BigOptiBase: Big Data Analytics for Base Station Energy Consumption Optimization

Evdokia Kassela, Nikodimos Provatas, Asterios Tsiourvas, Ioannis Konstantinou, Nectarios Koziris
Computing Systems Laboratory
National Technical University of Athens
Athens, Greece
{evie, nprov, atsiour, ikons, nkoziris}@cslab.ece.ntua.gr

***Abstract*—Mobile Network Operators develop new technologies, as the 5G network, to handle the constantly increasing network traffic, while they put less effort on optimizing their operations. However, more energy efficient approaches are essential for cost reduction and compliance with energy footprint principles. Network usage and IoT data produced in the base stations can be used to develop such approaches. Considering the above, the BigOptiBase platform has been designed. Through this platform we will offer a big data analytics subsystem developed to provide elastic energy efficient solutions for the base stations using data analytics and machine learning technologies.**

***Index Terms*—big data, machine learning, energy efficiency, elasticity, telecommunications**

I. Introduction

According to Gartner [1], 21 billion interconnected IoT devices will be operating up to 2020, which will have as an immediate effect the increase of the network traffic that mobile network operators (MNOs) need to handle. Industry has already developed the 5th Generation of Mobile Communications (5G), in order to cope with this constantly increasing traffic. However, radio networks are extended mostly to provide coverage and capacity to subscribers and less effort is given on optimizing their operations.

Moreover, with an annual consumption of 900 billion KW, information and telecommunication technologies are responsible for more than 10% of the global energy consumption. Besides, surveys have also indicated that more than half of the operational costs of a MNO are connected to energy consumption, while 60-80% of this energy is consumed in the base stations, which is mainly attributed to equipment that does not depend on the network traffic, as air conditions, etc [2]. MNOs are challenged to decrease their energy consumption for two crucial reasons. First of all, in this way they can reduce their operational costs and remain competitive. Last but not least, they will be compliant with principles created for the reduction of the energy footprint in CO2 emissions, as 3GP and ITU [3].

On the other hand, the network operation itself produces a large volume of data at extremely high speeds. Telemetry data which are produced by energy devices in the base stations, data from temperature and humidity sensors and data about the network usage and its performance (connected users, reasons for rejected calls, etc.) can be exploited in order to optimize the network operation in case they are collected, analyzed and properly combined.

To correctly utilize such kinds of data, traditional techniques for data management and analysis that depend on centralized systems cannot be used (e.g a database system or a single-node machine learning library), since they cannot scale to handle both the volume and the rate of the produced data. Moreover, they do not have the computational power and speed to manage the network operation in real time. The main reason is that increasing load may occur with the addition of more base stations or with an increase in the number of subscribers. Scalable solutions that can cope with such load in an elastic way, as with the addition of more computational units, are needed [4]. The aforementioned solutions utilize distributed ”shared-nothing” architectures that are executed on cloud computing infrastructures and are ”systems of distributed processing and management of big data” [5]. MNOs have started to recognise the benefits of such technologies and consider them in pilot systems [6]: until 2021 telecommunication providers are expected to invest more than 7 billion in computational infrastructure for industrial use [7].

Under the BigOptiBase project, which aims to optimize the energy consumption of base stations, we are currently building a big data subsystem that will be able to handle both the volume and the rate of the incoming data. For this purpose, we design a subsystem with the following features and utilities:

- **Scalability:** Our subsystem will be scalable in order to be able to cope with different sizes of workloads.
- **Analytics Library:** Statistics are very important for a provider to obtain an understanding of where the energy is consumed. Thus, we will provide a domain-specific analytics library that will serve the needs of this purpose.
- **Machine and Deep Learning Library:** It is very important to determine the energy policies that base stations should follow in order to be compliant with the principles mentioned before. Thus, we will build a series of predictive models that will be able to consume data and propose a series of actions for the base stations to make them more energy effiecient.
- **Elasticity:** As explained earlier, the incoming load in such systems is increasing due to the growth of base stations and subscribers. Therefore, we design our subsystem to be elastic in order to handle any incoming load.

978-1-7281-0858-2/19/$31.00 ©2019 IEEE

II. Design and Architecture

A. *BigOptiBase Platform Outline*

The overall architecture of the BigOptiBase platform, including the big data subsystem, is presented in Figure 1.

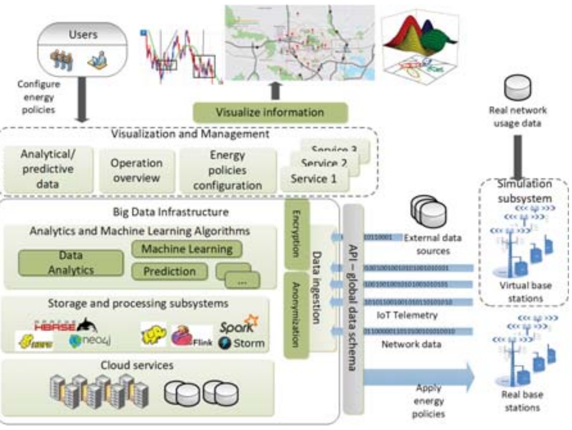

Fig. 1. BigOptiBase Architecture

Apart from the big data subsystem, BigOptiBase consists of other subsystems, as depicted in Figure 1. We briefly present these subsystems below, and how they interact with the big data subsystem. The definition of these interactions will be used to design some of our subsystem's features, which are extensively described in subsection II-B.

- **Vizualization and management subsystem**: This subsystem will be responsible for visualizing any data analytics performed on energy consumption data by the big data analytics subsystem. It will also be used to configure and inspect the energy policies of the base stations.
- **Simulation subsystem**: This part of BigOptiBase will generate base station data by taking into account real data of base stations. These will serve as example data used from the big data subsystem to create more accurate predictive models.
- **Base station agent**: This agent will be installed in base stations in the form of applications running on dedicated hardware. It will be able to send the network data to the big data subsystem and to apply the energy policies to the base stations.

B. *Big Data Subsystem Design and Features*

The big data subsystem will offer a full technological stack for analyzing a vast amount of data. Specifically, our subsystem will include state-of-the-art big data storage and processing frameworks that will be used to extract valuable knowledge aiming to develop new value-added services from the infrastructure until the algorithm layers. Moreover, we will develop big data software technologies that will be able to analyze a large amount of network and IoT energy data that are produced by base stations. Distributed Machine Learning techniques will be used to create prediction models and decisions regarding the energy policy of base stations will be derived and applied.

In order to collect and process incoming data, a data ingestion mechanism will be created (see Fig.1) which will provide an API, that will be used to receive data from any data source that will be connected to the platform. Among the data sources are both the base station agent and the simulation subsystem, mentioned in subsection II-A. Apart from the incoming data, our subsystem will export stored and processed data and predicted policies to the visualization and management subsystem. Thus, an API for data export is necessary.

Both batch and real-time data insertion techniques will be used, depending on the data type and the data generation rate. The latest trend in this area is the concept of the lambda architecture [8], which is adopted by multiple big data analytics companies worldwide [9] and will be used in the scope of this project to allow batch and streaming data storage and processing.

The subsystem will generate decisions regarding the energy policy of base stations taking into account the user configuration. Moreover, it will use connected sensors/monitoring devices to perform some initial data preprocessing and generate alerts in dangerous situations. The data will be transmitted, depending on the case, either in real time, or periodically, or upon request. Then, it will be processed either in a streaming or a batch manner on the appropriate execution engine and stored in the appropriate storage engines in order to visualize events and monitor the system's operation. In brief, the big data subsystem includes three different components as we can see in Figure 1:

1) Cloud platform: A cloud computing platform will be used as the computing and storage infrastructure where our system will be installed. Cloud techniques will be used to create a cluster of Virtual Machines with the appropriate automation and management tools (e.g. Openstack [10], Docker Swarm [11], etc.) to offer easy installation and resource configuration and elasticity through programming interfaces. For this purpose, algorithms and systems will be developed and used based on our existing research [4].

2) Big data storage and processing subsystems: A set of very popular big data frameworks will be installed on the cloud platform. These frameworks include state-of-the-art systems such as the Apache Hadoop ecosystem [12] for massive batch storage and processing, Apache Spark [13] for massive real-time processing, Apache Flink [14] and Apache Storm [15] for stream processing. These distributed frameworks will be the basic big data middleware that will ensure scalability and extensibility.

3) Analytics and Machine Learning Algorithms: A set of algorithms will be implemented to process the collected data in order to draw useful insights and for data characterization and visualization. These algorithms will be used from the value-added services and can be descriptive analytics algoritms, machive learning algorithms [16], [17] (anomaly detection, prediction, clustering), etc. Basic research [5] on network analytics with the use of big data as well as applied research (project SELIS [18]) that we have performed will be used and developed.

As mentioned in subsection II-A, dedicated hardware (based for example on Raspberry Pi microprocessor kits) will be installed and programmed in the base stations which will allow performing basic analytical computations on the data near the source where they are generated (in-situ computation – moving computation near the data), optimizing the overall system's response time.

C. Big Data Subsystem High Level Architecture

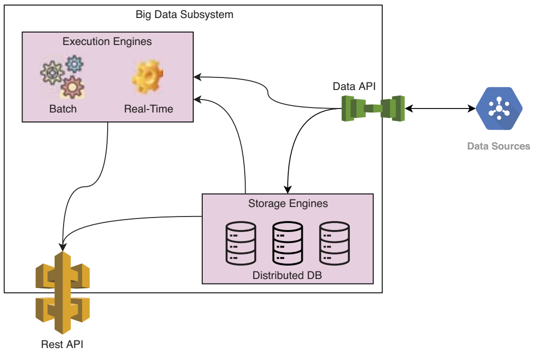

Fig. 2. Big Data Subsystem High Level Architecture

The high level architecture of the big data subsystem is presented in Figure 2. Taking into account the requirements for interacting with other data sources and subsystems, the big data subsystem offers two different APIs for interacting with it. A Rest API will be implemented for communicating with the visualization and management subsystem and a Data API will be used to receive data from various data sources, as we described in Section II-B. The storage and execution engines of the subsystem will interact with the Data API to store incoming data and process in real time input data streams in respect. The execution engines will also perform batch processing with data stored in the storage engines. The Rest API will serve data stored in the distributed databases and real-time predictions from the execution engines to external subsystems.

D. Expected Results

In this subsection, we present the expected results of the implementation of our designed subsystem. We expect to have both industrial and research results.

On the one hand, regarding the industrial results, note that our subsystem will be an independent innovative product which can be exploited in a conducive way. It is aimed for the global product marketplace for network optimization which blooms with the surge of the required network capacity. Furthermore, it can be used in the future in real networks.

On the other hand, in the latest years a lot of research is performed regarding the energy consumption optimization of radio networks. However, many existing reactive techniques are not enough to achieve better optimizations. For example, cell zooming techniques that are currently proposed are not capable of making predictions and mostly refer to low or average network traffic. Therefore, the use of machine learning to determine the energy consumption policies seems promising enough to overcome the lack in predictions that the existing approaches meet. Moreover, the integration of big data solutions will provide the real time change of the direction of antennas that we believe will further optimize the energy consumption.

E. Usability

In this subsection, we present how the results emerging from our subsystem will be useful for several target groups. First of all, the network providers can use them to optimize their networks. Moreover, ICT equipment constructors could find interest in using our algorithms to improve their products. Radio network optimization related software could also leverage the big data subsystem to provide more utilities. Last but not least is that other organizations from different areas of interest, e.g. natural gas, could use our subsystem to optimize the consumption is their own stations.

III. Acknowledgement

This research has been co-financed by the European Union and Greek national funds through the Operational Program ”Competitiveness, Entrepreneurship and Innovation”, under the call RESEARCH – CREATE – INNOVATE (project code:T1EDK-04605).

References

[1] “Gartner: 21 Billion IoT Devices To Invade By 2020 - InformationWeek,” http://www.informationweek.com/mobile/mobile-devices/gartner-21-billion-iot-devices-to-invade-by-2020/d/d-id/1323081, (Accessed on 10/31/2019).
[2] F. Han, Z. Safar, W. S. Lin, Y. Chen, and K. R. Liu, “Energy-efficient cellular network operation via base station cooperation,” in *2012 IEEE international conference on communications (ICC)*. IEEE, 2012, pp. 4374–4378.
[3] V. K. Bhargava and A. Leon-Garcia, “Green cellular networks: A survey, some research issues and challenges,” in *2012 26th Biennial Symposium on Communications (QBSC)*. IEEE, 2012, pp. 1–2.
[4] D. Tsoumakos, I. Konstantinou, C. Boumpouka, S. Sioutas, and N. Koziris, “Automated, elastic resource provisioning for nosql clusters using tiramola,” in *2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*. IEEE, 2013, pp. 34–41.
[5] D. Sarlis, N. Papailiou, I. Konstantinou, G. Smaragdakis, and N. Koziris, “Datix: A System for Scalable Network Analytics,” *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 5, pp. 21–28, Sep. 2015. [Online]. Available: http://doi.acm.org/10.1145/2831347.2831351
[6] “Pilot Project by Telefónica NEXT, South Pole Group and Teralytics: Successful test of Air Quality Analysis using mobile network data — Telefónica Deutschland.” https://www.telefonica.de/press/press-releases/news/6025/pilot-project-by-telefonica-next-south-pole-group-and-teralytics-successful-test-of-air-quality-analysis-using-mobile-network-data.html.
[7] “Mobile Broadband Operators Transform with Big Data and Machine Learning on the Journey to Digital Service Providers.” https://www.abiresearch.com/press/mobile-broadband-operators-transform-big-data-and-/.
[8] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications Co., 2015.
[9] “Metamarkets: Real-Time Analytics for Programmatic Advertising.” [Online]. Available: https://metamarkets.com/
[10] “Openstack.” https://www.openstack.org/.
[11] “Docker.” https://www.docker.com/.
[12] “Apache Hadoop.” http://hadoop.apache.org/.
[13] “Apache Spark.” https://spark.apache.org/.
[14] “Apache Flink.” https://flink.apache.org/.
[15] “Apache Storm.” http://storm.apache.org/.
[16] “Spark MLlib.” https://spark.apache.org/mllib/.
[17] “TensorFlow.” https://www.tensorflow.org/.
[18] “SELIS Project.” https://www.selisproject.eu/.