White Rose Research Online
eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/107255/

Version: Accepted Version

Proceedings Paper:

Nonde, L, Lawey, AQ, Elgorashi, TEH et al. (2017) Energy Efficient Cloud Networks. In: International Workshop on Green ICT and Smart Networking (GISN 2016). 12th International Conference on Network and Service Management, 31 Oct - 04 Nov 2016, Montreal, Quebec, Canada.. IEEE, pp. 412-417. ISBN: 978-1-5090-3236-5. EISSN: 2165-963X.

https://doi.org/10.1109/CNSM.2016.7818457

© 2016 IFIP. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Energy Efficient Cloud Networks

(Invited Paper)

Leonard Nonde, Ahmed Q. Lawey, Taisir E.H. Elgorashi and Jaafar M.H. Elmirghani
School of Electronic and Electrical Engineering University of Leeds, LS2 9JT, United Kingdom

Abstract— **Cloud computing is expected to be a major factor that will dominate the future Internet service model. This paper summarizes our work on energy efficiency for cloud networks. We develop a framework for studying the energy efficiency of four cloud services in IP over WDM networks: cloud content delivery, storage as a service (StaaS), and virtual machines (VMS) placement for processing applications and infrastructure as a service (IaaS).Our approach is based on the co-optimization of both external network related factors such as whether to geographically centralize or distribute the clouds, the influence of users' demand distribution, content popularity, access frequency and renewable energy availability and internal capability factors such as the number of servers, switches and routers as well as the amount of storage demanded in each cloud. Our investigation of the different energy efficient approaches is backed with Mixed Integer Linear Programming (MILP) models and real time heuristics.**

Keywords-Cloud; Content Distribution; Virtual Machine Migration; Network Virtualization; Virtual Network Embedding

I.INTRODUCTION

Backed by the provisioning of cloud based data intensive applications and services, Internet traffic has been growing exponentially. It is estimated that by 2020 there will be over 50 billion devices connected to the Internet all expected to leverage the availability of various cloud services and applications [1]. Cloud applications are challenging the cloud service delivery model in capacity and reach to interconnect the cloud to users' locations with guaranteed quality of service. One of the key challenges to maintain sustainability of cloud computing applications is reducing its energy consumption. Driven by the economic, environmental and societal impact, significant academic and industrial research effort has been focused recently on reducing the power consumption of cloud computing.

In this paper we give an overview of the different energy efficiency approaches for cloud services. We developed a framework for optimizing cloud services delivery in IP over WDM networks. We considered four cloud service offerings: (i) cloud content delivery, (ii) storage as a service (StaaS), (iii) virtual machines (VMS) placement for processing applications [2]-[4] and (iv) infrastructure as a service (IaaS) [5]-[7]. Our framework comprises MILP models and real time heuristics for the design of energy efficient cloud computing services taking into account two factors: (i) external network related factors including cloud geographical location (i.e. centralization vs. distribution), users demand distribution, content popularity, access frequency and renewable energy availability, (ii) internal capability factors including the number of servers, switches and routers and the amount of storage demanded in each cloud.

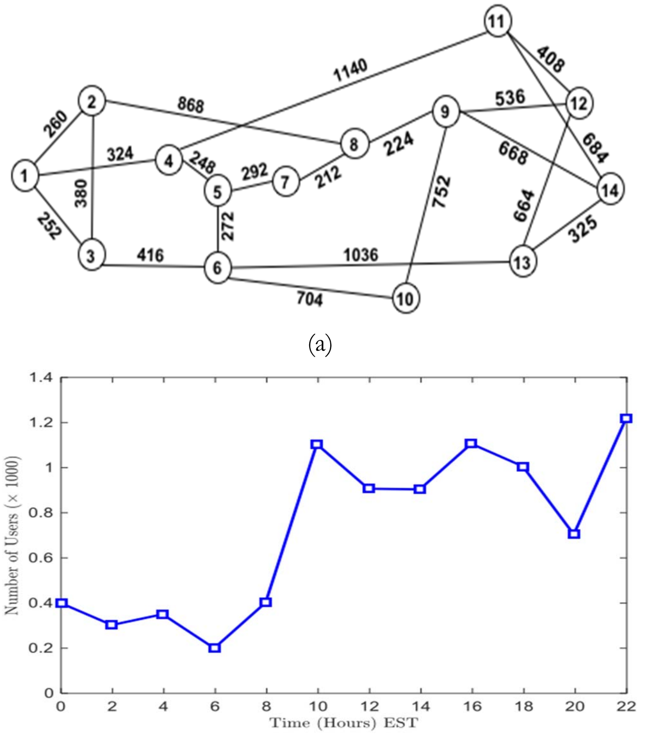
Our energy efficient cloud networks work was carried out as part of the £6m Engineering and Physical Sciences Research Council (EPSRC), INTelligent Energy awaRe NETworks (INTERNET) Programme Grant, 2010-2016. The framework for distributed cloud services and network virtualization was adopted by GreenTouch. , a 50+ member consortium including vendors: Alcatel-Lucent (and Alcatel-Lucent Bell Labs, now Nokia Bell Labs), Huawei, Fujitsu; and Service providers including France Telecom, AT&T, Swiss Com, China Mobile, NTT. It was part of a comprehensive research study, referred to as GreenMeter that investigated the overall impact and energy efficiency obtained from implementing a range of technologies, architectures, devices and protocols developed by GreenTouch and recommended measures for network operators worldwide. The outcomes of this study are published in [8].

II.ENERGY EFFICIENT CONTENT DELIVERY CLOUD

In [9], the authors studied optimizing content distribution for content providers from an energy efficiency point of view by comparing conventional and decentralized server based content delivery networks (CDN), content centric networks (CCN), and centralized server based CDN using dynamic optical bypass. In our work [2], [3] we further investigated the energy efficiency of content delivery networks by studying a range of content replication schemes for content delivery services and analyzed their performance in terms of network and cloud power consumption using MILP modelling. The investigated schemes include:

i. Maximum Full Replication (MFR): Users at each node are served by a local cloud that has a full copy of the content.
ii. Maximum No Replication (MNR): The content is distributed among all the clouds locations without replication.
iii. Maximum Popularity Based Replication (MPR): A cloud at each node, however, the number and locations of content replicas are optimized among all the clouds locations based on content popularity.
iv. Optimal Full Replication (OFR): Each cloud has a full copy of the content.
v. Optimal No Replication (ONR): Content is distributed among the optimum clouds without replication.
vi. Optimal Popularity Based Replication (OPR): Given the client requests and the popularity of the content, the optimum number of clouds and their locations in the network are selected as well as the capability of each cloud so that the total power consumption is minimized.

The network and cloud power consumption resulting from placing 50 popularity groups of content using these schemes in the NSFNET network, depicted in Figure 1(a), under the number of users profile in Figure 1(b) are evaluated. The popularity of the different objects of the content is assumed to follow a Zipf distribution, representative of the popularity distribution of several cloud content types such as YouTube and others. The results show that replicating content based on its popularity, i.e. the OPR scheme, is the optimum scheme for energy efficiency with 43% total (network and cloud) power saving compared to centralized content delivery in which datacenters use power un-aware servers management [3].



(a)

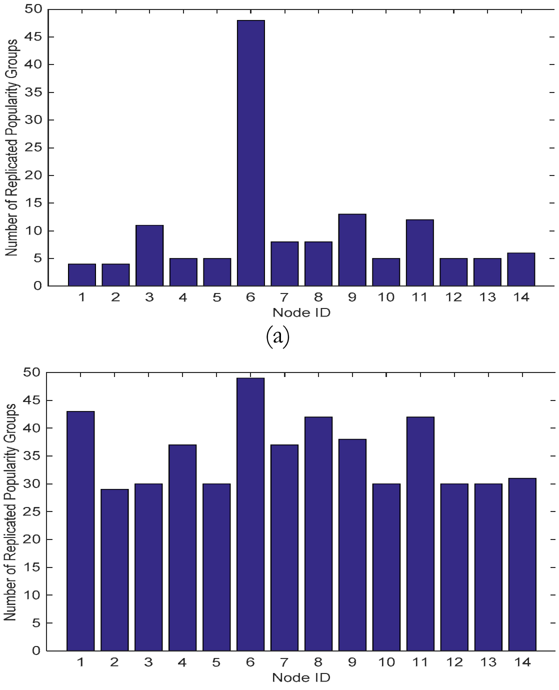
(b)

Figure 1: The scenario studied (a) The NSFNET network, (b) Number of users versus time of the day

Figure 2 shows the placement of the content groups under the OPR scheme at 06:00 and 22:00 which correspond to the lowest and highest load demands, respectively (See figure 1 (b)). The majority of the content is placed in the node that results in minimum average hop to other nodes (node 6 of the NSFNET network) at low traffic. On the other hand, at high traffic periods, more popularity groups are replicated into more clouds. This is because during high traffic, the power consumed in the clouds is compensated for by higher network power savings. We further proposed an energy efficient cloud content delivery heuristic based on the model insights, with analogous power efficiency to the MILP and lower computational complexity.

III. RENEWABLE ENERGY IN ENERGY EFFICIENT CONTENT DELIVERY CLOUD

We studied reducing the CO_2 emission of backbone IP over WDM networks by exploiting renewable energy sources in [10]. We further extended our study in [11] to investigate the problem of whether to locate datacenters next to renewable energy or to transmit renewable energy to datacenters. In [4] we extended our cloud content delivery model to study the impact of renewable energy availability on the optimization of cloud locations, internal capability and content replication patterns. Wind farms are considered as the source of renewable energy for their high production capacity and competitive price per megawatt hour compared to non-renewable energy.



(a)

(b)

Figure 2: Number of popularity groups replicated in each node (a) at 06:00 and (b) at 22:00

We studied a scenario where renewable energy is only available to power cloud data centres while the IP/WDM network is powered by non-renewable energy. This scenario calls for a trade-off between the non-renewable power savings achieved by powering the cloud using renewable energy and the power consumption of the network resulting from users accessing the content in cloud locations near to the renewable energy sources.

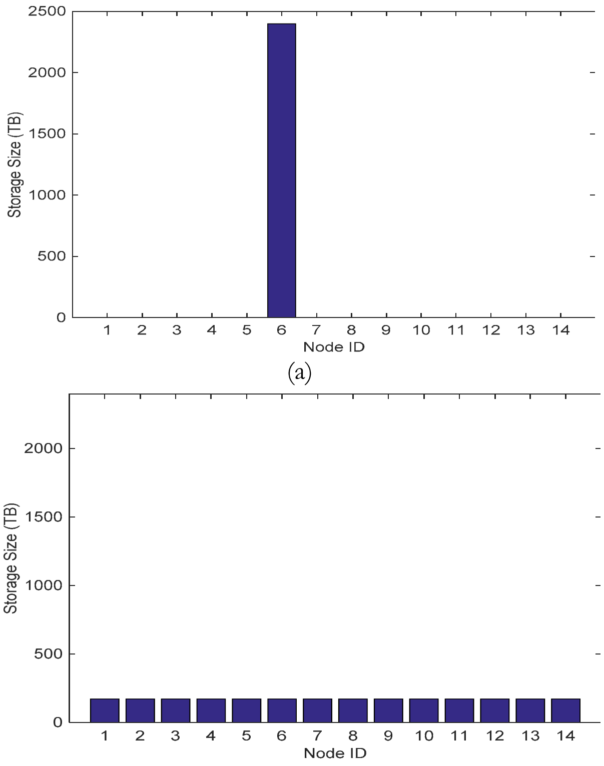
We developed a MILP to optimize the use of renewable energy under different scenarios of renewable energy availability. The model considers minimizing the IP/WDM network non-renewable power consumption, the cloud non-renewable power consumption and transmission power losses. Our results show that under limited renewable energy availability distributing content into a number of mini clouds based on content popularity is the optimum approach for energy efficiency.

The availability of abundant renewable energy sources to power all the clouds necessitates achieving a trade-off between transmission losses and non-renewable power consumption in the IP/WDM network. On one hand, building clouds in proximity of wind farms reduces transmission losses; but this reduction comes at the cost of consuming more non-renewable power. On the other hand, if the transmission losses are not considered but the renewable energy available is distance dependent, then building a cloud in each node results in saving network power consumption if enough renewable energy is available.

IV. ENERGY EFFICIENT STORAGE AS A SERVICE (STAAS)

We extend the content delivery model to optimize StaaS applications [3]. StaaS applications such as Dropbox, Google Drive, Skydrive, iCloud, and Box can be considered as a special case of the content delivery applications where content is only accessed by the owner of the content or a very limited number of authorized users. Our analysis looks at three different schemes to serve StaaS users: (i) Single Cloud where content owners are served from a single central cloud only; (ii) Optimal Clouds where content can be replicated into multiple clouds allowing the content owners to access their content either from the central cloud or from a nearby cloud; (iii) Max Clouds where the content is replicated at each node to serve nodes locally.

Our evaluation of these schemes considered a scenario where content owners with a storage quota of 2 GB are uniformly distributed over the NSFNET network accessing files of 45 MB and 22.5 MB at varying access frequencies. At low access frequencies, the Optimal Clouds scheme selects to serve all users from the central cloud, i.e. Single Cloud scheme (Figure 3 (a)) while content is copied into multiple clouds to be in proximity of the owner content to serve them locally at higher access frequencies (Figure 3(b)). The Optimal Cloud scheme resulted in 48% network power consumption savings compared to using a single central cloud to serve content (Single Cloud scheme) for an average file size of 45MB [3]. Limited power savings are obtained for smaller file sizes as the total traffic is a function of content size.



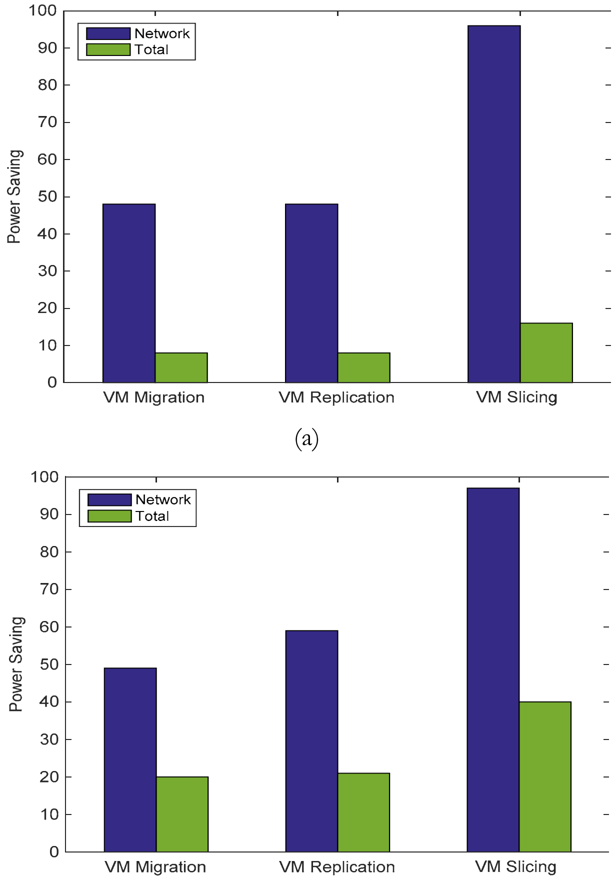
(a)

(b)

Figure 3: Storage size in each node under the StaaS optimal scheme at (a) 10 downloads/hour content access frequency (b) 130 downloads/hour content access frequency

V. VIRTUAL MACHINE PLACEMENT OPTIMIZATION

Our optimization approach is further extended to study optimizing the placement of virtual machines (VMs) [2]. A VM can be defined as a logical entity created in physical resources in response to a service request by one or more users sharing that VM. Machine virtualization opens the door for energy efficient dynamic infrastructure management through resources consolidation, i.e. multiple virtual machines coexist in a shared physical resource. Three different schemes for VMs placement are studied: (i) VM Replication where multiple copies of each VM can be created to serve the requests, (ii) VM Migration where a single copy of each VM serves all the requests, (iii)VM Slicing where a VM is sliced into smaller VMs each serving a smaller number of users. Our results show that VM migration and replication has similar power savings as we assume VMs are accessed uniformly across the network. However, by slicing the VMs into smaller VMs and placing them in proximity to their users, a saving of 25% of the total power can be achieved compared to a single virtualized cloud scenario [3]. Figure 4 summarizes the power savings achieved by the different VM placement schemes. We also developed a heuristic (DEER-VM) for real time VM placement which has achieved comparable power savings and lower computational complexity compared to the MILP.



(a)

(b)

Figure 4: Power Savings of the Different Schemes Compared to a Single Cloud (a) at 06:00 hours and (b) at 22:00 hours

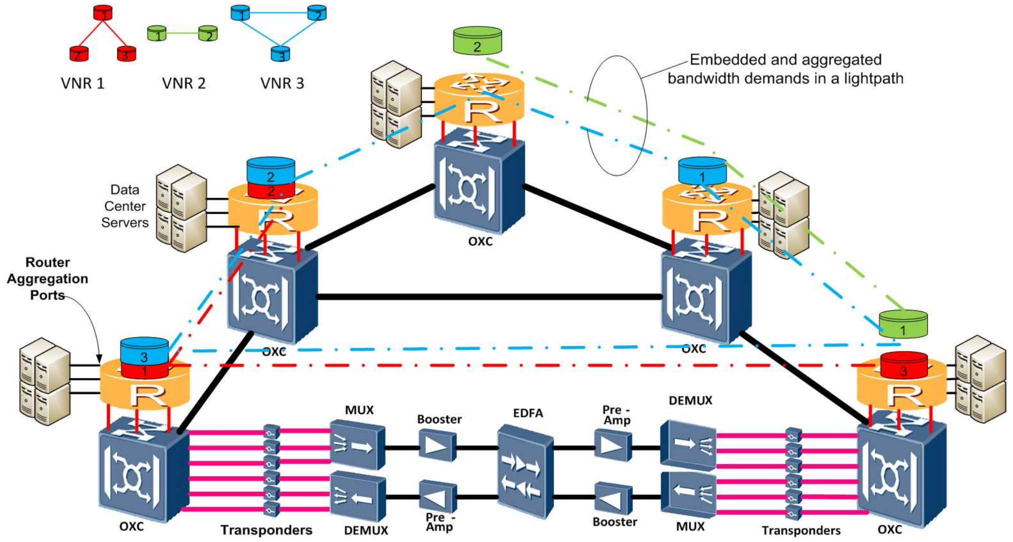


Figure 5: Virtual Network Embedding in IP over WDM Core Network with Data Centres [4]

VI.Energy Efficient Virtual Network Embedding for Infrastructure as a Service (IaaS)

Network Virtualization is extensively regarded as one of the main archetypes for the future Internet architecture as it provides a number of benefits including scalability, on request allocation of network resources and the potential of efficient use of network resources. In view of the ever-increasing demand for various types of Internet content and the acceptance of cloud computing as a broadly accepted computing paradigm, network virtualization will play a very vital role in the cloud service provisioning of infrastructure as a service (IaaS). IaaS is the delivery of virtualized and dynamically scalable computing, storage and networking on demand to clients on a pay as you use basis [12]. We have investigated the energy efficiency benefits which resource consolidation brings in network virtualization.

A. Energy Efficient Virtual Network Embedding Model and Heuristic

We have developed an efficient virtual resource allocation approach that takes power consumption into consideration during the provisioning of both networking and data centre resources [5]. We model our approach in an IP over WDM network with data centres using MILP and then we develop a heuristic, Real-time Energy Optimized Virtual Network Embedding (REOViNE) with power savings approaching those of the MILP model. The network architecture considered is as shown in Fig. 5. In our analysis, we have considered two different data centre power consumption profiles; an energy inefficient power profile where idle power consumption is accounted for and an energy efficient power profile where only power due to server processing is considered. Since we are considering Infrastructure as a Service, each client requests a portion of the network comprising both networking and processing requirements. The cloud service provider is accountable for service provisioning. Each request from an individual customer is referred to as a virtual network request (VNR) and the process of provisioning the VNR onto the physical network is referred to as virtual network embedding (VNE) [13].

Comparisons in terms of power consumption have been made against two other methods; one that focuses on minimizing the overall bandwidth resources used in the network during resource provisioning [14] and the other that simply reduces the number of activated links and data centers [15] during resource provisioning. In an energy inefficient data centre scenario, our developed MILP model, where the energy consumption is minimized by jointly optimizing the use of network resources and consolidating resources in data centers, has better power savings , as seen in Figure 6, compared to the approach that simply keeps the activated links and data centres to a minimum. This is because the network power consumption is driven by the wavelengths used rather than the number of active links. The approach that centers on minimizing the bandwidth used does not account for the number of activated data centers; therefore it has a poor power consumption performance in data centers. However, as the network becomes fully loaded and all the data centers are activated, bandwidth use becomes the key factor in the overall power consumption. Since data center power consumption is the main contributor to the total power consumption (given the power consumption of servers and network devices), the total power consumption, follows similar trends to that of the data centers. Overall maximum energy savings of 60% were achieved through MILP and heuristic [5].

When data centers with an energy efficient power profile are considered where only servers needed to serve a given workload are activated, i.e. unused servers are switched off, the power consumption performance of the various approaches provide a different picture. All the three approaches have similar power consumption trends in the data centers because the power consumption of energy efficient data centers is a function of the server utilization and not the number of activated data centers as there is no idle power. The total power consumption therefore follows the trends of the power consumption in networking components. This leads us to the conclusion that with the energy efficient data center power profile, the optimal VNE approach with the minimum power consumption is the one that only minimizes the bandwidth used in the network. This is however only true if requests are served sequentially and the new requests are accommodated in the spare capacity without reconfiguring the entire network.

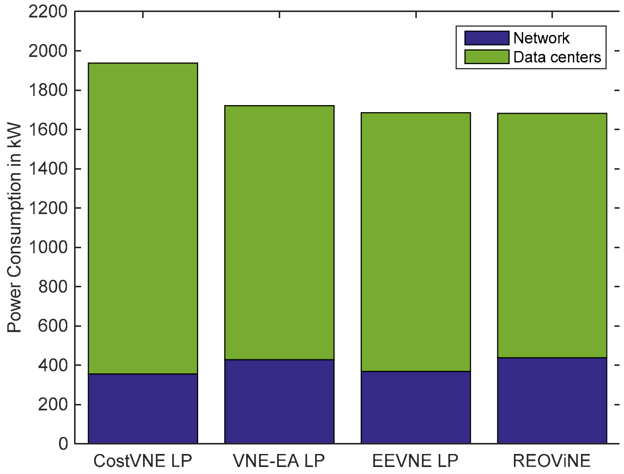


Figure 6: Network and data center power consumption of the different VNE approaches

B. *Energy Efficient Virtual Network Embedding in Optical OFDM Cloud Networks*

An OFDM-based elastic optical network can support various data rates through the use of bandwidth variable transponders (BVTs) and bandwidth variable optical cross-connects (BV-OXCs) deployed at each node in the core network. BVTs can adjust their actual transmission rate in response to the actual traffic demand through expansion or contraction of the optical path by varying the number of subcarriers used or the modulation format of subcarriers. By deploying optical OFDM (O-OFDM) as the underlying technology in the optical layer, replacing WDM, the virtual network embedding problem therefore extends all the way to the selection of the number of subcarriers and modulation format used during the link provisioning phase in the substrate network [5],[6].

Virtual network embedding in power minimized IP over O-OFDM saves 63% of the optical layer power consumption compared to VNE in IP over WDM while the spectrum minimized virtual network embedding saves 17% of the optical layer power consumption compared to the WDM implementation [6]. The power minimized O-OFDM network selects routes and modulation formats that result in minimum power consumption while the spectrum minimized O-OFDM selects routes that can support the highest modulation format to minimise the used spectrum. However, since the capacity of the links is limited, both approaches might consume the same amount of power to embed some of the VNRs at high loads under two scenarios. The first scenario is when the number of available subcarriers on the only available route is limited so the two models have to adopt the maximum modulation format to create enough capacity to embed the bandwidth requirement. The second scenario is when the bandwidth requirement has to be embedded over a longer route that can only support the lowest modulation format so the two models will have to adopt the lowest modulation format.

C. *Virtual Network Embedding Employing Renewable Energy Sources*

The use of renewable energy in cloud networks is becoming an urgent requirement for Infrastructure Providers (InPs) as the regulations surrounding the amount of CO_2 emissions are becoming stringent in this era of environmental sustainability. We have addressed the problem of reducing the greenhouse gas (GHG) emissions of cloud infrastructure networks hosting virtual networks [7]. An examination of the effective use of renewable energy during the resource allocation phase of virtual network embedding in cloud networks has been carried out as means of reducing the carbon footprint. We have developed a Green Virtual Network Embedding (GVNE) framework for minimizing the use of non-renewable energy through intelligent provisioning of bandwidth and cloud data center resources. The problem is modeled as a mixed integer linear program (MILP). The results show that it is better to instantiate virtual machines in cloud data centers that have access to abundant renewable energy even at the expense of traversing several links across the network. The GVNE model reduces the overall CO_2 emissions by up to 32% for the network considering solar power availability and data center locations [7].

VII.CONCLUSIONS

This paper has reviewed our recent work in energy efficient cloud networks. We have outlined our main findings in the optimization of four cloud services over IP over WDM networks: cloud content delivery, storage as a service (StaaS), and virtual machines (VMS) placement for processing applications and infrastructure as a service (IaaS). Our optimization is based on MILP models and real-time heuristics that were recommended by the GreenTouch Consortium to network operators worldwide.

ACKNOWLEDGEMENT

The authors would like to acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC), INTERNET (EP/H040536/1) and STAR (EP/K016873/1). All data are provided in full in the results section of this paper.

REFERENCES

[1] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020. White Paper 2016 2016; Available from: http://www.cisco.com/c/en/us/solutions/collateral/se

rvice-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

[2] Ahmed Q. Lawey, Taisir E. H. El-Gorashi and Jaafar M. H. Elmirghani, "Energy efficient cloud content delivery in core networks," in Proc. of IEEE IEEE Global Telecommunications Conference (GLOBECOM'13) Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA), Atlanta, GA, USA, 09– 13 Dec, 2013.

[3] Ahmed Q. Lawey, Taisir E. H. El-Gorashi and Jaafar M. H. Elmirghani, "Distributed energy efficient clouds over Core Network," IEEE/OSA Journal of Lightwave Technology, vol. 32, no. 7, pp.1261-1281, April 2014

[4] A. Q. Lawey, T. E. H. El-Gorashi and J. M. H. Elmirghani, "Renewable energy in distributed energy efficient content delivery clouds," 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 128-134.

[5] L. Nonde, T.E.H. El-Gorashi, and J.M.H. Elmirghani, "Energy Efficient Virtual Network Embedding for Cloud Networks", Journal of Lightwave Technology, 2015. 33(9): p. 1828-1849.

[6] L. Nonde, T.E.H. El-Gorashi, and J.M.H. Elmirghani, "Green Virtual Network Embedding in Optical OFDM Cloud Networks," in 16th IEEE Conference on Transparent Optical Networks (ICTON), 2014.

[7] L. Nonde, T.E.H. El-Gorashi, and J.M.H. Elmirghani, "Virtual Network Embedding Employing Renewable Energy Sources," Accepted in proc. IEEE Global Comminications Conference (GLOBECOM), 2016.

[8] "White Paper on GreenTouch Final Results from GreenMeter Research Study." A GreenTouch White Paper, Version 2, 15 Aug 2015. Available at: https://s3-us-west-2.amazonaws.com/belllabs-microsite-greentouch/uploads/documents/White%20Paper%20on%20Green%20Meter%20Final%20Results%20August%202015%20Revision%20-%20vFINAL.pdf

[9] Guan, Kyle, et al. "On the energy efficiency of content delivery architectures." 2011 IEEE International Conference on Communications Workshops (ICC). IEEE, 2011.

[10] Xiaowen Dong, Taisir E. H. El-Gorashi and Jaafar M. H. Elmirghani "IP Over WDM Networks Employing Renewable Energy Sources," IEEE/OSA Journal of Lightwave Technology, vol. 29, no. 1, pp. 3-14, Jan 2011.

[11] Xiaowen Dong, T. El-Gorashi and J. M. H. Elmirghani, "Green IP Over WDM Networks With Data Centers," in Journal of Lightwave Technology, vol. 29, no. 12, pp. 1861-1880, June15, 2011.

[12] R. Moreno-Vozmediano, R. S. Montero and I. M. Llorente, "IaaS Cloud Architecture: From Virtualized Datacenters to Federated Cloud Infrastructures," in Computer, vol. 45, no. 12, pp. 65-72, Dec. 2012.

[13] A. Fischer, et al., "Virtual Network Embedding: A Survey," Communications Surveys & Tutorials, IEEE, 2013. 15(4): p. 1888-1906.

[14] Houidi, I., et al., "Virtual network provisioning across multiple substrate networks," Computer Networks, 2011. 55(4): p. 1011-1023.

[15] J.F. Botero, et al., "Energy Efficient Virtual Network Embedding," IEEE Communications, 2012. 16(5): p. 756-759.