

Received November 18, 2019, accepted December 19, 2019, date of publication December 25, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962271

Service Embedding in IoT Networks

HAIDER QAYS AL-SHAMMARI¹, AHMED Q. LAWEY¹, TAISIR E. H. EL-GORASHI¹,
AND JAAFAR M. H. ELMIRGHANI¹, (Senior Member, IEEE)

School of Electrical and Electronic Engineering, University of Leeds, Leeds LS2 9JT, U.K.

Corresponding author: Haider Qays Al-Shammari (ml14hqas@leeds.ac.uk)

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC), in part by the INTERNET (EP/H040536/1), in part by the STAR (EP/K016873/1), and in part by the TOWS (EP/S016570/1) projects.

ABSTRACT The Internet of Things (IoT) is the cornerstone of smart applications such as smart buildings, smart factories, home automation, and healthcare automation. These smart applications express their demands in terms of high-level requests. Application requests in service-oriented IoT architectures are translated into a business process (BP) workflow. In this paper, we model such a BP as a virtual network containing a set of virtual nodes and links connected in a specific topology. These virtual nodes represent the requested processing and locations where sensing and/or actuation are needed. The virtual links capture the requested communication requirements between nodes. We introduce a framework, optimized using mixed integer linear programming (MILP), that embeds the BPs from the virtual layer into a lower-level implementation at the IoT physical layer. We formulate the problem of finding the optimal set of IoT nodes and links to embed BPs into the IoT layer considering three objective functions: i) minimizing network and processing power consumption only, ii) minimizing mean traffic latency only, iii) minimizing a weighted combination of power consumption and traffic latency to study the trade-off between minimizing the power consumption and minimizing the traffic latency. We have established, as reference, a scenario where service embedding is performed to meet all the demands with no consideration to power consumption or latency. Compared to this reference scenario, our results indicate that the power savings achieved by our energy efficient embedding scenario is 42% compared with the energy-latency unaware service embedding (ELUSE) reference scenario, while our low latency embedding reduced the traffic latency by an average of 47% compared to the ELUSE scenario. Our combined energy efficient low latency service embedding approach achieved high optimality by jointly realizing 91% of the power and latency reductions obtained under the single objective of minimizing power consumption or latency.

INDEX TERMS Energy efficiency, IoT, MILP, queuing, smart buildings, service oriented architecture (SOA), virtualization.

I. INTRODUCTION

In the near future, a considerably large number of physical objects will contain sensors and actuators and will have the ability to communicate, forming the basis for the Internet of Things (IoT) [1]. IoT has motivated many global establishments to research and invest in this area and in its promising use in healthcare, transportation, and other smart building applications [2]. However, these promises of IoT come with considerable challenges. One of these challenges is the energy used and its effects on the environment and the expenditure involved [3], [4]. Although each IoT device

consumes low power, it is predicted that the number of IoT nodes will reach approximately 50 billion by the year 2020 [5], a massive number that can cause a high aggregate power consumption, as smart cities and smart building applications for example are expected to use a large number of IoT devices across cities [6]. Therefore, minimizing the energy consumed by such applications can play a significant role in reducing the total energy consumed by IoT.

In this study, we investigated solutions that can enable IoT to enhance real-world applications in a smart building. A smart building setup consists of a system for the monitoring and control of certain specific applications in the public or private areas of the building. The monitoring and control system consists of distinct types of sensors and actuators

The associate editor coordinating the review of this manuscript and approving it for publication was Xuxun Liu¹.

such as motion detectors, sound detectors, light detectors, smoke detectors, alarms, and gate controllers. These sensors and actuators, by means of wireless nodes, are connected in the building through a mesh topology. In a smart building, there are distinct applications that use the same resources in the monitoring and control systems. For example, both a security application and an energy saving application may use motion detectors, radio frequency identification (RFID) tags, and light detectors for monitoring simultaneously. To be successful, the smart building concept needs to be supported by various applications and has to be adopted by a range of industries, service providers, and administrations. It has to be applied efficiently in a mutual pattern for these sectors [7]. An essential phase towards the realization of the smart building concept involves the improvement of the communication infrastructure. The IoT paradigm is capable of collecting data from a massive assortment of distinct devices uniformly and seamlessly. The decentralized and heterogeneous properties of IoT devices capable of providing multiple functions require an efficient architecture that hides such heterogeneity from higher-level applications and provides interoperability for information exchange with other IoT devices [8]. The majority of the existing studies on virtualization have focused on two approaches. The first is a vertical approach concerned with the virtualization of the IoT architecture. This approach has been considered to achieve decoupling of the service provider from the infrastructure provider. Thus, according to the virtualization concept, the service is not concerned with the infrastructure and the infrastructure is not assigned to a specific service. Researchers have proposed a number of architectures that solve the virtualization issues in IoT. The second approach is related to a horizontal architecture. This approach focuses on the concept of virtualization types. Thus far, researchers have proposed node virtualization and network virtualization. These are also known as node-level and network-level virtualization. A Service Oriented Architecture (SOA) is potentially a viable middleware between users' applications and the IoT physical layer and can achieve interoperability between these heterogeneous IoT devices [9]. SOA enables the abstraction of IoT device functions that can then be translated into basic services, which in turn can be composed of complex services and exploited in the upper application layer. Software defined networking (SDN) is a control and orchestration mechanism that can be applied in networks to achieve network virtualization. It is thus possible to use SDN to realize network virtualization. In contrast, building a framework with SOA based middleware provides a focus on services and offers an effective architectural principle for network and node virtualization and supports system integration with cloud service provisioning and virtualization in addition to other advanced properties of SOA.

Figure 1 depicts the SOA middleware for IoT, which is composed of three sub-layers [1], [2], [10]: (i) object abstraction layer that enables IoT devices to provide their functions to the upper layers; (ii) service management layer to enable dynamic object discovery, status monitoring, and mapping of

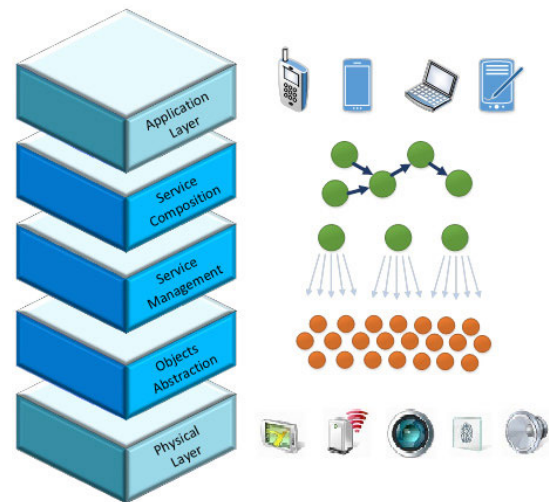


FIGURE 1. SOA-based architecture for IoT middleware [1].

the available services to the IoT devices' abstracted functions; and (iii) service composition layer where complex services, referred to as the business process (BP) workflow, are created from the basic services provided by the service management layer [11], [12].

With the use of SOA, devices can be reused or upgraded individually, leading to several advantages such as extensibility, scalability, and modularity along with the aforementioned interoperability among IoT devices [6], [13].

Because of these advantages of SOA, in [14], the authors presented an energy-centred and Quality of Service (QoS)-aware services selection algorithm (EQSA) for the composition of IoT services. They proposed a model that selects the services by using a lexicographic optimization strategy and a QoS constraint relaxation technique. The authors of [15] surveyed the recent development of SOA models for IoT and reviewed their fundamental technologies. The authors of [16] proposed a reference architecture based on SOA concepts by integrating the IoT, cloud, and edge technologies with the existing infrastructure. The authors of [17] surveyed the recent development of energy-efficient solutions for wireless sensors networks and reviewed some existing topologies that allow trade-offs between multiple requirements to be achieved for efficient and sustainable sensor networks. The authors of [18] presented a QoS message scheduling algorithm in IoT network-based SOA, which is targeted more toward service provisioning with the idea of service differentiation by classifying the messages into high-priority and best-effort messages. The authors of [19] surveyed the state of QoS methodologies in wireless terrestrial sensor networks to attain the low delay and reliability requirements needed in critical applications. These authors emphasized the main challenges in implementing QoS protocols in Wireless Sensors Networks (WSN) applications. The authors in [20] presented a neural network architecture based on a fog layered switch model that uses a neural network technique to manage intelligent congestion and QoS optimization. The proposed intelligent Software Defined Artificial Neural Network

(SD-ANN) model can improve QoS by minimizing the delay, reducing the data losses, and improving the reliability of complex network environments such as IoT-Fog computing networks. The authors in [21] presented a paradigm that uses Spine-leaf Fog Computing Network (SL-FCN) to reduce latency and network congestion of distributed multi-layered virtualized IoT data centres environments. The main objectives of the proposed paradigm are to maximize the bandwidth, maintain redundancy, and enhance the infrastructure resilience for critical applications. The authors in [22] introduced a QoS proactive auto scaling algorithm (PASCQA) modelled with a heuristic for cloud mesh Cyber-Physical Systems (CPS), the proposed model reduced scaling overheads, enhanced control of CPS resource utilization, provided higher flexibility in terms of meeting user requests, and minimized physical connections. The authors of [23], [24] developed strategies to improve the energy efficiency of Internet of Things, while [25], [26] considered the virtualization of such networks. Processing the sensor data and the use of data analytics based on big data streams was surveyed in [27], with [28] using these analytics for effective actuation in the network. Greening these big data networks was introduced and discussed in [29], [30] whereas improving the energy efficiency of the clouds and their interconnecting networks that process the IoT data was evaluated in [31], [32] with the energy efficiency of content sharing optimized in [32] and [33]. The energy efficiency of the networks supporting different services was optimized in [34]–[42]. Resilience is essential for a range of services, hence [43] and [44] introduced strategies to improve resilience with energy efficiency.

In the present study, we formulated the problem of finding the optimal set of IoT nodes and links to embed BPs into the IoT layer by considering the following three objective functions: i) minimizing only the network and processing power consumption, ii) minimizing only the mean traffic latency, and iii) minimizing a weighted combination of the power consumption and the traffic latency. This problem was formulated using mixed integer linear programming (MILP). We benefit from our track record in energy efficiency and networks virtualization, eg [12], [45]–[48] where we considered combinations of cloud, fog and communication links and devices in the energy efficient embedding problem. It should be noted that security is always an important consideration in IoT and was surveyed in [49]. While security is outside the scope of the current paper, it is worth considering the security of IoT service embedding and as a result the development of future architectures that provide security for the SOA architecture is important.

Reduction in latency is essential for a wide range of IoT applications where such strict QoS is needed. This may open new opportunities in smart cities where for example vehicular control is needed for improved traffic flow, industry 4.0 where robotic control is needed real time, in the Internet of energy where there may be variable levels of renewable energy and fast reaction is needed and in big data streaming [50]–[52].

The rest of this paper is organized as follows: In Section II, we introduce our framework of service embedding in IoT networks. Section III discusses the service embedding evaluation and its results. Finally, Section IV concludes this paper.

II. FRAMEWORK OF SERVICE EMBEDDING IN IoT NETWORKS

In the smart building setting, many services employ IoT nodes such as: security services employing motion detectors, RFID, display screens and alarms; energy saving services employing motion detection, temperature sensors; fire protection services employing temperature sensors, smoke detectors, water sprinklers and alarms; entertainment services employing noise detectors, and temperature sensors; administration services employing motion detectors, temperature sensors, door actuators, and alarms.

These services and other services can share the same sensing and actuating facilities like sensors for motion, temperature, sound, smoke detectors in addition to the processing modules of the IoT nodes. IoT networks are capable of providing multiple services but they require an efficient architecture that hides such heterogeneity from higher level services and provides interoperability for information exchange with other IoT devices. The SOA enables the abstraction of the IoT node functions and their translation into basic services which in turn can be composed into complex services and exploited by the upper application layer.

We developed a framework to embed service requests into a substrate network of IoT nodes. These requests are implemented following the SOA in the form of a BP. A BP is a virtual topology that consists of virtual nodes and links. The virtual nodes encapsulate the requested processing demand, sensing/actuating functions. We considered a setup where these virtual demands are within the capacity of the IoT physical resources. There is therefore no need to rely on relatively distant fog resources and distant cloud resources. In a previous early work, we considered a setup where fog and cloud resources are used supported by a passive optical network [12]. The virtual links carry traffic between virtual nodes. The embedding process maps the virtual nodes and virtual links of each BP into nodes and links of the IoT layer. Each BP is defined as a set of virtual nodes and links. Each virtual node has a function that requires processing and memory. Virtual nodes need to be embedded in certain geographical zones. Virtual links carry traffic demands between virtual nodes.

Each IoT node is characterized by the following modules as shown in Figure 2:

- A processing module hosting CPU and RAM.
- A network module hosting a wireless transceiver (Tx/Rx circuit and a Tx power amplifier).
- A function module that provides interfaces to a set of supported sensors and actuators.

Figure 3 gives an example of embedding two BPs. The framework embeds the virtual nodes of BP1 (A1-A2-A3) in

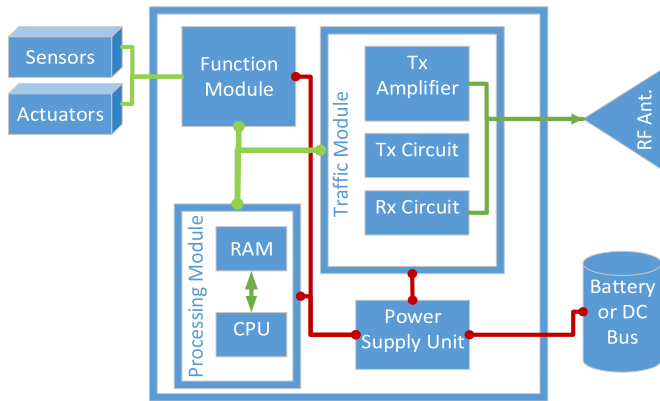


FIGURE 2. Block diagram of IoT node.

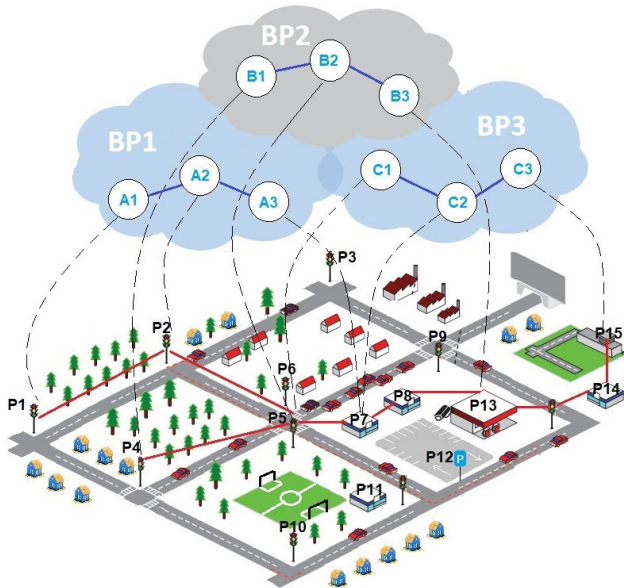


FIGURE 3. Service embedding layers in IoT networks.

the physical IoT nodes (P1-P2-P7), respectively; and chooses the path (P1-P2-P5-P7) to link the embedding IoT nodes. Each virtual node is embedded into an IoT node that satisfies the virtual node’s requirements. An IoT node that embeds a certain virtual node of a certain BP can at the same time work as a relay node for the traffic associated with another BP. This is shown in the second embedding example where IoT node P5 which is an embedding node for BP2 and at the same time works as a relay node for the traffic associated with BP1.

We consider a typical IoT setting where the power consumption of IoT nodes is mainly attributed to the processing and network modules while the sensing and actuating modules are externally powered.

As the traffic between IoT nodes is routed via a multi-hop network, we consider the queuing and transmission latency which dominate over the propagation delay as a network performance metric referred to it as traffic mean latency.

To study the power consumption and traffic mean delay resulting from embedding BPs into the IoT network, we formulate the embedding problem as a MILP model considering three different objective functions:

- Minimizing the total power consumption.
- Minimizing traffic mean latency.
- Minimizing both total power consumption and traffic mean latency in a multi-objective manner.

A. FRAMEWORK DEFINITIONS

Before we give these objective functions and the constraints imposed on the embedding of BPs, we introduce the sets, parameters and variables used in the formulations:

Sets

- B Set of business processes (BPs) in the virtual layer
- V Set of virtual nodes in each BP
- VN_{ia} Set of neighbors of each virtual node in each BP ($i \in B, a \in V$)
- P Set of IoT nodes in the physical layer
- PN_c Set of neighbors of IoT nodes ($c \in P$)
- F Set of functions supported by IoT nodes
- Z Set of zones in the IoT physical layer
- λ Set of arrival rates
- W_j Set of traffic mean latency per arrival rate ($j \in \lambda$) in ms per packet

Parameters

- V_{ian}^{FUNC} $V_{ian}^{FUNC} = 1$ if virtual node a in BP i requires the function n , $V_{ian}^{FUNC} = 0$ otherwise
- V_{iaz}^{ZONE} $V_{iaz}^{ZONE} = 1$ if virtual node a in BP i requires zone z , $V_{iaz}^{ZONE} = 0$ otherwise
- V_{ia}^{MCU} Processing requirement of the virtual node a in BP i in MHz
- V_{ia}^{RAM} Memory requirement of the virtual node a in BP i in kB
- V_{iab}^{TRFIC} Traffic demand between the virtual node pair (a, b) in BP_i in kb/s
- P_{cn}^{FUNC} $P_{cn}^{FUNC} = 1$ if IoT node c can provide the function n , $P_{cn}^{FUNC} = 0$ otherwise.
- P_{cz}^{ZONE} $P_{cz}^{ZONE} = 1$ if the IoT node c is located in zone z , $P_{cz}^{ZONE} = 0$ otherwise.
- P_c^{MCU} Processing capability of the IoT node c in MHz.
- P_c^{RAM} Memory capability of the IoT node c in kB.
- P_{ef}^{DIST} Distance between the neighboring IoT node pair (e, f) in meters.
- P_c^{IDLECP} Idle processor power in each IoT node c in mW.
- P_c^{MAXCP} Maximum processor power consumption in each IoT node c in mW.
- P_c^{IDLETP} Idle network power consumption in each IoT node c in mW.
- E_{ef}^{PBT} Energy per bit for each IoT link (e, f) in mW/kbps.
- M Large number ($= 10^8$).
- P_e^{CAPT} Link capacity for each IoT node (e) in kbps.
- F_{ef}^{TR} Transmit amplifier factor for each IoT link (e, f) in mW/kbps/m².

Variables

I_{iac}^{NE}	$I_{iac}^{NE} = 1$ if virtual node a in BP i has been embedded in IoT node c , $I_{iac}^{NE} = 0$ otherwise.
I_{iacn}^F	$I_{iacn}^{FI} = 1$ if IoT node c has the function n required by virtual node a in BP i , $I_{iacn}^{FI} = 0$ otherwise.
I_{iacz}^Z	$I_{iacz}^{ZI} = 1$ if IoT node c is located in zone z required by virtual node a in BP i , $I_{iacz}^{ZI} = 0$ otherwise.
I_{iabcd}^{LE}	$I_{iabcd}^{LE} = 1$ if the neighboring virtual nodes (a, b) in BP i have been embedded in IoT nodes (c, d), $I_{iabcd}^{LE} = 0$ otherwise.
X_{iabcd}^{XOR}	Dummy binary variable
R_{cd}^{TRFP}	Embedded traffic demand between IoT nodes (c, d) in kbps.
R_{cdef}^{ROUTE}	Traffic between IoT nodes (c, d) traversing the neighboring IoT nodes (e, f) in kbps.
I_{cdef}^R	$I_{cdef}^R = 1$ if the traffic demand between IoT nodes (c, d) traverses neighboring IoT nodes (e, f), $I_{cdef}^R = 0$ otherwise.
R_{ef}^{TRFL}	Traffic between neighboring IoT nodes (e, f) in kbps.
R_f^{TRFN}	Arrival rate of IoT nodes (f) in kbps.
LI_{ff}^{Lmbda}	Lambda indicator for each IoT node (f); (j) $LI_{ff}^{Lmbda} = 1$ if the arrival rate is (j), it is 0 otherwise.
W_f^{NODE}	Traffic mean latency for each node (f).
I_c^{PM}	$I_c^{PM} = 1$ if the processing module of IoT node c is powered on, $I_c^{PM} = 0$ otherwise.
I_c^{TM}	$I_c^{TM} = 1$ if the network module of IoT node c is powered on, $I_c^{TM} = 0$ otherwise.
TPP	Total processing power in the IoT network in mW.
TNP	Total network power in the IoT network in mW.
TL	Total traffic mean latency in ms.

B. FRAMEWORK OBJECTIVE FUNCTIONS

1) ENERGY EFFICIENT SERVICE EMBEDDING

This embedding scenario has an objective function whose goal is to minimize the total power consumption as follows:

$$\text{Objective : minimize } TNP + TPP \tag{1}$$

where TPP is total processing power and given by:

$$TPP = \sum_{c \in P} I_c^{PM} \cdot P_c^{IDLECP} + \sum_{c \in P} \sum_{i \in B} \sum_{a \in V} I_{iac}^{NE} \cdot P_c^{MAXCP} \cdot \frac{V_{ia}^{MCU}}{P_c^{MCU}} \tag{2}$$

where I_c^{PM} is a binary variable that indicates the activity of the processing module in IoT node c , P_c^{IDLECP} is the idle processing power parameter of IoT node c in mW, I_{iac}^{NE} is a binary variable that indicates if a virtual node a in BP i has been embedded in IoT node c , P_c^{MAXCP} is a parameter that gives the maximum CPU power consumption in each IoT node c in mW, V_{ia}^{MCU} is a parameter whose value gives the processing requirement of the virtual node a in BP a in MHz, and P_c^{MCU} is a parameter that specifies the processing capability of the IoT node c in MHz. The processing power consumption is considered to follow a linear profile versus the load with an idle power consumption. The total traffic power consumption of the network, TNP, and given by:

$$TNP = \sum_{e \in P} I_e^{TM} \cdot P_e^{IDLETP} + 2 \cdot \sum_{e \in P} \sum_{f \in PN_e} R_{ef}^{TRFIC} \cdot E_{ef}^{PBT} + \sum_{e \in P} \sum_{f \in PN_e} R_{ef}^{TRFIC} \cdot (P_{ef}^{DIST})^2 \cdot F_{ef}^{TR} \tag{3}$$

where f is neighbor IoT node of e and is included in PN_e , PN_e is the neighbors subset of IoT node e , I_e^{TM} is a binary variable that indicates the activity of the network module in the IoT node, P_e^{IDLETP} is the idle network power parameter of IoT node e , R_{ef}^{TRFIC} is a variable that specifies the traffic between neighboring IoT nodes e and f in kbps, E_{ef}^{PBT} is a parameter that gives the energy per bit for each IoT link e, f in mW/kbps, P_{ef}^{DIST} is a parameter that specifies the distance between the neighboring IoT nodes pair (e, f) in meters, and F_{ef}^{TR} is the transmit amplifier factor [18] for each IoT link e, f in mW/kbps/m².

The network power consumption is a function of the traffic and distance between the source and destination nodes. The network power consumption of each link consists of the idle power, the power consumed per bit by the electronics in the transmitter and the receiver, and the transmitter amplifier power consumption which is calculated based on the radio energy needed based on Frii's free-space equation in our setting (note that higher propagation factors beyond Frii's square law, e.g. cubic or higher, can be considered, and are a straight forward extensions of our equations, but are not considered here) [4], [48], [53].

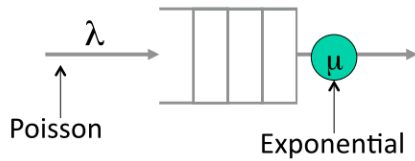
2) LOW LATENCY SERVICE EMBEDDING

The second scenario in our framework is concerned with minimizing the total traffic mean latency of the service embedding. The framework minimizes the traffic mean latency in the IoT network using the following objective function:

$$\text{Objective : minimize } TL \tag{4}$$

where TL P^{TL} is the total traffic mean latency in the network given by:

$$TL = \sum_{f \in P} W_f^{NODE} \tag{5}$$


FIGURE 4. Single server queuing system.

Our network is modelled as an open Jackson network of multiple M/M/1 queues where the utilization is less than 1 at every queue [54]. For simplicity, we consider each node as an M/M/1 queue. The M/M/1 model refers to a system with a single server, where arrivals are determined by a Poisson process and job service times have an exponential distribution as shown in Figure 4.

The mean latency is the average time that the packet takes to pass through queue and server, which is given by:

$$W_f^{NODE} = \frac{1}{(\mu_f^{NODE} - \lambda_f^{NODE})} \quad (6)$$

The arrival rate represents the average rate of successful packets transfer to the node through physical links per time unit. Mathematically, the arrival rate is the summation of data rates delivered to the node in the network.

In our framework, we considered that the service rate μ_f^{NODE} is fixed for each IoT node in the network. The service rate is the transmission rate of the network module. A variable, λ_f^{NODE} , is created to calculate the summation of packet arrival at each IoT device.

Since we are using linear programming, equation (6) must be converted to a linear format. To facilitate this, we use a lookup table indexed-variable to calculate the traffic mean latency. The lookup table indexed-variables method depends on generating lambda indicator as a binary variable according to the traffic value of λ_f^{NODE} for each node. Based on this indicator, the traffic mean latency for IoT nodes is given as the value corresponding to the indicator in the lookup table.

3) ENERGY EFFICIENT - LOW LATENCY SERVICE EMBEDDING

In this scenario, we consider a multi-objective MILP model to optimize the service embedding in IoT networks to achieve a trade-off between minimizing the power consumption and minimizing the traffic mean latency. The objective function is given as:

$$\text{Objective : minimize } \alpha \cdot \text{TL} + \beta \cdot \text{TNP} + \gamma \cdot \text{TPP} \quad (7)$$

where α , β and γ are weight factors with the following units 1/ms, 1/mW, 1/mW respectively used to harmonize the units and to emphasize the importance of the different components of the objective function.

C. FRAMEWORK CONSTRAINTS

The framework performs the embedding operation through two parts as follows:

1) EMBEDDING OF VIRTUAL NODES

$$\sum_{c \in P} I_{iac}^{NE} = 1 \quad \forall i \in B, \forall a \in V \quad (8)$$

$$\sum_{a \in V} I_{iac}^{NE} \leq 1 \quad \forall i \in B, \forall c \in P \quad (9)$$

Constraint (8) ensures that each virtual node in a BP is embedded in a single IoT node only. Constraint (9) states that each IoT node is not allowed to host more than one virtual node in each BP. This is considered as a coexistence constraint that is not used in all scenarios such as controller node virtualization.

$$\sum_{i \in B} \sum_{a \in V} I_{iac}^{NE} \geq I_c^{PM} \quad \forall c \in P \quad (10)$$

$$\sum_{i \in B} \sum_{a \in V} I_{iac}^{NE} \leq I_c^{PM} \cdot M \quad \forall c \in P \quad (11)$$

Constraints (10) and (11) build (include / add) a processing module in IoT node c if that node is chosen for embedding at least one virtual node a in BP i or more, where M is a large enough unitless number to ensure that $I_c^{PMI} = 1$ when $\sum_{i \in B} \sum_{a \in V} P_{iac}^{NE}$ is greater than zero.

$$\sum_{i \in B} \sum_{a \in V} V_{ia}^{MCU} \cdot I_{iac}^{NE} \leq P_c^{MCU} \quad \forall c \in P \quad (12)$$

$$\sum_{i \in B} \sum_{a \in L} V_{ia}^{RAM} \cdot I_{iac}^{NE} \leq P_c^{RAM} \quad \forall c \in P \quad (13)$$

Constraints (12) and (13) represent the processing and memory capacity constraints, respectively. They ensure that the embedded processing and memory workloads in an IoT node do not exceed the MCU and memory capacities, respectively.

$$I_{iac}^{NE} \cdot V_{ian}^{FUNC} = I_{iacn}^F \quad (14)$$

$$P_{cn}^{FUNC} \geq I_{iacn}^F \quad (15)$$

Constraints (14) and (15) ensure that the required function of each virtual node in BP is provided by its hosting IoT node.

$$I_{iac}^{NE} \cdot V_{iaz}^{ZONE} = I_{iacz}^Z \quad (16)$$

$$P_{cz}^{ZONE} \geq I_{iacz}^Z \quad (17)$$

Constraints (16) and (17) ensure that the required zone of each virtual node in a BP is matched by the zone of the hosting IoT node.

2) EMBEDDING OF VIRTUAL LINKS

$$I_{iac}^{NE} + I_{ibd}^{NE} = X_{abcd}^{LE} + 2 \cdot I_{abcd}^{LE} \quad (18)$$

$$\forall i \in B, \quad \forall a \in V, \quad \forall b \in VN_{ia} : a \neq b,$$

$$\forall c, d \in P : c \neq d$$

Constraint (18) ensures that neighboring virtual nodes a and b of i in B are also connected in the embedding IoT nodes c

and d . We achieve this by introducing a binary variable P_{iabcd}^{LE} which is only equal to 1 if I_{jqc}^{NE} and I_{ibd}^{NE} are exclusively equal to 1 otherwise it is zero, W_{iabcd}^{LE} is an auxiliary variable.

$$\sum_{i \in B} \sum_{a \in L} \sum_{b \in LNB_{ia}} I_{iabcd}^{LE} \cdot V_{iab}^{TRFIC} = R_{cd}^{TRFP} \quad \forall c, d \in P : c \neq d \quad (19)$$

Constraint (19) generates the path's traffic matrix resulting from embedding the virtual nodes a and b into the IoT nodes c and d .

$$\sum_{f \in PN_e} R_{cdef}^{ROUTE} - \sum_{f \in PN_e} R_{cdfe}^{ROUTE} \begin{cases} R_{cd}^{TRFP} & \text{if } e = c \\ -R_{cd}^{TRFP} & \text{if } e = d \\ 0 & \text{otherwise} \end{cases} \quad \forall c, d, e \in P : c \neq d \text{ and } e \neq f \quad (20)$$

Constraint (20) represents the flow conservation constraint for the traffic flows in the IoT network.

$$\sum_{c \in P} \sum_{d \in P} R_{cdef}^{ROUTE} = R_{ef}^{TRFL} \quad \forall e \in P, \forall f \in PN_e \quad (21)$$

Constraint (21) estimates the link's traffic between the neighboring IoT nodes e and d .

$$\sum_{f \in PN_e} R_{ef}^{TRFL} \leq P_e^{CAPT} \quad \forall e \in P \quad (22)$$

Constraint (22) states that the total traffic flows of the IoT node e does not exceed the node capacity.

$$R_{cdef}^{ROUTE} \geq I_{cdef}^R \quad \forall c, d, e \in P, \forall f \in PN_e : c \neq d, e \neq f \quad (23)$$

$$R_{cdef}^{ROUTE} \leq I_{cdef}^R \cdot M \quad \forall c, d, e \in P, \forall f \in PN_e : c \neq d, e \neq f \quad (24)$$

The constraints (23) and (24) build a path between the embedding IoT nodes c and d through the neighboring IoT nodes e and f , where $I_{cdef}^R = 1$ if there is a traffic path between the IoT nodes c and d that passes through the neighboring IoT nodes e and f , where M is a large enough unitless number which ensure that $I_{cdef}^R = 1$ when R_{cdef}^{ROUTE} is greater than zero.

$$\sum_{f \in PN_e} I_{cdef}^R \leq 1 \quad \forall c \in P, \forall d \in P, \forall e \in P \quad (25)$$

Constraint (25) ensures that traffic splitting is prevented for each path between the embedding IoT nodes c and d , such that the maximum number of physical links between neighboring IoT nodes e and f is one.

$$\sum_{c \in P} \sum_{d \in P} \sum_{f \in PNB_e} I_{cdef}^R \geq I_e^{TM} \quad \forall e \in P \quad (26)$$

$$\sum_{c \in P} \sum_{d \in P} \sum_{f \in PNB_e} I_{cdef}^R \leq I_e^{TM} \cdot M \quad \forall e \in P \quad (27)$$

Constraints (26) and (27) build a network module in IoT node e if that IoT node is chosen to send/receive traffic at least for

one link or more, where M is a large enough unitless number to ensure that $I_e^{TM} = 1$ when $\sum_{c \in P} \sum_{d \in P} \sum_{f \in PNB_e} I_{cdef}^R$ is greater than zero.

$$\sum_{e \in PN_f} R_{ef}^{TRFL} = R_f^{TRFN} \quad \forall f \in P : e \neq f \quad (28)$$

Constraint (28) estimates the arrival traffic for each IoT node.

$$\sum_{j \in J} LI_{ff}^{LMBDA} \cdot j = R_f^{TRFN} \quad \forall f \in P : e \neq f \quad (29)$$

Constraint (29) is an arrival rate indicator of arrival rate j for each IoT node f .

$$\sum_{j \in J} LI_{ff}^{LMBDA} \leq 1 \quad \forall f \in P \quad (30)$$

Constraint (30) ensures that each IoT node has no more than one arrival rate indicator.

$$\sum_{j \in J} W_j^{LIMDA} \cdot LI_{ff}^{LMBDA} = W_f^{NODE} \quad \forall f \in P \quad (31)$$

Constraint (31) estimates the mean traffic latency for each IoT (f).

The MILP optimization model was solved using CPLEX running on personal computer with processor core i5 -3.2 GHz and 16 GB RAM and on the university Polaris servers using 24 cores and 128GB RAM.

III. RESULTS AND EVALUATIONS

To evaluate the performance of the proposed model and heuristic, we consider a smart building scheme (for example in an enterprise campus) where the physical layer is composed of 30 IoT nodes connected by 89 bidirectional wireless links. These IoT nodes are distributed across an area 500 mx500 m and can carry various functions with the following assumptions:

- There is a set of 9 distinct functions, 4 sensing functions, one control function and 4 actuating functions. Each IoT node can provide 2 sensing functions, 2 actuating functions, and one controlling function (present only in one type of processor). The virtual node of each BP requests one function only.
- There is a set of five geographical zones that represent the sub-sections of the smart building (e.g. departments or sections in the enterprise campus). Each zone is equipped with six IoT nodes. All the functions and processor types exist in each zone. The virtual node requests an embedding location in one of these five zones.
- The IoT nodes processing capability is uniformly distributed among five processing capacities (8, 16, 16, 25, 25, 48 MHz) representing microcontrollers as shown in Table 1.
- Each virtual node has a specific processing demand that varies between 4 and 30 MHz hence we considered distributed data processing in the IoT computing environment. The scalability of our proposed architecture

TABLE 1. Processing modules power specifications and power consumption in active mode.

MCU Type	MCU CLK	RAM	Idle Power	Max. Power
MSP430F1	8 MHz	64 kB	1 mW	8 mW
MSP430FR5	16 MHz	64 kB	1 mW	14 mW
MSP430FR6	16 MHz	128 kB	1 mW	20 mW
MSP430F5	25 MHz	512 kB	1 mW	14 mW
MSP432P4	48 MHz	256 kB	1 mW	16 mW

TABLE 2. MILP model input parameters.

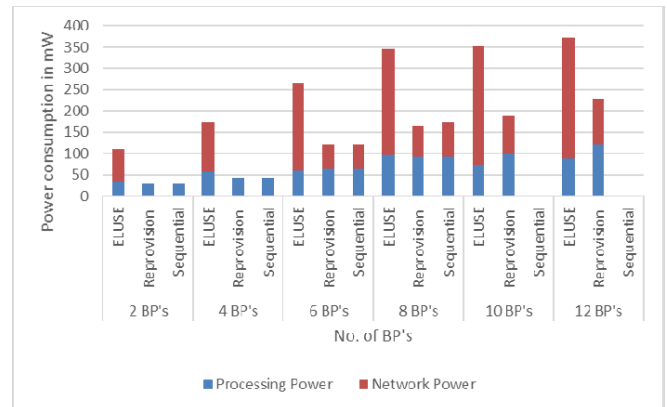
Parameter Description	Value and Unit
Energy per bit	50 nJ/bit
Maximum traffic capacity of node	250 kb/s
Packet size	128 byte
Maximum link distance	100 m
Transmitter amplifier power coefficient	255 pJ/bit. m ²
Scale factor with large value (M)	1000000

is ensured through its cellular architecture. Each building or section of the city represents an SOA cell and provides edge distributed data processing in addition to sensing and actuation in the IoT environment. The cells can be interconnection through a PON architecture as in our work in [12], where such cellular architectures are known to be very scalable as evidenced by cellular mobile communications which can provide nationwide coverage using limited resources. In situations where the IoT nodes have to collaborate extensively in processing tasks, a Spine-leaf model can be considered [21] as a form of distributed data processing. The cloud and fog can supplement the scalability of our cellular SOA by providing data processing, storage, resource management, service creation and service management.

- Each IoT node contains wireless transceiver modules [55]. The network modules used are low cost, low power, and are compatible with the ZigBee protocol stack for IoT networks [48]. The traffic demands of the virtual links vary from 50 to 200 packets per second with a packet size of 1 kb. Table 2 lists the model input parameters [48].

We study the embedding of 12 BPs arriving sequentially, two at a time. Each BP has three virtual nodes (sensor, controller and actuator) connected sequentially. The sensor is connected to the controller and the controller is connected to the actuator. The sensor virtual node requests a specific sensing function, the control virtual node requires processing capacity and the actuator virtual node requests a specific actuating function. The sensor and actuator virtual nodes of a BP need to be embedded in a specific zone while the controller virtual node can be embedded into any geographical zone.

We evaluate the power consumption and traffic mean latency resulting from embedding the BPs using the MILP model considering the three objective functions.

**FIGURE 5. Power consumption of energy efficient service embedding in same zone without coexistence constraint.**

A. ENERGY EFFICIENT SERVICE EMBEDDING

In this section, we evaluate the results of embedding BPs in terms of power consumption and traffic mean latency under three scenarios. In the first scenario, referred to as energy-latency unaware service embedding (ELUSE), BPs are embedded in physical nodes and links that satisfy their requirements where the objective function's goal is to ensure that all requests for embedding are met.

In the second and third scenarios, the objective is to minimize the total power consumption. However, in the second scenario, referred to as re-provisioning, each time a new BPs arrives, previously embedded BPs are re-embedded while in the third scenario, referred to as sequential embedding, arriving BPs are embedded without interrupting the existing BPs. We also study the coexistence constraints of the embedding and their effects on the results of the energy efficient service embedding.

1) SERVICE EMBEDDING ON SAME GEOGRAPHICAL ZONE

In this subsection, we considered that the sensor and actuator nodes of a BP need to be embedded in the same specific geographical zone. We also study embedding BPs with and without coexistence constraints. Under coexistence constraints, the virtual nodes of the BP cannot coexist in the same IoT node. The goal here is to improve the resilience of the BPs under single node failure.

Figure 5 shows the total power consumption of embedding BPs in which the sensing and actuating nodes are to be embedded in the same zone. The results show that the energy efficient re-provisioning embedding scenario resulted in saving an average of 63% of the power consumption compared to the ELUSE scenario. Under energy efficient embedding, fewer IoT nodes and links are activated to embed BPs compared to embedding under the ELUSE scenario. As no coexistence constraints apply, all the virtual nodes of a BP can be embedded in a single IoT node confining the virtual links traffic within this node and reducing the number of activated IoT nodes. The saving achieved by the energy efficient embedding decreases to 58% under the sequential scenario as the sequential approach builds on existing embedding

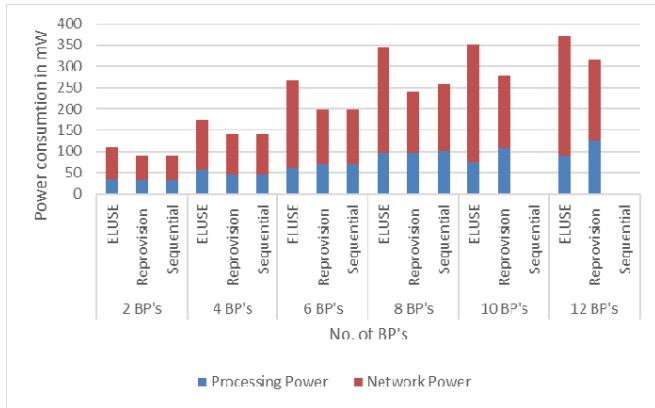


FIGURE 6. Power consumption of energy efficient service embedding in the same zone with coexistence constraint.

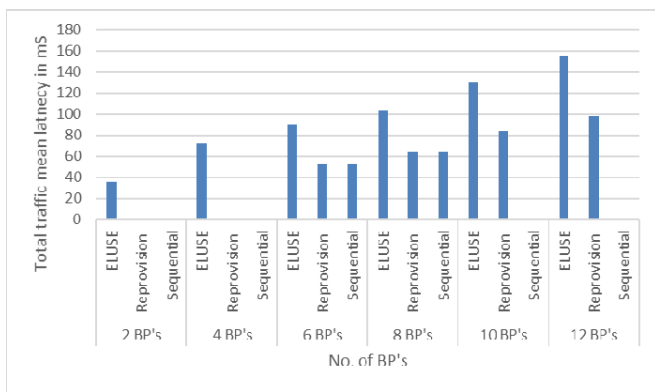


FIGURE 7. Average traffic mean latency of energy efficient service embedding in same zone without coexistence constraint.

decisions that become suboptimal with the arrival of new BPs. The optimal use of resources under the re-provisioning scenario resulted in successfully embedding 12 BPs while only 8 BPs were successfully embedded under sequential embedding.

Note that the power savings decrease as the number of embedded BPs increases. This is because the higher the load on the network the fewer the possible embedding solutions therefore narrowing the gap between energy efficient embedding and ELUSE.

Figure 6 shows the power consumption of embedding BPs in the same zone under coexistence constraints. The coexistence constraints reduce the power savings achieved by the energy efficient embedding scenarios to 36% and 29% for re-provisioning and sequential embedding, respectively. This reduction in power savings is due to the need to activate more IoT nodes to meet the coexistence requirements and the traffic between these nodes.

The results in Figure 7 display the average traffic mean latency resulting from embedding BPs without coexistence constraint. The re-provisioning embedding and the sequential embedding have reduced the average traffic mean latency by 62% and 60% respectively compared with ELUSE scenario. This is because energy efficient embedding selects routes of minimum hops and consequently lower traffic mean latency compared to random routing in ELUSE. However, energy

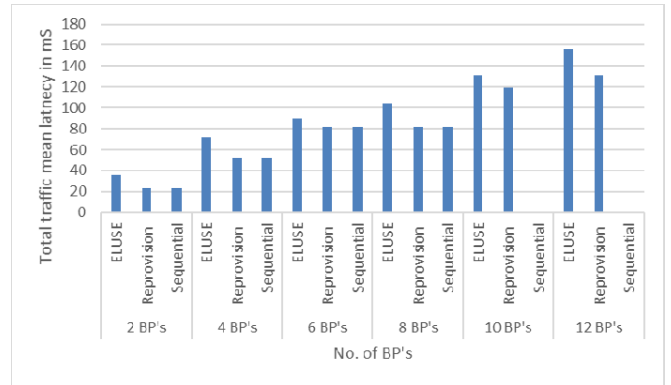


FIGURE 8. Average traffic mean latency of energy efficient service embedding in same zone with coexistence constraint.

efficient embedding does not produce the minimum traffic mean latency as energy efficient embedding tries to highly utilize the activated IoT nodes resulting in high traffic mean latency in these nodes.

Similar trends to those in Figure 7 are observed in Figure 8 for the average traffic mean latency resulting from embedding with coexistence constraints. The results show that the re-provisional embedding and the sequential embedding have reduced the average traffic mean latency by 27% compared with the ELUSE scenario. Comparing Figure 7 and Figure 8 shows that embedding BP on the same zone with coexistence constraint results in higher traffic mean latency compared to embedding without coexistence constraint. This is because without the coexistence constraint, the traffic of a BP can experience no traffic latency by embedding all the virtual nodes of the BP in a single IoT node.

2) SERVICE EMBEDDING ACROSS GEOGRAPHICAL ZONES

The previous results evaluated the power consumption and mean latency of embedding BPs where the sensor and actuator nodes need to be embedded in the same geographical zone. In this section we examine embedding BPs that require the sensor and actuator nodes to be embedded in distinct geographical zones. We study also the performance with and without coexistence constraints placed on the controller node. Under coexistence constraints, the controller cannot coexist in the same IoT node with the sensor or actuator node.

Figure 9 displays the power consumption of embedding BPs across different geographical zones without coexistence constraint. The power savings achieved by energy efficient embedding under the re-provisioning scenario and the sequential scenario when embedding across different zones are lower than those achieved for same zone embedding in Figure 6. This is because energy efficient embedding in the distinct zones cannot select to embed the sensor and actuator in the same node although coexistence constraints do not apply. The power savings achieved by the energy efficient embedding scenarios are 42% and 22% for re-provisioning and sequential scenarios, respectively.

The less efficient use of resources in embedding across zones reduces the number of BPs that can be embedded under

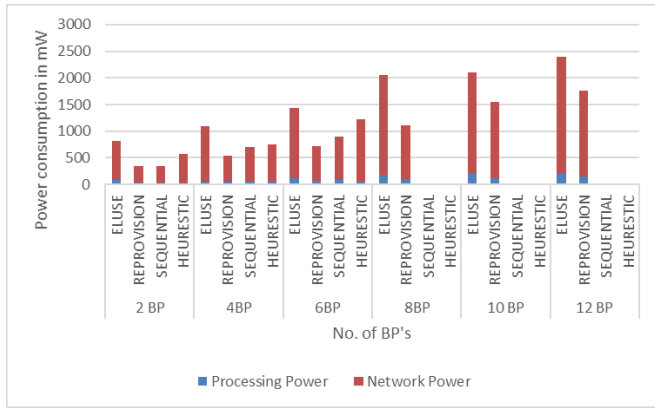


FIGURE 9. Power consumption of energy efficient service embedding across different zones without coexistence constraint.

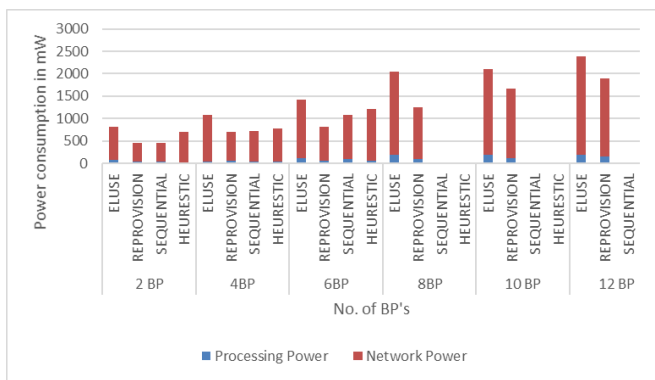


FIGURE 10. Power consumption of energy efficient service embedding across different zones with coexistence constraint.

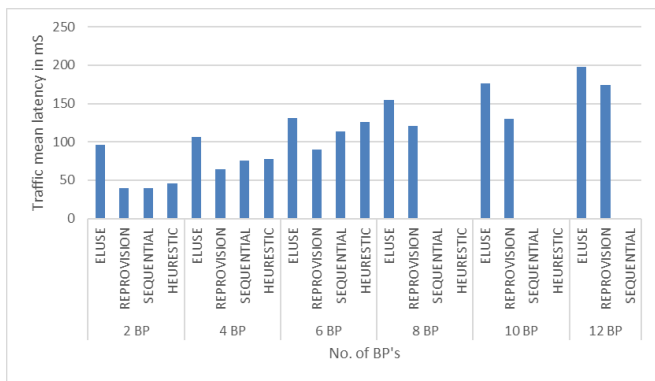


FIGURE 11. Average latency of energy efficient service embedding across different zones without coexistence constraint.

the sequential scenario to 6 BPs, while the re-provisioning embedding still succeeds to embed all the 12 BPs.

Figure 10 summarizes the power consumption results when embedding BP's into the physical IoT network with the coexistence constraint. The power savings achieved by the energy efficient embedding scenarios are reduced to 34% and 17% for re-provisioning and sequential cases, respectively. This reduction is due to embedding of virtual nodes of a BP in different IoT nodes as explained above.

The traffic mean latency resulting from embedding BPs across distinct zones without coexistence constraints are shown in Figure 11. The re-provisioning and sequential

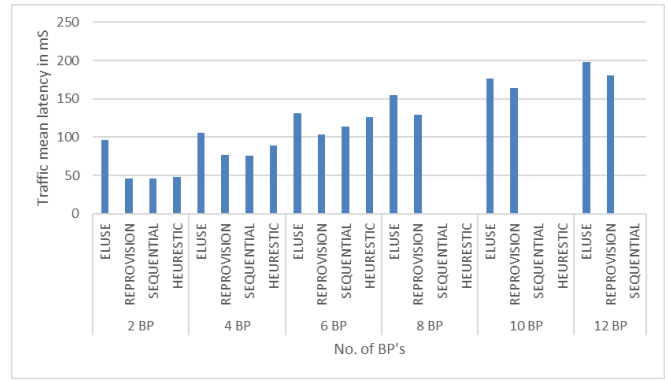


FIGURE 12. Average latency of energy efficient service embedding across different zones with coexistence constraint.

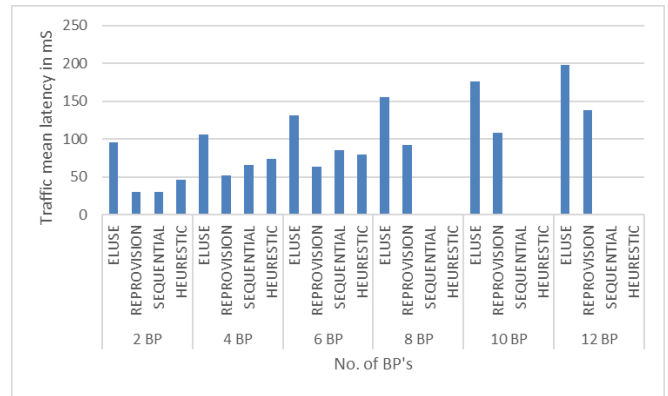


FIGURE 13. Average traffic mean latency of low latency service embedding across different zones without coexistence constraint.

embedding have reduced the average traffic mean latency by 32% and 15% compared with ELUSE scenario.

Figure 12 shows the traffic mean latency resulting from embedding BPs across distinct zones without coexistence constraints. The re-provisioning and sequential embedding have reduced the average traffic mean latency to 22% and 13% compared with ELUSE scenario.

B. LOW LATENCY SERVICE EMBEDDING IN IoT NETWORKS

In this subsection, we evaluate the low traffic mean latency when embedding BPs across different zones with and without the coexistence constraint and also assess the power consumption.

Figure 13 shows that the re-provisioning low latency embedding resulted in reducing the traffic latency by an average of 47% compared to the ELUSE scenario. The low latency embedding model optimizes the selection of IoT nodes and distributes the traffic so that the arrival rate at nodes and consequently the traffic latency is minimized. Under energy efficient embedding, fewer IoT nodes and links are activated to embed BPs compared to embedding under the ELUSE scenario.

The traffic latency reduction achieved by the energy efficient embedding decreases to 20% under the sequential scenario as the sequential approach builds on existing embedding decisions as explained in Section (A). The optimal use

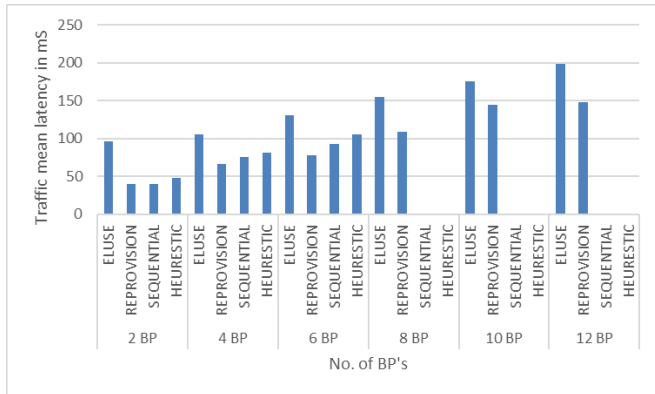


FIGURE 14. Average traffic mean latency of low latency service embedding across different zones with coexistence constraint.

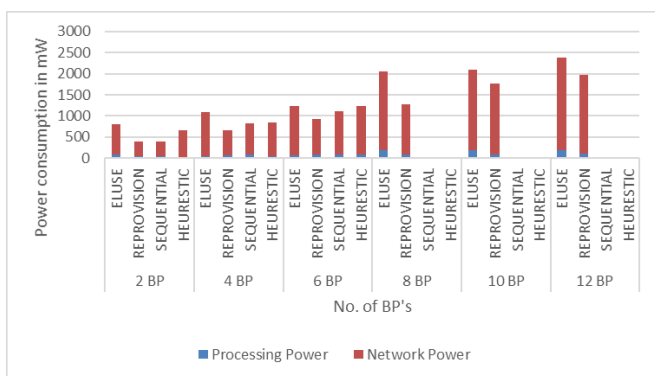


FIGURE 15. Power consumption of low latency service embedding across distinct zones without coexistence constraint.

of resources under the re-provisioning scenario resulted in successfully embedding 12 BPs while only 6 BPs were successfully embedded under sequential embedding.

Figure 14 displays the traffic mean latency of low latency BPs embedding across different zones with the coexistence constraint. Adding the coexistence constraint reduced the traffic latency achieved by the re-provisioning and sequential embedding to 34% and 19%, respectively compared to the ELUSE scenario as more traffic traverses the network due to the fact that multiple virtual nodes of the same BP cannot coexist in the same IoT node.

The results in Figure 15 show the power consumption resulting from low latency embedding across distinct zones without coexistence constraint. Distributing the traffic to reduce the delay increased the power consumption by 28% compared to the energy efficient re-provisioning embedding in Figure 9 as more nodes are activated. However, compared to the ELUSE scenario the power consumption is reduced by 18% and 10% under low latency re-provisioning and low latency sequential embedding, respectively.

Under the coexistence constraint in Figure 16, the increase in power consumption resulting from low latency embedding compared to the energy efficient embedding increased the power consumption by 20% compared to the energy efficient re-provisioning embedding in Figure 10. However, compared to the ELUSE scenario the power consumption is reduced

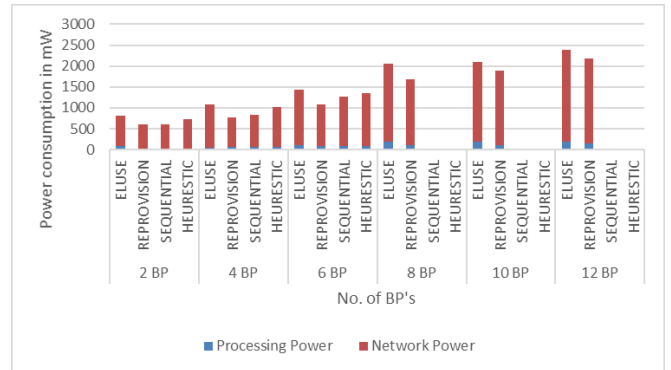


FIGURE 16. Power consumption of low latency service embedding across distinct zones with coexistence constraint.

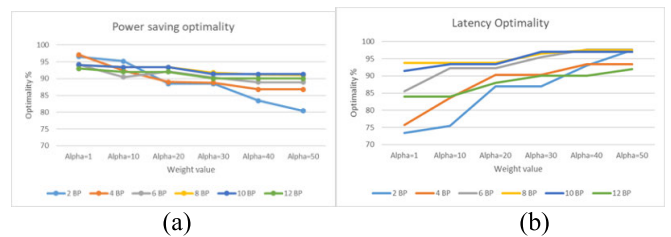


FIGURE 17. Optimality of (a) power saving and (b) traffic mean latency of embedding in distinct zones with coexistence constraint.

by 14% under low latency re-provisioning and sequential embedding.

C. ENERGY EFFICIENT-LOW LATENCY SERVICE EMBEDDING IN IoT NETWORKS

Minimum power consumption is achieved by consolidating the embedding of virtual nodes in the smallest possible number of energy efficient IoT nodes. On the other hand, minimum traffic mean latency is achieved by distributing the traffic into multiple paths to reduce the arrival rate at the individual IoT nodes. As explained in the previous section, the trade-off between minimizing the power consumption and minimizing the traffic mean latency is achieved through a multi-objective MILP model. We define a metric referred to as “embedding optimality” to compare the performance of the multi-objective embedding to single objective embedding. The embedding optimality is defined as follows:

$$Optimality_{QoS} = \frac{Optimal_{Multi-objective}^{QoS}}{Optimal_{Single-objective}^{QoS}} \quad (32)$$

Figure 17 displays the power saving (Figure 17-(a)) and traffic mean latency (Figure 17-(b)) average optimality of energy efficient–low latency service embedding scenario across distinct zones with coexistence constraint under $\alpha = 30$, $\beta = 1$ and $\gamma = 1$ in the multi-objective function (equation (7)). Note that the numerical value of power consumption and traffic latency are comparable, therefore the weight α is used to prioritize traffic latency, while the other two weights in equation (7) are set to one. We obtain equal optimality for power savings and mean traffic latency of 91% at $\alpha = 30$, i.e. this is the weight needed to achieve the trade-off.

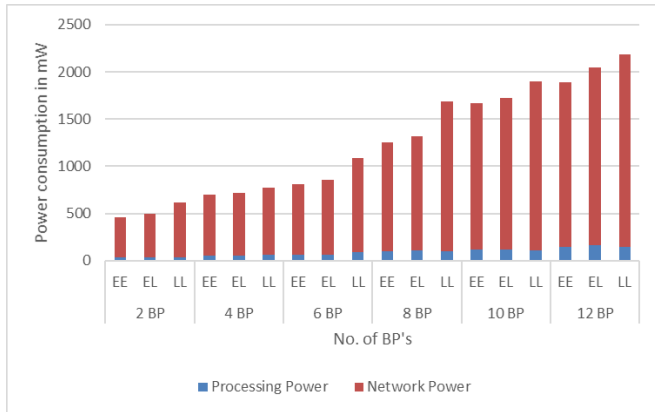


FIGURE 18. Power consumption of embedding in distinct zones with coexistence constraint.

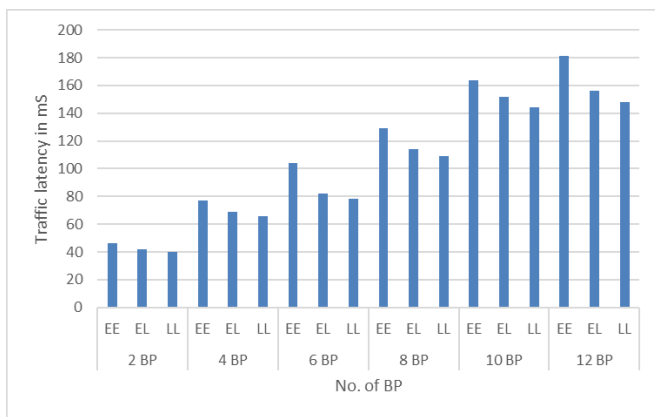


FIGURE 19. Average traffic mean latency of embedding in distinct zones with coexistence constraint.

Figure 18 and 19 compare the power consumption and delay, respectively of the energy efficient–low latency service embedding scenario with $\alpha = 30$ to those of the energy efficient service embedding and low latency service embedding scenarios. Note that the low latency scenario increases the power consumption by 20% compared to the energy efficient scenario (Figure 18) and the energy efficient scenario increases the traffic mean delay by 22% compared to the low latency scenario (Figure 19).

D. REAL TIME ENERGY EFFICIENT SERVICE EMBEDDING HEURISTIC

The RESE heuristic is shown in Figure 20 in pseudocode. The input to the heuristic is the IoT network topology and the BPs. The heuristic starts by sorting the IoT nodes according to the processing power efficiency in descending order and the BPs according to the processing demand of the controller node in ascending order.

The heuristic picks a BP from the ordered list and embeds its nodes one by one considering first the IoT node with the highest energy efficiency that satisfies the embedding requirements in terms of function, zone and coexistence. By doing so the heuristic tries to consolidate virtual nodes into the most energy efficient IoT nodes that meet the demand before activating another IoT node. The available processing

Inputs:

Initialize Virtual Demands
 BP.VN, BP.VN.CPU, BP.VN.RAM, BP.VN.ZONE,
 BP.VN.FUNC, BP.VN.TRF
 Initialize Physical Resources
 IoT.PNB, IoT.DIS, IoT.FUNC, IoT.ZONE, IoT.CPU,
 IoT.RAM, IoT.CPUPWR, IoT.IDLP, IoT.IDLT

Output:

Processing power consumption, Network power consumption, Traffic latency, Embedded BP and correspondent IoT node

Embedding of virtual nodes

```

For each IoT node m ∈ P
  For each BP node k ∈ BP
    For each virtual node l ∈ VN
      If all the physical resource of m satisfies virtual demands
        of VN l in BP l
          Then EMBD(k, l, m)=1
          Update the physical nodes utilization
        End if
      End for
    End for
  End for

```

Embedding of virtual links

```

Initialize SOURCE and DESTINATION for each BP
Select shortest path (TPOL, SOURCE,
DESTINATION)
Update the links utilization and network topology

```

Processing Power calculation

$$\begin{aligned}
 TPP = & \sum_{m \in P} I_m^{PM} \cdot P_m^{IDLECP} \\
 & + \sum_{m \in P} \sum_{k \in BP} \sum_{l \in VN} I_{mkl}^{NE} \cdot P_m^{MAXCP} \\
 & \cdot \frac{V_{kl}^{MCU}}{P_m^{MCU}}
 \end{aligned}$$

Network Power Calculation

$$\begin{aligned}
 TNP = & \sum_{i \in P} I_i^{TM} \cdot P_i^{IDLETP} + 2 \\
 & \cdot \sum_{i \in P} \sum_{j \in PN_i} R_{ij}^{TRFIC} \cdot E_{ij}^{PBT} \\
 & + \sum_{i \in P} \sum_{j \in PN_j} R_{ij}^{TRFIC} \cdot (P_{ij}^{DIST})^2 \cdot F_{ij}^{TR}
 \end{aligned}$$

Queuing latency calculations

```

For each IoT node I in P
  If arrival rate of node (i) > 0
    Q_i^{NODE} = \frac{1}{(\mu_i^{NODE} - \lambda_i^{NODE})}
  End if
End for
TL = \sum_{i \in P} Q_i^{NODE}

```

FIGURE 20. Service embedding heuristic.

capacity of the IoT nodes is updated after the embedding of a virtual node and another virtual node of the BP is selected

TABLE 3. Power consumption gap in mW between the RLSE heuristic and the sequential model.

	2 BP's		4 BP's		6 BP's	
	Sequential MILP	Real time Heuristic	Sequential MILP	Real time Heuristic	Sequential MILP	Real time Heuristic
Processing Power	35	23	48	44	96	56
Network Power	420	684	672	736	987	1149

TABLE 4. Traffic mean latency gap in ms between the RLSE heuristic and the sequential model.

	2 BP's		4 BP's		6 BP's	
	Sequential MILP	Real time Heuristic	Sequential MILP	Real time Heuristic	Sequential MILP	Real time Heuristic
Traffic mean Latency	40	48	76	81	93	106

to be embedded. After embedding all the virtual nodes of a BP, the traffic between the virtual nodes is routed based on shortest path routing [56]. This process is repeated for all BPs and the total power consumption (IoT nodes and network) and traffic mean latency resulting from embedding all the BPs are calculated.

Figure 9 to 12 show that the performance of the RESE heuristic approaches that of the sequential energy efficient MILP model for embedding across different zones. Table 3 summarizes the average performance gap between the RESE heuristic and the sequential model.

E. REAL TIME LOW LATENCY SERVICE EMBEDDING HEURISTIC

The RLSE heuristic reduces the traffic mean latency by setting a threshold on the node transmission capacity utilization. When routing the traffic between virtual nodes of a BP, the heuristic does not exceed this threshold which guarantees distributing the traffic over multiple links. The RLSE heuristic results are given in Figure 16. The threshold is set to 60% of the maximum node capacity. Different thresholds were examined and this threshold value was identified as the maximum threshold before the latency per node starts increasing fast.

Figure 13 to Figure 16 show that the performance of the RLSE heuristic approaches that of the sequential low latency MILP model for embedding across different zones. Table 4 summarizes the average performance gap between the RLSE heuristic and the sequential MILP model.

IV. CONCLUSION

This paper has investigated the power consumption and traffic mean latency of service embedding in the IoT network for a smart building setting and has introduced a framework for their minimization. The services to be embedded are represented by a virtual topology (virtual nodes and links) following a business process workflow dictated by the SOA paradigm. We developed a MILP framework and a real-time heuristic to optimize the selection of IoT nodes to embed the virtual nodes; and to route the traffic between virtual nodes considering three different objective functions:

(i) minimizing the total power consumption, (ii) minimizing traffic mean latency, (iii) minimizing both total power consumption and traffic mean latency in multi-objective manner.

We considered embedding BPs where all the sensor and actuator nodes exist in the same geographical zone and also considered embedding across different zones. We also studied embedding with and without constraints on the coexistence of virtual nodes in the same IoT node.

We used the MILP model to optimize the embedding in two scenarios: (i) re-provisioning scenario where each time a new BPs arrives, previously embedded BPs are re-embedded, (ii) sequential embedding where arriving BPs are embedded without interrupting the existing BPs.

In the energy efficient service embedding scenario, the re-provisioning scenario produces higher average power saving compared with the sequential embedding scenario. In the low latency service embedding scenario, re-provisional embedding reduced the average traffic mean latency compared with the sequential embedding scenario. The multi-objective optimization shows that it is possible to optimize the embedding of BPs to achieve high optimality of 91% for both power savings and traffic latency.

ACKNOWLEDGMENT

The author, Mr. Haider Al-Shammari, would like to thank the Higher Committee for Education Development (HCED) for funding his scholarship. All data are provided in full in the results section of this paper.

REFERENCES

- [1] D. Giusto, A. Iera, G. Morabito, and L. Atzori, *The Internet of Things: 20th Tyrrhenian Workshop on Digital Communications*. New York, NY, USA: Springer, 2010.
- [2] K. A. Evangelos, T. D. Nikolaos, and B. C. Anthony, "Integrating RFIDs and smart objects into a unified Internet of Things architecture," *Adv. Internet Things*, pp. 6–12, Apr. 2011, doi: [10.4236/ait.2011.11002](https://doi.org/10.4236/ait.2011.11002).
- [3] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, "A survey on facilities for experimental Internet of Things research," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 58–67, Nov. 2011.
- [4] Z. T. Al-Azez, A. Q. Lawey, T. E. El-Gorashi, and J. M. Elmighani, "Virtualization framework for energy efficient IoT networks," in *Proc. IEEE 4th Int. Conf. Cloud Netw. (CloudNet)*, Oct. 2015, pp. 74–77.

- [5] S. E. Collier, "The emerging enernet: Convergence of the smart grid with the Internet of Things," *IEEE Ind. Appl. Mag.*, vol. 23, no. 2, pp. 12–16, Mar./Apr. 2017.
- [6] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using big data analytics," *Comput. Netw.*, vol. 101, no. 4, pp. 63–80, Jun. 2016.
- [7] A. Cenedese, A. Zanella, L. Vangelista, and M. Zorzi, "Padova Smart City: An urban Internet of Things experimentation," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2014, pp. 1–6.
- [8] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, protocols, and applications," *J. Elect. Comput. Eng.*, vol. 2017, 2017, Art. no. 9324035, doi: 10.1155/2017/9324035.
- [9] L. Da Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.
- [10] P. Kumar, "Some observations on dependency analysis of SOA based systems," *Int. J. Inf. Technol. Comput. Sci.*, vol. 8, no. 1, p. 54, 2016.
- [11] H. Q. Al-Shammari, A. Lawey, T. El-Gorashi, and J. M. Elmirghani, "Energy efficient service embedding in IoT networks," in *Proc. Wireless Opt. Commun. Conf. (WOCC)*, Apr./May 2018, pp. 1–5.
- [12] H. Q. Al-Shammari, A. Lawey, T. El-Gorashi, and J. M. Elmirghani, "Energy efficient service embedding in IoT over PON," in *Proc. 21st Int. Conf. Transparent Opt. Netw. (ICTON)*, Angers, France, Sep. 2019, doi: 10.1109/ICTON.2019.8840429.
- [13] W. Zhiliang, Y. Yi, W. Lu, and W. Wei, "A SOA based IOT communication middleware," in *Proc. Int. Conf. Mechatronic Sci., Electr. Eng. Comput. (MEC)*, Aug. 2011, pp. 2555–2558.
- [14] M. E. Khanouche, Y. Amirat, A. Chibani, M. Kerkar, and A. Yachir, "Energy-centered and QoS-aware services selection for Internet of Things," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1256–1269, Jul. 2016.
- [15] P. Spiess, S. Karnouskos, D. Guinard, D. Savio, O. Baecker, L. M. S. de Souza, and V. Trifa, "SOA-based integration of the Internet of Things in enterprise services," in *Proc. IEEE Int. Conf. Web Services*, Jul. 2009, pp. 968–975.
- [16] S. Clement, D. W. McKee, and J. Xu, "Service-oriented reference architecture for smart cities," in *Proc. IEEE Symp. Service-Oriented Syst. Eng. (SOSE)*, Apr. 2017, pp. 81–85.
- [17] T. Rault, A. Bouabdallah, and Y. Challal, "Energy efficiency in wireless sensor networks: A top-down survey," *Comput. Netw.*, vol. 67, pp. 104–122, Jul. 2014.
- [18] S. Abdullah and K. Yang, "A QoS aware message scheduling algorithm in Internet of Things environment," in *Proc. IEEE Online Conf. Green Commun. (OnlineGreenComm)*, Oct. 2013, pp. 175–180.
- [19] I. Al-Anbagi, M. Erol-Kantarci, and H. T. Mouftah, "A survey on cross-layer quality-of-service approaches in WSNs for delay and reliability-aware applications," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 525–552, 1st Quart., 2016.
- [20] K. Okafor, G. Ononiwu, S. Goundar, V. Chijindu, and C. Udeze, "Towards complex dynamic fog network orchestration using embedded neural switch," *Int. J. Comput. Appl.*, pp. 1–18, Sep. 2018. [Online]. Available: <https://www.tandfonline.com/loi/tjca20>
- [21] K. C. Okafor, I. E. Achumba, G. A. Chukwudebe, and G. C. Ononiwu, "Leveraging fog computing for scalable IoT datacenter using spine-leaf network topology," *J. Elect. Comput. Eng.*, vol. 2017, Apr. 2017, Art. no. 2363240, doi: 10.1155/2017/2363240.
- [22] K. Okafor, "Dynamic reliability modeling of cyber-physical edge computing network," *Int. J. Comput. Appl.*, pp. 1–11, Apr. 2019, doi: 10.1080/1206212X.2019.1600830.
- [23] H. M. M. Ali, T. E. El-Gorashi, A. Q. Lawey, and J. M. Elmirghani, "Future energy efficient data centers with disaggregated servers," *J. Lightw. Technol.*, vol. 35, no. 24, pp. 5361–5380, Dec. 15, 2017.
- [24] J. M. H. Elmirghani, T. Klein, K. Hinton, L. Nonde, A. Q. Lawey, T. E. H. El-Gorashi, M. O. I. Musa, and X. Dong, "GreenTouch GreenMeter core network energy-efficiency improvement measures and optimization," *IEEE/OSA J. Opt. Commun.*, vol. 10, no. 2, pp. A250–A269, Feb. 2018.
- [25] M. O. I. Musa, T. El-Gorashi, and J. M. Elmirghani, "Bounds on Green-Touch GreenMeter network energy efficiency," *J. Lightw. Technol.*, vol. 36, no. 23, pp. 5395–5405, Dec. 1, 2018.
- [26] B. G. Bathula and J. M. Elmirghani, "Energy efficient optical burst switched (OBS) networks," in *Proc. IEEE Globecom Workshops*, Nov./Dec. 2009, pp. 1–6.
- [27] A. M. Al-Salim, T. E. El-Gorashi, A. Q. Lawey, and J. M. Elmirghani, "Greening big data networks: Velocity impact," *IET Optoelectron.*, vol. 12, no. 3, pp. 126–135, 2017.
- [28] S. Igder, S. Bhattacharya, and J. M. H. Elmirghani, "Energy efficient fog servers for Internet of Things information piece delivery (IoTIPD) in a smart city vehicular environment," in *Proc. 10th Int. Conf. Next Gener. Mobile Appl., Secur. Technol. (NGMAST)*, Aug. 2016, pp. 99–104.
- [29] A. M. Al-Salim, A. Q. Lawey, T. E. El-Gorashi, J. M. Elmirghani, and S. Management, "Energy efficient big data networks: Impact of volume and variety," *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 1, pp. 458–474, Mar. 2018.
- [30] M. S. Hadi, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Big data analytics for wireless and wired network design: A survey," *Comput. Netw.*, vol. 132, pp. 180–199, Feb. 2018.
- [31] A. Q. Lawey, T. E. El-Gorashi, and J. M. H. Elmirghani, "Renewable energy in distributed energy efficient content delivery clouds," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 128–134.
- [32] J. M. H. Elmirghani, L. Nonde, A. Q. Lawey, T. E. H. El-Gorashi, M. O. I. Musa, X. Dong, K. Hinton, and T. Klein, "Energy efficiency measures for future core networks," in *Proc. Opt. Fiber Commun. Conf. Exhib. (OFC)*, Washington, DC, USA: The Optical Society, Mar. 2017, pp. 1–3.
- [33] B. G. Bathula, M. Alresheedi, and J. M. Elmirghani, "Energy efficient architectures for optical networks," in *Proc. London Commun. Symp.*, 2009, pp. 1–3.
- [34] A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "BitTorrent content distribution in optical networks," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 3607–3623, Nov. 1, 2014.
- [35] N. I. Osman, T. El-Gorashi, L. Krug, and J. M. H. Elmirghani, "Energy-efficient future high-definition TV," *J. Lightw. Technol.*, vol. 32, no. 13, pp. 2364–2381, Jul. 1, 2014.
- [36] S. H. Mohamed, T. E. El-Gorashi, and J. M. Elmirghani, "Energy efficiency of server-centric PON data center architecture for fog computing," in *Proc. 20th Int. Conf. Transparent Opt. Netw. (ICTON)*, Jul. 2018, pp. 1–4.
- [37] M. Musa, T. Elgorashi, and J. Elmirghani, "Bounds for energy-efficient survivable IP over WDM networks with network coding," *J. Opt. Commun.*, vol. 10, no. 5, pp. 471–481, 2018.
- [38] X. Dong, T. El-Gorashi, and J. M. H. Elmirghani, "Green IP over WDM networks with data centers," *J. Lightw. Technol.*, vol. 29, no. 12, pp. 1861–1880, Jun. 15, 2011.
- [39] N. I. Osman, T. El-Gorashi, and J. M. H. Elmirghani, "The impact of content popularity distribution on energy efficient caching," in *Proc. 15th Int. Conf. Transparent Opt. Netw. (ICTON)*, Jun. 2013, pp. 1–6.
- [40] X. Dong, T. E. El-Gorashi, and J. M. Elmirghani, "On the energy efficiency of physical topology design for IP over WDM networks," *J. Lightw. Technol.*, vol. 30, no. 11, pp. 1694–1705, Jun. 1, 2012.
- [41] M. Musa, T. Elgorashi, and J. Elmirghani, "Energy efficient survivable IP-over-WDM networks with network coding," *J. Opt. Commun.*, vol. 9, no. 3, pp. 207–217, 2017.
- [42] T. E. El-Gorashi, X. Dong, and J. M. Elmirghani, "Green optical orthogonal frequency-division multiplexing networks," *IET Optoelectron.*, vol. 8, no. 3, pp. 137–148, 2014.
- [43] X. Dong, T. El-Gorashi, and J. M. H. Elmirghani, "IP over WDM networks employing renewable energy sources," *J. Lightw. Technol.*, vol. 29, no. 1, pp. 3–14, Jan. 1, 2011.
- [44] A. N. Al-Quzweeni, A. Q. Lawey, T. E. Elgorashi, and J. M. Elmirghani, "Optimized energy aware 5G network function virtualization," *IEEE Access*, Mar. 2019, doi: 10.1109/ACCESS.2019.2907798.
- [45] L. Nonde, T. E. Elgorashi, and J. M. Elmirghani, "Virtual network embedding employing renewable energy sources," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [46] L. Nonde, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Energy efficient virtual network embedding for cloud networks," *J. Lightw. Technol.*, vol. 33, no. 9, pp. 1828–1849, May 1, 2015.
- [47] Z. T. Al-Azez, A. Q. Lawey, T. E. El-Gorashi, and J. M. H. Elmirghani, "Energy efficient IoT virtualization framework with peer to peer networking and processing," *IEEE Access*, vol. 7, pp. 50697–50709, 2019.
- [48] J. Granjal, E. Monteiro, and J. S. Silva, "Security for the Internet of Things: A survey of existing protocols and open research issues," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1294–1312, Aug. 2015.
- [49] Z. Ke, S. Leng, Y. He, S. Maharjan, and Y. Zhang, "Mobile edge computing and networking for green and low-latency Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 39–45, May 2018.
- [50] Y.-Y. Shih, W.-H. Chung, A.-C. Pang, T.-C. Chiu, and H.-Y. Wei, "Enabling low-latency applications in fog-radio access networks," *IEEE Netw.*, vol. 31, no. 1, pp. 52–58, Jan. 2017.

- [51] N. H. Motlagh, M. Baga, and T. Taleb, "UAV-based IoT platform: A crowd surveillance use case," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 128–134, Feb. 2017.
- [52] J. Huang, Y. Meng, X. Gong, Y. Liu, and Q. Duan, "A novel deployment scheme for green Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 196–205, Apr. 2014.
- [53] J. Martin and Y. M. Suhov, "Fast Jackson networks," *Ann. Appl. Probab.*, vol. 9, no. 3, pp. 854–870, 1999.
- [54] (2017). *SimpleLink Wi-Fi CC3100 BoosterPack Reference Design*. [Online]. Available: <http://www.ti.com/tool/cc3100boost-rd?keyMatch=CC3100%20RF&tsearch=Search-EN-Everything>
- [55] (2019). *MathWork Inc.* Accessed: Mar. 4, 2019. [Online]. Available: <https://www.mathworks.com/help/matlab/ref/graph.shortestpath.html>



HAIDER QAYS AL-SHAMMARI received the B.Sc. and M.Sc. degrees in computer engineering from Al-Nahrain University, Baghdad, Iraq, in 2002 and 2006, respectively, and the Ph.D. degree in electronic and electrical engineering from the University of Leeds, U.K., in 2019. From 2006 to 2008, he worked as the Billing Engineer in telecommunication with ZTE Corporation, Iraq. From 2008 to 2011, he was a Data Communication Engineer in telecommunication with Huawei Technologies, Iraq. From 2011 to 2014, he was working as the Transmission Manager with NEC Partner, Iraq. In 2014, he was working as the Project Manager of Huawei technologies, Iraq. His current research interests include energy efficiency in the IoT networks and service virtualization.



AHMED Q. LAWEY received the B.S. and M.Sc. degrees (Hons.) in computer engineering from the University of Al-Nahrain, Iraq, in 2002 and 2005, respectively, and the Ph.D. degree in communication networks from the University of Leeds, U.K., in 2015.



From 2005 to 2010, he was a Core Network Engineer at ZTE Corporation for telecommunication, Iraq Branch. He is currently a Lecturer in communication networks with the School of Electronic and Electrical Engineering, University of Leeds. His current research interests include energy efficiency in optical and wireless networks, big data, cloud computing, and the Internet of Things.

TAISIR E. H. EL-GORASHI received the B.S. degree (Hons.) in electrical and electronic engineering from the University of Khartoum, Khartoum, Sudan, in 2004, the M.Sc. degree (Hons.) in photonic and communication systems from the University of Wales, Swansea, U.K., in 2005, and the Ph.D. degree in optical networking from the University of Leeds, Leeds, U.K., in 2010. She is currently a Lecturer in optical networks with the School of Electrical and Electronic Engineering, University of Leeds. Previously, she held a postdoctoral research position at the University of Leeds, from 2010 to 2014, where she focused on the energy efficiency of optical networks, investigating the use of renewable energy in core networks, green IP over WDM networks with datacenters, energy efficient physical topology design, energy efficiency of content distribution networks, distributed cloud computing, network virtualization, and big data. In 2012, she was a BT Research Fellow, where she developed energy efficient hybrid wireless-optical broadband access networks and explored the dynamics of TV viewing behavior and program popularity. The energy efficiency techniques developed during her postdoctoral research contributed three out of the eight carefully chosen core network energy efficiency improvement measures recommended by the GreenTouch Consortium for every operator network worldwide. Her work led to several invited talks at GreenTouch, Bell Labs, Optical Network Design and Modeling conference, Optical Fiber Communications Conference, International Conference on Computer Communications, and EU Future Internet Assembly and collaboration with Alcatel Lucent and Huawei.



JAAFAR M. H. ELMIRGHANI (M'92–SM'99) received the B.Sc. degree (Hons.) in electrical engineering from the University of Khartoum, in 1989, and was awarded all four prizes in the department for academic distinction, the Ph.D. degree in the synchronization of optical systems and optical receiver design from the University of Huddersfield, U.K., in 1994, and the D.Sc. degree in communication systems and networks from the University of Leeds, U.K., in 2014.

He joined Leeds, in 2007. Prior to that, he was the Chair in optical communications at the University of Wales Swansea, from 2000 to 2007. He founded, developed, and directed the Institute of Advanced Telecommunications and the Technium Digital (TD), a technology incubator/spin-off hub. He is currently the Director of the Institute of Communication and Power Networks, School of Electronic and Electrical Engineering, University of Leeds. He has provided outstanding leadership in a number of large research projects at the IAT and TD. He has coauthored the *Photonic switching Technology: Systems and Networks* (Wiley) and has published over 450 articles. He has research interests in optical systems and networks.

Prof. Elmirghani was the Chairman of the IEEE Comsoc Transmission Access and Optical Systems Technical Committee and was the Chairman of the IEEE Comsoc Signal Processing and Communications Electronics Technical Committee. He is a Fellow of the IET, Chartered Engineer, and a Fellow of the Institute of Physics. He received the IEEE Communications Society Hal Sobol Award, the IEEE Comsoc Chapter Achievement Award for excellence in chapter activities (both in international competition, in 2005), the University of Wales Swansea Outstanding Research Achievement Award, 2006; and received in international competition: the IEEE Communications Society Signal Processing and Communication Electronics Outstanding Service Award, 2009, and a Best Paper Award at the IEEE ICC'2013. Related to Green Communications, he received (i) the IEEE Comsoc Transmission Access and Optical Systems Outstanding Service Award 2015 in recognition of "Leadership and Contributions to the Area of Green Communications," (ii) the GreenTouch 1000x Award in 2015 for "pioneering research contributions to the field of energy efficiency in telecommunications," (iii) the IET 2016 Premium Award for Best Paper in the *IET Optoelectronics*, and (iv) shared the 2016 Edison Award in the collective disruption category with a team of six from GreenTouch for their joint work on the Green-Meter. He has been awarded in excess of €22 million in grants to date from EPSRC, the EU, and industry, and has held prestigious fellowships funded by the Royal Society and by BT. He was the Co-Chair of the GreenTouch Wired, Core, and Access Networks Working Group, an Adviser to the Commonwealth Scholarship Commission, a member of the Royal Society International Joint Projects Panel, and a member of the Engineering and Physical Sciences Research Council (EPSRC) College. He was the founding Chair of the Advanced Signal Processing for Communication Symposium which started at the IEEE GLOBECOM'99 and has continued since at every ICC and GLOBECOM. He was also the founding Chair of the first IEEE ICC/GLOBECOM Optical Symposium at GLOBECOM'00, the Future Photonic Network Technologies, as well as Architectures and Protocols Symposium. He chaired this Symposium, which continues to date under different names. He was the Founding Chair of the First Green Track at the ICC/GLOBECOM at GLOBECOM 2011, and is the Chair of the IEEE Green ICT initiative within the IEEE Technical Activities Board (TAB) Future Directions Committee (FDC), a pan IEEE Societies initiative responsible for Green ICT activities across the IEEE (2012–present). He is and has been on the technical program committee of the 34 IEEE ICC/GLOBECOM conferences, from 1995 to 2016, including 15 times as the Symposium Chair. He has given over 55 invited and keynote talks over the past eight years. He is currently an Editor of the *IET Optoelectronics* and the *Journal of Optical Communications*, and was an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS SERIES ON GREEN COMMUNICATIONS AND NETWORKING. He is also an Editor of the *IEEE Communications Magazine*. He was an IEEE Comsoc Distinguished Lecturer, from 2013 to 2016.

• • •