A Review of the Relationship Between Cyber-Physical Systems, Autonomous Vehicles and Their Trustworthiness

Craig Morrison[a], Elena Sitnikova[a], Shraga Shoval[b]

[a]University of NSW, Australian Centre for Cyber Security, ADFA Canberra, ACT, Australia
[b]Ariel University, Industrial Engineering & Management, Israel

craig.morrison@student.adfa.edu.au
e.sitnikova@adfa.edu.au
shraga@ariel.ac.il

Abstract: A concentration of development in cyber-physical systems containing advanced sensors, sub-systems and machine-learning algorithms over the past decade is equipping unmanned aerial and road vehicles with autonomous decision-making capabilities. The level of autonomy depends upon the make-up and degree of sensor sophistication and the vehicle's operational application, which can range from extremely high cost military-based combat vehicles, to commercially available civilian motor vehicles and very low-cost sub $100 hobby drones. As a result, the risk of autonomous vehicles being compromised and used as an improvised threat has become more serious and therefore trust-based methods to mitigate them are needed. This paper reviews the relationship between cyber-physical systems, automated and autonomous vehicles and their trustworthiness. It examines the significant role cyber-physical systems play in building sub-system, system, and system-of-system layers found in contemporary vehicles as well as the level of human involvement in decision control loops, and the trustworthiness challenges this creates across levels of vehicle automation and autonomy. Further, the paper reviews material on methods to authenticate a vehicle's "world-view" and the veracity of information it may communicate and share with other vehicles and infrastructure. It finds that to be trustworthy, automated and autonomous vehicles will need to implement data verification frameworks and methods. It recommends that future research be conducted into the effectiveness of implementing a pervasive trustworthiness hub where information flowing between heterogeneous autonomous vehicles and infrastructure within their surrounding environment can be authenticated.

Keywords:
Cyber-physical systems, autonomy, unmanned vehicles, human in/on the loop, trustworthiness, cyber-security

Introduction

Intensive development in advanced sensors, sub-systems and machine-learning algorithms over the past decade has equipped vehicle-based cyber-physical (VCP) systems with automated and autonomous decision-making capabilities. Depending upon the level of autonomy, humans can be "in-the-loop" for decision control, play a supervisory "human-on-the loop" role for escalated decisions, or be completely "out-of-the loop" where there is no human intervention or control over a system to reliably carry out its purpose (Nothwang et al., 2016, Issitt, 2016, Weber, 2014). Such high-level systems autonomously process and interpret an array of complex inputs to form a "world-view" contextual map of the surrounding environment and an awareness of its place within it without any operational human assistance (Anderson et al., 2014). The level of autonomy depends upon the mixture and degree of sensor sophistication and the vehicle's operational application which can range from extremely high cost military-based combat vehicles, to commercially available civilian motor vehicles and very low cost, sub $100 unmanned aerial vehicle (UAV) hobby drones. However, as predicted by Moore's Law, the gap between technology cost and sophistication is falling rapidly as the volume of component mass production increases to meet market demand (Mulay, 2016). As a result, there is a significant and present risk of easily accessible, autonomous VCP systems being able to be compromised and used as improvised threats in both theatres of conflict and civilian contexts. The threats are often unidentified and they will become more serious unless trust-based methods to mitigate them are further researched and better understood (Hartmann and Giles, 2016).

The purpose and significance of this paper is to explore and understand the trustworthiness problem inherent in automated and autonomous VCP systems through a review of contemporary literature and current

commercial trends. In the first section cyber-physical (CP) systems are explained within a vehicular context, their classification and application. The degree of vehicular automation and autonomy is then examined in section two where autonomous vehicles (AVs) are viewed from different human-machine interaction levels: 1) where humans are required to be "in-the-control-loop" for decision support; 2) where humans are required to be "on-the-loop" for escalated supervisory decision support; and 3) fully independent autonomous decision making where humans have no control and are "out-of-the-loop". Section three cross-examines the prior findings of each section to consider AV trustworthiness across three dimensions of trust: transparency, competence and management. Finally, the paper discusses methods to authenticate a vehicle's "world-view" and the veracity of information it may communicate and share with other vehicles and infrastructure. It finds that to be trustworthy, automated and autonomous vehicles will need to implement data verification frameworks and methods. It recommends that future research be conducted into the effectiveness of implementing a pervasive trustworthiness hub where information flowing between heterogeneous autonomous vehicles and infrastructure within their surrounding environment can be authenticated.

1. Cyber-physical systems and vehicles

A cyber-physical system is simply an integration of computation with physical processes (Lee and Seshia, 2014, Stojmenovic and Zhang, 2015) and the term is often used to describe the interplay between physical devices and sophisticated control system software (Pathan, 2016). Researchers further clarify this is as being embedded computers and networks monitoring physical processes, with feedback control loops affecting computations and, in-turn, physical responses (Lee and Seshia, 2014, Ekedebe et al., 2016). Lee and Seshia (2014) particularly note that CPS is about the interaction, not the union between physical and cyber systems and that it is not sufficient to just understand the physical and computational component separately.

The cyber in cyber-physical refers to cybernetics which is credited to Norbert Weiner, who used the term in 1948 to describe feedback concepts between men and machines (Wiener, 1948). He derived the term from the Greek kybernetes which aptly means to steer or govern. He described cybernetics as the conjunction of control and communications and his work continues to have a fundamental impact on the development of control systems theory which encompasses the conjunction of physical processes, computation and communication. (Krämer, 2014, Lee and Seshia, 2014). Early examples of CPS include a radio-based remote controlled submarine demonstrated by Nikola Tesla in 1898 and his 1926 vision of 'teleautomation' for a pocket-based instrument that wirelessly connects humans across the earth to form a "huge brain" (Stojmenovic and Zhang, 2015), exemplified in the ubiquitous smart phone and internet of today. The Z3 real-time computer invented by Konrad Zuse in 1941 is another example (Krämer, 2014). The Z3 was programmed to read the values from some forty analogue to digital sensors and processed them as variables to help calculate aerodynamics decision making in aircraft wing design.

Zuse's rudimentary sensors have evolved to create layers of functional cyber-physical control systems which are substantially represented in today's modern motor vehicles (Figure 1). At the First layer, components such as sensors, actuators and electronic control units are integrated with Second layer drive train control sub-systems. At the Third layer, drivetrain systems are tightly interconnected with each other to create a cohesive vehicle CPS architecture supervised by an integrated high-level vehicle control system (Carbone et al., 2014). Researchers note a Forth peer-to-peer layer in the latest generation of vehicles that operates beyond the vehicle to interactively sense or network with their external environment. This type of interconnection is known as vehicle to vehicle (V2V), and vehicle to infrastructure (V2I) communication (Gáspár et al., 2014, Gerla and Kleinrock, 2011, Gerla and Reiher, 2016). A tangible example of a V2V system is adaptive cruise control where the distance to, and the velocity of a vehicle ahead is sensed by a radar subsystem of the following vehicle. An example of V2I is a vehicle receiving navigation and guidance information about the condition of the road ahead. Finally, researchers discuss a Fifth internetworked layer where road vehicle ad hoc networks (VANETs) support various applications including allowing groups of vehicles to be controlled by a supervisory CPS in order to achieve a common purpose (Gáspár et al., 2014, Ekedebe et al., 2016, Gerla and Kleinrock, 2011, Gerla and Reiher, 2016). A real-world example of this is known as "highway platooning" where a group of road vehicles in close physical proximity to each other, one behind the other on highways, are assembled together to form a "road train". The road train of vehicles is then controlled by a supervisory autonomous system, usually within the lead vehicle. Platooning benefits include lower driver fatigue, increased fuel efficiency and improvements in safety outcomes (Crandall and Formby, 2016). Drivers of autonomous vehicles with compatible systems can relinquish control of their vehicle to join a platoon at a time of their choosing and can leave it at any time or when close to their destination.

When grouped, the multiple layers of cyber-physical systems create a complete vehicle CPS, and multiple interacting vehicle CPSs therefore represent a higher-level cyber-physical system-of-systems (Gerla and Reiher, 2016). Each functional level is dependent upon the delegated trusted performance of the level below for lower-order decisions and outcomes. Consequently, each layer relies upon a systems control loop ranging from high-frequency millisecond decision making at lower levels to intermittent event-based interrupts at higher levels. An example of delegated trust is the event based ability of a driver to relinquish control over a vehicle and join a highway platoon, which is dependent upon Global Navigation Satellite System (GNSS) positioning and adaptive cruise control, which then constantly rely on vehicle inputs such as current speed over ground and proximity to the vehicle ahead. They then rely upon drivetrain actuators to accelerate and brake the vehicle, which in turn, rely on GNSS, gyro and radar-based sensors to measure the relative position and speed of the vehicle ahead.

Importantly, a vehicle may not perform reliably, or as designed if any one of the dependent CPS layers cannot be trusted to operate accurately, particularly if the outcomes in one or more levels were corrupted or hacked to provide spurious inputs to the layer above. Such hacks could well cause a deceptive synthesis of the vehicles contextual "world-view" and therefore untrusted, anomalous decision making that could lead to collisions, misappropriation, congestion, and other undesirable or even fatal consequences (Gerla and Reiher, 2016). Consequently, it is important to understand the layering of vehicular CPSs, as it provides context for the discussion of autonomy and trust in the following sections as well as providing background to inform future research questions on how a pervasive hub for heterogeneous autonomous vehicle trust authentication could mitigate CPS-based operational risks.

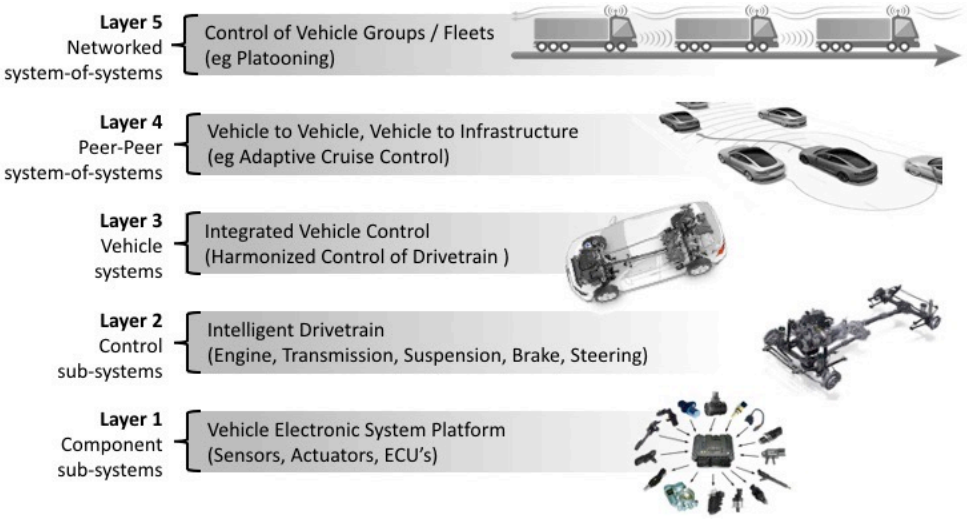

Figure 1: Dependent layers of vehicular cyber-physical functional control systems

2. Vehicles and autonomy

According to researchers of CPS, Artificial Intelligence (AI) and robotics, autonomy is about having the power for self-governance and determination (Antsaklis et al., 1989, Vamvoudakis et al., 2015, Zeigler, 1990). Self-governance is generally seen as enabled by a systems control loop which carries out three core tasks. The first is to sense an environment by measuring a context, the second is to plan by making algorithmic decisions based on the measurements, and the third is to take physical or logical action depending on the result set (Khatib et al., 2016, Shahzad, 2016, Zeigler, 1990). Figure 2 illustrates the CPS control loop which fundamentally supports components, sub-systems, systems and system-of-systems found across each of the CPS layers in Figure 1.

The CPS control loops in the lower two layers contain many specific, discrete primary internal control components running at very high frequency and often in parallel, such as Engine Management, Anti-lock Brake (ABS), and Electronic Stability / Traction Control (ECS/TCS) commonly found in modern motor vehicles. Anderson et al. (2014) note that in this case the planning stage is particularly short and the cycle is more akin to "sense-act". Layer three contains secondary systems, such as Automatic Steering (ASS), Lane Keeping and Adaptive Cruise Control (ACC) used for planning the vehicle's trajectory and decisions such as accelerating and braking. At layers four and five, new autonomous system-of-systems are being developed to plan and act "perfectly" (Anderson et al., 2014) in order to enable capabilities such as automatic driving and applications such as highway platooning discussed in the prior section.

Accordingly, the lower three CPS layers are relatively more reliable as they control discrete CPS functions that have been continuously developed and proven over decades when compared with layers four and five where higher-order system-of-systems are currently in development and yet to be proven as reliable in everyday applications. This illustrates a distinction between the "degree of autonomy" and "levels of autonomy" made by Huang et al. (2003) who points out that total autonomy in a low-level sub-system, such as anti-lock braking does not correspond to system-wide vehicle autonomy. Consequently, while layer one and two CPS components and sub-systems may well be discretely reliable and trustworthy, they form part of a system-wide sophistication found in layers three to five that are yet to prove their trustworthiness. Therefore it is these higher layers that are the focus of future research by the author.

Similar to CPS, autonomous vehicles (AVs) are not a new concept with the English engineer Robert Whitehead developing the first self-propelled naval torpedo in 1866 and by 1897 incorporating a gyroscope to control its course (Rigby, 2017, Yağdereli et al., 2015). The first unmanned land vehicle was a radio-controlled tricycle made by a Spanish inventor, Leonardo Torres-Quevedo in 1904 and experiments of unmanned aircraft were conducted in World War I (Everett, 2015). In 1925 the Hulett Motor Car Company demonstrated a radio controlled car in New York known as the "American Wonder", and self-driving trains have been in everyday use since the 1960's (Yağdereli et al., 2015).

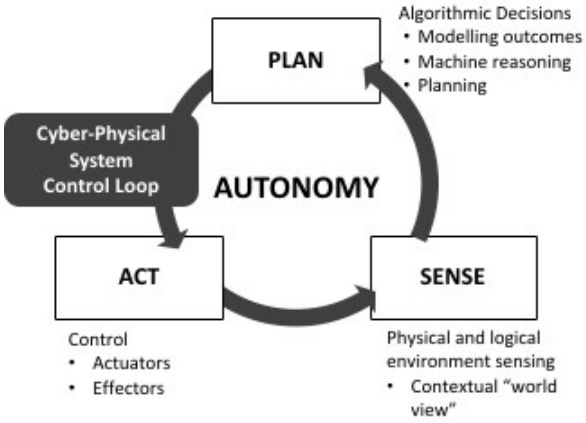

Figure 2: Architecture of autonomy employing a cyber-physical "sense, plan, act" control loop (influenced by Zeigler (1990), Anderson et al. (2014), Shahzad (2016))

Development in AVs over the past 40 years has been stimulated by exponential increases in computer processing technology as well as key initiatives backed by government to encourage further research within the domain (Issitt, 2016, Weber, 2014). Progress has been particularly noticeable in two main categories being autonomous road vehicles (ARVs) and unmanned aerial vehicles (UAVs) commonly known as drones. Advances in road vehicle autonomy has been predicted as being the biggest change since the motor cars replaced the horse and cart (Yağdereli et al., 2015), and in an "ultimate future" less predictable "untrustworthy" human driven vehicles may need to be identified apart from autonomous vehicles (Walsh, 2016).

While there has been intense growth and development of ARV and UAVs over the past decade, industry led standards and government legal frameworks have often been left behind which may delay widespread adoption over the next decade (Anderson et al., 2014, Wise, 2016, Wiggins, 2017). This was exemplified in a recent study by research firm Gartner Group who found that technology failures and security were the key reasons why respondents would not consider riding in a fully autonomous vehicle (Forni, 2017). Consequently, researchers and representative industry organisations have recognised the need for classification schemes to evaluate degrees of ARV and UAV autonomy for purposes of identification, certification and risk assessment.

Driven by the need to understand how independent an UAV is and how to measure it, ten autonomous control levels were proposed by researchers at the U.S. Air Force Research Laboratory's Air Vehicles Directorate in 2002 (Clough, 2002). The taxonomy describes single as well as multi-vehicle coordinated and cooperative levels of autonomy based upon the attributes "Perception/Situational Awareness, Analysis/Decision Making, and Communication/Cooperation" similar to the "Sense, Plan, Act" stages found in the aforementioned CPS control loop (Figure 2). Dalamagkidis (2015) suggests that the levels were drawn up for use in a military rather than commercial context and may not be entirely applicable for the regulation of civilian UAVs. Consequently, he suggests three simplified levels based solely upon the level of human involvement being 1) Remotely piloted, 2) Remotely operated (semiautonomous), and 3) Fully autonomous. However, these simplified definitions do

not easily encompass each of the CPS layers and Huang et al. (2003) broaden the simplified levels from Dalamagkidis (2015) and propose a model where the degree of autonomy for an unmanned system can be characterised by system scope, mission scope and complexity. The levels correspond to layers one to four of CPS ability shown in Figure 1. They conclude that reliance on human decision making and intelligence is higher and uncertainty assumptions are relatively low at lower CPS levels and that uncertainty rises proportionately with the level of mission scope and independence from human decision making.

Table 1: Levels of autonomy in highway vehicles (adopted from Yağdereli et al. (2015))

Level	Autonomy	Description	Example
1	**Driver Only**	*Entirely under human control.* *Some automated systems.*	*Cruise control, electronic stability contro anti-lock brakes.*
2	**Driver assistance**	*Steering and/or acceleration are automated.* *Driver controls other functions.*	*Adaptive cruise control, parking assistant, automated steering.* *Driver controls accelerator and brakes.*
3	**Partial autonomy**	*System controls steering and/or acceleration.* *Driver must take back control when required.*	*Adaptive cruise control with lane keeping.* *Traffic jam assistance.*
4	**High autonomy**	*Vehicle system operate autonomously for some portions of the journey.* *Transfer of control back to driver when warned.*	*Early prototypes of driverless car.* *Human takes control by braking our turning the steering wheel.*
5	**Full autonomy**	*Vehicle system capable of driving unaided for an entire journey with no human intervention.* *Potentially without a human in the car.*	*Driverless cars which have no need for a steering wheel, accelerator or brake pedal.* *100% autonomous.*

The system, scope and complexity approach to classification is complemented by levels two to five described by Yağdereli et al. (2015) which focuses on control by either the human driver or an autonomous driving system (Table 1). In their classification human drivers have full control at level one, which is incrementally ceded to driver assistance systems to reach full vehicle autonomy at the highest fifth level.

Global engineering standards organisation, the Society of Automotive Engineers (SAE) further expands on the delineation between human and systems control in the SAE J3016 standard which describes six levels of driving automation for on-road motor vehicles (SAE International, 2016a). However, the, SAE note that that the taxonomy relates to automation rather than autonomy as Automated Driver Systems (ADSs) do not necessarily have a capacity for self-governance, and that autonomy has varied inter-disciplinary meanings that may be misinterpreted. The SAE standard was adopted by the U.S. Department of Transportation (DOT) National Highway Traffic Safety Administration (NHTSA) in September 2016 within a Federal Automated Vehicles Policy (SAE International, 2016b). Table 2 summarises key elements of the standard.

The classification is based upon three actors being the human driver, the driving automation system, and other vehicle systems and components. It describes levels that perform all, or part of the dynamic driving task (DDT) ranging from level zero where there is no automation through to level five where the DDT is fully automated. Levels zero and one are exemplified by driving assistance systems that we see in everyday vehicles today, with responsibility for driving becoming incrementally automated from level two, through to five as indicated by the dashed arrow in the table showing incremental system control over 1) motion control, 2) object and event detection and response, 3) fall-back recovery, and in 4) an unlimited operational domain (for example the public road network).

What each of the frameworks have in common with the CPS layers, is that as system complexity increases to automate the driving task, the level of human control decreases and, if they are to be trusted, the need for "perfect" performance of systems within a particular contextual environment. This leads to a review of the trustworthiness of autonomous vehicles.

Table 2: SAE J3016 Summary of levels of driving automation (adopted from SAE International (2016a))

Level	Name	Narrative definition	Dynamic Driving Task (DDT)		DDT fallback	Operational Design Domain (ODD)
			Motion Control	Object, Event Detection and Response (OEDR)		
Human driver performs part or all of the Dynamic Driving Task (DDT)						
0	No Driving Automation	The performance by the *driver* of the entire *DDT*, even when enhanced by *active safety systems.*	Driver	Driver	Driver	n/a
1	Driver Assistance	The *sustained* and *ODD*-specific execution by a *driving automation system* of either the *lateral* or the *longitudinal vehicle motion control* subtask of the DDT (but not both simultaneously) with the expectation that the *driver* performs the remainder of the *DDT*.	Driver and System	Driver	Driver	Limited
2	Partial Driving Automation	The *sustained* and *ODD*-specific execution by a *driving automation system* of both the *lateral* and *longitudinal vehicle motion control* subtasks of the *DDT* with the expectation that the *driver* completes the *OEDR* subtask and *supervises* the *driving automation system*	**System**	Driver	Driver	Limited
Advanced Driving System (ADS) performance the entire DDT (while engaged)						
3	Conditional Driving Automation	The *sustained* and *ODD*-specific performance by an *ADS* of the entire DDT with the expectation that the *DDT fallback-read user* is *receptive* to *ADS*-issued *requests to intervene*, as well as to *DDT performance-relevant system failures* in other *vehicle* systems, and will respond appropriately.	System	**System**	Fallback-ready user (becomes the driver during fallback)	Limited
4	High Driving Automation	The *sustained* and *ODD*-specific performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene.*	System	System	**System**	Limited
5	Full Driving Automation	The *sustained* and unconditional (i.e., not *ODD*-specific) performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*.	System	System	System	**Unlimited**

3. Trustworthiness of autonomous vehicles

Trust is an important determinant of public acceptance and the future adoption and the active use of AVs as it stands between common attitudes toward automation and utilisation intentions. Researchers suggest that there are three dimensions to growing trust in functional AV systems (Kyu Choi and Gu Ji, 2015). The first is system transparency, a belief that the system is predictable and understandable; the second is technical competency, that tasks are performed accurately and correctly; and the third is situation management, that control can be recovered should it be necessary to do so. Two aspects that materially affect these dimensions are: 1) the extent of internal human involvement in driving automation tasks, and 2) how the external "world-view" environment is represented and seen by automation systems.

The level of human involvement in the control of an autonomous vehicle affects the dimension of system transparency. When the driving automation schema from the SAE (Table 2), levels of autonomy from Yağdereli et al. (Table 1) and the CPS layers (Figure 1), are compared they show that the role played by humans in driving tasks becomes less concentrated as driving system automation increases. Nothwang et al. (2016) suggests that this can be characterised as humans being either in the automation control loop (HIL), where they actively engage in control decisions to provide partial or conditional automation, on the control loop (HOL), where control decisions are escalated only when required and the human role is supervisory in nature, or finally, completely autonomous where the human plays no part in control loop decision making. Consequently, as shown in Table 3, humans are in total control at autonomy levels zero and one, humans form an active part and are in the control loop (HIL) at levels two and three, are on the supervisory control loop (HOL) at level four, and finally, at level five they are not in the control loop at all.

Table 3: Trust disposition by human control loop involvement at each automated vehicle level

Level	Desc	Human and the Control Loop	Trust Disposition	
			Current State Today	Future State + 15 yrs
0	No Driving Automation	**Human in control**	Medium	Low
1	Driver Assistance	**Human in control**	High	Medium
2	Partial Driving Automation	**Human in the loop**	High	High
3	Conditional Driving Automation	**Human in the loop**	High	High
4	High Driving Automation	**Human on the loop**	Low	High
5	Full Driving Automation	**Human not in control**	Low	Medium

When the extent of human involvement in the autonomous vehicle control loop is applied to current-state (present day) development, the higher levels of automation correlate with lower levels of trust (Nothwang et al., 2016). This is primarily caused by the current absence of long-term field data to inform response predictability at the higher levels. However, as the dashed arrow indicates in Table 3, over the next 15 years these systems will become more established as reliability data from higher adoption rates is analysed and incorporated as refinements into revised operating algorithms. Consequently, as systems development to improve transparency, technical competency, and situation management progresses, it is predicted that future state systems with humans on the loop (HOL) will be perceived as more efficient, reliable and therefore lower risk and more trustworthy than those with humans in total control or in-the-loop (HIL) (Nothwang et al., 2016, Kyu Choi and Gu Ji, 2015). A practical HOL example is current automated air traffic control systems where normal predetermined operation is automated and decisions are only escalated to air traffic controllers when an undefined exception is encountered. Further, some researchers predict a future where unpredictable human controlled vehicles will be seen as less trustworthy when mixed with high populations of predictable networked automated vehicles, and as with learner drivers today, they will need to be identified due to their higher risk profile. (Walsh, 2016).

While minimising human involvement in carrying out driving tasks may have tangible societal benefits such as fewer accidents, injuries and fatalities, lowered congestion, decreased fuel consumption and lower CO2 levels (Anderson et al., 2014, Beck, 2016, Litman, 2014, McDermott, 2016, Vachtsevanos and Valavanis, 2015), it also increases the number and complexity of cyber-physical systems within and across heterogeneous networks of AVs. Such complexity highlights the need for technical competency, the second dimension to grow trust in functional AV systems, by way of strong cyber-security defences to proactively identify and eliminate vulnerabilities to reduce the risk of cyber-attacks. This resilience is particularly required to support sophisticated capabilities found in the higher autonomy levels if they are to be trusted and well adopted (Anderson et al., 2014, Yağdereli et al., 2015).

Table 4: Types of cyber-attacks and possible results

Type of attack	Description	Result	Example outcome
Adv. persistent threat (APT)	Long term covert analysis of weaknesses	Undetected coordinated attack	Covert misuse of system by hackers
DDos	Service request inundation	Poor, no service responsiveness	Complete service disruption
Doxing	Publicly revealing records/documents	Privacy breaches	Reputational damage
Integrity attack	Change internal information	Misleading actions/output	False action/false data output
Malware	Packaged exploit of vulnerability	Incorrect statuses	Corrupted sub-systems
Man-in-the-middle (MITM)	Modified communication link to gain access	Unauthorised access	Control seized by attacker
Password theft	Repeated password attempts (brute force)	Unauthorised access	Control seized by attacker
Phishing	Fooling authorised user to give access	Unauthorised information collection	Information misused / blackmail
Spoofing	Presenting an alternate input to systems	Incorrect inputs	System responds to a false data
Trojan	Malware disguised in a software update	Incorrect statuses	Corrupted systems

Vulnerability risks for AVs must be identified early while systems are in the design and development stages. This relies on adequate policies, procedures and an information security aware culture from development through to manufacture and on-going product support (Da Veiga and Eloff, 2010, Mohan, 2016). Without a robust approach to mitigating vulnerability risks, possible attacks listed in Table 4 threaten the integrity and viability of an autonomous vehicle to such an extent that it may cause harm to both its own logical and physical components as well as its operators, passengers and innocent bystanders (Singer, 2014, Rani et al., 2016).

While each type of attack listed has the potential to undermine the trustworthiness of AVs, techniques such as man-in-the-middle and spoofing are of most concern as they can present an alternate reality to systems and sensors in order to manipulate and change an AVs normally expected behaviour. A Man-in-the-Middle (MITM) attack is a method used to spoof a system through the creation of false communication, in the form of a signal or similar means, surreptitiously sent from an unknown source, disguised as a legitimate and authorised source known to the receiver. The purpose is to provide an alternate reality to the receiver to cause it to act in an abnormal manner (Warner and Johnston, 2003, Chen et al., 2009). Two example cases of MITM spoofing relevant to AVs are the manipulation of GNSS signals and physical system access signals.

Firstly, AVs are highly reliant on location services enabled by the reception of weak GNSS radio signals and the positioning signal transmitted by these services being blocked or jammed by a higher strength signal. A MITM attack therefore involves feeding a GNSS receiver with fake signals so that it believes it is located elsewhere in order to make it act abnormally according to the spoofed alternate reality presented to it. Consequently, an AV may be fed false routing information causing it to be captured by an attacker (Warner and Johnston, 2003) and live cases have involved drug traffickers jamming GNSS signals to confuse and disable UAVs used by U.S. Customs to patrol the USA, Mexican border (Mohan, 2016).

Secondly, ARVs are also particularly open to MITM attacks that compromise their physical access methods. A well-known hack to automakers is where attackers spoof the signal from a wireless car key fob to open the doors of a vehicle. In this case, the signal from a fob held or stored by the owner some distance away (on a bed-side table within their home for example) is intercepted and read by scanning equipment. This signal is then re-transmitted by the scanning equipment to the locking CPS sub-systems of a vehicle parked in a driveway. Using a similar procedure, hackers can also access the ignition system to simply steal the vehicle (Greenberg, 2017).

While understanding the extent of human involvement in AV decision making and addressing AV system vulnerabilities to cyber-attacks is important, they focus discretely on AV functional and technical operation rather than its contextual situation. The third dimension to grow trust in functional AV systems, situation management involves interaction with the external environment including other vehicles, infrastructure, pedestrians and observers. Determining situational trust is therefore the process of ensuring that an AVs past, current and projected location and behaviour is contextually consistent and legitimate given its purpose.

Research reviewed in this paper has primarily focussed on the veracity of V2V and V2I messaging between vehicles and infrastructure to measure the trustworthiness of between AVs within an ad-hoc network of vehicles (VANET) (Gerla and Reiher, 2016). While technical approaches vary widely, the concepts are generally founded on whether individual AV messages agree with other available information sources. Other sources include messages from other vehicles within a VANET and their shared sensor readings, as well as vehicle measurement and control infrastructure and travel information sources. The content can be both real-time and historical so that a current track can be both verified and predicted. Results from processed data across many sources can then be compared and used to validate and report upon anomalous behaviours. For example, road congestion readings using mobile network cell hand-off data can be used to compare an ARVs reported speed on a particular road segment (BITRE, 2014, Bismeyer et al., 2012). On board short range radar found in ARV adaptive cruise control systems can also be used to determine the veracity of individual ARV speed and position messaging as other vehicle “nodes” in the VANET share data between each other (Yan et al., 2007). Some researchers have also extended the concept to include trust rating systems which maintain trust veracity values over time for vehicles they have encountered. In this case, vehicles that send inaccurate or false data would develop a low trustworthiness score over time (Dijiang Huang et al., 2010).

This section on the trustworthiness of autonomous vehicles suggests that a high degree of rigor across the three dimensions of trust is required as levels of automation sophistication increase and the levels of human decision making involvement decrease. However, autonomous systems rely on sensor component and sub-system level CPSs which can be spoofed with an alternate reality by a hacker impacting actual and reputational trustworthiness. A method to mitigate this risk is to authenticate information received from internal vehicle sensors with external information from other vehicles and services certified as trustworthy. This is highly relevant to the authors future work which proposes to focus on the effectiveness of a pervasive trustworthiness hub to authenticate information flowing between heterogeneous autonomous vehicles and infrastructure within their surrounding environment.

Conclusion and future research directions

Reviewing the relationship between cyber-physical systems, autonomous vehicles and their trustworthiness has revealed a number of areas where future research would prove to be beneficial.

- Firstly, current literature mainly focuses on the higher levels of autonomy without segmenting results by the levels of automation. This may be due to accepted autonomy level classification schemas from the likes of SAE being unavailable and adopted only recently by influential organisations such as the U.S. DOT NTSCA. However, practical implementations of SAE level three ARV automation are in production today and sophisticated level four autonomy is already well advanced in trials on public roads. Both levels would benefit from specific research and further analysis against the three dimensions of trust along with a review of other applicable frameworks. Continuous research is also needed to expose AV vulnerabilities for each level of automation and autonomy for both UAVs and ARVs. This will assist with the mitigation of future risks by exposing and eliminating weaknesses before hackers can employ exploits to take advantage of them.
- Secondly, much of the existing AV material is focussed on either UAVs or ARVs, with few articles on how they and other types of AVs could cooperate to authenticate shared information in a mixed vehicle and infrastructure environments. There is some discussion on linking AVs, VANETs and infrastructure to cloud based services within an Internet of Things (IoT), however securing these will offer particular research challenges if cloud based services are used to link them as they have unique features of location mobility and transience.
- Thirdly, the complexity of an AVs sensors and data analysis requires the storage of substantial amounts of information that may have privacy implications. Privacy concerns are also exacerbated when vehicle data is shared between surrounding vehicles and infrastructure or uploaded to networked services that have access to further external contextual information which can be used for situational data correlation, enrichment and interpretation.
- Finally, while peer level vehicle-to-vehicle trust has been considered there was no specific discussion in the materials reviewed for this paper on pervasive methods for real-time trustworthiness assessment and authentication based on data from heterogeneous automated and autonomous vehicles and their surrounding infrastructure. This is an area that can be investigated further to explore situational identification, attribution and responsibility rules for level three and four AVs where humans can be on a supervisory control loop. Such research could examine detection methods of illegitimate behaviours such as unauthorised entry to no-go geo-fenced areas that can then be handed to human controllers on-the-loop for risk assessment.

In summary, this paper has reviewed the relationship between cyber-physical systems, autonomous vehicles and trustworthiness. It has examined the crucial role CPS plays in building sub-system, system, and system-of-system layers found in present day automated level zero through to four vehicles and autonomous level five vehicles. It has described how the CPS control loop fundamentally supports systems autonomy within AVs and the levels of automation classification by leading industry organisations the SAE and USDOT. It recommends that to be successful, AVs will need to adopt frameworks such as the three dimensions of trust being transparency, competence and management, and research should be conducted into the effectiveness of implementing them within a pervasive trustworthiness hub to authenticate information shared between heterogeneous autonomous vehicles and infrastructure within their surrounding environment.

REFERENCES

ANDERSON, J. M., NIDHI, K., STANLEY, K. D., SORENSEN, P., SAMARAS, C. & OLUWATOLA, O. A. 2014. *Autonomous vehicle technology: A guide for policymakers*, Rand Corporation.

ANTSAKLIS, P., PASSINO, K. & WANG, S. 1989. Towards intelligent autonomous control systems: Architecture and fundamental issues. *Theory and Applications,* 1**,** 315-342.

BECK, J. 2016. Tomorrow's Driverless Convoy on the Road Today. *GPS World.* North Coast Media, LLC.

BISMEYER, N., MAUTHOFER, S., BAYAROU, K. M. & KARGL, F. 2012. Assessment of node trustworthiness in VANETs using data plausibility checks with particle filters.

BITRE 2014. New traffic data sources – An overview. *Exploring New Sources of Traffic Data.* Sydney: Bureau of Infrastructure, Transport and Regional Economics.

CARBONE, J. N., EROGLU, A., SUH, S. C. & TANIK, U. J. 2014. *Applied Cyber-Physical Systems*, New York, NY : Springer New York : Imprint: Springer.

CHEN, Y., XU, W., TRAPPE, W. & ZHANG, Y. 2009. Detecting and Localizing Wireless Spoofing Attacks. *In:* CHEN, Y., XU, W., TRAPPE, W. & ZHANG, Y. (eds.) *Securing Emerging*

Wireless Systems: Lower-layer Approaches. Boston, MA: Springer US.

CLOUGH, B. T. 2002. Metrics, schmetrics! How the heck do you determine a UAV's autonomy anyway. AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH.

CRANDALL, R. E. & FORMBY, S. K. 2016. Is that a driverless truck alongside you? *Industrial Engineer: IE.* Institute of Industrial Engineers.

DA VEIGA, A. & ELOFF, J. H. 2010. A framework and assessment instrument for information security culture. *Computers & Security,* 29**,** 196-207.

DALAMAGKIDIS, K. 2015. Classification of UAVs. *In:* VACHTSEVANOS, G. J. & VALAVANIS, K. P. (eds.) *Handbook of Unmanned Aerial Vehicles.* Dordrecht : Springer Netherlands : Imprint: Springer.

DIJIANG HUANG, M., XIAOYAN HONG, M. & GERLA, M. 2010. Situation-aware trust architecture for vehicular networks. *Communications Magazine, IEEE,* 48.

EKEDEBE, N., YU, W. & LU, C. 2016. Securing Transportation Cyber-Physical Systems. *In:* PATHAN, A.-S. K. (ed.) *Securing Cyber-Physical Systems*. Boca Raton: CRC Press.

EVERETT, H. R. 2015. *Unmanned systems of World Wars I and II,* Cambridge, Massachusetts, The MIT Press.

FORNI, A. 2017. Gartner Survey Reveals 55 Percent of Respondents Will Not Ride in a Fully Autonomous Vehicle. *Gartner Newsroom.* Stamford, USA: Gartner.

GÁSPÁR, P., ZSOLT, S. & SZILÁRD, A. 2014. Highly Automated Vehicle Systems. Available: http://www.mogi.bme.hu/TAMOP/jarmurendszerek_iranyitasa_angol/book.html [Accessed 2 Nov 2017].

GERLA, M. & KLEINROCK, L. 2011. Vehicular networks and the future of the mobile internet. *Computer Networks,* 55**,** 457-469.

GERLA, M. & REIHER, P. 2016. Securing the Future Autonomous Vehicle: A Cyber-Physical Systems Approach. *In:* PATHAN, A.-S. K. (ed.) *Securing Cyber-Physical Systems*. Boca Raton: CRC Press.

GREENBERG, A. 2017. Just a Pair of the $11 Radio Gadgets can steal a car. *Wired* [Online]. Available: https://www.wired.com/2017/04/just-pair-11-radio-gadgets-can-steal-car/ [Accessed 2 Nov 2017].

HARTMANN, K. & GILES, K. UAV exploitation: A new domain for cyber power. Cyber Conflict (CyCon), 2016 8th International Conference on, 2016. IEEE, 205-221.

HUANG, H.-M., MESSINA, E. & ALBUS, J. 2003. Toward a generic model for autonomy levels for unmanned systems (ALFUS). NATIONAL INST OF STANDARDS AND TECHNOLOGY GAITHERSBURG MD.

ISSITT, M. 2016. Autonomous Car. Salem Press.

KHATIB, O., SICILIANO, B. & SPRINGERLINK 2016. *Springer Handbook of Robotics*, Cham : Springer International Publishing : Imprint: Springer.

KRÄMER, B. J. 2014. Evolution of Cyber-Physical Systems: A Brief Review. *Applied Cyber-Physical Systems.* New York, NY : Springer New York : Imprint: Springer.

KYU CHOI, J. & GU JI, Y. 2015. Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *International Journal of Human-Computer Interaction*.

LEE, E. A. & SESHIA, S. A. 2014. *Introduction to Embedded Systems: A Cyber-Physical Systems Approach*, MIT Press.

LITMAN, T. 2014. Autonomous vehicle implementation predictions. *Victoria Transport Policy Institute,* 28.

MCDERMOTT, I. E. 2016. The Internet of Cars. *Online Searcher.* Information Today Inc.

MOHAN, M. 2016. Cybersecurity in drones. *In:* COLLGE, U. (ed.).

MULAY, A. 2016. *Sustaining Moore's law : uncertainty leading to a certainty of IoT revolution*, San Rafael, California : Morgan & Claypool Publishers.

NOTHWANG, W. D., MCCOURT, M. J., ROBINSON, R. M., BURDEN, S. A. & CURTIS, J. W. The human should be part of the control loop? Resilience Week (RWS), 2016. IEEE, 214-220.

PATHAN, A.-S. K. 2016. *Securing Cyber-Physical Systems,* Boca Raton, CRC Press.

RANI, C., MODARES, H., SRIRAM, R., MIKULSKI, D. & LEWIS, F. L. 2016. Security of unmanned aerial vehicle systems against cyber-physical attacks. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology,* 13**,** 331-342.

RIGBY, R. 2017. *The Whitehead Torpedo (c. 1866 -)* [Online]. Melbourne, Australia: University of Melbourne. Available: http://www.torp.esrc.unimelb.edu.au/biogs/E000001b.htm [Accessed 2 Nov 2017].

SAE INTERNATIONAL 2016a. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *Surfaced Vehicle Recommended Practice.* Warrendale PA: SAE International.

SAE INTERNATIONAL 2016b. U.S. Department of Transportation's New Policy on Automated Vehicles Adopts SAE International's Levels of Automation for Defining Driving Automation in On-Road Motor Vehicles. Warrendale PA: SAE International.

SHAHZAD, K. 2016. Cloud Robotics and Autonomous Vehicles. *In:* ZAK, A. (ed.) *Autonomous Vehicle.* Rijeka: InTech.

SINGER, P. W. 2014. Cybersecurity and cyberwar : what everyone needs to know. *In:* FRIEDMAN, A. (ed.). New York, NY : Oxford University Press.

STOJMENOVIC, I. & ZHANG, F. 2015. Inaugural issue of 'Cyber-physical systems'. *Cyber-physical systems.,* 1.

VACHTSEVANOS, G. J. & VALAVANIS, K. P. 2015. *Handbook of Unmanned Aerial Vehicles*, Dordrecht : Springer Netherlands : Imprint: Springer.

VAMVOUDAKIS, K. G., ANTSAKLIS, P. J., DIXON, W. E., HESPANHA, J. P., LEWIS, F. L., MODARES, H. & KIUMARSI, B. Autonomy and machine intelligence in complex systems: A tutorial. 2015 American Control Conference (ACC), 1-3 July 2015. 5062-5079.

WALSH, T. 2016. Turing's Red Flag. *Communications of the ACM.* Association for Computing Machinery.

WARNER, J. S. & JOHNSTON, R. G. 2003. GPS spoofing countermeasures. *Homeland Security Journal,* 25**,** 19-27.

WEBER, M. 2014. Where to? A History of Autonomous Vehicles. Available from: http://www.computerhistory.org/atchm/where-to-a-history-of-autonomous-vehicles/ [Accessed 2 Nov 2017].

WIENER, N. 1948. *Cybernetics or Control and Communication in the Animal and the Machine*, MIT press.

WIGGINS, J. 2017. Australia 'unprepared' for arrival of driverless cars. *Australian Financial Review*, 24 Nov 2017, p.2.

WISE, D. J. 2016. VEHICLE CYBERSECURITY DOT and Industry Have Efforts Under Way, but DOT Needs to Define Its Role in Responding to a Real-world Attack. *GAO Reports.* U.S. Government Accountability Office.

YAĞDERELI, E., GEMCI, C. & AKTAŞ, A. Z. 2015. A study on cyber-security of autonomous and unmanned vehicles. *Journal of Defense Modeling and Simulation,* 12**,** 369-381.

YAN, G., CHOUDHARY, G., WEIGLE, M. & OLARIU, S. 2007. Providing VANET security through active position detection.

ZEIGLER, B. P. 1990. High autonomy systems: concepts and models.