A Vision of Cyber-Physical Internet

Anis Koubâa*† and Björn Andersson*

* CISTER Research Unit, Polytechnic Institute of Porto (ISEP/IPP), Portugal.
† Al-Imam Mohamed bin Saud University, College of Computer and Information Sciences, Riyadh, Saudi Arabia.

Emails: aska@isep.ipp.pt, bandersson@dei.isep.ipp.pt.

Abstract—When the Internet was born, the purpose was to interconnect computers to share digital data at large-scale. On the other hand, when embedded systems were born, the objective was to control system components under real-time constraints through sensing devices, typically at small to medium scales. With the great evolution of the Information and Communication Technology (ICT), the tendency is to enable ubiquitous and pervasive computing to control everything (physical processes and physical objects) anytime and at a large-scale. This new vision gave recently rise to the paradigm of Cyber-Physical Systems (CPS). In this position paper, we provide a realistic vision to the concept of the Cyber-Physical Internet (CPI), discuss its design requirements and present the limitations of the current networking abstractions to fulfill these requirements. We also debate whether it is more productive to adopt a system integration approach or a radical design approach for building large-scale CPS. Finally, we present a sample of real-time challenges that must be considered in the design of the Cyber-Physical Internet.

I. INTRODUCTION

The vision towards large-scale distributed computing systems is currently evolving to a new frontier, where computation is no longer decoupled from its environment. This sight stems from the need to integrate external physical data and processes with computations for sake of pervasive and ubiquitous control of the surrounding environment. However, it is commonly known that this integration is not a new concept as it has always been the case with embedded systems. In fact, embedded computing systems are intrinsically dependent on their environment where they are deployed through sensing physical processes. As computing becomes increasingly integrated into our environment, traditional embedded systems has found their limits in satisfying the new requirements of massively networked embedded systems. On the other hand, the Internet has been providing a worldwide infrastructure for data sharing and information retrieval. However, Internet applications have been driven by the need to exchange logical information at large-scale; nevertheless, the mapping between the physical environment and the logical information has not been considered in the design of those applications. Thus, the convergence of the Internet with embedded systems is an important milestone for enabling large-scale distributed computing systems that are tightly coupled with their physical environment. On the one hand, a first step towards this convergence has been put into practice by Radio-Frequency Identification (RFID) based systems, which have enabled the concept of network of physical objects, commonly known as Internet-of-Things [1]. In November 2005, the International Telecommunication Union (ITU) produced an executive report in the World Summit on the Information Society that summarizes the basic concepts of Internet of Things, related technologies, challenges and concerns, market opportunities, etc. Clearly, RFID has been considered as the key technology for Internet-of-Things.

On the other hand, the Wireless Sensor Networks (WSNs) paradigm has emerged as another alternative to networks of physical events, which supports the control and monitoring of physical phenomena in the environment through sensing. Some other alternatives considered the use of sensor-based mobile phones for monitoring everyday items through cellular networks [2]. All of these approaches fall into the concept of Cyber-Physical Systems (CPS), which are systems deployed in large geographical areas and generally consist of a massive number of distributed computing devices tightly coupled with their physical environment. Fig. 1 presents the main components of cyber-physical systems.

The frontier between CPS and Internet-of-Things has not been clearly identified since both concepts have been driven in parallel from two independent communities (i.e. sensor networks and RFID, respectively), although they have always been closely related. The history returns itself as this situation
may be thought to be similar to the design of the Internet (driven by the TCP/IP community) and Telecommunication networks (driven by International Telecommunication Union - ITU) in the early eighties. However, with the emergence of the recently released 6LoWPAN [3], the convergence between CPS and Internet-of-Things becomes a real fact as it enables to use the Internet as supportive infrastructure to sensor networks, similarly to its integration with RFID systems.

The current status of these new emerging cyber-physical systems recall to the mind the age just preceding the birth of the Internet, when networks were scattered and private mainly due to lack of standards. Similarly, CPS are currently scattered and private networks, each performs specific tasks related to the environment where it operates. The main challenge in the design of CPS is how to enable the interconnection and interoperability of all these scattered networked embedded systems into a single large-scale network that satisfies all their requirements. There are a number of handicaps that hinder the set-up of a unified network for cyber-physical systems. In addition, one important question is whether it would be better that the design of CPS follows a system integration approach, which consists in integrating heterogeneous networks together to form a CPS, or a radical design approach (as it was claimed in several papers [4]–[6]), which consists in building CPS from scratch.

In this position paper, we first present and criticize the recent vision towards the design of CPS radically from scratch and we show that realism imposes not to lose legacy. We show that the real current trends address the challenge of interoperability between existing heterogeneous systems to form a universal network interconnecting not only data but also objects and physical events. Finally, we present the networking abstraction and challenges for the design of the Cyber-Physical Internet (CPI).

II. THE DESIGN DILEMMA

As of the emergence of CPS, there were many calls to rethink the computation foundations to cross a new frontier towards future cyber-physical networks [4]–[6]. This is a pretty nice statement to trigger new theoretical research challenges; however, in the practical sense, it may face serious limitations. Although it is obvious that computation paradigms must be adapted to the new requirements that arise with the emergence of the cutting-edge cyber-physical technologies, adopting a radical design approach, as it might be understood from the literature [5], [6], seems to be not pragmatic, at least in medium to short terms. During the modeling process of large-scale complex systems, it is always important that designed models ensure the best trade-off between their different requirements; however, when taking a look to history we can realize that real systems that spread out into the market does not really fulfill the objectives of the theoretically expected models. A straightforward question is: "Will Cyber-Physical Systems face the same fate?"

As a matter of fact, two worldwide standard technologies perfectly embed this belief: IP (Internet Protocol) and IEEE 802.11 (hereafter, WiFi). These two standard protocols are commonly known to be rather poor in terms of efficiency and Quality-of-Service (QoS). Several patches have been proposed for IP (such as Integrated Services, Differentiated Services, etc.) as well as for WiFi (e.g. IEEE 802.11e extension) to enhance their performance. In spite of efficiency and QoS shortage, these two protocols have been widely and quickly spreading since their release. On the other hand, other more sophisticated protocols such X.25, ATM or HyperLan have been designed with more care to achieve higher efficiency and better QoS, but did not gain too much space in the commercial market. It appears that there is always a gap between how new systems are expected to operate and how they do operate in reality. The reason is that the vision of the market stakeholders is different from the vision of academic researchers, as the former do not care about optimized efficiency, but rather reduce the time-to-market and cost of real products. Hence, it can be easily noticed that, in practical terms, the modeling paradigm is to quickly design, implement and put-into-market simple solutions that (1) just work, (2) fulfill basic requirements and (3) can be patched to plug new functionalities or to improve their behaviors. Therefore, it seems that rethinking the current computation foundations to build large-scale CPS is not pragmatic. Instead, it seems that it is more natural to take profit from the legacy infrastructure to achieve the large-scale CPS objectives. We thus believe that research efforts must focus on the system integration approach for enabling very large-scale CPS, which we refer to as the Cyber-Physical Internet. Consequently, interoperability is the key challenge to build large-scale heterogenous cyber-physical networks. In addition, it is necessary to take into consideration the specificities of CPS in the integration process of existing networks, in particular the nature of the manipulated data, as discussed in the next section. One question may thus arise: "What will be the core protocol for the prospective CPI?". Definitely, IP is the legacy protocol that will play the key role in the future Cyber-Physical Internet. IP has been thought for so long — since the birth of the sensor network paradigm — as being non compatible with the requirements of sensor-based systems. However, this thought has been recently revisited [7]–[9] and with the emergence of 6LoWPAN [3], [9] that embeds IPv6 on top of the IEEE 802.15.4 [10] as an alternative to ZigBee [11] Network Layer. The challenge has been won by the IETF 6LoWPAN Working Group and 6LoWPAN becomes a serious competitor to ZigBee as it enables to seamlessly merge the sensor network world with the Internet, which ZigBee is not able to. However, if 6LoWPAN wins in terms of high degree of interoperability with existing networks, it still needs to justify its efficiency in terms of energy-efficiency and real-time guarantees. In fact, the design objectives of 6LoWPAN are likely to put more weights on interoperability and integration with the Internet rather than on the typical requirements of sensor networks. Optimizing the trade-off between those design objectives remains a research challenge.
III. NETWORKING ABSTRACTIONS OF THE CYBER-PHYSICAL INTERNET

In this section, we define the requirements that have to be considered in the specification of the networking abstractions of the Cyber-Physical Internet, which can roughly be viewed as the large-scale universal network that interconnects several heterogeneous CPS. We consider a CPS as a mixture of several and different networks that monitor physical objects and events, including WSNs, RFID-based systems, mobile phones, etc.

Future massively networked embedded systems require new standards for achieving interoperability. In fact, prospective large-scale CPS should not be foreseen as separated and dispersed systems, but as a unified system that seamlessly interconnects heterogeneous cyber-physical components. The challenge will be to build a global network that interconnects all cyber-physical devices and provide plug-and-play services to the end-users in a completely transparent way. It is therefore necessary to rethink the current networking abstractions to ensure a worldwide interoperability of cyber-physical devices. This requirement imposes the design and the development of new standardized protocols for cyber-physical systems. These protocols have to be designed while taking into account the properties of the environment, where the CPS will be deployed. The IP protocol stack model and the WSN protocol stack model [12] feature fundamental limitations that must be addressed in the design of the Cyber-Physical Internet. In fact, the current protocol layers make a total abstraction on the nature of data to be processed, thus it is not possible to design protocols tightly coupled with their external environment. It is therefore necessary to propose an extended protocol stack model for CPS that also integrates the properties of the physical environments. Fig. 2 presents the potential reference architecture for the CPI.

The protocol stack architecture for CPI must include an additional layer, the Cyber-Physical Layer (CY-PHY layer), which provides an abstract description of the properties and nature of cyber-physical data. This layer must provide the set of protocols to universally represent data in a unified and structured way. In addition, the CY-PHY layer should provide services for lower layer protocols to support an efficient cross-layer design of the underlying application and communication protocols. This means that all protocol layers have to adapt their behavior according to the information provided by the CY-PHY layer. For instance, in the context of health care monitoring, the information provided by the body sensors will have a significant impact on the behavior of the protocol suite designed for this application. In fact, depending on the type and the nature of cyber-physical data, several changes may be imposed in the protocol layers, namely:

Physical Layer: The input from the CY-PHY layer can lead to changing some properties of the physical channel such as the channel frequency band and the modulation scheme depending on the requirements of the cyber-physical data. It can be conceived that some frequency bands can be allocated for critical cyber-physical traffic using robust modulation schemes. Note that there have been great advances in radio technologies with the design of software radios and cognitive radios, which offer better flexibility to adapting radio properties to the user requirements.

Medium Access Layer: The MAC layer usually grants different operational modes and services for higher-layer protocols such as synchronized or unsynchronized modes, different levels of Quality-of-Service (QoS), energy management services (e.g. duty cycle) etc. In CPS, the dynamic behavior of the environment would significantly impact the operational behavior of the MAC layer, which must be adaptive to the CY-PHY data. For instance, the decision to switch from synchronized operational mode to unsynchronized mode (or vice-versa) or the adaptation of the duty cycle must be driven by the CY-PHY layer. This can be achieved by some existing technologies such as the IEEE 802.15.4 protocol, which offer the beacon-enabled mode (synchronized) and the non beacon-enabled mode (unsynchronized) in it MAC layer. This interaction between the MAC layer and CY-PHY layer is very important to enable a close-loop control of the QoS based integrated with the status of the monitored environment.

Network Layer: The Network Layer provides routing and data aggregation services. The cross-layer interaction between the Network Layer and the CY-PHY Layer is necessary to define the adequate routing strategies and data aggregation mechanisms. For example, the aggregation functions used for processing temperature information would be completely different from those used for accelerometer or bio-medical sensory data. In addition, the selection of the routing mechanism or the parameters affecting a given routing protocol may depend on the nature of the data and also from the status of the environment.

Transport Layer: Transport protocols have not been extensively investigated for CPS (e.g. wireless sensor networks, embedded systems), although it is of a paramount importance to specify different degrees of reliability with respect to the end-to-end delivery of data. This naturally implies three

![Fig. 2. Protocol Stack Architecture for CPI](image-url)
classical tasks in the transport layer including (i) reliable transport, (ii) flow control and (iii) congestion control mechanisms. The Internet already provides the connection-based TCP and connectionless UDP transport protocols for providing guaranteed and best-effort services, respectively. These heavyweight protocols are not suitable for CPS applications, which raise the need to rethink new transport protocols that cope with the requirements and properties of CPS. A real challenge with regards to transport protocols is to design reliable transport protocols without the need to send back acknowledgements to the source nodes to avoid drowning the network with increasing control traffic.

**Application Layer:** In CPS, the application layer is responsible of processing data and extracting useful information with respect to the application objectives. One main challenge is to provide standard distributed signal processing algorithms/protocols for each potential CPS applications. This will encourage the development of CPS applications and reduce the time-to-market and cost.

**IV. REAL-TIME CHALLENGES FOR CPS**

Real-time usually imposes serious challenges in the design of cyber-physical systems. However, this issue must be carefully analyzed and some of the concepts must be revisited. There are several promising research directions in the real-time area. In what follows, a sample of potential directions are presented:

- **Operating systems:** It is important that operating systems supports real-time, although it induces additional design complexity. The main challenge is to achieve an optimal balance between several important features needed by a CPS operating system, including modularity, effective hardware/software split, hardware abstraction, energy efficiency, and real-time [6]. The most widely used operating system, TinyOS, represents a promising solution for CPS as it addresses most of those requirements. However, the lack of real-time support represents a serious limitation in TinyOS for developing real-time protocols and synchronization mechanisms [13]. The lack of pre-emption and prioritization in TinyOS is a main handicap for providing predictable timing behavior at node level. In many cyber-physical applications, where timing constraints must be respected, reliability and real-time are very much coupled. It is therefore fundamental to consider timing guarantees for building reliable systems. As a matter of fact, the lack of real-time in TinyOS has prevented the release of standard-conforming IEEE 802.15.4 protocol stack for both open-ZB [14] and TKN implementations [15], [16]. In [17], the authors demonstrated that the behavior of the IEEE 802.15.4 implementation has been much reliable when implemented over ERIKA [18], a promising real-time operating system for embedded devices.

- **Networking protocols:** From the networking perspective, real-time imposes several challenges still open to research. Distributed and adaptive resource allocation in synchronized multi-hop sensor networks, where resources must be adequately allocated depending of the physical/logical network changes, represents one interesting research problem. In synchronized WSNs, it is naturally more efficient to grant resources (bandwidth, memory) to active sensor nodes involved in critical tasks. The use of static allocation plans is clearly not efficient for highly dynamic and mobile systems as they do not adapt to the system changes. On the other hand, the centralized adaptive synchronization induces a significant amount of computation and communication overheads, which cannot really work in resources-constrained WSNs, due to its complexity and non-responsiveness. We need to find new approaches for adaptively managing resources for synchronized and mobile multi-hop WSNs in a distributed, efficient, transparent, and most importantly real-time way.

- **Timing Guarantees:** The provision of deterministic real-time guarantees in unpredictable wireless ad-hoc and sensor networks is considered as a questionable issue. In the literature, most of the papers dealing with deterministic guarantees assume that channels are error free. While this assumption might be correct for very extreme and rare cases, where wireless links are very stable and of a high quality, most of the real-world applications refute this assumption, since the majority of the wireless links are typically located in the gray region, where links are highly variable and unstable. The notion of real-time guarantees must be revisited, as we need to find new means to characterize deterministic performance under channel uncertainty. Claiming that a wireless network deterministically provides a delay bound would not make sense. One interesting characterization is to associate a confidence level with each guaranteed delay bound. The objective of the associated confidence level is to quantify the uncertainty on the guaranteed delay bound, as illustrated in Fig. 3. The main idea consists in computing the delay bounds taking into account the number of possible retransmissions, and to statistically determine the distribution of the number of retransmissions in a given channel. Fig. 3 shows the delay bounds and the cumulative distribution function (CDF) for different values of number of retransmissions due to channel errors. The CDF helps on bounding the maximum number of retransmissions with a certain probability, which defines the confidence level. In this case, it is possible to determine the delay bound that corresponds to the the maximum number of retransmissions, with an associated confidence. For instance, in Fig. 3 we can observe that the delay bound is equal to 0.35 time unit with a confidence level of 97%.

- **Performance Compositionality:** The end-to-end delay analysis in CPS is a complex and stimulating problem,
in particular due to their heterogeneity. Traditional end-to-end delay analysis basically derives system-level delay from component-level delays, and this paradigm has been known to be inefficient. More sophisticated techniques use composition theorems to transform a multi-component system into a one-component system, thus reducing the complexity of the end-to-end analysis [19], [20]. In [19], [20] the authors presented a complete framework to systematically transform distributed real-time systems into a single system, which is used to infer the end-to-end schedulability of the original system. Although these works represent a pioneer contribution towards performance compositionality in distributed real-time systems, their traffic model — based on the schedulability analysis theory — is rather restricted to periodic/aperiodic streams and constant execution times (i.e. worst-case execution time), which is not enough generic for modeling heterogenous CPS applications. On the other hand, Network Calculus formalism, which relies on more generic traffic models (defined by their upper bound curves) [21], provides concatenation methodologies for reducing the analysis of a multi-hop system to a single-hop system by determining an equivalent service curve for the whole system. The concatenation analysis, however, has two drawbacks: (1) The equivalent service curve for a given stream depends on a parameter $\theta$, whose optimization is quite complex [22]. (2) The analysis relies on rate-latency service curves, which is not enough generic for modeling services in heterogenous CPS applications. For that reason, we need to find adequate system reduction techniques that rely on general model abstractions for representing traffic and services, and that take into account system heterogeneity.

- **Data aggregation:** As already stated in Section III, the current Internet protocol layers make a total abstraction on the nature of data to be processed. However, many CPS applications are not interested in the data itself but they are rather interested in high-level queries about the physical world. It is possible for a user to request that each sensor delivers its sensor reading and then computes the result based on all those sensor readings. Nevertheless, such an approach generates an enormous amount of data traffic something that (i) increases the time required to obtain the result of the query and (ii) wastes energy of sensor nodes. Performing information processing inside the network, for example allowing routers to also process incoming packets before forwarding, can lead to significant improvements however. This is often referred to as data aggregation, content-based network, in-network processing or data distillation network [23]. Regardless of its labeling, three important issues remain for the use of such an approach in Cyber-Physical Internet.

1) **Query language specification.** There is a need to define a language in which users can define their queries and these queries should be injected into the network. The community of wireless sensor networks is currently using slightly modified variants of SQL. However, these SQL variants are not sufficiently expressive. For example, in a scenario where we desire to detect whether a route is ice-free, we may wish that sensor nodes perform signal processing locally something that is difficult to perform efficiently with SQL. One approach could be however that a sensor allows users to install a device driver on that sensor node and this device driver acts as a virtual sensor; a virtual sensor performs a computation based on physical sensors. For example, a virtual sensor may deliver a Boolean value "true" if there is ice close to this sensor and "false" otherwise. Then, SQL may be used to express queries based on the virtual sensors. In fact, support for such virtual sensors are already available in a software package called Global Sensor Networks (GSN) [24] but it is used to be run on a gateway interfacing with a wireless sensor network rather than to be run on sensor nodes themselves.

2) **Query planning and optimization.** The research community of databases has produced an extensive literature on query processing of SQL queries. It typically assumes that the cost (for example time) of a query should be minimized and the query planning/optimization attempts to find a way of executing the query such that the cost is minimized. These works assume that the cost is dominated by disk accesses or CPU processing. However, for data aggregation in CPS, we expect the limited capacity of the (wireless) communication channel to be the main bottleneck and therefore query optimization should strive to minimize that cost instead. This is non-trivial because (i) knowledge about the (wire-
less) network topology and interference relationships between nodes are needed in order to exploit parallel transmissions and (ii) some operations can be performed at great efficiency (such as MIN, MAX) with a prioritized MAC protocol [25]; the query optimizer must be aware of that potential when taking decisions in how to decompose a user-query into operations.

3) Data integrity. When a user asks a query he wants to be sure that the sensor readings are authentic. Since data aggregation allows routing nodes to modify the data payload, normal end-to-end encryption/authentication methods do not work. Therefore, ensuring data integrity must be an integral part of the network.

V. ACKNOWLEDGEMENTS

This work was partially supported by the Portuguese Science and Technology Foundation (Fundacao para Ciencia e Tecnologia - FCT), the ArtistDesign Network of Excellence on Embedded Systems Design, funded by the European Commission under FP7 with contract ICT-NoE- 214373 and CONET, the Cooperating Objects Network of Excellence, funded by the European Commission under FP7 with contract number FP7-2007-2-224053.

REFERENCES


