IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 11 December 2023, accepted 5 January 2024, date of publication 10 January 2024, date of current version 10 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3352024

RESEARCH ARTICLE

Resilient Energy Efficient IoT Infrastructure With Server and Network Protection for Healthcare Monitoring Applications

IDA SYAFIZA M. ISA[1], **TAISIR E. H. EL-GORASHI**[2], **MOHAMED O. I. MUSA**[3], **(Member, IEEE), AND JAAFAR M. H. ELMIRGHANI**[2], **(Fellow, IEEE)**

[1]Faculty of Technology and Engineering Electronic and Computer, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka 76100, Malaysia
[2]Faculty of Natural, Mathematical and Engineering Sciences, King's College London, WC2R 2LS London, U.K.
[3]School of Electrical and Electronic Engineering, University of Leeds, LS2 9JT Leeds, U.K.

Corresponding author: Ida Syafiza M. Isa (idasyafiza@utem.edu.my)

This work was supported in part by the PJP through Universiti Teknikal Malaysia Melaka (UTeM), Engineering and Physical Sciences Research Council (EPSRC), under Grant PJP/2022/FTKEE/S01883; in part by the INTelligent Energy aware NETworks (INTERNET) under Grant EP/H040536/1; in part by the SwiTching And tRansmission (STAR) under Grant EP/K016873/1; and in part by the Terabit Bidirectional Multi-user Optical Wireless System (TOWS) Project under Grant EP/S016570/1. The work of Ida Syafiza M. Isa was supported by the Ministry of Education, Malaysia, and Universiti Teknikal Malaysia Melaka (UTeM), for his Ph.D. studies, and for the approved fund which makes this important research viable and effective.

ABSTRACT Fog computing has been introduced to extend the cloud services by bringing the services near to the user's proximity. However, the distributed location of the fog servers requires a proper management to ensure the network to provide a service resilience during disruption while preserving the energy consumption of the networking and processing equipment. In this paper, a 1+1 server protection scheme where a primary and a secondary processing server are used to serve Electrocardiogram (ECG) monitoring IoT applications concurrently has been considered at the fog networking infrastructure. The infrastructure is designed to be resilient against server failures related to the geographic location of primary and secondary servers and against both server and network failures. A Mixed Integer Linear Programming (MILP) model is developed to optimize the number and locations of the processing servers for energy-efficient resilient fog infrastructure. The results reveal that considering server protection without geographical constraints resulted in network and processing energy penalties as the traffic is doubled compared to the non-resilient scenario. Meanwhile, considering geographical constraints for server protection at low demands resulted in high network energy penalty as more nodes are used to host the processing servers. Interestingly, increasing the resilience level to consider network protection with link and node disjoints selection at high demand resulted in low network energy penalty due to the activation of a large part of the network in any case to serve the demands. The results also reveal that the network energy penalty was reduced when more processing servers are allowed at each fog node while the same processing energy is consumed regardless of the increased resilience level. A heuristic was developed for each resilience scenario for verification and to enable real-time operation of the network, servers and IoT devices, and the results of the heuristic approach those of the MILP.

INDEX TERMS ECG monitoring, energy consumption, fog computing, GPON, health monitoring, Internet of Things, machine-to-machine (M2M) communication, network protection, resilience, server protection.

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

I. INTRODUCTION

Cloud computing technologies can provide computation and storage services anytime and anywhere. However, offloading massive amounts of data generated by end Internet of Things (IoT) devices to the cloud for computation requests increases the network's congestion. Also, it can increase both networking and processing energy consumption. Many researchers focused on improving the energy efficiency of the architectures for the core network and cloud data centres under increasing applications' traffic [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. Different methods and technologies are used to increase the energy efficiency of the network. These include virtualization [6], [11], [17], designing and optimizing the network architecture [14], [18], [19], [20], optimizing content distribution [12], [13], [21], progressive big data processing [3], [5], [7], [22], network coding [4], [16] and using renewable energy to reduce the carbon footprint of the network further [15]. Also, fog networks that integrate distributed edge servers in a decentralized architecture have been proposed to reduce the burden on central data centers [15], [23], [24]. In our previous work in [25], we have shown that there is 68% total energy saving when fog computing is used to serve electrocardiogram (ECG) monitoring IoT applications to save heart patients while considering the time restriction enforced by the American Heart Association (AHA) compared to the traditional cloud computing approach. Meanwhile, in [20], we further extend our work in [25] to consider a realistic case study under both low and high data rate applications which resulted in 36% and 52% energy efficiency improvements, respectively, when the health data are processed and analyzed at the fog compared to the cloud.

Trustworthiness is one the challenges elements when dealing with the Internet of Things (IoT) system [26]. This consists of the system requirements concerning several aspects including resilience to recover the communication services between the network equipment due to network failures [27]. For instance, ensuring the data trustworthiness in the IoT system will also ensure the resilient of the data flow. Many approaches have been introduced to improve service resilience at the cloud networking infrastructure as surveyed in [28], ranging from designing and operating the cloud facilities, servers and networks with resilience in mind to their integration and virtualization. In [29], the concept of virtualization was used to allow the backup servers to be shared in geo-distributed data centers, which improved the utilization of backup servers by 40%. However, the proposed shared protection scheme requires high reserved bandwidth and can increase the latency between the backup and primary servers. Meanwhile, the work in [30] studied the impact of the relocation of the backup and primary servers on the cost of the network and its capacity and the servers. The study revealed that considering protection against single link failures with relocation reduces the cost associated with the capacity of the servers and network. Furthermore, the work also showed that the benefits of repositioning the backup and primary servers are more noticeable for sparser network topologies. In [31], the authors proposed a recovery technique which is the self-triggered update scheme that dynamically perform the rerouting process of the communications nodes at the network level. This is to ensure that the network will still be functioning even the network nodes do not operate due to faults or attacks.

Fog computing can also improve service resilience by performing data processing at the edge of the network. For instance, in [32], the authors show that using fog computing to process the data at the network edge can improve network resilience besides reducing the latency compared to processing the data at cloud computing, mainly for an interactive demand. In [33], the authors proposed a set of mechanism to serve the service requests for chained Virtual Functions to process specific requirement of application while increasing their resilience using Integer Linear Programming Model. The results show that, the proposed method has the possibility to increase the resilience of the chained Virtual Function while balancing the services requests of the infrastructure nodes. In [34], the author considered the used of Virtual Network Function (VNFs) and Service Function Chain (SFC) to deliver the service demands. A multi-path protection (MP) scheme has been proposed to protect the SFC in which the MP enable the SFC to be split on multiple disaster zone (DZ) disjoint working path. The results reveal that compared to the dedicated protection (DP) schemes, the proposed MP scheme has improved the resource consumption by up to 20%. Meanwhile, in [35], the author proposed a fault tolerance framework for micro service execution considering the fog-IoT orchestration to ensure the efficiency of the network to recover after failure occurs in the system. The results show that, the proposed framework able to perform the seamless transfer of micro services in a fog-IoT ecosystem. In [36], an adaptive traffic signal control algorithm is introduced to dynamically adjust the timing's signal based on the current traffic density to reduce the average wait times of vehicles while maximizing the bandwidth utilization at the fog and cloud layer. To implement this system, the real-time videos captured at the fog layers will be processed locally and send the processed data i.e. count of vehicles to the cloud layer. Meanwhile, at the cloud layer, a KKN-based Machine Learning model will train the processed data from the fog layer to estimate the traffic density of that particular time. In case of a fog node failure, the predicted values will be used to set the signal time. The results show that compared to the static method, the average wait times of vehicles and the bandwidth utilization is reduced by 17.624% and 99.99%, respectively. The work in [37] proposed a fog-to-cloud (F2C) scheme to organize the management strategies to control resources from the cloud to the edge for resilience purposes. Three strategies, Zero Knowledge (ZK), Keep Updating (KU) and High-Layer download (HLD) have been used to evaluate the proposed scheme in which a particular mode failed in the F2C architecture. The results show that, the keep updating (KU) methods requires lower synchronization time as it stores a copy of the leader database. However, the KU method increases the overhead in the network. They also studied the impact of considering multiple backup options and the results reveal that, the higher the number of backup required, the higher the network utilization. Table 1 summarizes the finding from the literature review.

However, to the best of our knowledge, no work has focused on increasing the networking and processing equipment's energy efficiency while improving the service resilience. As mentioned above, in [20], the results have shown that 36% and 52% of energy efficiency of networking equipment are achieved when considering fog computing in the network infrastructure, compared to the cloud computing infrastructure for low and high data rate applications, respectively. Therefore, in this work, we furthermore consider server and network protection at the fog networking infrastructure level. In [38], we have proposed a resilient fog computing infrastructure for health monitoring applications with a 1+1 protection scheme. In this scheme, two processing servers (PSs), primary and secondary, are utilized to aid the ECG monitoring IoT applications simultaneously in a West Leeds, United Kingdom setting. The patients send the necessary data to the primary and secondary PSs to be processed, analyzed and for decision making. A Mixed Integer Linear Programming (MILP) model was used to obtain the optimal number and locations of the PSs to reduce the energy consumed by the networking and processing equipment. We use MILP modelling to mathematically represent the proposed architecture. Using the MILP model, we obtain the optimal the number and locations of processing servers for an energy-efficient resilient fog infrastructure which sets the upper bound on the performance of the proposed architecture [39]. In [38] preliminary results, were presented that show that only network energy consumption is affected when resilience is increased to consider geographical constraints compared to a scenario with no geographical constraints for server protection.

The current paper is based on chapter 6 in [40] which makes several new contributions beyond those presented in [38] as summarized below:

i. Compared to [38], it provides for the first time the developed MILP model for both scenarios without and with geographical constraints for server protection.
ii. Evaluation on the network and processing energy consumption under both scenarios: without and with geographical constraints, considering a wide range of parameters and scenarios in terms of the total PSs allowed at each candidate fog node (i.e. from 1 up to 8 processing server (PS) per candidate fog node).
iii. Evaluation on the energy penalty for network and processing in a resilient scenario with no geographical constraints compared to the non-resilient scenario.
iv. Development on two new heuristics: our new energy optimized resilient infrastructure fog computing without geographical constraints (EORIWG) heuristic, and the energy optimized resilient infrastructure fog computing with geographical constraints (EORIG) heuristic. Also, the evaluation on the performance gaps between the heuristic and the MILP model results in terms of the total energy consumed by the networking and processing equipment are presented. The heuristics are simpler than the MILP, provide validation for the MILP and enable real-time operation.
v. Development of further MILP model that extends our infrastructure so that it is resilient against the server and network failures, to determine the optimal number of primary and secondary PSs and their optimal locations while minimizing the energy consumed by networking and processing equipment. We consider geographical constraints for server protection while introducing disjoint links and node selection for network protection, offering a design with higher levels of resilience. In this design, the different nodes required to host both the primary and secondary PSs and the links and nodes used to transmit the data to and from primary and secondary PSs are disjoint, as node and link failures in the network are not improbable. We consider disjoint links and nodes only at the access layer, as the PSs can only be placed in the access layer. An Energy optimized resilient infrastructure fog computing with geographical constraints, and link and node disjoint (EORIGN) heuristic is developed for real-time implementation, a third new heuristic in this paper.

The rest of the paper is structured as follows: Section II introduces the proposed resilient fog computing IoT infrastructure architecture for healthcare monitoring applications. Section III provides the MILP mathematical modelling of the proposed approach considering a Gigabit Passive Optical Network (GPON) network at the access layer. Next, Section IV introduces the parameters considered in this work, while section V presents the health monitoring IoT application's performance. The development of each heuristic for each considered protection scenario is given in Section VI, while Section VII discusses their performance evaluation. Finally, Section VIII concludes this paper.

II. THE PROPOSED RESILIENT FOG COMPUTING ARCHITECTURE FOR IoT HEALTH MONITORING APPLICATIONS

In this work, we consider the resilient fog computing IoT architecture with a GPON access network for health monitoring applications in [38] and [40]. However, for the reader's convenience, we re-introduce the architecture in Figure 1. The architecture consists of four layers. The details of each layer are as explained in [38] and [40].

III. MILP MODEL FOR ENERGY-EFFICIENT AND RESILIENT INFRASTRUCTURE FOG COMPUTING IoT HEALTH MONITORING APPLICATIONS

In [20], we focused on energy minimization while this paper extends the previous work by introducing resilience. Therefore, in this section, a MILP model is developed to minimize the energy consumed by the networking and

TABLE 1. A comparison of the contributions of this paper with related works in the literature.

Ref	Location of server		Type of Resilient					Method		Results			
	Processing Layer		Server Failure		Server & Network Failure		Controll-ability	Mixed Integer Linear Progra-mming	Integer Linear Program-ming	Optimization Consideration			
	Fog	Cloud	Without geograp-hical	With geograph y-yical	Link disjoint	Link and Node disjoint				Energy	Cost	Quality of Services	Fault toleran-ce
[29]		√		√					√		√		
[30]		√			√						√		
[31]		√			√							√	
[32]	√											√	
[33]	√	√				√			√			√	
[34]		√			√						√		
[35]	√						√						√
[36]	√	√					√					√	
[37]	√	√					√					√	
[38]	√	√	√	√				√		√			
Our work	√	√	√	√		√		√		√			

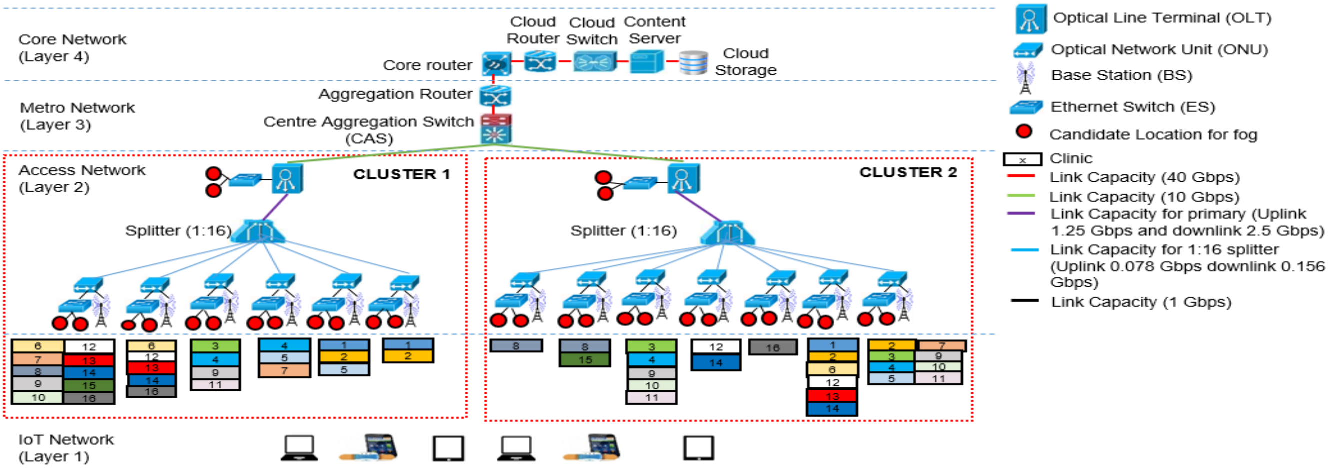


FIGURE 1. Resilient fog computing IoT infrastructure for health monitoring applications [38].

processing equipment for the resilience scenarios that consider a geographic location for server protection where scenario 1 consider server protection with no geographical constraint while scenario 2 consider server protection with geographical constraint. We further extend the MILP model to consider server and network protection with the geographic location constraints and link and node disjoint design which is scenario 3, with the same objective function. Note that energy consumption of networking equipment comprises of the energy consumed by all networking devices at the access, metro and core network. In contrast, the energy consumption of processing includes the energy consumed by both primary and secondary PSs. It is worth noting that Long Term Evolution Machine (LTE-M) base stations (BSs) are used in the proposed architecture to aggregate traffic at the IoT network.

A. SCENARIO 1: PROTECTION FOR SERVER WITHOUT GEOGRAPHICAL CONSTRAINTS

To model the energy consumption minimization approach considering server protection with no geographical constraints, i.e. Scenario 1, we utilized the same sets, parameters, variables and objective function as in [20]. We furthermore introduce additional variables as in Table 2. Note that we considered a setting where each candidate fog node can host one or more PS. Therefore ϕ_d, the number of processing servers at node d, is set as a variable.

The considered power consumed by the networking devices and PSs are decomposed into two parts: idle and linear proportional part that increases with load. The Ethernet switches and the PSs are dedicated only for healthcare applications (i.e. unshared) while various applications share the other devices. Hence, we only consider a portion of the idle power contributed by the healthcare application (x) for the shared devices while for the unshared devices, the maximum idle power is considered. Also, the energy consumption of all devices is proportional to the time the devices are utilized and the load placed on the devices to serve the workload. The equations used to calculate the energy consumed at the access network and the energy consumed at the core network are as in [20]. However, as the proposed resilience architectures

TABLE 2. Sets, parameters and variables used in MILP.

Variables	
ωa_{sd}	Total patients at clinic s assisted by primary PSs placed at fog node d, $s \in CL, d \in FN$
ωb_{sd}	Total patients at clinic s assisted by secondary PSs placed at fog node d, $s \in CL, d \in FN$
Ya_d	$Ya_d = 1$, if there is primary PS placed at fog node d, otherwise $Ya_d = 0$, $d \in FN$
Yb_d	$Yb_d = 1$, if there is a secondary PS located at fog node d, otherwise $Yb_d = 0$, $d \in FN$
z_d	z_d is a dummy variable with a value $Ya_d \oplus Yb_d$, where $\oplus$ is an XOR operation, $d \in FN$
y_d	y_d is a dummy variable with value $Ya_d . Yb_d$, where . is an AND operation, $d \in FN$
ϕa_d	Total primary PSs hosted at fog node d, $d \in FN$
ϕb_d	Total secondary PSs hosted at fog node d, $d \in FN$
τpa_d	Time to process and analyze raw data (in seconds) at primary PS at fog node d, $d \in FN$
τpb_d	Time to process and analyze raw data (in seconds) at secondary PS at fog node d, $d \in FN$

considers two clusters with additional secondary servers, we redefined the equations to calculate the energy consumed at the metro network, the energy consumed at the cloud, and the fog nodes' energy consumption.

1) ENERGY CONSUMPTION OF METRO NETWORK

The energy consumption of the metro network (EMN) includes the energy consumed by centre aggregation switches (CASs) and aggregation routers. The aggregation router is only used to transmit the analyzed health data storage traffic as the candidate fog nodes allowed to host the PSs are at the access network. Meanwhile, the CASs are used to send the raw health data traffic, analyzed health data feedback traffic and analyzed health data storage traffic between different clusters. Therefore, the energy consumed at the metro network is as follows:

$$EMN = (ECASP + ECASF + ECASS + EARS)\,\eta \quad (1)$$

The details calculation to determine the energy consumption of the CASs and aggregation routers to perform the tasks is the same as in [20].

2) ENERGY CONSUMPTION OF CLOUD

The energy consumed in the cloud is calculated as in [20]. However, for cloud storage, the storage traffic, S_i is divided by '2' as only one analyzed health data from both primary and secondary PSs is stored as shown in (2).

$$ECSTS = 2 \sum_{i \in CST} \left(ICSTx\zeta c_i + \frac{S_i}{2} \tau c \frac{(PCST - ICST)}{CCST} \right) \tau c \quad (2)$$

3) ENERGY CONSUMPTION OF FOG NODES

The energy consumed at fog node (EFN) includes the energy consumption of PSs (EPS) and the energy consumption of Ethernet switches ($ETES$) as in [20]. However, as the secondary PS is considered for resilience purposes, the energy consumed by PSs (EPS) is as given in (3).

$$EPS = \sum_{d \in FN} \left(IPS(\phi a_d + \phi b_d)(\tau a + \tau b + \tau c) + PPS(\tau pa_d + \tau pb_d) \right) \quad (3)$$

The following are the modified and additional constraints used in addition to the constraints in [20], to model the energy consumption minimized approach considering server protection with no geographical restrictions:

Subject to:

$$\omega a_{sd} \leq Pt_s Ya_d; \quad \forall s \in CL, \forall d \in FN \quad (4)$$

$$\omega b_{sd} \leq Pt_s Yb_d; \quad \forall s \in CL, \forall d \in FN \quad (5)$$

Constraints (4) and (5) are used to allocate patients from clinic s, to be served by the primary and secondary PSs located at node d, respectively. Note that if a patient is assigned to a candidate location, this location should have fog servers.

$$\sum_{d \in FN} \omega a_{sd} = Pt_s; \quad \forall s \in CL \quad (6)$$

$$\sum_{d \in FN} \omega b_{sd} = Pt_s; \quad \forall s \in CL \quad (7)$$

Constraints (6) and (7) ensure that all patients at clinic s, are assigned to the primary and secondary PSs located at any node d, respectively.

$$P_{sd} = (\omega a_{sd} + \omega b_{sd})\,\delta a; s \quad \in CL, d \in FN \quad (8)$$

$$F_{sd} = (\omega a_{ds} + \omega b_{ds})\,\delta b; \quad \forall s \in FN, d \in CL \quad (9)$$

$$S_{sd} = \sum_{i \in CL} (\omega a_{is} + \omega b_{is})\,\delta c \delta_{sd}; \quad \forall s \in FN, d \in CST \quad (10)$$

Constraint (8) determines the total raw health data traffic, while constraints (9) - (10) determine the total analyzed health data traffic for feedback and storage, respectively. This is done by considering the association of patients from the clinic to the PS (i.e. ωa_{sd} for primary PS and ωb_{sd} for secondary PS) and the data rate allocated per patient depending on the tasks (i.e. $\delta a/\delta b/\delta c$) to perform the transmission. Note that, as in [20], single cloud storage is considered, hence in (10), $\delta_{sd} = 1$.

$$Ya_d + Yb_d = 2Y_d - z_d; \quad \forall d \in FN \quad (11)$$

$$\phi a_d + \phi b_d \leq N; \quad \forall d \in FN \quad (12)$$

Equation (11) determines the nodes that are used to place the PSs where $Y_d = 1$ if any of Ya_d and Yb_d are equal to 1 ($Ya_d + Yb_d$), otherwise, it is zero. This is achieved by using a binary variable z_d which will be 1 if Ya_d and Yb_d are exclusively equal to 1 ($Ya_d \oplus Yb_d$). Otherwise, z_d is equal to zero. Constraint (12) makes sure that the total number of PSs at the selected fog node d does not exceed the maximum number of PSs permitted at each fog node N.

$$\sum_{s \in CL} \omega a_{sd} \leq \Omega max \phi a_d; \quad \forall d \in FN \quad (13)$$

$$\sum_{s\in CL} \omega b_{sd} \leq \Omega max \phi b_d; \quad \forall d \in FN \tag{14}$$

Constraints (13) and (14) ensure the total patients aided by each primary and secondary PS at node d, respectively, do not exceed the maximum allowable users (Ωmax). However, the model also allows more than one PS (i.e. ϕa_d for primary PS and ϕb_d for secondary PS), to be deployed at the same node d if the number of users is higher than Ωmax.

$$\tau pa_d = \sum_{s\in CL} m\omega a_{sd} + \acute{c}\phi a_d; \forall d \in FN \tag{15}$$

$$\tau pb_d = \sum_{s\in CL} m\omega b_{sd} + \acute{c}\phi b_d; \forall d \in FN \tag{16}$$

Equations (15) and (16) determine the time consumed to process and analyze the health data at the primary PS and secondary PS hosted at node d, respectively. This is done by considering the total patients aided by the PS (i.e. ωa_{sd} for primary PS and ωb_{sd} for secondary PS) and the number of PSs utilized (i.e. ϕa_d for primary PS and ϕb_d for secondary PS), where $m$and $\acute{c}$ represent the gradient and y-intercept, respectively.

$$\sum_{s\in CL} \omega a_{sd}\alpha \leq \Lambda max \phi a_d; \quad \forall d \in FN \tag{17}$$

$$\sum_{s\in CL} \omega b_{sd}\alpha \leq \Lambda max \phi b_d; \quad \forall d \in FN \tag{18}$$

Constraints (17) and (18) ensure that the storage capacity of a primary PS and secondary PS at node d, do not exceed its maximum capacity (Λmax), respectively. This is done by considering the total patients aided by the PS (i.e. ωa_{sd} for the primary PS and ωb_{sd} for the secondary PS) as well as the size of analyzed health data per patient (α). Note that the model also allows more than one primary PSs (ϕa_d) and secondary PSs (ϕb_d) to be deployed at the same fog node d, if the data's size is larger than Λmax.

B. SCENARIO 2: PROTECTION FOR SERVER WITH GEOGRAPHICAL CONSTRAINTS

This section considers server protection with geographical constraints, i.e. Scenario 2 in which a single fog node is not allowed to host both the primary PS and secondary PS at the same time. Typically, most service providers place their primary and secondary services in distant locations to increase resilience. For example, BackupVault, which is a leading provider of online cloud backup for businesses in the United Kingdom (UK), locate their primary data centre in Slough, UK; while the secondary data centre for redundancy is located in Reading, UK [41]. Therefore, this work considers a setting where the nodes housing the primary PSs are not permitted to house any secondary PSs. The same parameters, variables, constraints and objective functions in the previous scenario in Section III-A are utilized. However, to ensure the locations of both primary PSs and secondary PSs are different, constraint (11) is replaced with equation (19), as shown below:

$$Ya_d + Yb_d = Y_d + 2y_d; \forall d \in FN \tag{19}$$

where constraint (19) ensures that either primary or secondary PSs can be placed at one location d. This is achieved by presenting a new binary variable y_d which will be equal to 1 if Ya_d and Yb_d are equal to 1 ($Ya_d.Yb_d$). Otherwise, y_d is equal to zero.

C. SCENARIO 3: PROTECTION FOR SERVER CONSIDERING GEOGRAPHICAL CONSTRAINTS AND PROTECTION FOR NETWORK CONSIDERING LINK AND NODE DISJOINT DESIGN

This section considers server protection with geographical constraint and network protection with link and node disjoints, i.e. Scenario 3, in which the primary PSs and secondary PSs are not permitted to be hosted at the same candidate fog node, and the links and nodes used to relay the traffic to and from primary PSs and secondary PSs are disjoint. Beyond the optical line terminal (OLT) and heading to the cloud, the network is not protected. This is because the server that did the processing has a copy of the data to be stored and can retain it until the network beyond the OLT recovers. Note that we considered the disjoint links and nodes to be only at the access layer. The same sets, parameters, variables, constraints and objective functions in Section III-B are utilized, and additional sets and variables are introduced in Table 3 to determine the optimal number of the primary and secondary PSs with their optimal locations while considering the geographical constraints and link and node disjoint resilience so that the minimum total networking and processing equipment energy is consumed.

In addition to the constraints presented in Section III-B, the following new constraints are considered:

$$Pa_{sd} = \omega a_{sd}\delta a; \quad s\in CL, d \in FN \tag{20}$$

$$Pb_{sd} = \omega b_{sd}\delta a; \quad s\in CL, d \in FN \tag{21}$$

Constraints (20) and (21) calculate the traffic of the raw health data from clinic s to the primary and secondary PSs located at node d, respectively. This is done by considering the association of patients from clinic to PSs (i.e. ωa_{sd} and ωb_{sd}) and the provisioned data rate per patient (δa) to perform the transmission.

$$Fa_{sd} = \omega a_{sd}\delta b; \quad s\in FN, d \in CL \tag{22}$$

$$Fb_{sd} = \omega b_{sd}\delta b; \quad s\in FN, d \in CL \tag{23}$$

Constraints (22) and (23) calculate the feedback traffic (i.e. analyzed health data) from primary and secondary PSs placed at node s to the clinic d, respectively. This is done by considering the association of patients from clinic to PSs (i.e. ωa_{sd} and ωb_{sd}) and the provisioned data rate per patient (δb) to perform the transmission.

$$Sa_{sd} = \sum_{i\in CL} \omega a_{is}\delta c\delta_{sd}; \quad s\in FN, d \in CST \tag{24}$$

TABLE 3. Additional variables used for the MILP model.

Set	
ND	Set of BSs, ONUs and OLTs (access layer)
Variables	
Pa_{sd}	Total traffic of raw health data from clinic s to primary PSs at target node d (bps), $s \in CL, d \in FN$
Pb_{sd}	Total traffic of raw health data from clinic s to secondary PSs at target node d (bps), $s \in CL, d \in FN$
Pa_{ij}^{sd}	Total traffic of raw health data from clinic s to primary PSs at target node d passing the link between nodes i and j (bps), $s \in CL, d \in FN,\ i, j \in N$
Pb_{ij}^{sd}	Total traffic of raw health data from clinic s to secondary PSs at target node d passing the link between nodes i and j (bps), $s \in CL, d \in FN,\ i, j \in N$
Fa_{sd}	Total traffic of analyzed health data feedback from primary PSs at node s to clinic at node d (bps), $s \in FN, d \in CL$
Fb_{sd}	Total traffic of analyzed health data feedback from secondary PSs at node s to clinic at node d (bps), $s \in FN, d \in CL$
Fa_{ij}^{sd}	Total traffic of analyzed health data feedback from primary PSs at node s to clinic at node d passing the link between nodes i and j (bps), $s \in FN, d \in CL, i, j \in N$
Fb_{ij}^{sd}	Total traffic of analyzed health data feedback from secondary PSs at node s to clinic at node d passing the link between nodes i and j (bps), $s \in FN, d \in CL, i, j \in N$
Sa_{sd}	Total traffic of analyzed health data storage from primary PSs at node s to cloud storage at node d (bps), $s \in FN, d \in CST$
Sb_{sd}	Total traffic of analyzed health data storage from secondary PSs at node s to cloud storage at node d (bps), $s \in FN, d \in CST$
Sa_{ij}^{sd}	Total traffic of analyzed health data storage from primary PSs at node s to cloud storage at node d passing the link between nodes i and j (bps), $s \in FN, d \in CST, i, j \in N$
Sb_{ij}^{sd}	Total traffic of analyzed health data storage from secondary PSs at node s to cloud storage at node d passing the link between nodes i and j (bps), $s \in FN, d \in CST, i, j \in N$
La_{ij}	$La_{ij} = 1$, if the incoming and/or outgoing traffic of primary PSs passes through the link between nodes i and j otherwise $La_{ij} = 0$
Lb_{ij}	$Lb_{ij} = 1$, if the incoming and/or outgoing traffic of secondary PSs passes through the link between nodes i and j otherwise $Lb_{ij} = 0$
ρa_i	$\rho a_i = 1$, if the incoming and/or outgoing traffic of primary PSs passes through node i, otherwise $\rho a_i = 0$
ρb_i	$\rho b_i = 1$, if the incoming and/or outgoing traffic of secondary PSs passes through node i, otherwise $\rho b_i = 0$

$$Sb_{sd} = \sum_{i \in CL} \omega b_{is} \delta c \delta_{sd};\ \ s \in FN, d \in CST \tag{25}$$

Constraints (24) and (25) calculate the storage traffic (i.e. analyzed health data) from primary and secondary PSs placed at node s to cloud storage, d, respectively. This is done by considering the association of patients from clinic to PSs (i.e. ωa_{is} and ωb_{is}) and the provisioned data rate per patient (δc) to perform the transmission. Note that, as explained above, δ_{sd} is set as a parameter that is equal to 1. Next, the traffic flow constraints are considered as follows:

$$\sum_{j \in Nm[i]: i \neq j} Pa_{ij}^{sd} - \sum_{j \in Nm[i]: i \neq j} Pa_{ji}^{sd} = \begin{cases} Pa_{sd}\, if i = s \\ -Pa_{sd}\, if i = d \\ 0 \ otherwise \end{cases}$$
$$s \in CL, d \in FN, i \in N \tag{26}$$

$$\sum_{j \in Nm[i]: i \neq j} Pb_{ij}^{sd} - \sum_{j \in Nm[i]: i \neq j} Pb_{ji}^{sd} = \begin{cases} Pb_{sd}\, if i = s \\ -Pb_{sd}\, if i = d \\ 0 \ otherwise \end{cases}$$
$$s \in CL, d \in FN, i \in N \tag{27}$$

$$\sum_{j \in Nm[i]: i \neq j} Fa_{ij}^{sd} - \sum_{j \in Nm[i]: i \neq j} Fa_{ji}^{sd} = \begin{cases} Fa_{sd}\, if i = s \\ -Fa_{sd}\, if i = d \\ 0 \ otherwise \end{cases}$$
$$s \in FN, d \in CL, i \in N \tag{28}$$

$$\sum_{j \in Nm[i]: i \neq j} Fb_{ij}^{sd} - \sum_{j \in Nm[i]: i \neq j} Fb_{ji}^{sd} = \begin{cases} Fb_{sd}\, if i = s \\ -Fb_{sd}\, if i = d \\ 0 \ otherwise \end{cases}$$
$$s \in FN, d \in CL, i \in N \tag{29}$$

$$\sum_{j \in Nm[i]: i \neq j} Sa_{ij}^{sd} - \sum_{j \in Nm[i]: i \neq j} Sa_{ji}^{sd} = \begin{cases} Sa_{sd}\, if i = s \\ -Sa_{sd}\, if i = d \\ 0 \ otherwise \end{cases}$$
$$s \in FN, d \in CST, i \in N \tag{30}$$

$$\sum_{j \in Nm[i]: i \neq j} Sb_{ij}^{sd} - \sum_{j \in Nm[i]: i \neq j} Sb_{ji}^{sd} = \begin{cases} Sb_{sd}\, if i = s \\ -Sb_{sd}\, if i = d \\ 0 \ otherwise \end{cases}$$
$$s \in FN, d \in CST, i \in N \tag{31}$$

Constraints (26) – (31) ensure that the incoming and outgoing traffic are equal for all network nodes for processing, feedback, and storage tasks, respectively. However, this does not apply to the source and destination nodes. The traffic flowing through links is governed by the following constraints in our resilient architecture:

$$\sum_{s \in CL} \sum_{d \in FN} Pa_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CL} Fa_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CST} Sa_{ij}^{sd} \geq La_{ij}$$
$$i \in N, j \in Nm[i] \tag{32}$$

$$\sum_{s \in CL} \sum_{d \in FN} Pa_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CL} Fa_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CST} Sa_{ij}^{sd} \leq M La_{ij}$$
$$i \in N, j \in Nm[i] \tag{33}$$

$$\sum_{s \in CL} \sum_{d \in FN} Pb_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CL} Fb_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CST} Sb_{ij}^{sd} \geq Lb_{ij}$$
$$i \in N, j \in Nm[i] \tag{34}$$

$$\sum_{s \in CL} \sum_{d \in FN} Pb_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CL} Fb_{ij}^{sd} + \sum_{s \in FN} \sum_{d \in CST} Sb_{ij}^{sd} \leq M Lb_{ij}$$
$$i \in N, j \in Nm[i] \tag{35}$$

Constraints (32) and (33) ensure that $La_{ij} = 1$ if the incoming and/or outgoing traffic of primary PSs pass through the link between nodes i and j; otherwise, the value is zero. Meanwhile, constraints (34) and (35) ensure that $Lb_i = 1$, if the incoming and/or outgoing traffic of secondary PSs pass through the link between nodes i and j, otherwise the value is zero.

$$La_{ij} + Lb_{ij} \leq 1;\ \ i \in ND, j \in ND \tag{36}$$

$$La_{ij} + Lb_{ji} \leq 1;\ \ i \in ND, j \in ND \tag{37}$$

$$La_{ji} + Lb_{ij} \leq 1;\ \ i \in ND, j \in ND \tag{38}$$

Constraints (36) - (38) are used to ensure that the incoming traffic and/or outgoing traffic of the primary and secondary PSs traverse different links.

$$\sum_{j \in Nm[i]: i \neq j} La_{ij} \geq \rho a_i;\ \ i \in ND \tag{39}$$

$$\sum_{j \in Nm[i]: i \neq j} La_{ij} \leq M \rho a_i;\ \ i \in ND \tag{40}$$

$$\sum_{j \in Nm[i]: i \neq j} Lb_{ij} \geq \rho b_i;\ \ i \in ND \tag{41}$$

$$\sum_{j \in Nm[i]: i \neq j} Lb_{ij} \leq M \rho b_i;\ \ i \in ND \tag{42}$$

$$\rho a_i + \rho b_i \leq 1;\ \ i \in ND \tag{43}$$

Constraints (39) and (40) and constraints (41) and (42) determine the nodes that are used to relay the incoming and/or outgoing traffic of the primary PSs and secondary PSs, respectively. Meanwhile, constraint (43) ensures that the nodes used to relay the incoming and/or outgoing traffic of primary and secondary PSs are different.

IV. PARAMETER SELECTIONS

In this work, we considered patients with postoperative atrial fibrillation in the ECG monitoring IoT application. Also, we consider a setting where each patient transmits their ECG signal with a duration of 30-second to the network, as suggested in [42]. Note that this 30-second ECG signal needs high processing capabilities for processing and analysis. The following subsections describe the methodologies used to determine the model input parameters utilized in this work.

A. NETWORK LAYOUT

The patients' locations considered in this work are at the clinic where they are registered. A total of 37 clinics were available in West Leeds in 2014/2015 [43]. However, the complexity of the MILP model rises exponentially with increase in the number of nodes in the network. Therefore, a scenario with 16 clinics and 13 LTE BSs in West Leeds is considered for our case study. We choose the 13 LTE BSs with the shortest distance between the considered BSs and clinics. Note that the locations (i.e. latitude and longitude) of the clinics and BSs are the actual locations in West Leeds, as explained in [20].

TABLE 4. Number of monitored patients in clinic for ECG IoT monitoring applications.

Clinic	Number of Patients
Craven Road Medical Practice	20
Leeds Student Practice	68
Hyde Park Surgery	13
Burton Croft Surgery	15
Laurel Bank Surgery	16
Kirkstall Lane Medical Centre	11
Burley Park Medical Centre	23
Thornton Medical Centre	18
Beech Tree Medical Centre	16
Hawthorn Surgery	4
Priory View Medical Centre	20
Abbey Grange Medical Centre	9
Vesper Road Surgery	16
The Highfield Medical Centre	25
Dr G Lees & Partners	10
Whitehall Surgery	16

As shown in Figure 1, we considered two clusters as a case study where each clinic can be associated with a maximum of two nearest BSs in each cluster. For instance, clinic 10 (i.e. light yellow) is associated with one base station (BS) in cluster 1 and two BSs in cluster 2. Note that the BSs in each cluster are assigned as follows: We determine the BSs with the largest distance and set them as the central point for each cluster. Then we assigned the remaining BSs to the cluster with the lowest distance. For each cluster, we choose only one OLT provided by the BT Wholesale network [44] that has the lowest total distance to the BSs in the selected cluster.

B. TOTAL MONITORED PATIENTS FOR ECG IoT MONITORING APPLICATIONS

In this work, we consider patients that may experience postoperative atrial fibrillation (AF) in West Leeds, UK as the respondents, which relates to the total traffic considered in the network. As explained in detail in [20], 0.176% of the UK population is considered to have heart surgeries. Therefore, we used this percentage to estimate the total number of patients monitored in each clinic [43]. Table 4 shows the total estimated number of monitored patients logged in every 16 clinics that may experience postoperative AF.

C. LINK CAPACITIES

It is essential to highlight that the link capacities at all layers (access, metro and core layers) are considered to serve traffic for all other applications. As in our previous works [20], [25], [38], in this work, we consider only 0.3% of the maximum capacities at all layers to be dedicated to healthcare applications. The detailed calculation of this percentage can be found in [20], [25], and [38].

D. TIME FOR PROCESSING AND ANALYSIS

In this work, the same 30-second ECG recording signal (i.e. $\Pi = 252.8$ kbits) in [20] is utilized. Each patient sends their 30-second ECG signals to both primary and secondary

PSs located at the access layer for processing and analysis. The correlation between the time to process and analyse the ECG signal (τp) at the PSs (i.e. primary and secondary PS) using the Pan-Tompkins algorithm and the number of patients (Pat) are as obtained from the experiments we conducted in [20] using MATLAB with parallel processing which is $\tau p = 0.002Pat + 4.685$.

E. PATIENT DATA RATE

In this work, the number of patients allowed to be aided by a single PS, Pat is limited to investigate the distribution of both primary and secondary PSs in the network, with increasing demands. Therefore, from the total patients in the 16 clinics studied, we only consider 20% of them as the maximum Pat that can be aided by each server, which is the lowest demand evaluated in the network. Based on our experimental results in [20], the size of the analyzed ECG data after processing and analysis using the Pan-Tompkins algorithm is 256 bits (α). This analyzed data is sent to the cloud from the primary PSs and secondary PSs for permanent storage purposes. However, cloud storage will only store one copy, and the same concept is applied to the data sent from the primary and secondary PSs to the clinic.

The timing restrictions set by the AHA [25] are used to calculate the network and processing energy consumption. Hence, 4 minutes (i.e. $\tau t = 4minutes$) is considered as the maximum duration to save heart patients. The 4 minutes compris; (i) the duration time to record the ECG signal (i.e. 30 seconds) (τm), (ii) the transmission time to send the 30 seconds ECG signal to primary and secondary PS for processing task (τmax), (iii) the time to process and analyze the 30 seconds ECG signal (τp), and (iv) the transmission time to send analyzed ECG signal for feedback task (τb). The available time to transmit the recorded 30-second ECG signal to the primary and secondary PSs(τmax), is determined based on the processing and analysis time (τp), considering the maximum patients allowed to be aided by a single PS (Pat) as well as the transmission time to send the analyzed health data feedback traffic to the clinics (τb) while considering the 30 seconds of ECG recording (τm) from the patient for τt equal 4 minutes.

The feedback time is determined based on the maximum patients the PSs can aid at each candidate fog node, $MaxP$ and the minimum shared link capacity between the LTE BS to the candidate fog node at the access network (i.e. the link from optical network unit (ONU) to OLT), Cb_{min}. Thus, $MaxP$ is given as

$$MaxP = PatN \tag{44}$$

where N refers to the maximum PSs allowed per candidate fog node. The number of PSs at every candidate fog node is limited by the space available at the fog nodes. Note that the maximum patients that the PSs can aid at each fog node will share the minimum link capacity. Therefore, the provisioned data rate for each patient to transmit their feedback data (i.e. analyzed health data) to clinics (δf), is calculated by diving the minimum link capacity by $MaxP$ as follows:

$$\delta f = Cb_{min}/MaxP \tag{45}$$

It is worth noting that limiting the feedback data rate to the data rate offered for healthcare applications in GPON links reduces the number of active BSs, as explained in [20]. Also, in this study, we consider LTE-M BS utilizing quadrature phase-shift keying (QPSK) as the modulation technique as in [20], which gives a single physical resource block (PRB) equal to 336 bps. We used the same approach in [20] to determine i) the allocated number of PRBs per patient to transmit the analyzed feedback data, (Rb) ii) the provisioned data rate per patient to transmit their analyzed health data for feedback task (δb) and its transmission time (τb), iii) the remaining time for each patient to transmit their raw health data to the PSs (τmax), iv) the allocated number of PRBs per patient to send their raw health data (Ra), and v) the provisioned data rate per patient to transmit their raw health data for processing (δa) and its transmission time (τa). Meanwhile, the data rate to transmit the analyzed health data storage for permanent storage for each patient (i.e. from primary and secondary PS to the cloud storage) was determined by dividing the minimum shared uplink capacity or node capacity from the PS to the cloud storage that is provisioned by the IoT healthcare application (Cc_{min}) and the maximum patients allowed to be aided by the PSs at every candidate fog node ($MaxP$) as below:

$$\delta c = Cc_{min}/MaxP \tag{46}$$

The time needed to send the analyzed health data storage traffic from the PSs to the cloud storage is calculated as in [20].

F. POWER CONSUMPTION OF NETWORKING AND PROCESSING EQUIPMENT

As mentioned in Section III-A, for all networking devices and PSs, the power consumption consists of two part; idle and linear proportional. The details of the input parameters for all networking devices and PSs, including their maximum power consumption, maximum working capacity and the considered portion of the idle power attributed to our IoT healthcare application (x), are as explained in [20].

V. PERFORMANCE EVALUATION

This section discussed the results and the MILP model analysis for the three scenarios: i) Scenario 1 that consider protection for the server with no geographical constraints, ii) Scenario 2 that consider protection for the server with geographical constraints, and iii) Scenario 3 that consider protection for both the servers and network with geographical constraints and link and node disjoint design, respectively. It is worth to note that, the MILP model is solved using the AMPL software equipped with CPLEX 12.8 solver that runs on a high-performance computing cluster (HPC) with 12 core CPU and 64G RAM. The evaluation is divided into three steps. For each step, an analysis is carried out to determine

(i) the optimum locations of the PSs in each scenario, (ii) the energy consumed by the networking equipment following MILP optimization, and (iii) the network energy penalty due to the increased level of resilience. The first step compares the non-resilient scenario with a resilient scenario 1. Secondly, it compares the resilient scenario 1 with the resilience scenario 2. Thirdly, it compares the resilient scenario 2 with the resilient scenario 3. Also, note that the performance of each scenario is investigated based on the demand level (i.e. the percentage of the patients considered in the network) and the number of PSs each candidate fog node can host.

A. SCENARIO 1: PROTECTION FOR SERVERS WITHOUT GEOGRAPHICAL CONSTRAINTS

In this section, the performance of the non-resilient scenario is used as a benchmark to evaluate the resilient scenario 1 in terms of both network and processing energy consumption for ECG monitoring applications. Table 5 presents the considered input parameters for ECG monitoring applications for different numbers of PSs per candidate fog node (N) when each PS (i.e. primary and secondary) can serve 20% of the total patients from all considered clinics. Note that the given data rate and time to transmit the raw ECG signal to the PSs for each number of PSs per candidate fog node in Table 5 are identical for all percentages of patients evaluated in the network (i.e. 20%, 40%, 60%, 80% and 100% of the total patients in the 16 clinics). This is because the data rate is given based on the number of PRBs while ensuring the total data rate provided by the total number of PRBs per patient is higher than or equal to the minimum data rate required to ensure the system can work within the 4 minutes. Therefore, the same number of PRBs is given to each patient under a different number of PSs per candidate fog node, although their required minimum data rate is different.

The results in Figure 2 show that the number of PSs for the resilient scenario 1 is double that of the non-resilient scenario. This is because the non-resilient scenario only has primary PSs, while the resilient scenario 1 consists of a secondary PS for each primary PS for server protection purposes. The results show that increasing the percentage of patients served resulted in increasing the number of PSs. For the non-resilient scenario, at demand levels of 20%, 40%, 60%, 80% and 100%, the number of PSs required to serve all patients are one, two, three, four and five, respectively. Meanwhile, for the resilient scenario 1, the number of PSs for each demand level is double that of the number of PSs for the non-resilient scenario.

Figure 2 also reveals that the OLT is always selected to place the PSs since the OLT is the nearest shared location to the patients (i.e. the OLT is associated with all BSs in each cluster). Therefore, the total number of PSs needed to serve the patients and the total number of hops to send the raw ECG signal to the PSs is reduced. The results in Figure 2-(a) and Figure 2-(b) show that at a demand level equal to 60% or less, the PSs are placed at only one cluster. This is because the BSs can aid all patients in a single cluster. Therefore, the ONU is chosen to host the PSs that cannot be assigned at the OLT at the same cluster for the resilient scenario 1 to reduce the utilization of the networking equipment. In contrast, for the non-resilient scenario, the PSs are only located at the OLT.

However, when the percentage of patients increased to 80% and 100%, the BSs, ONUs and OLTs in both clusters are utilized. For the non-resilient scenario, increasing the demand level to 80% and 100% has resulted in placing the primary PSs at the OLT and ONU of different clusters. This is because the OLT does not have enough capacity to support all of the traffic. Therefore, the OLT of cluster 2 is employed first, followed by sending the remaining demands to the ONU of cluster 1 to reduce the total traffic passing in the network as patients are connected directly to the ONUs. For the resilient scenario 1, when the considered demand level is 80%, and a single candidate fog node can host three PSs, the ONUs and OLT of cluster 1 are employed first, and the ONU of cluster 2 is used to serve the remaining demand. This is due to the same reason, as explained for the non-resilient scenario. However, increasing the demand level to 100% has led to the usage of the OLTs of all clusters and only the ONU of one cluster. The model did not use multiple ONUs to host the PS to reduce the utilization of the Ethernet switches. When five PSs per candidate fog node are allowed, the PSs are placed

TABLE 5. Data rate and related time for different PSs per candidate fog node, *N*, for ECG IoT monitoring applications.

Type of Data	Number of PSs at each Candidate Fog Node					
	3	4	5	6	7	8
Data rate to transmit raw ECG signal to PS, δa (kbps)	1.344	1.344	1.344	1.344	1.344	1.344
Time to transmit raw ECG signal to PS, ta (s)	188.1	188.1	188.1	188.1	188.1	188.1
Data rate to transmit analyzed health data to clinics, δb (kbps)	1.008	0.672	0.672	0.336	0.336	0.336
Time to transmit analyzed health data to clinics, tb (s)	0.254	0.381	0.381	0.762	0.762	0.762
Data rate to transmit analyzed health data to cloud storage, δc (kbps)	1.28	0.96	0.768	0.64	0.548	0.48
Time to transmit analyzed health data to cloud storage, tc (s)	0.2	0.267	0.333	0.4	0.467	0.533

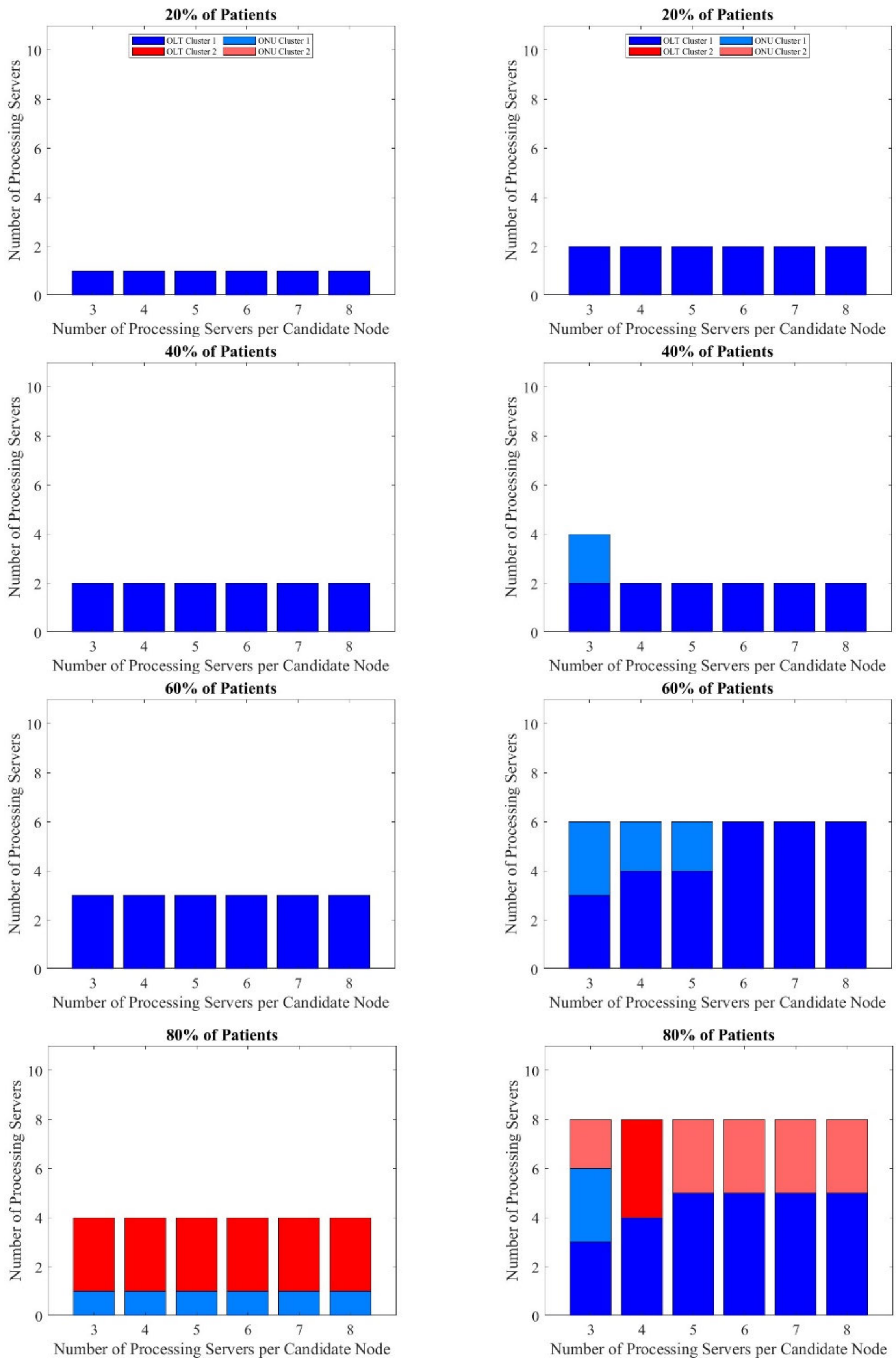
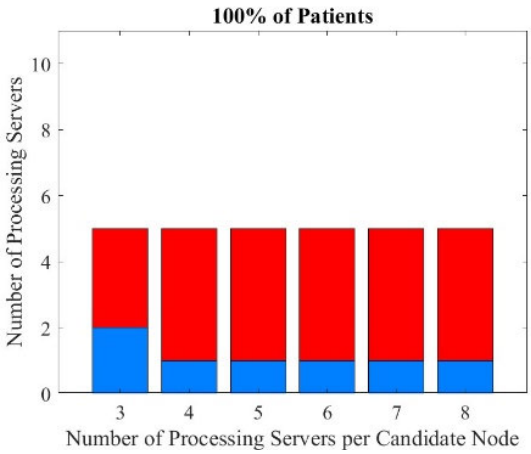
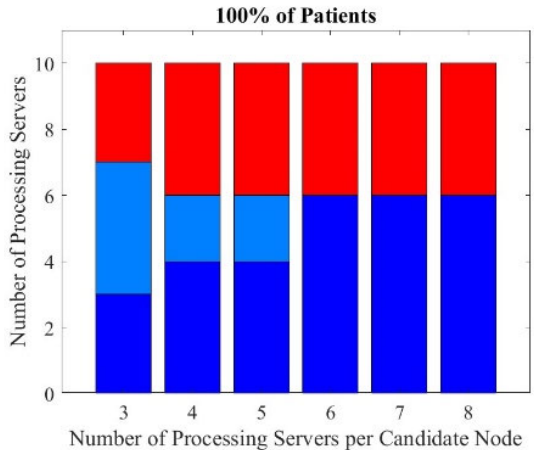


FIGURE 2. Optimal location of processing servers for (a) non-resilient scenario and (b) resilient scenario 1.



(a)



(b)

FIGURE 2. ***(Continued.)*** **Optimal location of processing servers for (a) non-resilient scenario and (b) resilient scenario 1.**

at both OLTs and one ONU to minimize the number of BSs utilized, as a BS consumes more energy than an Ethernet switch.

The results in Figure 3 show that the increase in the rate of network energy consumption due to the increase in demand for the resilient scenario 1 is higher than the non-resilient scenario. The results also show that for the resilient scenario 1, the network energy consumption is always higher than the non-resilient scenario for all levels of demand allowed to be served at each candidate fog node and all number of PSs. This is because, for the resilient scenario 1, the total traffic traversing the networking equipment is doubled compared to the non-resilient scenario, hence increasing the energy consumed by the networking devices in the resilient scenario. The increase in network energy consumption is one of the key penalties as a result of having resilience.

Figure 3 also shows that at a demand level equal to or more than 40%, the network energy consumption of the resilient scenario 1 reduced significantly when the total PSs allowed at each candidate fog node increased from three to eight. This is because allowing more PSs to be hosted at each candidate fog node has resulted in placing the PSs at their optimal locations in addition to reducing the number of utilized nodes.

The results in Figure 4 show that the energy penalty (defined as the difference in energy consumption between the resilient and the non-resilient cases) increases when the level of demand rises from 20% to 80%. This is because, at demand levels of 20% to 80%, the total number of BSs utilized to aid all patients to transmit their raw ECG signal to the PSs for the non-resilient scenario is the same. In contrast, for the resilient scenario 1, the total number of utilized BSs increases with increase in demand, as shown in Figure 6. The increase in the number of utilized BSs under the resilient scenario 1 is because each patient sends two ECG signals to both primary and secondary PSs, requiring many BSs to aid all patients and this number increases as the demand increases.

For the non-resilient scenario, each patient only sends one raw ECG signal to the primary PSs, and the same number of BSs are used, as they can accommodate the increasing demand by up to 80%. Nevertheless, at a demand level of 100%, the energy penalty is lower than 80%. This is because when the demand level is 100%, the number of BSs utilized for the non-resilient scenario increases, hence increasing the network energy consumed in the non-resilient scenario. Figure 4 also shows that increasing the total number of PSs per candidate fog node can significantly reduce the energy penalty when the demand is equal to or higher than 40%. This is due to the reduced number of fog nodes used to host the PSs for the resilient scenario 1, as shown in Figure 5, where more PSs are hosted at the same fog node when the number of PSs allowed at each candidate fog node increases.

The results in Figure 7 reveal that the processing energy consumption of the resilient scenario 1 is higher than the non-resilient scenario. This is because the number of utilized PSs for the resilient scenario 1 is double that of the non-resilient scenario. The results also show that the processing energy consumption increases as the demands increase for both scenarios. This is because increasing the total patients in the network increases the number of PSs proportionally. However, the same number of PSs is used in both scenarios under constraints on the number of PSs per candidate fog node, as patients are optimally consolidated in the servers. Also, there is a slight increase in energy consumption for both scenarios when more PSs are allowed to be hosted at every candidate fog node. The increase in energy is due to the increase in utilization time of the PSs to transmit both the feedback and storage traffic, with the growing number of PSs per candidate fog node, as shown in Table 5.

B. SCENARIO 2: PROTECTION FOR SERVERS WITH GEOGRAPHICAL CONSTRAINTS

In this section, the performance of the resilient scenario 1 is used as a benchmark to analyze the energy implications of the increased level of resilience gained by considering the geographical constraints, i.e. scenario 2. The energy

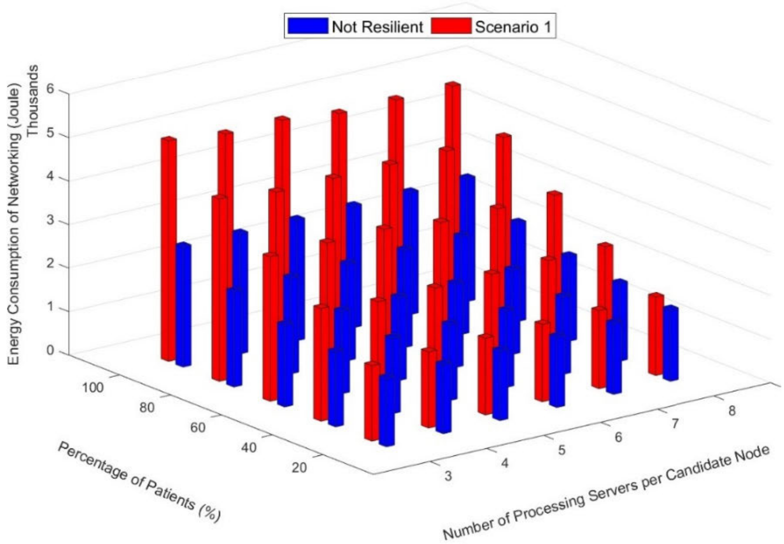


FIGURE 3. Energy consumption of networking equipment for non-resilient scenario and resilient scenario 1.

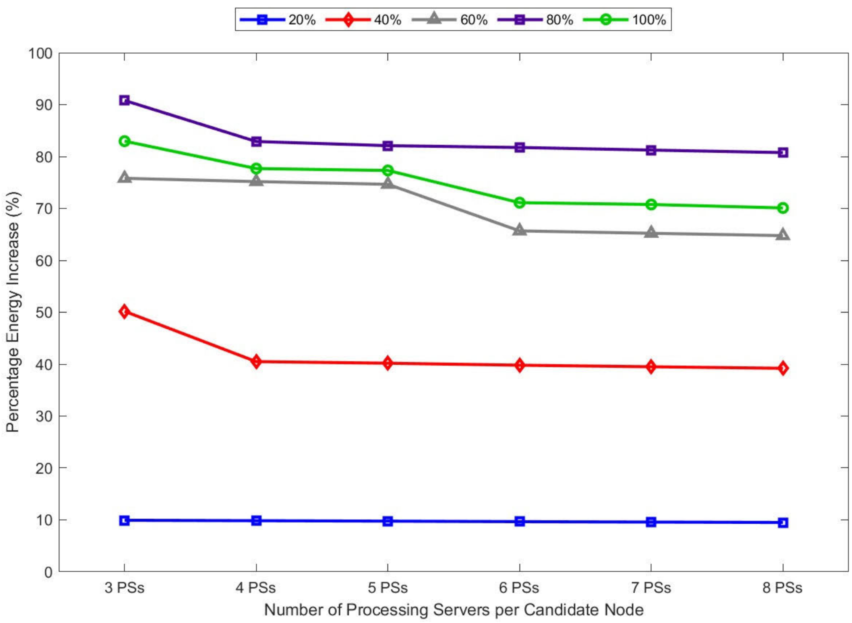


FIGURE 4. Percentage energy penalty of networking equipment for resilient scenario 1 compared to non-resilient scenario.

evaluation is based on both networking and processing energy consumption.

The results in Figure 8 reveal that the OLT is always used to host the PSs as in the previous scenarios. The results also indicate that only one cluster is used to place the PSs when the percentage of patients is equal to or less than 60%. This is to reduce the utilization of networking equipment. However, due to geographical constraints, at least two locations are required to place the primary and secondary PSs. Therefore, both OLT and ONU of the same clusters are selected to place the PSs, separately.

Figure 8 also shows that at a high demand level (i.e. 80% and 100%), the BSs, ONUs and OLTs from both clusters are utilized. The results show that at a demand level of 80% for all PSs per candidate fog node, the ONUs and OLT of cluster 1 are employed first, and due to the limited number of resources of the BSs in cluster 1 to serve the patients, the ONU of cluster 2 is used to serve the remaining demand. This reduces the amount of traffic traversing the network as ONUs are connected directly to the patients. The results also show that, at the demand level of 80%, when the total PSs allowed at every candidate fog node increases to four, the number of

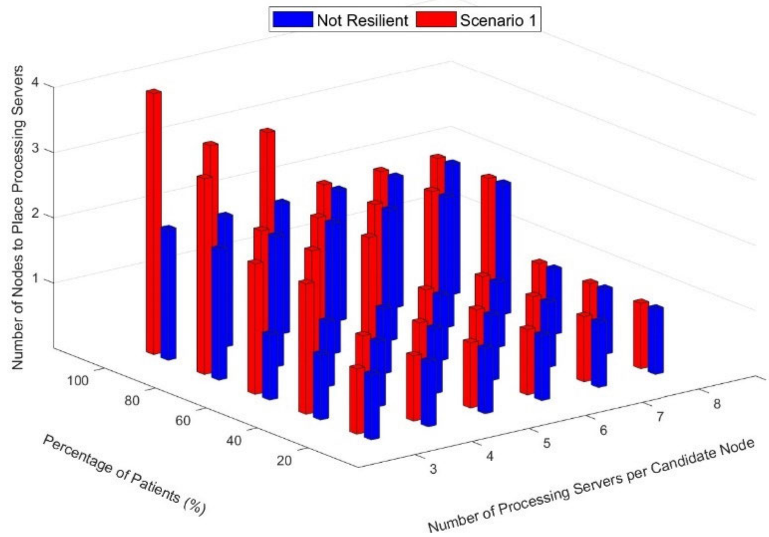


FIGURE 5. Number of nodes used to host processing servers for the non-resilient scenario and the resilient scenario 1.

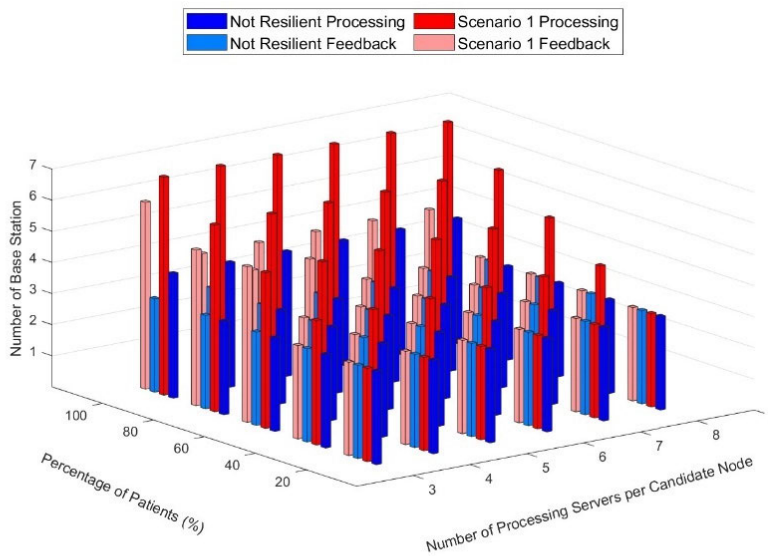


FIGURE 6. Number of base stations used to send the raw ECG signal for processing and the analyzed ECG signal for feedback, for non-resilient scenario and resilient scenario 1 under different percentages of patients and number of processing servers per candidate fog node.

fog nodes used to host the PSs is reduced as more PSs are hosted at the OLT.

However, increasing the demand level to 100% resulted in utilizing the OLT and the ONU of both clusters to accommodate the growing number of PSs in the network for all considered PSs per candidate fog node. The results indicate that allowing more PSs to be hosted at every candidate fog node does not affect the location to place the PSs, as optimal locations are selected.

The results in Figure 9 show that at demand levels of 40%, 60%, and 100%, and when the number of available PSs is three, the networking energy consumption for both scenarios is the same. This is due to the fact that the same amount of networking equipment is utilized, where the same number and location of candidate fog nodes are utilized to host the PSs, and the same number of BSs are utilized to aid the patients to transmit their raw ECG signal to the PSs, for both scenarios, as shown in Figure 10 and Figure 11, respectively.

However, at a demand level of 60% and when four and five PSs are allowed at every candidate fog node, the network energy consumption for the resilient scenario 2 is slightly higher than the resilient scenario 1, although the same number

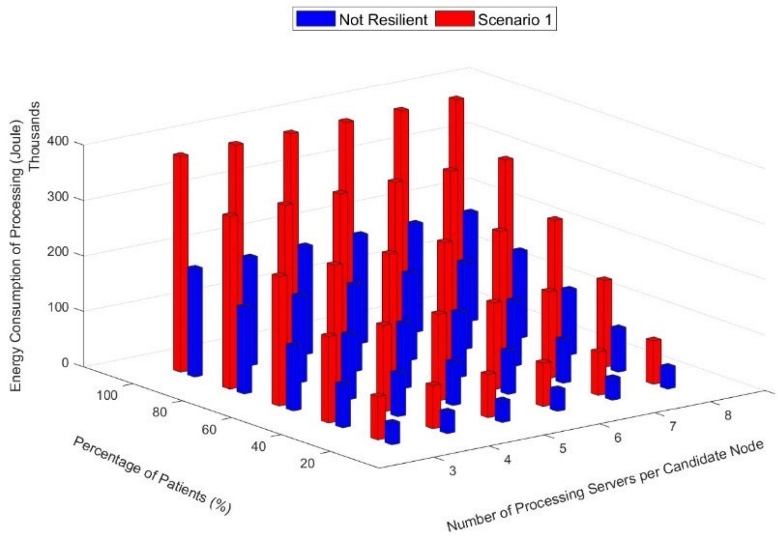


FIGURE 7. Energy consumption of processing for non-resilient scenario and resilient scenario 1.

of BSs and fog nodes are utilized to host the servers for both scenarios. This is due to the different placement of the PSs in the network for both scenarios. For the resilient scenario 2, the location of PSs has resulted in more data traversing the networking equipment compared to the resilient scenario 1.

Meanwhile, for the other demand levels and for the different numbers of PSs per candidate fog node, the energy consumed with the resilient scenario 2 is higher than the resilient scenario 1, as shown in Figure 9. This is because, considering the geographical constraint increases the total fog nodes needed to host the PSs, as shown in Figure 10. Hence, the utilization of the networking equipment under the resilient scenario 2 increases. This increase in network energy consumption is the penalty for increasing the resilience level.

Figure 12 shows no energy penalty incurred when the resilience level is increased to consider geographical constraints at demand levels of 40%, 60% and 100%; when each candidate fog node can serve three PSs. This is due to the same number of utilized networking equipment in both scenarios (i.e. fog nodes to host the PSs and BSs to send the processing traffic). However, at demand levels of 20% and 80%, increasing the resilience level to consider geographical constraints in resilient scenario 2 has resulted in an energy penalty. This is because, at these specific demands, more fog nodes are used to host the PSs for the resilient scenario 2 than the resilient scenario 1, as shown in Figure 10.

Figure 12 also reveals that the network energy penalty at a demand level equal to or more than 20% increases when the total number of PSs allowed at every candidate fog node increased. The increase in energy penalty is due to the decreasing number of fog nodes utilized to host the PSs with a resilient scenario 1, as shown in Figure 10. However, at demand levels of 20% and 80%, increasing the total number of PSs per candidate fog node does not significantly impact the energy penalty. This is due to the same amount of fog nodes used to host the PSs, and the same number of BSs used to send the raw ECG signal to the PSs in both scenarios at this specific demand, as shown in Figure 10 and Figure 11, respectively.

Figure 12 also reveals that when the total number PSs each candidate fog node can host is equal to or higher than six, the energy penalty decreases as the demand increases from 20% to 80%. This is because the same number of BSs are used in both scenarios to transmit the raw ECG signal to the PSs, as shown in Figure 11. However, the energy penalty at a demand level of 100% is higher than 40%, as the utilization of the fog nodes to host the PSs in the resilient scenario 2 is doubled compared to the resilient scenario 1, as illustrated in Figure 10.

The results in Figure 13 reveal that the processing energy consumed increases with increase in the level of demand for all number of PSs allowed at every candidate fog node for both scenarios. This is because increasing the demand level will proportionally increase the number of PSs. For all PSs per candidate fog node, equal energy is consumed for both resilience levels. This is because the same amount of PSs will be used irrespective of their location since the patients are optimally consolidated in the servers. Also, there is a slight increase in the processing energy consumption when the total allowable PSs per candidate fog node increases. This is because the time utilized by the PSs to transmit the feedback and storage traffic increases with the growing number of PSs allowed at every candidate fog node, as explained previously.

C. SCENARIO 3: PROTECTION FOR SERVERS CONSIDERING GEOGRAPHICAL CONSTRAINTS AND NETWORKS WITH LINK AND NODE DISJOINT DESIGNS

In this section, the performance of the resilient scenario 2 is used as a benchmark to evaluate the increased level of resilience (in disjoint link and node resilience), i.e. scenario 3, in terms of both networking and processing energy consumption for ECG monitoring applications.

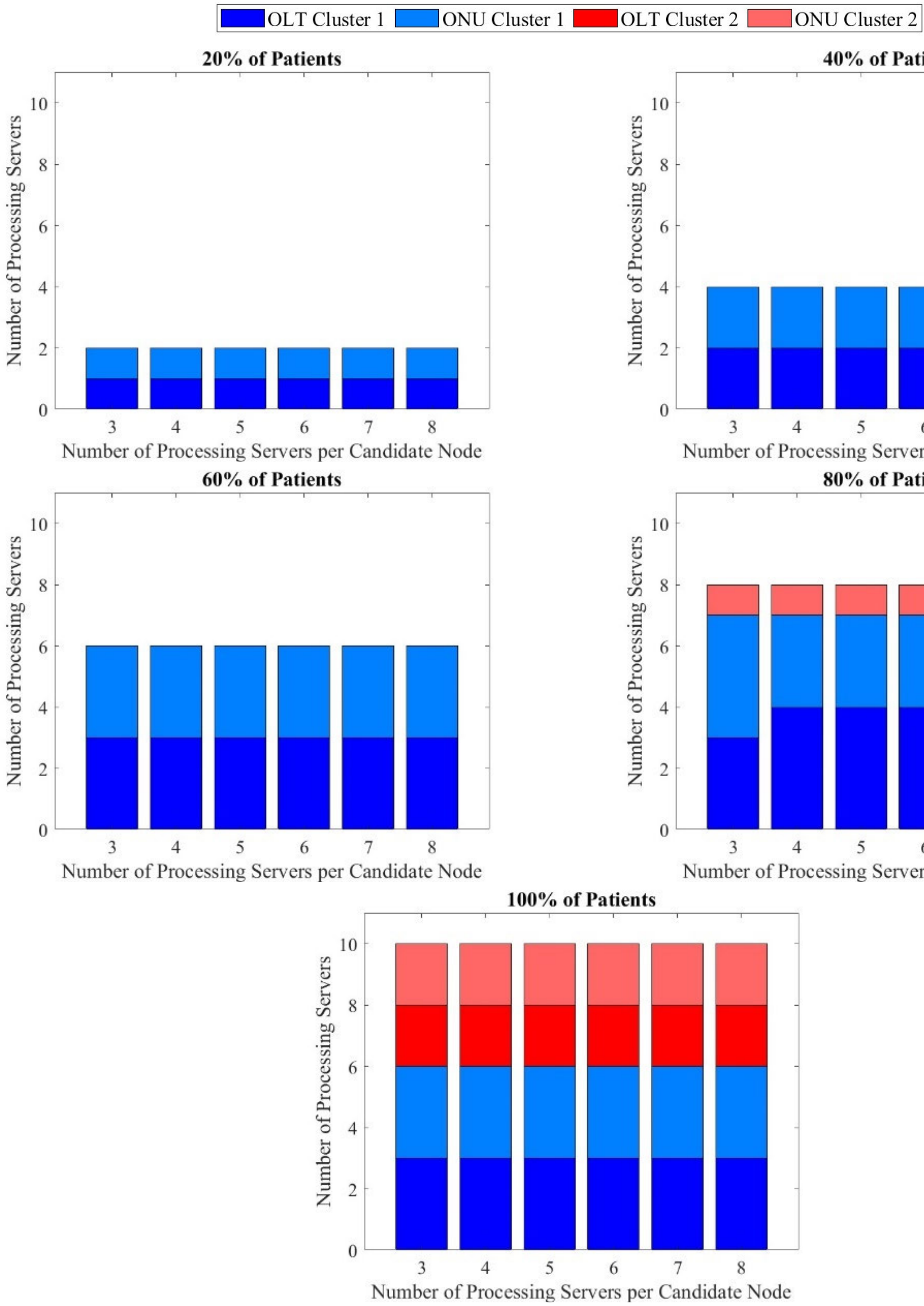


FIGURE 8. Optimal location of processing servers for resilient scenario 2.

The results in Figure 14 show that the PSs are only placed at the (optical line terminals) OLTs in both clusters when the number of PSs allowed at every candidate fog node is higher than or equal to the total number of primary or secondary PSs required in the network. This is due to two reasons. The first is to reduce the number of fog nodes (i.e. Ethernet switches) used to host the PSs, as the OLTs are the nearest shared location to all patients. The second is because each cluster is used to host the same set of PSs. For instance, cluster 1 is used to host only primary PSs, while cluster 2 is used to host only secondary PSs. Therefore, when the total PSs allowed at every candidate fog node is less than the number of primary and

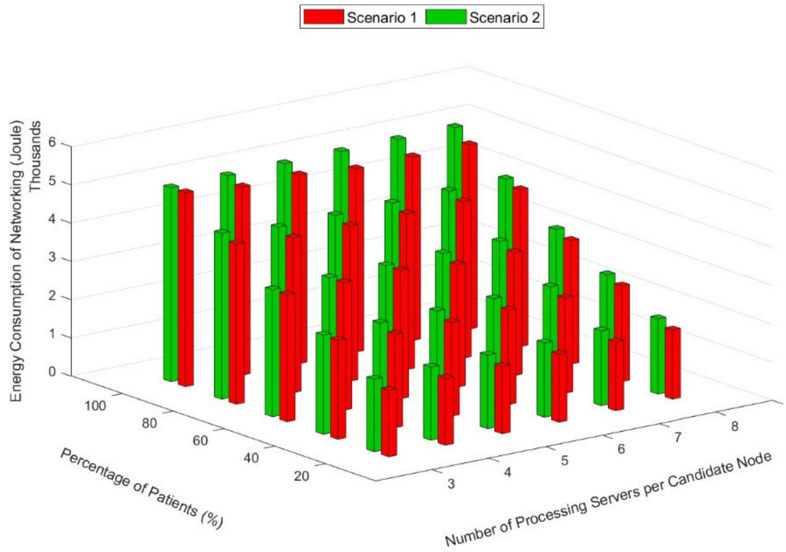


FIGURE 9. Energy consumption of networking equipment for resilient scenario 1 and resilient scenario 2.

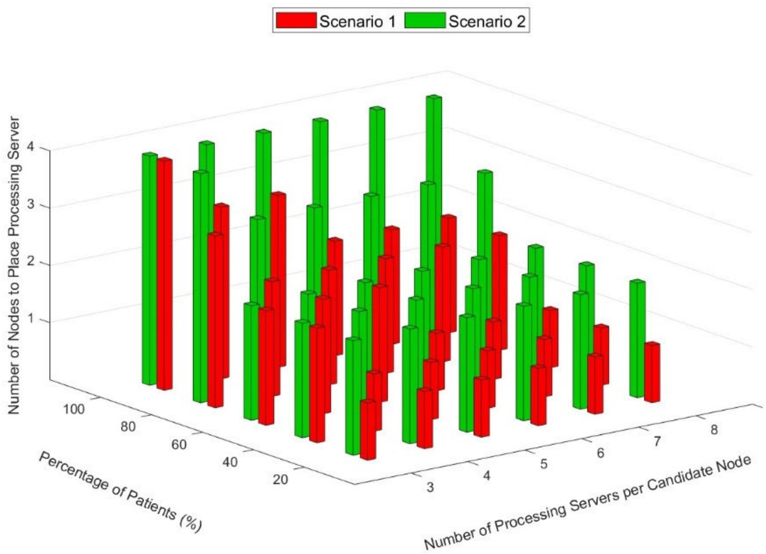


FIGURE 10. Number of nodes used to host processing servers for resilient scenario 1 and resilient scenario 2.

secondary PSs required, the (optical network units) ONUs in both clusters are used to host the remaining PSs under increasing demands.

The results in Figure 15 reveals that the energy consumed by the networking equipment for both scenarios increases as the demand level increases regardless of the total numbers of PSs allowed at every candidate fog node. This is because of the increasing traffic in the network, which increases the total utilization of the networking devices in the network. The results also show that for all levels of demands and for different numbers of PSs allowed at each candidate fog node, the energy consumed by the network for the resilient scenario 3 is always higher than the resilient scenario 2 that only considers geographical constraints. This is due to the high number of BSs utilized in the resilient scenario 3, as shown in Figure 16. Note that each base station is connected to only one OLT in the network. Therefore, considering disjoint links and nodes for network protection has increased the number of BSs without maximizing the utilization of their resources to send the processing traffic to both primary and secondary PSs.

It is worth noting that the number of fog nodes utilized to host the PSs at demand levels of 80% and 100% in the resilient scenario 3 is lower than the resilient scenario 2 when the number of PSs at every candidate fog node is equal to or more than four and five, respectively, as shown in Figure 17.

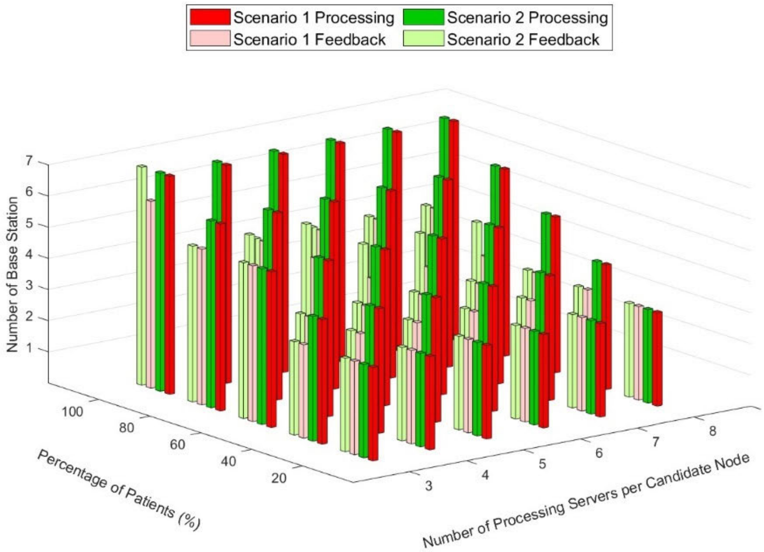


FIGURE 11. Number of base stations used to send the raw ECG signal for processing and the analyzed ECG signal for feedback, for the resilient scenario 1 and resilient scenario 2 under different percentages of patients and number of processing servers per candidate fog node.

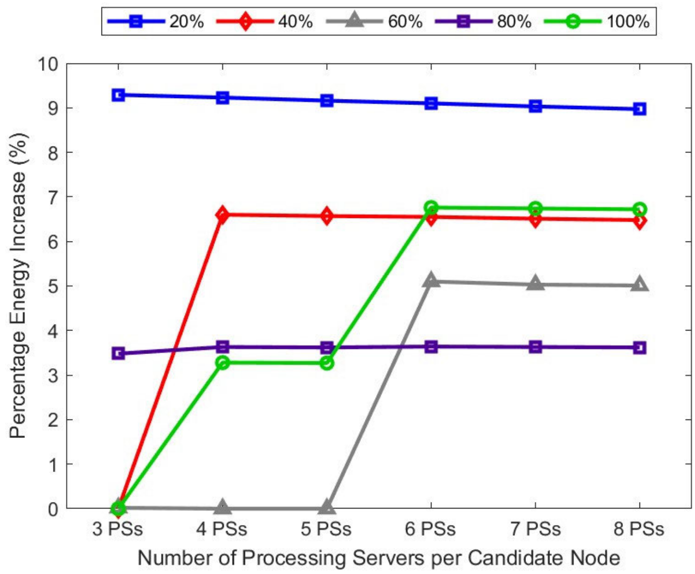


FIGURE 12. Percentage energy penalty of networking equipment for the resilient scenario 2, compared to the resilient scenario 1.

However, as the energy consumed by a single BS is approximately 1.5x higher than the energy consumed by a single node (i.e. Ethernet switch) to place the PSs, therefore there is an energy penalty with resilience scenario 3.

The results in Figure 18 reveal that the energy penalty with the resilience scenario 3 decreases as the demand level increases from 20% to 60% and 80% to 100%. This is because the total number of BSs utilized in the resilient scenario 2, which only consider the geographical constraint, increases with the increasing demand level considered in the network, as illustrates in Figure 16. This increases the network energy consumption for the resilient scenario 2 as the demand levels increase. However, at a demand level of 60%, the energy penalty is lower than at a demand level of 80%. This is because when the demand level is 80%, the number of BSs utilized for the resilient scenario 3 starts to increase and therefore this increases the energy consumption of the networking equipment of the resilient scenario 3.

Figure 18 also shows that, at demand levels of 80% and 100%, increasing the number of PSs at every candidate fog node to 4 and 5, respectively, decreases the energy penalty. This is because the number of fog nodes (i.e. Ethernet

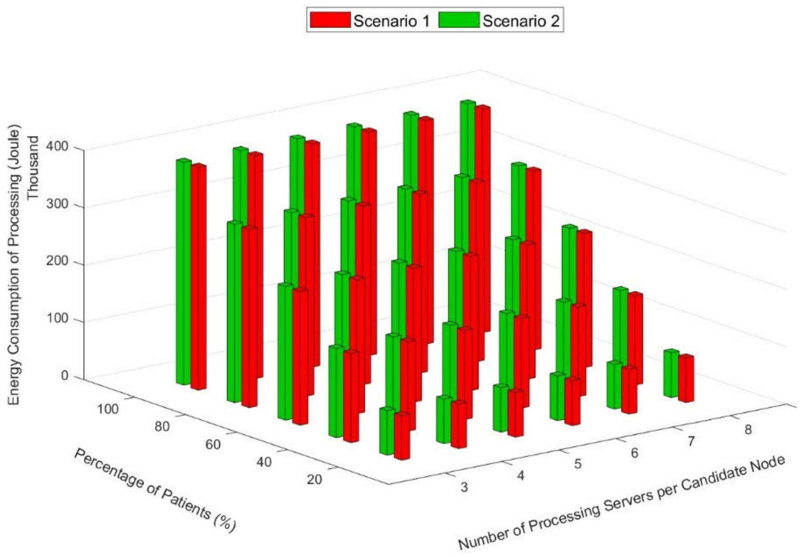


FIGURE 13. Energy consumption of processing for resilient scenario 1 and resilient scenario 2.

switches) utilized to place the PSs for the resilient scenario 3 reduces while the same number of nodes are used for the resilient scenario 2 that only considers geographical constraints, as shown in Figure 17.

The results in Figure 19 reveal that the energy consumed by the processing for both scenarios increases with increase in the demand level. This is because of the increased number of utilized PSs in the network, as explained previously. Figure 19 also shows that increasing the resilience level does not increase the processing energy consumption. Also, the energy consumption of processing has slightly increased when a single candidate fog node can host more PSs. The same energy usage for processing in both scenarios, and the increase in energy for processing in both scenarios, with the increased number of PSs at each candidate fog node, is due to the same reasons as explained previously.

VI. HEURISTIC MODELS

The complexity of MILP models and therefore the running time grows exponentially with the size of the problem (number of nodes in the network). This makes MILP models unsuitable for providing real time solutions. Therefore, we develop heuristic algorithms that can provide solutions in real time. This section presents the heuristic algorithms developed using MATLAB for the three resilient scenarios. Note that the heuristics are independent of the MILP and hence act to validate the MILP models and their results. The flow charts and detailed explanation of each heuristic are also provided in this section.

A. ENERGY OPTIMIZED RESILIENT INFRASTRUCTURE FOG COMPUTING WITHOUT GEOGRAPHICAL CONSTRAINTS (EORIWG) HEURISTIC

The EORIWG heuristic is designed to determine three main items as follow; i) the BSs to aid patients in transmitting raw health data, ii) the BSs to aid patients in receiving feedback analyzed health data, and iii) the fog nodes to host the primary and secondary PSs at the access network; with the aim of minimizing the total energy consumed by the networking and processing equipment. The details pseudo code of the developed EORIGW heuristic is shown in Figure 20.

The inputs of the EORIWG heuristic are the number of patients, the number of PSs and the number of BSs while the outputs are the BS to transmit raw health data, the BSs to receive feedback data, the number of PS used, the location of PS used and total energy consumption. The heuristic starts by grouping the clinics while considering the number of BSs in cluster 1 the clinics can connect to and sorts the groups in ascending order. For each group, the clinics are sorted according to the total number of BSs in both clusters the clinic can connect to, in ascending order. The heuristic allocates first the clinic with the least number of connections to the BSs in cluster 1 and the least number of connections to all BSs in both clusters, to help in reducing the utilization of OLTs. Also, it ensures that all clinics are assigned to BSs.

The steps to assign the clinic patients to a BS are as follows: First, the heuristic sorts the BSs attached to the clinic under consideration, beginning with BSs with available resources that are formerly utilized by the healthcare application. These BSs are then arranged in ascending order according to the number of clinics the BS can aid, followed by the unutilized BSs in cluster 1 in descending order and followed by the unutilized BSs in cluster 2 also in descending order. Sorting the activated BSs in ascending order is used to minimize the total utilization of the BSs. Meanwhile, sorting the unutilized BSs in descending order in cluster 1 followed by the unutilized BSs in cluster 2 ensures that choices are left open until late in the allocation process while minimizing the number of utilized OLTs. Next, the patients under the considered clinic are merged into the minimum number of BSs to minimize the

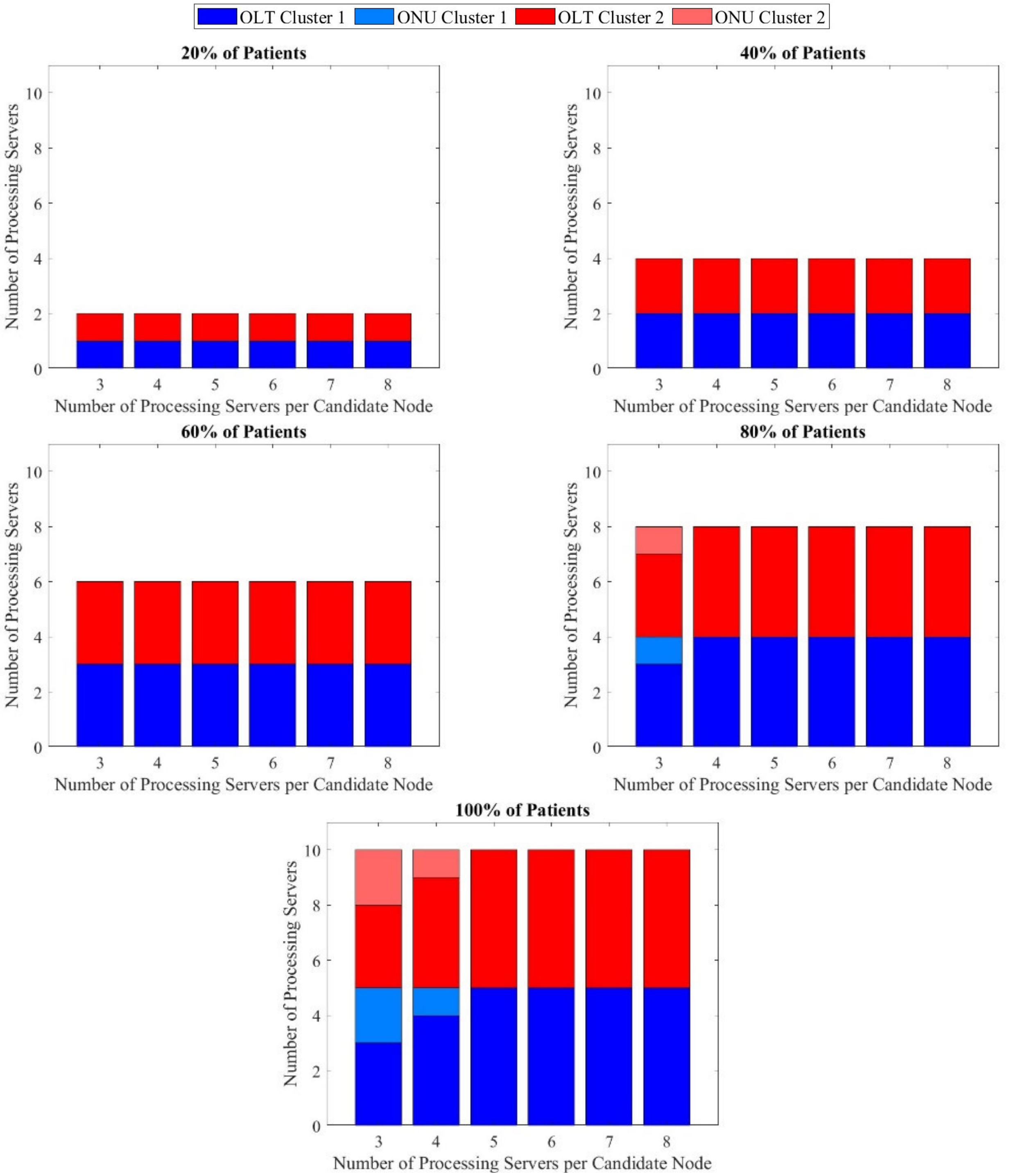


FIGURE 14. Optimal location of processing servers for the resilient scenario 3.

total number of BSs utilized for the healthcare application. Note that, for each patient, the heuristic assigned double resources to clinics to send their health data to both primary and secondary PSs.

The heuristic then calculates the total number of primary PSs and secondary PSs needed to aid all patients. Also, it determines the total number of fog nodes required to host the PSs. Note that, in this heuristic, the OLTs and the ONUs co-located at the BSs selected to aid the patients are used as the candidate fog nodes that can host the PSs. Next, the heuristic determines the combination of candidate fog nodes to host the primary and secondary PSs with minimum energy

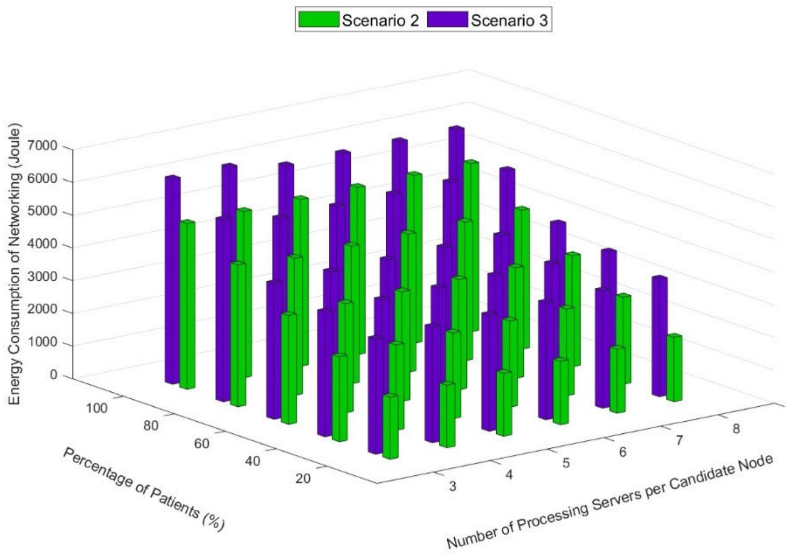


FIGURE 15. Energy consumption of networking equipment for the resilient scenario 2 and for the resilient scenario 3.

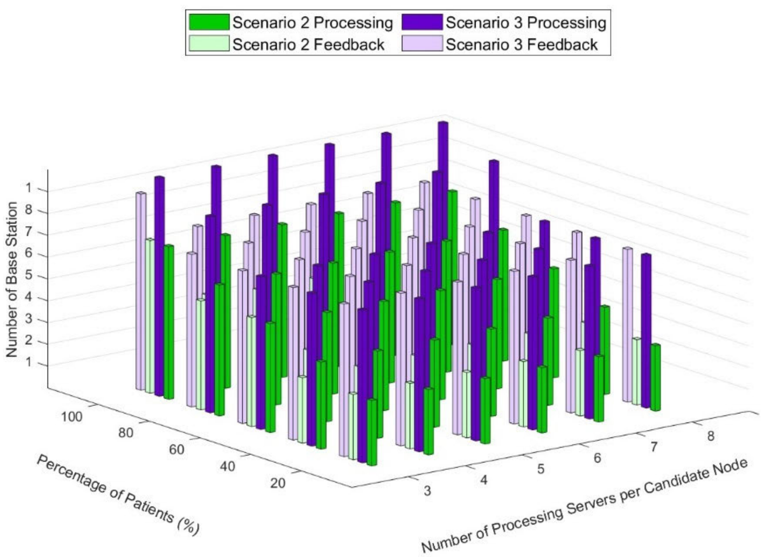


FIGURE 16. Number of base stations used to transmit the raw ECG data for processing and the analyzed ECG data for feedback, for the resilient scenario 2; and for the resilient scenario 3 under different percentages of patients and number of processing servers per candidate fog node.

consumption considering the minimum number of fog nodes needed to aid all patients (i.e. based on the total PSs allowed at every candidate fog node). Limiting the number of fog nodes to host the primary PSs and secondary PSs reduces the number of utilized Ethernet switches to host the PSs.

The energy consumption resulting from placing the primary and secondary PSs at the selected combination of candidate fog nodes is determined by transmitting the traffic of the raw health data from BSs (i.e. starting with the BS serving the highest number of patients) to the nearest fog node with available processing capacity; for the combination of the considered candidate fog nodes with minimum hop routing.

Next, the heuristic chooses the BSs to transmit the feedback traffic (i.e. analyzed health data traffic) from the considered combination of fog nodes to the clinic, using the same methods employed to choose the BSs to transmit the raw health data traffic. However, note that different BSs may be used to transmit the raw and feedback health data traffic. The difference is because the raw health data size is bigger than the analyzed data for feedback.

The combination of fog nodes to be used to host the primary and secondary PSs to aid all patients (based on the minimum number of fog nodes) with minimum energy consumption is chosen. Then, the heuristic increases the number of candidate fog nodes to host both primary and

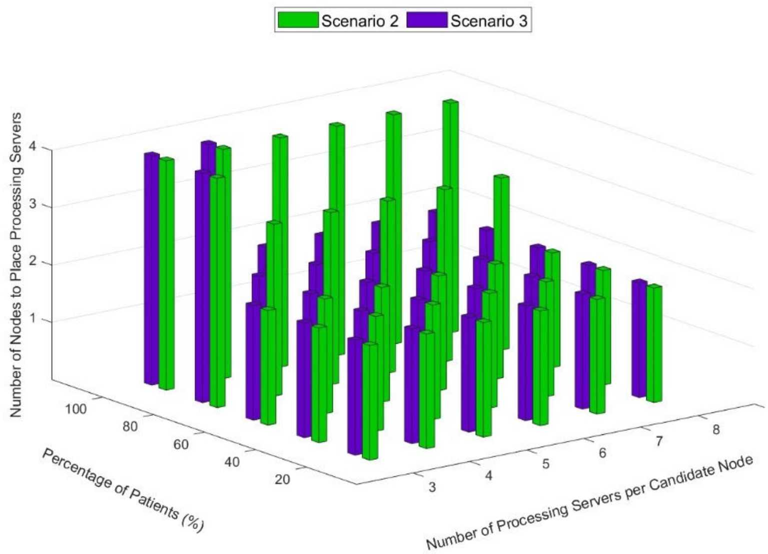


FIGURE 17. Number of nodes used to host the processing servers for the resilient scenario 2; and for the resilient scenario 3.

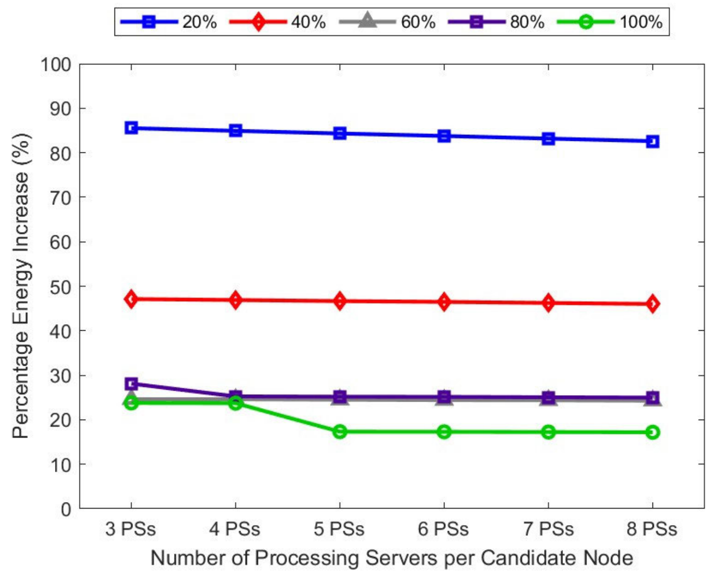


FIGURE 18. Percentage energy penalty of networking equipment for the resilient scenario 3 compared to the resilient scenario 2.

secondary PSs and performs the same process above. Suppose low energy is consumed with this combination of fog nodes compared to the energy consumed considering the combination of fog nodes with minimum fog node to host PSs. In that case, the heuristic attempts placing PSs in more candidate fog nodes. Else, the combination of fog nodes with the minimum number of fog nodes needed to host PSs is chosen to place PSs.

B. ENERGY OPTIMIZED RESILIENT INFRASTRUCTURE FOG COMPUTING WITH GEOGRAPHICAL CONSTRAINTS (EORIG) HEURISTIC

As in the EORIWG heuristic, the EORIG heuristic is designed to determine the three main items listed above to minimize the networking and processing energy while ensuring the primary and secondary PSs are node disjoint (geographical constraints). Below is the list of the changes made for the EORIG heuristic compared to the EORIGW heuristic:

1. The number of fog nodes needed to host PSs is determined considering the maximum number of fog nodes where the primary and secondary PSs are placed in disjoint nodes.
2. Assigning patients from BSs to the primary PSs is done first, and the fog nodes used to place the primary PSs are removed from the combination of fog nodes before assigning the same patients from the BSs to the secondary PSs.

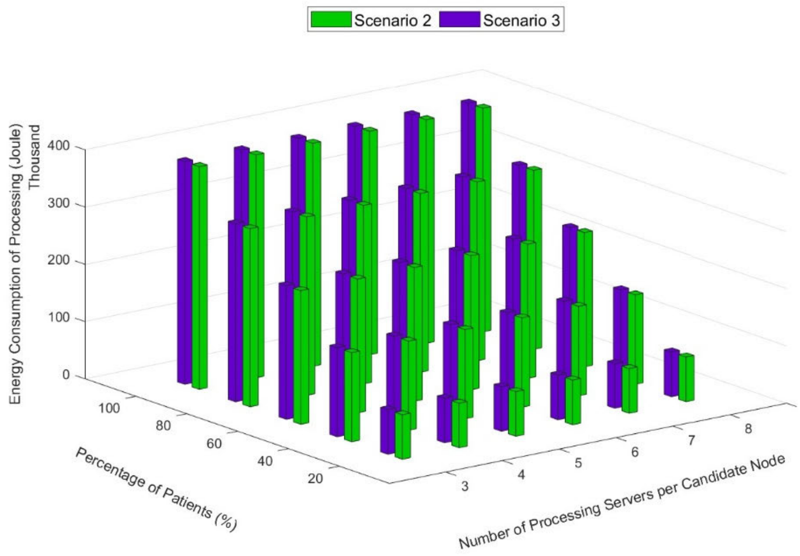


FIGURE 19. Energy consumption of processing for the resilient scenario 2; and the energy consumption of the resilient scenario 3.

C. ENERGY OPTIMIZED RESILIENT INFRASTRUCTURE FOG COMPUTING WITH GEOGRAPHICAL CONSTRAINTS AND LINK AND NOD DISJOINT DESIGN (EORIGN) HEURISTIC

As in the previous heuristics, the EORIGN heuristic is designed to determine the three main items mentioned above to minimize both the networking and processing energy while ensuring server protection with geographical constraints and network protection with link and nodes disjoints. The details pseudo code of the EORIGN heuristic is presented in Figure 21 where the inputs of the heuristic are the number of patients, the number of PSs and the number of BSs while the outputs are the BS to transmit raw health data, the BSs to receive feedback data, the number of PS used, the location of PS used and total energy consumption.

In the EORIGN heuristic, the selection of the locations to host the primary and secondary PSs are made separately to ensure that the traffic to the primary and secondary PS are routed separately (link and node disjoint). In this process, the heuristic begins by selecting a cluster to assign the patients in the clinics to the primary PS. The heuristic groups the clinics considering the number of BSs in the chosen cluster that the clinic can connect to and sorts the groups in ascending order. Then for each group, the clinics are arranged in ascending order according to the number of patients the clinic serves. The heuristic allocates first the clinic with the least number of connections to the BSs in the selected cluster and the least number of patients it serves, to the BSs to ensure each clinic can be served by at least one BS and help in packing the BSs. Note that packing is optimum for devices with high idle power consumption.

The assignment of the clinic patients to a BS are is carried out as follows: The heuristic sorts the BSs in the chosen cluster that has a connection to the considered clinic, beginning with the BSs with available resources and are formerly utilized by the healthcare application. These BSs are arranged in ascending order based on the total number of clinics the BS can aid, followed by the unutilized BSs in the chosen cluster in descending order. The reason for sorting the activated BSs and the unutilized BS in the selected cluster in ascending and descending order, respectively, are as explained for the EORIGW heuristic. Then, the considered clinic patients are merged into the minimum number of BSs so that the total number of utilized BSs for the healthcare application is reduced.

Next, the heuristic determines the number of primary PSs needed to aid all patients, as well as the number of fog nodes to host them. As in the previous heuristics, the candidate fog nodes to be used to host the PSs are the OLT of the chosen cluster and the ONUs associated with the BSs chosen to aid the patients. Then, the heuristic selects the combination of candidate fog nodes to host the primary PSs with minimum energy consumption considering the minimum number of fog nodes needed to host the primary PSs to aid all patients (i.e. based on the total PSs allowed per candidate fog node). Limiting the number of fog nodes utilized to host the primary PSs is due to the same reason as explained for EORIWG heuristic. The energy consumption due to hosting PSs at a combination of candidate fog nodes in the selected cluster is calculated as explained for the EORIWG heuristic.

Next, the heuristic chooses the BSs to transmit the feedback traffic (i.e. analyzed health data traffic) from the considered combination of fog nodes to the clinic using the same methods to choose the BSs to transmit the raw health data traffic. Note that different BSs may be utilized to transmit raw and feedback health data traffic due to the same reason as explained for the EORIWG heuristic. The combination of fog nodes needed to host primary PSs to aid all patients that result in minimum energy consumption is chosen. As in

Input: Number of patients, number of PSs, number of BSs.
Output: Number of BSs to transmit raw health data, number of BSs to receive feedback data, number of PSs used, location of PSs used, total energy consumption.
Algorithm :
Group clinics based on the number of connections to BS in cluster 1 and sorted in ascending order.
Sort the clinic in each group based on the number of connections to all BSs in ascending order.

If All the clinics are served.

Select a clinic with the smallest number of connections to BSs in cluster 1 and the smallest number of connections to all BSs.

If All patients in the selected clinic are served.

Sort used BSs that have a connection to the selected clinic based on the total number of clinics each BS can serve in ascending order followed by unused BSs in cluster 1 in descending order and the unused BSs in cluster 2 in descending order (List A).
Select the first BS in List A to assign patients.
Update the available resource of the BSs.

Endif
Endif

Determine the minimum number of candidate nodes to place both primary and secondary processing servers (n).
Calculate the energy consumption.

If n nodes result in lower energy consumption

Increase the number of nodes required to host the primary and secondary processing servers (n=n+1).
Calculate the energy consumption.

Endif
Select n-1 nodes to place the servers.

FIGURE 20. **Pseudo code of EORIWG heuristic.**

Input: Number of patients, number of PSs, number of BSs
Output: Number of BSs to transmit raw health data, number of BSs to receive feedback data, number of PSs used, location of PSs used, total energy consumption.
Algorithm :
Select a cluster to place the primary processing servers (PS).
Group clinics based on the number of connections to BS in the selected cluster and sorted in ascending order.
Sort the clinics in each group based on the number of patients it serves in ascending order.

If All the clinics are served.

Select the clinic with the smallest number of connections to BSs in the selected cluster and the smallest number of patients served.

If All patients in the selected clinic are served.

Sort used BS in the selected cluster that has a connection to the selected clinic based on the total number of clinics, it can serve in ascending order followed by unused BSs in that cluster in descending order (List A).
Select the first BS in List A to assign patients.
Update the available resource of the BS.

Endif
Endif

Determine the minimum number of candidate nodes to place primary PS (n).
Calculate the energy consumption.

If n nodes result in lower energy consumption.

Increase the number of nodes required to host the primary processing servers (n=n+1).
Calculate the energy consumption resulting from this placement.

Endif

Select n-1 nodes to place the servers.
Remove the used links and nodes and select another cluster to place secondary PS with minimum energy consumption.

FIGURE 21. **Pseudo code of EORIGN heuristic.**

the EORIWG heuristic, the heuristic increases the number of candidate fog nodes utilized to host PSs. The selection of candidate fog node to host PSs is as explained in EORIWG heuristic.

Next, the EORIGN heuristic removes the links and nodes used to send the traffic to or from primary processing servers and selects another cluster to assign patients in the clinics to the secondary processing servers. Different clusters are used to host the primary PSs and secondary PSs, which is due to the link and node disjoint resilience mandated for network protection. The same process is used to allocate patients to the BSs to transmit raw health data and receive analyzed feedback health data. It is also used to select locations to host the secondary PSs and determine the optimal place to host the secondary processing server.

VII. RESULTS AND ANALYSIS OF THE HEURISTIC MODELS

This section evaluates the performance of the developed heuristics for server protection, the EORIWG heuristic and EORIG heuristics, and the heuristic for server and network protection, the EORIGN heuristic. We compare the heuristic results to the MILP results in terms of networking and processing energy consumption. Note that the heuristics are simple and run fast (few seconds) on a PC with a 3.2 GHz CPU and 16 GB RAM. This is in contrast to the MILP which took few hours in each case to run in the high-performance computing cluster (HPC) with 12 core CPU and 64G RAM, mentioned in section V.

A. ENERGY OPTIMIZED RESILIENT INFRASTRUCTURE FOG COMPUTING WITHOUT GEOGRAPHICAL CONSTRAINTS (EORIWG) HEURISTIC

Figure 22 reveals that the total energy consumption in the EORIWG heuristic is equal to that of the MILP model (i.e. resilient scenario 1) when the demand levels are 20% and 40% for all PSs per candidate fog node. The same energy is due to the ability to use the minimum number of primary

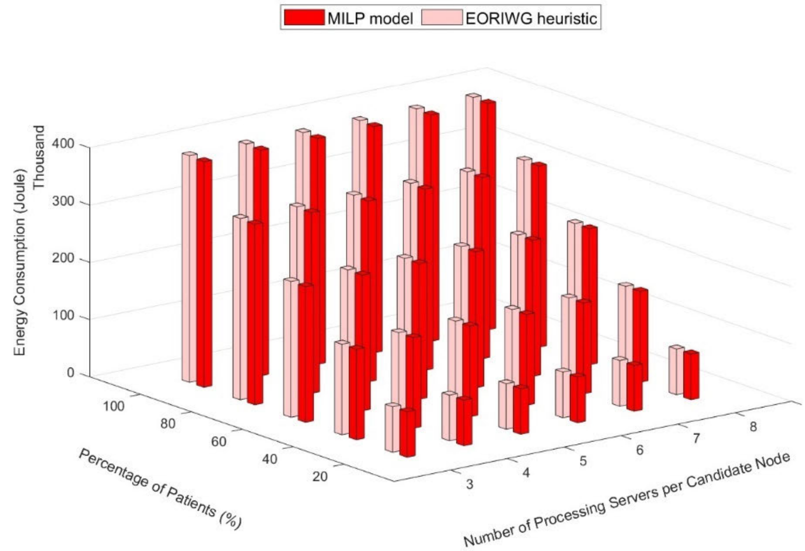


FIGURE 22. Total energy consumption of networking and processing equipment for the MILP model and the EORIWG heuristic with different percentages of total number of patients for different number of processing servers per candidate fog node.

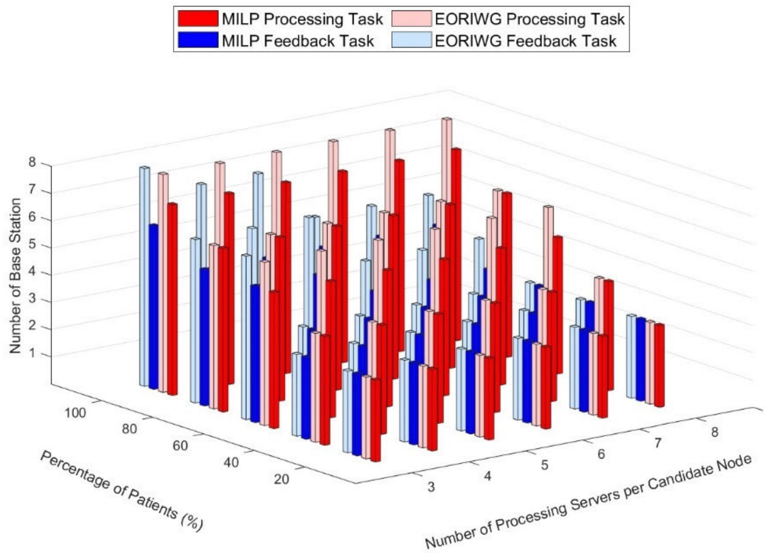


FIGURE 23. Number of base stations used to serve the processing and feedback tasks for the MILP model and the EORIWG heuristic with different percentages of total number of patients for different number of processing servers per candidate fog node.

and secondary PSs and the minimum number of fog nodes to host the PSs that are built into the EORIWG heuristic while assigning the patients from clinics to the PSs.

Figure 22 also reveals that the total energy consumed in the EORIWG heuristic is higher than the MILP model with an average of 0.17%, 0.42% and 0.44%, at demand levels of 60%, 80% and 100%, respectively. The higher energy consumed by the EORIWG heuristic is because at demand levels of 60% and 100%, increasing the total patients has resulted in utilizing more BSs to transmit the raw ECG data to the PSs, as shown in Figure 23. In the EORIGW heuristic, all BSs in cluster 1 are utilized. Due to the different connections of each clinic to the BSs, the utilization of the resources in the selected BSs is not maximized. Therefore, the BSs in cluster 2 are also used to serve the patients from the remaining clinics.

Also, at demand levels of 80% and 100%, the number of BSs utilized in the EORIGW heuristic to send the feedback traffic is higher than in the MILP model, as illustrated in Figure 23. Hence more networking equipment energy is consumed in the EORIWG heuristic compared to the MILP model. It is worth noting that increasing the number of

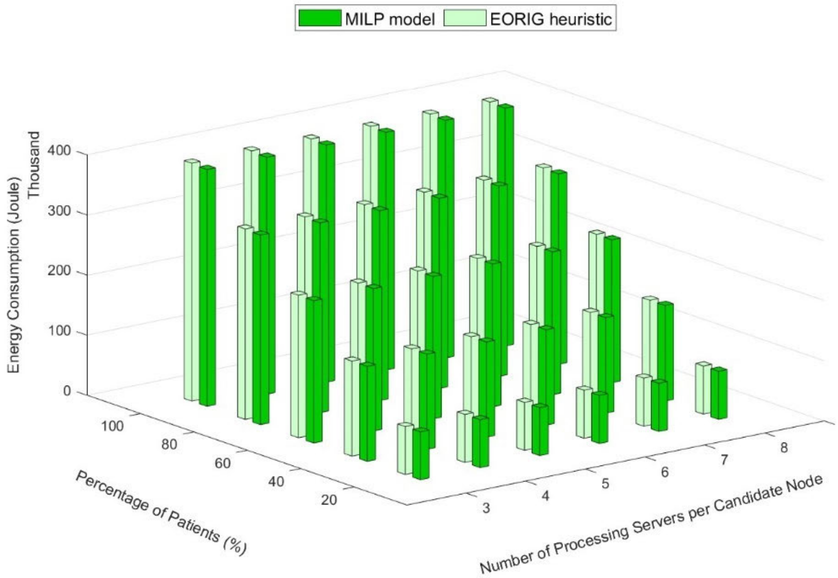


FIGURE 24. Total energy consumption of both networking equipment and processing for the MILP model and the EORIG heuristic with different percentage of patients for different number of processing servers per candidate fog node.

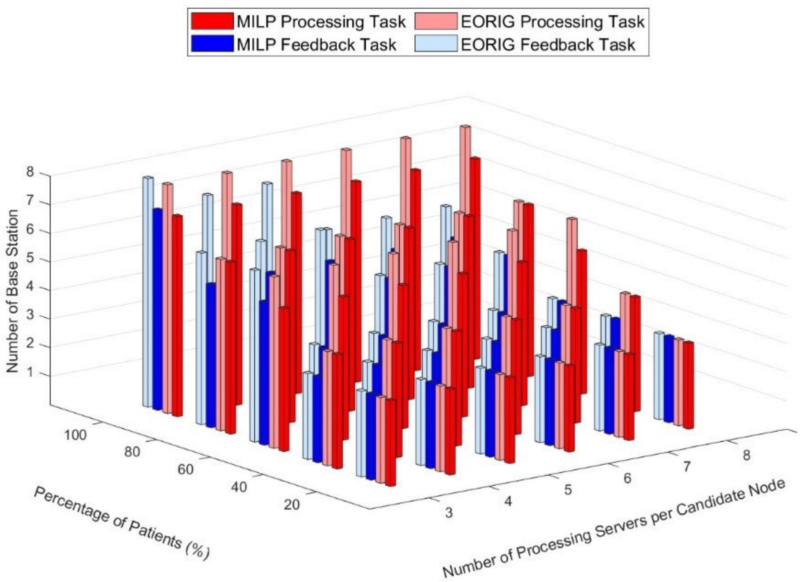


FIGURE 25. Number of base stations used to serve the processing and feedback tasks for the MILP model and the EORIG heuristic with different percentage of patients, for different number of processing servers per candidate fog node.

utilized BSs to transmit the processing traffic results in higher impact on the network energy consumption than the growing number of BSs utilized to transmit the feedback traffic. Also, in the EORIWG heuristic, the total number of fog nodes used to host the PSs is equal to the minimum number of nodes required to host the primary and secondary PSs. Therefore, due to the restricted number of fog nodes to host the PSs, the CAS is activated in the EORIWG heuristic to transmit the ECG signal to the PSs located at different clusters when the demand levels increase to 60% or are more than 60%. The utilization of the CAS has increased the network energy consumption in the EORIWG heuristic.

B. ENERGY OPTIMIZED RESILIENT INFRASTRUCTURE FOG COMPUTING WITH GEOGRAPHICAL CONSTRAINTS (EORIG) HEURISTIC

The results in Figure 24 reveal that the total energy consumption of the EORIG heuristic is equal to that produced optimally by the MILP model (i.e. resilient scenario 2) at demand levels of 20% and 40% for all PSs allowed at each

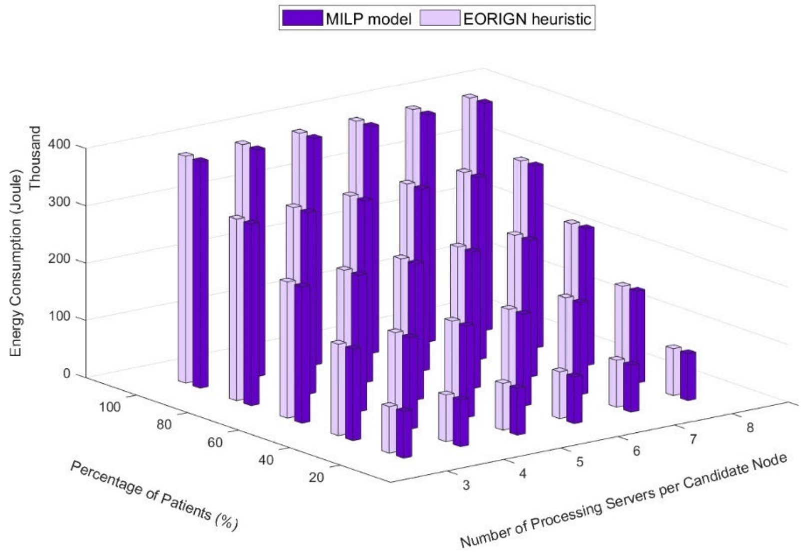


FIGURE 26. Total energy consumption of networking equipment and processing for the MILP model and the EORIGN heuristic with different percentages of the total number of patients for different number of processing servers per candidate fog node.

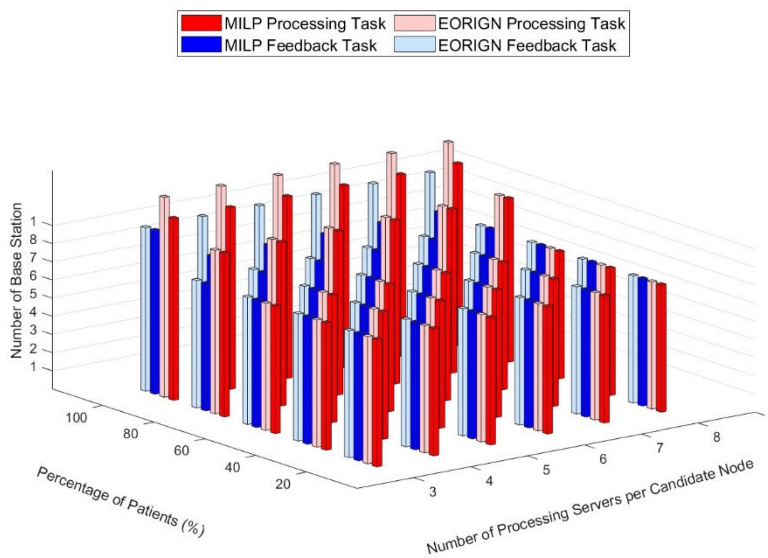


FIGURE 27. Number of base stations used to serve the processing and feedback tasks for the MILP model and the EORIGN heuristic with different percentages of the total number of patients for different number of processing servers per candidate fog node.

candidate fog node. This is mainly due to the ability to utilize the minimum number of primary and secondary PSs and the minimum number of fog nodes to host the PSs that are built in the EORIG heuristic while assigning the patients to the PSs. Also, as the size of demand is small, the total utilized networking equipment in both the EORIG heuristic and MILP model is the same.

Figure 24 also reveals that the total energy consumed in the EORIG heuristic is higher than the MILP optimization model with an average increase of 0.17%, 0.39% and 0.39% when the demand levels are 60%, 80% and 100%, respectively. The higher energy consumption of the EORIG heuristic at demand levels of 60% and 100% is due to the higher number of BSs utilized to send the processing and feedback traffic, as shown in Figure 25. Note that the BSs in cluster 1 and cluster 2 are used to serve the processing traffic due to the limitation of the connection between the clinics and the BSs. Also, at 80% and 100% of the maximum demand level, the higher energy consumption of the EORIG heuristic is because of the utilization of the CAS to relay the processing traffic between the clusters to the PSs.

TABLE 6. Summary of the finding for three sceanrios.

Type of Scenario	Findings
S1	• Compared to the non-resilient scenario, the energy penalty is due to the increasing energy consumption of both the networking and processing equipment as more networking and processing devices are utilized to serve the increasing traffic. • The energy penalty is low at low demand level (i.e. <10%). • Increasing number of processing servers that can be served at each candidate node for all demand level can either decrease or maintain the energy penalty.
S2	• Compared to S1, the energy penalty is due to the increasing energy consumption of the networking equipment due to the high utilization of fog nodes to host the processing server. • The energy penalty is high at low demand level (i.e. <10%). • Increasing number of processing servers served at each fog node at high demand level (i.e. 40% to 100%) can either decrease or increase the energy penalty.
S3	• Compared to S2, the energy penalty is due to increasing energy consumption of the networking equipment as more nodes are activated to serve the traffic. • The energy penalty is low at high demands level. • Increasing the number of processing servers served at each fog node at high demand can reduce the energy penalty.

C. ENERGY OPTIMIZED RESILIENT INFRASTRUCTURE FOG COMPUTING WITH GEOGRAPHICAL CONSTRAINTS AND LINK AND NODE DISJOINTS (EORIGN) HEURISTIC

The results in Figure 26 reveal that the total energy consumed in the EORGN heuristic is equal to the energy consumption reported by the MILP optimization model (i.e. resilient scenario 3) at demand levels of 20%, 40%, 60%, and 80% for all PSs per candidate fog node. The same energy is due to the same number of utilized networking devices and PSs in both models. Figure 26 also shows that, at a demand level of 100%, the total energy consumed by the EORIGN heuristic is slightly higher than the MILP model with an average difference of about 0.1%. This is due to the limited number of connections between the BSs and the clinics in each cluster. Therefore, more BSs are utilized in the EORIGN heuristic, as illustrated in Figure 27, to serve the processing traffic without maximizing the utilization of its resources.

VIII. CONCLUSION

This work has optimized the placement of processing servers so as to minimize the impact of increasing the resilience level on the energy consumption of networking and processing equipment when serving fog-based IoT health monitoring applications. Three resilience scenarios were considered in this work. The first two scenarios, scenario 1 and scenario 2, focus on server protection while considering the geographic locations of servers. In contrast, the third scenario, scenario 3, considers server and network protection with geographical constraints and link and node disjoint design, respectively.

The results show that considering a resilience scenario with no geographical constraints, scenario 1 has increased the energy consumed by the networking and processing equipment compared to the non-resilient scenario. This is primarily due to the high number of utilized networking equipment and PSs as adding resilience resulted in doubling the level of traffic resulting from sending the data to the primary and secondary PSs. Meanwhile, increasing the resilience level while considering geographical constraints, scenario 2 has led to a high energy penalty at low demand. This is due to the high utilization of the fog nodes to host the PSs under geographic constraints. However, increasing the resilience level does not contribute to an energy penalty when the demand level rises from 40% to 100%, and this depends on the number of PSs that can be hosted at every candidate fog node. Also, increasing the total number of PSs per candidate fog node at a demand level of more than 20%, can either decrease or increase the energy penalty. The increase in the energy penalty is because of the reduced number of fog nodes needed to host the PSs in the resilient scenario, with no geographical constraints. On the other hand, the decrease in the energy penalty is because of the reduction in the number of fog nodes needed to host the PSs in the resilient scenario 2, with geographical constraints. However, the energy penalty due to considering geographical constraints at a demand level of more than 20% is less than 7%. The results also reveal that the processing energy consumption in both resilient scenarios is equal for all PSs per candidate fog node. This is because the total number of PSs used in both scenarios are the same, as the patients were optimally consolidated in the processing servers. Increasing the level of resilience by considering geographical constraints for server protection; and link and node disjoint resilience for network protection, scenario 3, gives the same energy consumption of processing while increasing the network energy consumption. The results indicate that considering additional disjoint link and node resilience at high demands has resulted in a low network energy penalty. This is because, in any case, a large part of the network is activated due to the high demands. Also, increasing the number of PSs allowed at each candidate fog node, can reduce the network's energy penalty at high demand levels. Table 6 shows summarizes the finding of the three scenarios.

We also developed a heuristic for each scenario; EORIWG, EORIG and EORIGN for the scenarios with no geographical constraints, with geographical constraint and with geographical and link and node disjoint, respectively. The results reveal that the performance of the heuristic algorithms approaches the optimum performance obtained through the use of the MILP models. In our upcoming research endeavors, we aim to expand our current work to consider the embedding for multiple healthcare applications. Additionally, we intend to enhance energy efficiency of both networking equipment and processing by allowing one network slice to share the processing and networking infrastructure with other network slice for protection. We aim to develop a MILP model for resilient energy efficient health applications embedding.

ACKNOWLEDGMENT
All data is provided in the results section of this article.

REFERENCES
[1] 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," *J. Opt. Commun. Netw.*, vol. 10, no. 2, pp. A250–A269, Feb. 2018, doi: 10.1364/JOCN.10.00A250.
[2] M. O. I. Musa, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Bounds on GreenTouch GreenMeter network energy efficiency," *J. Lightw. Technol.*, vol. 36, no. 23, pp. 5395–5405, Dec. 2018, doi: 10.1109/JLT.2018.2871602.
[3] A. M. Al-Salim, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Energy efficient big data networks: Impact of volume and variety," *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 1, pp. 458–474, Mar. 2018, doi: 10.1109/TNSM.2017.2787624.
[4] M. Musa, T. Elgorashi, and J. Elmirghani, "Bounds for energy-efficient survivable IP over WDMnetworks with network coding," *J. Opt. Commun. Netw.*, vol. 10, no. 5, pp. 471–481, May 2018, doi: 10.1364/JOCN.10.000471.
[5] A. M. Al-Salim, T. E. H. El-Gorashi, A. Q. Lawey, and J. M. H. Elmirghani, "Greening big data networks: Velocity impact," *IET Optoelectronics*, vol. 12, no. 3, pp. 126–135, Jun. 2018, doi: 10.1049/iet-opt.2016.0165.
[6] A. N. Al-Quzweeni, A. Q. Lawey, T. E. H. Elgorashi, and J. M. H. Elmirghani, "Optimized energy aware 5G network function virtualization," *IEEE Access*, vol. 7, pp. 44939–44958, 2019, doi: 10.1109/ACCESS.2019.2907798.
[7] 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, doi: 10.1016/j.comnet.2018.01.016.
[8] X. Dong, A. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Energy-efficient core networks," in *Proc. 16th Int. Conf. Opt. Netw. Design Modelling (ONDM)*, 2012, pp. 1–9, doi: 10.1109/ONDM.2012.6210196.
[9] 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. 2011, doi: 10.1109/JLT.2011.2148093.
[10] H. M. M. Ali, T. E. H. El-Gorashi, A. Q. Lawey, and J. M. H. Elmirghani, "Future energy efficient data centers with disaggregated servers," *J. Lightw. Technol.*, vol. 35, no. 24, pp. 5361–5380, Dec. 2017, doi: 10.1109/JLT.2017.2767574.
[11] 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 2015, doi: 10.1109/JLT.2014.2380777.
[12] 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. 4209–4225, Nov. 2014, doi: 10.1109/JLT.2014.2351074.
[13] 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. 2014, doi: 10.1109/JLT.2014.2324634.
[14] X. Dong, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "On the energy efficiency of physical topology design for IP over WDM networks," *J. Lightw. Technol.*, vol. 30, no. 12, pp. 1931–1942, Jun. 2012. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI=jlt-30-11-1694
[15] 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. 2011, doi: 10.1109/JLT.2010.2086434.
[16] M. Musa, T. Elgorashi, and J. Elmirghani, "Energy efficient survivable IP-over-WDM networks with network coding," *J. Opt. Commun. Netw.*, vol. 9, no. 3, pp. 207–217, Mar. 2017.
[17] Z. T. Al-Azez, A. Q. Lawey, T. E. H. 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, doi: 10.1109/ACCESS.2019.2911117.
[18] B. G. Bathula, M. Alresheedi, and J. M. H. Elmirghani, "Energy efficient architectures for optical networks," in *Proc. IEEE London Commun. Symp.*, London, U.K., Sep. 2009, pp. 5–8, doi: 10.1117/12.852292.
[19] B. G. Bathula and J. M. H. Elmirghani, "Energy efficient optical burst switched (OBS) networks," in *Proc. IEEE Globecom Workshops*, Nov. 2009, pp. 1–6, doi: 10.1109/GLOCOMW.2009.5360734.
[20] I. S. B. M. Isa, T. E. H. El-Gorashi, M. O. I. Musa, and J. M. H. Elmirghani, "Energy efficient fog-based healthcare monitoring infrastructure," *IEEE Access*, vol. 8, pp. 197828–197852, 2020, doi: 10.1109/ACCESS.2020.3033555.
[21] A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Distributed energy efficient clouds over core networks," *J. Lightw. Technol.*, vol. 32, no. 7, pp. 1261–1281, Apr. 2014, doi: 10.1109/JLT.2014.2301450.
[22] M. S. Hadi, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Patient-centric cellular networks optimization using big data analytics," *IEEE Access*, vol. 7, pp. 49279–49296, 2019.
[23] F. Jalali, K. Hinton, R. Ayre, T. Alpcan, and R. S. Tucker, "Fog computing may help to save energy in cloud computing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1728–1739, May 2016, doi: 10.1109/JSAC.2016.2545559.
[24] F. Jalali, A. Vishwanath, J. De Hoog, and F. Suits, "Interconnecting Fog computing and microgrids for greening IoT," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Eur.*, Nov. 2016, pp. 693–698, doi: 10.1109/ISGT-Asia.2016.7796469.
[25] I. S. M. Isa, M. O. I. Musa, T. E. H. El-Gorashi, A. Q. Lawey, and J. M. H. Elmirghani, "Energy efficiency of fog computing health monitoring applications," in *Proc. 20th Int. Conf. Transparent Opt. Netw. (ICTON)*, Jul. 2018, pp. 1–5, doi: 10.1109/ICTON.2018.8473698.
[26] F. M. R. Junior and C. A. Kamienski, "A survey on trustworthiness for the Internet of Things," *IEEE Access*, vol. 9, pp. 42493–42514, 2021, doi: 10.1109/ACCESS.2021.3066457.
[27] V. Prokhorenko and M. Ali Babar, "Architectural resilience in cloud, fog and edge systems: A survey," *IEEE Access*, vol. 8, pp. 28078–28095, 2020, doi: 10.1109/ACCESS.2020.2971007.
[28] C. Colman-Meixner, C. Develder, M. Tornatore, and B. Mukherjee, "A survey on resiliency techniques in cloud computing infrastructures and applications," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2244–2281, 3rd Quart., 2016, doi: 10.1109/COMST.2016.2531104.
[29] R. S. Couto, S. Secci, M. E. M. Campista, and L. H. M. K. Costa, "Server placement with shared backups for disaster-resilient clouds," *Comput. Netw.*, vol. 93, pp. 423–434, Dec. 2015, doi: 10.1016/j.comnet.2015.09.039.
[30] C. Develder, J. Buysse, B. Dhoedt, and B. Jaumard, "Joint dimensioning of server and network infrastructure for resilient optical grids/clouds," *IEEE/ACM Trans. Netw.*, vol. 22, no. 5, pp. 1591–1606, Oct. 2014, doi: 10.1109/TNET.2013.2283924.
[31] D. M. Senejohnny, S. Sundaram, C. De Persis, and P. Tesi, "Resilience against misbehaving nodes in asynchronous networks," *Automatica*, vol. 104, pp. 26–33, Jun. 2019, doi: 10.1016/j.automatica.2019.02.035.
[32] A. Modarresi and J. P. G. Sterbenz, "Toward resilient networks with fog computing," in *Proc. 9th Int. Workshop Resilient Netw. Design Modeling (RNDM)*, 2017, pp. 1–7.
[33] D. P. Abreu, K. Velasquez, L. Paquete, M. Curado, and E. Monteiro, "Resilient service chains through smart replication," *IEEE Access*, vol. 8, pp. 187021–187036, 2020, doi: 10.1109/ACCESS.2020.3030537.

[34] M. A. Madani, F. Zhou, and A. Meddahi, "Deploying disaster-resilient service function chains using adaptive multi-path routing," in *Proc. 19th Int. Conf. Netw. Service Manage. (CNSM)*, Oct. 2023, pp. 1–5, doi: 10.23919/cnsm59352.2023.10327915.

[35] M. Whaiduzzaman, A. Barros, A. R. Shovon, M. R. Hossain, and C. Fidge, "A resilient fog-IoT framework for seamless microservice execution," in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Sep. 2021, pp. 213–221, doi: 10.1109/SCC53864.2021.00034.

[36] A. Matange, V. Taneja, A. Chaumal, P. Buwa, and J. Abraham, "Resilient fog-based adaptive traffic," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol.*, 2023, pp. 1–7, doi: 10.1109/ICCCNT56998.2023.10306713.

[37] X. Masip-Bruin, S. Sanchez-Lopez, A. Jurnet, E. Marin-Tordera, A. Jukan, and G. J. Ren, "Towards a resilient control architecture for combined fog-to-cloud systems," in *Proc. IEEE 8th Int. Conf. Cloud Netw. CloudNet*, Nov. 2019, pp. 3–6, doi: 10.1109/CloudNet47604.2019.9064116.

[38] I. S. M. Isa, M. O. I. Musa, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Energy efficient and resilient infrastructure for fog computing health monitoring applications," in *Proc. Int. Conf. Transparent Opt. Netw.*, Jul. 2019, pp. 1–5, doi: 10.1109/ICTON.2019.8840438.

[39] I. Kantor, J.-L. Robineau, H. Bütün, and F. Maréchal, "A mixed-integer linear programming formulation for optimizing multi-scale material and energy integration," *Frontiers Energy Res.*, vol. 8, p. 49, Apr. 2020, doi: 10.3389/fenrg.2020.00049.

[40] I. S. B. M. Isa. (Oct. 2019). *Energy Efficient and Resilient Internet of Things Networks*. [Online]. Available: http://etheses.whiterose.ac.uk/26627/

[41] B. Vault. (2021). *U.K. Datacentres*. [Online]. Available: https://www.backupvault.co.uk/uk-datacentres/

[42] N. Lowres, S. B. Freedman, R. Gallagher, A. Kirkness, D. Marshman, J. Orchard, and L. Neubeck, "Identifying postoperative atrial fibrillation in cardiac surgical patients posthospital discharge, using iPhone ECG: A study protocol," *BMJ Open*, vol. 5, no. 1, Jan. 2015, Art. no. e006849, doi: 10.1136/bmjopen-2014-006849.

[43] Public Health England. (2015). *Statistic of Patients in West Leeds*. [Online]. Available: http://healthierlives.phe.org.uk/topic

[44] S. Knows. *FTTC Exchanges*. London, U.K.: UK Broadband Availability, 2022. [Online]. Available: https://availability.samknows.com/broadband/exchanges/bt/fttc

IDA SYAFIZA M. ISA received the Ph.D. degree from the University of Leeds, U.K., in 2020, with a focus on energy-efficient access networks design for healthcare applications. She is currently a Senior Lecturer with Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Her research interests include network architecture design, energy efficiency, network optimization, mixed integer linear programming, and healthcare systems. She has published several papers in this area.

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. From 2010 to 2014, she was a Postdoctoral Researcher with the University of Leeds, 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 data centers, 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. She is currently a Lecturer in optical networks with the School of Electronic and Electrical Engineering, University of Leeds. 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, the Optical Network Design and Modeling Conference, the Optical Fiber Communications Conference, the International Conference on Computer Communications, EU Future Internet Assembly, IEEE Sustainable ICT Summit, IEEE 5G World Forum, and Collaboration with Nokia and Huawei.

MOHAMED O. I. MUSA (Member, IEEE) received the B.Sc. degree (Hons.) in electrical and electronic engineering from the University of Khartoum, Khartoum, Sudan, in 2009, and the M.Sc. degree (Hons.) in broadband wireless and optical communication and the Ph.D. degree in energy efficient network coding in optical networks from the University of Leeds, Leeds, U.K., in 2011 and 2016, respectively. His current research interests include ICT energy optimization, network coding, and energy-efficient routing protocols in optical networks.

JAAFAR M. H. ELMIRGHANI (Fellow, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the University of Khartoum, in 1989, the Ph.D. degree in 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 2012. From 2000 to 2007, he was the Chair of Optical Communications with the University of Wales, Swansea. In 2007, he joined the University of Leeds as a Full Professor. He is currently the Director of the Institute of Communication and Power Networks and a Professor of communication networks and systems with the School of Electronic and Electrical Engineering, University of Leeds. He coauthored *Photonic Switching Technology: Systems and Networks* (Wiley). He was a PI of the £6 million EPSRC Intelligent Energy Aware Networks (INTERNET) Program Grant from 2010 to 2016. He is a PI of the EPSRC £6.6m Terabit Bidirectional Multi-User Optical Wireless System (TOWS) for 6G LiFi from 2019 to 2024. He has published over 550 technical articles. He has been awarded more than 30 million in grants to date from EPSRC, the EU, and industry. He has held prestigious fellowships funded by the Royal Society and BT. He leads a number of research projects. His research interests include communication networks, and wireless and optical communication systems. He is a fellow of IET and the Institute of Physics. He was elected as a fellow of IEEE for "Contributions to Energy-Efficient Communications," in 2021. He was a member of the IEEE ComSoc Technical Activities Council (TAC). He was awarded all four prizes at the Department of Electrical Engineering for academic distinction. He received the IEEE Communications Society 2005 Hal Sobol Award for exemplary service to meetings and conferences, the IEEE Communications Society 2005 Chapter Achievement Award, the University of Wales Swansea Inaugural Outstanding Research Achievement Award, in 2006, the IEEE Communications Society Signal Processing and Communication Electronics Outstanding Service Award, in 2009, the Best Paper Award from IEEE ICC'2013, the IEEE Comsoc Transmission Access and Optical Systems Outstanding Service Award in recognition of "Leadership and Contributions to the Area of Green Communications,'' in 2015, the GreenTouch 1000x Award for ''Pioneering Research Contributions to the Field of Energy Efficiency in Telecommunications'' in 2015, the IET 2016 Premium Award for Best Paper in *IET Optoelectronics*, and the 2016 Edison Award in the collective disruption category with a team of six from GreenTouch for their joint work on the GreenMeter, the IEEE Communications Society Transmission, Access, and Optical Systems Technical Committee in 2020 Outstanding Technical Achievement Award for outstanding contributions to the ''energy efficiency of optical communication systems and networks.'' He was the Chairman of the IEEE U.K. and RI Communications Chapter, the IEEE Comsoc Transmission Access and Optical Systems Committee, and the IEEE Comsoc Signal Processing and Communication Electronics (SPCE) Committee. He is and has been on the Technical Program Committee of 41 IEEE ICC/GLOBECOM conferences, from 1995 to 2020, including 19 times as the Symposium Chair. He was the founding Chair of the Advanced Signal Processing for Communication Symposium which started at IEEE GLOBECOM'99 and has continued since at every ICC and GLOBECOM, the first IEEE ICC/GLOBECOM Optical Symposium at GLOBECOM'00, and the Future Photonic Network Technologies, Architectures and Protocols Symposium. He chaired this Symposium, which continues to date. He was also the founding Chair of the first Green Track at ICC/GLOBECOM at GLOBECOM 2011, the Co-Chair of the GreenTouch Wired, Core and Access Networks Working Group, an Adviser of the Commonwealth Scholarship Commission, and a member of the Royal Society International Joint Projects Panel and the Engineering and Physical Sciences Research Council (EPSRC) College. He has been the Chair of the IEEE Sustainable ICT Initiative, a pan IEEE Societies Initiative responsible for Green ICT activities across IEEE, since 2012. He was an Editor of IEEE Communications Surveys and Tutorials and IEEE Journal on Selected Areas in Communications series on Green Communications and Networking. He is currently an Area Editor of IEEE Journal on Selected Areas in Communications Series on Machine Learning for Communications, and an Editor of IEEE Journal of Lightwave Technology, *IET Optoelectronics*, Journal of Optical Communications, and *IEEE Communications Magazine*. He has given over 90 invited and keynote talks over the past 15 years. He was named among the top 2% of scientists in the world by citations, in 2019, in Elsevier Scopus, Stanford University database which includes the top 2% of scientists in 22 scientific disciplines and 176 sub-domains. He was an IEEE Comsoc Distinguished Lecturer, from 2013 to 2016.