Recent Advances and Trends in On-board Embedded and Networked Automotive Systems

Lucia Lo Bello, Senior Member, IEEE, Riccardo Mariani, Member, IEEE, Saad Mubeen, Senior Member, IEEE, and Sergio Saponara, Senior Member, IEEE

Abstract—Modern cars consist of a number of complex embedded and networked systems with steadily increasing requirements in terms of processing and communication resources. Novel automotive applications, such as, automated driving, rise new needs and novel design challenges that cover a broad range of hardware/software engineering aspects. In this context, this paper provides an overview of the current technological challenges in on-board and networked automotive systems. The paper encompasses both the state-of-the-art design strategies and the upcoming hardware/software solutions for the next generation of automotive systems, with a special focus on embedded and networked technologies. In particular, the work surveys current solutions and future trends on models and languages for automotive software development, on-board computational platforms, in-car network architectures and communication protocols, and novel design strategies for cybersecurity and functional safety.


I. INTRODUCTION

The size of embedded systems market is growing at a drastic pace. According to an estimate, it will be 258.72 billion USD by 2023 [1]. It is further estimated that automotive applications comprise of over 20% of this market. An embedded system consists of hardware (HW), processor and peripherals, and software (SW) that runs on the embedded processor [2]. In a modern car, the size of the embedded SW is in the order of millions of code lines. Many automotive embedded systems are real-time (RT) constrained, i.e., they must provide logically correct responses at correct times that are dictated by time-critical functionalities. Particularly, hard RT requirements apply to Autonomous Driving (AD) or Autonomous Machinery Operation, according to a pre-planned path/statement-of-work. Such functionalities are demanding in terms of both computational and environmental conditions. Challenging environmental requirements [3] have to be faced, such as temperature from -40°C to 125°C, mechanical and chemical stress and moisture resistance over 15-year lifetime, and electrostatic discharge (ESD) protection of several kV.

The current electric/electronic (E/E) on-board automotive architectures, with up to 100 dedicated Electronic Control Units (ECUs), is no longer capable of answering the computing, communication and memory requirements coming from innovations with increased needs of fail-operational, functional safety (FuSa), cyber security and RT behavior [4]–[7]. Such innovations include: transition from internal combustion engines to full electric cars; introduction of advanced driver assistance systems (ADAS) and AD functions; increased level of on-board connectivity, mainly wired (e.g. FlexRay [8], CAN/CAN-FD [9], [10], Automotive Ethernet [11]); vehicle to everything (V2X) wireless connectivity for advanced services such as fleet management, platooning, over-the-air SW updates; stringent constraints in terms of FuSa and cyber security.

Fig. 1. Computation needs vs. AD/ADAS functions [4]

Rather than just increasing the number of basic ECUs, using 32-bit microcontrollers (MCUs), new on-board E/E-architectures will exploit embedded High Performance Computing (eHPC) platforms. Computational power in the order of Tera Operations per Second (TOPS), see Fig. 1, is needed to implement RT perception and AD tasks in modern vehicles, particularly for high automation (Level 4, L4), in which the system can perform the driving task without human intervention, and full automation (Level 5, L5), in which the system takes over all the aspects of driving full time. As shown in Fig. 2, the automotive eHPC should sustain in RT the following functions:

1) Observation: building a model of the surrounding environment, where inputs are the direct observations produced by sensors [12] (cameras, radars, sonars, lidars) or V2X wireless data, and output is a geometrical and topological representation of the environment.

2) Perception: localization of the car, i.e. estimating its path, position and orientation within a map, by fusion of global
trends and solutions for RT in-vehicle communications are presented in Section IV. Cybersecurity issues related to in-vehicle networking and possible countermeasures are analyzed in Section V, while design strategies for on-board functional safety are dealt with in Section VI. Finally, conclusions are drawn in Section VII.

II. AUTOMOTIVE SOFTWARE DEVELOPMENT

Automotive industry has undergone a drastic shift from mechanic-intensive to SW-intensive applications in the last couple of decades [29]. According to [30], more than 80% of innovation in cars come from computer-controlled functionalities. The increasing demand for such functionalities and data-intensive applications in modern cars has led to the increasing size and complexity of automotive SW. According to an estimate in 2014, the amount of SW in a regular four-door car increased ten times in eight years reaching the size of approximately 1 GB [31]. Another estimate in 2009 predicted that a modern premium car shall contain nearly 100 million lines of code (MLoC) and was expected to reach 200-300 MLoC in the coming years. This estimate seems accurate, as Ford showcased their car containing 150 MLoC at the Consumer Electronics Show (CES) in 2016 [32].

Model-based Engineering (MBE) [33] and Component-based SW Engineering (CBSE) [34] have emerged as an attractive and cost-effective option to deal with the size and complexity of the SW. MBE uses models to describe functions, structures and design artifacts throughout the SW development. CBSE allows to build large SW systems by reusing pre-existing SW components and their architectures. It is estimated that up to 90% of automotive SW can be reused from previous releases or other projects if MBE and CBSE are used [35]. There exist several modeling languages and component models in the automotive domain that employ the principles of MBE and CBSE for the SW development [36], [37].

EAST-ADL [38] is an architecture description language for automotive embedded systems. It has developed and evolved based on several European projects and research works such as [39]–[42]. The EAST-ADL methodology allows to model the SW architecture at four abstraction levels. These levels, together with some of the models, languages, and tools that are used for the SW development at each level, are depicted in Fig. 3. At the top level, called the Vehicle Level, end-to-end requirements on the automotive functionality are captured. At the Analysis Level, the requirements are refined and expressed formally. Moreover, several different analyses can be performed including the requirements consistency analysis and the functions analysis. The Design Level defines the SW architecture, HW architecture and SW to HW allocation model by abstracting implementation details. The concrete implementation of SW architecture is performed at the implementation level. The language supports the modeling of automotive SW architecture only at the top three levels in Fig. 3. The methodology recommends to employ standard or proprietary architectures and component models at the Implementation Level, e.g., AUTOSAR [43] and Rubus Component Model (RCM) [44], [45]. EAST-ADL is also aligned with the PuSa standard for road vehicles, ISO26262 [46], [47]. There are several variants and specific implementations of the language that are currently used in the automotive industry, e.g., Systemweaver and its variants SE Tool, Rubus-EAST, Fraunhofer ESK [48], that are also shown in Fig. 3. A detailed comparison of these tools and models is discussed in [49], [50].
There are several middleware approaches and component models that are used for the development of automotive SW at the implementation level. For example, CORBA [51] and iLAND [52]. COMDES [53] and ProCOM [54] represent examples of component models from academia, whereas RCM and AUTOSAR are the examples of industrial component models. AUTOSAR is a worldwide development partnership of vehicle manufacturers, suppliers and companies from the electronics and ICT industry [43]. AUTOSAR-based SW is widely used by all Original Equipment Manufacturers (OEMs) in Europe and is gaining momentum in the U.S., Japan and Korea. Initial version of AUTOSAR did not account for modeling timing information, which is of utmost importance in vehicular safety-critical systems. The support for timing modeling in AUTOSAR is provided by the TADL [55] and TADL2 [56] languages. These languages were developed within two large EU initiatives, i.e., TIMMO and TIMMO2USE [57]. The AUTOSAR standard comprises a way to define the in-car network infrastructure and communication matrix, the necessary exchange formats as well as an operating system infrastructure for embedded ECUs (Classic Platform) and performance ECUs (Adaptive Platform). To support automotive requirements, the SW environment and development kits must provide these functionalities, see Fig. 4.

![Fig. 4. AD architecture with AUTOSAR Platform.](image)

To develop SW for the eHCP platform, a SW development kit together with well-defined exchange formats has to be provided. AUTOSAR ARXML, as the industry-standard format for exchanging information about ECUs, ECU communication, integrated self-describing SW-services and SW-components, makes the system complexity manageable. In the future, the term “system” in automotive will be redefined from single ECUs up to complete cars, and even extended to fleets.

Prototyping environments for AD development extend the Automotive eHPC SW environment into a production environment for AD. Examples of such prototyping environments include Robot Operating System (ROS) 2.0 from Open Source Robotics Foundation and EB robinos, which is a SW framework and architecture for highly automated driving based on open interfaces implementing the Open robinos specifications. Optimized application libraries should be provided for use by the perception tasks, sensor fusion and situative behavior analysis. Lidar sensor processing, which involves representation of data and segmentation into objects, requires efficient implementations of the PCL (Point Cloud Library) and FLANN (Fast Library for Approximate Nearest Neighbors). Camera processing implies a variety of computer visions that are prototyped with the OpenCV library and moved to the OpenVX programming environment to meet the performance requirements. Sensor fusion and other high-performance functions of computer vision are implemented in OpenCL when CUDA is not available. Dense linear algebra libraries such as BLAS/BLIS and Eigen (C++ templated library) must be available and optimized, as they are required by machine learning algorithms and standard deep learning frameworks (e.g., Caffe and TensorFlow).

### III. High Performance Computation Platforms

Today’s embedded automotive-qualified processors, with capabilities of hundreds of MOPS, see Fig. 1, can not handle AD
functions. There is a need for more powerful HW platforms such as eHPC data fusion platform, see Fig 5, which are designed by combining an automotive-certified real-time MCU with general purpose HPC processors and accelerators. The latter are used to increase the power efficiency and to act as safe number crunchers with direct access (not shown in Fig 5) to sensor data through Ethernet or Low Voltage Differential Signaling (LVDS). A multi-Gbps TSN link should be used to connect the safe MCU supervisor, the accelerators and the general purpose HPC processors. This type of interaction will require reliable and secure communication channels, proper identity management and assurance, while providing adequate data and identity privacy. Next-generation AD systems require that the whole perception process be qualified at ASIL-D level according to the ISO 26262 automotive FuSa standard [47]. This can be achieved by performing redundant computations with possibly dissimilar implementation techniques on the “safe number crunchers”. These safe number crunchers are qualified at ASIL-B by implementing a range of safety mechanisms such as Error Detection and Correction Codes (EDC/ECC) in memory, parity in caches, CRC in network-on-chip (NoC). The redundant results are then compared by the safe MCU qualified for ASIL-D, which monitors the computations and decides whether the results can be trusted. The eHPC platform will be connected to the car backbone with a run-time environment compliant with AUTOSAR. For the safe MCU, already available 32-bit cores, like Infineon Aurix or ST SP5/Freescale MPC56, can be adopted. The ST SP5 and Freescale MPC56 families are built in 40 nm technology on 32-bit PowerPC instruction set, with up to 4 cores (with dual lock-step approach), single instruction multiple data (SIMD) floating point unit, 8 MB of embedded flash, multi-channel 12-bit analogue to digital converter (ADC), interfacing data-rate up to 10 Mbps with FlexRay, I2C, LIN [58], CAN, SPI [59]. Similarly, the Infineon Aurix ranges from a 300 MHz triple-core device with 720 MIPS and 8 MB of embedded Flash down to an 80 MHz single-core with 130 MIPS and 0.5 MB embedded flash.

Fig. 5. Automotive eHPC platform.

Instead, for the HPC units in Fig 5, massively parallel platforms are appearing in the car market. Mass production of the Renesas R-Car H3 in 16 nm technology is expected in 2018 [60]. R-Car H3 includes 9 ARM Cortex cores (8 64-bit A57/A53 engines with L1/L2 cache plus a 32b R7 with L1 cache), offering 40k MIPS plus a PowerVR GX6650 graphics engine with 192 ALU cores for 3D graphics (more than 100 GOPS and 4K video display/streaming) and dedicated video co-processors (H.26x/MPEGx codec, distortion compensator, IMP-X5 image recognition). The R-Car H3 is ASIL-B and has a rich set of high-rate interfaces such as Ethernet, USB, DVD/blue-ray SATA, SD card, Audio/video I/O, besides I2C and CAN. The power consumption amounts to tens of Watts. New actors are entering this application domain like Nvidia and Intel, to bring on-board TOPS capability and AI technologies with multi-chip automotive supercomputers. They have signed core partnerships with mass market car makers, Nvidia with Audi, Intel with BMW, to have on the roads AI cars by 2020. To this aim, a new European Processor Initiative for embedded HPC in autonomous driving has recently been started [61]. NVIDIA has recently presented the Xavier AI car’s computer, which features 30 TOPS capability for a power consumption of 30 W, thanks to 8 ARM 64-bit cores plus a 512-core Volta GPU, a Video Processing Unit supporting 8K video decode and encode and High Dynamic Range, as well as a computer vision accelerator. The Xavier AI is fabricated in 16 nm TSMC FinFET technology with an estimated complexity of 7 billion transistors. The power consumption of such HPC platforms will be in the range from tens to hundreds of Watts, e.g. from 30 W of Xavier chip to 500W of the 320 TOPS Drive PX Pegasus board announced at GTC Europe 2017 [62]. Due to the high environmental temperature of under-the-hood car electronics, passive cooling systems are not enough. Hence, the design of low-cost/low-size active cooling systems for HPC ECUs is a new emerging challenge.

Also Intel is developing several platforms for automated driving: a first multi-chip platform, so called Intel Go, has been made available using an Aurix ASIL-D 32-bit MCU enhanced for computation capability by an ATOM C3000 core in 14 nm technology, and by Arria 10 FPGA [3]. The FPGA accelerator includes an embedded dual-core 1.5 GHz ARM A9 core, more than 1M logic elements (a 64-bit 6-LUT with 4 FFs at the output), and 1.7M user flip-flops, and 64 MB of embedded memory. The Arria 10 family includes hardened single-precision IEEE 754 floating point units, with an aggregate throughput of 1.3 TOPS. This platform supports driving automation L3, in which the system performs the driving task, but a human driver will intervene when requested. The next evolution of the platform, suitable for all AD levels, will combine one or more EyeQ5 [63] accelerators (by Mobileye, an Intel company) and one or more ATOM-based general-purpose processors (e.g. Denverton). The EyeQ5 will enable processing of more than sixteen multi-mega-pixel cameras and other sensors. Its computational power targets 15 TOPS, while drawing only 5-6 Watts in a typical application. It implements high performance NoC interconnect and multi-channel low power DDR interfaces, to support high computational and data bandwidth requirements. Another INTEL platform is announced that will use powerful Xeon processors and two multi-chip boards connected with a 16-port 10 GB Ethernet to sustain L4 and L5 AD levels, mainly targeting fleets.

According to the scheme in Fig. 5, high-performance functions that need time-predictability, such as perception functions in automated cars, need to be implemented on high-performance accelerators that also provide response-time guarantees. Time-predictability capabilities start with the core, then the local memory hierarchy, then the global interconnect, and finally external memory and I/O interfaces. In the Intel Go proposal the accelerator role is managed through an Arria10 FPGA or EYEQ5 chip. As alternative, the architecture extensions of the RISC-V accelerator cores have already proven to be suitable for scalable computing capabilities with high power
efficiency [64], including also machine-learning tasks [65].
RISC-V is developed according to an open HW-SW model, thus easing interoperability of eHPC solutions.

The accelerator should ensure timing compositional property, which means that any global worst-case execution time (WCET) is composed of local WCETs. This also implies that the WCET in a core experiencing resources conflicts, e.g., accesses to the memory hierarchy, is safely approximated by adding the resource interference times to the WCET on the core executing without interferences [66]. The timing compositional property requires in-order instruction pipeline and is compatible with caches, provided they have a LRU replacement policy. The basis for the RT accelerator architecture in Fig. 5 can be a Very-Long Instruction Word (VLIW) extension of the RISC-V ISA. VLIW execution, opposed to superscalar execution, is a core implementation technique that enables multiple instruction issues, while being compatible with the timing compositional property. This VLIW extension approach ensures that any standard RISC-V binary will execute correctly, but in single-issue mode on such a VLIW core. A simple recompilation will enable to achieve multiple instruction issues on this core. In both the accelerator and the general-purpose HPC architecture in Fig. 5, a NoC interconnect is responsible for arbitrating access to shared resources, such as an I/O or a memory. One main issue when using multi-core or many-core architectures for designing safety critical systems is to master the impact of contentions that can arise due to parallel requests for a shared resource, on the estimation of the WCET of tasks. Current approaches either rely on a HW approach, for instance Time Division Multiple Access (TDMA), to ensure no contention can arise at runtime, or on SW approaches through the use of specific execution models, such as PRedictable Execution Model (PREM) that explicitly separates data accesses from computation. The need to integrate functionalities with different level of criticalities on such multi or many-core architectures has led to the design of mixed-criticality systems. Extension of these approaches to such mixed-criticality systems is currently based on a technique that drops non-critical tasks whenever a given threshold contention level has been reached. However, more flexible strategies are required at the interconnect level to maximize the utilization level of such multi or many-core platforms. At the multi-core level, the introduction of a HW contention manager to monitor the slack activity at the interconnect level will improve the system capability to allocate resources to non-critical tasks and adapt the scheduling of requests to shared resources, such that critical tasks still meet their deadline while the number of requests from non-critical tasks is maximized.

At the many-core level, a NoC is used to interconnect cores or tiles. NoC is often designed for a given type of application and specific characteristics when targeting RT systems should be developed. Experimentally checking the behavior of a NoC in case of contention between flows is still an open topic. Designing a way to execute routers of a NoC in which a stream would systematically compete with other flows would facilitate the observation of contentions within a NoC. The HW mechanisms for regulating streams in contention could then be enriched to interface with the system SW, in order to dynamically adapt the control performed vs. the target latency.

The eHPC platform in Fig. 5 should be equipped with HW resources to sustain V2X connectivity needed for real-time HD map download and infotainment, over-the-air diagnostic and SW update, sensor-data upload from the vehicle for machine learning. To this aim two solutions can be adopted [5], [6]: IEEE 802.11p or Cellular-V2X. IEEE 802.11p uses 10-MHz channels within the (5.85-5.925 GHz) band to achieve data rates of several Mbps for V2X. IEEE 802.11p transceivers are already available on the market (e.g. STM-Autotalks chipset) and, as discussed in [5], they can be implemented at low-cost in mature and already automotive-qualified CMOS technologies. With 33 dBm of effective isotropic radiated power, a single-hop connectivity of 1 km can be achieved. Cellular-V2X connectivity can be achieved with multilayer MIMO transmission according to emerging 5G transceivers. Operating both in sub-6 GHz and 28 GHz millimeter wave (mmW) bands, data rates of up to several Gbps can be achieved [67]. However, high-end technology nodes are required to sustain mmW and massive MIMO 5G operations and the way to achieve low-latency guaranteed performance is still an open issue. A first 5G modem has been announced by Intel at CES 2017 [68], although its automotive qualification is still ongoing and the 5G standardization is still not settled.

IV. IN-VEHICLE NETWORK ARCHITECTURES

Vehicles are becoming increasingly smart, connected and part of the Internet. While new functionalities such as natural speech recognition and cloud-based services are developed, in-vehicle legacy systems have to be maintained and integrated with the new developments for the sake of cost-effectiveness. As a consequence, traditional signal-based communication, mainly consisting of cyclic message broadcasts, like in LIN, CAN/CAN-FD [69], [70], FlexRay, has to co-exist with service-based communication, made up of event-based message unicasts, such as the ones typical of IP-based networks (e.g., Ethernet, Wi-Fi) [71]. SOME/IP (Scalable service-Oriented Middleware over IP) allows the introduction of service-oriented transmission of information, in which a sender only transmits data when at least one receiver in the network needs this data, thus avoiding to load the network and all connected nodes with unnecessary traffic.

Dynamic distribution of functions, virtualization of ECUs and the network controlled by Virtual Machine and network hypervisor are in the roadmap of future automotive network architectures, which today are migrating from the current central-gateway structure to a domain-based architecture. The vehicle E/E-architecture of tomorrow will be therefore characterized by an automotive Ethernet backbone connecting different domains, isolated and protected by domain controllers [72]. The central Ethernet switch will be also connected to a smart antenna, being LTE/5G, WiFi/BLE, V2X/DSRC the most likely technologies. Following the development of the IEEE standards within the TSN Working Group, the vision for the automotive Ethernet backbone connecting different domains, each with its TSN control unit, provides for master MCUs integrating Ethernet PHYs and TSN switch functionalities with security modules and various protocol converters for local legacy serial networks.

Ethernet switches implement separate collision domains and offer several features that can be used to increase security: VLANs, unicast filtering, multicast filtering and access control lists. However, many state-of-the-art attacks from the Information Technology (IT) world can be applied to in-vehicle Ethernet, so special care must be taken and multiple
levels of defense, should be in place. For instance, performing deep packet inspection in the switches represents an efficient solution to avoid forwarding malicious packets to the host controller, that would be therefore entrusted with security checks only on a second stage of inspection, that would be required for specific frames only, e.g., those coming from the external of the vehicle.

The automotive Ethernet backbone will likely be a TSN-enabled implementation of 802.3 Ethernet. The recent standards 100BASE-T1 [73] and Gigabit PHY (IEEE 802.3bp-2016) [74], already allow the use of a light unshielded twisted pair of copper wires for automotive usage. Also, a broad spectrum of bitrates are envisaged for automotive Ethernet nowadays, also including 10 Mbps and 2.5, 5 and 10 Gbps (for the backbone and for raw sensor data transmission). For Multi-Gig Automotive Ethernet PHY, various options, from shielded cables to coax to optical fiber (for 10 Gbps) are under consideration.

The recent Layer 2 TSN standards are expected to be dominating the scene for autonomous driving. In fact, although current ADAS systems already require processing of high-resolution data originating from video cameras, radars and lidars, self-driving cars require a significantly higher number of sensors, more network connections and better networking solutions for video links than the current technologies based on point-to-point connections, that will not be able to support the packet-based data transport needs of self-driving cars.

A. Automotive Ethernet from AVB to TSN

The IEEE 802.1 AVB is a set of technical standards that provides the specifications for time-synchronized low-latency streaming services through IEEE 802.1Q [75] networks. The AVB documents include: the IEEE 802.1AS-2011 [76] - Timing and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks, (whose revision is in progress as IEEE P802.1AS-Rev project); the IEEE 802.1Qav-2009, Forwarding and Queueing Enhancements for Time-Sensitive Streams, which specifies the Credit-Based Shaper (CBS); the IEEE 802.1Qat-2010 and Stream Reservation Protocol (SRP). The last two amendments have been rolled into the IEEE 802.1Q-2014 standard [75]. Finally, the IEEE Std 802.1BA-2009 specifies a set of usage-specific profiles to help interoperability between networked devices using the AVB specifications. AVB introduces a number of new and important concepts to IEEE 802.1 networks to provide Quality of Service. The first is the support for priority, to distinguish between time-sensitive flows and ordinary traffic and handle them differently. The second is bandwidth reservation, to set aside a certain amount of guaranteed bandwidth across a portion of the network for handling the high-priority traffic. Last but not least, AVB provides a set of protocols to manage the network time for supporting synchronized operations (i.e., A/V playback). For seven hops within the network, AVB guarantees a fixed upper bound for latency. In particular, two Stream Reservation (SR) classes are defined, i.e., Class A, that provides a maximum latency of 2ms and Class B, that provides a maximum latency of 50ms. With AVB, the IEEE has moved Ethernet into the real-time applications domain. AVB is expected to replace (or is already gradually replacing) the Media Oriented Systems Transport (MOST) protocol [77] in the multimedia/infotainment domain, and the LVDS cables in camera-based ADAS. The paper [78] deals with the Credit Based Shaper of AVB and the use of priorities as defined in IEEE 802.1Q in automotive cases studies. The AVB suitability for automotive usage is addressed in [79] and [80]. In particular, [79] provides a comparative performance evaluation of AVB and TTEthernet, a well-known technology standardized by SAE (Society of Automotive Engineers) as AS6802 [81], for ADAS, multimedia and infotainment traffic. The comparison was obtained through OMNeT++ simulations based on realistic traffic patterns on star-based networks under a high and varying workload. Results show that both AVB and TTEthernet meet the requirements of ADAS and multimedia flows. The two technologies complement each other, as TTEthernet allows for completely deterministic transmission and offline verification of time-triggered messages for safety-critical applications, while AVB allows for online stream reservation, thus fitting entertainment applications with varying bandwidth demand. The problem of routing AVB streams to minimize their worst-case end-to-end delay is addressed in [82], which proposes an effective solution, based on a search-space reduction technique and a Greedy Randomized Adaptive Search Procedure (GRASP)-based heuristic.

Despite its advantages, AVB does not provide support for scheduled traffic, i.e., high-priority small-size time-sensitive traffic (e.g., control traffic) that has to be transmitted according to a time schedule without interference from other traffic. In fact, as AVB provides only two real-time traffic classes, a mutual interference problem raises if multiple time-critical traffic flows in the same network are mapped on the same SR Class, with non-negligible effects on delay. Moreover, SR Class frames undergo the CBS algorithm, and shaping blocks frame transmission for a given class if the credit of the class is below zero. For this reason, a more effective way of handling scheduled traffic in AVB networks was proposed in [83] and [84]. The new approach, called AVB_ST, adds a new, separate traffic class on top of the AVB SR Classes A and B, which is called the Scheduled Traffic (ST) Class. ST frames are tagged with the highest priority TAG according to the IEEE 802.1Q standard, while SR Classes A and B take the second and the third highest priority, respectively. ST traffic is handled in a separate queue and does not undergo credit-based shaping, thus avoiding the undesirable effects of shaping on the flow latency. SR Class A and B are handled by CBS, while best-effort traffic by strict priority. Comparative performance assessments between standard AVB and AVB_ST in a realistic automotive scenario in [84] showed that the AVB_ST is able to support scheduled traffic, offering low and predictable latency values without significantly affecting SR traffic. This is thanks to: the introduction of a separate class for scheduled traffic combined with offset-based scheduling for ST flows; the temporal isolation provided by the Time-Aware Shaper mechanism; strict priority scheduling, that offers low and bounded latencies to scheduled traffic even under a high SR traffic load. A response time analysis for multi-hop AVB ST networks that is also applicable to multi-hop AVB networks is presented in [85]. The analysis uses a bandwidth over-reservation concept and overcomes the limitations of previous analysis approaches for AVB networks [86], [87], which, in most of the cases, do not lead to a schedulable result due to
the tight bandwidth allocation imposed by the AVB standard. AVB_ST is similar to the IEEE 802.1Qbv-2015 standard [88], with differences in the way that the ST window is sized and for the rate at which the increase of one of the CBS parameters (i.e., the Idleslope) for the SR Classes is determined.

B. Time-Sensitive Networking

The TSN standard family provides precise time synchronization, deterministic communications, ultra-low latency, zero congestion loss, reliability, and fault-tolerance. These properties are foundational for the next generation of AD vehicles. TSN offers other notable advantages. One is the ability to support both real-time and best-effort traffic over the same network in a flexible way. Changes in the time-critical flows can be accommodated without the need for offline reconfigurations and best-effort traffic can use any bandwidth left over by TSN flows. In addition, TSN offers fast startup, thanks to pre-configured values for timing and bandwidth reservation, and faster firmware updates time than other protocols (e.g., CAN), thanks to the higher datarate. Table I summarizes the TSN standards published so far and some of the ongoing projects that are relevant to automotive applications. A brief description of each of them is provided in the following.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Title</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>P802.1AS-Rev</td>
<td>Robust time synchronization</td>
<td>In progress</td>
</tr>
<tr>
<td>802.1Qbu-2016</td>
<td>Frame Preemption (amendment to 802.1Q)</td>
<td>Published</td>
</tr>
<tr>
<td>802.1Qbv-2015</td>
<td>Enhancements for Scheduled Traffic (amendment to 802.1Q)</td>
<td>Published</td>
</tr>
<tr>
<td>802.1Qca-2015</td>
<td>Path Control and Reservation (amendment to 802.1Q)</td>
<td>Published</td>
</tr>
<tr>
<td>802.1CB-2017</td>
<td>Frame Replication and Elimination for Reliability</td>
<td>Published</td>
</tr>
<tr>
<td>802.1Qcc</td>
<td>Stream Reservation Protocol (SRP) Enhancements and Performance Improvements (amendment to 802.1Q)</td>
<td>In progress</td>
</tr>
<tr>
<td>802.1Qcb-2017</td>
<td>Cyclic Queuing and Forwarding (amendment to 802.1Q)</td>
<td>Published</td>
</tr>
<tr>
<td>802.1Qcr-2017</td>
<td>Per-Stream Filtering and Policing (amendment to 802.1Q)</td>
<td>Published</td>
</tr>
<tr>
<td>P802.1Qcr</td>
<td>Asynchronous Traffic Shaping (amendment to 802.1Q)</td>
<td>In progress</td>
</tr>
</tbody>
</table>

V. CYBERSECURITY ISSUES AND COUNTERMEASURES FOR IN-VEHICLE NETWORKING

Since this survey addresses on-board embedded and networked automotive systems, this Section is focused on cybersecurity issues and countermeasures for in-vehicle networks. Other cybersecurity aspects, such as those related to cars external connectivity, cloud-based traffic and fleet managements, just to name a few, are out of scope of this work.

Secure by design in-vehicle networking should ensure several properties, such as data integrity, confidentiality, authentication, and availability. However, several security vulnerabilities [101]–[108] characterize current in-vehicle networking technologies, using CAN and/or CAN-FD as a backbone, and
a plethora of other interconnecting technologies for specific subsystems (e.g., LIN for local interconnection of low data-rate nodes, MOST for infotainment with USB and Bluetooth user interfaces, and FlexRay for latency-critical functions).

The net-spanning data exchange via various gateway devices potentially allows access to any vehicular bus from every other existing bus system. In principle, each LIN, CAN or MOST controller is able to send messages to any other existing car controller [109], [110]. Without particular preventive measures, a single compromised bus system endangers the whole vehicle communication network. Whereas attacks on LIN or multimedia networks may result in the failure of power windows or navigation software, successful attacks on CAN or FlexRay networks may result in malfunctioning of some important driving assistance functions, which leads to serious impairments of driving safety [111], [112].

While the use of Cyclic Redundancy Check (CRC) ensures data integrity, the broadcast nature of CAN/CAN-FD or FlexRay is a risk in terms of confidentiality, as an attacked ECU can monitor all data passing on the bus. Moreover, since new ECUs can be added in a plug-and-play way (assigning them a new identifier) without modifying the already installed ECUs, and since the data link layer does not provide any signature mechanism, there is a high risk of authentication vulnerability. Similarly, the multi-master feature with an arbitration based on identifier priority poses risks in terms of availability. For example a hacker can attack a bus and behave as a new ECU, reading all data on the bus and generating false packets. Using a high priority identifier, the malicious ECU can win the arbitration and then continuously send invalid messages thus making a jamming attack. Even though these invalid frames will be discarded by the receiving controllers, the attack makes the bus unavailable to other ECUs connected to the bus. Denial of service attacks may affect the backbone bus or the local bus. In the first case, they will lead to system failure, in the second case they will lead to functional failure. The malicious ECU, after reading a message from the bus, can also impersonate another ECU for replay attacks, with a potential for harmful consequences for the vehicle occupant.

Due to the lack of signature mechanisms for authenticity and transmission encryption, it is easy for an attacker to emulate a protocol-compliant behavior. As a consequence, controllers are not able to verify whether an incoming message comes from an authorized or unauthorized and/or malicious sender. Controllers just check rules, such as bit-stuffing, CRC, data length code consistency, which may be enough for data integrity, but not for cybersecurity. Moreover, utilizing the CAN mechanisms for automatic fault localization, malicious CAN frames can determine the disconnection of every single controller by posting several well-directed error flags. Similar to the CAN automatic fault localization, the bus guardian in FlexRay can be utilized for the well-directed deactivation of any controller by appropriate faked error messages. Attacks on the common time base, which would make the FlexRay network completely inoperative, are also feasible by posting proper malicious SYNC messages on the bus. Moreover, the introduction of well-directed sleep frames deactivates the corresponding power-saving capable FlexRay controllers.

As possible countermeasures, the following techniques are foreseen and are likely to appear in the new generation of car connectivity devices:

- To cluster the subnetworks and related subsystems in security islands, separated by gateways with embedded cybersecurity functionalities, so that an attack on a non-safety related bus, like LIN or MOST, cannot propagate to the safety-related functions connected to Flexray or CAN [103]. This approach will also be applied to the future architectures based on Automotive Ethernet [113].
- To embed cybersecurity hardware accelerators in new automotive computing units to sustain message encryption in real-time. This is the reason why in the literature new digital macrolayers are appearing, that implement in real-time security techniques like the Advanced Encryption Standard (AES), with different cipher modes, used in symmetric cryptography [114] or more complex algorithms like the Elliptic Curve Digital Signature Algorithm (ECDSA) for asymmetric cryptography [115], [116]. The use of HW-based co-processors is required by stringent latency and energy-efficiency requirements that are not achievable with software-based implementations.
- To embed signature mechanisms for controller authentication in new automotive computing units. Authentication of all senders is needed to ensure that only valid controllers are able to communicate on automotive bus systems [103], [115], [117], [118]. All unauthorized messages may then be processed separately or immediately discarded. Every controller therefore needs a certificate to authenticate itself against the gateway as a valid sender. For example, as proposed to [103], a certificate may consist of the controller identifier ID, the public key and the authorizations of the respective controller. The gateway, in turn, should securely hold a list of public keys of all accredited OEMs for the considered vehicle. Each controller certificate is digitally signed by the OEM with the relevant secret key. The gateway again uses the corresponding public key of the OEM to verify the validity of the controller certificate. If the authentication process succeeds, the relevant controller is added to the gateways list of valid controllers.
- To cluster the ECUs in different trustable classes depending on how easily they can be attacked. For example, in [119] a security framework for vehicular systems, called VeCure, is proposed, which can fundamentally solve the message authentication issue of the CAN bus. Each node that sends a CAN packet needs to also send the message authentication code packet (8 bytes). The ECUs are split into two categories, namely, the Low-trust and the High-trust groups. ECUs that have external interfaces, e.g., OBD-II or telematics are put in the low-trust group. The High-trust group ECUs share a secret symmetric key to authenticate each incoming and outgoing message.
- To implement intrusion detection mechanisms based on the physical or packet layer features. For example, a clock-based intrusion detection system at physical layer is proposed in [105]. Similarly, an in-vehicle network traffic monitoring technique is proposed in [120] to detect the increased transmission rates of manipulated message streams.
- To implement gateway firewalls. For example, as proposed in [103], if the vehicular controllers are capable of implementing digital signatures, the firewall rules are based on the authorizations given in the certificates of every controller. Therefore, only the authorized controllers are able to send valid messages to the high safety-critical in-
vehicle bus systems. If the vehicular controllers do not have the abilities to use digital signatures, the firewall can be established only on the authorizations of each subnet. However, controllers of less restricted networks such as LIN or MOST should generally be prevented from sending messages to the high safety-relevant bus systems as CAN or FlexRay. Simplified firewall-like functionalities can be also implemented in each end-node and not only in the gateways, with the so-called digital data diode [121]. The idea is to interpose a digital unit between the CAN controller and the CAN transceiver to detect and block unauthorized access. When a frame is detected as malicious, the digital unit corrupts the CRC sequence modifying the CRC-field bits. Therefore, the transmission and reception of a frame that is targeted as malicious generates an error condition which is detected by all the nodes in the CAN network (i.e., each node that has received the corrupted malicious frame transmits an error frame). Furthermore, the digital unit conceals the corruption operation from the sender of the malicious frame. As a result, the sender cannot detect the CRC sequence corruption. Hence, the sender will not attempt to retransmit the malicious frame.

VI. Functional and Responsibility Safety

The new world of SW-defined autonomous things brings both technical challenges and liability concerns [122]. Particularly, AD vehicles are composed of electronic platforms with many sensing inputs and many complex processing elements (see Fig. 2), which involves millions of SW lines of code. As a consequence, HW and SW may go wrong and this may cause hazards if no countermeasures are taken. On top of HW and SW failures, cars operate in a very complex environment with many variants, e.g., AD cars share the road with human-driven vehicles. Last but not least, the increase in connectivity through V2X opens possibility for security attacks. Consequently, several potential issues and requirements need to be considered by the automotive manufacturers. One such requirement is FuSA, which is mainly concerned with making the safe from HW failures and SW bugs.

A. FuSa in the Context of the ISO26262 Standard

The first edition of the ISO26262 safety standard consisted of 9 normative parts and a guideline as the 10th part. The second edition of the standard, to be published within 2018, will consist of 10 normative parts and two guidelines, one (the part 11) is specific to the application of ISO26262 to semiconductor components. The goal of the standard is to provide an automotive safety lifecycle (management, development, production, operation, service and decommissioning) and support tailoring of the necessary activities during the lifecycle. The standard also covers the functional safety aspects of the entire development process (requirements specification, design, implementation, integration, verification, validation and configuration). Moreover, the standard provides requirements for validation and confirmation measures to ensure that an acceptable level of safety is achieved. The standard covers both systematic and random failures. The systematic failure (either in HW or SW) is related in a deterministic way to a certain cause that can only be eliminated by changing the design, manufacturing process, operational procedures, documentation or other relevant factors. Whereas, the random HW failure is one that can unpredictably occur during the lifetime of a HW element and that follows a probability distribution.

The standard provides an automotive-specific risk-based approach for determining risk classes (ASIL), where “D” and “A” represent the highest and lowest safety integrity levels respectively. Note that ASIL is as a classification for the overall system, but the safety requirements specified to the HW and SW elements, in general, inherit the same level. For example, today SW-defined cockpit systems require ASIL-B (trending to ASIL-C) while ADAS and AD require ASIL-D. To give an idea of the implications, in terms of HW random failures, ASIL-D means that 99% of the faults potentially violating the safety goal shall be either detected or safely managed and that the overall system shall have a probability of residual (i.e. unmanaged) HW random failures less than 10 FIT (10 faults in one billion hours of operation). An important concept of ISO26262 is the safety mechanism, which is a technical solution implemented to detect and mitigate (tolerate, control or avoid) failures in order to achieve/maintain the intended functionality or a safe state in the case of a failure without an unreasonable level of risk. The second edition of the standard emphasises not only on fail-safe systems but also on fault-tolerant systems. Here the goal is to guarantee the normal (or reduced) operation after a fault has occurred.

Despite FuSa is measured at system level, there are specific requirements for semiconductors. The second edition of ISO26262 will include a new part (part 11) with more than 150 pages of guidelines for digital and analog macrocells, FPGAs and sensor circuits. Herein, some of the most important topics and challenges are as follows.

- How to consider safety aspects of semiconductor components? The aspects of interest include in-context vs. Safety Element Out of Context (SEooC) and definition of the Assumption of Use (AoU). The AoU refers to the usage modes or countermeasures that the system maker has to consider if using the safety related semiconductor component.
- How to define the level of details of the safety analysis as a function of the safety concept, the stage of the analysis and the safety mechanisms used?
- How to determine the correlation between fault, error and failures? The relationship among the fault, error and failures is depicted in Fig. 6(A). This challenge is also concerned with the definition of fault models, failure modes and distribution of failure rate across failure modes. In order to address this challenge, guidelines are required to derive a consistent set of failure modes and consider new fault models (e.g. multiple stuck-at) caused by modern technologies.
- How to handle all kinds of macrocell (hard or soft) with or without embedded safety mechanisms embedded? This challenge also extends to legacy macrocells.
- How to determine base failure rate for both permanent and transient faults? Another challenge in this regard is to deal with non-constant failure rates and advanced packaging.
- How to perform fault injection? The scope of this challenge spans over different abstraction levels that support evaluation of the hardware architectural metrics, pre-silicon verification of safety requirements and detection of faults and control their effects.
- How to identify dependent failure initiators (DFI, see Fig. 6(B))? A related challenge is how to perform the dependent failure analysis (DFA).
• How to define and apply fault models, failure modes, safety mechanisms, and avoidance of systematic failures, with respect to ISO26262, for HW platforms? The platforms include digital and analogue components, memories, PLDs/FPGAs, sensors/MEMS, multi-cores, and modern SoCs. The SoCs used in the automotive domain include a combination of the following HW and SW features.

- EDC/ECC for memories, including caches and registers.
- Built-In Self-Test (BIST) for arrays and logic, that are operated both at key-on/off and at periodic intervals.
- Safety mechanisms for on-chip interconnects, including coherent fabrics (e.g., information redundancy, data/address codes, firewalls and timeouts).
- Different redundancy types for processing cores, see Fig. 7.
- End-to-End safety protocols for peripherals. These protocols are combinations of CRC, time stamp and frame counter.
- SW Test Libraries (STL) to address permanent failures in the logic not covered by other safety mechanisms.
- Dedicated HW cores for fault handling (e.g., Safety Island).

\[ \text{Fig. 6. A) Link between fault, error and failure; B) Dependent failures} \]

B. Responsibility-Sensitive Safety

The most recent trend in FuSA is Responsibility-Sensitive Safety (RSS). Introduced by [123], the RSS model formalizes the common sense of human judgment under a comprehensive set of road situations. It sets clear definitions for what it means to drive safely versus to drive recklessly. With human drivers, the interpretation of responsibility for collisions and other incidents is fluid. Today, in the case of an accident, the blame is determined based on imperfect information and other factors interpreted afterwards. With machines, the definitions can be formal and mathematical. Machines have highly accurate information about the environment around them; they always know their reaction time and braking power, and are never distracted or impaired. We do not need to interpret Machines’ actions after the fact. Instead, we can program them to follow a determined pattern – as long as we have the means to formalize that pattern. At its core, the RSS model is designed to formalize and contextualize today’s driving dilemmas, like notions of safe distance and safe gaps when merging and cutting in, which agent cuts in and thus assumes responsibility to maintain a safe distance. Moreover, this model allows to specify the right of way, define safe driving with limited sensing (e.g., when road users are hidden behind buildings or parked cars and might suddenly appear), and more. Clearly, human judgment includes avoiding accidents and not merely avoiding blame. The RSS model attempts to build a formal foundation that sets all aspects of human judgment in the context of driving with the goal of setting a formal “seal of safety” for autonomous cars. More details on the RSS model can be found in [123].

\[ \text{Fig. 7. Different redundant architecture solutions} \]

VII. Conclusion

This paper analyzed recent technological challenges and HW/SW solutions for on-board embedded and networked automotive systems. In this context, the paper mainly focused on automotive SW, advanced execution platforms, on-board network communications, on-board cybersecurity and functional safety with respect to SW and HW. The paper identified the need for new E/E architectures, exploiting eHPC and number-crunching accelerators, supervised by a safe and secure MCU, to meet the computation and memory requirements in the order of TOPS and TB, respectively, for perception and fusion tasks. Beside HW, also the automotive SW complexity has drastically increased in the recent years. Model- and component-based SW development techniques have proven helpful and cost effective in managing the size and complexity of automotive SW, which is often in the range of several tens to hundreds million lines of code. The SW complexity is expected to grow further in time. Hence, there is a strong need to develop efficient models and languages for the automotive SW development. Moreover, the existing standard technologies for the SW development (e.g., AUTOSAR) need to adapt according to the evolution in the CAR industry, with respect to advanced computer-controlled functionality, AD, ADAS and V2X. There is also a strong need to support interoperability and automation among the state-of-the-art and state-of-the-practice languages, models and tools that are used for the automotive software development at various abstraction levels. In-car communications require new network architectures. Ethernet, with a broad choice of data-rates, and the TSN standards will be key enablers for upcoming automotive scenarios, including autonomous driving. Current in-vehicle networks suffer from several vulnerabilities in terms of confidentiality, authentication, and availability. While some possible countermeasures have been already found, vehicular communications (V2X) and automated driving are fostering the steady rise of novel challenging vehicular cybersecurity issues. In addition to security, other open research topics that deserve investigation include traffic planning response-time analysis of TSN networks, and the use of Ethernet for 5G mobile fronthaul.

ACKNOWLEDGEMENT

The work in this paper is supported by the Swedish Governmental Agency for Innovation Systems (VINNOVA) through the project DESTINE, the Swedish Knowledge Foundation (KKS) via the projects HERO and DPAC, University of Catania through the project CHANCE, University of Pisa through the grant PRA2017 and EPI H2020 programs.
REFERENCES


[39] Peter Thorngren, keynote Talk: Experiences from EAST-ADL Use, EAST-ADL Workshop, Munich, Germany.
