Massive-Scale Automation in Cyber-Physical Systems: Vision & Challenges

D. W. McKee, S. J. Clement, J. Almutairi, Jie Xu
School of Computing
University of Leeds
Leeds, UK
{D.W.McKee, S.J.Clement, ml13jfja, J.Xu}@leeds.ac.uk

Abstract—The next era of computing is the evolution of the Internet of Things (IoT) and Smart Cities with development of the Internet of Simulation (IoS). The existing technologies of Cloud, Edge, and Fog computing as well as HPC being applied to the domains of Big Data and deep learning are not adequate to handle the scale and complexity of the systems required to facilitate a fully integrated and automated smart city. This integration of existing systems will create an explosion of data streams at a scale not yet experienced. The additional data can be combined with simulations as services (SIMaaS) to provide a shared model of reality across all integrated systems, things, devices, and individuals within the city. There are also numerous challenges in managing the security and safety of the integrated systems. This paper presents an overview of the existing state-of-the-art in automating, augmenting, and integrating systems across the domains of smart cities, autonomous vehicles, energy efficiency, smart manufacturing in Industry 4.0, and healthcare. Additionally the key challenges relating to Big Data, a model of reality, augmentation of systems, computation, and security are examined.


I. INTRODUCTION

The current wave of computing is the era of the Internet of Things (IoT) [1]; Edge [2], Cloud [3] and Fog [4] computing; as well as Big Data [5] with Deep Learning [6] and high performance computing (HPC) [7]. However in order to look towards applications such as Industry 4.0 [8], the Internet of Everything (IoE) and Anything (IoA) [9], and beyond where every part of society and industry is digitally integrated there are significant challenges that must addressed. Therefore, this paper presents the set of core challenges that must be addressed to achieve this level of digitisation and automation.

Although the technologies that have been developed in each of these domains have provided significant advances in enabling System of Systems (SoS) to be integrated together in a holistic fashion, there are still significant limitations. In order for systems to be integrated across smart cities, autonomous vehicles, IoT, smart manufacturing, healthcare, as well as the aerospace, defence, and finance industries there must be a concerted effort to develop techniques to handle the explosion of big data streams [10], [11].

Further, the integration of these systems in an automated environment requires a shared model of reality. Specifically in order to enhance the cyber-physical systems that exist within each of the domains there must be a method for providing a set of shared perspectives on reality that can be integrated with simulation and decision support systems via the Internet of Simulation (IoS) [12]. Achieving this will require a significant undertaking to provide a set of unifying standards to integrate both the existing and future technologies [13].

Additionally the service economy [14] will continue to act as the cornerstone for these developments, specifically referring to services and micro-services from a Service Oriented Architecture (SOA) perspective [15]. These services may be hardware systems or devices, human individuals, Cloud hosted software (SaaS), or even simulations (SIMaaS). The aggregation or composition of these services into workflows and subsequently the workflows into services (WFaaS) will provide a scalable approach to augmenting existing systems [12].

To facilitate each of these aspects the trends of Cloud, Edge, and Fog computing [2]–[4] will have to be pushed to their limits with extensive virtualisation to abstract away from individual cloud or HPC providers [7]. The communication infrastructure between systems, such as 5G and LTE [16], along with Software Defined Networks (SDNs) will have to be advanced to improve reliability, bandwidth, and security.

The remainder of this paper is structured as follows: in Section II the motivation for SoSs integration and automation is presented with the state-of-the-art across a range of domains. In Section III the challenge of the Data Explosion that will be experienced is discussed which is due to the expansion of autonomous IoT systems. Additionally, the need for a Model of Reality for autonomy is discussed in Section IV. Following is the challenge of Augmenting existing systems. Then in Sections VI and VII the respective computational and security challenges are discussed. Finally some conclusions are presented in Section VIII.

II. BACKGROUND AND MOTIVATION

In recent years there has been a paradigm shift in the computing landscape towards distributed computing, both in the forms of low power IoT devices [17] and the availability of cloud computing [18], [19]. The IoT is described by Gubbi et al. [1] as digital technologies facilitating the interconnection of components, devices, and services at a
large-scale across a network. Context-aware computation and smart connectivity allow intelligence to be incorporated into IoT [17], and growing to become an Internet of Anything and Everything [9]. This augments Cloud computing with the notion of Edge computing [20] mitigating the need for transferring and processing data in the cloud, instead data is processed much closer to its source.

Advances in cooperative robotics towards autonomous systems [21] are also augmented by the widespread availability of cloud computing. Cloud robotics is an emerging field combining the research areas of cloud computing and robotics to provide services to robots and facilitate robot interaction [21]–[23]. This includes utilising cloud services for robotics [24]–[26] and robots themselves providing services [27]. These research areas are already being applied to a number of domains and the future prospect of these applications is a wide-scale adoption of automated, intelligent systems as part of public life and economic development. The primary domains focussed on in this paper are manufacturing and infrastructure though there is also significant scope and research for automation in the domains of defence and security, aerospace and finance.

Within the next 10-15 years we anticipate that there will be ubiquitous, intelligent networks and computing managing and augmenting most of systems we interact with on a daily basis.

A. Smart Cities

Despite the concept being the latest trend for urban planning, smart cities have no concrete definition in the literature [28]. In general a smart city describes a cyber-physical SoS heavily reliant on intelligent autonomy and IoT [29], [30]. These systems making up the smart city will automatically manage power and communication infrastructure, environment, traffic and other aspects of the city for the benefit and well-being of its inhabitants through ubiquitous sensing and embedded intelligence. This includes robotics for repair and maintenance [31], [32], driverless transportation [33] and power management [30] among others.

B. Autonomous Vehicles

Driverless cars are probably the most publicly visible autonomous systems that are currently being developed. Some of the foundational technologies for these systems are already deployed as part of Advanced driver assistance systems (ADASs) while others are still under development [34]. Of particular note is the possibility for interconnected vehicles, an Internet of Vehicles or a Vehicular Cloud [35]. The interconnection of these systems has the potential to allow for holistic traffic management and the use as a service of the data generated by the vehicle. This trend is not only limited to vehicles however, ships [36] and aerial vehicles [37] are also being automated for similar tasks.

C. Power and Energy Efficiency

Two key areas in this domain are smart grids and efficiency. Intelligent, distributed power generation is the evolution of the existing power grid infrastructure [38]. The vision of smart grids is one where intelligent, demand-side systems manage smart, renewable energy generation combined with energy storage.

Given that many of these autonomous systems depend on data-centers for processing, the efficiency of data centers is crucial as demand increases [39]. Improving the efficiency of data-centers requires intelligent scheduling [40] and modelling of workload patterns [41]. Further efficiency gains are possible by utilising the waste heat generated by the data-center [42].

D. Smart Manufacturing & Industry 4.0

A 4th industrial revolution dubbed Industrie 4.0 is emerging, fuelled by the integration of intelligent automation into the manufacturing value chain [43], [44]. The key characteristics and technologies driving this change are the adoption of IoT devices in the manufacturing process, also known as the Industrial IoT [45]. This leads to smart factories that are able to flexibly adapt to changing demands in the marketplace [8]. The data streams generated by the interconnection of large numbers of autonomous systems within a factory will allow it to gain a level of self-awareness, calculating machine health, behaviour and self-optimising operations [46].

E. Health and Well-being

The application of autonomous systems to the domain of healthcare is growing. The adoption of evidence based medicine [47] and the widespread record keeping of the medical community provides opportunities to apply big data analytics to the field [48]. There is also large amounts of additional health data being generated by the marketplace of wearable health devices within IoT [49]. Security of this online, personalised health data has become an increasing concern [50] and the move towards blockchain record systems [51] aims to facilitate the secure sharing of patient records. Additionally, there has been a move to utilise robotic systems in patient care to reduce the demand on healthcare services [52].

III. CHALLENGE: DATA EXPLOSION

The increasing variety and number of data collecting devices joining the IoT have fuelled the Big Data trend. Big data analytics provide techniques for the analysis and visualisation of extremely large datasets [53]. Specifically these data sets are too big to store on a single machine and so must be distributed. Already the growth of data is exponential [54] and increasing data collection and further cloud services will only accelerate this further [55]. Very quickly this could lead to a situation where we are no longer able to process the vast amount of data being collected.

This data explosion is being driven partly by the growth in IoT and the large-scale collection of data. It is envisioned that the ubiquitous collection of data will enable machine learning techniques to provide models that can respond to the growing demand for intelligent autonomous systems [11]. IoT promises
ubiquitous sensing and a network of data driven devices that is unprecedented today.

A number of problems are presented by the oncoming explosion in data generation. Firstly, the size of data that is being generated may be too large to store in a dataset for further processing. The amount of potential data being generated by a ubiquitous IoT will easily overwhelm current network and storage infrastructure. Instead, we may be forced to rely on stream processing to collect relevant information from sources and discard the rest [10], [56]. Secondly, the large number and variety of data sources may invoke the curse of dimensionality where it is unclear which data streams to process and which to discard. Given the large number of potential data sources in the IoT it may also be difficult to manually generate meaningful features for conventional machine learning techniques. Instead, automatic generation of features and intelligent dimensional reduction to filter data for relevant information will be vital research areas in the coming years to mitigate the effect of the big data explosion.

IV. CHALLENGE: MODEL OF REALITY

As more autonomous systems are deployed into the various domains detailed in Section II, the demand for intelligent automation increases. This effect is seen most clearly in the manufacturing domain where intelligent automation is now being applied to production. With the development of intelligent cyber-physical systems it is no longer enough for the system to automatically respond to the environment. In a broad sense, these systems must now anticipate future scenarios in dynamic environments. For example, autonomous vehicles must predict the future positions via trajectories of all moving objects around them in order to avoid collisions; smart factories must predict demand and equipment failure; and smart homes might predict their inhabitants behaviour.

The basis of the field of machine learning is the training of models based on available data in order to predict or classify inputs. However this may not be possible for all sceneries and environments that autonomous cyber-physical systems are being deployed in. In cases such as autonomous vehicles, collision avoidance will be based on predictions grounded in physical models. Where physical systems are complex enough, simulation may be required to support the decisions made by these systems. Simulations have the benefit of being able to model more complex interactions than simple mathematical models.

With simulations there are trade-offs between the detail of the simulation, the speed of execution and the accuracy of the results, for example 1D vs 3D simulation. In a safety
critical system that responds in milliseconds the detail and scope of the simulation might be reduced to ensure a timely response. In other applications a more detailed simulation may be employed, though this may require large amounts of computing power. In certain cyber-physical systems, especially mobile systems, where power usage or weight is a concern, it may be necessary to utilise cloud or HPC computing for simulation. Just as the IoT allows the interconnection of devices, the IoS [12] could allow the interconnection of simulation and provide the detailed decision support and predictive power that intelligent automation systems require (see figure 2). One proposed benefit of this approach is the ability to construct large co-simulations from constituent parts, mitigating the difficulty of development associated with large scale simulation [13], [57].

There are however a number of unresolved barriers to the implementation of IoS. Primary among them is the problem of simulation integration [58]. Bringing together an arbitrary set simulations remains infeasible for a number of reasons. Differing levels of fidelity in simulations mean that accuracy may be sacrificed. Simulations might utilise incompatible data types or representations. The simulation may not scale to the required level [59]. Even the execution methods or timesteps of simulations may not be compatible with each other or the proposed platform. Additionally, simulations are often created from a specific viewpoint and any two differing viewpoints may not be compatible. There are a number of standards that have been developed for this problem such as DIS [60], HLA [61], [62], FMI [63] and FDMU [64]. However, none of these satisfy all requirements for the proposed usage above [65], [66].

V. CHALLENGE: AUGMENTING EXISTING SYSTEMS

The use of services and the shared model of reality between the various systems provides a foundation for augmenting existing systems with additional functionality. But a crucial aspect of enhancing existing systems with automation and intelligence is their continued operation. Most significantly, from a smart city perspective, the augmentation of existing systems must facilitate the city’s growth without interfering with the operation of any of its vital systems.

Currently a service marketplace can be used to facilitate the discovery and integration of web services into workflows [15]. However as shown in Figure 1 the services, systems, devices, and individuals from across the different layers of a city must be digitally integrated together.

The challenge of integrating already existing systems remains challenging due to the lack of compatible standards, as discussed in the previous section, and becomes even harder with the need to augment those existing systems with the model of reality and the huge amount of data being derived from these systems. As depicted in Figure 1, the existing city and compute infrastructure must be integrated. This includes the conceptual and business layers of SOAs [15] - along with the layers of Cloud computing: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). An additional challenge is the ability to automatically re-factor services to ensure continuous compatibility with future versions and systems [67], [68].

Further the augmentation of the existing systems with simulations as services (SIMaaS), from the model of reality, facilitates decision support as well as prototyping and product testing from an Industry 4.0 perspective [12]. Combining these with workflows as services (WFaaS) provides an extensible means for augmenting the existing systems.

VI. CHALLENGE: COMPUTATIONAL SYSTEMS

The Cloud computing era provided a mechanism for computation and data processing to be performed off-site at a low-cost. However, the centralised nature of the Cloud introduces significant limitations due to communication bandwidths and latencies. Therefore, with the availability of smaller and cheaper but yet powerful compute devices, the Edge was born to bring the processing back to the devices themselves [2]. However, the future level of processing required along with the need to connect with and share with other system’s data and models of reality requires the power of centralised Cloud and also high performance computing (HPC) [69].

Therefore we now have the emerging hybrid paradigm of Fog computing [4], [70]. In order for this paradigm to successfully achieve integration of these systems the reliability of every aspect must be managed and guaranteed with a level of Quality of Service (QoS). This includes managing the communication infrastructures, particularly wireless communication technologies such as the successors to 5G [16]. The Fog paradigm of providing a virtual layer between the data centre and the IoT devices must be extended to provide a virtual cloud that hides the identity or location of all data centers, but also encapsulates other compute resources such HPC facilities, as shown in Figure 2. Such an approach could be used to mitigate the issues of scalability, fault-tolerance, elasticity [2] as well as facilitating management services to detect failures [70].

The computational infrastructure is therefore going to have to evolve to become a self-adaptive ecosystem that learns and predicts system performance. Somehow it must also be to a certain degree technology agnostic, allowing both digital and physical systems as well as human individuals to act and be modelled as services.

VII. CHALLENGE: SAFETY AND SECURITY

Some of the major challenges in existing systems are those which relate to the safety and security of those systems. From the security perspective numerous approaches have been proposed for either increasing the level of system security or improving the practicality of those security approaches with regards to performance limitations. A detail review of security for IoT is provided by Jing et al. [71].

One particular challenge with IoT and the continued increase in use of data-centers will be to find methods to inhibit DDOS attacks from IoT devices. For example the Mirai attack demonstrated the use of unsecured services - via HTTP, tent,
Fig. 2. Cloud layers of abstraction for IoT and IoS

and SSH server - to gain remote control of devices and load malware into memory [72]. Additionally Kirner demonstrated the ability to remotely execute code on IP CCTV and DVR devices [73]. These highlight the need for a concerted effort to resolve these, and many other, security issues.

A current trend in Cloud security is performing computation on encrypted data, using homomorphic encryption, this can provide a significant performance improvement by removing the need to encrypt and decrypt data in the Cloud [74]. These methods have been extended with “somewhat” homomorphic encryption to improve performance, but at the cost of limiting the data values and types that can be processed [75]. These approaches are currently limited to tasks such as searching, sorting, and arithmetic operations where the encryption process is order-preserving [76]. It is anticipated that future techniques may be based on homomorphic encryption and involve a mixed level of hardware and software processing. It is vital to consider that in order to facilitate the wide-scale integration of systems there would have to be a clear set of security standards shared across systems and devices.

Another area that should be considered is the use of Quantum-Key distribution techniques [77] where it can be immediately identified if an individual is listening in on the communications, and as developments allow for these techniques to be used over greater distances they are likely to play a part in defining security standards and protocols. Additionally the use of blockchains [78] is another approach that is gaining interest in order to improve the security of communication systems in the domains of finance and healthcare in particular [51].

As processing is moved towards virtual clouds there remains a challenge to certify the security and certain conditions that will be maintained for data and processing. One such is the multi-tenancy of those systems whereby guarantees are required to be in place such that certain organisations cannot not be multi-tenant on servers with other specific organisations, or organisations from the same domain [79], which is particularly prevalent in the banking industry.

Finally there is also a challenge of guaranteeing the safety of these systems, in adherence with standards appropriate for each domain. Therefore there must be mechanisms to consistently and automatically evaluate the safety of any given system [80] and at a SoS level adapt as the safety expectations degrade. The use of provenance and data analysis to evaluate the performance of the system will be critical to providing an effective safety assurance mechanism which is able to identify potential faults before they become problems [81].

VIII. CONCLUSION

The existing technologies of Cloud, Edge, Fog computing and Big Data across the domains of IoT, smart manufacturing with Industry 4.0, smart cities, autonomous vehicles, and healthcare are facilitating the integration of anything and everything. However, these technologies are not currently
adequate to facilitate the integration of all the systems due to incompatible standards and protocols.

Additionally the successful integration of these cyber-physical systems in an automated fashion will require handling an explosion in data, in particular the rate and scale of data streams that must be processed. Further the collection of the data along with the integration of simulations and workflows as services (SIMaaS & WFaaS) requires a shared model of reality. This in turn may facilitate the automated augmentation of existing systems, across all existing layers of a city and the computational infrastructure, supporting the drive towards Industry 4.0.

The computational systems - including Cloud, Edge, and HPC - must be homogenised as a virtual hybrid cloud which also manages both wired and wireless communication infrastructures for required levels of reliability and QoS. And finally the significant challenges of managing data security, using techniques ranging from homomorphic encryption to quantum-key distribution, must be urgently addressed. There must also be mechanisms for ensuring the compliance with necessary safety protocols and the development of automated techniques for continuous evaluation of compliance as service performance may degrade over time.

It is anticipated the next step towards facilitating the complete integration of systems and services from across smart cities will involve the extension of the Internet of Things (IoT) with the Internet of Simulation (IoS).

ACKNOWLEDGMENT

This work is partially supported by the China National Key Research and Development Program (No. 2016YFB1000101 and 20016YFB1000103).

REFERENCES


