ML algorithms on the work-cell camera*: The collaborative Vision ML algorithms on the work-cell camera should be able to identify the emergency and communicate the problem to the rest of the CPSoSaware system. In order to validate this, images with emergency events will be injected to the surveillance systems, while the CPSoSaware system will be validated with an accuracy percentage on the successfully recognized events. The accuracy percentage will be also enriched with results on precision and recall measures.

- *Reduce or avoid possible emerging failures in the assembly line*: In order to mitigate the risk and avoid or reduce the possible emerging failures in the assembly line (due to an emergency or original failure), the robot CPS collaborative ML algorithms will have to collect feedback and address the issue by changing functionality (e.g. rearrange process tasks). In order to validate this, two methods will be followed: (i) inject emergency events and calculate accuracy percentage (including accuracy and precision); (ii) retrieve feedback from questionnaires answered from Human operators.
- *Reaction and protection against malicious attacks*: CPSoSaware is providing a security-by-design approach including cybersecurity protection and reaction to the system. In order to validate this, two methods will be followed: (i) inject malicious attacks to the system and detect the number and the time of detecting the attacks; (ii) retrieve feedback from questionnaires answered from Human operators who were using the system during the attacks.

2.2.2.3 Data acquisition

Data for this use case evaluation will be collected during actual use of the cell with specific functional tests performed by multiple operators in the cell equipped with the different sensors. The data that need to be collected are following:

- Data from the cell's safety components:
  - SafetyEYE system: Safety Zone Violation (TCP/IP strings)
  - Safety light curtain: Safety Zone Violation (TCP/IP strings)
- Smartwatch – Operator State Monitoring System data
  - ECG
- PLC
  - Cell status
- Other cell components
  - Camera
  - Gravity Shelf
  - Robot Controller

In addition to collected data, very important part of this scenario validation will be performed with questionnaires. Several validated questionnaires developed in prior work have proven useful for the assessment of psychological safety in HRI:

- *Godspeed* questionnaire (Bartneck et al., 2009Y) – standardized and validated questionnaire allowing for assessment of five pillars of psychological safety in HRI: perceived safety, anthropomorphism, animacy, likeability, and perceived intelligence.
- *Negative Attitude toward Robots Scale* (NARS; Nomura et al., 2006) - method to quantify humans' satisfaction during HRI. Psychological scale measuring negative attitudes toward robots with the use of three sub-scales.
- *BEHAVE-II* (Joosse et al., 2013) – method to assess attitudinal and behavioural human responses to robot behaviour. The attitudes measured include trust towards robots.

Subjective workload during the interaction with robot can be assessed with the use of a widely used tool:

- *NASA-TLX* (Hart & Staveland, 1988) - subjective, multidimensional assessment tool developed by the Human Performance Group at NASA that rates perceived workload, that can be used in order to assess the HRI quality.

Additionally, a dedicated questionnaire can be developed to assess trust and stress experienced in the context of HRI in manufacturing. The development of the questionnaire consists of several steps (see Figure 7):

- Phase I - interviews/surveys among a group of CRF workers about the most common sources distrust in robots and stress during interaction.
- Phase II – expert-based classification of the issues into distrust and stress factors categories.
- Phase III –pilot study on a sample of workers, based on which reliability analyses will be performed to choose the set of most powerful items to be included in the final version of the questionnaire.

Figure 7. Overview of the HRI in manufacturing trust and stress questionnaire development process.

The questionnaire will be administered to the CRF employees participating in the trial, they will be asked to relate to each of the items on 5-point Likert scale.

2.2.2.4 Outcome measures

The goal of this use case evaluation is to measure the satisfaction level of the human operator who will be using the CPSoSaware system. The satisfaction level will be extracted from questionnaires that the human operators will have to answer during the pilots. Furthermore, there will be some technical evaluation results which will include the following:

- *Providing means to intervene to the Human operator*: by the distribution of questionnaires the Human operator will answer on how many times he successfully managed to intervene when needed using the CPSoSaware HMI device.
- *Identify emergency and communicate problem using Vision ML algorithms on the work-cell camera*: the validation results in this case will include an accuracy percentage as well as precision and recall measures after the injection of images with emergency events to the surveillance systems.
- *Reduce or avoid possible emerging failures in the assembly line*: the validation results in this case will include:
    - The satisfaction level of the human operator who is facing emergency events (information gathered from questionnaires),
    - An accuracy percentage as well as precision and recall measures after the injection of emergency events.
- *Reaction and protection against malicious attacks*: the validation results in this case will include:
    - The satisfaction level of the human operator who is facing malicious attacks,
    - A percentage of the successfully detected malicious attacks,
    - The time of detecting malicious attacks.

2.3 Connected and Autonomous L3-L4 Vehicles

2.3.1 Human in the loop control use case in single vehicle scenario

2.3.1.1 Detailed use case concept

The goal of this human in the loop control scenario is to facilitate the cooperation between the vehicle systems and the human driver (e.g., during the request to intervene) and improve subjective attitudes of the drivers towards the instrumented L3-L4 vehicles.

The ongoing process of transport autonomization, consists of several automation stages, on which the transition processes from automated to manual control (level 0) and vice versa might take place. Transitions of control can be of dual origin: driver-initiated or system-initiated (Figure 8) presents the overview of the possible origins of transition depending on the type of driving engaged before the transition starts.

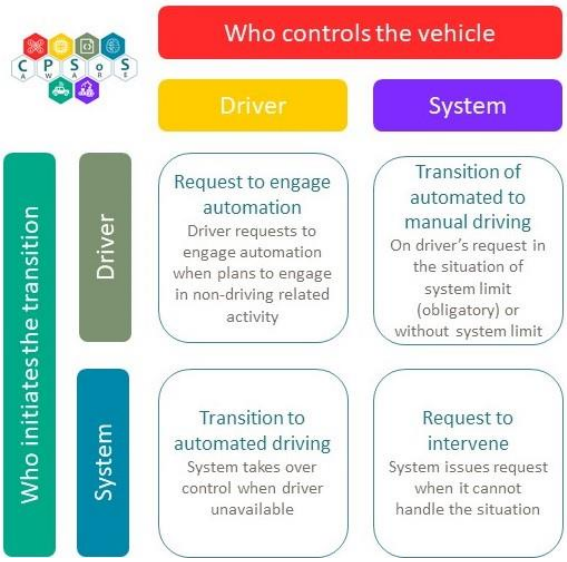

Figure 8. Manual-to-automated and automated-to-manual driving transitions depending on the driving mode engaged at the starting moment and the origin of transition [source: ISO/TR 21959-1:2020].

CPSoS will consider scenarios in which Autonomous Driving Systems (ADS), that are able to work unattended only under mild conditions, while they require a human driver to take control in situations that cannot be handled in an automatic way by issuing the so-called Request to Intervene (i.e., SAE level 3 vehicles), meaning the situation when system controls the vehicle and the system issues request to change driving state. Such systems require the cooperation of a Driver State Monitoring System (DSM), that assesses the state of the human driver (e.g., by performing pose-estimation, or emotion recognition, possibly utilizing multiple modalities) and a sub-system that performs an analysis of the scene outside the vehicle and controls the vehicle to move autonomously on a predefined path (e.g., recognizing vehicles ahead, estimating their velocity/trajectory, forecasting future vehicle locations).

Proper cooperation between human driver and the automated driving system requires constant monitoring of the driver state and assessing driver availability. To evaluate the scenario, it is important to

provide the framework for safety-critical states that can lead to road accidents. To this end, CPSoSaware will focus on the following driver states: distraction, drowsiness, and emotional state. Each of the targeted driver states might be confusing since they are influenced by a variety of factors. Moreover, variety of terms is used by manufacturers and researchers while referring to them. The proposed framework considers the fact that driver inattention has no universal definition, but it can be classified into two basic and distinctive categories – misdirected attention and fatigue (Figure 9).

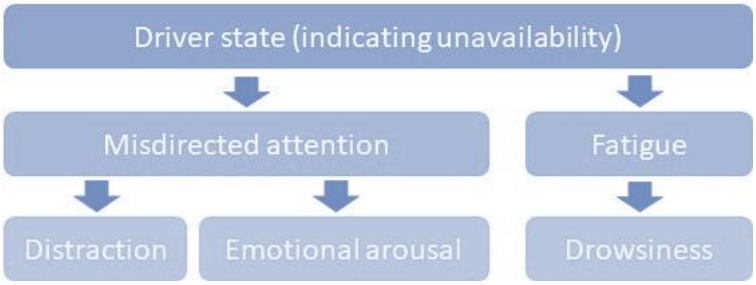

Figure 9. Types of driver inattention.

Drowsiness

Drowsy driving is assumed to be one of the major problems in the traffic safety. As many as 20% of crashes in Europe are believed to be due to driver drowsiness (around 100,000 crashes a year according to Barr et al., 2014. Although, it is difficult to quote precise numbers due to incomplete data most researchers agree that drowsiness may be actually a direct cause of a vast number of drivers' mistakes leading to the crashes (RoSPA, 2001).

Drowsiness is the state between wakefulness and sleep defined as a state of progressive impaired awareness associated with a desire or inclination to sleep (Kozak et al., 2006). Drowsy driving is defined as the instance of driving when the driver wishes to sleep (Barr et al., 2014), or driving while being impaired by a lack of sleep (Higgins & Fette, 2011). In the literature, the term drowsiness is often used interchangeably with some other terms, such as fatigue, vigilance, inattention, sustained driving, etc. NHTSA (Barr et al., 2014) differentiates between the terms drowsiness and fatigue as follows: drowsiness means wishing to sleep while fatigue is wishing to cease performing the current task. Drowsiness results in large variability of performance impairment over relatively short time intervals (Anderson et al., 2013) which may include: slow reaction time, decreased situational awareness, impaired judgement and microsleeps, as well as the danger of falling asleep.

Fatigue is an effect of physical labour or prolonged task defined as disinclination to further perform the task at hand (Yang et al., 2009). It is usually defined as a gradual process of progressive disinclination towards effort. It can be divided into sleep-related (drowsiness) and task-related. Consequently, although the drowsiness may be partially caused by the task-related fatigue, they can both appear separately (May & Baldwin, 2009). The task-related driver fatigue might be then subdivided into active and passive fatigue (Gimeno et al., 2006). It often happens that the driver falls asleep when subject to passive fatigue as it unmasks the already present sleepiness (e.g., sleep deprivation - induced; Yang et al., 2009).

Sleepiness can be defined as a difficulty in remaining awake even while carrying out activities (Dement & Carskadon, 1982).

Vigilance is a state of high efficiency in detecting and responding to environmental stimuli (Yang, 2007).

Distraction

Misdirected attention occurs when the demands of activities which are currently critical for safe driving are not matched due to the allocation of resources to other activities (Engström, 2013). One of the key aspects of misdirected attention is distraction. It can be further divided into *visual distraction* when the driver directs his sight into areas non-critical for safe driving task, while omits those that are critical and *cognitive distraction* related to mental load while the driver might be looking at the road, but not actually seeing it.

Driver distraction is one of the most important safety problems in the automotive domain. The statistics of WHO indicate driver distraction as a rising problem, leading to approximately 20% of accidents according to NHTSA estimates. The most commonly targeted type of distraction is the situation when the driver diverts their visual attention from the safety-critical areas, because they "allocate resources to a non-safety-critical activity while the resources allocated to the activities critical for safe driving do not match the demands of these activities" (Engström & Monk, 2013). The non-safety-critical activities are usually classified regarding the modalities engaged in task performance:

Visual-manual task –engages two important senses: visual and haptic channels (e.g., infotainment system operating, texting). This task not only requires cognitive resources, but also degrades road observations and steering wheel control.

Auditory-vocal task – engages mainly cognitive resources (e.g., conversations) while the sight may be focused freely due to the driving task demand.

Emotional arousal

Emotion is a certain feeling accompanied by some level of physiological arousal and some positive/negative valence. According to Plutchik (Plutchik, 1980) emotions can be divided to basic emotions as well as secondary or tertiary emotions which consist of some blended characteristics of one or more basic emotions (i.e., anger, fear, sadness, disgust, contempt, surprise, enjoyment-happiness; see Figure 10). Moreover, cross-cultural research, show the seven emotions are universal (Ekman, 2003; Ekman & Friesen, 1978).

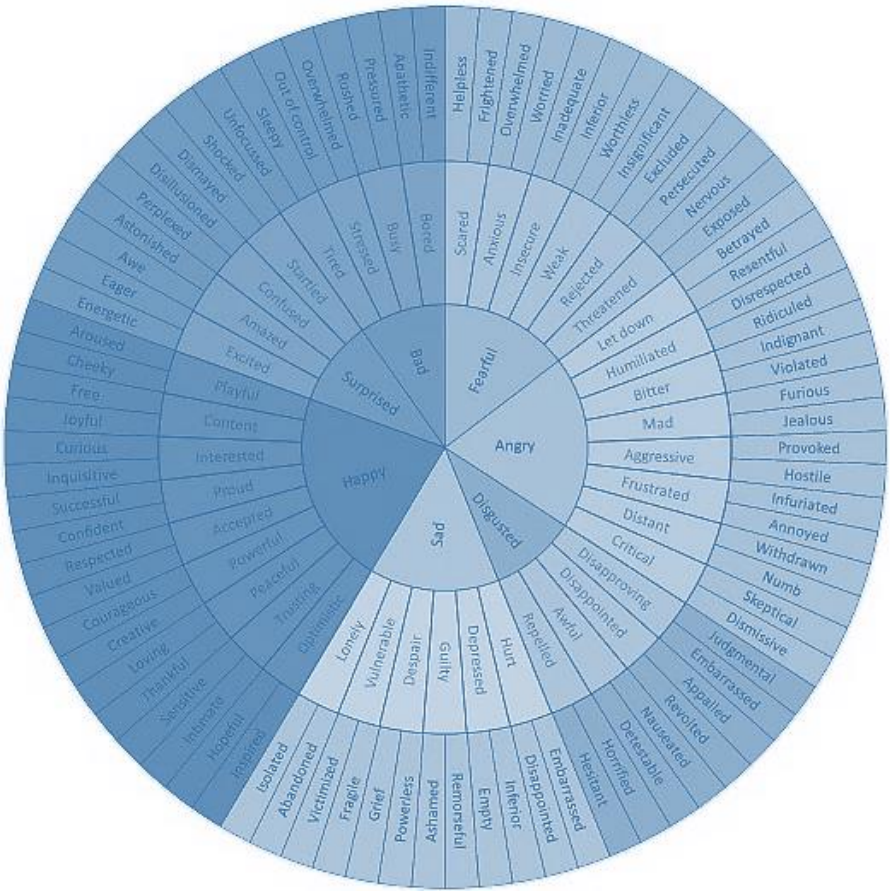

Figure 10. Wheel of emotions [source: Birmingham Education Partnership].

Ekman also managed to describe unique and universal signals of these seven emotions, based on Facial Action Coding System (FACS) research. For these seven emotions, he describes the characteristic physiological changes (like brow rise, eye openness, etc.) which constitute the due facial expression. For this reason, most emotion recognition research refer to recognizing seven emotional categories. The achievements of FACS also enabled research on facial emotion recognition through image analysis, making the area accessible for vide-based driver monitoring systems, which recognize driver emotional state on the basis of relative position of key facial landmarks and provide basis for concluding about mental states and emotional arousal level which is susceptible of hindering driver performance, e.g. due to high stress load (one of the most frequent driver-state-related reasons for car crashes and traffic accidents (Cartwright, Cooper & Barron, 1996).

Driver inattentiveness measures

The most frequently used driver inattentiveness measures can be divided into three groups according to the type of observation to be conducted: Controller Area Network data, data from external sensors, video-based measures, and physiological signals.

CAN-bus data include steering wheel behaviour data which include steering wheel angle as well as the frequency and intensity of steering wheel movements; pedals usage power (i.e., gas and brake); speed records.

Data from external sensors include: vehicle position and lane monitoring thee data are monitored with the use of external camera; LIDAR data allowing for scene analysis and obstacle detection as well as road signs recognition; GPS data enabling vehicle localization and movement tracking, and IMU data used to calculate orientation, angular rate, velocity, and acceleration of the vehicle.

Video-based measures include gaze location and gaze tracking, head position and orientation tracking, blinking behaviour, and emotions recognition.

Physiological signals include electrocardiography allowing for monitoring of heart activity; electroencephalography enabling brain activity monitoring, electrooculography used to monitor eyeball movements, galvanic skin reaction reflecting the level of arousal, as well as pulsometry and actigraphy allowing for polysomnographic monitoring and circadian rhythm monitoring.

All of the abovementioned measures can be found among the most often used driver monitoring methods (both in research and application field), section 2.3.1.3 presents the methods for data acquisition and provides insight into their usefulness for particular driver states detection. Some of the measures are more relevant in the situations when the driver controls the vehicle while they lose their meaning for driver monitoring while automation is engaged (i.e., pedals usage, steering wheel behaviour, lateral control). Other, remain valid regardless of the driving mode (e.g., video-based driver monitoring).

2.3.1.2 Validation procedure

For the purpose of validating the human-in-the-loop scenario three types of experimental procedures can be used varied in the level of control over the validation procedure and the level of natural feelings of the participant. The procedures involve laboratory testing allowing for driver state manipulations with no hazard to the participants safety, test track testing with safety driver engaged allowing for moderate level of driver state manipulation, and naturalistic driving procedure allowing for observation of the driver behavior in natural situations. Figure 11 presents the overview of differences between various study procedures.

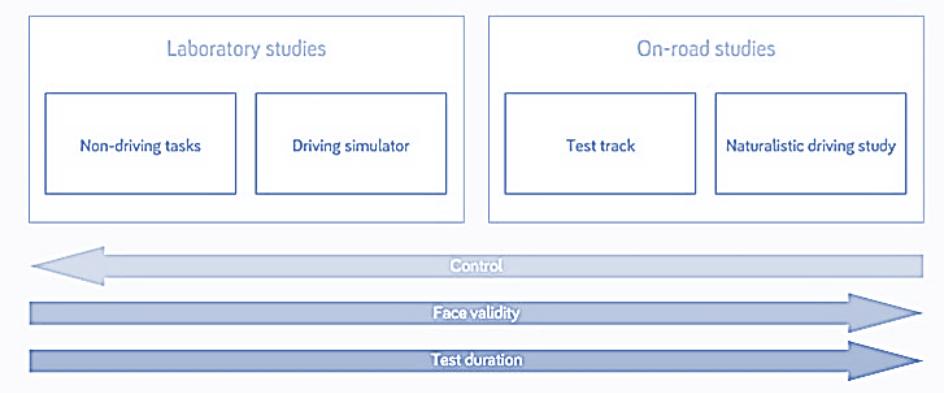

Figure 11. Overview of various validation procedures.

Experimental testing

Experimental methods for driver distraction testing include controlled manipulation of driver attention with the use of the so called secondary task (not related to driving), below, a set of most common tasks which can be used for such purpose is presented. This type of procedures can be used in laboratory studies with the use of driving simulator or, to some extent, in tests on a closed area (i.e., test track) when safety driver takes care of participant safety. The design of the experiments should follow guidelines formulated in ISO/TR 21959-2:2020.

Classic methods of testing for visual distraction usually base on differentiating between visual attention paid to the areas defined as critical for the driving task and those not related/critical to perform the drive. To increase the number of observations and the level of confidence about driver attention allocation, experimental tasks are used. Continuous Tracking Task (CTT; ISO/TS 14198:2012) is a task, that requires continuous control of a moving target on a tablet screen from the participant whose task is to keep it in the central position (see Figure 12). Surrogate Reference Task (SURT; ISO/TS 14198:2012) is a visual-manual task, which requires the participant to report whether a pre-specified target is embedded in a multi-item display. The task allows for applying various levels of demand on the participant with the use of several difficulty levels. In the task, the participant needs to find the target among similar distractors (see Figure 12).

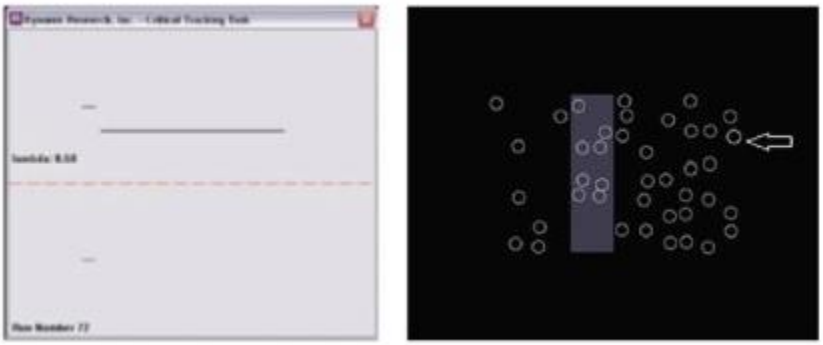

Figure 12. Typical screens of CTT (left) and SURT (right) tasks.

Another, more natural for the participant, but still highly controlled distraction manipulation is the Radio Tuning Task (RTT). In this task, the participant is asked to tune the radio band (on a tablet or smartphone) according to the instructions. The task is specified in the AAM guideline (Driver Focus-Telematics Working Group, 2006).

Emotional state manipulation is usually based on emotional cues and mood control. One of the methods is to present the participant with a video inducing emotional response. The other includes road advertising with emotional cues being presented in the driving simulation scenario. Regardless of the type of method, emotions-oriented trials require the initial mood control with a dedicated questionnaire administered before and after the manipulation. In case of observatory study short survey before and after the drive shall be filled by the participant.

Most popular drowsiness testing procedures refer to long, monotonous tasks (e.g., Anund, 2018), circadian rhythm (National Sleep Foundation, 2007 after: Bowman et al., 2012), and subjective feelings of

the subjects. Consequently, drowsiness related studies are usually conducted during the night (i.e., between 11 PM and 5 AM) with the use of driving simulator with dedicated long, monotonous scenario allowing the drowsiness and microsleep events to occur in safe conditions. The level of sleepiness is usually controlled with physiological measures and subjective evaluation mostly with the use of Karolinska Sleepiness Scale (KSS, Åkerstedt & Gillberg, 1990) - a subjective scale, in which the drivers assess their own sleepiness on a 9-point Likert-type scale. Yet another procedure to be employed in drowsiness studies is to use actigraphy during the naturalistic observatory study and choose drowsiness-related sequences of data during data postprocessing.

Naturalistic driving testing

The scenario will be also validated in real driving environment. A cohort of 25 drivers with a range of demographic and driving characteristics will be recruited. Their behavior will be monitored both in safety and/or ecofriendly advice. During the observation period, also the reliability of the scene analysis system will be monitored. The testing is planned to take place over a period of 6 months. The experimental group can be split in several groups, one for each specific type of cue prompted while driving (for instance, shift gear as soon as possible). After a short adaptation phase of 4 weeks, the subjects will drive once on a specific road without any advice. Then, the drivers will drive for 20 weeks on the same road after following specific advice. A multitude of experimental data, acquired either by online daily monitoring their vehicle usage and driving behavior through several in-car sensors or offline (before/after trips via online questionnaires), will ground the model with sufficient information for the quantification of the CPSoSaware user models & risk prediction framework.

2.3.1.3 Data acquisition

Data for this use case evaluation will be collected during real drive performed by multiple drivers in the vehicle equipped with multiple sensors (both exterior and interior monitoring system). The data that need to be collected are following:

- Ego Vehicle (CAN bus) data:
  - Statistics of pedals – show the level of control over the vehicle, sudden changes, and high frequency of use reveal inattention.
  - Speed statistics – reveal the level control over the vehicle speed and the smoothness of changes, allowing for calculation of variables indicating inattentiveness.
  - Steering behaviour – variables power and frequency of steering wheel moves, steering wheel angle, allow for calculation of certain variables like Steering Wheel Reversal Rate – highly correlated with drowsiness.
- Video-based driver Monitoring:
  - Gaze tracking – derived from head position and orientation as well as ocular activity, allows to calculate gaze vector, and related measures like gaze dispersion, gaze transitions, Total Eyes Off Road Time, Percentage Eyes Off Road Time, etc. Similarly to all camera-based measures, sensitive to changing illumination and possible occlusions.
  - Blinking behaviour derived from the distance between eyelids, allows to calculate eye closure ratio, blinking frequency and amplitude, Percentage of Eye Closure, and other related measures. Similarly to all camera-based measures, sensitive to changing illumination and possible occlusions.

- Head tracking – allows for calculating gaze-related measures, as well as detecting nods and other extreme positions of the driver's head. Apart from video-based measures, head tracking can be performed with IMU and motion capture systems. Similarly to all camera-based measures, sensitive to changing illumination and possible occlusions.

- Exterior sensing data:
  - LIDAR (Light Detection and Ranging) - is the active sensor that illuminates the target with laser light and measures the reflection with the sensor. Received signals are postprocessed and 3D representation of environment is created. LIDARs are broadly used in automotive for external sensing, received 3D Point Cloud can be used as input for many algorithms like object detection and classification, localization, and mapping.
  - RGB Cameras – RGB Video camera is a passive sensor collecting series of images creating 2D representation of the environment. External cameras are widely used for lateral behaviour monitoring, i.e. vehicle position and lane keeping monitoring, allowing for indirect monitoring of driver state. Moreover cameras, like LIDARs can be used for scene understanding algorithms, but because of missing depth coordinate in recorded data, estimation of distance to detected objects is less accurate. Another problem of camera-based external sensing is low quality of collected data in low illumination conditions.
  - GPS (Global Positioning System) - is a satellite-based positioning system. Thanks to known locations of the satellites and timestamps transmitted in each message, location of the receiver can be calculated. Unfortunately, in some situations, accuracy of GPS-based localization in not sufficient for automotive applications, especially in tunnels and densely built-up urban environments.
  - IMU (Inertial Measurement Unit) - is an electronic device consisting of accelerometers, gyroscopes and sometimes magnetometers. In automotive it is used to calculate orientation, angular rate, velocity, and acceleration of the vehicle.
- Physiological data:
  - Electroencephalography (EEG) - electrophysiological monitoring method to record electrical activity of the brain. Usually non-invasive performed with the use of electrodes attached placed on the participants head. EEG allows for very precise measurements but requires constant control of professional technician and involves the use of complicated headset.
  - Electrocardiography (ECG) - the process of producing a graph of voltage versus time of the electrical activity of the heart using electrodes placed on the skin. ECG allows for very precise measurements but requires constant control of professional technician and involves the use of complicated hardware.
  - Electrooculography (EOG) – a technique for measuring the corneo-retinal standing potential that exists between the front and the back of the human eye. EOG allows for very precise measurements but requires constant control of professional technician and involves the use of complicated hardware attached in the eye area and hinders the use of camera-based driver monitoring systems.
  - Galvanic Skin Response (GSR) – a method for measuring electrical skin conductance changing with physiological arousal. GSR allows for collecting very detailed data nut involves electrodes attached to the participant's body and the resultant data are often difficult to interpret.
  - Actigraphy - a non-invasive method of monitoring human rest/activity cycles. Wristwatch-like actigraphy unit is worn for a week or more to measure gross motor activity.

Assuming that the usefulness of various sensors differs between driver monitoring use cases (i.e., whether the vehicle is driver- or system controlled), and taking into consideration the limitations of each of the systems, CPSoSaware plans to use multiple sensors allowing for certain degree of redundancy, and at the same time securing proper functioning of the components even if one of the sensors experiences temporary drop down. Table 1 presents the overview of potential use of the proposed metrices for monitoring driver state.

Table 1. Overview of data sources for driver state monitoring.

Driver state	Driver monitoring measure															
	CAN bus			Video-based				External sensors			Physiological					
	Pedals use	Steering wheel use	Speed statistics	Blinking behaviour	Gaze tracking	Head tracking	Face expression	Lateral control	LIDAR	IMU	EEG	ECG	EOG	GSR	Actigraphy	
Drowsiness		+	+	+		+	+	+	+	+	+	+			+	
Distraction	+		+	+	+	+		+	+				+			
Emotional arousal							+		+		+	+		+	+	
Driving style	+	+	+					+		+						

Questionnaires

Apart from data collected with multiple sensors, another important part of validation of Human in the loop scenario is the trust assessment using questionnaires. A dedicated questionnaire will be developed for this purpose.

The questionnaire development will consist of several steps (see Figure 13) building up a reliable method of automation thrust and acceptance testing. The first step includes short interviews/surveys among a group of drivers aimed at pointing out the most common sources of doubts and distrust for automation, unacceptance or willingness to stick to manual driving mode. The results will be collected in the form of a questionnaire, in which one item is devoted to one issue reported in the first step. This first iteration of the questionnaire will undergo pilot procedure with a number of experts, who will be asked to assign each of the items to one of the three categories (thrust, acceptance, willingness to use). The third step, will involve pilot on a group of 60 drivers, asked to relate to each of the items on a 5-point Likert scale. The data will be used to choose the strongest items with the use of Cronbach's Alpha coefficient. In this step there will be also a confirmatory factor analysis performed leading to the final structure of the questionnaire. The questionnaire will consist of statements related to autonomous vehicle thrust, acceptance, and willingness to use, approximately 5-15 statements each. The respondents will be asked to relate to each of the statements on 5-point Likert scale.

In case of conducting the study on samples speaking in various languages, backtranslation with English as a source language will be used as a part of questionnaire preparation procedure to ensure comparability of the results.

Figure 13. Overview of the car automation trust questionnaire development process.

The questionnaire will be administered to a group of real drivers in order to assess social attitudes towards automation. In case of the group of drivers that took part in the validation procedure, test-retest procedure will be applied. The questionnaires will be administered twice: before the driver starts participating in the validation, and after the validation trials. The data collected will be then processed statistically, regarding demographic questions assisting the questionnaire. Repeated measures analyses will be applied to the test-retest sample.

Apart from the dedicated CPSoSaware questionnaire, the trials participants will go through widely used questionnaires to assess the quality of HRI in terms of their subjective workload during the interaction with car HMI:

- *NASA-TLX* (Hart & Staveland, 1988) - subjective, multidimensional assessment tool developed by the Human Performance Group at NASA that rates perceived workload, that can be used in order to assess the HRI quality.

Datasets

In addition to the data collected during the trial, multiple datasets for driver behaviour monitoring are available and can be used for models training and validation:

- General:
  - *SHRP2 Naturalistic Driving Study* - SHRP2 safety data from Virginia Tech Transportation Institute (VTTI) consists of two large databases: the naturalistic driving study (NDS) database and the roadway information database (RID). The databases include recordings of more than 5 400 000 trips recorded during the study taken by 3 147 volunteer drivers (ages 16-90+) over a 1- or 2-year period. The datasets provide over 36 000 crash, near crash, and baseline driving events. The data includes detailed video of the driver and the roadway, as well as data on the vehicles' speed, acceleration, braking, and other manoeuvres. Information such as seatbelt use and the presence of alcohol is also available. NDS trip data can be linked to roadway data from the RID, such as the roadway location, curvature, grade, lane widths, and intersection characteristics. The RID also provides environmental data such as time of day and weather.
  - *The Drive&Act* - the dataset is a state of the art multi modal benchmark for action recognition in automated vehicles from Fraunhofer IOSB and Karlsruhe Institute of Technology (KIT) collected in a static driving simulator. It offers 12h of video data in 29 sequences. The dataset includes: calibrated multi view camera system with 5 views (multi modal videos: NIR, depth and colour data), marker less motion capture: 3D Body Pose and Head Pose, model of the static interior of the car, 83 manually annotated activity labels.
  - *The Warwick-JLR DMD - Driver Monitoring Dataset* - the DMD is a joint venture between the Department of Computer Science at the University of Warwick and Jaguar Land Rover. It was collected and analysed as part of work for the PhD thesis, "Data Mining for Vehicle Telemetry Data" (Phillip Taylor, 2015). The data were collected with the use of real vehicle driven on a closed test track. The aim of the project was to investigate the driver monitoring from a data mining perspective, utilizing data from sensors which are readily accessible via the Controller Area Network (CAN)-bus. The authors formulate a classification problem with ground truth taken from physiological data including Heart Rate,

Heart Rate Variance, Electrodermal Magnitude and Frequency of Electrodermal Responses, and secondary task timings.

- Drowsiness:
  - *The University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD)* - the dataset was created for the task of multi-stage drowsiness detection, targeting not only extreme and easily visible cases, but also subtle cases when subtle micro-expressions are the discriminative factors. The dataset consists of around 30 hours of RGB videos of 60 healthy participants. For each participant there is one video for each of three different classes: alertness, low vigilance, and drowsiness (180 videos in total). There were 51 men and 9 women, from different ethnicities (10 Caucasian, 5 non-white Hispanic, 30 IndoAryan and Dravidian, 8 Middle Eastern, and 7 East Asian) and ages (from 20 to 59 years old with a mean of 25 and standard deviation of 6). The subjects wore glasses in 21/180 videos, and had considerable facial hair in 72/180 videos. Each video was self-recorded by the participant, using a cell phone or web camera of the participant. The frame rate was always less than 30 fps, which is representative of the frame rate expected of normal cameras used by the general population.
  - *NTHU Driver Drowsiness Detection Dataset* - the video dataset from National Tsuing Hua University consists of recordings of 36 participants both male and female drivers, from different ethnicities, in 5 kinds of scenarios. The scenarios contain BareFace, Glasses, Sunglasses, Night-BareFace and Night-Glasses. Each frame in the video is labelled with either drowsy or non-drowsy state. The videos are captured under different situations, including day and night, with different types of drowsy and non-drowsy activities. The videos are in 640x480 pixels, 30 frames per second AVI format without audio. The subjects were recorded while sitting on a chair and playing a plain driving game with simulated driving wheel and pedals. The total time of the entire dataset is about 9 and a half hours.
- Distraction:
  - *Distracted Driver Dataset* - a dataset from State Farm of 2D dashboard camera images, each taken in a car with a driver performing some action (texting, eating, talking on the phone, makeup, reaching behind, etc).
  - *RobeSafe Driver Monitoring Video (RS-DMV)* - the dataset is a set of video sequences of drivers, recorded with cameras installed over the dashboard. The dataset contains 10 video sequences. The drivers were fully awake, talked frequently and were asked to look regularly to rear-view mirrors and operate the car sound system. Sequences contain occlusions, illumination changes and other elements that are problematic to face tracking and driver monitoring systems using computer vision. Frames are recorded in gray-scale, at 30 frames per second, and stored as RAW video. Frame size of outdoor videos is 960x480 pixels, and 1390x480 for indoor videos. Faces in the videos have been marked with 20 points.
  - *A Multimodal Dataset for Various Forms of Distracted Driving* - the set includes data for n=68 volunteers that drove the same highway under four different conditions: No distraction, cognitive distraction, emotional distraction, and sensorimotor distraction. The experiment closed with a special driving session, where all subjects experienced a startle stimulus in the form of unintended acceleration—half of them under a mixed distraction, and the other half in the absence of a distraction. The recorded variables include: speed, acceleration, brake force, steering, and lane position signals, perinasal electrodermal

activity (EDA), palm EDA, heart rate, breathing rate, and facial expression signals; biographical and psychometric covariates as well as eye tracking data were also obtained.

- *EEE BUET Distracted Driving (EBDD)* - the clips of the database depict cautious driving as well as activities that cognitively distract drivers during driving. A Sony Cyber Shot 14.1 mega pixels camera was affixed on the front windshield facing the driver inside the vehicle. Videos of a number of drivers were captured in daylight on city roads and university campus in Dhaka, Bangladesh. The formed dataset is diverse in terms of landscape, illumination, vehicle, or road condition (smooth or bumpy). The age and experience of the drivers also vary significantly. The videos in the database can be broadly categorized as cautious or distracted driving. The distracted driving can be of four types - talking on cell phone, eating, texting on cell phone, and operating cabin equipment (inattentive). The videos have frame size of 854x480 pixels with a frame rate of 30 fps.

- Emotion recognition:
  - *Oulu-CASIA NIR&VIS facial expression database* - Oulu-CASIA NIR&VIS facial expression database contains videos with the six typical expressions (happiness, sadness, surprise, anger, fear, disgust) from 80 subjects captured with two imaging systems, NIR (Near Infrared) and VIS (Visible light), under three different illumination conditions: normal indoor illumination, weak illumination (only computer display is on) and dark illumination (all lights are off).
  - *MMI Facial Expression Database* - the database consists of over 2900 videos and high-resolution still images of 75 subjects. It is fully annotated for the presence of AUs in videos (event coding), and partially coded on frame-level, indicating for each frame whether an AU is in either the neutral, onset, apex or offset phase. A small part was annotated for audio-visual laughers. The MMI Facial Expression Database contains both six prototypical expressions and expressions with a single FACS Action Unit (AU) activated, for all existing AUs and many other Action Descriptors. Recently recordings of naturalistic expressions have been added. The database contains recordings of the full temporal pattern of a facial expressions, from Neutral, through a series of onset, apex, and offset phases and back again to a neutral face.
  - *Spontaneous Affective and Mental States* - the dataset contains 1184 multimodal facial video clips collected from 31 subjects in MP4 format. The 1184 video clips contain spontaneous facial expressions and speech of 13 emotional and mental states happiness, anger, sadness, disgust, fear, surprise, boredom, contempt, confusion, neutral, thinking, concentrating, and bothered.
  - *ISED Indian Spontaneous Expression Database* - the database consists of 428 segmented video clips of the spontaneous facial expressions of 50 participants. Emotions were induced among the participants by using emotional videos and simultaneously their self-ratings were collected for each experienced emotion. Facial expression clips were annotated carefully by four trained decoders, which were further validated by the nature of stimuli used and self-report of emotions. An extensive analysis was carried out on the database using several machine learning algorithms and the results are provided for future reference.
  - *KMU-FE Keimyung University Facial Expression of Drivers Database* - the dataset contains sequences captured in a real vehicle driving environment with an NIR camera. The KMU-FED database consists of drivers' facial expressions captured using an NIR camera installed on the dashboard or steering wheel. It contains 55 image sequences from 12 subjects,

which include various changes in illumination (front, left, right and back light) and partial occlusions caused by hair or sunglasses.

- *CMU-PITTSBURGH AU-Coded Face Expression Image Database* - facial behaviour was recorded in 210 adults (age 18 - 50 y). They were 69% female, 31% male, 81%% Euro-American, 13% Afro-American, and 6% other. They were observed in an observation room equipped with a chair on which to sit and two Panasonic WV3230 cameras, each connected to a Panasonic AG-7500 video recorder with a Horita synchronized time-code generator. One of the cameras was located directly in front of the subject, and the other was positioned 30 degrees to the subject's right. For approximately one third of subjects, ambient room lighting augmented by a high-intensity lamp was used. For the other two thirds, two high intensity lamps with reflective umbrellas were used to provide uniform lighting. Subjects performed a series of 23 facial displays; these included single action units and combinations of action units. 60 subjects performed head rotation to 30 degrees with facial expression, which was recorded with both cameras. 1917 image sequences from 182 subjects have been FACS coded for either target action units or the entire sequence. Thirty-degree views are available on videotape.

2.3.1.4 Outcome measures

The main goal of validation of human in the loop scenario is to assess process of taking control over vehicle by human driver and to evaluate driver behaviour and driving style of human driver. For this purpose, following metrics will be considered:

- Time-related (expressed in seconds; ISO/TR 21959-1:2020):
  - Take over time - time interval between the onset of RtI and user-initiated intervention, understood as deactivation of the engaged automation feature.
  - Decision time - time interval between detection of the RtI and the decision to disengage the automation feature.
  - Intervention time - time interval required by the driver to handle the imminent take-over situation by performing an appropriate driving manoeuvre.
  - Driving recovery time – the sum of take over time and intervention time.
  - Control stabilisation time - time duration it takes for an individual user to reach a similar or comparable quality level of manual driving performance as in ordinary level 0 driving by an average driver.
- Quality-related (expressed in seconds; ISO/TR 21959-1:2020):
  - Safety-oriented, objective take-over quality measures - measures to assess safety effects on the individual and on other traffic participants (e.g., collision avoidance, omission of visual checks, operating errors, pedals use, minimum time to collision, minimum time to lane crossing).
  - Sensitivity-oriented, objective take-over quality measures - measures related to lateral and longitudinal control (e.g., standard deviation (SD) of lateral position, SD of steering wheel angle, distance to other vehicles or objects, time headways, speed behaviour)
- Expert-based observer ratings – expert assessment of driver behaviour and its safety-related quality. Applicable to take-over situations as well as driving style modelling.
- Ego-Vehicle ADAS metrices:
  - Driver monitoring data from camera-based system

- Feedback Loop coming from ADAS functions (lateral deviations from Lane-Marking topology; driving behaviour modelled through statistical processing of signals (pedals, wheel, brakes; obstacle alert statistics - speed adaptation in association to object distance).

2.3.2 Cybersecurity issues in connected cars scenario

2.3.2.1 Detailed use case concept

A very important aspect is the cybersecurity protection of the CPSoS system. Nowadays vehicles have transformed into hyper connected and digital artefacts. Manufacturers and engineers need to innovate in this area in order to step ahead of the rest, which implies integrating more and more digital assets. This also opens the vehicles to the cyber threat landscape, which for this domain has a big surface attack and opens new cyberthreats with each innovation. Therefore, the communication and safe state of the vehicles is a critical issue as it could have an impact in human lives. For that reason, when an attack is detected it is necessary to protect the system, inform the users and, if necessary, react for the safety and security of the users. We plan to demonstrate a scenario of the connected vehicles that evaluates the protection against malicious attacks. These attacks range from stealing confidential data (e.g., position of the vehicle, status of the motor, etc.) to take control of IT devices. We will use CPSoSaware for modelling a secure and safe infrastructure of V2V or V2P where several cybersecurity requirements and needs will be defined and how they are fulfilled by the CPSoSaware cybersecurity solutions. We will analyse more in detail the more critical needs of cybersecurity of this domain and the reaction capabilities expected and how they can be integrated at design time using the capabilities of CPSoSaware.

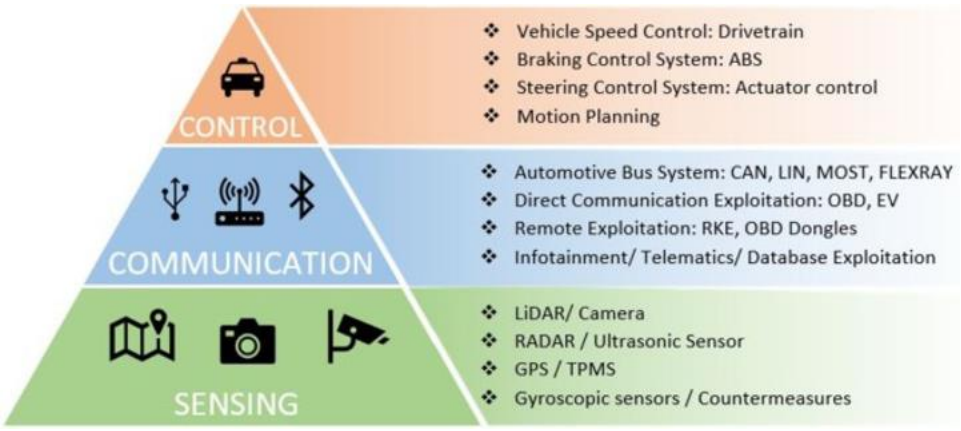

Figure 14. AV/ADAS vehicle sensing-communication-control framework [Source: El-Rewini et al., 2020].

- List of possible Sensor exploitation Cyber Threats:
  - On board Camera exploit,
  - GPS Sensor spoofing,
  - Lidar Sensor exploit,
  - Adversarial attack,

- Man-in-the-Middle based sensor information stealing (e.g., Car relay attacks, car position stealing).

- List of possible Vehicle Control Module Cyber Threats:
  - CAN bus data manipulation,
  - Malicious Firmware reprogramming,
  - OBD malicious access control.
- List of possible Network based V2V/V2X Cyber Threats:
  - Distributed Denial of Service,
  - Internet Enabled Exploits.

2.3.2.2 Validation procedure

All cyberattacks in this scenario will be validated in simulation environment. Various signal deformations will be applied to sensor data to simulate possible attacks affecting sensors. All perception data will be collected from the photorealistic simulation and resistance to cyber-attacks will be assessed using specific metrics. Raw localization GNSS data will be also stored to simulate and validate GPS spoofing detection algorithms.

Similarly, the attacks on CAN bus and Firmware and their detection using CPSoSaware will be validated using a simulated, control environment where packet sniffers will be provided in order to capture relevant data and Man in the Middle attack software will be used for injecting packages to the CAN bus automotive simulated network.

Attacks on communication layer of V2V/V2X will be validated in simulation environment including V2X Simulation module. Signals transmitted between vehicles and the infrastructure will be modified for fooling extended perception and cooperative localization algorithms.

In the entire scenario regarding cybersecurity issues in connected vehicles, the following CPSoSaware components will be used and validated:

- V2X Simulator,
- GPS localization,
- Deep Multimodal Scene Understanding,
- Security Runtime Monitoring.

2.3.2.2.1 GPS Spoofing:

A location spoofing attack attempts to deceive a GNSS/RTK receiver by broadcasting incorrect satellite signals, structured to resemble a set of normal satellite signals (e.g., GPS, GLONASS, GALILEO, etc.). These spoofed signals may be modified in such a way as to cause the receiver to estimate its location to be somewhere other than where actually is. One common form of a location spoofing attack begins by broadcasting signals synchronized with the genuine signals observed by the target receiver. The power of the counterfeit signals is then gradually increased and drawn away from the genuine signals.

This type of attack has already been successfully carried out in several scenarios, i.e., against boats or Unmanned Aerial Vehicles (UAVs). Following such philosophy, in CPSoSaware, the attack is carried out thanks to fake satellite signals transmitted by Software Defined Radio (SDR) hardware. Figure 15 shows a

possible implementation where a UAV (e.g., commercial-grade drone) is used for transmitting counterfeit signals. Other candidate implementations for demonstration include static transmitters carried by the attacker covering a specific target area, e.g., a crossroad.

For this attack, the CPSoSaware system will be able to detect when the satellite signals are spoofed thanks to a parallel stream of vehicle locations that does not rely on satellite signals, but rather in-car measurements readily available through the vehicle's CAN bus, which we call hereinafter the CPSoSaware secondary location stream. Such secondary location stream is based on a Bayesian filtering technique, which consists of two basic steps: (i) the prediction step and (ii) the update step. With Bayesian filtering, the motion of the vehicle is described through the characterization of the underlying physical laws, e.g., with a bicycle model, and the prediction on the future vehicle locations is obtained through on-board sensors, e.g., with the Inertial Measurement Unit (IMU) readings. In the update step, the forecasted vehicle locations are then fused with satellite-free global location measurements.

Figure 15. A possible implementation for the location spoofing attack.

Detection Approach:

In order to detect the location spoofing attack, in CPSoSaware, the secondary location stream will be compared with the obtained satellite-based locations. When the difference between the two sets of location measurements exceeds a predefined threshold, an alarm will be raised, and the location spoofing attack will be detected. The alarm will be communicated through the CPSoSaware agent/sensor infrastructure within the autonomous car and will be propagated to the CPSoSaware Runtime Monitoring system, which then will take appropriate countermeasures

The secondary location stream will be computed by an application within a security strengthen ECU (or a similar security dedicated antihacking device) provided by the CPSoSaware system, which will be installed inside the vehicle. Such an application will retrieve information regarding on-board sensor readings from the CAN bus and will fuse such information with satellite-free global positioning measurements thanks to a Bayesian filtering technique. An attack will be identified, and an alarm will be triggered, by comparing the coherence of the obtained secondary location stream with the location

information obtained by the satellite receiver (Figure 16). To this end, the system will take the secondary location stream as the ground truth and will exploit the knowledge of its error's covariance to validate the satellite-based location. If the probability of the measured distance between the averages of the two multivariate distributions is above a given pre-defined threshold, an alarm will be raised, i.e., a location spoofing attack will be detected.

Notably, the solution adopted by the CPSoSaware system is modular, and each block could be modified based on the available on-board sensors and on the available satellite-free global positioning measurements. As an example, the vehicle state model could be defined depending on the most accurate sensors present in the vehicle or on the sensors leading to the most accurate location predictions. In the same way, any satellite-free Global Positioning Measurement could be used, as far as it is possible to determine its error covariance matrix.

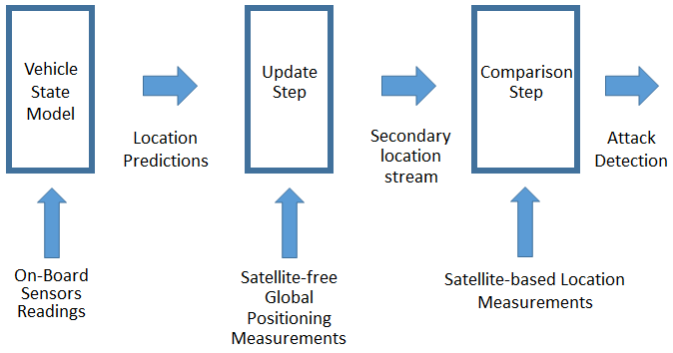

Figure 16. Block Diagram of the satellite-based location integrity check application.

Possible methods of cooperative localization that can be used as secondary location source in GPS spoofing attack detection are described in detail in section 2.3.3.2.1.

2.3.2.2.2 Data Integrity and Authenticity Threats

ECU CAN Bus protocol issues

The underlying CAN protocol has several inherent weaknesses that are common to any implementation. The key among these:

- Broadcast Nature

Since CAN packets are both physically and logically broadcast to all nodes, a malicious component on the network can easily snoop on all communications or send packets to any other node on the network. CARSHARK leverages this property, allowing us to observe and reverse-engineer packets, as well as to inject new packets to induce various actions.

- Fragility to DoS

The CAN protocol is extremely vulnerable to denial-of-service attacks. In addition to simple packet flooding attacks, CAN's priority-based arbitration scheme allows a node to assert a "dominant" state on the bus indefinitely and cause all other CAN nodes to back off. While most controllers have logic to avoid accidentally breaking the network this way, adversarially-controlled hardware would not need to exercise such precautions.

- No Authenticator Fields

CAN packets contain no authenticator fields — or even any source identifier fields— meaning that any component can indistinguishably send a packet to any other component. This means that any single compromised component can be used to control all of the other components on that bus, provided those components themselves do not implement defenses.

- Weak Access Control

The protocol standards for our car specify a challenge-response sequence to protect ECUs against certain actions without authorization. A given ECU may participate in zero, one, or two challenge-response pairs.

- Reflashing and memory protection

One challenge response pair restricts access to reflashing the ECU and reading out sensitive memory. By design, a service shop might authenticate with this challenge-response pair in order to upgrade the firmware on an ECU.

- Tester capabilities

Modern automobiles are complex and thus diagnosing their problems requires significant support. Thus, a major use of the CAN bus is in providing diagnostic access to service technicians. In particular, external test equipment (the "tester") must be able to interrogate the internal state of the car's components and, at times, manipulate this state as well.

In CPSoSaware we consider the ECUs are CPSs interconnected in a CPSoS. In the experimental setup, we are going to introduce in a controlled environment ECU to ECU CAN communication (emulating the ECU functionality in an of the self-embedded system) and using some CAN bus sniffing and manipulating software in order to mount an attack. The attack architectures, will follow the approach that is described in (https://projectswiki.eleceng.adelaide.edu.au/projects/index.php/Projects:2019s1-105_Hacking_CAN_Bus) will consider man-on-the-side (MOTS) and man-in-the-middle (MITM), where the attacking device is physically connected to the CAN bus. The MOTS architecture attaches the attacker's device to the CAN bus directly to read and insert new messages on the network (see Figure 17). In MITM, the attacking device is inserted between an existing ECU and the CAN bus, allowing it to listen, broadcast and intercept messages (see Figure 18).

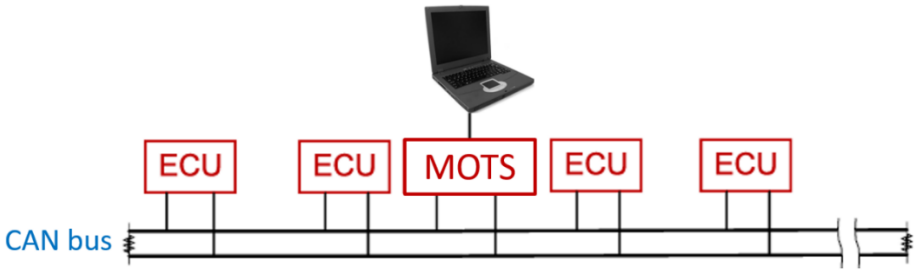

Figure 17. Man-on-the-side attack architecture.

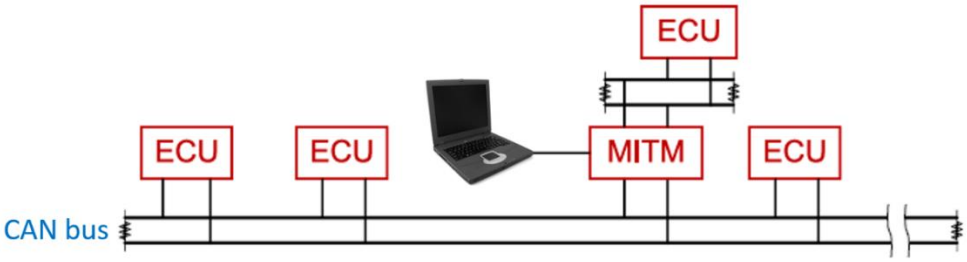

Figure 18. Man-in-the-middle attack architecture.

The attack architecture will be used in order to perform evaluation of Man in the Middle-based sensor information stealing (e.g., Car relay attacks, car position stealing) and CAN bus data manipulation as well as possible unauthorized access control to the CAN bus protocol information.

Malicious Firmware:

Changing ECU firmware has large implications as it can completely reprogram the vehicle's behaviour, resulting in it becoming a potential threat to public safety. The firmware could be modified or replaced by performing a physical and valid update via the On-Board Diagnostics (OBD) port. In principle this is a relatively easy procedure to perform; however, the use of asymmetric cryptographic (public-private key) architecture to ensure that the firmware came from a genuine source can mitigate this vulnerability.

However, ECUs within the vehicle might not be able to effectively handle the overhead of the Public Key Cryptographic schemes in order to match the automotive constrains. Also, any cryptography scheme needs to be implemented in such a way that it can be considered trusted. This means that it should be protected against various microarchitectural and side channel attacks that may be mounted through the OBD in order to obtain secret information (like the private keys).

Detection Approach:

In CPSoSaware we are aiming to redesign and reevaluate the concept of CPSs inside the CPSoS system. Therefore through an experimental CPS setup that can support the project's Model based design approach we can support a hardware assisted/accelerated security/cryptography mechanism that is able to retain the performance requirements of the Automotive domain (on ECUs) and still provide strong cryptography support to the CAN bus protocol by adding authentication and integrity on top of the protocol.

The security enhanced CPSoSaware CPSs (acting as next generation ECUs) will be able to identify potential data security problems and log them to the CPSoSaware Runtime monitoring system.

Also, in CPSoSaware we are designing appropriate CPS security agents/sensors that will be deployed as a decentralized attack detection mechanism within a connected car and forward detected alarms to the CPSoSaware runtime monitoring system. The security agents and security strengthen ECUs will have dedicated trusted hardware accelerated security primitives that will act as Public Key Infrastructure and will be able to handle security/cryptography related issues in order to enforce (data) security within the autonomous car. In Figure 19 such a detection example can be observed for a CAN bus packet injection attack scenario. By adding appropriate data integrity information on top of the CAN bus (handled by the security strengthened ECUs) we are able to detect the possible attack and propagate an alarm to the CPSoSaware Runtime monitoring system.

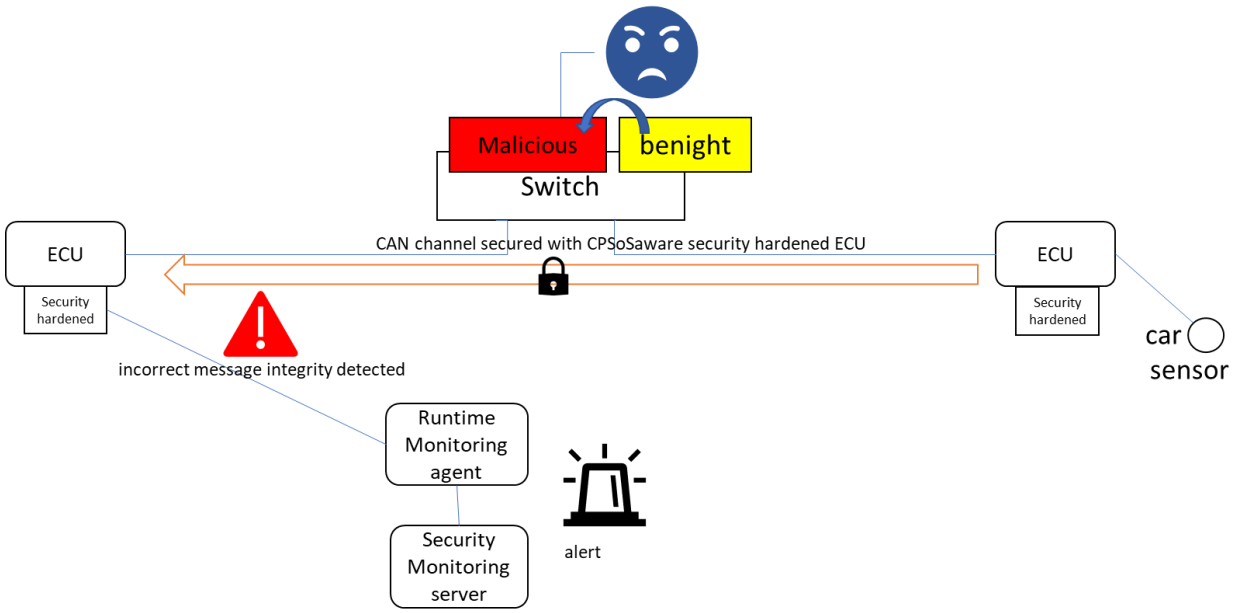

Figure 19. CPSoSaware detection of a Man-on-the-side attack.

2.3.2.2.3 V2V or V2X security exploits

V2X communication refers to broad remote communication technologies such as Dedicated Short Range Communications (DSRC), cellular networks, Bluetooth, Wi-Fi, Ultra-Wide Band (UWB) and Radio Frequency Identification (RFID) or even Remote Keyless Entry Systems (RKES).

In the scope of the CPSoSaware project and V2V/V2X security exploits consortium will focus on especially wireless access technologies and attacks detection taking into account threat model proposed by Boddupalli and Ray (2019). This taxonomy takes into account diverse types of attacks (e.g., masquerade, wormhole, man-in-the-middle) and assumes three main vectors of attacks for wireless communication: frequency of malicious communication, the effect of the attack on V2X (e.g., injection of fabricated message, message mutation or even preventing delivery of the message) and effect on the vehicle (e.g., compromising safety or loss of efficiency of the targeted cooperative application).

Attack Orchestration

To present the potential of the Security Runtime Monitoring system specific connected car application will be considered with relevant types of attacks e.g.,:

- Security of cooperative awareness system and specifically detecting threats to UWB-based communication (Zhang et al., 2006). UWB is a type of wireless communication that is known for its ability to transmitting high data flows for relatively low transmission power. UWB is also proposed as one of the very promising technology for precise localization of agents (vehicles, pedestrians, and others). Although UWB is one of the least cyber-vulnerable technologies Hennessy (2016) presents that the threat for UWB might be eavesdropping attacks. Security Runtime Monitor should secure from attacks also UWB wireless network of sensors.
- Security of cooperative automated driving scenario relying also on V2V communication for maintain safety gap between vehicles through targeting specific time headway values (e.g., platooning). The leading vehicle can communicate falsified acceleration values (higher than actual acceleration values, lower than actual acceleration values, stopping reporting acceleration values). The attack orchestration will assume mutating and changing the messages for V2X communication to account for different types of messages that will be received by the safety-critical cooperative automated system of the vehicle.

Detection Approach:

CPSoSaware project will develop relevant CPS components that will strengthen hardware assisted/accelerated security/cryptography mechanisms for V2V/V2X security threats. These components will be evaluated in the experimental setup that is based on AV-connected car simulator. Two sub use cases and testing scenarios will be prepared and assessed. All detected V2V/V2X security problems will be logged into CPSoSaware Runtime monitoring system.

In the UWB use case security agents will ensure secure localization scheme for UWB networks. One of the most widely discussed security case for UWB is the one associated with range-based systems (based on RSSI – Received-Signal-Strength-Indicator, AoA – Angle of Arrival, ToA-Time of Arrival and TDoA Time-Difference-of-Arrival for localization) where nodes cooperate to determine accurate locations. Attackers can exploit weakness of localization algorithms and e.g., attempt to reduce and enlarge distance estimates by impersonating one of the sensors and jamming the later genuine response. This will result in increased inaccuracy of location.

In the case of falsifying V2X messages designed detection approach will leverage the recent work in the machine learning field. Especially anomaly detection methods will be developed and tested in relevant cooperative scenarios. Anomalies (outliers, rare events, deviants) are pattern in data that do not match normal expected behaviour. Anomaly detection flag these patterns and unusual cases in the data, using knowledge from previous observations. Anomaly detection is being considered as one of the most promising methods for detecting potential threat events in a reliable manner. Anomaly detection approaches can involve supervised, semi-supervised and unsupervised learning. Most popular anomaly detection methods that will be considered in CPSoSaware project include clustering, nearest neighbour, statistical, classification and deep learning (autoencoders, GANs, variational autoencoders and sequence-to-sequence models). Important metrics of the anomaly detection model quality includes measures of precision and recall based on true positives and false positives.

2.3.2.2.4 Sensor layer attacks

Connected and automated vehicles depend on the collection large volumes of sensors data and processing them to operate safely by maintaining the field of safe travel. Contactless sensors layer attacks can disturb operations of the perception layer of the AV/ADAS stack of the vehicle and can cause hazardous road situations. Practically all types of sensors can be attacked e.g. Lidar might be attacked by recording outbound optical signal and sending it back to optical receiver, camera can be disturbed with pointed laser beam (that can also permanently damage its CMOS/CCD sensors) or direct illusional attack on specific classification machine-learning algorithm and ultrasonic sensor can be jammed by generating ultrasonic noise, spoofed by crafted fake ultrasonic echo pulses or even quieted.

Attack Orchestration

In the context of CPSoSaware project two types of attacks are being considered for testing Runtime Monitoring agent part and dealing with cybersecurity issues related to sensor layer attacks:

- Saturating/Blinding attack on sensor (Lidar, ultrasonic, camera) saturating renders the victim sensor unable to reflect the input signal changes. The victim systems can most often quite easily detect the attack but cannot prevent the sensor from saturating.

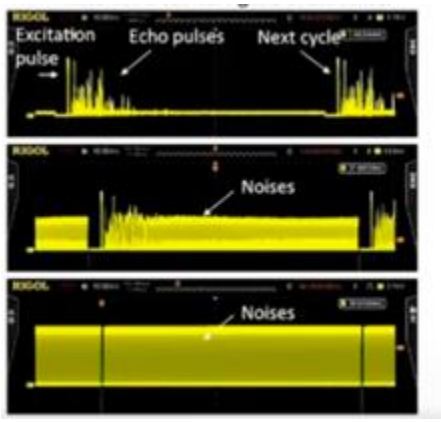

Figure 20. Received electrical signals at the ultrasonic sensors with artificially generated noise (no jamming, weak jamming, and strong jamming) [Source: Yan et al., 2016].

- Scenery attacks / illusion attacks - during scenery attack, attackers manipulate/modify physical objects present in the environment and driving scene in a way that they can result in incorrect classification by implemented in vehicles perception models. Most popular attacks include perturbations in traffic signs or road markings that are safety-critical e.g. attack can include adding dedicated peturturbions to the stop sign as shown below to misguide machine learning model such as ResNet. This can result in classification of stop traffic sign as speed limit of 60 km/h.

Figure 21. Stop sign before perturbation, perturbation and adversarial example (adding perturbation) classified as speed limit of 60 km/h by ResNet model [Source: Suo et al., 2019].

Detection Approach:

Research community indicates following risk mitigation solutions for sensor layers attacks: using robust optimization techniques while developing detection and warning vehicle algorithms (especially during the training phase for models based on machine learning) and applying multimodal sensor fusion techniques for sanity checks of received input and detections (fusion of camera, lidar and radar datasets).

Novel detection approaches for managing sensors data include using signal characterisation and anomaly detection for threat evaluation as shown by Bezemskij et al. (2016, 2016a). Anomaly detection methods can improve sensors layer attack identification by learning sensors behavioural pattern model. However, it should be noted that models should learn how to identify normal deviations caused by real environment, anomalies in the signature characteristics (relevant for particular sensors data source). Final model should detect anomalies using multiple data sources (sensors) to identify if vehicle operates normally. Having confirmation of multiple data sources data points can be classified as abnormal and detect cyber-physical threat to the sensor.

2.3.2.3 Data acquisition

For validation of cybersecurity scenarios, relevant data need to be collected in simulation tool used for autonomous scenario of CPSoSaware project. Data that will be stored and used for cybersecurity threats detection are following:

- Sensors data (GPS, Camera, Lidar, etc),
- CAN bus network packages,
- V2X messages,
- Experimental Firmware files,
- IP network traffic,
- OBD data.

2.3.2.4 Outcome measures

Following metrics will be used for assessment of cyber-attacks detection algorithms in the CPSoSaware platform:

- Fooling ratio: the amount of spoofed data that manage to be undetected by the CPSoSaware security Infrastructure will be a validation outcome measure against GPS and sensor spoofing attacks.
- Data integrity/authenticity failed attempts: various types of attacks will be mounted using MiTM and MoTS attack architecture. The detection rate of those attacks will be monitored, and an

appropriate acceptable threshold will be used to assess if the validation experiments are successful or not

- **Information gathering (eg. Port scanning) attempt number:** information gathering software (from the Information Technology Security Domain) will be used (eg. Nmap, OpenVAS etc.) will be used in order to perform internet-based exploit identification. The number of such attempts should be monitored and evaluated against the detected attempts by the CPSoSaware Runtime Monitoring System.
- **OBD unauthorized access control attempt number**: attack scenarios on the OBD port will be identified and will be launched. The CPSoSaware ECU strengthened ECUs and the CPSoSaware sensors/agents in the automotive CPSs will have to detect those attempts. The outcome metrics will be the number of such attacks that have been detected by the CPSoSaware security infrastructure.
- **False Positive/Negative ratio:** In the above scenarios we will also evaluate the CPSoSaware system based on the detected anomalies by the CPSoSaware Runtime Monitoring system and amount of those anomalies that constitute false positive or false negative.

2.3.3 Cooperative awareness scenario

2.3.3.1 Detailed use case concept

Principle of the cooperative awareness scenario is using data transmitted between multiple traffic agents and the infrastructure to improve performance of algorithms supporting autonomous driving (Figure 22). In field of cooperative awareness in CPSoSaware project, analysed scenarios can be divided in two groups:

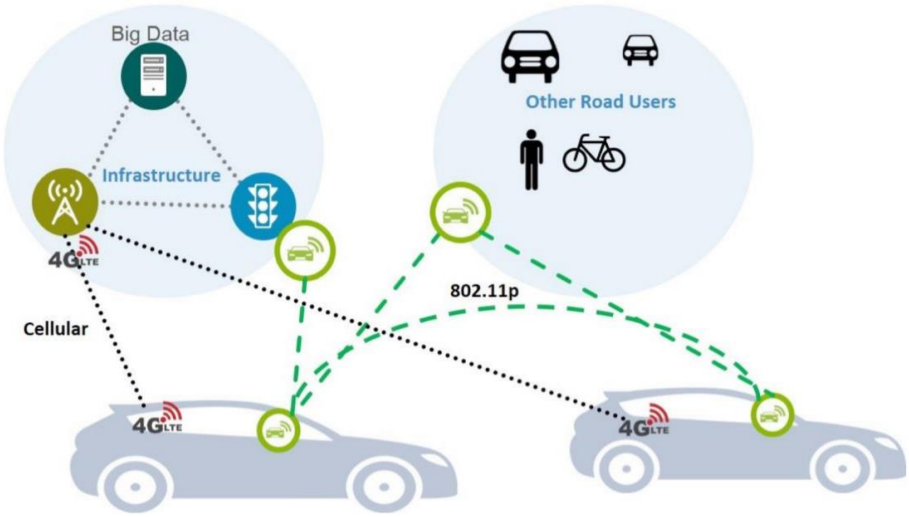

Figure 22. Principles of cooperative awareness scenario.

- Cooperative localization

Cooperative localization in automotive is based on Vehicular ad-hoc network (VANET). Consider a 2-D region where N interconnected, via V2X, vehicles of a VANET, are moving and collecting measurements. An example of VANET, is shown in Figure 23.

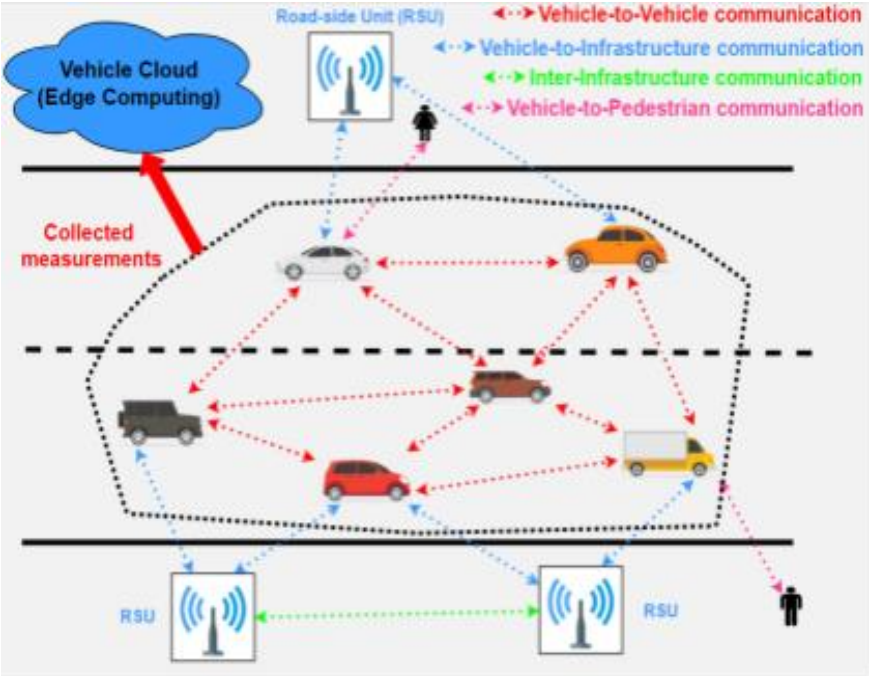

Figure 23. Example of VANET.

Each vehicle is able to know its absolute position from GPS and to measure its relative distance and angle of arrival to connected neighbouring vehicles using LIDAR, RADAR or UWB technology.

A common approach in cooperative localization is to formulate (based on the measurement models) an objective cost function C(x) (according to Maximum Likelihood Estimation or MLE) and to minimize it with respect to vehicle locations, in order to reduce the error of absolute position measurement. According to MLE, the relative or self-measurements depend only on the locations of nodes or location of the self-node involved. Thus, the desired cost function will have a number of different terms that will consider matching the estimated position from different modalities with the actual one.

We focus on the robustification of the cooperative localization approach, that can be considered as a maximum likelihood approach, assuming that the noise in the different modalities can be modelled as a normal gaussian noise. It can be considered as an unsupervised AI approach. The robustification strategies are based either: i) on estimating vehicles locations using also input from cameras and range measurement from geotagged images, ii) by imposing constraints on additional optimization variables that correspond to the GPS spoofing attacks.

- Driver situational awareness

Driver's situational awareness is widely acknowledged to be one of the key factors of driving safety (Gugerty, 1997). According to the Society of Automotive Engineers (SAE J3114), situational awareness means driver's understanding of the driving environment, including perception of environment, its comprehension, and anticipation of upcoming changes. One of the goals of Cooperational Awareness scenario is to increase driver's situational awareness by providing access to information otherwise out of the driver's field of view. In the connected car, the driver will have access to the traffic- and safety-related information from other vehicles, as well as from X2V and P2V systems, allowing for earlier reaction or more accurate decision making. The delicacy of these information providing, however, lies in the balance between providing the highest rate of useful information and avoiding the risk of overloading the driver's cognitive processing capacity, which would in turn result in longer and less accurate reactions (Engström et al., 2017).

- Path planning optimization

Another goal of cooperative awareness scenario is to improve path planning and safety systems of vehicles using extended perception (Figure 24) based on communication with other traffic agents and infrastructure. In cooperative awareness scenario, decisions that are taken by autonomous car can consider not only objects in field of perception of the sensors, but also not visible traffic agents, thanks to exchanging all useful data through V2X communication.

Cooperative awareness is the basis for most safety Intelligent Transportation Systems (ITS) applications proposed by standardization bodies. Using the information provided by cooperative messaging, vehicles and Road Side Units (RSUs) are able to create a map of their surroundings, which is then used as input for safety applications that detect potentially hazardous situations. To enable cooperative awareness, standardization bodies have proposed specific messages for that purpose: in the European Union (EU), Cooperative Awareness Messages (CAMs) have been specified as part of the standard, whereas in the United States (U.S.), the same functionality is enabled by the Basic Safety Message (BSM). These messages are exchanged periodically and contain location, speed, and direction of the vehicle, among other information.

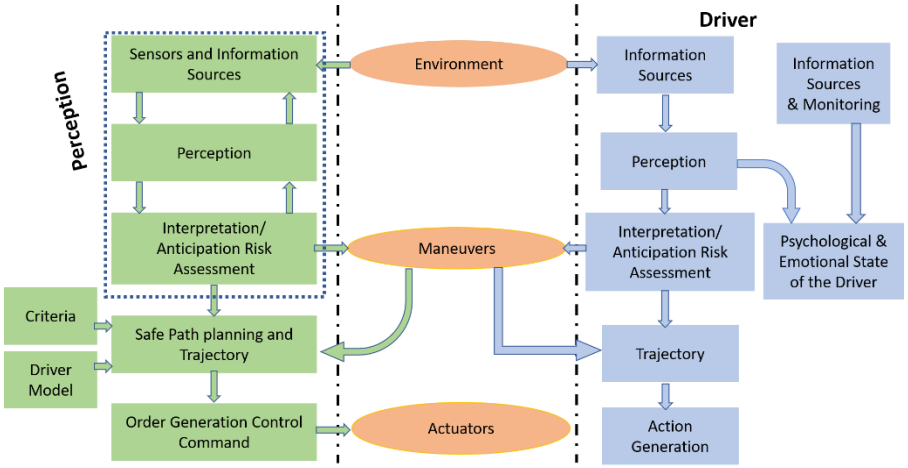

Figure 24. Extended perception schema.

2.3.3.2 Validation procedure

This scenario will be validated in simulation environment consisting of realistic 3D simulation and V2X simulator working in parallel. All useful data will be transmitted between traffic agents and the infrastructure to improve localization accuracy, extend the range of sensing, and enable improved local path planning and safety functions.

For offline analysis, all the exchanged data will be stored for each vehicle and benefits of cooperative awareness will be measured using predefined metrics. In cooperative scenarios defined in autonomous vehicles part of CPSOSaware project, scenario that can be analysed offline is cooperative localization.

For scenarios that require active modifications of vehicles trajectories, stored data becomes invalid after first decision about path of the vehicle is made. Because of that, offline analysis of collected datasets is not possible to correctly assess such cases and simulation-based validation will be performed instead.

CPSoSaware components that will be used and validated in cooperative awareness scenarios are following:

- V2X Simulator,
- Deep Multimodal Scene Understanding,
- Localization API,
- Path Planning API.

Validation procedures for each of the scenarios related to cooperative awareness are described below.

2.3.3.2.1 Cooperative localization

Goal of cooperative localization validation is to measure benefits of including additional data in localization algorithms, compared to pure GNSS based localization of vehicles. Validation procedure of this scenario will be performed by running simulation and collection of the following data:

- GNSS localization of each vehicle,
- Perception data,
- V2X messages.

With all the data collected, localization related metrics can be calculated for pure GNSS localization and cooperative localization algorithms and compared to assess the improvement of localization accuracy.

2.3.3.2.2 Driver situational awareness

To measure influence of extended perception on driver situational awareness, multiple validation scenarios will be defined. The scenarios will be tested in with the use of driving simulator, to ensure observation of driving characteristics as well as safe conditions for testing. For set of scenarios, take over, cognitive load, and situational awareness while performing the task will be measured to assess the benefit of cooperative awareness in considered scenarios.

Simulation scenarios should cover various complex scenarios with obstacles or other traffic agents occluded from ego vehicle perspective. In such scenarios, the benefit of cooperative awareness is expected to be the most significant due to valuable information received through V2X about potentially dangerous situations.

2.3.3.2.3 Path planning optimization

To measure influence of extended perception on path planning optimization and safety of traffic agents, multiple validation scenarios will be defined. For set of scenarios, path planning and safety related metrics will be measured for cases of only ego-vehicle perception and perception extended with V2X communication, to assess the benefit of cooperative awareness in considered scenarios (Figure 25).

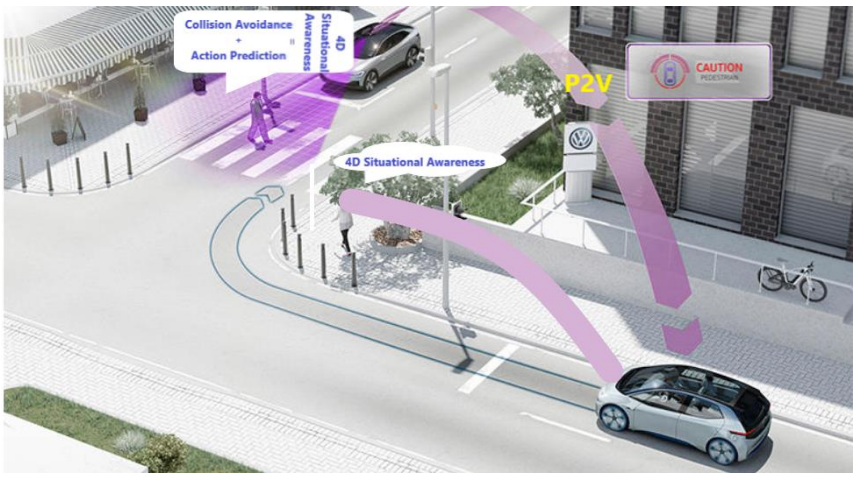

Figure 25. Example of cooperative awareness scenario.

Simulation scenarios should cover various complex scenarios with obstacles or other traffic agents occluded from ego vehicle perspective. In such scenarios, the benefit of cooperative awareness is expected to be the most significant due to valuable information received through V2X about potentially dangerous situations.

2.3.3.3 Data acquisition

Data for this use case will be collected in simulation environment, due to the limited number of test vehicles. Simulation will contain 3D photorealistic simulation with multiple simulated traffic agents and V2X simulation module for information exchange between them. The data that need to be collected for each vehicle are following:

- Ego vehicle:
  - Speed,
  - Trajectory,
  - Location,
  - Perception data.
- Surrounding vehicles data transmitted through V2X:
  - Speed,

- Trajectory,
- Location,
- Perception data.

Additional data need to be collected to measure driver situational awareness:

- Ego Vehicle (CAN bus) data:
  - Statistics of pedals – show the level of control over the vehicle, sudden changes, and high frequency of use reveal inattention.
  - Speed statistics – reveal the level control over the vehicle speed and the smoothness of changes, allowing for calculation of variables indicating inattentiveness.
  - Steering behaviour – variables power and frequency of steering wheel moves, steering wheel angle, allow for calculation of certain variables like Steering Wheel Reversal Rate – highly correlated with drowsiness.
- Cognitive load data:
  - *NASA-TLX* (Hart & Staveland, 1988) - subjective, multidimensional assessment tool developed by the Human Performance Group at NASA that rates perceived workload.
- Situational awareness assessment:
  - *Situation Awareness Global Assessment Technique* (SAGAT) (Endsley, 1990) - an objective measure of SA. SAGAT employs periodic, randomly-timed freezes in a simulation scenario during which all of the operator's displays are temporarily blanked. At the time of the freeze a series of queries are provided to the operator to assess his or her knowledge of what was happening at the time of the freeze.
  - *Situational Awareness Rating Technique* (SART) (Taylor, 1990) - a ten-dimensional self-rating technique which elicits the subjective opinion on how aware a person was during task performance.

For training and validation of all algorithms considered in cooperative awareness scenario, multiple datasets can be used. Due to the specific conditions of the cooperative awareness scenario, some datasets have to be created during the pilot. Available datasets that can be used for training and validation of specific parts of cooperative awareness use case are following:

- Scene Understanding/Perception:
  - Berkeley DeepDrive BDD100k – large dataset containing HD video sequences with GPS locations, IMU data and timestamps. Dataset consist of 1100 hours of recorded driving in New York and San Francisco area. Images are annotated with 2D bounding boxes for multiple categories: bus, traffic light, traffic sign, person, bike, truck, motor, car, train, and rider.
  - KITTI dataset - multipurpose autonomous driving dataset that can be used for 2D/3D object detection and tracking, visual odometry and stereo vision. Recorded data consist of video streams, 3D Velodyne point clouds, GPS and IMU data.
  - NuScenes – large public dataset for autonomous driving developed by Aptiv Autonomous Mobility. Dataset contains camera images, LIDAR point clouds, RADAR data and annotated 23 classes with accurate 3D bounding boxes.
- Cooperative Localization:

- Ford Multi-AV Seasonal Dataset - dataset collected by a fleet of Ford autonomous vehicles in different driving scenarios and environmental conditions. Each vehicle recorded data from multiple sensors: Velodyne Lidars, IMU, cameras. Additionally, ground truth locations, pose trajectories and 3D maps are available in the dataset.

2.3.3.4 Outcome measures

Goal of cooperative awareness is to assess the benefit of gathering additional data from other traffic agents and infrastructure for various algorithms. To measure it quantitively, following metrics will be compared for algorithms with only ego-vehicle data and for algorithms based on cooperative awareness:

- Localization accuracy:
  - Mean localization error – average error of calculated location, calculated from set of validation scenarios. Mean error gives insight about algorithm's mean accuracy but analysing only this metric is not sufficient to evaluate reliability of the algorithm - several observations with big localization error does not affect mean accuracy in the test containing lots of measurements.
  - CDF of localization error - Cumulative Distribution Function of localization error shows actual performance of the algorithms. It shows distribution of error in all observations, so analysis of CDF gives information about performance of algorithm over whole dataset. Despite CDF is fully informative, it is a function, not a single number, so it is not trivial to compare performance of multiple algorithms using this metric.
  - CEP parameters – Circular error probability is a metric widely used in localization problems. Value of CEPXX means: "what is the distance that XX% of observations has error lower than". It can be easily read from CDF, but it makes comparison of algorithms using single CEP value more straightforward than using complete CDF function.
- Safety related metrics:
  - Number of collisions - collisions count in assumed time frame or assumed rout length. Relative collision ratio with and without CPSoS components engaged will be calculated as performance assessment metrics.
  - Time to collision - time (seconds) from needed for the ego vehicle to meet the front forward object considering the ego vehicle and the object speed. Other variations are also possible, e.g. time headway - time (seconds) from needed for the ego vehicle with its current speed to reach the position of the front forward object.
- Path optimization related metrics:
  - Mean path distance (difference between Ground Truth and reference path) – mean of the distance between the observed path and the optimal path for covering the given distance calculated with account of individual preferences (to drive on the left, centre, or right of the lane).
  - Relative Standard Deviation of Path Distance – variable calculated on the basis of Mean path distance according to the following formula: $100*s/|\bar{x}|$ (where s is the standard deviation and $|\bar{x}|$ is the mean path distance).
  - Path Anomaly – deviations from the estimated optimal path larger than one standard deviation.

- Estimated time of traversal - this metric includes the change of ego vehicle movement direction and the total rotational movement the vehicle has to perform.

- Driver situational awareness related metrics:
  - Time-related (expressed in seconds; ISO/TR 21959-1:2020):
    - Take over time - time interval between the onset of RtI and user-initiated intervention, understood as deactivation of the engaged automation feature.
    - Decision time - time interval between detection of the RtI and the decision to disengage the automation feature.
    - Intervention time - time interval required by the driver to handle the imminent take-over situation by performing an appropriate driving manoeuvre.
    - Driving recovery time – the sum of take over time and intervention time.
    - Control stabilisation time - time duration it takes for an individual user to reach a similar or comparable quality level of manual driving performance as in ordinary level 0 driving by an average driver.
  - Quality-related (expressed in seconds; ISO/TR 21959-1:2020):
    - Safety-oriented, objective take-over quality measures - measures to assess safety effects on the individual and on other traffic participants (e.g. collision avoidance, omission of visual checks, operating errors, pedals use, minimum time to collision, minimum time to lane crossing).
    - Sensitivity-oriented, objective take-over quality measures - measures related to lateral and longitudinal control (e.g. standard deviation (SD) of lateral position, SD of steering wheel angle, distance to other vehicles or objects, time headways, speed behaviour)

3 Conclusions

Quantification trials in CPSoSaware project will provide initial assessment of the components developed in the project. Validation of the platform will be performed from use cases perspective, based on 2 scenarios: automotive and human-robot interaction in manufacturing environment. In this document, validation procedures, collected data and used metrics were introduced and the validation of CPSoSaware components will be performed in the next tasks of WP6 accordingly. Next steps in WP6 are following:

- Definition of Evaluation trials,
- Small-scale trials,
- Preliminary evaluation of CPSoSaware platform,
- Final full-scale trials.

The presented document constitutes the first iteration of Quantification Trials Definition and Planning which serves as a guide for field trials deployment. The document is constructed as a catalogue of methods and procedures valuable in trials devoted to testing and evaluation of CPSoS components, ensure the best performance of CPSoSaware system. The proposed methodologies rely on research literature, industrial standards, and state-of-the- art methods. However, for the sake of choosing best-fitted methodologies for the current project challenges and needs, the contents of the document will be updated through the WP6 activities resulting in the second iteration of the report providing the report of CPSOS evaluation trials together with detailed methodology and analysis of achieved results.

References

14198 ISO/TS (2012). *Road vehicles — Ergonomic aspects of transport information and control systems — Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems*.

21959-1 ISO/TR (2020). *Road vehicles — Human performance and state in the context of automated driving — Part 1: Common underlying concepts*.

21959-2 ISO/TR (2020). *Road vehicles — Human performance and state in the context of automated driving — Part 2: Considerations in designing experiments to investigate transition processes*.

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2), 29-37.

Anderson, C., Chang, A. M., Sullivan, J. P., Ronda, J. M., & Czeisler, C. A. (2013). Assessment of drowsiness based on ocular parameters detected by infrared reflectance oculography. *Journal of Clinical Sleep Medicine*, 9(09), 907-920.

Anund, A. (2018). *Intra-individual difference in sleepiness and the effect on driving performance – a three-times repeated driving simulator study*. Paper presented at The 6th International Conference on Driver Distraction and Inattention, DDI2018. Gothenburg, Sweden.

Barr, L., Popkin, S., & Howarth, H. (2014). *An evaluation of emerging driver fatigue detection measures and technologies* (No. FMCSA-RRR-09-005). United States. Federal Motor Carrier Safety Administration.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71-81.

Bezemskij, A., Anthony, R.J., Loukas, G. and Gan, D. (2016). Threat evaluation based on automatic sensor signal characterisation and anomaly detection. ICAS2016, IARIA

Bezemskij, A., Loukas, G., Anthony, R.J. and Gan, D. (2016a). Behaviour-based anomaly detection of cyber-physical attacks on a robotic vehicle. CSS-2016, IEEE

Birmingham Education Partnership. Retrieved from https://bep.education/wp-content/uploads/2017/05/10-Plutchiks-Wheel-of-Emotions.jpg

Boddupalli, S. Ray, S. (2019). REDEM: *Real-Time Detection and Mitigation of Communication Attacks in Connected Autonomous Vehicles Applications in: Internet of Things. A Confluence of Many Disciplines*, Second IFIP International Cross-Domain Conference, IFIPIoT 2019, Tampa, FL, USA, October 31 – November 1, 2019, Revised Selected Papers.

Bowman, D., Hanowski, R. J., Alden, A., Gupta, S., Wiegaud, D., Baker, S., . . . Wierwille, W. 2012. *Development and Assessment of a Driver Drowsiness Monitoring System* (Raport no: FMCSA-RRR-12-008). Washington, DC : Federal Motor Carrier Safety Administration, 2012.

Cartwright, S., Cooper, C. L., & Barron, A. (1996). The company car driver, occupational stress as a predictor of motor vehicle accident involvement. *Human Relations*, 49(2), 195-208.

Dement, W. C., & Carskadon, M. A. (1982). Current perspectives on daytime sleepiness: the issues. Sleep: *Journal of Sleep Research & Sleep Medicine*.

Driver Focus-Telematics Working Group. (2006). *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems Including 2006 Updated Sections*. Washington, DC.

El-Rewini, Z., Sadatsharan, K., Selvaraj, D. F., Plathottam, S. J., & Ranganathan, P. (2020). Cybersecurity challenges in vehicular communications. *Vehicular Communications*, 23, 100214.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, NY: Henry Holt.

Ekman, P., & Friesen, W. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologist Press.

Endsley, M. (1990). A methodology for the objective measurement of pilot situation awareness. *AGARD, Situational Awareness in Aerospace Operations* 9 p (SEE N 90-28972 23-53).

Engström, J. (2013). *Driver inattention and distraction in UDRIVE*. Presentation on 5th Driver Distraction and Inattention Conference. Paris, France.

Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human factors*, 59(5), 734-764.

Engström, J., Monk, C. A., Hanowski, R. J., Horrey, W. J., Lee, J. D., McGehee, D. V., ... & Victor, T. (2013*). A conceptual framework and taxonomy for understanding and categorizing driver inattention*. Brussels, Belgium: European Commission.

Gimeno, P. T., Cerezuela, G. P., & Montanes, M. C. (2006). On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures. *International Journal of Vehicle Design*, 42(1/2), 67-86.

Gugerty, L. J. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, 3(1), 42.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Hennessy, A.P. (2016) *Implementation of Physical Layer Security of an Ultra-Wideband Transceiver*, Ph.D. thesis.

Higgins, L., & Fette, B. (2011). *Drowsy Driving*. College Station, TX : Texas Transportation Institute, 2011

J3114 SAE (2016), *Human Factors Definitions for Automated Driving and Related Research Topics*, 2016

Joosse, M., Sardar, A., Lohse, M., & Evers, V. (2013). BEHAVE-II: The revised set of measures to assess users' attitudinal and behavioral responses to a social robot. *International Journal of Social Robotics*, 5(3), 379-388.

Kozak, K., Pohl, J., Birk, W., Greenberg, J., Artz, B., Blommer, M., ... & Curry, R. (2006, October). *Evaluation of lane departure warnings for drowsy drivers. In Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 22, pp. 2400-2404). Sage CA: Los Angeles, CA: Sage Publications.

Lasota, P. A., Fong, T., & Shah, J. A. (2017). A survey of methods for safe human-robot interaction. *Foundations and Trends in Robotics*, 5(4), 261-349.

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), 218-224.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437-454.

Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis*. New York, NY: Harper and Row.

Royal Society for the Prevention of Accidents. (2001). *Driver fatigue and road accidents a literature review and position paper*. Birmingham : Royal Society For The Prevention Of Accidents, 2001

Suo, D. Chen, S.J. Goggin, L. (2019) State-of-the-art Security Attacks on Sensor-Rich Automated Vehicles with Wireless Linkage for Remote Access and Control: A Survey. MIT CSAIL online projects repository: https://courses.csail.mit.edu/6.857/2019/project/13-Suo-Chen-Goggin.pdf

Taylor, R. M. (1990). Situation awareness rating technique (SART): the development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations* (Chapter 3). France: Neuilly sur-Seine, NATO-AGARD-CP-478.

Yan, C. Liu, J. Xu, W. (2016) Can You Trust Autonomous Vehicles: Contactless Attacks against Sensors of Self-Driving Vehicles, Def Con 2016.

Yang, J. H., Mao, Z. H., Tijerina, L., Pilutti, T., Coughlin, J. F., & Feron, E. (2009). Detection of driver fatigue caused by sleep deprivation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(4), 694-705.

Zhang, Y., Liu, W., Fang, Y., & Wu, D. (2006). Secure localization and authentication in ultra-wideband sensor networks. *IEEE Journal on Selected areas in communications*, 24(4), 829-835.