Challenges when bringing IoT into Industrial Automation

Tomas Lennvall
RISE SICS Västerås, Sweden

Mikael Gidlund
Mid Sweden University, Sweden

Johan Åkerberg
ABB Corporate Research, Sweden

***Abstract*—The Internet of Things (IoT) is captivating the society because of its potential to rapidly transform businesses and people's lives. It is widely believed that IoT will also transform the industrial automation business in terms of improved productivity, less cost, flexibility, and increased revenues. Hence, there are some challenges that needs to be addressed when IoT is introduced to the industrial automation domain. This paper aims to present realistic requirements and highlights identified challenges such as security, interoperability, deterministic and low latency communication, and how the required availability (uptime) can be kept. Moreover, the paper also point out the need of standardization and sustainable business models. The conclusion is that introducing IoT devices and connecting them directly to cloud services is not straightforward for process automation.**

I. Introduction

Internet of Thing (IoT) has been a trend for many years and is being hailed as the next industrial revolution and the next Internet. Business analysts foresee huge potentials from IoT solutions for business, governments and civilians alike. Initial estimations of the number of connected *things* was in the range of 50 billions by 2020. This estimate has now been refined to closer to 30-35 billion things [1]. No matter what, it is a staggering number of things which are expected to be connected to the next Internet. Examples of IoT applications are set-top boxes, home automation systems (e.g., thermostats, white goods), smart meters, environmental monitoring (e.g., humidity, temperature, acidity), smart cities solutions (e.g., traffic monitoring, free parking locations, garbage monitoring) and intelligent transportation [2], [3], [4].

Recently IoT has started to move into other domains and in recent years, the trend in industry and academia is to "connect the unconnected" and the vision is that millions of resource constrained embedded devices used for time- and mission-critical applications will be connected to the Internet. This network of ubiquitous smart objects it called the Industrial Internet of Things (IIoT). In [5], it is estimated that shipments of wireless devices for industrial applications including both network and automation equipment reached 3.7 million units worldwide in 2014. Growing at a compound annual growth (CAGR) rate of 23.2 percent, shipments are expected to reach 12.9 million by 2020. The installed base of wireless devices in industrial applications is forecasted to grow with a CAGR of 27.2 percent from 10.3 million connections at the end of 2014 to 43.5 million devices by 2020. Even though IoT seems to promise such remarkable opportunities, a justified question is why it is being very slowly adopted into industrial automation?

Industrial automation is a domain in which applications (i.e., processes), e.g., making steel, paper, oil, chemicals, are very complex, critical, and hazardous. If such a process encounters problems there are usually severe consequences, such as financial loss due to production problems, dangers to people and the environment due to for example toxic spill. An industrial plant is therefore designed and constructed to support the tough requirements posed by the application, e.g., industrial grade devices (e.g., IP and EX protection), high reliability and redundant solutions for communication and control systems. Most IoT solutions are designed for the consumer domain, thus, not at all supporting harsh industrial requirements. Therefore, plant owners hesitate using IoT solutions for any critical operations in their process, maybe considering using it for non critical parts only. In [12] the authors have developed an industrial cloud architecture which according to the authors support various applications, ranging from control to enterprise management. The architecture includes methods to deal with the fact that running a control loop in the cloud is not as timely and reliable as running it locally in the plant automation system. Many control loops have too tight requirements on time and reliability to run in the cloud today. One question arises: is it easier in terms of engineering, maintenance, and service to have parts of the plant automation system run locally and other parts in the cloud?

This paper describes which challenges needs to be overcome before IoT solutions can penetrate the industrial automation domain in a wider context. Firstly, the IoT systems needs to meet the strict requirements that are given in the industrial automation domain, especially the IoT systems needs to offer availability, reliable and deterministic communication of data, and low latency and low jitter is another requirement. Furthermore, the IoT solutions needs to offer device interoperability and a high level of security. It is important that security is addressed throughout the device lifecycle, from the initial design to the operational environment. Moreover, before IoT can be successfully launched in industrial automation there needs to be a sustainable business model in place. All of the aforementioned topics will be discussed within this paper.

The reminder of the article is organized as follows: Section II describe different use cases and requirements for industrial automation. Section III outlines and discuss the main challenges to adopt IoT in industrial automation, and finally Section

IV concludes the paper.

II. Industrial Applications and Requirements

Industrial automation can be divided into several sub-areas; building automation (BA), process automation (PA), factory automation (FA), and substation automation (SA), which have different distinct requirements on communication, see Table I. By tradition the communication medium in industrial automation is provided by wires and it has shown to be successful. However, with advancement of technology and new requirements on mobility, flexibility, cost, and easy maintenance, wireless communication has become a viable option but still several challenges remain to solve before it can be adopted in full scale [6].

In Fig. 1 the general architecture model for industrial automation is shown (It is also called the 'Automation Pyramid'). The automation pyramid is divided into several layers with a different set of networks, each with different demands and importance of various properties. In the bottom of the hierarchy is the field network which typically consist of sensors and actuators. At this level the main requirements on communication is real-time behavior, low latency and low jitter for control applications. The next level is the control network which typically consists of controllers and connectivity servers. The top levels are server and plant networks, which basically consist of operator workplaces, engineering and monitoring stations and servers, and at the highest layers business decision systems. In general, the higher layers of the pyramid have more relaxed constraints on latency and real-time properties compared to the lower layers. The bottom two layers consists of Operations Technology (OT) equipment and protocols, which are the core critical part of the plant automation system. All the above layers consists of Information Technology (IT) equipment and protocols. The top level is usually the only level which can be connected to Internet through VPNs and firewalls. There are firewalls in place bordering all levels in the pyramid to protect the mission critical lower layers of the automation system. It should be noted that less obvious requirements on the communication systems are to enable high resolution time synchronization in the system in order to have distributed measurements at the same time for later control where the predictable latency can be compensated for.

A. Commercial versus Industrial Networks

The most essential difference between commercial and industrial networks is that the commercial ones are traditionally built for best-effort traffic, while industrial networks focus on predictability, determinism, safety, and real-time data transfer. Below we list some of the main differences between commercial and industrial networks.

Time sensitivity Industrial Networks typically provides real-time properties to support the critical industrial automation applications. Applications are time sensitive and thus have deadlines which are critical and considered hard. Thus, a single deadline miss due to a packet arriving late can put the automation system in an unsafe state.

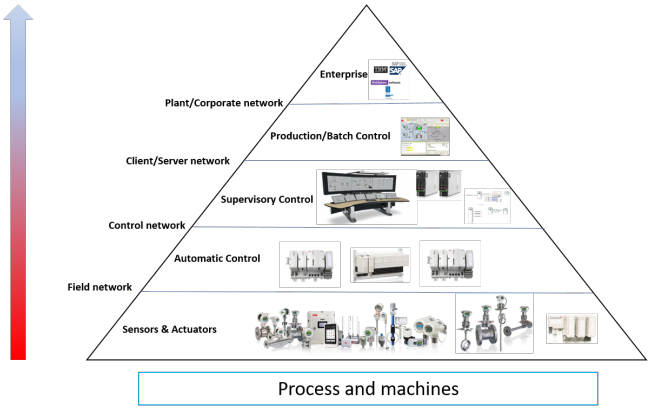

Fig. 1. General architecture model for industrial automation [6].

However, the $(m, k) - firm$ [13] approach is used when designing automation systems to loosen the hard requirement of zero deadline misses, i.e., the industrial network is guaranteed to meet m out of k consecutive deadlines before it is considered to fail. This implies that data is considered useless once its deadline has passed since fresh data has been produced at the source. Deadlines are typically in the sub-second range for the more critical parts of the automation system (i.e., lower levels of the pyramid). Commercial networks tend to have zero predictability and no or possibly soft deadline requirements. Commercial networks are designed for best effort and focuses on achieving a good average latency (which is not to be confused with a real-time deadline!). Many commercial networks provides high reliability by using protocols such as TCP, which guarantees packet delivery. However, using TCP for an industrial network, where data is considered old and useless after a missed deadline is contradictory to keeping on retransmitting until successful. Industrial networks therefore employ other means to provide reliable service.

Predictability & Determinism Besides supporting real-time deadlines, industrial networks must be both predictable and deterministic. It must be possible to predict how the network will behave before it is up and running, i.e., to know that all application requirements will be met. The network itself must be deterministic, i.e., behave the same way at runtime, to support predictability. There are many different ways to achieve a high reliability for communication, e.g., retransmissions (e.g., TCP), error correcting codes, multi-path routing. However, many of these solutions do this at the cost of sacrificing predictability and determinism, e.g., keeping on retransmitting a packet in order to successfully transfer it may cause it to miss its deadline! Commercial networks have no or extremely limited support for predictability and determinism since they are designed for best effort applications.

Safety Safety is one of the most important requirements of an automation system. Safety is a broad topic, but

TABLE I
REQUIREMENTS FOR SOME TYPICAL AUTOMATION DOMAINS

Application Domain	Update Frequency	Nodes / m^2	Telegram loss rate
Building Automation	Seconds	1 - 20	$< 10^{-3}$
Process Automation	10 - 1000 ms	1-20	$< 10^{-5}$
Factory Automation	500 μs - 100 ms	20 - 100	$< 10^{-9}$
Substation Automation	250 μs - 50 ms	1 - 10	$< 10^{-9}$
High Voltage DC control	10 - 100 μ s	300 - 500	$< 10^{-9}$

for automation systems the aspects of intrinsic safety (e.g. electric/electronic failures should not ignite gases) and functional safety (failure in the system should not cause dangerous operations). Many industrial networks support special functional safety protocols. These protocols ensures that the automation system is kept in a safe state at all times to avoid failure and malfunctions, which otherwise may result in serious harm to people, equipment, or the environment. It should be noted that real-time guarantees are necessary for functional safety in order to guarantee that the safety function is executed within the safety function response time, since a deadline miss can expose a dangerous system behavior. There are dedicated safety protocols to be used in combination with the communication protocols in order to have safe end-to-end communication over non safety certified networks. Commercial networks can be used if they are deterministic and provide sufficient error detection mechanisms.

QoS Industrial networks must be able to support data with various requirements on QoS. Data can be transmitted both periodically, which is typical for process values, or aperiodically, which are used for alarms, configuration data, and reports. Depending on the importance of the data (dictated by the application), the network must be able to prioritize important traffic over less important traffic (e.g., time critical over best effort). Higher priority always have precedence over lower priority. There is no notion of fairness! Commercial networks are designed around the notion of fairness and tries to avoid starvation for applications at all costs.

Data size One major difference is the size of data. Typically the data in an industrial network is only a few bytes (the actual value being measured and some overhead information). Occasionally, configuration data or time series are transmitted. Commercial networks typically transmits regular data of Terabytes, Megabytes, or Kilobytes with a typical data size per packet of around 1400 - 1500 bytes. Such differences affects how the network protocols are designed, as well as its performance with respect to reliability and security.

III. CHALLENGES OF USING IOT IN INDUSTRIAL AUTOMATION

The vision for IoT is to connect all the industrial devices (things) located on level one (Sensors & Actuators) to the Internet. This means that the most critical part of the plant will be exposed to the Internet, thus creating some challenges which must be overcome. Challenges are not only technical, since the adoption of new technologies must bring added business or cost benefits in order to succeed.

A. Installing and retrofitting IoT

Adding IoT connectivity to level 1 of an industrial system can potentially introduce security holes. Legacy industrial systems have not been designed to have external connectivity directly to the "floor" of a plant, i.e., the critical lower levels of the pyramid. Internet connectivity is usually at the highest level, and each level below is protected by firewalls and security measures at each level borders. Thus, adding IoT devices or connectivity to legacy devices requires careful consideration of new security threats, how they will affect the whole industrial system, and how a required security level could be maintained through the plant lifetime. Besides the security aspect, one the communicated advantages of IoT is the harmonization of protocols avoiding different 'silos' which is enabling interoperability and technology penetration. This is true from an IT perspective, where the advantages from the IT domain can be reused. However, the success of industrial automation systems heavily relies on reliable and available deterministic communication and execution in order to reach the main objective of an automation system. The rest of this chapter will elaborate further on the fundamental requirements that needs to be in place in order to have safe 24/7 operation of large scale industrial production plants. Before going into the technical details of the different requirements and challenges it is important to give an overview of the current state-of-practice in the area of industrial automation and how the subsystems interact.

Industrial automation systems are designed to control processes in such a way that there are minimum quality variations of the final product over time, with high availability due to the return-of-investment, lower the business risks and minimize the operational costs and environmental impact. Therefore, the requirements of availability, determinism and timely functional response are essential. If IoT should be installed in large scale industrial plants at connectivity level one, without any silos, the IoT technologies needs to provide better or similar performance as the current automation systems. The main reason for this statement is that otherwise a parallel IoT silo has to be installed next to the OT devices. If not, the information needed for control has to be sent and executed in the OT silo, and the supervisory information in the IoT silo.

Integrating the IoT system into the automation system and feeding information from an IoT system into a PLC might have safety-related problems as the quality (timeliness, reliability, etc.) is not known when the information is available in the automation system. That information (process data) can later on be used for process control by a control engineer as the data is available in the control system, but it's not visible in the system that this information is not intended for control purposes. In addition, the process instrumentation and its sensor parts are usually not mounted as a thermostat on a wall, but most often installed in the process (pipes, tanks, boilers, etc.) and are installed in a harsh environment and are usually expensive and difficult to install. This implies that for reduced capital and operational expenses, and to gain the benefits of IoT, the process equipment should not be duplicated.

B. Security

One major concern for adopting IoT within an industrial automation context is how to protect privacy and create a secure environment [10], [11]. Unfortunately, privacy and security are the challenges that has received least attention in the past but with recent attacks this topic has escalated. For example IP cameras being used to create denial of service attacks, smart appliances leaking WiFi passwords, unauthorized remote control of TV streaming devices, and unauthorized control of an Internet connected car. These attacks will in the near future have the potential to include industrial equipment, since it is becoming more connected and exposed. Adopting existing information security concepts to Industrial IoT systems is not straightforward. It needs to be stressed that there are many differences between classical IT systems and those used in Industrial IoT. In classical IT enterprise systems, integrity and confidentially are primary protection goals but within Industrial IoT, availability is a fundamental requirement. For instance, if a cyber-attack occurs, affected IT systems are typically temporarily disabled and then restored after the attack. However, this approach cannot be applied in industrial IoT due to resource constraints and also from an availability point of view. Although there do not exist any silver bullet that can effectively mitigate every possible cyber threat but it is important to implement security in multi-layer, i.e., from device level to system level.

C. Reliability and availability

The most important requirement for an industrial automation system is availability. I.e., the plant must be able to produce or construct what it is intended to do with minimal downtime. Downtime means no revenue and only cost for the plant owner(s). In order to maximize the availability, it is critical that the industrial network must support high reliability. Todays Distributed Control Systems (DCS) can in the extreme case deliver a reliability of 99.9999% [16]. It is important to remember that availability is the most important requirement, but in order to achieve safe and economical viable production there are also requirements on deterministic systems in order to achieve safe production facilities as well as to maintain extremely low variations in production quality to be attractive on the global market. This high availability is required from the operational layers (two bottom layers) of the automation pyramid. This is the critical core functionality of the plant automation system and exactly where IoT devices are envisioned to be injected. If IoT is to be introduced at this level or as part of the process control, it must fulfill these requirements. This is the main reason why cheap "lick and stick" IoT devices have difficulty replacing industrial process control devices.

D. Latency and jitter

It is predicted that more or less all IoT edge connectivity will be through wireless communication. Moreover, it is expected that hundreds or thousands sensors, actuators and PLCs will be connected to the Internet in the future and the grand challenge will then be to achieve reliable and deterministic data transfer and analysis in real-time. With the introduction of Internet connectivity it is also important to understand what is going on at different protocol layers. Currently most existing low-power wireless devices are based on the IEEE 802.15.4 standard (WSNs) or IEEE 802.11 (WLAN) standards, which adopts carrier sense multiple access with collision avoidance (CSMA/CA) as medium access control (MAC). It is well known that CSMA/CA is not designed for low-latency applications due to the unpredictable time delay being generated by random distribution of the back-off time [15]. Hence, the major challenge lies in the Transport protocol layer which major goal is to guarantee reliability and to perform end-to-end congestion control. The most common connection-oriented Transport protocol used in the Internet is the Transmission Control Protocol (TCP), which is inadequate for IoT for several reasons [2] and introduces quite substantial amount of latency. For instance, the connection setup will most likely be unnecessary since many sensors and actuators only transmit a small amount of data, and the setup phase would then last for a large portion of the session time. Furthermore, TCP uses end-to-end congestion control and within IoT this might render in performance reduction since most of the IoT communication is predicted to take place over the wireless medium, and it is well-known that TCP suffers performance losses over wireless. Furthermore, if the amount of data to be exchanged in a single session is very small, TCP congestion control is useless, given that the whole TCP session will be concluded with the transmission of the first segment and the consequent reception of the corresponding acknowledgment. Normally, this could be easily avoided by using UDP, which does not have these problems. However, all of the major cloud providers (e.g., Amazon, Microsoft, Google) requires the use of communication protocols built on top of TCP to connect to their cloud services (e.g., MQTT, AMQP, HTTPS).

To perform closed-loop control with IoT the aforementioned problems needs to be solved. Moreover, today's control systems are designed in a way to offer deterministic communication and computation, but with the introduction of IoT a certain level of unpredictability in sensor readings, packet delivery or

processing time will complicate closed-loop performance and stability.

E. Scalability

The rapid growth of embedded technologies is leading to enormous deployment of miniaturized devices (sensors, actuators, etc.). As the number of devices grows, the data produced by these devices grow unbounded. Thus, handling the growth of number of devices and information they produce is a massive challenge in IoT. Moreover, future IoT systems must be adaptive and scalable through software or added functionality that integrates with the overall solution and this is a major challenge. It will become crucial to have an identification, addressing and naming scheme that supports and scale with large number of devices.

F. Interoperability

Today the industry is dominated by proprietary solutions (and interfaces) from many different vendors and domains. In a fully interoperable environment, any IoT device would be able to connect to any other device or system and exchange information as desired. In practice, interoperability is far more complex. Interoperability between IoT devices and systems happens in varying degrees at different layers within the communication protocol stack between devices. Furthermore, full interoperability across every aspect of a technical product is not always feasible, necessary, or desirable. The heart of interoperability discussion for IoT is the standardization and adoption of protocols that specify communication details. An example of interoperability standardization is the OP unified architecture (OPC-UA) [14], which addresses interoperability throughout the automation pyramid. The level of interoperability varies, but on level 1 of the automation pyramid (Sensors & Actuators), interoperability is often defined as completely interchangeable devices. That is, a temperature sensor from vendor A is completely interchangeable with one from vendor B. This interoperability is enforced by standards and required by plant owners. This is very different from interoperability defined as using the IP protocol, which does not guarantee devices from vendor A and B can be interchanged.

G. Standardization

The rapid growth of IoT makes standardization efforts difficult although standardization plays an important role for further development and spread of IoT. Standardization in IoT aims to lower the entry barriers for the new service providers and users, to improve the interoperability of different applications/systems and to allow products or services to better perform at a higher level. If we look at the IoT standardization landscape today, it is clear there is significant fragmentation of effort and overlapping activities. International standards should be the preferred approach for standard activities that cross domains, functionalities and requirements elaborated at an international level. Today there is various ongoing standardization activities such as Internet standards via IETF, horizontal standards from ITU, IEC, ETSI and IEEE, web standards from W3C, and horizontal common service standards from oneM2M. In addition, industry-specific guidelines or standards for implementing IoT in industrial environments are also recommended for easier integration of various services.

H. Sustainable Business Models

In order for IoT for industrial automation to become a success, the need of sustainable business models will be very important [7] and several things such as supply of the technology itself, proven business models which link supply to demand, and a strong market demand needs to come together. With the increasing trend of using data mining or machine learning to create new values for the customers, an interesting and important question arises - when the equipment manufacturer connects the devices to Internet, who owns the data and/or how should the manufacturer charge his customer for the the data?

Most of the existing IoT solutions today are divided into three value-creating layers: i) Sensor/Actuator; ii) Connectivity (wired/wireless); and iii) data analytic (big data). To make an attractive business model, these layers cannot be isolated, they need to be somewhat integrated to create additional value for the customer. It is essential that hardware is developed in close interconnection with Internet solutions. Another growing trend in the context of healthcare and smart cities is the use of *zero-capex* business model [8]. For example in healthcare, with decreasing reimbursement and increasing economic uncertainty, there is a growing trend for hospitals to opt for a zero capital expenditure, managed service model with the purchase of IT systems in healthcare. Managed services refer to a model where the vendor owns the IT infrastructure, with the hospital paying a fixed fee per month based on projected examination volumes. The vendor is also fully responsible for maintaining this infrastructure, providing data storage and software on a subscription basis. The benefits of managed services include reducing the need for heavy capital investment in IT, such as costly in-house IT support staff and IT infrastructure investment. It is questionable is such business model will work for industrial automation.

In future it also anticipated that cellular IoT which falls under the umbrella of 5G will be used within the context of industrial automation. By tradition, mobile operators business model is *cost per bit* but many IoT industrial applications will have a rather low per-bit value since the application do not send data frequently. However, one can consider three different models [8]:

1) The ”Bluetooth” model, i.e. the consumer and industry purchase the IoT equipment and handle connectivity themselves.
2) The ”WiFi model”, i.e. local connectivity is provided and managed by an IoT operator which charges the customer on a regular basis.
3) The ”Cellular operator model”, i.e., global connectivity by 4G/5G is provided by a mobile operator.

All the aforementioned models have their pros and cons. Hence, the cellular operator model will most likely become