On the Energy Efficiency Prospects of Network Function Virtualization

Rashid Mijumbi†, Joan Serrat†, Juan-Luis Gorricho‡ and Javier Rubio-Loyola‡
†Universitat Politècnica de Catalunya, 08034 Barcelona, Spain
‡CINVESTAV, Tamaulipas, Mexico

Abstract—Network Function Virtualization (NFV) has recently received significant attention as an innovative way of deploying network services. By decoupling network functions from the physical equipment on which they run, NFV has been proposed as a way to enhance network agility and efficiency, better time-to-market, and reduced Capital Expenses (CAPEX) and Operating Expenses (OPEX). One of the main selling points of NFV is its promise for better energy efficiency resulting from consolidation of resources as well as their more dynamic utilization. However, there are currently no studies or implementations which attach values to energy savings that can be expected, which could make it hard for Telecommunication Service Providers (TSPs) to make investment decisions. In this paper, we utilize Bell Labs’ GW ATT tool to estimate the energy savings that could result from the three main NFV use cases—Virtualized Evolved Packet Core (VEPC), Virtualized Customer Premises Equipment (VCPE) and Virtualized Radio Access Network (VRAN). We determine that the part of the mobile network with the highest energy utilization is expected to lead to several advantages. First, since the NF can be deployed, chained and updated remotely, this promises more flexibility, agility and reduced time-to-market. Second, it also allows for easier software upgrades and cheaply to changes in market conditions. For example, as represented by the number of servers allocated to each function in Fig. 2, there might be a need to increase user plane resources without affecting the control plane. In this case, VNFRs such as a virtual MME may scale independently according to their specific resource requirements. In the same way, VNFRs dealing with the data plane might require a different number of resources than those dealing with signaling only. This flexibility is expected to lead to more efficient utilization of resources. Finally, it also allows for easier software upgrades on the EPC network functions, which would hence allow for faster launch of innovative services.

I. INTRODUCTION

TSPs have been facing an explosion in the data that their networks must support. And this will not stop soon. According to Cisco [1], the annual global IP traffic will nearly triple from 2014 to 2019. Specifically, the traffic will pass the zettabyte threshold by the end of 2016, and reach 2 zettabytes per year by 2019, at which time two-thirds of all this traffic will originate from non-PC devices. With this traffic burst, TSPs must correspondingly expand and maintain their networks both of which lead to increased CAPEX and OPEX. Yet, due to competition both among themselves and from services provided over-the-top on their infrastructure, TSPs cannot respond to these increased costs by increasing subscription fees. This has lead to reductions in TSP revenues [2], and forced them to search for ways of deploying and maintaining networks at reduced costs.

NFV [3], [4] has been proposed as a viable path towards achieving this goal. By leveraging recent advances in virtualization technology, the main idea of NFV is to decouple Network Functions (NFs) from the physical equipment on which they run. This way, the NFs can be virtualized, and run on industry standard servers and switches which could be located in either data centers or network nodes. The Virtual Network Functions (VNFRs) may then be relocated and instantiated in different network locations (for example aimed at introduction of a service targeting customers in a given geographical location) without necessarily requiring purchase and installation of new hardware [5].

One of the network parts which is expected to benefit from NFV is the EPC, which is the core network for Long Term Evolution (LTE) as specified by 3GPP [6]. In Fig. 1 we show a basic architecture of LTE without NFV. The User Equipment (UE) is connected to the EPC over the LTE access network (E-UTRAN). The evolved NodeB (eNodeB) is the base station for LTE radio. The EPC performs essential functions including subscriber tracking, mobility management and session management. It is made up of four NFs: Serving Gateway (S-GW), Packet Data Network (PDN) Gateway (P-GW), Mobility Management Entity (MME), and Policy and Charging Rules Function (PCRF). It is also connected to external networks, which may include the IP Multimedia Core Network Subsystem (IMS). In current EPC implementations, all its functions are based on propriety equipment spread across the core network. Therefore, even minor changes to a given function may require a replacement of the equipment. The same applies to cases when the capacity of the equipment has to be changed.

Fig. 2 shows the same architecture in which the EPC is virtualized. In this case, either all functions in the EPC, or only a few of them are transferred to a shared (cloud) infrastructure. Virtualizing the EPC could potentially lead to better flexibility and dynamic scaling, and hence allow TSPs to respond easily and cheaply to changes in market conditions. For example, as represented by the number of servers allocated to each function in Fig. 2, there might be a need to increase user plane resources without affecting the control plane. In this case, VNFRs such as a virtual MME may scale independently according to their specific resource requirements. In the same way, VNFRs dealing with the data plane might require a different number of resources than those dealing with signaling only. This flexibility is expected to lead to more efficient utilization of resources. Finally, it also allows for easier software upgrades on the EPC network functions, which would hence allow for faster launch of innovative services.

_____

†A zettabyte is equal to 1000 exabytes, which is equal to 1000 billion gigabytes.
be significantly reduced. Finally, as a result of deploying NF on virtualized resources, NFV promises the ability to scale resource allocations up and down as traffic demands ebb and flow. This is expected to potentially reduce the number of physical devices operating at any point, and hence reduce TSPs’ energy bills. Since energy bills represent more than 10% of TSPs’ OPEX, reduced energy consumption is one of the strong selling points of NFV. This means that energy efficiency will be critical key performance indicator for NFV.

In this paper, we use Bell Lab’s GW ATT tool to estimate the energy savings that could result from the three main uses cases (VEPC, VRAN and VCPE) of NFV. The tool uses state-of-the-art network and traffic models to estimate not only the expected changes in user traffic, but also the resulting effect on network equipment in terms of energy consumption. It also includes possibilities to set different network architectures, including possibilities to select virtualized networks. We believe that an understanding of the actual savings in energy that can be derived from virtualizing different parts of the network is very important in guiding TSPs to make investment decisions. In particular, at this early stage of NFV where virtualized functions will exist along side those running on specialized hardware, it could be important for TSPs to determine which part of the network could be virtualized first based on the expected gains.

The rest of this paper is organized as follows: section II discusses related work, while the estimation tool used in described in detail in section III. In section IV we present and discuss results from utilizing the tool, before concluding the paper in section V.

II. RELATED WORK

China Mobile recently published their experiences in deploying a Cloud Radio Access Network (C-RAN). One of the tests was performed on their 2G and 3G networks, where it was observed that by centralizing the RAN, power consumption could be reduced by 41% due to shared air-conditioning. In addition, Shehab et al. analyzed the technical potential for energy savings associated with shifting U.S. business software to the cloud. The results suggested a substantial potential for energy savings. In fact, the authors noted that if all U.S. business users shifted their email, productivity software, and CRM software to the cloud, the primary energy footprint of these software applications could be reduced by as much as 87%.

DROP is a middleware platform which was originally aimed at creating software routers on top of commodity servers, while hiding the complexity of its modular architecture to control-plane applications and system administrators. It has been recently extended to DROPv2 which is focused on implementing more efficient power management in SDN. The idea of DROPv2 is to periodically calculate, based on total traffic, the number of forwarding elements that should be put into a standby state, and how the remaining ones should share the available traffic. Both these research activities do not specifically look at virtualized NFs.

Yathirah et al. identified a need to extend Openstack to include a scheduler which is energy-aware. The idea is to use analytics to determine the current status, and use it to schedule Openstack resources. However, the authors do not give any details about solution or the actual resulting energy savings. Instead of relying on generic servers, uses Application Specific Instruction-set Processor (ASIP) to achieve an energy-efficient deployment of a virtualized deep packet inspection (DPI). Based on experiments, the authors argue that they are able to achieve superior energy efficiency compared to DPIs deployed on commodity servers. However, this would be against the general concept of NFV as it runs the functions on specialized servers. In addition, it only considers a specific, light function. Results could be different on a large scale, for more processing intensive functions or chains of functions.

To the best of our knowledge, the work presented in this paper is the first study and discussion of the forecast energy efficiency for any of the NFV use cases. Given the importance of energy efficiency as a key performance indicator for NFV, we hope that the results of this study can help TSPs make necessary investment decisions with regard to the evolution towards NFV. We also hope that this can direct researchers towards proposing energy-aware algorithms to those parts of the network where savings may be highest. It is important to note that the GWATT tool used in this paper has NOT been developed by the authors. Our contribution is in using the tool—which is freely available, and aimed at determining energy efficiency in multiple network scenarios—to study, simulate, summarize and discuss results specific to chosen NFV use cases.

\[ \text{This figure can raise up to 45% for TSPs in less developed countries.} \]
III. Bell Labs’ GWATT

Bell Labs’ GWATT tool [7] is an important step in determining the possible effect, on energy consumption, of the evolution to NFV. Based on some forecast for traffic growth, the tool is able to show the effect of virtualizing different network functions. It also gives a cumulative forecast on energy savings over a five year period. As shown in Fig. 3, the tool is able to show the effect of virtualizing different network models and technologies and hence analyze its energy impact. As shown in the top left corner of Fig. 3, the figure is from an analysis of the energy consumption of a network in which the EPC is virtualized. The results from the analysis are shown at the bottom of the figure. To discuss these results, we start by noting that GWATT is based on two main models:

1) A traffic data model. For each domain of the network, the tool uses a traffic volume expressed in exabytes/month. The traffic volume is based on

| Domain                  | Traffic (Exabytes per month) | Total Efficiency (MBITS/J) | Total Power (MWATTS) |
|-------------------------|------------------------------|----------------------------|----------------------|
| Home & Enterprise       | 59.85                        | 0.012807                   | 14,422.00            |
| Access & Aggregation    | 62.36                        | 0.009887                   | 27,501.50            |
| Metro                   | 163.07                       | 0.760444                   | 661.85               |
| Edge                    | 83.26                        | 0.325516                   | 756.66               |
| Core                    | 166.53                       | 1.770111                   | 290.36               |
| Service Core & Data Centre | 381.47                   | 0.033012                   | 29,608.69            |

Fig. 3. Network Details from GW ATT for Virtualized EPC

| Domain                  | Traffic (Exabytes per month) | Total Efficiency (MBITS/J) | Total Power (MWATTS) |
|-------------------------|------------------------------|----------------------------|----------------------|
| Home & Enterprise       | 59.85                        | 0.012807                   | 14,422.00            |
| Access & Aggregation    | 62.36                        | 0.009887                   | 27,501.50            |
| Metro                   | 163.07                       | 0.760444                   | 661.85               |
| Edge                    | 83.26                        | 0.325516                   | 756.66               |
| Core                    | 166.53                       | 1.770111                   | 290.36               |
| Service Core & Data Centre | 381.47                   | 0.033012                   | 29,608.69            |
extrapolated traffic projections performed by Bell Labs and based on real data. For example, as shown in Figure 3 when the EPC is virtualized, the traffic associated to the network core domain is 166.53 exabytes/month, while that associated with the data center and service core is 384.93 exabytes/month.

2) A network element efficiency model. An efficiency expressed in megabits/joules is associated with each domain. For instance, after virtualizing the EPC as shown in Figure 3 it can be observed that the power efficiency for the core network domain is 1.77011 megabits/joules while that for the data center and service core domain is 0.043241 megabits/joules. We also note that with this efficiency, the total power consumed by all network elements in each of these domains is 290.36 Megawatts and 29,608.70 Megawatts respectively.

Based on the traffic and efficiency data above, and taking into consideration the underlying network model, GWATT computes the total end-to-end network power (expressed in megawatts), energy consumption (in gigajoules). In addition, by comparing all these values with those from a baseline network the tool is able to determine the forecast savings in energy by changing a given part of the end-to-end network. It can be observed that by virtualizing the EPC, 32.82 Megawatts of power are saved at the edge while 20,442.95 Megawatts are saved in the data center and service core. These traffic, consumption and efficiency values for the baseline network are shown in Table I.

In order to use the tool, some choices need to be made. These settings relate to the network size, traffic growth rate, number of applications and network architecture. The default network size worldwide, but this can be set to one of North America, Small Country, Large Country etc. The tool also allows one to set the rate at which traffic grows in the selected network size as a percentage. The traffic applications setting can be used to select one or more applications of the available application which include online gaming, file sharing, video, web, business video, business file sharing and business web. The default setting utilizes all the applications, which represents 100% of the set traffic. Finally, it is possible to select a particular network architecture. While the tool includes many architectures, the focus of this paper is on the virtualization of the CPE, RAN and EPC. The VEPC virtualizes mobile network functions by implementing a cloud-based EPC architecture that uses resources more efficiently and improves agility and scale. The VRAN virtualizes and centralizes the consolidation of eNodeB Baseband Units (BBU) from multiple cell sites to improve scale and efficiency. Finally, the VCPE virtualizes the residential gateway of the CPE to reduce complexity, energy consumption and deployment cost (CAPEX).

3A baseline network is one where all functions are run in physical equipment, using the tool’s default technologies and settings.

4Unless stated otherwise, the results presented in this paper are based on GWATT’s default settings.
IV. RESULTS

The results for the three NFV use cases are summarized in Table I. The power savings are determined in comparison with the results from the baseline network in I. These results, including others with different setting are presented in graphical form in Figs. 4 - 12.

From Fig. 4 and 5, it can be observed virtualization of network functions leads to a reduced power consumption. However, it can be noted that VEPC achieves the highest reduction in power consumption, followed by VRAN and then VCPE. The reason for superiority of virtualizing the EPC could be due to the higher traffic that goes through the EPC domain as shown in Table I. This means that there are many network elements in the core, which presents more possibility to reduce energy not only by consolidating these network elements, but by also reducing the required capacity through the better resource management presented by virtualization. In fact, this can be further proved by looking at Fig. 6 which shows how the power savings change if the rate of traffic growth is increased. It can be noted while the VEPC saves 21% of power even when traffic is reducing, the savings increase to about 23% in conditions of traffic growing at 50%.

In Figs. 7, 8 and 9 we show how virtualizing a given network domain affects the energy consumption of the other domains. As expected, it can be observed that each of the three use cases considered have the strongest effect in terms of power savings on the respective domain which has been virtualized.

In Fig. 10 we show the total network traffic carried in each use case, which is broken up across all the contributing network domains in Fig. 11. As earlier noted, it is evident that the service core and data center domain has the highest traffic in all cases. However, it is interesting to note that while this traffic is almost constant across all the other network domains for the 4 cases evaluated (including the baseline), it varies considerably in the metro domain. In fact, the total traffic traffic in Fig. 10 takes on a similar profile to the traffic in the metro network. This could mean that the virtualization of any of the considered use cases could benefit more from virtualization of the metro network. The idea is that if the traffic increases, there is potential to benefit for improved resource management through virtualization of resources.

Finally, Fig. 12 shows the total efficiency resulting from virtualizing of the considered domains. Just like the savings in power consumption, and probably for the same reasons, the EPC benefits more from virtualization than both the RAN and CPE. The fact that the CPE lags in terms of energy efficiency and/or power savings in all these results is not so surprising. CPEs normally consume little energy and the advantages from virtualizing the CPE could be much more on the side of CAPEX (as the boxed would become cheaper on a large scale). The OPEX savings will mainly be due to ability to make remote configurations as well as upgrades.

However, while the results presented in this paper show promising values on energy savings, a lot of work will likely need to be done to achieve them in practice. The fact that NFV makes data centers an indispensable part of telecommunication networks raises questions on the actual value of savings in energy consumption (if any) that will be achieved. According to an analysis in the SMARTer 2020 report from GeSI [15], the total electricity requirements of the cloud (including data centers and networks, but not devices) in 2011 was 684 billion kWh. The cloud, if it were a country, would rank 6th in the world in terms of its energy demand, and yet this demand is expected to increase by 63% by 2020 [16]. While some progress on energy efficient cloud computing has been made, the fast growing energy needs of data centers continue to
receive a lot of attention [17], [18]. Therefore, there is an urgent need to evaluate whether NFV will be able to meet these energy saving expectations, or whether—the NFs—the energy consumption will just be transferred to the cloud.

V. CONCLUSION

In this paper, we have attached values to the energy saving expectations of some NFV use cases. Using Bell Lab’s GWATT tool, we have determined that by virtualizing the EPC, RAN and CPE we can reduce network power consumption by 22%, 17% and 3% respectively in comparison to a baseline worldwide network consumption of 93,722.80 Megawatts. We have also shown that this saving in energy consumption could even be higher in cases of increasing traffic across the network. Finally, we have noted that virtualizing each of these network domains does not significantly affect the traffic passing through the domain, instead affecting the traffic through the metro network. For this reason, it’s important to consider the virtualization of the metro network as a necessary aspect of virtualizing other network domains.

However, while GWATT is an important step in attaching numbers to the energy savings expected from NFV, it can still be improved. In particular, it currently does not have a detailed technical documentation. In addition, Cisco’s visual networking index forecasts that annual global IP traffic will reach 1000 exabytes in 2016. Based on this, the (monthly) traffic values in Tables A1 and A2 seem to be so high, yet the tool does not provide a possibility to know how these values are derived. Without a doubt, the energy efficiency of cloud based NFs will continue to receive attention. NFV will put InPs under even more pressure to manage energy consumption not only to cut down energy expenses, but also to meet legal, regulatory and environmental requirements. Therefore, approaches for placing network functions [19], [20] as well as those that dynamically manage the network resources [21], [22], [23], [24] must be geared towards energy efficiency just as they are focused on efficient resource utilization.

ACKNOWLEDGMENT

This work is partly funded by FLAMINGO, a Network of Excellence project (318488) supported by the European Commission under its Seventh Framework Programme, and project TEC2012-38574-C02-02 from Ministerio de Economía y Competitividad.

REFERENCES

[1] Cisco, “The Zettabyte Era: Trends and Analysis,” Tech. Rep., May 2015.
[2] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, “Network function virtualization: State-of-the-art and research challenges,” 2015.
[3] R. Guerzoni, “Network Functions Virtualisation: An Introduction, Benefits, Enablers, Challenges and Call for Action. Introductory white paper,” in SDN and OpenFlow World Congress, June 2012.
[4] R. Mijumbi, J. Serrat, J.-L. Gorricho, S. Latré, M. Charalambides, and D. Lopez, “Management and orchestration challenges in network function virtualization,” 2015.
[5] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and S. Davy, “Design and Evaluation of Algorithms for Mapping and Scheduling of Virtual Network Functions,” in IEEE Conference on Network Softwarization (NetSoft). University College London, April 2015.

[6] “The 3rd Generation Partnership Project (3GPP),” http://www.3gpp.org/about-3gpp/about-3gpp 2015. Accessed: February, 10 2015.
[7] Bell Labs, Alcatel Lucent, “G.W.A.T.T. (Global What if Analyzer of Network Energy Consumption),” Bell Labs application able to measure the impact of technologies like SDN & NFV on network energy consumption. White Paper. http://gwatt.net/intro/ 2015.
[8] ETSI Industry Specification Group (ISG) NFV, “ETSI GS NFV 001 V1.1.1: Network Function Virtualization. Use Cases,” www.etsi.org/deliver/etsi_gs/NFV/001/059/000/01.01.01_601gs_NFV091v010101p.pdf; October 2013.
[9] C.-L. I. J. Huang, R. Duan, C. Cui, J. Jiang, and L. Li, “Recent Progress on C-RAN Centralization and Cloudification,” Access, IEEE, vol. 2, pp. 1030–1039, September 2014.
[10] E. Masanet, A. Shehabi, L. Ramakrishnan, J. Liang, X. Ma, B. Walker, V. Hendrix, and P. Mantha, “The Energy Efficiency Potential of Cloud-Based Software: A US Case Study. Lawrence Berkeley National Laboratory, Berkeley California.” 2013.
[11] R. Bolla and R. Bruschi, “An open-source platform for distributed linux software routers,” Comput. Commun., vol. 36, no. 4, pp. 396–410, Feb. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2012.11.002.
[12] R. Bolla, C. Lombardo, R. Bruschi, and S. Mangialardi, “DROPv2: energy efficiency through network function virtualization,” Network, IEEE, vol. 28, no. 2, pp. 26–32, March 2014.
[13] Yathiraj, Udupi and Dutta Debo and Krishnan, Ram, “Increasing Network and Energy Efficiency via Optimized NFV Placement in Openstack Clouds,” OpenStackSummit,Africa. May 2014.
[14] K. Yamazaki, T. Osaka, S. Yosuda, S. Ohteru, and A. Miyazaki, “Accelerating sdn/nfv with transparent offloading architecture,” in Presented as part of the Open Networking Summit 2014 (ONS 2014). Santa Clara, CA: USENIX, 2014. [Online]. Available: https://www.usenix.org/conference/ons2014/technical-sessions/presentation/yamazaki.
[15] “Global e-Sustainability Initiative (GeSI) SMARTer2020,” http://gesi.org/SMARTer2020 2015. Accessed: June, 08 2015.
[16] GREENPEACE, “‘Clicking Clean: How Companies are Creating the Green Internet,’” www.greenpeace.org April 2014.
[17] The Natural Resources Defense Council (NRDC), “Data Center Efficiency Assessment. Scaling Up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers. Issue Paper.” August 2014.
[18] A. Beloglazov, R. Buyya, Y. C. Lee, and A. Y. Zomaya, “A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems.” Advances in Computers, vol. 82, pp. 47–111, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=3968122