An Energy Efficient Architecture for 5G Network Management

Kieran Sullivan*†, Michael Taynnan Barros*†, Angel Martin‡, Haytham Assem§

*TSSG, Waterford Institute of Technology, Ireland
†Department of Computing, Math and Physics, Waterford Institute of Technology (WIT), Waterford, Ireland
‡ Digital Media department, Vicomtech-IK4, San Sebastian, Spain
§ Cognitive Computing Group, Innovation Exchange, IBM, Ireland
Email: ksullivan@tssg.org, mbarros@tssg.org, amartin@vicomtech.org, haythama@ie.ibm.com

***Abstract*—5G networks will require high data rates, high quality-of-service (QoS), and high energy efficiency in a package that can accommodate virtualized network functions and software-defined networks. Current energy efficient approaches for wireless networks were not designed for such high performance demands, meaning network resource allocation is often inefficient. In this paper, a novel energy efficient architecture is proposed that can dynamically allocate resources based on both current traffic demands and pre-defined energy policies. This approach is empowered by machine learning algorithms and predictive analysis. Several 5G demands are addressed by the proposed architecture.**

***Index Terms*—5G; Energy Efficiency; Architecture; Machine Learning; Cognitive Networks**

I. Introduction

Wireless networks have massively increased by a factor of 100 in the past decade, and even more growth is predicted in the new generation of mobile networks - the **fifth generation (5G)** [1]. A significant change to the mobile communication industry and related research is expected with respect to service demands from users. A natural evolution of 4G is unlikely to suffice for next-generation applications and services currently envisioned (and even developed in some instances). Therefore, new technologies and novel designs offer future networks an opportunity to make this required transition. In parallel however, a more fundamental concern relates to the energy requirements for this increase in user services and the 5G networks that will support them.

Among the many proposed architectures for 5G, the most appealing is *SoftAir* [2], [3]. Decoupling between data and control plane is presented with high programmability and flexibility. The data plane is open, programmable and virtualizable network forwarding infrastructure. And the control plane presents network managements tools and customized applications of service providers or virtual network operators. SoftAir architecture has the potential to guarantee the properties required for 5G networks [3].

An estimate of 190 exabytes in 2018 and 500 exabytes in 2020 of IP data transfer has been suggested [4]. The wireless research community must not only provide quality-of-service (QoS) provisioning, but also energy efficiency for this huge amount of data [1]. Three main issues are to be addressed regarding energy efficient communications networks: (1) Environmental concerns, where finite resources increase harmful emissions, (2) Operating costs, which telecommunication providers seek to reduce in order to offer more competitive services to their customers, and (3) High performance of the network itself.

Historically, improving hardware efficiency helped increase energy efficiency at device and infrastructure levels in mobile communications. Such gradual hardware advances will not reduce energy consumption sufficiently for 5G, given the expected increase in devices, data rates, and coverage. The hardware-based approach fails to address issues (2) and (3) above. A software-based approach offers a better solution to improve overall network management. Additionally, such an approach better fits the proposed 5G architecture, where data and control planes are decoupled, such as with SoftAir.

In this paper, an energy efficient architecture for 5G network management is presented, based on an extensive analysis of the network's requirements. This software-based approach will be localized in the control plane of the 5G network. The infrastructure resources will be adjusted according to (a) energy policies, (b) QoS requirements of the data being transferred, and (c) network resource conditions; thereby, addressing issues (1), (2) and (3) outlined above.

5G networks rely on cognition in the network to obtain flexibility. Therefore, the architecture will be based on the following pillars:

- **Machine Learning:** Certain components of the architecture will be able to learn about the patterns of the data traffic in the networks and provide efficient models to precisely quantify the infrastructure demands.
- **Predictive Analysis:** The architecture will run a predictive analysis on possible resource allocation setups evaluating its performance in a long run, providing more efficient and safer outputs.

The added cognition to the energy efficiency architecture will enhance the performance of power consumption management with minimum interference on the network data transfer performance. Dynamic resource allocation is thus achieved, being an advantage over the existing energy efficient resource allocation approaches [5], [6], [7].

This paper is laid out as follows: Section II outlines what engineering requirements 5G is expected to encompass; section III describes the various current approaches to energy efficiency in mobile networks; section IV combines the previous sections and offers protocols to ensure 5G is energy efficient; section V concludes the work.

II. Expected 5G Requirements

A. What will 5G Look Like

From the end-user point-of-view, 5G will look and feel no different to the current networks. It will however, significantly enhance their mobile experience, in terms of stable connectivity, new services, data rates, costs, and devices they can connect to the network.

A simple evolution of LTE/4G will not meet these user-driven demands. Instead, 5G will require a paradigm shift that encompasses significant densification of base stations and devices, high carrier frequencies with increased bandwidth, and a core backhaul that offers flexibility and intelligence previously thought to be unmanageable. A heterogeneous, virtualised approach would fit well with these requirements. With respect to the numbers involved, the amount of IP data handled by mobile networks is expected to go from 3 exabytes in 2010, to 500 exabytes in 2020, with video and multimedia traffic the chief drivers of this 100-fold increase [8].

For the increases expected in devices and data, the energy required to run the network could reasonably be expected to increase also. However, current trends indicate that energy levels in 5G must decrease both from the capital investment (CAPEX) and operating costs (OPEX) point-of-view.

B. Sample 5G Scenarios

In this section, we describe some scenarios to illustrate what 5G might look like for end-users.

1) Large-scale Event: In the run up to Christmas, many towns host traditional markets which take place across outdoor and indoor environments. The number of people attending varies during the day and the evening, depending on the weather, attractions such as live music and street performances, and external events such as televised sports, to name but a few. People will require a seamless connection to the network, and as a result, the number of devices requesting connection will increase and decrease at various times. Thus, the core network must provide the relevant cells with extra resources as required, without bluntly over-provisioning when the number of connected devices reduces during off-peak times.

2) Machine-2-Machine (M2M): While often considered the Internet-of-Things, a more realistic environment for M2M communication would be a smart factory where the factory itself is essentially considered a black box. In this scenario, the owner wants to ensure seamless interaction between his factory and its surroundings, where supply chain management is a key function. A number of independent SMEs supply various products and services to the factory, and each engages in a common platform which virtualises all factory components. This allows the factory owner to track his suppliers processes (i.e. his inputs), while the virtual machine infrastructure is shared among all federated members. The factorys network must facilitate the secure sharing of data, the integration of service platforms and devices, and subsequent changes in suppliers and the factorys level of trust in their reliability.

3) Dense Smart Cities: A massive number of devices are expected to connect to the 5G network. The diversity of connected devices is also expected to increase, and each category of device will have its own very different communications requirements. Issues will include when to connect, how to connect, with which priority, how long to connect, etc. This will be especially relevant in smart cities which will deploy large numbers of stationary sensors and actuators, in addition to moving sensors in cars and public transport. User devices will also vary from smart-phones, watches, IoT wearables, etc. The network must also support devices linked to private networks (e.g. enterprise networks, manufacturing, logistics), and devices related to security and safety (e.g. webcams, traffic detection). Over-provisioning cannot continue due to spectral efficiency and energy efficiency, plus the network architecture must consider which operations to execute at the edge of the network and which require centralised control.

4) Emergency Communication: Consider the unfortunate collapse of a stage at a music concert. Paramedics attending the scene use augmented reality services that help deliver medical aid, but this requires considerable bandwidth. The network must facilitate this high priority service via dedicated authentication and authorisation. It must also dynamically allocate bandwidth and related network resources to guarantee the service. This may include removing bandwidth from other users, but disruption should be minimised where possible. Similar to the Large-scale Event scenario, the network must return to normal operation once the accident scene is cleared.

5) Personal Security: Currently, connected devices are protected essentially by installing specific tools on each device. This raises several issues including the computational power required to run such tools, variation in platform capability, and varying levels of access required on the same device. In 5G, network providers may offer a suite of security functions to users. For example, the provider defines the business and legal policies applicable to a set of users and a set of different locations. Corporate administrators define their policy within their corporate network (e.g., office, hotel, mall, airport). Users specify personal security preferences, which follow them in public spaces and update automatically when they arrive to work. The network deploys and configures security functions at designated locations, and must react to specific changes in the conditions (e.g. new policies, raised level of attack) at the different locations.

6) Connected Cars: Autonomous connected cars are the future of transport infrastructures over cities roads or highways. Congestion control of cars traffic will be available, in which automatic rerouting of cars will decrease average trip time in such dynamic mobility scenario. Roads also will be safer using movement prediction techniques among cars and between cars and citizens. Also, high quality multimedia streaming is required for several applications within the range of entertainment to health-care. These examples of functionalities for the connected cars scenario will only be possible with a technology that allows high data rate alongside with efficient QoS provisioning. 5G is potentially fitted to the connected cars scenarios demands as an alternative for such transmission requirements. This technology will allow intense information exchange between cars while also maintain a reliable and effective communication system.

C. Requirements

While the above scenarios provide illustrations of various 5G scenarios where the network is expected to react and adjust, the following defines some envisioned parameters within which the 5G network must operate.

1) Data rates: Supporting mobile data traffic is a huge requirement for the 5G network, and a number of specific metrics are expected [1]. These include:

- *Aggregate capacity*: The total amount of data the network can serve, expressed in bits/second per unit area. Figures here will depend on the providers cells, including pico/femto vs. macrocells, etc.
- *Edge rate*: The minimum data rate a user should expect. While estimates vary depending on load, cell size, etc. the consensus seems to hover around 100Mbps.

2) Latency / Quality-of-Service (QoS): Current 4G latencies are approximately of the order of 15ms, but consumer-led demands in 5G, such as multi-user gaming, means a latency of less than 10ms will be required. This will impinge on the design of a number of layers of the protocol stack.

3) Energy Efficiency: At a minimum, energy costs should not increase in 5G networks, despite the massive increase expected in devices connecting to the network and the subsequent rise in network deployments. This includes more active base stations and the additional strain put on the backhaul from edge to the core network. If data rates increase x100, then the cost-per-bit must decrease x100.

4) Flexible Functionality: The scenarios above all highlight the need for additional intelligence in the network. The extra services and data envisioned for 5G means the network must take account of and react to successive changes in user demand. To do this, a centralised view of the network will considerably assist matters as will enhance machine-learning (analysing traffic patterns and making provisioning predictions based on same). The mainstream incorporation of network function virtualization (NFV) and software-defined networks (SDN) must be considered since they offer increased flexibility and deployment options.

It should be noted that different scenarios and instances in 5G will have different requirements. That is, in some cases, a higher latency might be acceptable as long as the connection is guaranteed (e.g. Emergency Communication). At other times, low latency may take primacy over the networks aggregate data capacity (e.g. Machine-2-Machine).

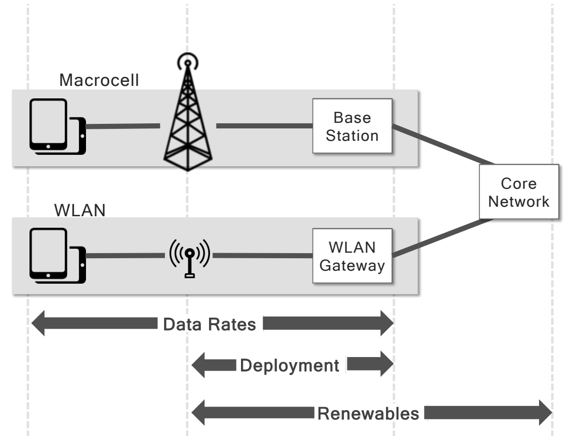

Fig. 1: Approximate location of current software-led energy efforts

III. Current Approaches to Energy Efficiency in Mobile Networks

Improvements in hardware efficiency play a significant role in increasing energy efficiency at device and infrastructure levels in mobile communications. However, incremental hardware advances will not reduce energy consumption sufficiently for 5G, given the expected increased data rates, coverage and number of devices. Thus, a major challenge is to use a software-based approach to improving overall network management. Current approaches to energy efficiency in mobile networks attempt to make the network smarter, and they fall broadly into three main categories. Samples from each category are described below, and Figure 1 shows the approximate location of these efforts along two basic network types.

A. Data Rates

Work by [6] across a range of maximum transmit powers showed that relatively small reductions in the data rate resulted in significant energy savings. Results indicated that maximum energy efficiency approximately equates to maximising spectral efficiency for low transmit power. However, this technique fails to provide latency/QoS and data rate requirements for the 5G network.

B. Deployment

Cell wilting (sleep mode) and cell blossoming (re-awakening) are proposed by [9], as a means to switch off and switch on subsets of base stations within a wider a cell architecture. This technique also does not consider the latency/QoS and data rate requirements of the 5G network.

C. Renewables

Harvesting a combination of wind and solar energy to power individual base stations is one of a number of renewable-related energy options for 5G [10]. Similarly, optimising circuit consumption, battery storage and data rates while combining renewable and non-renewable energy sources are considered by [11]. Even tough these techniques are important, they do not provide efficient flexible functionalities to the 5G network.

IV. Energy Efficiency Architecture for 5G

An architecture is proposed to address both 5G network requirements and energy efficient demands. This software-based approach takes account of the different topics discussed in the paper, incorporating current energy efficiency efforts as well as the potential offered by the various technologies envisioned for 5G.

One of the main requirements of the 5G technology is to provide energy efficient resource management even with the high volume of users in the network. Network resources in the data plane are dynamically adjusted based on the network information data, processed in the control plane. This cognitive feedback system allows improvements in performance, since machine learning-based predictive analysis alongside with network policies in the control plane adjust precisely the data plane.

In order to implement energy efficient systems in 5G, one must consider the insertion of such in the control plane. The control plane will provide both processing capabilities and important input information about the data plane. This is required to consider the current network state and its history for the machine learning predictive analysis, and also for any cognitive-based functionality. High performance energy efficiency can, therefore, be achieved due to near-optimum resource management obtained from the cognitive-based system.

An architecture was created to substantiate and validate the aforementioned idea. Fig. 2 presents the energy efficient architecture. It is divided into five different blocks listed below:

- *Traffic Classifier* – Current traffic passing through the network needs to be classified before a QoS class is attributed to it.
- *QoS Class* – A QoS Class is attributed to the classified traffic and passed to the energy efficient protocol. In this way, the energy efficient protocol knows the data rate, latency and QoS requirements of the application.
- *Monitoring/Performance Tool* – The performance of the network is going to be monitored by some packet-sniffing tool.
- *Network Configuration* – The network configuration is also going to be monitored.
- *The energy efficient (EE) protocol* – Commands the data plane with a network protocol that regulates energy usage in the network infrastructure, network interfaces and network nodes. Regulating messages with power in active modes, power in low power modes and time in active modes, can enhance energy policies performance in the network. Such information will be defined as a result of network information processing and predictive analysis. Information about the current network traffic, network history, and service layer agreements will help decision making of the energy efficient protocol.

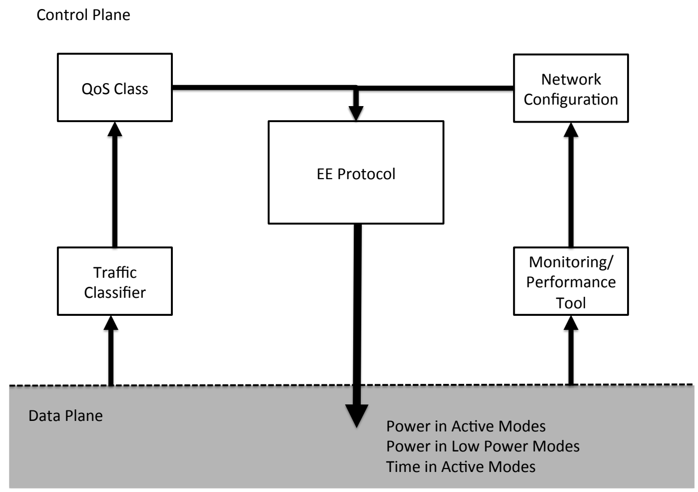

Fig. 2: Control plane accommodates the energy efficiency architecture (EEA) in a 5G platform. The EEA have inputs from the control plane that process data acquired from the data plane

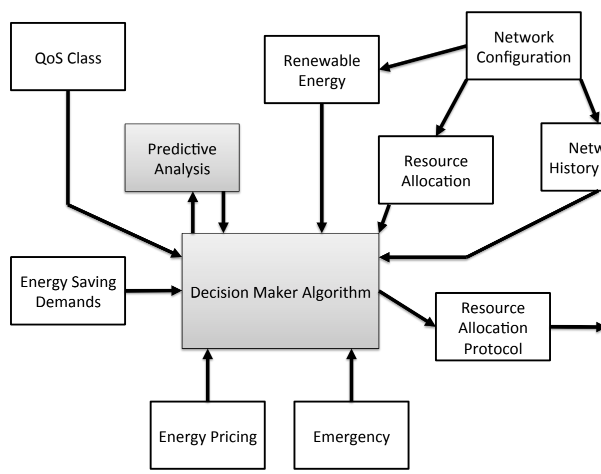

Fig. 3: Closer look at the EE protocol. Nine elements give relevant input information for a decision maker algorithm set up a desire resource allocation configuration. Machine learning components are in grey.

In the next subsection, the energy efficient (EE) protocol is explored in detail.

A. The energy efficient (EE) protocol

To ensure flexibility and scalability, the EE protocol has nine unique elements. The most important is the decision maker algorithm which will group information about the other elements and give a precise response to the resource allocation protocol. This will be enabled via machine learning algorithms. A close look at the protocols is presented in Fig. 3, with components based on machine learning highlighted in grey. The components of the EE protocol are listed below.

- *Decision Maker Algorithm* - Responsible for gathering all the data about the network and aligning with both the energy and performance demands. The output will be forwarded to the resource allocator.
- *Resource Allocation Protocol* - Based on the resource set up given by the Decision Maker Algorithm, the resource allocation protocol will transmit control information to the data plane to adjust the network configuration.
- *QoS Class* - Different applications will generate different traffic requirements that are translated into QoS classes. The Decision Maker Algorithm needs to know which traffic is going through the network for setting up the required energy and performance demands by the application.
- *Network Configuration* - The network configuration from the network in the data plane.
- *Renewable Energy* - The Decision Maker Algorithm needs Information about the renewable energy resources available in the data plane network. This type of energy can be used depending on the QoS Class and the Energy Saving Demands.
- *Resource Allocation* - The Decision Maker Algorithm for overviewing different setup possibilities needs Information about the current resource allocation of the data plane network.
- *Network History Profile* - Periodic traffic traces collected for regular traffic analysis can be used to create a network history profile. This will allow the Decision Maker Algorithm to conciliate the current network traffic with the regular traffic profile in the network in a given time.
- *Predictive Analysis* - Because the control plane is located in high performance computing data centres, prediction algorithms can be used to quick analyse different options of resource allocation in the network. This will help the Decision Maker Algorithm with choosing the best resources for the network and the applications running through it.
- *Energy Pricing* - Energy pricing policies can be created and independently updated from the Decision Maker Algorithm. This will allow dynamic negotiation of prices between costumers, providing faster energy auctioning.
- *Emergency* - Emergency policies influence the resource allocation of the data plane network for greater network management provisioning.

Based on the 5G requirements in Section II-C, a discussion about the architecture design and those requirements is now presented. Data rate, latency and QoS requirements are addressed by the feedback mechanism of the QoS class. This will feed enough information to the Decision Maker Algorithm to address the 5G network performance requirements. Energy Pricing and Renewable Energy are going to provide energy policies to the Decision Maker Algorithm to consider the best power saving requirements. Finally, the other components, especially the emergency one, provide enough flexibility to the architecture that will be needed for a programmable and scalable scenario such as the 5G network.

Using the proposed energy efficient architecture, the 5G network is then capable of providing high data rate transmission while managing the network resources for energy saving to a near-optimum performance. Also, the aggregated components provide enough capabilities to be integrated with other network function models in the control plane.

V. Conclusion

Current energy efficient approaches for wireless networks are unlikely to address the expected requirements of 5G - there is a need for a software-based architecture to better manage energy. A novel energy efficient architecture was designed based on 5G requirements, and relying on network functionality such as decoupling between data and control plan. The proposed architecture is based on machine learning algorithm and predictive analysis that dynamically allocate network resources to resolve both the current traffic demands and energy policies. Efficient data rate, latency and QoS can be achieved with openness, programmability and flexibility. Future work includes implementation and validation of the proposed architecture.

Acknowledgements

The work described in this paper is facilitated by the CogNet project (671625), which is funded under the European Commissions Horizon 2020 programme.

References

[1] J. G. Andrew, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal Selected Areas in Communications*, vol. 32, no. 7, pp. 1065–1082, 2014.

[2] I. F. Akyildiz, S.-C. Lin, and P. Wang, "Wireless software-defined networks (w-sdns) and network function virtualization (nfv) for 5g cellular systems: An overview and qualitative evaluation," *Computer Networks*, vol. 93, Part 1, pp. 66 – 79, 2015.

[3] I. F. Akyildiz, P. Wang, and S.-C. Lin, "Softair: A software defined networking architecture for 5g wireless systems," *Computer Networks*, vol. 85, pp. 1 – 18, 2015.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?" *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1065–1082, 2014.

[5] F. Meshkati, H. V. Poor, and S. C. Schwartz, "Energy-efficient resource allocation in wireless networks," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 58–68, May 2007.

[6] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink ofdma networks with base station coordination," *IEEE Transactions Wireless Communications*, vol. 14, no. 1, pp. 1–14, 2015.

[7] S. Buzzi and H. V. Poor, "Joint receiver and transmitter optimization for energy-efficient cdma communications," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 3, pp. 459–472, April 2008.

[8] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update," 2013-2018.

[9] A. Conte, A. Feki, L. Chiaraviglio, D. Ciullo, M. Meo, and M. Marsan, "Cell wilting and blossoming for energy efficiency," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 50–57, 2011.

[10] H. S. Dhillon, Y. Li, P. Nuggehalli, Z. Pi, and J. Andrews, "Fundamentals of heterogeneous cellular networks with energy harvesting," *IEEE Transactions Wireless Communications*, vol. 13, no. 5, pp. 2782–2797, 2014.

[11] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in ofdma systems with hybrid energy harvesting base station," *IEEE Transactions Wireless Communications*, vol. 12, no. 7, pp. 3412–3427, 2011.