

Article

Energy Efficiency and Network Performance: A Reality Check in SDN-Based 5G Systems [†]

Adriana Fernández-Fernández * , Cristina Cervelló-Pastor  and Leonardo Ochoa-Aday 

Department of Network Engineering, Universitat Politècnica de Catalunya (UPC), Esteve Terradas, 7, 08860 Castelldefels, Spain; cristina@entel.upc.edu (C.C.-P.); leonardo.ochoa@entel.upc.edu (L.O.-A.)

* Correspondence: adriana.fernandez@entel.upc.edu; Tel.: +34-93-413-7027

[†] This paper is an extended version of our paper published in Fernández-Fernández, A.; Cervelló-Pastor, C.; Ochoa-Aday, L. Evaluating the Impact of Energy-Aware Routing on Software-Defined Networking Performance. In Proceedings of the XIII Jornadas de Ingeniería Telemática (JITEL 2017), Valencia, Spain, 27–29 September 2017.

Received: 14 November 2017; Accepted: 11 December 2017; Published: 14 December 2017

Abstract: The increasing power consumption and related environmental implications currently generated by large data networks have become a major concern over the last decade. Given the drastic traffic increase expected in 5G dense environments, the energy consumption problem becomes even more concerning and challenging. In this context, Software-Defined Networks (SDN), a key technology enabler for 5G systems, can be seen as an attractive solution. In these programmable networks, an energy-aware solution could be easily implemented leveraging the capabilities provided by control and data plane separation. This paper investigates the impact of energy-aware routing on network performance. To that end, we propose a novel energy-aware mechanism that reduces the number of active links in SDN with multiple controllers, considering in-band control traffic. The proposed strategy exploits knowledge of the network topology combined with traffic engineering techniques to reduce the overall power consumption. Therefore, two heuristic algorithms are designed: a static network configuration and a dynamic energy-aware routing. Significant values of switched-off links are reached in the simulations where real topologies and demands data are used. Moreover, the obtained results confirm that crucial network parameters such as control traffic delay, data path latency, link utilization and Ternary Content Addressable Memory (TCAM) occupation are affected by the performance-agnostic energy-aware model.

Keywords: 5G; software-defined networking; energy-aware routing; in-band control traffic

1. Introduction

Over the last decade, the exponential demand growth and the ever-increasing number of connected devices have forced the necessity to look to the next evolution of wireless data communications [1]. The adoption of 5G networks, expected by 2020, will allow handling more traffic in dense environments, providing higher data rates and reduced end-to-end latency [2]. In this scenario, achieving energy efficiency becomes even more concerning and challenging. For instance, according to [3], by 2025, the global Internet will be responsible for more than 10% of the world's electricity consumption. To face this steadily rising power consumption, a strongly related Key Performance Indicator (KPI) to be addressed by the 5G generation is precisely the energy efficiency [4]. Accordingly, an effective energy management, as well as enhanced network performance, are essential design goals to fulfill the requirements of future 5G systems for heterogeneous applications and services. However, some of these requirements may be in conflict, and specific strategies must be developed. Precisely, the existing trade-off between energy efficiency and network performance was investigated as part of a conference paper in [5] and is now further expanded in this article.

Energy consumption concern in 5G system has currently attracted a great deal of attention from networking researchers. Several papers have been proposed with solutions enabling significant energy efficiency gains in the mobile networks division [6–8]. Although 5G is mostly perceived as wireless access by the user, different communication facilities and users need to be connected through backhaul networks—either using optical fiber networks or radio links—and backbone networks. Apart from being key enablers for a successful deployment of this complex architecture, these network segments are of paramount importance to reduce the energy consumption of 5G systems. Therefore, this paper is conceived to tackle the energy consumption problem in 5G backbone networks.

Software-Defined Networking (SDN) is expected to play a major role in 5G systems in order to provide a more intelligent use of the underlying transport networks [9]. Moreover, this technology, together with Network Function Virtualization (NFV), can provide the required tools to support network slicing in order to accommodate the wide range of demanded services over a common infrastructure simultaneously [10–12]. The basic idea of SDN [13]—control and data planes separation—makes network environments more manageable. The logically centralized control plane in SDN provides global knowledge of the network state information, which allows for end-to-end visualization. Moreover, it can manage network tasks and perform device programming without needing any additional software or hardware-based intelligence in each one of the switching elements. Meanwhile, interconnection devices follow the rules set by the controller to forward the traffic. Consequently, the use of an underlying SDN architecture will facilitate the introduction and deployment of new applications and services, making it easier than with classical hardware-dependent standards. Another advantage of exploiting SDN is the possibility to dynamically adapt control decisions to comply with diverse QoS requirements to handle heterogeneous application-driven networks.

Given that in practice, the energy consumption of network equipment is not in proportion with their traffic load, the reduction of the number of active elements is an effective and widely-accepted strategy to decrease the consumption of data networks [14]. This feature can be implemented by putting into sleep mode (i.e., a low-power state) unused networks elements such as line cards or port interfaces. Although turning off entire interconnection devices improves the energy efficiency, this possibility is not considered in this work given the resilience concerns in the case of network events. Nevertheless, due to the link over-provisioning typically considered in the design and operation of backbone networks, substantial energy can still be saved putting into sleep mode port interfaces that are not transferring data. Within this context, SDN architecture is very well-suited to perform an energy-aware routing and to manage the state of unused switch interfaces in a coordinated and centralized way. Therefore, the implementation of an energy-aware solution in the control plane is a valuable opportunity to solve the power consumption problem in data networks.

Despite consistent efforts to improve the network power efficiency, energy-aware techniques may lead to performance degradations if QoS requirements are neglected [15,16]. Inspired by this reality, this paper introduces a new energy-aware strategy and evaluates its impact on different performance metrics. Instead of restricting the path selection and potential improvements in terms of energy efficiency to meet some specific metric bound, this work aims to gain insight into potential energy savings and to quantify the existing trade-off between power consumption and several performance indicators, as one crucial issue for communication systems nowadays.

Throughout this work, we consider an SDN architecture with multiple controllers and, similar to previous works [17–19], in-band control traffic. In this operational mode, links are shared between data and control plane traffic. Hence, control messages are exchanged without the need for additional links. In this way, the proposed energy-aware routing can be applied when implementing a dedicated control network is not feasible either for physical or cost-related restrictions. In large backbone networks, this is a more realistic scenario since additional links dedicated to directly connect controllers and forwarding devices are impractical and cost-inefficient.

Specifically, the major contributions of this work are as follows:

- An Integer Linear Problem (ILP) is formulated to optimize the number of active links in SDN, considering multiple controllers and links shared between data and control plane traffic.
- For large network topologies, we propose a novel energy-aware mechanism that reduces the time complexity of our approach and allows allocating the traffic demands in real time as it comes.
- Two solution modules were conceived of in this mechanism, exploiting knowledge of the network topology and traffic engineering techniques to reduce the overall power consumption.
- Using real topologies and traffic demands, we provide a performance comparison analysis of our proposal with another routing approach.

The remainder of this paper is structured as follows. In Section 2, previous related studies about different strategies to tackle the problem of power consumption are discussed. In Section 3, the energy consumption optimization problem is formalized through a general mathematical formulation fully compatible with SDN environments using multiple controllers and in-band control traffic. In Section 4, we explain the main characteristics of our low-complexity energy-aware approach together with a detailed description of its two comprised modules. The simulations strategies and the obtained results are presented and analyzed in Section 5. Finally, in Section 6, we provide the conclusion of our work and outline future research guidelines.

2. Related Works

Energy-aware techniques based on the reduction of active network elements can be divided into traffic-based and topology-based solutions, according to the elements considered in the model. In this section, we analyze in more detail works that deal with each one of these two approaches.

2.1. Traffic-Based Solutions

Under some assumption of expected traffic behavior, traffic-based solutions are routing mechanisms that aggregate traffic over a network subset in over-provisioned systems, in order to switch off the unused network components.

For instance, Zhang et al. [20] propose an intra-domain, centralized traffic engineering mechanism, called GreenTE, that finds a set of links that can be turned off under a given traffic load or matrix. The approach is based on a Mixed-Integer Linear Programming (MILP) formulation where traffic demands are routed using a set of pre-computed k-shortest paths. Performance requirements such as Maximum Link Utilization (MLU) and network delay are considered as constraints in the problem. However, the implementation of such a coordinated strategy is a difficult task given the distributed nature of network control in traditional networks.

More recently, in [21], the authors introduced a state-of-the-art study including different energy efficiency strategies in SDN. This paper addresses the importance of implementing green routing methods in SDN, taking advantage of the flexible control provided by dynamic configuration and centralized network view capabilities. A summary of some existing energy-aware techniques in SDN with their key properties (benefits and drawbacks) is presented.

The problem of saving energy in partially deployed SDN is addressed by the authors of [22]. They formulated an optimization problem to determine network subsets in these hybrid systems with minimum power consumption. Giroire et al. [23] proposed an energy-aware routing approach considering as a constraint of the model the limited rule space of Ternary Content Addressable Memory (TCAM) in SDN devices. For this purpose, an ILP formulation, as well as an efficient heuristic are proposed. The authors of [24] provided two greedy algorithms to reduce the power of used line-cards and integrated chassis. To do so, they used the connections between forwarding devices to derive an expanded network topology. Markiewicz et al. [25] formulated an MILP model that aims to switch on a minimum number of routers and links to handle the traffic. To solve the problem for large networks, they present a heuristic method, called Strategic Greedy Heuristic (SGH), that iteratively selects a pre-computed shortest path for each request, according to four different strategies of processing order

of requests. Nevertheless, in all these works, a dedicated network for control plane communications was considered.

In [26], the authors proposed a model for controller-switch associations, called GreCo, which improves the energy efficiency of the network. In this work, the routing of control traffic is included. However, they considered that controllers act as well as forwarding devices, allowing that data plane communications are routed through network controllers. In this way, only links belonging to control paths are activated while data traffic demands are routed using these links under a given MLU bound. In our work, the routing of data plane traffic through network controllers is avoided, since this will represent an additional unnecessary load in these devices.

The work in [27] addressed the problem of minimizing the number of required links in large-scale SDN considering an in-band operation mode. To accomplish this, an ILP model and a heuristic algorithm are presented, integrating the routing requirements for data and control traffic. In this model, the distribution of switches between controllers is also optimized in terms of energy efficiency and load balancing. In [28], a distributed routing algorithm optimizing the energy consumption in large-scale SDN with multiple domains is proposed. This solution, called Distributed Energy-Aware Routing (DEAR), finds the minimum number of links needed to satisfy a given traffic matrix. Despite being efficient models, the complexity of considering the entire topology for the selection of the most suitable routes can be very expensive in networks with major path redundancy. To the contrary, in this work, after pruning the network topology, the number of paths and the consequent computation complexity are significantly reduced.

2.2. Topology-Based Solutions

The lack of awareness of traffic conditions in typical operative networks has led to several research works that, in order to reduce the number of active links, are oriented to control the network topology. Basically, these approaches modify the existing topology considering different requirements such as the resulting connectivity.

In [29], the authors present an Open Shortest Path First (OSPF)-based routing mechanism that considers the topological information exchanged among routers. The proposed EAR algorithm is based on the definition of the “exportation” mechanism where a Shortest Path Tree (SPT) is shared between neighbor nodes. The routers with the highest node-degree, called “exporters”, calculate the SPTs that are used to route the traffic and force the use of these paths to all their neighbors, which reduces the overall set of active links. The exportation mechanism is enhanced in [30], where the concept of “move” was introduced turning the energy saving routing problem into a formulation of the well-known maximum clique problem in an undirected weighted graph.

The authors in [31] proposed a routing algorithm denoted as Energy Saving based on Algebraic Connectivity (ESACON), using the algebraic connectivity as a metric to control the resulting network topology. Based on this metric, ESACON is able to identify and switch off the network links that affect the network connectivity less, keeping this value over a given threshold.

Similarly, the topology-based solution reported in [32] also takes into account the algebraic connectivity as a requirement to preserve the overall network connectivity. This work also considers the edge betweenness as a metric to measure the links role in the network, placing the links least frequently used as the first candidates to be pruned. However, this approach is conceived of to be implemented in a distribute way into each IP router.

The work in [33] also aims to improve the energy efficiency reducing the number of links that will remain active. For this purpose, the authors propose four different versions of the algorithm called Energy Saving based on Occurrence of Links (ESOL) that show the trade-off between complexity and efficiency in putting into sleep mode a great number of links. The parameters used in this approach to select the network interfaces to be switched off are the occurrences of nodes and links in shortest paths, extracted from the network topology by using the classical Dijkstra algorithm.

The analysis of including QoS requirements in an energy-aware topology-based solution is discussed by the authors of [34]. Their approach, called Energy Saving IP Routing (ESIR), is also based on the concepts of SPT exportation and move, but constrained to a maximum load boundary on network links in the traffic-aware scenario. However, in their study, fixed requirements are considered, which is impractical and limits the suitability of their work for dynamic and heterogeneous network services.

All the previously-described works mainly tend to minimize the number of active network elements in the current topology restricting the path selection to meet some specific metric bound or connectivity rate and fail to extensively examine the impact of energy-aware routing on SDN performance. Moreover, their lack of awareness about the requirements of incoming connection requests can lead to performance degradations, which is highly undesired.

A summary of the discussed energy-aware proposals is presented in Table 1. Each row in the table refers to a different approach. Meanwhile, columns refer to a particular feature: proposal description, performance metrics used as constraints and networking scenario considered. Our approach is also included in the table, where a novel point of view, fully compatible with current dynamic networking environments, is exploited.

Different from the aforementioned works, the aim of this paper is to provide a low-complexity energy-aware strategy, which will be used to evaluate its impact on crucial performance metrics, considering an SDN architecture with multiple controllers and in-band control traffic.

Table 1. Comparison between the proposed approach and other existing related works. MLU, Maximum Link Utilization; OSPF, Open Shortest Path First; MPLS, Multiprotocol Label Switching; TCAM, Ternary Content Addressable Memory; SPT, Shortest Path Tree; LSA, Link State Advertisements.

Reference No.	Proposal Description	QoS Constraints	Networking Environment
[20]	Intra-domain, centralized green traffic allocation under a given traffic matrix	MLU and packet delay	OSPF/MPLS-compliant IP networks
[22]	Centralized green traffic allocation using a fewer number of SDN elements under a given traffic matrix	Traffic priority is considered	Hybrid networks (IP/SDN)
[23]	Centralized green traffic allocation and TCAM size reduction by using the concept of default rule	Subject to TCAM capacities	SDN with centralize and out-of-band control
[24]	Centralized green traffic routing based on an expanded topology of chassis and line-cards under a given traffic matrix	MLU and packet delay	SDN with centralize and out-of-band control
[25]	Green path selection of precomputed shortest paths with different processing order of a given traffic matrix	Only k-shortest paths are used	SDN with centralize and out-of-band control
[26]	Green traffic allocation with Ct-Sw association under traffic knowledge and controllers acting as switches	MLU and control paths delay	In-band SDN with multiple controllers
[27]	Off-line green control and data path computation with Ct-Sw association under traffic knowledge	MLU and control paths delay	In-band SDN with multiple controllers
[28]	Distributed intra-domain green routing with inter-domain path selection based on the exchange of performance metrics	MLU	Multi-domain SDN with predefined Ct-Sw association
[29]	Selection of a subset of router SPTs as routing paths based on node-degree and the exchange of LSA messages	N/A	OSPF-compliant IP networks
[31]	Off-line detection and switching off of links keeping algebraic connectivity above a given threshold	N/A	OSPF-compliant IP networks
[32]	Off-line detection and switching off of links based on link betweenness and algebraic connectivity thresholds	N/A	OSPF-compliant IP networks
[33]	Off-line detection and switching off of links based on their occurrence in network shortest paths	N/A	OSPF-compliant IP networks
[34]	Selection of green routing paths based on the concepts of SPT exportation and move with traffic-awareness	MLU	OSPF-compliant IP networks
Current paper	Topology-based control plane setup and network pruning with online green data and control traffic allocation	Subject to controllers capacity	In-band SDN with multiple controllers

3. The Energy-Aware Routing Problem in SDN

To formalize the energy consumption optimization problem, we consider an SDN represented by a directed graph $G = (V, E, C)$, being V , E and C the set of nodes, links and controllers respectively, where $C \subset V$. We use $c_{i,j}$ to denote the capacity of a link $(i, j) \in E$. The set of interconnection devices is defined as $S = \{n \mid n \in V \wedge n \notin C\}$.

Considering F as the entire set of traffic flows existing in the network between any pair of nodes, we use D to denote the subset of data plane communications. For the control plane, let T denote the subset of communications between network controllers and switches, while H is used as the subset of communications between controllers. Accordingly, $F = D \cup T \cup H$. Each flow $f \in F$ from source s_f to destination t_f , has its associated throughput, denoted by b_f .

To minimize the number of links needed to route a given traffic demand matrix, we develop an ILP model, according to the following binary variables:

$x_{i,j}$: describes the state of a link $(i, j) \in E$.

$$x_{i,j} = \begin{cases} 1 & \text{if } (i, j) \text{ is active,} \\ 0 & \text{otherwise.} \end{cases}$$

$t_{i,j}^f$: describes the selection of a link $(i, j) \in E$ to route a flow $f \in F$.

$$t_{i,j}^f = \begin{cases} 1 & \text{if } (i, j) \text{ is selected to route } f, \\ 0 & \text{otherwise.} \end{cases}$$

$\lambda_{n,c}$: describes the association of each forwarding device $n \in S$ with a controller $c \in C$.

$$\lambda_{n,c} = \begin{cases} 1 & \text{if } n \text{ is associated with } c, \\ 0 & \text{otherwise.} \end{cases}$$

Considering the complete set of demands fixed and known in advance, a global optimization process can jointly compute all the required optimal control and data paths minimizing the energy efficiency. Using the aforementioned binary variables, the optimization model can be formulated as:

$$\text{minimize } \sum_{(i,j) \in E} x_{i,j} \quad (1)$$

To manage each forwarding device in the network $n \in S$, a single controller is selected.

$$\sum_{c \in C} \lambda_{n,c} = 1 \quad \forall n \in S \quad (2)$$

Additionally, the number of switches associated with each controller cannot exceed the controller capacity. In this expression, we use R_c to denote the computational and networking resources, in terms of number of devices that can be supported for a controller $c \in C$.

$$\sum_{n \in S} \lambda_{n,c} \leq R_c \quad \forall c \in C \quad (3)$$

To avoid additional traffic load through network controllers, data plane communications (i.e., $f \in D$) cannot be routed through these devices. Furthermore, control traffic between controllers and switches (i.e., $f \in T$) will not pass through any other controller (i.e., except for those being the source or target of the traffic). The same must hold true for communications between controllers

(i.e., $f \in H$). In these constraints, we use $N(i)$ to denote the set of neighbors of a node i and n_f to identify the forwarding device involved in the source/target pair of traffic flow $f \in T$.

$$\sum_{j \in N(i)} t_{i,j}^f \leq \begin{cases} 0 & \forall f \in D, \forall i \in C \\ \lambda_{n_f,i} & \forall f \in T, \forall i \in C \\ 0 & \forall f \in H, \forall i \in C \setminus \{s_f, t_f\} \end{cases} \quad (4)$$

The routing of data plane communications and control traffic exchange between controllers follows the traditional flow conservation constraints.

$$\forall i \in V, \forall f \in D \cup H : \sum_{j \in N(i)} t_{i,j}^f - \sum_{j \in N(i)} t_{j,i}^f = \begin{cases} 1 & \text{if } i = s_f \\ -1 & \text{if } i = t_f \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Meanwhile, for the subset of communications between controllers and switches $f \in T$, these constraints are modified to assure that every switch exchanges control messages only with its controller. Similarly, the forwarding device and controller involved in the source/target pair of traffic flow $f \in T$, are denoted with n_f and c_f , respectively.

$$\forall i \in V, \forall f \in T : \sum_{j \in N(i)} t_{i,j}^f - \sum_{j \in N(i)} t_{j,i}^f = \begin{cases} \lambda_{n_f,c_f} & \text{if } i = s_f \\ -\lambda_{n_f,c_f} & \text{if } i = t_f \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

A link (i, j) is active if it is used by some traffic flow $f \in F$. Furthermore, the total traffic in each active link must be less than its assigned capacity.

$$\sum_{f \in F} t_{i,j}^f b_f \leq c_{i,j} x_{i,j} \quad \forall (i, j) \in E \quad (7)$$

Using this model, the centralized controller can determine all the optimal routes and set the corresponding flow rules on each interconnection device in the network before the traffic arrival. Although this formulation allows the attainment of optimal solutions for the energy consumption problem in SDN, it becomes highly challenging to solve on large or even medium-scale topologies. This is because the complexity of the energy-aware routing problem is NP-hard [35], so the required computational time and resources grow exponentially with the network size.

4. Energy-Aware Approach

Given the high complexity of the proposed optimization problem in large real-world networks, in this section, we present a hybrid solution for the energy efficiency problem in SDN comprising the main advantages of the two aforementioned solution types. More precisely, we exploit specific network topological properties combined with the use of traffic engineering to reduce the overall power consumption.

An illustrative diagram of this strategy is shown in Figure 1. The first component, denoted as Static Network Configuration Algorithm (SNetCA), is a topology-based solution intended to be statically activated at specific instances as a planned operation. On the other hand, the traffic-based module, denoted as Dynamic Energy Saving Routing Algorithm (DESRA), is activated by the arrival of each incoming traffic demand. Therefore, an accurate prediction of incoming traffic is not needed.

In essence, this approach finds the routes between network elements that minimize the number of active links used, being the links shared between data and control plane traffic (i.e., in-band mode). Therefore, control paths between controllers and switches (in both senses) and between controllers are also established.

Additionally, given the controllers placement in the network topology, an ideal distribution of switches between controllers is determined by our model. This is done in terms of energy efficiency, while considering as well the controllers capacity to support forwarding devices. Moreover, our energy-aware approach avoids the routing of additional traffic load through controllers. Therefore, admissible control paths do not traverse any other controller, but the source or target of the traffic and data plane communications cannot be routed through any of these devices.

The two main parts enclosed within the proposed energy-aware approach are described in more details in the following subsections.

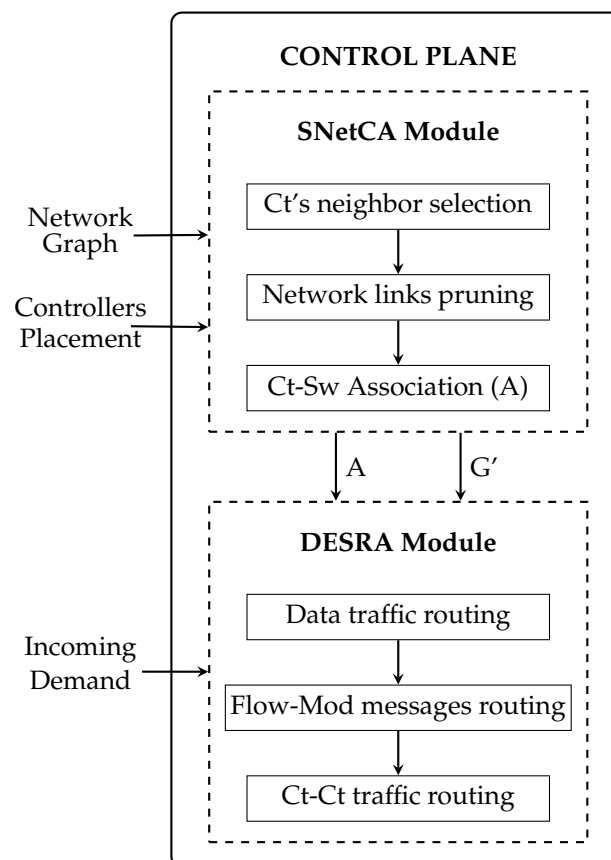


Figure 1. Illustrative diagram of the proposed approach. SNetCA, Static Network Configuration Algorithm; DESRA, Dynamic Energy Saving Routing Algorithm.

4.1. Static Network Configuration Algorithm

By considering the typical link redundancy of backbone networks, we design a Static Network Configuration Algorithm, denoted as SNetCA, which aims to prune as many links as possible in order to stress the importance of energy saving. Additionally, the most favorable switch-controller associations in terms of energy efficiency and load balance are determined in this stage.

The algorithm, described in the Algorithm 1 pseudo-code, is composed of three steps:

1. selecting one of the controller's neighbors, as the node that will remain connected to it in the outcome topology;
2. identifying the links that do not disconnect the graph to be put into sleep mode;

3. associating each node with one controller and computing the control path between them.

Algorithm 1 SNETCA

Require: $G = (V, E, C)$

Ensure: $G' = (V, E', C), A, P_{sc}$

```

1:  $N_c \leftarrow$  Set of neighbors of controller  $c \in C$ 
2:  $Sel\_N \leftarrow$  NULL ▷ Array of selected switch for each controller
3:  $G' \leftarrow G$ 
4: for  $c \in C$  do
5:    $B \leftarrow$  NULL ▷ Array of betweenness values
6:    $SP_c \leftarrow$  Set of shortest paths from controller  $c \in C$ 
7:   for  $n \in N_c$  do
8:     if  $n \in C$  then
9:       continue
10:    end if
11:     $B_n = 0$ 
12:    for  $p \in SP_c$  do
13:      if  $n \in p$  then
14:         $B_n = B_n + 1$ 
15:      end if
16:    end for
17:  end for
18:   $L \leftarrow N_{c\_Sorted}$  according to decreasing order of  $B$ 
19:  for  $s \in L$  do
20:    if  $c$  not already in  $Sel\_N$  then
21:       $Sel\_N = Sel\_N \cup (s, c)$ 
22:    end if
23:    Remove links  $(s, c)$  and  $(c, s)$  from  $G'$ 
24:  end for
25: end for
26: for  $i, j \in E'$  do
27:   if  $i \in C$  or  $j \in C$  then
28:     continue
29:   end if
30:    $G'' \leftarrow G'$ 
31:   Remove controllers  $c \in C$  from  $G''$ 
32:   Remove link  $i, j$  from  $G''$ 
33:   if  $G''$  remains strongly connected then
34:     Remove link  $i, j$  from  $G'$ 
35:   end if
36: end for
37: for  $s, c \in Sel\_N$  do
38:   if  $s$  is already associated with another  $c' \neq c$  then
39:     continue
40:   end if
41:   PATHSELECTOR( $s, c$ )
42:   Update  $P_{sc}, A, X, U$ 
43: end for
44: for the rest of  $s \in S$  do
45:   PATHSELECTOR( $s, C$ )
46:   Update  $P_{sc}, A, X, U$ 
47: end for

```

The input of Algorithm 1 is the network topology with controllers' placement, and its outputs are a pruned network with a reduced number of links, denoted as G' , an array keeping the controller-switch associations, denoted as A , and the control paths from each node to its controller, denoted as P_{sc} . Additionally, we use X to denote the set of active links $X \subseteq E$ and U to store the utilization of network links.

In the first step, the algorithm iterates over the set of network controllers in order to evaluate each one of its neighbors. The selection of one neighbor node for each controller is based on the betweenness centrality (B_n), which measures the intermediary role of a node in the network. In the proposed approach, we use a simplified version of this metric considering only the shortest paths from a controller to every switch.

In particular, after computing the shortest paths from one controller as the single source, the algorithm determines whether a neighbor node belongs to each path and increases the B_n associated with that node (Lines 7–17). For each controller, a list of neighbor devices, sorted in decreasing order of B_n , is stored in L . This list is used to identify the neighbor with the highest betweenness centrality. This will be the node that will remain connected to the controller in the resulting pruned topology. Thus, the controller considered in the current iteration and the selected neighbor, are stored in Sel_N .

For the remaining nodes in L , the links between them and the controller are removed from the resulting network graph. This means that they are put into sleep mode in the original graph. Notice that when a controller's neighbor is another controller, the link between them is not considered as a candidate to be pruned (Lines 8–10).

In the next step, the algorithm iterates over the set of directional links in the pruned network that do not have any controller as its extreme nodes. At each iteration, the algorithm attempts to increase the number of switched-off edges.

A new link is removed only when the resulting graph remains being strongly connected, i.e., at least one path exists between every pair of nodes in the network. To accomplish this, a temporal graph without any controller, denoted as G'' , is created. This graph is used to check the required connectivity between all the forwarding devices. After validating that the possibility to reach any node in the network is not affected, the considered link is removed from the resulting graph.

The last step of the algorithm is intended to determine a control path from each forwarding device to one controller. To achieve this goal, the algorithm starts evaluating the pairs of controller-switch already stored in Sel_N (Line 37). For each pair, the algorithm first determines whether the considered switch is still available. If this is the case, an admissible control path minimizing the number of active links is computed using the method PATHSELECTOR described in Algorithm 2, which will be further explained below. As stated previously, admissible control paths do not pass through any other controller that does not act as the source or target of the considered traffic. The remaining forwarding devices are then considered. Notice that in this case, the algorithm takes into account the control paths to all controllers in the network. Precisely, the path computed by the PATHSELECTOR in this step defines the controller for the rest of forwarding devices.

Using this initial control plane configuration, switches send to the controller packet_in requests when a new traffic flow arrives, as well as statistics and failure notifications. Consequently, there is an initial set of active links in the network before the ingress of traffic flows, as well as some link utilization.

The PATHSELECTOR method, described in Algorithm 2, performs the energy-aware path selection. In essence, this function is used to select the best admissible route between a pair of nodes, aiming to minimize the total number of active links in the network. The key idea of this function is to perform a low-complexity greedy evaluation between all the admissible paths to select the most suitable route in terms of energy-efficiency, while guaranteeing a balanced load of switches between controllers and capacity constraint of links. Since this method works over the pruned network with a reduced number of links (i.e., G'), the set of admissible paths considered is significantly smaller than in the original topology, and the solution can be found after fewer iterations. When this function is asked to find the path between each forwarding device and one controller (i.e., using the set of controllers as the traffic

destination), the controller load is considered (Line 4–8). In addition, a candidate path is selected as long as it has sufficient link capacity to route the required traffic volume.

Algorithm 2 PATHSELECTOR(a, b)

```

1:  $L \leftarrow \infty$ 
2:  $SeP \leftarrow None$ 
3: for  $p \in \text{Get\_All\_Admissible\_Paths}(G', a, b)$  do
4:   if  $b = C$  then
5:     if  $p$  is to an already loaded controller then
6:       continue
7:     end if
8:   end if
9:    $off \leftarrow$  number of links in  $p$  that are not in  $X$ 
10:  if  $off < L$  and  $p$  has sufficient bandwidth then
11:     $L \leftarrow off$ 
12:     $SeP \leftarrow p$ 
13:  end if
14: end for

```

4.2. Dynamic Energy Saving Routing Algorithm

When a new traffic demand arrives, a routing request is sent from the input node to its associated controller using the path between both devices previously computed during the static network configuration phase. Based on its global knowledge of the network topology, this controller calculates the required data path minimizing the number of links that need to be activated for this connection request and creates the flow forwarding rules.

The proposed dynamic energy-aware routing is shown in Algorithm 3. For an incoming demand d from source s_d to destination t_d , the algorithm starts storing in Ct_1 the controller associated with the source node. This controller is the main responsible of managing this traffic request.

Using the PATHSELECTOR method, the most favorable admissible data path in terms of energy consumption is computed. This is done considering that admissible paths do not go through any controller in the network. Then, a loop is used to establish the required control plane communications for each node along this path.

Algorithm 3 DESRA

Require: G', A, d incoming traffic request

Ensure: P_{ss}, P_{cs}, P_{cc} data and control paths, X active links, U links utilization

```

1:  $Ct_1 \leftarrow A[s_d]$ 
2:  $p = \text{PATHSELECTOR}(s_d, t_d)$ 
3: Update  $P_{ss}, X, U$ 
4: for  $n \in p$  do
5:    $Ct \leftarrow A[n]$ 
6:    $\text{PATHSELECTOR}(Ct, n)$ 
7:   Update  $P_{cs}, X, U$ 
8:   if  $Ct_1 \neq Ct$  then
9:      $\text{PATHSELECTOR}(Ct_1, Ct)$ 
10:    Update  $P_{cc}, X, U$ 
11:   end if
12: end for

```

After identifying the controller associated with each node in the data path, a control path is computed between them. These paths are used to set the flow forwarding rules in each switch using the flow_mod messages. Given the multi-domain scenario considered, the nodes traversed by the data traffic may be associated with different controllers. When a node is not associated with C_{t1} , an additional control message is sent from this controller to the other, in order to inform the second controller of the flow forwarding rule that needs to be installed in one of its managed nodes.

4.3. Complexity Analysis

Considering that the computation of the shortest paths from each single controller is done in $O(N \log N)$ using the Dijkstra algorithm, the selection of one controller's neighbor in the first step of Algorithm 1 has a worst run-time complexity equal to $O(N \log N + N^2 + S \log S + S)$, where N is the number of total network nodes and S is the number of forwarding devices. Given that this operation is performed C times, being C the number of controllers, the complexity of Step 1 becomes $O(N^2 C)$. Pruning as many links as possible without disconnecting the network graph during the second step has a complexity equal to $O(E(N + E))$, where E is the number of links and $O(N + E)$ refers to the connectivity checking process [36]. The complexity of last step is determined by $O(SM)$, where M indicates the worst-case complexity of Algorithm 2. It should be noticed that after pruning the network topology, the number of admissible paths and the consequent computation complexity of this method are significantly reduced with respect to the original graph. Therefore, the SNetCA complexity can be expressed as $O(N^2 C)$.

The overall complexity of the proposed dynamic routing solution, mixing together both Algorithms 2 and 3, is $O(SM^2)$ since the maximum length of a data path is precisely the number of forwarding devices given the routing restriction avoiding data traffic through network controllers. Given that M is usually a small number after pruning the network topology and will not grow rapidly along with the network topology size, running the algorithm upon each flow request is a reasonable approach.

5. Simulations and Results

In this section, we describe the evaluation of our energy-aware approach and analyze the achieved results. We used the linear programming solver Gurobi Optimizer [37] for the ILP model. Meanwhile, the proposed control framework described in Section 4 was implemented using the programming language Python to develop the heuristic algorithms. A computer with 3.30 GHz Intel Core i7 and 16GB RAM was used to carry out all computations. To conduct our simulations, we use real-world network topologies and traffic demands collected from SNDlib [38]. Each router in the network was considered as an SDN node or as a tentative controller location. Since the topologies used in our experiments are backbone networks, for the sake of simplicity and without loss of generality, we opted to compute the communications delay as the propagation latency.

Specifically, we use three of the most link-redundant network topologies in SNDlib in order to assess the effectiveness of the proposed scheme. The mentioned topologies are New York, Geant and Norway, and their main characteristics are summarized in Table 2.

Table 2. Main characteristics of considered topologies.

Topology	V	E	D	Average Node Degree	Total Amount of Traffic Demand (Mbps)	Mean Traffic Demand (Mbps)
New York	16	98	240	6.12	1774.0	7.39
Geant	22	72	430	3.27	42565.53	98.99
Norway	27	102	702	3.78	5348.0	7.62

Table 2 presents, apart from the topological properties, a general description of the provided traffic load for each studied topology. Specifically, for New York and Norway, we use the default traffic

matrices provided, while for Geant, we select the subset of available traffic matrices measured at 00:00 on 5 May 2005.

As we are considering an in-band scenario, after placing the controllers in each simulation instance, we have deleted the data traffic demands from/to them. The network load being an important parameter that impacts the efficiency of any energy-aware solution, we also compute the average network load ρ , defined as follows:

$$\rho = \frac{\sum_{(i,j) \in E} \sum_{f \in D} t_{i,j}^f b_f}{\sum_{(i,j) \in E} x_{i,j} c_{i,j}} \quad (8)$$

Where the traffic flowing on each link was obtained routing each traffic demand using the shortest path. Results from this early analysis confirm that the real scenarios considered present a low network traffic load. More in detail, the average network loads of the three network topologies are less than 0.05, New York links being particularly lightly loaded (i.e., $\rho < 0.01$). This behavior is typical in real backbone topologies where capacity planning strategies aim to always ensure a significant over-provisioning in core links relative to the offered average load. This is done as an attempt to avoid the congestion in case of peak load and to allow the fulfillment of Service Level Agreement (SLA) requirements. Intuitively, low loaded networks are suitable scenarios for deploying energy-aware solutions that concentrate traffic and turn off unused network links. Therefore, this analysis suggests that substantial energy savings are possible. Moreover, we can deduce that potential energy benefits will be more limited by the required connectivity and topological properties than by the network traffic load.

An average rate of 1.7 Mbps was assumed for the control traffic [39]. Considering a homogeneous scenario, where all controllers have the same computational and networking capabilities, in our simulations, we set the maximum number of forwarding devices that can be associated with each controller as follows:

$$R_c = \left\lceil \frac{|S|}{|C|} \right\rceil \quad \forall c \in C \quad (9)$$

In this way, switches are evenly distributed, and the load of switches is balanced among controllers.

To analyze the performance of our energy-aware approach, five groups of evaluations are presented varying the number of controllers in the considered topologies. In the first part, the performance of our heuristic algorithms with respect to the optimal model is analyzed in order to fix an upper bound for the energy saving capabilities of proposed solutions. In the second part, the potential of SNetCA to prune a network topology is investigated. Then, we compare our solution with the four strategies of another energy-aware method proposed in the literature. Next, we analyze the impact of our model on crucial network performance metrics, such as latency, link utilization and TCAM occupation. Finally, we provide an initial analysis exploring the existing trade-off between energy savings and network resilience.

5.1. Optimal vs. Heuristic Solutions

To assess the performance of proposed heuristic algorithms against the optimal solution achieved by the ILP model, we show their behaviors in Figure 2 using the New York and Geant topologies.

In both networks, we consider all admissible controller placements and compute the average energy savings for different numbers of controllers. Notice that a controller placement is admissible when the assumptions established in this proposal to avoid the routing of additional traffic load through network controllers can be kept (i.e., the network graph without any controller is strongly connected). We compute energy savings as the proportion of links in sleep mode with respect to the total amount of network links.

As shown in Figure 2, the energy savings reached by the optimization model are up to 64% and 83% in Geant and New York topologies, respectively. On the other hand, the heuristic strategy (denoted in the figure as SNetCA/DESRA) allows attaining close-to-optimal energy savings, with a maximum gap of 9%.

We can also see that while the number of controllers grows energy savings decrease. This is due to the fact that in our approach data plane communications are not routed through network controllers. Therefore, as the number of network controllers is increased, a higher number of links (i.e., links directly connected to the controllers) is likely to be used to route control traffic, but not for data plane communications. A similar decreasing behavior can be noticed in the gap between optimal and heuristic results since, as the amount of network controllers grows, a lesser number of feasible possibilities can be considered by the linear solver.

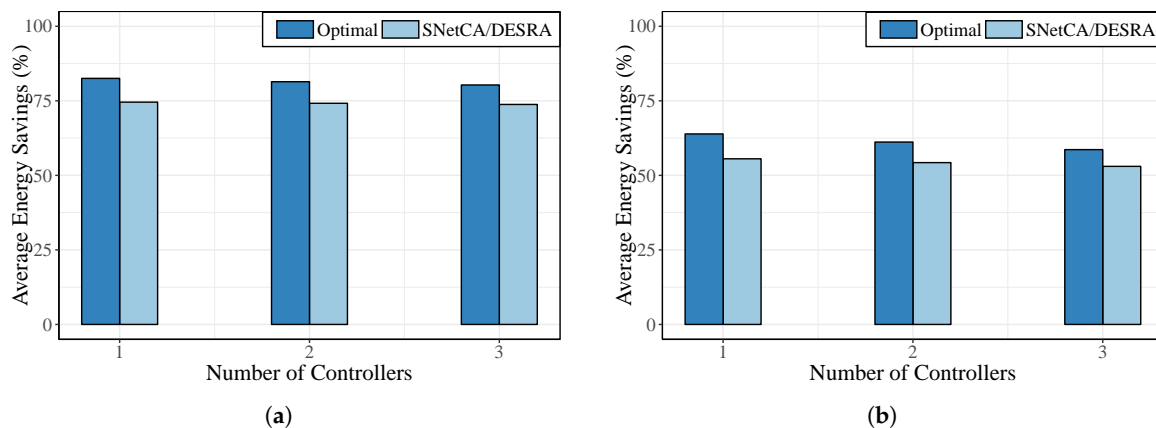


Figure 2. Average energy savings comparison between optimal and heuristic solutions for different numbers of controllers in: (a) New York topology and (b) Geant topology.

Table 3 shows the average execution times of the optimal model and heuristic algorithms in the three network topologies considered in our experimental simulations. While the SNetCA column shows the execution times required to prune the network, the two others contains how many seconds are spent by each routing approach to compute all the required control and data paths according to the incoming traffic. As is shown, in both approaches, computation times mostly tend to decrease as the number of controllers grows. The reason for this is that in these energy-aware approaches, the routing of additional traffic load through network controllers is avoided. Therefore, an increase of network controllers implies that fewer alternate paths between each pair of nodes are considered in the simulation.

Table 3. Average execution time (s) using different numbers of controllers on real topologies.

Topology	C	Optimal	SNetCA	DESRA
New York	1	172.76669	0.06118748	0.7708125
	2	191.75141	0.05415126	0.6329832
	3	78.89208	0.04126375	0.5140916
Geant	1	114.88002	0.05245455	2.4348636
	2	91.40695	0.05004147	1.9789217
	3	49.54307	0.04750903	1.5744784
Norway	1	9946.57420	0.10151852	8.9833704
	2	5227.32024	0.09847852	7.2682321

Although in all cases, the proposed strategy outperforms the optimal model in terms of computation time, a higher improvement is achieved as the network size grows. For instance, in the Norway topology, the processing times required by the optimal model increase dramatically. The ILP model in this topology can take more than two hours on average to find a solution, which is a great limitation in current networking environments. Meanwhile, it is always less than 10 s for the heuristic strategy, i.e., almost a three order of magnitude improvement. This comparison validates the improvements achieved by the heuristic proposal in terms of computation time and clearly justifies its necessity.

For the Norway topology, only the case of having one and two controllers is shown. The required convergence time for solving the exact model beyond this limit became unfeasible, since the number of possible combinations of nodes as controllers' placement grows exponentially.

5.2. SNetCA Performance

In order to evaluate the effectiveness of the proposed topology-based module, Figure 3 shows an example of the performance of SNetCA on the Norway topology, considering two network controllers placed at nodes denoted as 1 and 2 and emphasized with a different color in the figure. The distribution of switches between controllers is depicted through the use of labels in each node, indicating the controller number to which the node is associated. Here, we focus our attention on the Norway topology, but similar results have been obtained for the two remaining topologies considered.

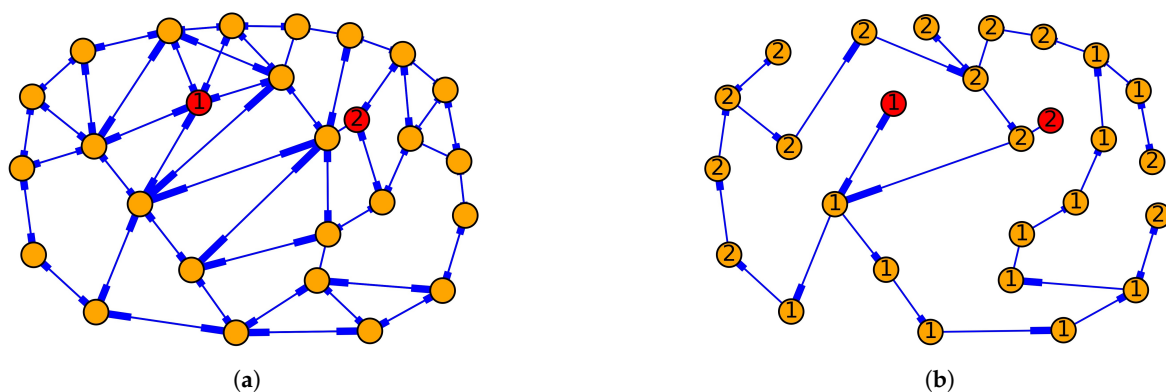


Figure 3. (a) Original Norway graph. (b) Resulting Norway graph after applying SNetCA.

A comparison between the original network and the resulting graph illustrated in Figure 3a,b respectively, shows a difference of 67 edges, which represents more than 65% of total network links. These links are pruned by our algorithm guaranteeing that the resulting graph remains strongly connected and avoiding additional traffic load through network controllers.

Additionally, as a result of applying SNetCA on the Norway topology, switches are distributed between controllers minimizing the number of required active links and ensuring a balanced controllers load. For instance, 12 switches are associated with Controller 1, while the remaining 13 are managed by Controller 2.

To provide a more general perspective, Figure 4 shows, for the three considered topologies, the average number of links pruned by SNetCA, which contribute directly to the energy efficiency achieved by this proposal. In this analysis, we also consider all the admissible placements of one to three controllers.

As is shown, an important number of links is pruned in all the topologies considered, but the highest energy savings are achieved in the New York topology. The reason for this is that New York has much more link redundancy than Geant and Norway. Therefore, a higher number of links can be

pruned while guaranteeing the network connectivity. In general, the more redundant the network, the higher number of links are put into sleep mode applying this strategy.

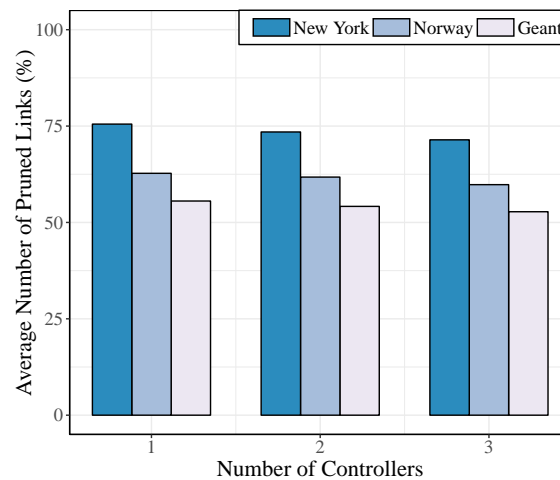


Figure 4. Average number of pruned links in the three topologies varying the number of controllers.

5.3. Assessment of Energy Saving Potential

SGH [25] is an energy-aware routing solution based on knowledge of the incoming traffic requests. This proposal selects, using a pre-calculated set of shortest paths, the most suitable path in terms of energy savings to allocate each traffic demand, as long as it has enough capacity. To do so, traffic demands are ordered according to the following four different strategies:

- Node pairs with the shortest Shortest Path First (SPF)
- Node pairs with Longest shortest Paths First (LPF)
- Node pairs with Smallest Demand First (SDF)
- Node pairs with Highest Demand First (HDF).

Figure 5 shows the average performance in terms of energy savings of the two-module based strategy SNetCA/DESRA with respect to the four different versions of SGH in the three real topologies analyzed for different amounts of controllers.

As we are considering an in-band SDN with multiple controllers, the traffic matrix provided to SGH includes, together with the data demands, a control traffic flow for each pair of associated controller-switch and for each pair of controllers in the network. In this way, required control plane communications are also established by SGH. This is done taking into account the distribution of switches between controllers obtained from SNetCA. Additionally, the routing restrictions established in this proposal to avoid additional traffic load through network controllers are considered in the computation of the pre-calculated shortest paths. On the other hand, given that DESRA is an online routing strategy, the connection requests are sequentially allocated as they appear in the considered traffic matrix.

In Figure 5a–c, we can see that SNetCA/DESRA performs better than the remaining algorithms. While similar results are obtained by the four different versions of SGH, the proposed strategy achieves notable improvements in the three considered topologies. For instance, differences up to 35%, 29% and 45% of energy savings are reached in the New York, Geant and Norway topologies, respectively. Such significant differences are mostly due to the operation of SNetCA before the traffic arrival, which is able to prune a great number of links without affecting the network capabilities to manage the incoming requests. Additionally, the routing decisions performed by SGH are limited to a predefined number of pre-computed shortest paths and fail to extensively exploit the energy saving potential of each topology. Therefore, even without a prior arrangement of demands based on accurate knowledge

about the incoming traffic, the proposed scheme in this paper puts to sleep mode a higher percentage of links, being able to save substantially more energy.

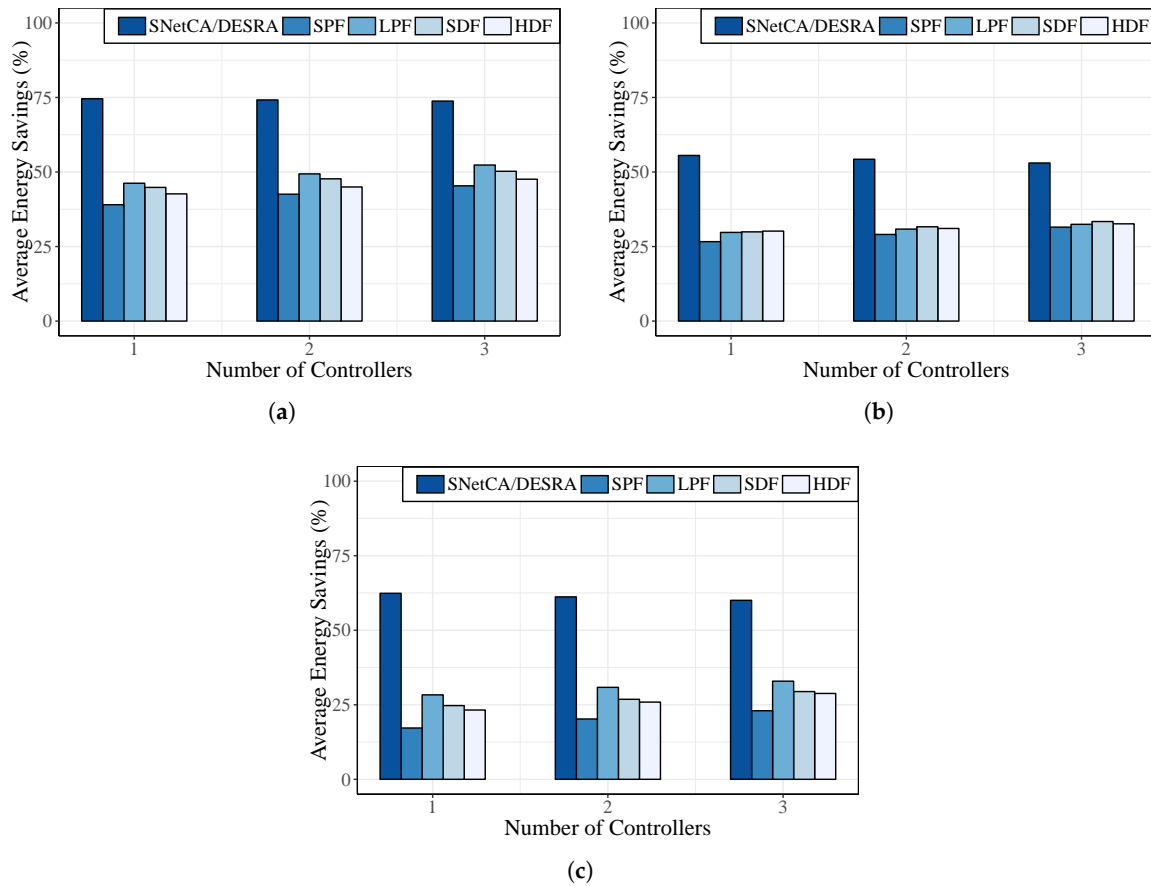


Figure 5. Average energy savings comparison between proposed strategy and SGH solutions for different numbers of controllers in the: (a) New York topology; (b) Geant topology; and (c) Norway topology.

5.4. Impact on Network Performance

It is to be emphasized that in our energy-aware approach, Quality of Service (QoS) constraints and performance metric boundaries are not considered. This is not a limitation, but a choice; since we intend to measure the impact of our proposal on the network performance metrics as a trade-off with the energy saving improvements. In fact, we are presenting an effective and easy to implement green routing mechanism that emphasizes the importance of energy efficiency in the operation of current data networks.

In order to assess the impact of our energy-aware approach on the network performance, we adapt two well-known state-of-the-art routing algorithms: Shortest Path Routing (SPR) and Load Balancing (LB) for their use in the considered in-band SDN environment. Additionally, the rule space being a significant issue of concern in SDN, we include in this analysis an algorithm balancing the number of rules installed in each forwarding device, denoted here as TCAM Occupation Balancing (TOB). In general, these algorithms are greedy heuristics that follow the procedure described in Algorithm 4.

More precisely, these algorithms will evaluate every candidate admissible path and find the one prioritizing some performance metric such as: traffic latency, link utilization or TCAM occupation. According to the approach used, the selected path will be:

- SPR** The one minimizing the propagation latency
LB The one minimizing the maximum link utilization
TOB The one minimizing the maximum TCAM occupation

It is clear that, in each of these metrics, the corresponding baseline algorithm will have better performance than our energy-aware approach. However, the purpose of this evaluation is to use these algorithms as a reference point to illustrate the energy-aware solution impact on network performance. All of them follow the assumptions established in this proposal to avoid the routing of additional traffic load through network controllers. Similarly, SNetCA is still used to determine the distribution of switches between controllers. Although we may focus our attention on some specific network for the different performance metrics, the derived conclusions are general and independent of the specific considered topology. Thus, they hold whichever network is examined.

Algorithm 4 GREEDY_BASELINES

Require: G, A, D

```

1: for  $d \in D$  do
2:    $p_d = \text{Find\_Best\_Data\_Path}$ 
3:   Update network metrics
4:   for  $s \in p_d$  do
5:     if no control path is already established between  $s$  and its controller in  $A$  then
6:        $p_c = \text{Find\_Best\_Control\_Path}$ 
7:       Update network metrics
8:     end if
9:   end for
10: end for
  
```

5.4.1. Traffic Latency

In the first set of simulations, we analyze how the data and control paths latency is affected by routing decisions made.

To evaluate the impact of our algorithm on control path delay, we collect, for each traffic demand, the length of its associated control paths and the corresponding shortest paths. The notation Maximum Over-length is used to denote the maximum number of additional hops (with respect to the length of the corresponding shortest path) among the routing solutions for the required control paths. For instance, when this value is equal to zero, it means that every control traffic is routed using exactly the shortest path.

Figures 6–8 show this behavior for the three studied topologies considering all possible placements for different amounts of controllers.

As is shown, in all cases when the number of controllers grows, the control traffic is routed using a larger number of hops for a higher fraction of demands. Norway being the largest one in terms of network size (number of nodes and links), control paths in this topology are increased by a higher number of hops.

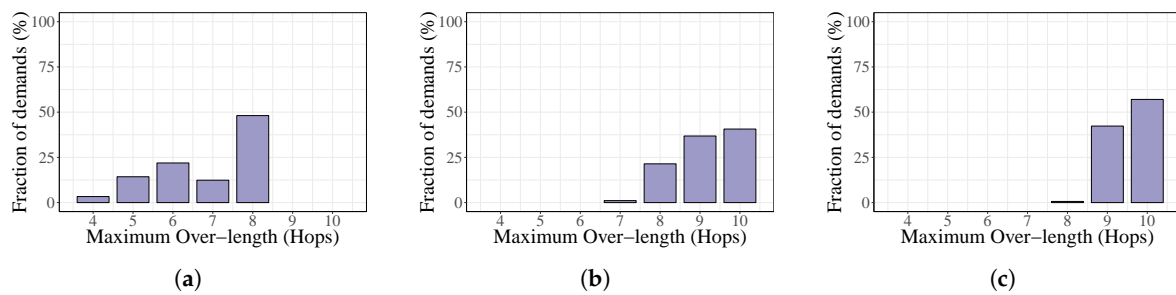


Figure 6. Distribution of maximum control traffic over-length in the New York topology for: (a) 1 controller, (b) 2 controllers and (c) 3 controllers.

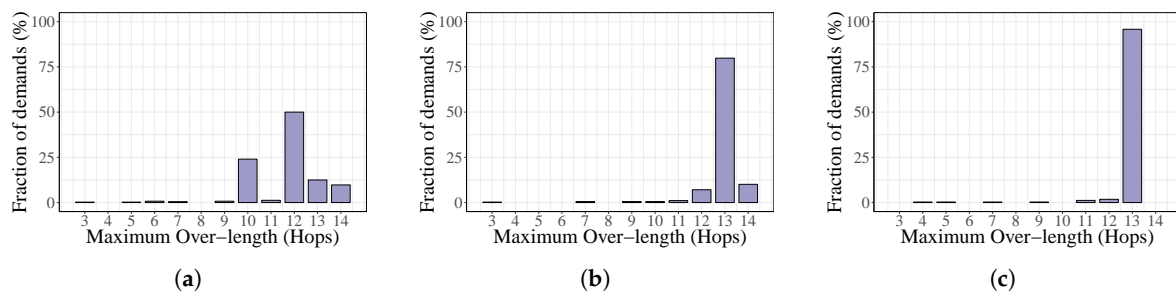


Figure 7. Distribution of maximum control traffic over-length in the Geant topology for: (a) 1 controller, (b) 2 controllers and (c) 3 controllers.

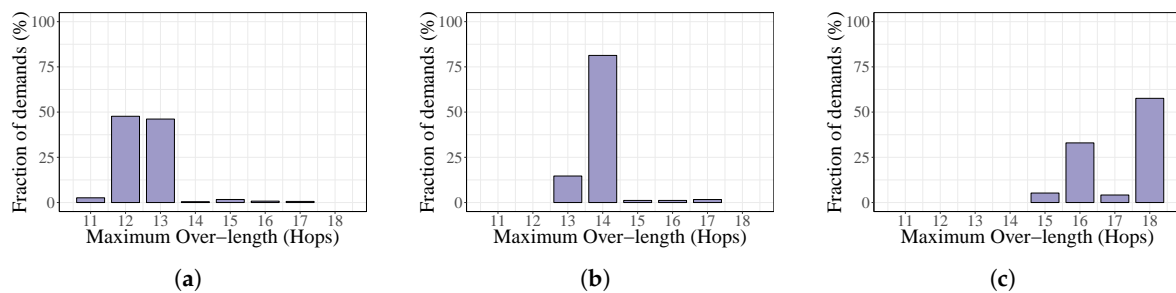


Figure 8. Distribution of maximum control traffic over-length in the Norway topology for: (a) 1 controller, (b) 2 controllers and (c) 3 controllers.

To take a closer look at the data plane, we draw in Figure 9 the Cumulative Distribution Function (CDF) of data paths' latency for the three topologies considering all possible locations of one to three controllers.

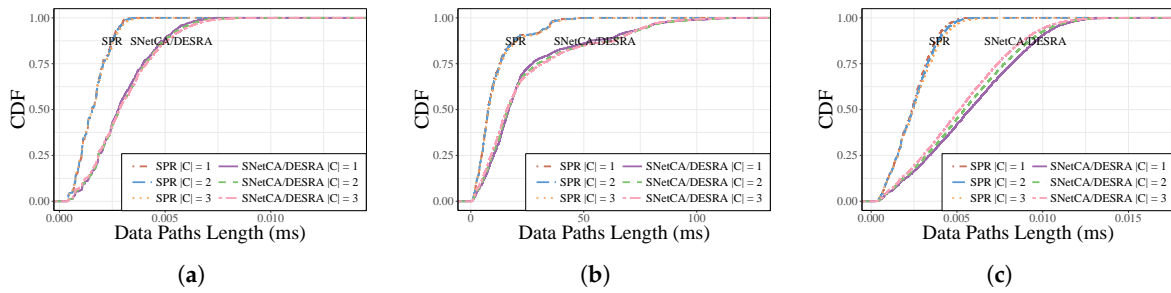


Figure 9. Cumulative distribution function of data paths latency varying the amount of controllers in the: (a) New York topology; (b) Geant topology; and (c) Norway topology.

As shown in Figure 9a–c, the CDFs of data paths' latencies for different amounts of controllers are quite similar. However, we can see that under the energy-aware routing, data path delay is affected since larger data paths are used in order to minimize the number of active links. For instance, in Figure 9b, only 87% of data paths exhibit delays lower than 50 ms, meanwhile all control paths in the SPR case are under this value. This performance degradation is less critical in the two other topologies, which are deployed in smaller geographic areas compared to Geant. In general, the larger the network (in terms of geographic length), the more increase in latency is incurred.

Despite the presented latency degradations with respect to SPR, the solution performances observed in Figure 9a,c are suitable for supporting latency critical services in 5G networks demanding end-to-end delays lower than 10 ms [40], such as robotics and telepresence, virtual reality, health care, among others. Likewise, less demanding applications and use case scenarios to be addressed in 5G networks, such as intelligent transport systems and smart grid, with latency requirements up to 100 ms [40], could be conceived of and deployed in Geant-like environments.

These latency degradations confirm that, according to traffic requirements, specific performance bounds may be required. In order to ensure the suitability of the proposed algorithm for delay-critical services, we include in this analysis the evaluation of a delay-constrained version. To do so, we now restrict the number of links that can be initially pruned by SNetCA. Specifically, network links belonging to the shortest path between any pair of nodes are not removed during the static network configuration phase. In this way, the shortest path will always be available if it is needed for allocating the incoming traffic with DESRA. In both modules, the path latency is now taken into account for the PATHSELECTOR mechanism in order to satisfy given delay bounds.

Figures 10 and 11 show the performance of the delay-constrained version considering data and control paths latency bounded by the factor r . This latency threshold is used to denote the relation among the delay requirement and the shortest path propagation latency for every established path. For instance, when $r = 2$ every path latency will be, at most, twice that of the shortest path.

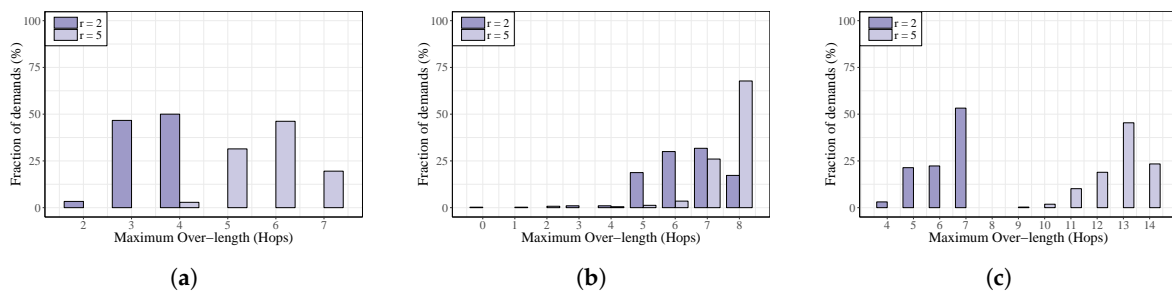


Figure 10. Distribution of maximum control traffic over-length for one controller under delay constraint in the: (a) New York topology; (b) Geant topology; and (c) Norway topology.

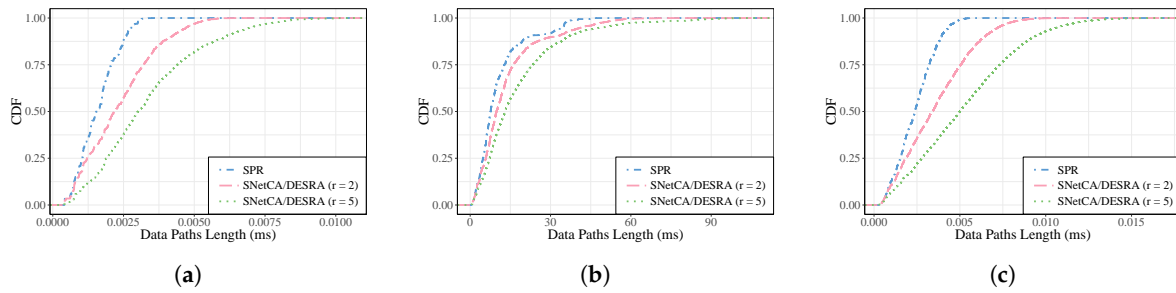


Figure 11. Cumulative distribution function of data paths latency for one controller under delay constraint in the: (a) New York topology; (b) Geant topology; and (c) Norway topology.

As is shown in Figure 10 under the delay-constrained approach, the control traffic can be routed incurring smaller over-lengths with respect to the performance-agnostic approach. Similarly, Figure 11 shows that the data paths latency are less compromised when more restrictive possibilities in terms of latency ($r = 2$ and 5) are considered. In general, the more restrictive is the latency bound used, the better the performance in terms of delay can be achieved.

On the other hand, considering a latency restriction during paths' selection will negatively impact the potential improvements in terms of energy efficiency. To validate this affirmation, the energy savings achieved by the delay-constrained version using $r = 2$ are shown in Table 4. In this table, optimal and heuristic values are depicted as well as the energy saving differences between the performance-agnostic results shown in Figure 2 and the delay-constrained version.

Table 4. Energy savings with paths' delay constrained to $r = 2$.

Topology	C	Optimal	Delay-Constrained Heuristic	Difference
New York	1	58.16327	39.92347	34.63010
	2	61.40456	41.11645	33.05608
	3	62.37759	42.04418	31.74105
Geant	1	29.67172	17.67677	37.87879
	2	33.81976	18.65719	35.61828
	3	34.72549	19.85249	33.16422

Optimal values were obtained after adding to the ILP model presented in Section 3 the following constraint, where $d_{i,j}$ and L_f denote the link propagation delay and the traffic maximum latency bound, respectively.

$$\sum_{(i,j) \in E} t_{i,j}^f d_{i,j} \leq L_f \quad \forall f \in F \quad (10)$$

Significant reductions in terms of achieved energy saving values can be seen in Table 4 for the both considered topologies as a consequence of restricting the path selection to meet the delay requirements. However, even considering such restricted thresholds, where no traffic can be routed using a path longer than twice the shortest path length (in terms of propagation delay), substantial energy savings can still be reached by the proposed heuristic strategy.

5.4.2. Links Utilization

The selection of routing paths minimizing the energy consumption has a direct influence on the traffic load of all the network links. To better showcase this situation, we use the New York and Geant topologies and the LB algorithm. Figure 12 provides the CDF of link utilization under both algorithms considering all possible locations of one to three controllers in both topologies.

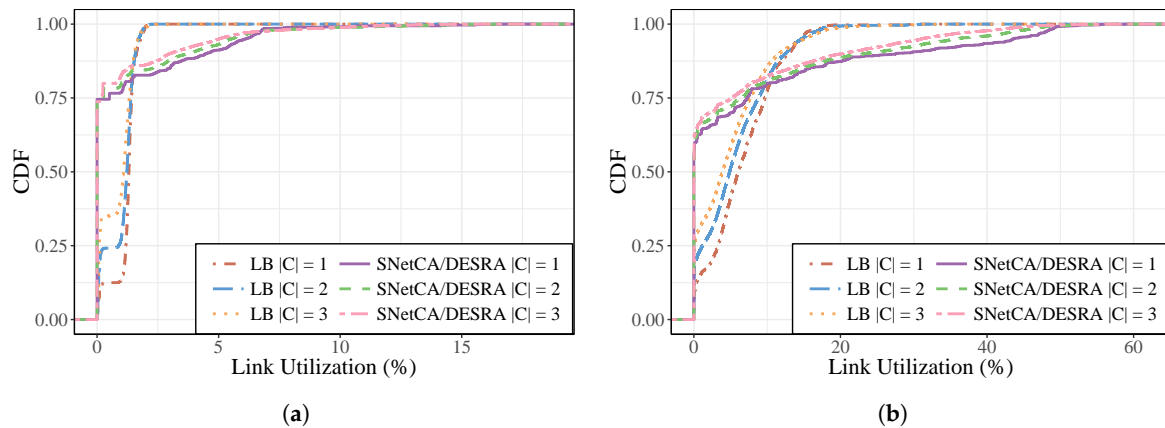


Figure 12. Cumulative distribution function of link utilization varying the amount of controllers in the: (a) New York topology; and (b) Geant topology.

As expected, the fairness of traffic distribution is altered by the energy-aware routing, since under this approach, traffic is concentrated in a fewer number of links. Therefore, there is a subset of active links that is more overloaded than the others. For instance, in Figure 12b, the link utilization of some links in Geant is increased to more than twice the value achieved by the LB algorithm. Nevertheless, even in the more loaded cases, the link utilization in this topology is under 60%. A less concerning situation can be observed in Figure 12a since the given traffic load in the New York topology is very low.

5.4.3. TCAM Occupation

Intuitively, energy-aware routing would affect the allocation of flow rules, a practical constraint in OpenFlow devices, given that traffic flows are redirected to minimize the number of active links.

In Figure 13, we evaluate the impact of our approach on TCAM occupation with respect to the TOB algorithm using the Geant and Norway topologies and all possible locations of two network controllers.

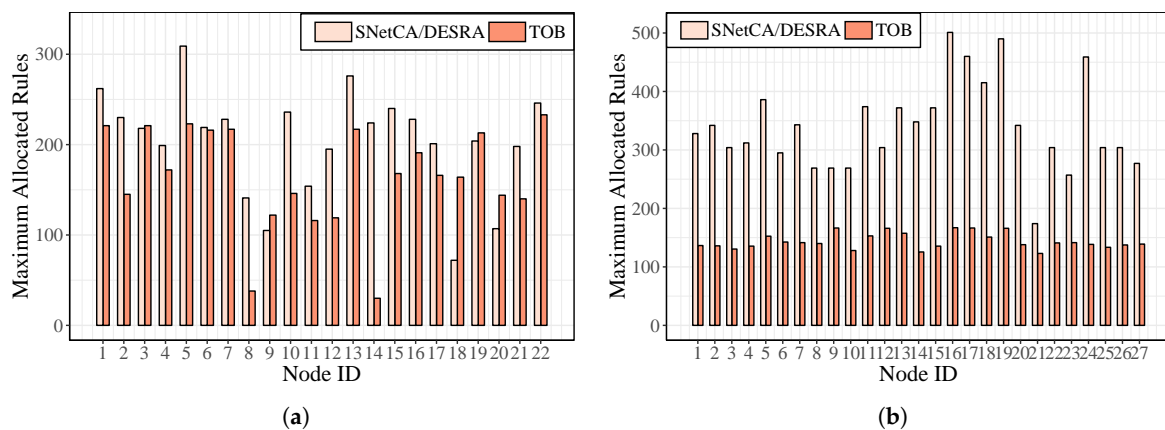


Figure 13. Average TCAM occupation with $|C| = 2$ in the: (a) Geant topology; and (b) Norway topology.

As expected, the number of installed rules is raised by the energy-aware routing since new paths are more likely to be allocated over the most loaded devices instead of activating a new one. For instance, this increase is observed in all network devices in Norway topology and in 17 out of 22 nodes in Geant topology, in some cases being more than twice the value obtained by the TOB algorithm. However, an accurate explanation about these numbers is difficult since the energy-aware

proposal does not care at all about the use of rule capacity. Furthermore, the SNetCA/DESRA performance in both topologies is still physically acceptable considering that a routing table can support around a few thousands of rules [23].

5.4.4. Energy Savings

Finally, to get a sense of the other side of the trade-off between energy efficiency and network performance, Figure 14 shows the average energy performance of all the considered routing models in the New York and Norway topologies for the case of one centralized controller in the network.

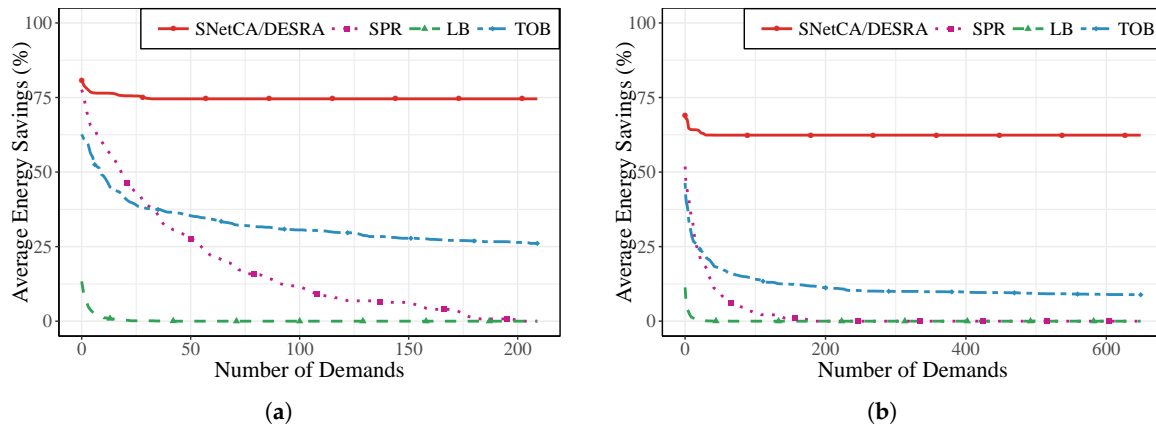


Figure 14. Average energy savings with $|C| = 1$ in the: (a) New York topology; and (b) Norway topology.

As expected, in all cases, energy savings decrease while the number of demands grows. This is due to the fact that new paths need to be established to accommodate such traffic. The flat tendency of energy savings achieved by the power-aware solution despite the increase in allocated demands is possible given the low network load discussed at the beginning of this section. Moreover, the proposed strategy greatly outperforms SPR, LB and TOB in terms of energy saving. In general, SNetCA/DESRA is able to achieve significant energy savings, but bigger improvements with respect to the other approaches are reached when the traffic grows.

5.5. Resilience Concerns

While SNetCA allows important gains in terms of energy efficiency by pruning as many links as possible and leaving available only the minimum number of links needed to support the incoming traffic, the remaining subnetwork is more vulnerable to resource failures and sudden traffic bursts. To ensure an adequate network reliability while keeping low energy consumption, resilience constraints should be taken into account.

Given that the largest Laplacian eigenvalue of a graph is a widely-accepted metric to assess network robustness with respect to link and node removals [41], we use it (referred to here as λ_{max}) to control the resulting resilience after applying SNetCA. Several papers about graph theory [42,43] sustain that a network is more resilient the higher the largest eigenvalue of its Laplacian matrix. In general, these networks are more robust since they have small diameters, higher numbers of nodes and link disjoint paths and are likely to expand faster.

In order to provide an initial investigation exploring the existing trade-off between energy savings and network reliability, in this analysis, we evaluate a resilience-constrained version of SNetCA, which enables us to limit the admissible λ_{max} reduction rate due to the link removal process. Additionally, to improve the redundancy for the control paths, the amount of neighbors that will remain connected to the controller is relaxed using different bounds (from one to the controller degree).

Table 5 presents some of the obtained results using the three considered topologies with one centralized controller. The presented results validate the impact of varying these two criteria (i.e., controller connectivity and λ_{max} reduction rate) on energy savings and network robustness.

Table 5. Average performance under connectivity and resilience constraints. ES, Energy Saving.

C_{con}	δ	New York		Geant		Norway	
		ES (%)	λ_{max}	ES (%)	λ_{max}	ES (%)	λ_{max}
1	0	54.08163	11.49839	29.79798	9.516197	51.67030	8.056087
	1	74.55357	10.96566	55.55556	7.601231	62.38199	6.764033
2	0	52.04082	11.62907	27.02020	9.642021	49.70951	8.094395
	1	72.32143	11.15838	52.71465	7.783621	60.34858	6.846996
3	0	50.27211	11.72088	31.01852	9.608442	47.88235	8.104170
	1	70.13605	11.36235	49.07407	8.046415	58.23529	6.907458
4	0	48.23129	11.84817	37.26852	9.462972	46.90799	8.087313
	1	67.95918	11.49919	46.99074	7.765802	55.73152	6.966380
5	0	46.35569	12.00953	33.33333	9.478610	45.75163	8.067831
	1	65.74344	11.62032	44.16667	7.988221	54.24837	7.046465
6	0	44.89796	11.93040	27.31481	9.424882	44.11765	8.146692
	1	63.77551	11.51884	42.59259	8.024959	51.47059	7.572282

While the controller connectivity (denoted as C_{con}) is bounded by a number of neighbors between one and the controller degree, the allowed λ_{max} reduction rate (denoted as δ) was normalized according to the following expression, where the term SNetCA' denotes the resilience-constrained version:

$$\delta = \frac{\lambda_{max}(\text{Original}) - \lambda_{max}(\text{SNetCA}')}{\lambda_{max}(\text{Original})} \quad (11)$$

As is shown in Table 5 when $C_{con} = 1$ and $\delta = 1$, the resilience-constrained version of SNetCA behaves exactly as the resilience-agnostic one. However, less energy can be saved as more restrictive values of C_{con} and δ are imposed, since each of these elements determines that a fewer number of links could be put into sleep mode by the pruning function of SNetCA.

Inversely, when no resilience degradation is allowed (i.e., $\delta = 0$), higher values of λ_{max} are obtained. More importantly, we can observe that considerable energy savings can still be reached by the proposed heuristic strategy, while ensuring the original network reliability.

Although higher values of λ_{max} are expected when the bound of nodes that will remain connected to the controller is increased, this is not always the behavior that can be appreciated in the table for the three network scenarios considered. The reason for this is that every network node, which is considered in one particular instance of the simulation as the controller, does not have the same node degree. Thus, for each considered possibility of C_{con} , different amounts of nodes are part of the presented average values.

We stress that the impact of switching off network links on SDN reliability is even more critical for networks with in-band control traffic where any link/node failure will affect not only data plane communications, but the connection with the controller. Hence, a further analysis about this crucial issue will be done in follow-on work.

6. Conclusions

In this work, 5G is considered as the broad context where several recent technologies, such as SDN, will be leveraged to fulfil functional and performance requirements of heterogeneous services and devices. At the same time, the unprecedented number of devices together with the rate-demanding services and the high energy efficiency expected in 5G networks are decisive drivers motivating

this research. Undoubtedly, to gain insight into the existing trade-off between energy efficiency and network performance is a valuable contribution on the road to 5G. Precisely, this is the aim of this paper, conceived to highlight this crucial issue and to tackle the energy consumption problem in 5G backbone networks.

More in details, in this paper, we proposed an energy-aware strategy that reduces the number of active links used to handle the incoming traffic suitable for SDN environments with in-band control traffic and multiple controllers. To achieve such a goal, we first provided a link-based formulation of the optimization problem, integrating the routing requirements for data and control traffic. Given the overall complexity of the exact model in large-scale topologies and the fact that new devices with less predictable traffic patterns will join the 5G broadband network, a heuristic hybrid approach is conceived, comprising two algorithms: a static network configuration and a dynamic energy-aware routing. In this way, the number of links to be considered in the paths' computation is significantly reduced by the first component, which contributes to decreasing the computation times, and traffic demands can be allocated by the second module in real-time as they come. Based on experimental simulations with real-world topologies and traffic matrices, energy savings between around 50% and 80% are reached by the proposed energy-aware approaches. The heuristic strategy attains results very close to the optimal values. In addition, the proposed algorithms are able to converge much faster and handle larger network sizes for which the exact model fails to find solutions in a reasonable time.

Apart from providing an effective power-aware scheme able to achieve notable improvements in terms of energy saving, the most significant added value of this proposal is the insightful analysis presented to evaluate the impact on network performance. Extensive simulations validate that crucial network parameters such as control traffic delay, data path latency, link utilization and TCAM occupation are affected by the performance-agnostic energy-aware model. Therefore, this paper stands as a valuable proposal in the design of online energy-aware strategies suitable for control planes of current programmable networks. Moreover, the intrinsic trade-off between environmental and performance concerns is disclosed and investigated using the proposed approach. As future work, we plan to extend the analysis about the impact of energy-aware solutions based on the reduction of active network elements on SDN reliability, considering in-band control traffic.

Acknowledgments: This work has been supported by the Ministerio de Economía y Competitividad of the Spanish Government under Project TEC2016-76795-C6-1-R and AEI / FEDER, UE, and through a predoctoral FPI scholarship.

Author Contributions: Adriana Fernández-Fernández, Cristina Cervelló-Pastor and Leonardo Ochoa-Aday conceived of and designed the proposed scheme. Adriana Fernández-Fernández, Cristina Cervelló-Pastor and Leonardo Ochoa-Aday conceived and designed the simulations. Adriana Fernández-Fernández performed the simulations. Adriana Fernández-Fernández, Cristina Cervelló-Pastor and Leonardo Ochoa-Aday analyzed the data. Adriana Fernández-Fernández wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, White Paper. Available online: <https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/white-paper-listing.html> (accessed on 31 October 2017).
2. Agiwal, M.; Roy, A.; Saxena, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1617–1655.
3. Tucker, R.S. Energy consumption in telecommunications. In Proceedings of the IEEE Optical Interconnects Conference (OI), Santa Fe, NM, USA, 20–23 May 2012; pp. 1–2.
4. 5G Infrastucture Public Private Partnership, KPIs. Available online: <https://5g-ppp.eu/kpis/> (accessed on 31 October 2017).
5. Fernández-Fernández, A.; Cervelló-Pastor, C.; Ochoa-Aday, L. Evaluating the Impact of Energy-Aware Routing on Software-Defined Networking Performance. In Proceedings of the XIII Jornadas de Ingeniería Telemática (JITEL), Valencia, Spain, 27–29 September 2017; pp. 241–248.

6. Bousia, A.; Kartsakli, E.; Antonopoulos, A.; Alonso, L.; Verikoukis, C. Multiobjective Auction-Based Switching-Off Scheme in Heterogeneous Networks: To Bid or Not to Bid? *IEEE Trans. Veh. Technol.* **2016**, *65*, 9168–9180.
7. Oikonomakou, M.; Antonopoulos, A.; Alonso, L.; Verikoukis, C. Evaluating Cost Allocation Imposed by Cooperative Switching Off in Multi-Operator Shared HetNets. *IEEE Trans. Veh. Technol.* **2017**, *PP*, doi:10.1109/TVT.2017.2719404.
8. Soliman, S.S.; Song, B. Fifth generation (5G) cellular and the network for tomorrow: cognitive and cooperative approach for energy savings. *J. Netw. Comput. Appl.* **2017**, *85*, 84–93.
9. Sun, S.; Gong, L.; Rong, B.; Lu, K. An intelligent SDN framework for 5G heterogeneous networks. *IEEE Commun. Mag.* **2015**, *53*, 142–147.
10. Ordonez-Lucena, J.; Ameigeiras, P.; Lopez, D.; Ramos-Munoz, J.J.; Lorca, J.; Figueira, J. Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges. *IEEE Commun. Mag.* **2017**, *55*, 80–87.
11. Vincenzi, M.; Antonopoulos, A.; Kartsakli, E.; Vardakas, J.; Alonso, L.; Verikoukis, C. Multi-Tenant Slicing for Spectrum Management on the Road to 5G. *IEEE Wirel. Commun.* **2017**, *24*, 118–125.
12. Antonopoulos, A.; Kartsakli, E.; Perillo, C.; Verikoukis, C. Shedding Light on the Internet: Stakeholders and Network Neutrality. *IEEE Commun. Mag.* **2017**, *55*, 216–223.
13. Kreutz, D.; Ramos, F.M.; Verissimo, P.; Esteve Rothenberg, C.; Azodolmolky, S.; Uhlig, S. Software-Defined Networking: A Comprehensive Survey. *Proc. IEEE* **2015**, *103*, 14–76.
14. Gupta, M.; Singh, S. Greening of the Internet. In Proceedings of the ACM SIGCOMM, Karlsruhe, Germany, 25–29 August 2003; pp. 19–26.
15. Tuysuz, M.F.; Ankarali, Z.K.; Gözüpek, D. A survey on energy efficiency in software defined networks. *Comput. Netw.* **2017**, *113*, 188–204.
16. Bolla, R.; Bruschi, R.; Carrega, A.; Davoli, F. Green network technologies and the art of trading-off. In Proceedings of the 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Shanghai, China, 10–15 April 2011; pp. 301–306.
17. Schiff, L.; Schmid, S.; Kuznetsov, P. In-Band Synchronization for Distributed SDN Control Planes. *SIGCOMM Comput. Commun. Rev.* **2016**, *46*, 37–43.
18. Fernández-Fernández, A.; Cervelló-Pastor, C.; Ochoa-Aday, L. A Multi-Objective Routing Strategy for QoS and Energy Awareness in Software-Defined Networks. *IEEE Commun. Lett.* **2017**, *21*, 2416–2419.
19. Hock, D.; Hartmann, M.; Gebert, S.; Jarschel, M.; Zinner, T.; Tran-Gia, P. Pareto-optimal resilient controller placement in SDN-based core networks. In Proceedings of the 25th International Teletraffic Congress (ITC), Shanghai, China, 10–12 September 2013; pp. 1–9.
20. Zhang, M.; Yi, C.; Liu, B.; Zhang, B. GreenTE: Power-Aware Traffic Engineering. In Proceedings of the 18th IEEE International Conference on Network Protocols (ICNP), Kyoto, Japan, 5–8 October 2010; pp. 21–30.
21. Assefa, B.G.; Ozkasap, O. State-of-the-art Energy Efficiency Approaches in Software Defined Networking. In Proceedings of the International Symposium on Advances in Software Defined Networks (SoftNetworking), Barcelona, Spain, 19–24 April 2015.
22. Wang, H.; Li, Y.; Jin, D.; Hui, P.; Wu, J. Saving Energy in Partially Deployed Software Defined Networks. *IEEE Trans. Comput.* **2016**, *65*, 1578–1592.
23. Giroire, F.; Moulhierac, J.; Phan, T.K. Optimizing Rule Placement in Software-Defined Networks for Energy-Aware Routing. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Austin, TX, USA, 8–12 December 2014; pp. 2523–2529.
24. Wang, R.; Jiang, Z.; Gao, S.; Yang, W.; Xia, Y.; Zhu, M. Energy-Aware Routing Algorithms in Software-Defined Networks. In Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Sydney, Australia, 16–19 June 2014; pp. 1–6.
25. Markiewicz, A.; Tran, P.N.; Timm-Giel, A. Energy Consumption Optimization for Software Defined Networks Considering Dynamic Traffic. In Proceedings of the IEEE 3rd International Conference on Cloud Networking (CloudNet), Luxembourg, 8–10 October 2014; pp. 155–160.
26. Ruiz-Rivera, A.; Chin, K.W.; Soh, S. GreCo: An Energy Aware Controller Association Algorithm for Software Defined Networks. *IEEE Commun. Lett.* **2015**, *19*, 541–544.
27. Fernández-Fernández, A.; Cervelló-Pastor, C.; Ochoa-Aday, L. Achieving Energy Efficiency: An Energy-Aware Approach in SDN. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016; pp. 1–7.

28. Fernández-Fernández, A.; Cervelló-Pastor, C.; Ochoa-Aday, L. Energy-Aware Routing in Multiple Domains Software-Defined Networks. *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.* **2016**, *5*, 13–19.
29. Cianfrani, A.; Eramo, V.; Listanti, M.; Marazza, M.; Vittorini, E. An Energy Saving Routing Algorithm for a Green OSPF Protocol. In Proceedings of the IEEE Conference on Computer Communications (INFOCOM), San Diego, CA, USA, 14–19 March 2010; pp. 1–5.
30. Cianfrani, A.; Eramo, V.; Listanti, M.; Polverini, M. An OSPF enhancement for energy saving in IP networks. In Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Shanghai, China, 10–15 April 2011; pp. 325–330.
31. Cuomo, F.; Abbagnale, A.; Cianfrani, A.; Polverini, M. Keeping the connectivity and saving the energy in the internet. In Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Shanghai, China, 10–15 April 2011; pp. 319–324.
32. Cuomo, F.; Cianfrani, A.; Polverini, M.; Mangione, D. Network Pruning for Energy Saving in the Internet. *Comput. Netw.* **2012**, *56*, 2355–2367.
33. Cuomo, F.; Abbagnale, A.; Papagna, S. ESOL: Energy saving in the Internet based on Occurrence of Links in routing paths. In Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Lucca, Italy, 20–24 June 2011; pp. 1–6.
34. Cianfrani, A.; Eramo, V.; Listanti, M.; Polverini, M.; Vasilakos, A.V. An OSPF-Integrated Routing Strategy for QoS-Aware Energy Saving in IP Backbone Networks. *IEEE Trans. Netw. Serv. Manag.* **2012**, *9*, 254–267.
35. Giroire, F.; Mazauric, D.; Moulierac, J.; Onfroy, B. Minimizing Routing Energy Consumption: From Theoretical to Practical Results. In Proceedings of the IEEE/ACM International Conference on Green Computing and Communications (GreenCom), Hangzhou, China, 18–20 December 2010; pp. 252–259.
36. Nuutila, E.; Soisalon-soininen, E. On Finding the Strongly Connected Components in a Directed Graph. *Inf. Process. Lett.* **1994**, *49*, 9–14.
37. Gurobi Optimizer (Version 6.5). Available online: <http://www.gurobi.com/> (accessed on 31 October 2017).
38. Orłowski, S.; Pióro, M.; Tomaszewski, A.; Wessäly, R. SNDlib 1.0-Survivable Network Design Library. *Networks* **2010**, *55*, 276–286.
39. Li, J.; Yoo, J.H.; Hong, J.W.K. Dynamic control plane management for software-defined networks. *Int. J. Netw. Manag.* **2016**, *26*, 111–130.
40. Parvez, I.; Rahmati, A.; Guvenc, I.; Sarwat, A.I.; Dai, H. A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions. *ArXiv* **2017**, arXiv:cs.NI/1708.02562. Available online: <http://arxiv.org/abs/1708.02562> (accessed on 31 October 2017).
41. Li, C.; Wang, H.; Van Mieghem, P. Degree and Principal Eigenvectors in Complex Networks. In *NETWORKING 2012: 11th International IFIP TC 6 Networking Conference, Prague, Czech Republic, May 21–25, 2012, Proceedings, Part I*; Bestak, R., Kencl, L., Li, L.E., Widmer, J., Yin, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 149–160.
42. Mahadevan, P.; Krioukov, D.; Fomenkov, M.; Dimitropoulos, X.; Claffy, K.C.; Vahdat, A. The Internet AS-level Topology: Three Data Sources and One Definitive Metric. *SIGCOMM Comput. Commun. Rev.* **2006**, *36*, 17–26.
43. Brede, M.; de Vries, B.J. Networks that optimize a trade-off between efficiency and dynamical resilience. *Phys. Lett. A* **2009**, *373*, 3910–3914.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Reproduced with permission of copyright owner. Further reproduction
prohibited without permission.