The Case for Energy-Aware Simulation and Modelling of Internet of Things (IoT)

(Invited Paper)

Matthew Forshaw, Nigel Thomas
School of Computing Science
Newcastle University
Newcastle upon Tyne, UK
{matthew.forshaw,nigel.thomas}@ncl.ac.uk

A. Stephen McGough
School of Engineering and Computing Sciences
Durham University
Durham, UK
stephen.mcgough@durham.ac.uk

ABSTRACT

Existing approaches to energy management of large scale distributed systems are ill-equipped to handle the challenges introduced by the dynamic and self-adaptive nature of the Internet of Things. In this position paper we motivate the need for energy-aware modelling and simulation approaches for IoT infrastructures, to facilitate what-if analyses, and support design decisions and runtime optimisation. We identify open challenges and research opportunities in the energy-aware simulation and modelling of IoT.

Keywords

Simulation, Modeling, Internet of Things, Energy

1. INTRODUCTION

The energy impact of IT infrastructures is a significant resource issue for many organisations, as energy costs now dominate IT infrastructure total cost of ownership. The Natural Resources Defence Council estimates that US data centres alone consumed 91 billion kilowatt-hours of electrical energy in 2013 – enough to power the households of New York twice-over – and this is estimated to grow to 139 billion kilowatt-hours by 2020. These financial and ecological challenges are further compounded by social and political factors and strict environmental legislation governing organisations, making improvements to energy-efficiency of paramount importance.

As a consequence of increased scrutiny of the energy impact of these systems, aggressive power management policies are often employed to conserve energy, but in doing so these policies severely restrict the operation of such systems. One example of such infrastructures is the Internet of Things (IoT), a paradigm gaining increasing traction in the area of modern wireless communications. In IoT systems, *things* or *objects* – which may include RFID tags, sensors, and actuators – are organised into self-configuring and adaptive systems capable of cooperating with each other to reach common goals. Numerous potential application areas for IoT are identified, including environmental monitoring, e-health, intelligent transportation systems, military and industrial plant monitoring [1]. It is anticipated that 50 billion devices will be linked by 2020.

The rest of this position paper is organised as follows. In Section 2 we motivate the need for energy-aware management in IoT, and describe a typical IoT infrastructure in Section 3. Section 4 presents existing performance evaluation methodologies for large-scale systems, evaluating their applicability to IoT systems. Section 5 identifies open research challenges in the modelling and simulation of IoT systems, before we offer future directions in Section 6.

2. MOTIVATION

In large-scale sensor network deployments, whose operation are constrained by battery lifetime, and replacement is infeasible or prohibitively costly, energy conservation is a primary optimisation goal yielding significant operational and financial benefits. The energy consumption and lifetime of devices is highly dependent on a number of factors, including sampling rate, computational workload, and environmental factors. Hence, the efficient management of the heterogenous resources comprising an IoT environment are of key importance.

Early works in energy-efficient computing focused on such battery- and computationally-constrained devices; however, these approaches are ill-equipped to handle the challenges introduced by the dynamic and self-adaptive nature of Internet of Things (IoT) infrastructures. Many existing approaches, including Dynamic Voltage and Frequency Scaling (DVFS) and Dynamic Power Management (DPM), focus on energy-aware decision making at a single sensor level. However, in an IoT setting, ensemble-level decision making leveraging global knowledge promises favourable results in federated environments spanning organisational boundaries. Further complexity is present in systems exhibiting hard or soft real-time requirements e.g. *critical infrastructures*, such as healthcare, fire detection, and industrial controllers.

Determining optimal management policies is a complex process as system behaviour can often be difficult to predict *a priori*, and often exhibit emergent patterns only present at large scale. Hence, the ability to model and evaluate the efficacy of proposed energy-aware approaches in a quantifiable, repeatable and controllable manner is sought.

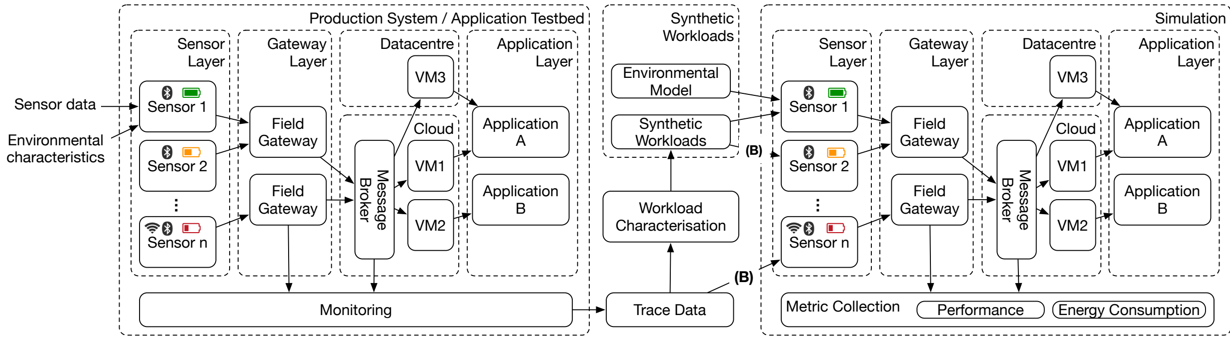

Figure 1: Simulation and modelling of a typical IoT architecture.

3. IoT INFRASTRUCTURES

Here we consider a typical IoT infrastructure, comprising sensors observing the natural environment, field gateways aggregating data from sensors for transmission to cloud resources. We also consider a scenario where there is flexibility over where a given computation is performed, where sensors may or may not include on-board processing capabilities.

Figure 1 exemplifies the potential interplay between a typical IoT architecture and modelling/simulation approaches. A runtime/production system deployment on the left is instrumented to collect monitoring information. Collected trace data can then be either; *(A)* replayed through a simulation environment or *(B)* be characterised to form the basis for synthetic workloads.

4. EVALUATION METHODOLOGIES

Monitoring and changing a live system is often not a practical solution. There are cost implications in doing so, configuration overheads limit the number of scenarios to test, and it normally requires significant time in order to fully ascertain the long-term trends. There is also the risk that any changes could lead to detrimental impacts, either in terms of the functionality of the system or in the energy consumed.

Three primary methodologies dominate the performance evaluation literature; namely experimental testbeds, emulation and simulation [6]. Here we discuss the relative merits and limitations of each approach applied in the context of Internet of Things (IoT) systems.

Experimental testbeds: Experimental testbeds are frequently considered for the performance evaluation of large-scale distributed systems. Testbeds seek to facilitate the evaluation of IoT solutions on actual hardware, operating under realistic environmental conditions.

Practitioners may opt to use an existing experimental testbed, or construct their own private testbed. There exists a trade-off between the capital investment to acquire the required hardware infrastructure and operational expenditure of using an external service. However, when considering the use of testbeds for the evaluation of energy efficiency, the domain is dominated by private testbeds, with very few public infrastructures reporting energy metrics [7].

Emulation: A further approach considered in a number of works is the emulation of large-scale systems. In an emulation approach, performance evaluation is carried against the concrete implementation of the system under test, rather than a simulated implementation. Such an approach boasts a number of key benefits, alleviating the need for an abstract model for the system required in simulation or analytical approaches, and allowing the same code used for experimentation to be deployed into a production environment.

A significant constraint on emulation-based experiments is that of scale, with emulations frequently shown to be capable of evaluating systems with orders of magnitude fewer entities. In our context of large-scale IoT systems, many of the operating decisions and policies we propose may only be evaluated meaningfully at scale, so we do not pursue an emulation approach further.

A Case for Simulation and Modelling: The use of a modelling and simulation approach is of particular interest in this context of IoT infrastructures comprising heterogeneous resources and decentralised decision making, where neither invasive instrumentation of the infrastructure, nor longitudinal measurement, may be assumed. It also offers the potential for much faster turn-around and feedback, along with the ability to evaluate the impact of many different options simultaneously.

4.1 Phases of Evaluation

We consider the potential for evaluation efforts to support IoT systems at a number of levels; *design*, *policy*, *capacity planning* and at *runtime*.

At the *design* phase, evaluation efforts are capable of providing insights into emerging system configurations which have not yet been implemented. At the *policy* level, evaluation considers management policies for an existing architecture. *Capacity planning* allows system designers to reason over the number of resources required to meet certain performance guarnatees and SLAs. Further opportunities exist for simulations to run in tandem with production environments to inform runtime optimisation decisions.

Table 4 summarises the applicability of each approach to these settings.

Phase	Testbed	Emulation	Simulation & Modelling
Design	✗	✗	✓✓
Policy	✓	✓	✓✓
Capacity	✓	✗	✓✓
Runtime	✗	✓	✓

Table 1: Comparison of Evaluation Methodologies

5. OPEN RESEARCH CHALLENGES

Although energy-aware modelling and simulation-based approaches have been applied elsewhere in large scale computing, e.g. Wireless Sensor Networks (WSNs) [10], many issues have not been fully addressed. Further issues arise as a consequence of the increased scale and heterogeneity of IoT. Here we identify a number of key challenges in adapting simulation and modelling approached to IoT contexts. The list here is not intended to be exhaustive, but rather it represents a set of current considerations that will have an increasing impact on IoT systems as scale, complexity and heterogeneity increases.

5.1 Modelling challenges

Here we present challenges in applying existing workload characterisation and modelling approaches to IoT.

5.1.1 Holistic cross-layer modelling

Many existing approaches consider energy saving schemes at a single level of the infrastructure; e.g. sensor, field gateway or cloud infrastructure levels. Hence, it is often difficult to quantify the effect of an energy policy applied one level, on the rest of the infrastructure. We advocate a holistic approach, capable of quantifying the energy and functional impact of a policy throughout the entire infrastructure. While many of the constituent models exist within the literature (e.g. predictive models for energy usage in commodity and server hardware), these must be combined to provide this unified view. Particular attention must be paid to the complex dependencies between devices and applications at each layer of the system.

For example, typical cloud resources host systems such as message brokers (e.g. MQTT, Kafka, ActiveMQ), distributed stream processing engines (e.g. Storm, Spark) and complex event processing (CEP) engines (e.g. Esper), whose behaviours are complex and strongly linked to the offered workload [2]. Understanding the impact of workload on these systems, and consequently on hardware subsystems (CPU, network, disk) is the focus of active research interest.

Such an enhancement to the state of the art offers compelling applications in supporting numerous infrastructure management decisions, including; *a)* energy-aware autonomic deployment of IoT workloads [8], *b)* energy harvesting [11] leveraging alternative energy sources in the environment to enhance service life, *c)* offloading computation [8] from sensors to field gateways or cloud resources.

5.1.2 Application modelling

Many candidate approaches to reduce the energy consumption of a sensing component (e.g. batching and aggregation of samples, lowering sampling rates, discarding samples, relaxing precision of computation) have a direct impact on the volume, quality and timeliness of data made available for analysis by IoT applications.

The tolerance of a given application to such delays has a demonstrable impact on the potential for energy saving across an IoT infrastructure. For this reason, the challenge of energy-efficient management is exacerbated in applications with real-time requirements. It is crucial to understand the impact of energy-aware management schemes on applications running across IoT infrastructures. We seek to quantify workload sensitivity, hereby maximising energy-savings while retaining inferential quality of the data collected.

5.1.3 Workload models in Data-intensive systems

The characteristics of real-world workloads in IoT systems is currently poorly understood. There is a strong need for a focus on workload modelling, to acquire a corpus of trace data from production IoT systems, against which synthetic workloads may be derived [3].

Key to the successful characterisation and modelling of systems is data collection. While for traditional server-based systems, workload logs are routinely stored, this is rarely each case at all levels of IoT infrastructure. There may also be the desire to capture data pertaining to special environmental effects, adding context to infrastructure decisions. For data-intensive IoT systems, involving thousands of messages and events per second, persisting all data to storage would be infeasible, and would lead to observational effects and *interference* to the running system.

A considered approach to the instrumentation of systems under test to reduce the impact of monitoring is required. Approaches to subsample an offered workload while retaining features of interest rarely result in representative traces, and in many cases would still result in significant data volume or extra energy requirements.

Workload characterisation at each level of the system poses unique challenges, particularly for bandwidth-, energy- and computationally-constrained devices. In many situations, it would be insufficient to collect data only at the field gateway or cloud level. For example, smart sensors may discard measurements to reduce data transmission, which would then otherwise be lost from traces. One solution would be to perform preliminary characterisation of the workload trace data during runtime, transmitting only this summary data, with consideration the computational cost of the aggregation, observer interference, and extracting data at the level of detail required to recreate representative traces.

5.1.4 Generalisable modelling

The energy consumption of sensor and server hardware has been studied extensively in the literature. Early works leveraged low-level metrics such as performance counters when developing predictive models of energy consumption [5]. Many approaches required significant architectural knowledge and typically were not generalisable to other hardware, nor scalable to model large infrastructures.

Further issues are evident when conceptual misunderstandings arise. Naicken *et al* [9] observed significant inconsistencies between results produced by multiple simulation frameworks modelling the same distributed environment, and attribute this variability to inconsistencies between underlying abstract models and implementations.Hence, it is highly desirable to derive a taxonomy of IoT device abstractions, incorporating a generalisable models of energy consumption and performance for IoT devices.

The introduction of abstract models has important consequences for enabling energy-aware autonomic deployment of IoT workloads [8]. Where, conventionally, applications were unaware of their real-time energy impact, this would provide the possibility for a common abstraction against which applications can be developed. This could provide fine-grained metrics for energy-consumption, battery, and harvestable energy availability. Applications may then adapt their behaviour dynamically in response to these factors, and system events, to better conserve energy and optimise their activities on the specific platform to which they are deployed.

5.2 Simulation Challenges

Characterisations of offered workloads experienced by production environments and abstract models of system components may now be combined in a simulation context to support *what-if* analysis of new deployment configurations. We now consider challenges in the simulation of IoT systems.

5.2.1 Scalability

Given the anticipated scale of IoT systems, a critical challenge to any simulation approach will be that of scalability. In order to meaningfully evaluate the behaviour of simulated systems, it must be possible to evaluate large-scale systems comprising hundreds of nodes across complex scenarios.

The scalability challenge may be partly addressed by advances in distributed simulation across multiple compute resources. However, a critical issue is that of level of abstraction. It may be infeasible to scale a sensor-level simulation at the finest level of abstraction to analyse thousands of simulated nodes. Similarly, to raise the level of abstraction may obscure effects which are significant at scale.

A methodology is needed to allow the composition of large-scale simulations, incorporating results derived from detailed simulations on smaller components of the overall system.

5.2.2 Composability

The challenges of the integration and composition of multiple simulations are emerging in many IoT application areas. For example, consider a Cyber-Physical Systems (CPS) context, where sensing infrastructure and embedded computational elements are integrated with physical processes and actuators [4]. Continuous-time (CT) simulations responsible for describing physical systems (e.g. industrial actuators modelled as a system of differential equations) lack the required abstractions for software structure and concurrency. Meanwhile, discrete-event (DE) simulations are not always suitable for modelling physical systems.

Here the need for specialised co-simulation approaches arises. Challenges exist in facilitating the coordinated running of multiple simulations, each employing different simulation formalisms, and operating at different levels of abstraction. Understanding the trade-off between fidelity and simulation runtime is required.

5.2.3 Coverage

An overarching goal of simulation-based studies is to achieve good *coverage*, evaluating a system under the various situations and conditions which may lead to deviations in performance and energy consumption. Achieving good coverage is particularly difficult in IoT settings, where there is potential for unanticipated emerging behaviours, and significant impact from uncontrollable environmental factors. This challenge is amplified in the early design stages, and evaluating systems which do not yet have a real-world deployment.

6. SUMMARY AND DIRECTIONS

Internet of Things (IoT) systems exhibit clear operational and financial incentives to reduce energy consumption. In this paper, we identify several key challenges in the modelling and simulation of IoT systems. Key to these challenges is the heterogeneous nature of devices, and the complex dependencies between devices at different levels. We have presented opportunities where a modelling and simulation approach may be used to inform the design, capacity planning and runtime optimisation of IoT infrastructures.

Our ongoing work seeks to address these challenges by developing a methodology for predictive modelling, and reasoning over, energy considerations within heterogeneous IoT infrastructures, with consideration for performance and dependability. These models shall be incorporated into an IoT simulation environment (under development), facilitating the prototyping and evaluation of energy-aware management schemes, and validated within real IoT systems.

Acknowledgment

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/L015358/1].

7. REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[2] T. Cooper. Doctoral Symposium: Proactive Scaling of Distributed Stream Processing Work Flows using Workload Modelling. In *Proceedings of the 10th ACM International Conference on Distributed Event-Based Systems*. ACM, 2016.

[3] D. G. Feitelson. *Workload modeling for computer systems performance evaluation*. Cambridge University Press, 2015.

[4] J. Fitzgerald, C. Gamble, R. Payne, P. G. Larson, S. Basagiannis, and A. Mady. Collaborative Model-based Systems Engineering for Cyber-Physical Systems – a Case Study in Building Automation. In *Proceedings of the 26th Annual INCOSE International Symposium*, 2016.

[5] M. Forshaw. *Operating Policies for Energy Efficient Large Scale Computing*. PhD thesis, Newcastle University, UK, 2015.

[6] M. Forshaw, N. Thomas, and A. S. McGough. HTC-Sim: A trace-driven simulation framework for energy consumption in High Throughput Computing systems. *Concurrency and Computation: Practice and Experience*, 2016.

[7] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo. A survey on facilities for experimental internet of things research. *IEEE Communications Magazine*, 49(11):58–67, 2011.

[8] P. Michalák, S. Heaps, M. Trenell, and P. Watson. Doctoral Symposium: Automating Computational Placement in IoT Environments. In *Proceedings of the 10th ACM International Conference on Distributed Event-Based Systems*. ACM, 2016.

[9] S. Naicken, B. Livingston, A. Basu, S. Rodhetbhai, I. Wakeman, and D. Chalmers. The state of peer-to-peer simulators and simulations. *SIGCOMM Comput. Commun. Rev.*, 37(2):95–98, Mar. 2007.

[10] G. Z. Papadopoulos, K. Kritsis, A. Gallais, P. Chatzimisios, and T. Noel. Performance evaluation methods in ad hoc and wireless sensor networks: a literature study. *Communications Magazine, IEEE*, 54(1):122–128, 2016.

[11] F. K. Shaikh and S. Zeadally. Energy harvesting in wireless sensor networks: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 55:1041–1054, 2016.