

3

Integrating IoT + Fog + Cloud Infrastructures: System Modeling and Research Challenges

Guto Leoni Santos, Matheus Ferreira, Leylane Ferreira, Judith Kelner, Djamel Sadok, Edison Albuquerque, Theo Lynn, and Patricia Takako Endo

3.1 Introduction

There is widespread recognition, from academia, industry, and policymakers, that social media, cloud computing, big data, and associated analytics, mobile technologies, and the Internet of Things (IoT) are transforming how society operates and interacts with technology and each other [1, 2]. Often referred to as the Internet of Everything (IoE) or “Third IT Platform,” these technologies presage a future of greater inter-dependencies between people, devices, and the infrastructure that supports these relationships. Cisco estimates that there are IoE, with estimates of 8–10 billion connected today [3].

While cloud computing has been a key enabling technology for the IoT, a small increase in the percentage of connected or cyber-physical objects represents dramatic change in the feature space of computing and a potential tsunami of computation and hyper-connectivity, which today’s infrastructure will struggle to accommodate at historic levels of quality of service (QoS). Large-scale distributed control systems, geo-distributed applications, time-dependent mobile applications, and applications that require very low and predictable latency or interoperability between service providers are just some of the IoT application categories that existing cloud infrastructures are not well-equipped to manage at a hyperscale [4]. Traditional cloud computing architectures were simply not designed with an IoT, characterized by extreme geographic distribution, heterogeneity and dynamism, in mind. As such, a novel approach is required to meet the requirements of IoT including transversal requirements (scalability, interoperability, flexibility, reliability, efficiency, availability, and security) as well as **cloud-to-thing (C2T)**-specific computation, storage and communication needs [5].

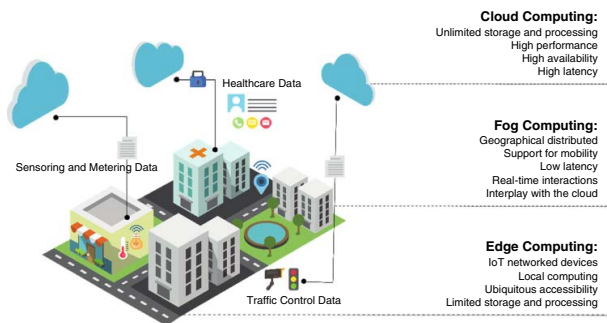


Figure 3.1 Integration of IoT devices with fog and cloud computing.

In response to the need for a new intermediary layer along the C2T continuum, fog computing has emerged as a computing paradigm situated between the cloud and connected or smart end-devices where intermediary compute elements (fog nodes) provide data management and/or communications services to facilitate the execution of relevant IoT applications [6], as shown in Figure 3.1. The ambition for fog computing is greater support for interoperability between service providers, real-time processing and analytics, mobility, geographic distribution, and different device or fog node form factors, and as a result the achievement of QoS expectations [4]. Despite these advantages, fog computing adds a layer of complexity that operators across the C2T continuum need to account for, not least resource orchestration and management [4, 7]. Fog computing represents both an opportunity to exploit but also a risk to mitigate. Not only do failures in the cloud and end points need to be considered but the potentiality and impact of failures across the entire C2T continuum.

In this chapter, we review the literature with regarding to the use of modeling techniques to represent and evaluate an integrated C2T system comprising cloud computing, fog computing, and the IoT (C2F2T). The remainder of this chapter is organized as follows. The next section describes the methodology adopted to guide this literature review and provides a descriptive analysis of the final works selected for use in this chapter. Section 4.3 presents an analysis of existing system modeling techniques used in cloud computing, fog computing, and IoT research against four categories – analytical models, Petri Net models, integer linear programming, and other approaches. Section 4.4 discusses the main scenarios modeled in extant research while section 4.5 discusses the metrics used in evaluation. The chapter concludes with a discussion of research challenges and future directions for research.

3.2 Methodology

The objective of this systematic literature review is to present an overview of academic literature on (i) the use of modeling techniques to represent and evaluate an integrated C2F2T system; (ii) the main scenarios modeled; and (iii) the metrics used to evaluate models. In general, the literature review follows the methodology outlined in [8] and illustrated in Figure 3.2.

While [9] suggests authors aim for complete coverage, such coverage is not feasible. Thus, we limit this review to the computer science discipline and only publication outlets featured in three repositories, namely: Science Direct, IEEE Xplore, and ACM Digital Library. The literature search was limited to modeling, cloud computing, and the IoT. We did not include “fog” in the string, as it was deemed likely that papers on integrated C2F2T featuring fog computing would need to feature cloud computing and IoT keywords. As such,

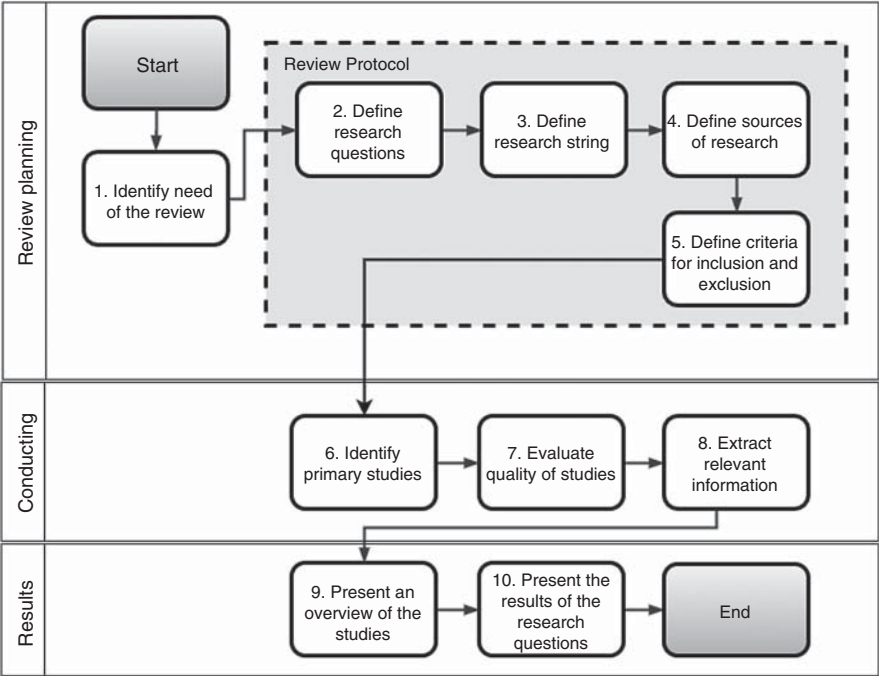


Figure 3.2 Systematic review steps. Adapted from [8].

the search was limited to publications resulting from the following search expression: model AND (cloud OR cloud computing) AND (IoT OR Internet of Things).

Our initial search yielded 1,857 publications from 2013 to 2017. We considered conference and journal papers, but books, PhD theses, and industry publications were excluded. Abstracts were further scrutinized and a final list was produced with 23 relevant articles. Papers were primarily omitted on the grounds that (i) their main focus was not concerned with an integrated C2T system; (ii) the papers were concerned with business modeling; (iii) architectures were misrepresented as models; or (iv) there is a lack of models or modeling techniques. Full texts for the final 23 articles were evaluated to validate that the articles meets the literature search criteria. Table 3.1 presents the number of articles in the initial search and the final selection; while Table 3.2 shows the list of articles selected by year and publication source.

As can be seen in Table 3.2, the largest number of papers was published in 2016 (11), followed by 2017 (5). As cloud computing, IoT, and fog computing are all relatively new fields, high-ranking outlets dealing specifically with the topic are scarce, and those that do exist may not be affiliated with IEEE or may require longer turnaround times for acceptance. Given the increasing number

Table 3.1 Summary of systematic search results.

Repository	Science Direct	IEEE Explore	ACM Digital Library	Total
Initial search	1,244	426	187	1,857
Final selection	10	12	1	23

of conference papers since 2013, one would expect a greater number of journal articles in the coming years.

3.3 Integrated C2F2T Literature by Modeling Technique

In this section, we present an analysis of the modeling techniques used in the sample of papers identified in the conducting phase. A wide range of techniques was identified in the papers analyzed. The analysis presented in Figure 3.3 suggests that analytical models followed by Petri Nets and Integer Linear Programs are the most common techniques used for modeling an integrated C2F2T system.

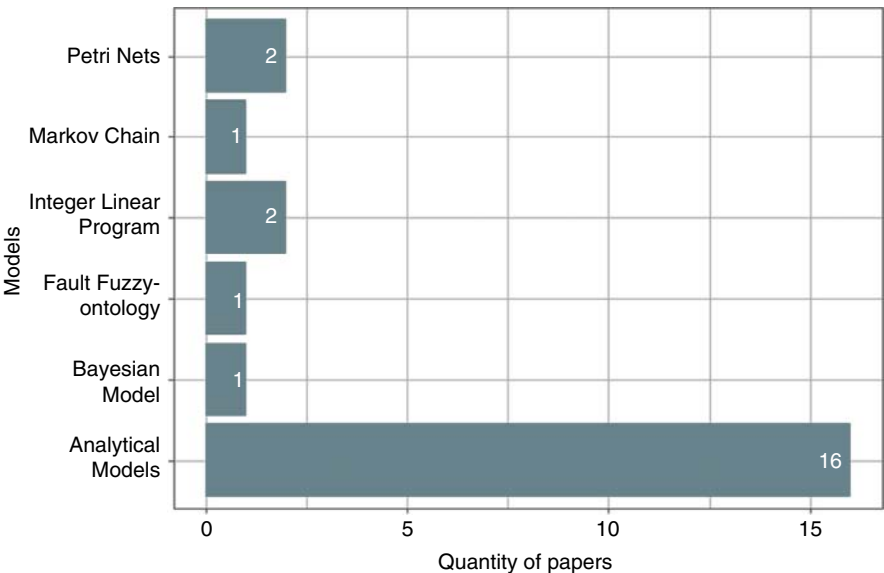


Figure 3.3 The most approaches used to model the integration among cloud, fog, and IoT.

Table 3.2 Articles about modelling of IoT, fog, and cloud integration.

Year/ Source	IEEE	Science Direct	ACM
2013		High-performance scheduling model for multisensory gateway of cloud sensor system-based smart-living [10] QoS-aware computational method for IoT composite service [11]	
Total	0	2	0
2014	Energy efficient and quality-driven continuous sensor management for mobile IoT applications [12]	A fault fuzzy-ontology for large-scale fault-tolerant wireless sensor networks [13]	
Total	1	1	0
2015	Virtualization framework for energy efficient IoT networks [14] Opacity in IoT with cloud computing [15] Reliability modeling of service-oriented Internet of Things [16]		
Total	3	0	0
2016	Query processing for the IoT: Coupling of device energy consumption and cloud infrastructure billing [17] A location-based interactive model for IoT and cloud (IoT-cloud) [19] An IoT-based system for collision detection on guardrails [21]	System modeling and performance evaluation of a three-tier cloud of things [18] Collaborative building of behavioral models based on IoT [20] Deviation-based neighborhood model for context-aware QoS prediction of cloud and IoT services [22]	

	<p>Towards distributed service allocation in fog-to-cloud (F2C) scenarios [23]</p> <p>Interconnecting fog computing and microgrids for greening IoT [25]</p> <p>Theoretical modeling of fog computing: a green computing paradigm to support IoT applications [27]</p>	<p>Event prediction in an IoT environment using naïve Bayesian models [24]</p> <p>Mobile crowdsensing as a service: A platform for applications on top of sensing clouds [26]</p>	
Total	6	5	0
2017	<p>Application-aware resource provisioning in a heterogeneous IoT [28]</p> <p>Leveraging renewable energy in edge clouds for data stream analysis in IoT [31]</p>	<p>Wearable IoT data stream traceability in a distributed health information system [29]</p> <p>Incentive mechanism for computation offloading using edge computing: A Stackelberg game approach [32]</p>	<p>A novel distributed latency-aware data processing in fog computing-enabled IoT networks [30]</p>
Total	2	2	1

3.3.1 Analytical Models

Analytical models are mathematical models that have a closed-form solution, i.e. the solution to the equations used to describe changes in a given system can be expressed as a mathematical analytic function. In general, analytical models can be used to predict computing resource requirements related to workload behavior, content, and volume changes, and to measure effects of hardware and software changes [33]. However, most of analytical models rely on approximations, and it is important to realize how these impact on any given models' results. It can be observed that papers using analytical models dominate C2F2T literature featuring modeling techniques. In this systematic review, we found 16 articles using analytical models.

The authors in [18] define the architecture, where each physical and virtual component of layers are described as a vector of associated features. A set of equations is defined to calculate metrics such as power consumed and time latency of scenarios.

In [30], the authors consider the number of gateways (devices that receive data from IoT devices and forward to cloud infrastructure or fog devices) present in an architecture, the total data received for each gateway, and the time spent by each gateway to process the data, in their modeling. Equations are proposed to represent the available buffer and the occupancy efficiency of gateway buffers. Depending on the available space on the gateways, the data are transferred to higher layers, increasing the delay for data processing. Several equations are defined for the calculation of the delay and the authors propose an optimization of the gateways' efficiency, improving the occupancy and the response time efficiency for all gateways in the system.

The authors in [20] propose an analytical model to represent a scheduling mechanism in IoT environment. An equation is defined to represent the addition of a load to be processed in a device. This process takes into account several variables, such as processor load, free memory, and free bandwidth, among others.

The work presented in [22] proposes a method to predict the QoS of IoT services. The model is based on a neighborhood collaborative filter and allows an efficient global optimization scheme. A set of equations is defined to predict correctly the QoS of a service, based on aspects such as latency, response time, and user network condition.

A scheduling model is proposed in [10] to manage sensor applications by a gateway. The requirement resources of applications are taken into account in scheduling problem. Equation (3.1) shows the problem formulated by [10]. In a scenario with n applications $A = \{a_i(R_i, r_i, w_i(t), s_i(t))\}$, where R_i is a resource requirement vector, $r_i \in [0, 1]$ is a priority, $w_i(t)$ is a required work status, and $s_i(t)$ is an actual working status of a_i .

$$\begin{aligned}
\min. \varphi &= \sum_{i=1}^n \int_0^T r_i w_i(t) [1 - s_i(t)] dt \\
s. t. (1) \quad &\sum_{i=1}^n u_i s_i(t) \leq \alpha_c u_c \\
(2) \quad &\sum_{i=1}^n c_i s_i(t) \leq \alpha_m c_m \\
(3) \quad &\sum_{i=1}^n b_i^I s_i(t) \leq \alpha_l b_l \\
(4) \quad &\sum_{i=1}^n b_i^O s_i(t) \leq \alpha_o b_o
\end{aligned}$$

Equation 3.1 Scheme proposed in [10].

The integral calculates the total waiting time of the tasks, where T is evaluation time. u_c, c_m, b_l and b_o are, respectively, total CPU utilization, total memory capacity, input bandwidth and output bandwidth the gateway can offer. and the vector $[\alpha_c, \alpha_m, -\alpha_l, \alpha_o]$ represents the overloading factors for the CPU, memory, input and output bandwidth, respectively, and indicates that the system can run smoothly to some extent.

In [17], a set of analytic expressions are proposed for representing the expected energy consumption of devices, as well as a cloud billing method for a group of devices on an IoT aggregator. Considering n devices, Equation (3.2) computes the energy consumption of each device over the monitoring period, T . Where $E[\Psi_e]$ is a query data volume of devices, g_e is consumption rate (joule-per-bit), and i_e is “idle” energy consumption by each device. The integral of the second term represents the expected energy consumption of a device in idle mode. $c_e E[\Psi_e]$ is a threshold where the application activates “idle” mode, and $P_e(\omega_e)$ represent the probability density function of Ψ_e .

$$E_{\text{exp}} = E[\Psi_e]g_e + i_e \int_0^{c_e E[\Psi_e]} (c_e E[\Psi_e] - \omega_e) P_e(\omega_e) d\omega_e$$

Equation 3.2 Energy consumption model presented in [17].

Equation (3.3) shows the calculation of expected cloud billing cost, taking into account n aggregated query volumes from n devices. $E[\Psi_b]g_b$ represents the transfer/storage costs, the first integral represents the idle billing cost, and the second integral corresponds to the active billing cost. i_b is the cost of transferring one bit (dollars-per-query-bit), and c_b is a coupling point that define the expect billing cost.

In [25], the utilization of a strategy using fog computing with microgrids to reduce the energy consumption of IoT applications is considered. Two

$$B_{\exp} = E[\Psi_b]g_b + i_b \int_0^{c_b} (c_b - \omega_b)P(\omega_b)d\omega_b \\ + p_b \int_{c_b}^{\infty} (\omega_b - c_b)P(\omega_b)d\omega_b$$

Equation 3.3 Billing model presented in [15].

equations are proposed to evaluate energy consumption. Equation (3.4) calculates the energy consumed by an IoT service using cloud computing. This equation takes into account the energy consumed by IoT gateways when receiving data from IoT devices and sensors (E_{GW-r}), the energy consumed by IoT gateways to transmit data to the cloud data center (E_{GW-t}), the energy consumed by transport network between IoT gateways and cloud (E_{net}), and the energy consumed by components of data center (E_{DC}).

$$E_{IoT-cloud} = E_{GW-r} + E_{GW-t} + E_{net} + E_{DC}$$

Equation 3.4 Calculate of energy consumption between IoT and cloud infrastructure [25].

$$E_{IoT-fog} = E_{GW-r} + E_{GW-c} + \beta(E_{GW-t} + E_{net} + E_{DC})$$

Equation 3.5 Calculate of energy consumption between IoT and fog infrastructure [25].

Equation (3.5) calculates the energy consumption of communication between the IoT and the fog. This equation takes into account the same components of previous equation plus two other components: E_{GW-c} , the energy consumed by IoT gateways for local computation and processing, and β , a ratio of the number of updates from the fog to the cloud for synchronization.

In [16], an analytical modeling for estimating the reliability of an IoT scenario is proposed in a smart home context. An algorithm is proposed to estimate the reliability of the IoT service, which is formed by n subsystems. The calculation of the IoT system reliability is defined in Equation (3.6). It considers the availability of the all k programs running on the virtual machines (P_{pr}), the availability of the f input files for the programs (P_f), and the reliability of each subsystem (ISR), i.e. the reliability of the VM being executed.

$$R_s(t_b) = \prod_{i=1}^N ISR_i \times \prod_{i=1}^f P_f(i) \times \prod_{i=1}^k P_{pr}(i)$$

Equation 3.6 Reliability equation presented in [16].

In [15], the authors presented a model to evaluate the security level of C2T systems. The focus of the model is the flow of information, where an initial state of the system is defined, and a set of operations are performed. Thus, after performing these actions, the system reaches new states.

Works presented in [11, 23], and [32] use analytical models in optimization problems. In [23], authors used a knapsack problem (MKP) to find the optimal service allocation in C2F2T scenarios. For this, they consider a number of application aspects: load balancing, delay and energy consumption. So, the service allocation is defined as an MKP problem, where the objective is three-fold: minimizing the energy consumption by devices, minimizing the overload in terms of processing capacity, and minimize the overall allocation of services in infrastructure. In [32], authors address the interactions among cloud operator and IoT service provide as optimization problem. They formulated analytically the problem and maximize the utilities of cloud service with the purpose of obtaining optimal payment and computation offloading. In [11], authors propose analytical modeling to represent QoS of IoT composite services, taking into account such metrics as availability, reliability, and response time. An optimization algorithm is proposed to find optimal cost with QoS constraints.

In [27] and [31], authors use analytical models to compare two layers of architectures proposed. In [31], analytical models are used to decide if offloading computing will be processed in IoT devices or in cloud, taking into account the desired QoS and energy level available in IoT devices. By contrast, in [27], analytical models are proposed to compare the performance between the fog architecture against traditional cloud computing. Authors consider several aspects, such as location of devices, operation mode, hardware details, and type of events.

In [12] and [19], authors present analytical models to represent mobile nodes connected to cloud computing. In both articles, the proposed models consider the movement of the devices that are connected to the cloud. In [19], the model details the architecture in a selection of components: wireless sensor network (WSN), the cloud infrastructure, applications, and mobile users. While in [12], authors consider that the mobile devices are connected to the cells to send data to the cloud.

3.3.2 Petri Net Models

According to [34], a Petri Net is a well-known model to represent systems with respect to evaluate performance and dependability. To solve a Petri Net, one can use two options: (i) analytic solution by using Markov chain (in which case all transitions must follow exponential distributions); or (ii) simulations using theory of discrete event simulation. Although Markov chains are also indicated to represent the availability of a system, Petri Nets allows a more fine-grained

representation of the system, utilizing Markovian and non-Markovian distributions, and represents the system behavior with a fewer number of states [34]. We found two works that used Petri Net models: [26] and [29].

In [26], a mobile crowdsensing framework, integrating mobile devices into services hosted in the cloud is proposed. In order to demonstrate that the proposed framework is better than the standard mobile crowdsensing architectures, two Petri Net models are proposed. Figure 3.4 illustrates these two models. Figure 3.4a shows the proposed framework, while Figure 3.4b shows a common mobile crowdsensing architecture.

The places represent four possible states of a contributing node: contributing node availability (Av/NAv) and position on the interest area (In/Out). The transitions represent the probability of the devices entering or exiting these states. The framework model is more complex as extra modules were added, such as contributor enrollment and churn management.

In [29], the problem of managing the traceability of data in C2T scenario is evaluated. A Petri Net model is proposed (Figure 3.5) to map and match device data to users that assists tracking a transparent data trace route, and possible detection of data compromises.

Here, the Petri Net represents the behavior of a proposed wearable IoT architecture. The places represent the different sources where data are generated or

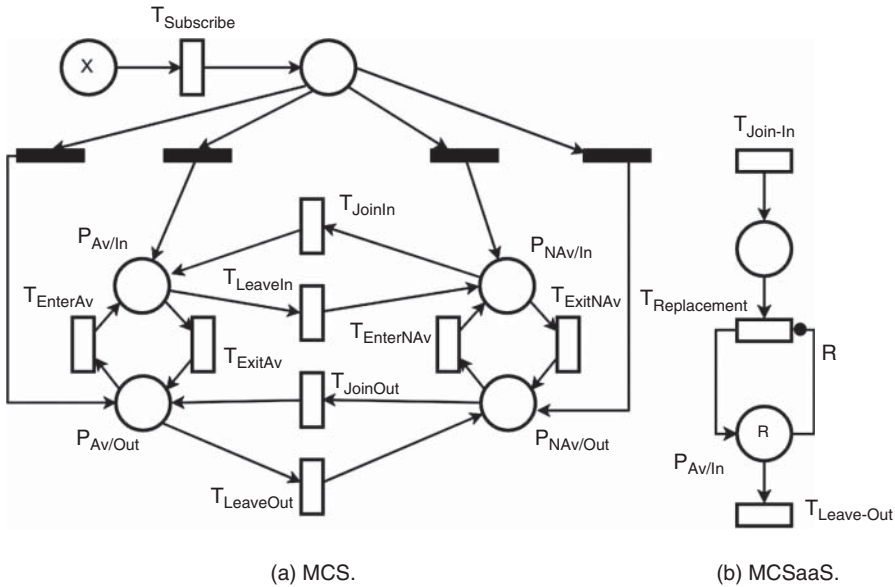


Figure 3.4 Petri Net model proposed in [26]. © Elsevier. Reproduced with the permission of Elsevier.

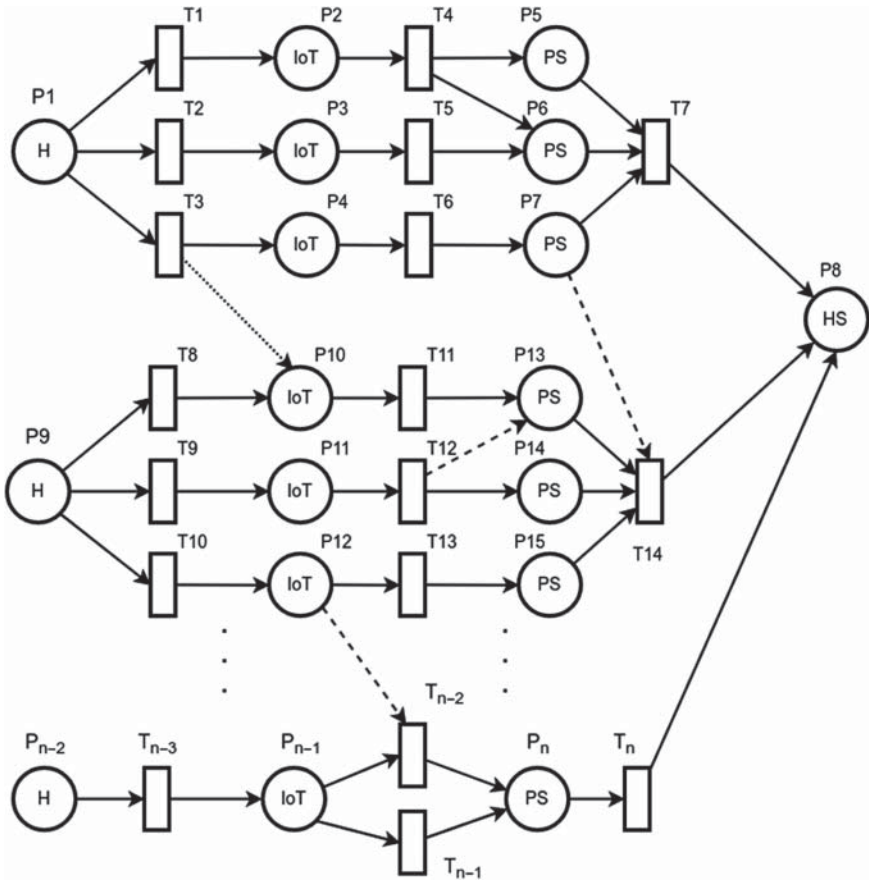


Figure 3.5 Petri Net model presented in [29]. © Elsevier. Reproduced with the permission of Elsevier.

collected. Transitions represent events that may occur, such as new medical data readings, for example vitals, from the wearable IoT device.

3.3.3 Integer Linear Programming

Some optimization problems can be modeled using integer linear programming (ILP). Here, the objective function and all the constraints are expressed as linear functions [35]. However, if the problem involves continuous and discrete variables, a mixed-integer linear programming (MILP) approach can be used to solve it. In this systematic review, two works modeled their problems using ILP and MILP, [28] and [14], respectively.

In [28], an ILP model is proposed for calculating the financial cost of a fog-cloud architecture located in a metropolitan area network (MAN). The authors represent application profiles and characteristics of each node in the MAN as two vectors. The ILP model minimizes the operational cost necessary to support the traffic in network topology while satisfy the application constraints. Equation (3.7) is the objective function that needs to be minimized. $Cost_t$ represents the total cost, $Cost_p$ is the processing cost, $Cost_s$ is the storage cost, $Cost_u$ are the total upstream and downstream costs respectively, and $Cost_c$ is the total MAN link capacity.

$$Cost_t = Cost_p + Cost_s + Cost_u + Cost_d + Cost_c$$

Equation 3.7 Objective function used in ILP model presented in [28].

Authors in [14] model the energy consumption of an IoT architecture by using MILP approach. This architecture is composed of mini clouds. The objective of the model is to minimize the energy consumption, where this consumption is composed of traffic-induced energy consumption and processing-induced energy consumption.

3.3.4 Other Approaches

According to [36], Markov chains model a sequence of random variables, which correspond to the system states, in which a state at one time depends only on the state in the previous time. Markov chains are being widely applied as statistical models of real-world problems. One article was identified in this category [21]. In this work, the authors proposed a Markov chain model with the goal of representing the energy consumption of an IoT-based system for collision detection on guardrails. This system is composed of a WSN, where gateways collect the sensors information and send it to the cloud. The Markov chain model is presented in Figure 3.6. Each state of the Markov chain represents a state of the system, with an associated energy consumption level. It is possible to estimate the energy consumption of the system by calculating the probability of a given state materializing.

Another modeling approach that can be used to represent the C2F2T integration is a probability-based one. While probability is useful to express the likelihood of an occurrence of an event, Bayesian probability represents a conditional measure of uncertainty associated with the occurrence of a given event, considering available information and prior beliefs [37]. The authors in [24] propose a Bayesian model for predicting events that may occur in an IoT application that is connected with the cloud. For this, the model calculates the probability of future events occurring based on historical event data. In addition, the authors assume that the occurrence of an event in the application

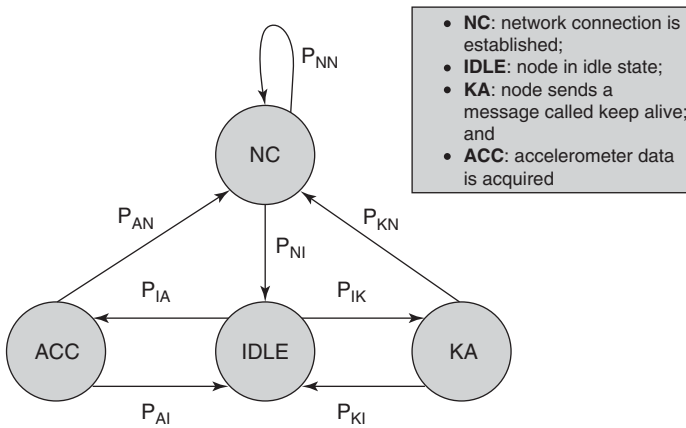


Figure 3.6 Markov model presented in [21]. Adapted from [21].

may imply the occurrence of several chain events. Prediction of flight delays resulting from problems in airplane equipment is used as a use case scenario. Thus, the Bayesian model calculates the conditional probabilities of such events occurring.

Beyond the approaches presented previously, different techniques can be combined to achieve a goal. For instance, in [17], a fault fuzzy-ontology is proposed for analyzing faults in WSN scenarios. In this work, the authors combine an ontology with fuzzy logic, arguing that ontology is appropriate to describe fault, error, and failure domains of systems, while fuzzy logic is a good approach for fault diagnosis. In this way, this approach provides a representation of heterogeneity of faults that allows one understand failures from different perspectives, e.g., applications, devices, and communications. The schema proposed allows us to detect and categorize the faults that occur in WSN. Figure 3.7 illustrates the fuzzy-ontology to detect faults in a hardware system. From right to left, the first layer of the fuzzy-ontology model represents a selection of fault possibilities that can occur in a wireless sensor network. The second layer represents the fault categories. The subsequent layers represent the propagation of fault in wireless sensor network.

3.4 Integrated C2F2T Literature by Use-Case Scenarios

In this section, we describe the use case scenarios modeled in the identified papers e.g. resource management, smart cities, WSN, health and other/generic. These are summarized in Table 3.3.

A number of studies present models to represent applications in health use case scenarios. Authors in [15] use a medical application as a case study for

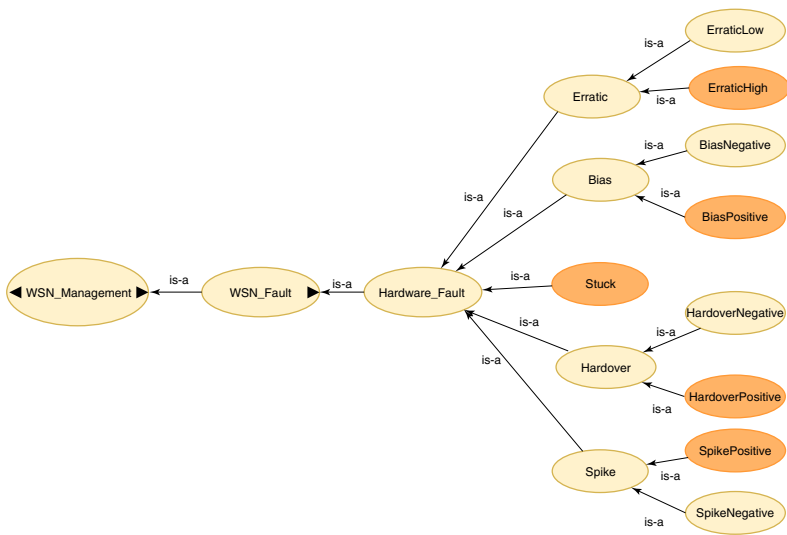


Figure 3.7 Fuzz-ontology model proposed in [13]. © Elsevier. Reproduced with the permission of Elsevier.

their proposed model that analyzes the security of the information flow in IoT systems integrated with cloud infrastructures. In [20], authors propose a framework that enables multiple applications to share IoT computational devices for health monitoring. The use case scenario in [29] is a wearable IoT architecture for healthcare systems.

Use case scenarios that address the WSN context can be found in [17] and [19]. In [13], authors propose a fault fuzzy-ontology that can be used to verify fault tolerance in large-scale WSN using service-oriented applications while [19] proposes a WSN model for sensing as a service that integrates IoT and cloud infrastructures.

The literature widely identifies cases used in smart cities. Authors in [26] and [12] explore a crowdsensing application. The former uses a smart traffic application to evaluate an architecture proposed for mobile crowdsensing-as-a-service [24]. The latter presents an air quality sensor application using mobile sensors and devices as inputs to the proposed model [12]. In [31], authors present a vehicle-to-cloud monitoring application that interacts with edge computing; their analytic model defines where the data will be processed – e.g. in the edge or in the cloud infrastructure. They use a video-streaming application to validate their model. The work presented in [16] is a fire alarm system. Authors evaluate it by using an IoT reliability model. The authors in [24] propose an airplane monitoring scenario, where data generated by C2T application are analyzed through a model in order to

Table 3.3 Scenarios presented in articles.

Scenarios	Applications	Articles
Generic	Generic IoT/fog/cloud applications	[12, 18, 30]
	Edge offloading computing	[32]
Health	Health monitoring	[20, 29]
	Medical research	[15]
WSN	Sensing as a service	[13]
	Fault tolerance	[19]
Smart cities	Mobile crowd sensing as a service	[26]
	Smart traffic	[26, 31]
	Smart living	[10, 16, 20]
	Airplane monitoring	[24]
Resource management	Resource allocation	[23, 28]
	Energy efficiency	[14, 17, 21, 25]
	Quality of service (QoS)	[11, 22]

estimate flights delay. Finally, in [10], authors use a smart living space in their scheduling model with gateways to access cloud resources.

In the resource management category, works were identified that analyze energy consumption. Authors in [17] and [14] study energy efficiency of IoT devices in a cloud-IoT context. Renna et al. [17] analyzes the relation between energy consumption and cloud infrastructure billing cost. Benazzouz and I. Parissis [13] presents a model to improve the energy efficiency of the IoT devices. In [21], authors use a model to analyze the energy consumption of a WSN. Similarly, [25] studies energy consumption but, in this case, the authors analyze how the integration of IoT and fog can reduce this consumption. Papers that address resource allocation in a C2F2T context were also identified, i.e. [23] and [28]. Finally, [11] and [22] discuss QoS in IoT and cloud-IoT environments, respectively. Ming and Yan [11] utilize a mathematical model to calculate the QoS of a set of IoT services (IoT sensor network). In [11], the authors propose a neighborhood model that makes QoS predictions for IoT and cloud services.

Some works do not describe a specific use case scenario or application for their models. For instance, Li et al. [18] presents a three-tier model that address the C2F2T scenario; and Desikan, Srinivasan, and Murthy [30] proposes a model to improve the response time of an IoT system that uses gateways to communicate to the cloud, using fog computing as an intermediary platform. Similarly, the work in [27] is a generic fog architecture and [32] proposes an architecture focused on offloading computing to the edge to improve the mobile user's experience.

3.5 Integrated C2F2T Literature by Metrics

In this section, we explore the sample of papers identified by the evaluation metrics (Table 3.4) used to derive insights in to the main concerns relating to C2F2T scenarios. We do not describe the metrics addressed in some of the papers in this section. Benazzouz and Parissis [13] do not evaluate the model proposed; while in [20, 22], and [24], the authors only evaluate the efficiency of the proposed models.

3.5.1 Energy Consumption

By design, most IoT devices are limited by power source, typically a battery, which comprises application performance. Unsurprisingly, energy consumption features in most (10) of the articles examined. Authors in [18] propose a model to represent a C2F2T architecture. They evaluated the energy consumption in each of the application layers (IoT, fog, and cloud), identified

Table 3.4 Metrics observed in articles.

Metric	Variations	Articles
Energy consumption	Devices power consumption	[14, 17, 18, 25, 27, 31],
	Energy efficiency	[19]
	Percentage energy saved	[12, 23]
Performance	Latency	[18, 27, 28]
	Effective arrival rate	[30]
	Average system response	[30]
	System effectiveness	[26, 30]
	Cache performance	[10]
Resource consumption	Usage of devices	[23, 32]
	Average buffer occupancy	[30]
Costs	Itemized costs	[28]
	Operational costs	[17]
QoS	Quality of pattern recognition in images	[31]
	Response time	[11]
	Reliability	[11, 16, 29]
Security	Opacity	[15]

the main sources of consumption of each, and verified the energy consumption in relation to the increase in the number of devices in the architecture.

In [17], the energy consumption of devices in an IoT infrastructure are presented in a different way. The model represents the device energy cost in three modes: active, idle, and when the application switches from idle to active state. In [12], the authors also address the energy consumption of idle and active devices sending data to a cloud environment, but in this work, the energy consumption for transmitting, receiving, listening, sensing, and computing are taken into account. In [21], the authors consider the operational mode of devices, however, in this work the energy consumption depends on the distance between the communicating nodes and the number of bits that will be transmitted. In addition, the authors take into account the medium where the communication will occur – for instance, open space.

In [12, 14, 27], the authors consider that energy consumption is directly related to the traffic between hierarchical devices layers. In [14], the energy consumption is also impacted by processing-induced VMs located in the networking elements at the upper three layers. In [12], the movement of devices is considered as an additional variable.

Jalali et al. [25] evaluate the energy consumption with respect to two perspectives: the synchronization between IoT devices and the fog, and between IoT devices and the cloud. Moreover, they consider how the data flow of applications affect energy consumption.

The work presented in [31] assesses the energy consumption of fog devices and cloud equipment. The fog devices are more similar to the IoT devices with respect to computational capacity. This paper is relatively unique in that it presents a renewable energy source where the fog devices are powered by photovoltaic panels. They evaluate the consumption of energy produced for photovoltaic panels, and other energy sources, such as batteries.

In [23], Souza et al. try to minimize the excessive energy consumption of C2F2T architectures to find on optimal service allocation.

3.5.2 Performance

Another metric that has been evaluated in many articles is application performance. In [18] and [27], the system performance is evaluated with respect to the latency of the application. This latency can be divided into (i) processing latencies and (ii) transmission latencies. The processing latency is the time taken by an application to process all tasks. The transmission latency is the communication delay to send a unit of data to the destination. In [19], the packet delivery latency is evaluated. This is calculated by the number of hops to the destination, the sleep interval of a node, and the time to transmit a data packet. In [28], the delay of processing applications requests is defined analytically as a combination of computational complexity of devices and the average flow size of application.

In [10], Lyu et al. evaluate the performance of a scheduling mechanism. For this, they assess the scalability of the scheduling and caching solver and add several sensors connected to cloud servers. In addition, the average waiting time and throughput applications are analyzed.

In [26], a mobile crowdsensing scenario is used to evaluate the performance and the effectiveness of the system using a Petri Net model proposed.

In [30], the authors examine the performance of C2F2T applications using a variety of means. The response time of system is defined as the time elapsed from the data was generated until it is processed by a gateway. Efficiency processing is formulated mathematically and takes account of the time to process the application data, the occupancy buffer, and the response time of all gateways.

3.5.3 Resource Consumption

Due of the limited capabilities of IoT devices, careful attention is required when allocating tasks on these devices. An analysis of this workload allows assigning tasks more optimally, taking into account the resources currently available. In

addition, because fog and cloud computing offer additional capabilities for IoT devices, it is necessary to examine where such workloads should be processed.

In [23], authors aim to obtain the optimal allocation of services on the available resources, taking into account resources of devices available in C2F2T infrastructures. They present the advantages of using fog computing and evaluate the number of allocated resources and the usage of resource devices.

In [32], authors assess the level consumption in C2T scenario. They evaluated the consumption about two aspects: cloud infrastructure consumption and IoT devices consumption.

Due to the limited computing and storage capacity of IoT devices, [30] evaluates the effectiveness of gateways by examining the buffer occupancy of the gateways, where the arrival of data to the gateways was represented by a queuing system.

3.5.4 Cost

Applications that require a lot of cloud computing resources can dramatically increase the cost to the service provider and ultimately the consumer. Obviously, there is overlap between energy consumption studies and cost studies.

Renna et al. [17] examine the billing costs when computational resources are reserved to process queries uploaded by IoT applications. Therefore, billing costs are directly proportional to the expected query volume generated by IoT devices.

Sturzinger et al. [28] attempt to minimize the operational cost of provisioning IoT traffic to the cloud, while satisfying all application constraints. The total cost is composed of the sum of the cost of all devices including as processing, storage, upstream and downstream cost, local and global traffic, and the link capacity cost.

3.5.5 Quality of Service

In IoT scenarios, many factors can impact on QoS such as network-related delays, available computational resources, and the number of IoT devices consuming resources.

In [11], composite services are evaluated in an IoT environment. In order to evaluate the QoS, the authors consider it is necessary to divide composite services into simpler more granular services, and then evaluate the QoS of each one separately.

The use case presented in [31] analyses of videostreams in the cloud generated by vehicles on a road. Authors represent QoS as the detection accuracy of an object in video images. They vary the resolution of video in order to decrease the CPU, memory and bandwidth consumption. In addition, considering the scenario where there are $2n+1$ cars, the accuracy is defined

as the probability of a result (object detected in image) appears exceeds $n + 1$ times among $2n + 1$ results.

In [16], reliability of the IoT system depends on a variety of factors, including the availability of the service, the availability of the input files to service, and the reliability of each sub-system that composes the overall IoT system. In [29], reliability was defined as the ability of the traceability model to detect and prevent attacks on applications.

3.5.6 Security

Only one article evaluated security issues. In [15], the authors evaluate the concept of opacity in data flow in the C2T context. Opacity is a uniform approach to describe security properties expressed as predicates. The authors used analytical models to verify the opacity in the medical research application scenario.

3.6 Future Research Directions

Fog and cloud computing address a number of problems encountered in the IoT; however, they also increase management complexity. Despite fog and cloud computing offering greater availability and resilience, they can also be viewed as vulnerabilities or potential points of failure. As such, in addition to device/end point failure, we now also need to pay attention to fog node and cloud infrastructure failures. While cloud and fog integration is relatively well known and shares common technologies, the integration/extension with IoT is a nontrivial task, mostly due to massive device heterogeneity and service requirements.

As presented in previous sections, many studies examined have proposed computational models to understand how IoT, fog, and cloud infrastructures can be integrated in order to improve the overall system performance and availability. Those works consider a wide range of applications and scenarios, modeling, and analysis regarding the reduction of application energy consumption resulting from greater C2F2T integration. Future research may address the use of gateways to distribute/balance requests to be processed in the cloud infrastructure or by fog devices.

Another area for future research concerns failure management and, in particular, the minimization of failure, whether on the device or fog node, on application availability. Some applications have high criticality, such as those in the connected health space e.g. health monitoring. In these case, any downtime can lead, in the worst case, to death. Here, the main goal is to identify the bottlenecks in this integrated system and propose strategies to minimize the application downtime, prevent failure, and guide investment decision-making.

Given the complexity of the resource pool in the C2T continuum, resource management is a fruitful area for future research, as requests can be allocated

locally, in the fog, or in one or more clouds. There is a wide range of data for informing resource allocation including device location, user information, application throughput and scalability, to name but a few. As the IoT feature space becomes more standardized per use case, more fine-grained data can be used to inform and model user and device behavior to inform these decisions. More sophisticated QoS mechanisms can then be assigned application priorities and resources can be allocated appropriately.

The cloud-fog-IoT space is complex, and such complexity is exacerbated by lack of standardization and extreme heterogeneity. From a modeling perspective, this results in much more computationally complex models. State-based models, such as Markov chains and Petri Nets, grow exponentially with the size of the model and can suffer from the so-called state-space explosion problem [38]. As such, one needs to pay attention to the models' scalability and run-time performance. Moreover, the complexity and, to some extent, the uncertainty resulting from a rapidly changing environment and the chain of service provision in the cloud-fog-IoT space can result in significant challenges in validating models against real scenarios. Researchers need to consider how best to improve the effectiveness and accuracy of models from a methodological perspective. This may require additional effort e.g. prototyping to some other methodology to validate the model's performance accuracy.

3.7 Conclusions

According to Cisco,¹ approximately 20 billion objects will be connected in 2020. These devices generate a massive volume of data at high velocity and varying formats, and require additional processing and storage capacity. In this aspect, cloud computing provides "unlimited" capacity to IoT devices based on the pay-as-you-go model, and one completes the other. However, this integration is complex to manage for many reasons, not least security, performance, communication delay, QoS, and so on. Fog computing adds a layer between cloud and IoT to solve problems related to communication, since the fog devices are located geographically closer to the IoT devices.

A wide variety of applications depend on C2F2T integration. Smart cities, WSN, e-health, traffic management, and smart buildings are some of the use case scenarios featured in the literature examined. These applications have their own domain-specific requirements and associated limitations that need to be evaluated on a case-by-case basis, such as data security and integrity, availability, reliability, real-time data, etc. Models have a useful role to play in the

¹ https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

evaluation of many of the components, variables, and aspects of integrated C2TF2 systems.

In this chapter, we presented a systematic literature review about models that represent C2F2T integrations. We analyzed other relevant aspects including typical scenarios and metrics evaluated in the articles. We identified that the most used modeling techniques in the articles were analytical models, with 16 articles, followed by Petri Net (2 articles), and ILP (2 articles). In itself, this suggests a greater need for exploration of the C2T topic through a wider range of methodological lenses. Similarly, our descriptive analytics suggest that energy consumption is the topic of most concern based on the number of articles focusing on this particular unit of measurement. While the literature examined covers a relatively small number of techniques and units of measurement, there is greater variability in the use case scenarios as illustrated in the previous paragraph. Our descriptive analytics reflect a substantial opportunity for future academic research through the identification of studies whose uniqueness can be designed from the interstices of C2F2T components and systems, modeling approaches, units of measurement or metrics, and use case scenarios.

Some aspects were not considered in the articles analyzed. The absence of a significant number of studies regarding application availability in C2F2T scenarios is noteworthy. For example, unavailability of a health monitoring system or healthcare application, for even a short period, can have unacceptable outcomes. Furthermore, greater examination is needed regarding the impact of each C2F2T layer (cloud, fog, and IoT) on the overall availability of the application is needed. Such studies will lead to novel strategies to improve the availability of applications in the C2F2T scenarios, which in themselves will require further evaluation.

The IoT ecosystem is a significant economic and societal opportunity whose evolution may lead to the Internet of Everything, where people, processes, things, data, and networks are all connected and interconnected. It is a vision for the future that is not without challenges. System modeling can play an important role in understanding and optimizing C2F2T systems and accelerating the evolution and maturity of the IoT for everyone's benefit.

Acknowledgments

This work is partially funded by the European Union's Horizon 2020 and FP7 Research and Innovation Programmes through RECAP (<http://www.recap-project.eu>) under Grant Agreement No. 732667.

The authors would like CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the support.

References

- 1 F. Gens, TOP 10 PREDICTIONS. IDC Predictions 2012: Competing on the 3rd Platform. <https://www.virtustream.com/sites/default/files/IDCTOP10Predictions2012.pdf>. March, 2018.
- 2 S. Aguzzi, D. Bradshaw, M. Canning, M. Cansfield, P. Carter, G. Cattaneo, S. Gusmeroli, G. Micheletti, D. Rotondi, R. Stevens. Definition of a Research and Innovation Policy Leveraging Cloud Computing and IoT Combination – Digital Agenda for Europe – European Commission. <https://ec.europa.eu/digital-agenda/en/news/definition-research-and-innovation-policy-leveraging-cloud09-0computing-and-iot-combination>. March, 2018.
- 3 J. Bradley, J. Barbier, D. Handler. Embracing the Internet of Everything to Capture Your Share of \$4.4 Trillion. Cisco IBSG Group. http://www.cisco.com/web/about/ac79/docs/innov/IoE_Economy.pdf. March, 2018.
- 4 F. Bonomi, R. Milito, P. Natarajan, et al. Fog computing: A platform for Internet of Things and analytics, *Big Data and Internet of Things: A Roadmap for Smart Environments*. Springer, Switzerland, 2014.
- 5 A. Botta, W. De Donato, V. Persico, and A. Pescapé. Integration of cloud computing and internet of things: A survey. *Future Generation Computer Systems*, 56: 684–700, March 2016.
- 6 L. Iorga, L. Feldman, R. Barton, M. Martin, N. Goren, C. Mahmoudi. The NIST definition of fog computing – draft, NIST Special Publication 800 (191), 2017.
- 7 P-O. Östberg, J. Byrne, P. Casari, P. Eardley, A. Fernandez Anta, J. Forsman, J. Kennedy, T. Le Duc, M. Noya Marino, R. Loomba, M.A. Lopez Pena, J. Lopez Veiga, T. Lynn, V. Mancuso, S. Svorobej, A. Torneus, S. Wesner, P. Willis, and J. Domaschka. Reliable capacity provisioning for distributed cloud/edge/fog computing applications. European Conference on Networks and Communications (EuCNC), Oulu, Finland, June 12–15, 2017.
- 8 FQB da Silva, M. Suassuna, A. César C. França, et al. Replication of empirical studies in software engineering research: a systematic mapping study. *Empirical Software Engineering*, 19(3): 501–557, June 2014.
- 9 F. Rowe. What literature review is not: diversity, boundaries and recommendations. *European Journal of Information Systems*, 23(3): 241–255, May 2014.
- 10 Y. Lyu, F. Yan, Y. Chen, et al, High-performance scheduling model for multisensor gateway of cloud sensor system-based smart-living. *18th International Conference on Information Fusion (Fusion)* 21: 42–56, January 2015.
- 11 Z. Ming and M. A. Yan. QoS-aware computational method for IoT composite service. *The Journal of China Universities of Posts and Telecommunications* 20 (2013): 35–39.

- 12 L. Skorin-Kapov, K. Pripuzic, M. Marjanovic, et al, Energy efficient and quality-driven continuous sensor management for mobile IoT applications, *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2014 International Conference on, Miami, Florida, October 22–25, 2014.
- 13 Y Benazzouz and I. Parissis, A fault fuzzy-ontology for large scale fault-tolerant wireless sensor networks, *Procedia Computer Science* 35 (September 2014): 203–212.
- 14 Zaineb T. Al-Azez, et al. Virtualization framework for energy efficient IoT networks. *4th IEEE International Conference on Cloud Networking (CloudNet)*, Niagra Falls, Canada, October 5–7, 2015.
- 15 W. Zeng, K. Maciej, and P. Watson. *Opacity in Internet of Things with Cloud Computing*. University of Newcastle Upon Tyne, Newcastle upon Tyne University Computing Science, Newcastle, England, 2015.
- 16 R. K. Behera, K. Ranjit Kumar, K.R. Hemant, K. Reddy, and D.S. Roy. Reliability modelling of service-oriented Internet of Things. *Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)*, 2015 4th International Conference on, Noida, India, September 2–4, 2015.
- 17 F. Renna, J. Doyle, V. Giotsas, Y. Andreopoulos. Query processing for the Internet-of-Things: Coupling of device energy consumption and cloud infrastructure billing. *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*. Berlin, Germany, April 4–8, 2016.
- 18 W. Li, I. Santos, F.C. Delicato, P.F. Pires, L. Pirmez, W. Wei, H. Song, A. Zomaya, S. Khan. System modelling and performance evaluation of a three-tier cloud of things. *Future Generation Computer Systems* 70 (2017): 104–125.
- 19 T. Dinh, K. Younghan, and L. Hyukjoon. A location-based interactive model for Internet of Things and cloud (IoT-cloud). *Ubiquitous and Future Networks (ICUFN)*, 2016 Eighth International Conference on, Vienna, Austria, July 5–8, 2016.
- 20 J.F. Colom, H. Mora, D. Gil, M.T. Signes-Pont. Collaborative building of behavioural models based on internet of things, *Computers & Electrical Engineering*, 58: 385–396, February 2017.
- 21 T. Gomes, D. Fernandes, M. Ekpanyapong, J. Cabral. An IoT-based system for collision detection on guardrails. 2016 IEEE International Conference on Industrial Technology (ICIT), Taipei, Tawan, May 14–17, 2016.
- 22 H. Wu, K. Yue, C. H. Hsu, Y. Zhao, B. Zhang, G. Zhang. Deviation-based neighborhood model for context-aware QoS prediction of cloud and IoT services. *Future Generation Computer Systems* 76: 550–560, November 2017.

- 23 V. B. Souza, X. Masip-Bruin, E. Marin-Tordera, W. Ramirez, and S. Sanchez. Towards Distributed Service Allocation in Fog-to-Cloud (F2C) Scenarios, *Global Communications Conference (GLOBECOM)*, 2016 IEEE, Washington, DC, USA, December 4–8, 2016.
- 24 B. Karakostas. Event prediction in an IoT Environment using Naïve Bayesian models. *Procedia Computer Science*, 83: 11–17, 2016.
- 25 F. Jalali, A. Vishwanath, J. de Hoog, F. Suits. Interconnecting fog computing and microgrids for greening IoT. *IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*, Melbourne, Australia, 28 November–1 December, 2016.
- 26 G. Merlino, S. Arkoulis, S. Distefano, C. Papagianni, A. Puliafito, and S. Papavassiliou. Mobile crowdsensing as a service: a platform for applications on top of sensing clouds. *Future Generation Computer Systems*, 56: 623–639, March 2016.
- 27 S. Sarkar and M. Sudip. Theoretical modelling of fog computing: a green computing paradigm to support IoT applications. *IET Networks* 5.2: 23–29, March 2016.
- 28 E. Sturzinger, T. Massimo, and M. Biswanath. Application-aware resource provisioning in a heterogeneous Internet of Things. *IEEE 21th International Conference on Optical Network Design and Modeling(ONDM)*, Budapest, Hungary, May 15–18, 2017.
- 29 R. K. Lomotey, J. Pry, and S. Sriramoju. Wearable IoT data stream traceability in a distributed health information system. *Pervasive and Mobile Computing*, 40: 692–707, September 2017.
- 30 K. E. Desikan, M. Srinivasan, and C. Murthy. A Novel Distributed Latency-Aware Data Processing in Fog Computing-Enabled IoT Networks. In *Proceedings of the ACM Workshop on Distributed Information Processing in Wireless Networks*, Chennai, India, July 10–14, 2017.
- 31 Y. Li, A.C. Orgerie, I. Rodero, M. Parashar, and J.-M. Menaud. Leveraging Renewable Energy in Edge Clouds for Data Stream Analysis in IoT. In *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, Madrid, Spain, May 14–17, 2017.
- 32 Y. Liu, C. Xu, Y. Zhan, Z. Liu, J. Guan, and H. Zhang. Incentive mechanism for computation offloading using edge computing: a Stackelberg game approach. *Computer Networks* 129: 399–409, 2017.
- 33 Gregory V. Caliri. Introduction to Analytical Modeling, *Int. CMG Conference*, Orlando, USA, December 10–15, 2000.
- 34 F. Bause and P. S. Kritzinger. *Stochastic Petri Nets*, Springer Verlag, Germany, 2002.

- 35 E. Oki. *Linear Programming and Algorithms for Communication Networks: A Practical Guide to Network Design, Control, and Management*, CRC Press, USA, 2012.
- 36 W. Ching, X. Huang, Michael K. Ng, and T.-K. Siu. *Markov Chains: Models, Algorithms and Applications*, Springer, USA, 2006.
- 37 T. Ando. *Bayesian Model Selection and Statistical Modeling*, CRC Press, USA, 2010.
- 38 E.M. Clarke, W. Klieber, M. Novacek, P. Zuliani. Model checking and the state explosion problem. *Tools for Practical Software Verification*, Germany, 2011.