UNIVERSITY OF PIRAEUS

DEPARTMENT OF DIGITAL SYSTEMS

POSTGRADUATE PROGRAMME:

“Postgraduate Programme in "Technology Education and Digital Systems"”

AERA OF STUDY:

Digital Communications and Networks

MASTER THESIS

TITLE:

Design and Management of Next generation networks (HetNets-small cells)

Student	**George Gkoudinakis**
Professor	**Panagiotis Demestixas – Konstantinos Tsagaris**

Athens, October 2014

Table of Contents

Table of Figures

Acknowledgments

I wish to express my deepest gratitude to my supervisor Dr. P. Demestichas and to Dr. K. Tsagkaris for many suggestions and constant supports during this thesis work. His detailed involvement in my research has been an invaluable asset for me.

I would also like to thank my parents and my girlfriend for their encouragement and love. This thesis would not have been possible without their support.

Abstract

The increasing demands for data mobile traffic leads to new challenges for mobile operator and networks in terms of user capacity and data throughput. In order to be able to serve these demands, Heterogeneous Networks (HetNets) consist by macro and small cells have seemed like a promising solution. By adopting a HetNet solution we have more cells in the network, with lower coverage than the ones that already exist, which leads to better user experience for the same area of coverage compared with a macro cell. However, the introduction of more cells also leads to high-complexity scenarios. Thus, it is necessary to create tools that help to mitigate this complexity and be able to help engineers dimensioning the expansion of the network, more efficiently.

The study here presented focus on the use of "what if" scenarios for a given traffic load forecast. The results are given in terms of the network user throughputs. They show that in order to serve a huge amount of data at the future, it is better to embraces HetNet solutions instead of upgrading your existing and expensive macro cell network.

It is also shown that although more complex, a coordinated architecture does not necessarily lead to a better solution, when using schedulers that try to treat fairly all the users. In this case it is better to apply policy control for the users and assign your users to policy profiles, with this approach a better Quality of Experience for the users will be achieved.

With the evolution of mobile networks and the growth of Smartphone's, more and more applications having Quality of Service requirements are coming up. The study of users being classified under a CoS (Class of Service) profile has been also done. Based on the results obtained in this paper through a variety of scenarios and simulations, managing future data traffic and dealing with users that are classified to CoS profiles, in HetNets can be seen as a positive solution.

1 Introduction

This chapter presents a brief introduction to the research area along with a description of the problems addressed in this Master Thesis.

1.1 Area of Study

The number of mobile broadband subscriptions continues to grow at a huge rate as the internet goes mobile since the last two decades. Indeed, the number of mobile subscribers is expected to reach around 3.5 billion by 2015, being in their majority smartphone-based subscribers[1]. This potential growth also implies a considerable increase in mobile data traffic. According to Cisco, global mobile data traffic is expected to grow up to 10.8 Exabyte's per month by 2017[2]. With the evolution of the hardware technology and the development of software, Smartphone's are now capable of displaying high quality videos or real time video traffic, this is clearly a challenge for cellular networks in terms of capacity needs. At the below graphical image the user will be able to notice this forecast of Cisco.

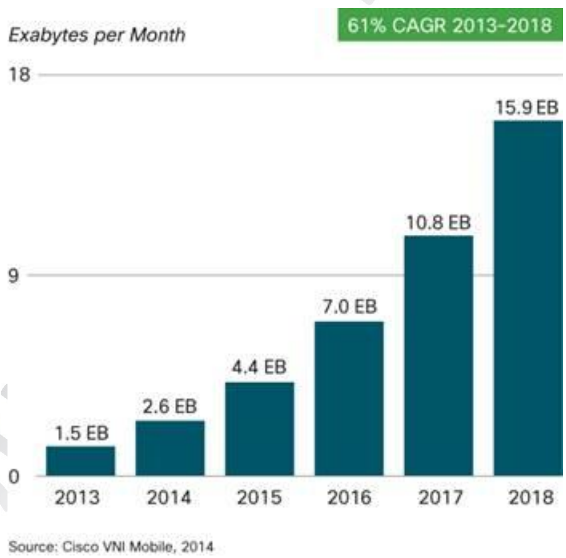

Figure 1 Cisco VNI, mobile traffic forecast

In order to fulfill the previously mentioned traffic storm demands, new mobile networks should be deployed and towards this direction the mobile operators are going. During the last years, deployment of Third Generation Partnership Project (3GPP)'s Long Term Evolution (LTE) has become

[1] Ericsson White Paper. Differentiated mobile broadband.

[2]Cisco White Paper. Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016. www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf , February 2012

more and more present[1]. However, even with the enhancements offered by the new mobile networks, the continuous growing demands in terms of capacity and increased data throughput are still not satisfied. A typical example is the coverage of an extremely dense area (e.g. shopping malls or big stadiums) where the base station has to serve a high number of users at the same time and in most cases the available capacity of the node is not enough to cover all these needs. It is more than obvious that a macro cell node and the traditional thought of network architecture that we had until today, is not capable of serving the needs of tomorrow.

As far as the exponential growth of subscribers and data traffic volume, in order to improve the capacity and coverage the architecture of the existing Radio Access Networks (RANs) should evolve. Different solutions have already been proposed so as to improve RANs, below are two examples being worth mentioning:

- ✓ Increase the density of the macro layer by increasing the number of macro base stations in the same cell site area. However, this solution based on the experience of mobile operators doesn't seem so efficient. It is know that the acquisition of new macro base stations is expensive especially in urban areas, additionally if we decide to follow this scenario we have to think about the redesign of the frequency share which makes the whole process much more complex.
- ✓ Combining macro cells with low power nodes (LPNs) with different types of transmission power, backhaul connectivity, etc. usually such solutions take place in order to overcome the problem of coverage holes and with this improve the overall capacity of this network portion.

The latter option, also referred to as a heterogeneous network (HetNet), is considered a promising way of increasing the average user capacity and coverage[2]. From now, the study carried out along this thesis will focus on the use of HetNets made up by small cells and carrier Wi-Fi hotspots placed in dense areas and deployed alternatively to macro cells.

Additionally, a major challenge for the operators with the fast growth of mobile applications market and services like "Youtube", is to maintain Quality of Service (QoS), meaning that the given network services from the provider must be guaranteed. A variety of methods could be followed in order to maintain QoS, even though the new IP network technology compared with the traditional

---

[1] Harri Holma and Antti Toskala. LTE for UMTS: Evolution to LTE-Advanced, 2nd edition . Wiley, New York, 2009.
[2] Harri Holma and Antti Toskala. LTE - Advanced 3GPP Solution for IMT-Advanced, 2nd edition . Wiley, New York, 2012

SDH network may seem more flexible and efficient to handle higher requirements, but it always remains a challenge to deliver the quality of all requirements with a guarantee. Basically, QoS in terms of network is translated into some specific variables (KPIs) that define the performance experienced by users, which is so called Quality of Experience (QoE) by the telecom industry. For this reason, different QoS parameters are assigned to each pool of users depending on the application data they use, enabling therefore differentiation among them. To this purpose, different classes of services (CoS) have been defined by means of CoS values for each application session that is to be served by the network. Let's suppose that we can mark every service application with a value, i.e 30 for skype and another value for youtube i.e 15, by doing this for the most common used application we can build a CoS profile and categorize the users based on these CoS profiles. The previously mentioned value assignment that we have done for Skype and Youtube, can also be used as a priority flag. Such user categorization can be taken as a reference from engineers during the network dimensioning, so as to achieve more precise results.

1.2 Scope of the Thesis

This Thesis is an analysis of mobile data traffic forecasts. It examines and investigates the way data traffic of RAN (radio access networks) is forecasted both for 3G and 4G by looking at the carrier Wi-Fi hotspots like the perfect HetNet solution that could serve, the provisioning of services from the network, which demand more and more data at a higher bit rate requires new network architecture deployment for mobile operators.

The decision to implement small cells for indoor sites and ultra dense areas, makes infrastructure and interference issues easier to handle. Heterogeneous networks can also be integrated easily with Wi-Fi technology, which is a mature and a much cheaper solution for, indoor mobile broadband coverage for enterprise and home users than femtocells.

Considering the fact that femtocells are a more costly and complex way to deliver coverage to case studies homes were the problem that we are called to solve is the lacking of good macro coverage. Another issue is that femtocells are not fully integrated with the macro network, femtocells reuse macro carrier frequency and this could lead to interference, which affects user QoE across the entire network. Addressing frequency sharing by splitting spectrum and using a separate dedicated carrier for these cells, it increases the total capacity of the network, but makes too complicated the provisioning of RF resources, this is a known problem to engineers that are planning the RF resources. So it is obvious that Wi-Fi is, a better, much simpler and flexible HetNet choice, as it can be used for serving huge data and among other assets new Smartphone's are Wi-Fi-capable. Separately, unlicensed or licensed spectrum is also available for Wi-Fi, allowing operators to avoid problems like interference or spectrum sharing that can result when using femtocells. Alternatively, the carrier Wi-Fi hotspots can also be used to be as an offloading solution for mobile cellular networks, due to the fact that 802.11 WLANs are widely deployed and they are using an unlicensed band of the frequency spectrum (an advantage due to the fact that there is no need to spent money in buying frequencies) .

2 HetNets - Small cells

2.1 What is HetNet

The term HetNet, indicates the use of more than one type of access nodes in a wireless network, in our case the network of a mobile provider. A mobile provider, can use either macrocells, picocells, or femtocells and Wi-Fi hotspots in order to offer wireless coverage in an environment with a variety of wireless coverage zones, this coverage zones could be either open outdoor environment to office buildings, homes, and underground areas like metro stations. With the growth of Smartphone's and the amount of connected devices, wireless operators face increasing demand for mobile data capacity. Long -Term Evolution (LTE) also know as 4G mobile network, allows operators to use new spectrum and supplements existing 3G networks to deliver more mobile traffic to users. At this point we must mention that, with the improvement of the "air" link approaching the theoretical limit and the licensed spectrum available to operators is often limited and so expensive that sometimes it seems prohibitive. The next performance and capacity evolution will come from network architecture change. This change will occur by using of a mix between of macro cells and small cells, also known to the telecom industry as a Heterogeneous Network (HetNet). The adoption of a HetNet solution will lead to a denser network topology closer to the user, capable of serving the future traffic needs. At the figure below the user will be able to see a typical example of HetNet architecture.

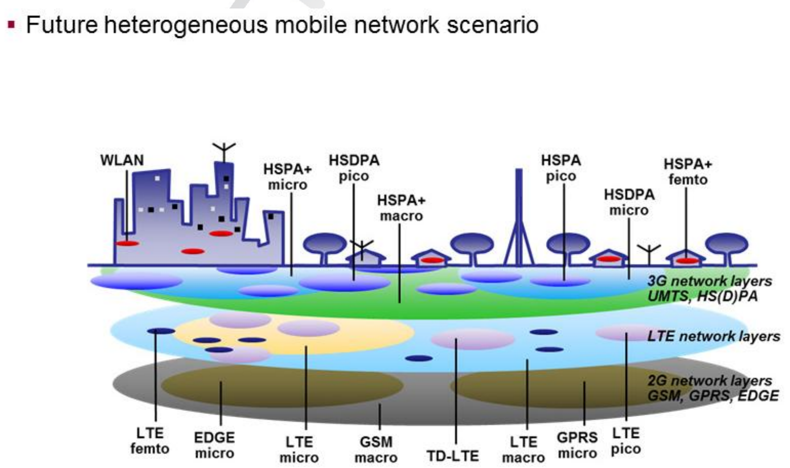

Figure 2 Nokia HetNet Architecture

The HetNet techniques introduced in LTE Advanced which is the evolution of first introduced LTE, according to Qualcomm this techniques will be "namely enhanced inter-cell resource and interference coordination (eICIC) in the network, and advanced terminal receivers with interference cancellation (IC) , enable operators to deploy low power small cells in addition to macro cells in the same channel", but this is a fact only when we speak about pico and femto cells which are sharing the same spectrum with macro cells. By adopting the previously described techniques, mobile operators can increase the spectrum efficiency per unit area (bps/Hz/km2) and provide higher network capacity and improved user QoE. As we have mentioned again above, since Wi-Fi is considered a HetNet solution it can also used by the mobile operators where it is available and it can deliver the same user experience with complexity from the network side. Since all smart phones have built-in Wi-Fi access and in some cases of tablets there is only Wi-Fi access and no other mobile access, so mobile operators can offload data from macro networks to carrier Wi-Fi networks at homes implementations, in office buildings or at public hotspots. But in either way, for public hotspots operators need to either implement or operate their own Wi-Fi networks, or reach an agreement with other carrier Wi-Fi network operators.

Heterogeneous Networks (HetNet), will help the mobile providers to serve data "hungry" Smartphone's and tablets. By experiencing streaming services, uploading pictures and using cloud-based services (Dropbox), will create high traffic demands for mobile networks. Heterogeneous Networks will play a key role to create an optimal user experience, especially in dense urban areas. In order to have a better level of understanding for the data traffic that is expected at the coming years, the user will be able to have a look at the figure below.

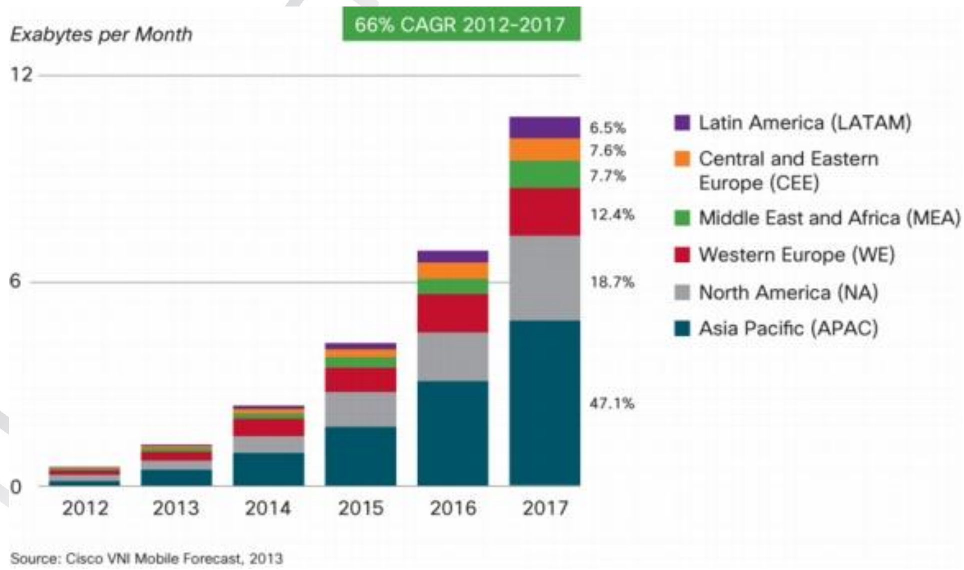

Figure 3 Cisco Geographic Mobile traffic forecast.

A Heterogeneous Network (HetNet) based on a 3GPP-standardized and coordinated radio network with integrated Wi-Fi hotspots, an efficient traffic management platform and high-performance backhaul capabilities close to 1GB/s and sometimes higher will help to deliver a robust, high-quality and of course a seamless mobile broadband experience. This capability of HetNet solution to deliver a significantly higher mobile user experience can be seen below in 3 steps.

Improving existing macro cells by enhancing macro cells with more spectrum (more carriers), advanced antennas that have baseband processing capacity within and between nodes. Continuing with the evolution of 3G and 4G network technology, will increase macro nodes efficiency through specific features, such as higher-order modulation, higher sectorization, and multi-carrier and multi-antenna solutions, as well as the redesign of the available RF, using hybrid radio solutions. Adopting this solution in order to increase the capacity and data rates does not require new sites so the site acquisition cost issue is not affecting this solution.

Densify the macro node's capacity and data rates achieved by evolving only the macro network will eventually be not enough to cover the future traffic demands. Trying to add cells only to specific areas where the performance is poor can improve the capacity issue. By adopting this solution we can keep the total number of sites relatively low. A common way to increase the density of an existing network could be a cell-split, which enables a site from transitioning to i.e. a six-sector site from a three, sector site. Coming back to the previously described cells, they be could macro cells or even micro cells, but of course at the end we will be confronted with a more complex solution due to frequency sharing.

Add small cells by integrating macro cells with small cells and dedicated indoor solutions. This approach can include the use of micro cells, pico cells, as well as Wi-Fi hotspot solutions. It leads to higher capacity volume per user and rate coverage in areas covered by the small cells, with the ability to improve performance in the macro cells network by offloading traffic generated dense urban areas (big stadiums and shopping malls). The rate of integration that can be achieved throughout the Heterogeneous Network will determine the overall network performance.

2.2 Small Cells

Today, small cells are used mainly for capacity improvements in dense urban areas, where high subscriber densities exhaust the capacity of the macro cell. It is expected that mobile traffic will increase by 1000x in the next decade. Small cells have the characteristic of high scalability and low cost, it is a total new network model that can fulfill future traffic requirements. The model capitalizes on existing consumer sites and backhaul to reduce both CAPEX and OPEX while achieving significant offloading of users from the macro network, providing huge throughput improvement and with it better user QoE. The user can see below the telecom industry challenge, that is to be fulfilled through small cells and it is called the “1000x challenge”.

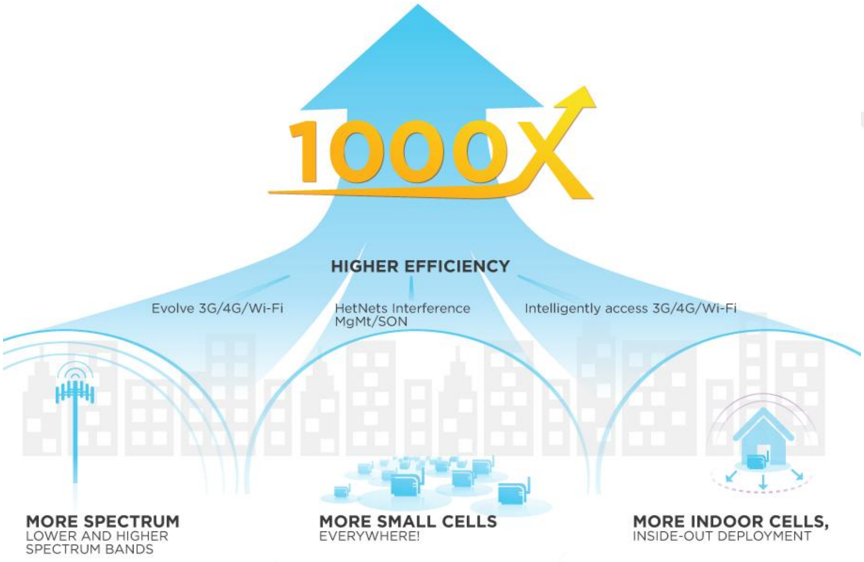

Figure 4 Nokia Vision for mobile data traffic storm.

In general, a small cell solution is defined as a base station plus antenna with lower output power, smaller form factor and lower weight compared to a macro cell. The main purpose of small cells is twofold:

- To fill coverage holes of the existing macro cell network
- To improve capacity in areas where macro cells are drain out.

The smaller form factor opens opportunities for operators to install small cells in areas where, macro cells cannot be used. Small cells are typically found in urban and dense-urban areas. High-rise buildings in cities are the usual cause for coverage holes in certain parts of the topology. High user densities with the combination of new services and application growth also cause the capacity of macro cells to be insufficient. This leads mobile operator to place small cells at strategic points in these areas helps to overcome coverage holes and insufficient capacity issues. In addition, small cells are often seen as a low-cost alternative to macro cells in rural and hard-to-reach areas because of their ease of installation and provisioning.

Although it may seem that a very big amount of cells are needed to meet the 1000x challenge, each small cell comes with a far lower price tag than does a macro cell, i.e when we are speaking for a Wi-Fi hotspot we are talking about 5.000-9.000 euro’s but in case of a macro cell this number sometimes it reaches to 50.000-80.000 euro's. SON (Self-Optimized Network) techniques enable small cells to autonomously adapt their transmit power and therefore to ensure adequate coverage, this leads to much simpler RF planning. Plug and play capabilities of small cells mean end-users can install them without any assistance from the operator. There is no site acquisition required for the deployment of these small cells as end-users' premises can be used. Additionally, existing backhaul at those locations can be leveraged, xDSL technology can be used for backhauling solutions.

A combination of the above solutions, allows operators to grow their network capacity significantly while keeping the deployment costs much lower compared to traditional macro cell solution or HetNet deployment. It is clear that small cells can be more effectively and meet future user demands for capacity as they are selectively deployed where users and thus data demand exist. Bellow the user can observe a characteristic example of the capacity capabilities and the evolved user QoE, which the combination of small and macro cells, could provide to end users.

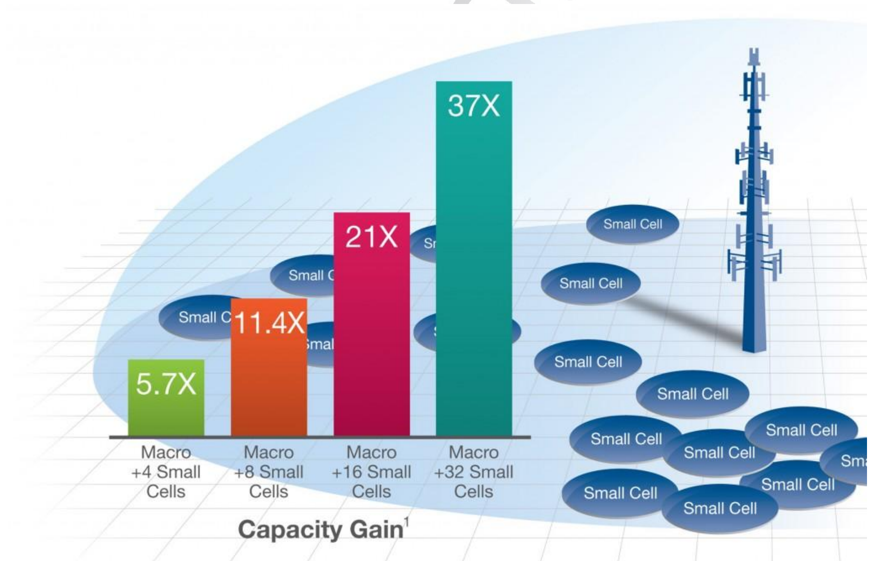

Figure 5 Small Cell deployment

Although we have mentioned in detail all the advantages of small cells, they still have some major challenges, like backhauling, site acquisition and maintenance, power supply and inter-cell-interference.

Backhauling: Each small cell needs to be connected to the network. It needs to be decided on a case-by-case and site-by-site basis, what backhaul solution is best. As the numbers of small cells are increasing, backhauling becomes a major headache for operators in terms of transport network complexity, the variety of rolled-out backhauling options and finding the most suitable and cost effective solutions.

Site acquisition and maintenance: Site acquisition for macro cells is almost impossible for many operators in dense urban areas. Even though the requirements for small cells are easier, the cost of such site acquisition including lease agreements and fees is often down-played in the industry. Site maintenance in a continuously growing grid adds further commercial component to small cell business case.

Power Supply: Power supply is often a downplayed topic for small cells. The application process can take a long time and the availability of power is not always given. Innovative examples show that good cooperation with municipalities can reduce contractual complexity and reach volume agreements.

Inter-cell interference (ICI):Small cells create new sectors within the macro cell. This causes more inter-cell interference. To reduce ICI, careful network design and planning is necessary.

Although many different definitions exist, small cells are generally categorized in terms of their output power, number of users they can serve and the cell sizes they support. Figure 6 shows a typical categorization of small cells. Some sources also categorize small cells in terms of their usage in home, business, venue (e.g. airport) or out-doors solution. In this paper, a new type of small cell is introduced that fits within the micro / metro category. In terms of their hardware architecture, no easy categorization of small cells is possible. In general, they consist of a low-power RF front end with antenna that may be paired with some baseband processing or a simplified BBU (base band unit). As a connection to the operator network, either an IP connection (S1 interface in the case of LTE) or a communications protocol like CPRI (common public radio interface) is used.

Small Cell Type	Usage	Output Power	Users	Cell Radius
Femto	Indoor	1 mW ... 100 mW	> 20	10 ... 100 m
Pico	Indoor	100 mW ... 250 mW	30 ... 100	> 200 m
pico	Outdoor	1 W ... 5 W	> 100	> 200 m
Micro / Metro	Outdoor	5 W ... 10 W	> 100	2 km

Figure 6 Small Cell Power Consumption

2.3 The Wi-Fi solution

As it is mentioned at the above chapters, Wi-Fi is considered to be a realistic solution for the small cell category. Carrier-class 802.11n-based Wi-Fi can offer user experience that can easily exceed the experience of HSPA (3G) and delivers data rates close to LTE networks. Today a single 802.11n Wi-Fi

Access Point can deliver 100 or even 150 Mbps of data rates using smart antenna techniques and interference mitigation.

Nowdays we see more and more the term 'offload' at the telecom industry, this term was chosen to describe the interworking between Wi-Fi and mobile 3G / LTE technology. But the word "offloading" as expression, covers a small portion of the opportunities that are arising with the integration of mobile (3G, 4G) and Wi-Fi technologies. Mobile Operators are moving beyond the use of Wi-Fi just for achieving access to their users, or data offload, they are strategically making Wi-Fi technology a tool that will help them to support higher quality for their services and offer broadband experience everywhere. This change of strategy is being driven by technologies which enable public Wi-Fi to be integrated far more seamlessly with other networks such as 3G/4G mobile networks. An illustrated picture of such an example can be seen below.

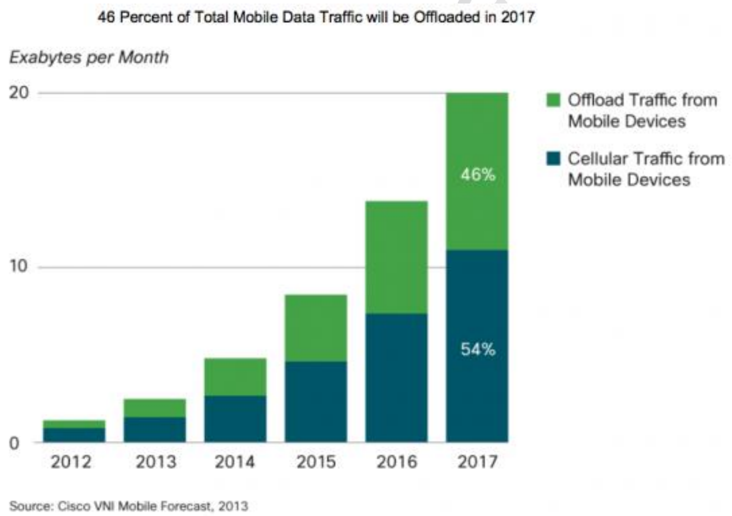

Figure 7 Cisco VNI Mobile Data traffic forecast

Seamless Wi-Fi offload is the new market trend and opportunity for mobile operators. With the smooth integration of Wi-Fi offloading and the existing network infrastructure mobile operator can provision combined carrier-class Wi-Fi and mobile services and profit from offering consumers improved user experience and higher data rates always available. With clientless EAP-SIM / AKA authentication and carrier-class Wi-Fi offload, Smartphone and tablet users can enjoy seamless connectivity/mobility from one technology to another. While mobile operators will see the "production" cost of their services being lowered which will give a significant competitive advantage against the other mobile operators in the market.

Carrier-class WI-Fi is a convenient complement solution to 3G / LTE small cells. Some vendors such as Ericsson have launched multi-radio small cell products with Wi-Fi and 3G / LTE. Vendor Models show that carrier-class Wi-Fi deployments allow mobile providers to save 30-40% [1]on network CAPEX while gaining several thousand percent in radio network capacity in typical indoor deployments. With a solution like seamless Wi-Fi offload, the operator minimizes each own need for expensive mobile spectrum and in parallel has the ability to serve any amount of data traffic that the users will produce into the future.

Seamless Wi-Fi is introduced to the market as an offload solution but this is not the only asset of the of carrier-class Wi-Fi, mobile operator are see beyond the offloading ability for their macro node’s. It is also about offering their users enhanced data service and better user QoE. In a market where competition between mobile operators is fierce and is based on prices, such a solution could give big competitive advantages against the other operators. So it is obvious that the 1st mobile operator that will embrace seamless Wi-Fi to Smartphone subscribers will be the 1st one that will “harvest” a significant competitive advantage.

A recent study of Western markets[2] indicates that 70% of smartphone owners supplement their data services that they have from their provider with Wi-Fi. It is a fact that more than half of the world’s smartphone users have mentioned that they prefer Wi-Fi over mobile. On the other hand the 80% of laptop, tablet users tend to prefer Wi-Fi to mobile services.

Another surprising fact is that the majority of users also think of Wi-Fi as easier to use, faster, and – this may surprise a few MNOs – more robust than mobile data services. The same study shows that 85% of users consider seamless handover (offload) between mobile and Wi-Fi networks to be necessary.

The growing cost of mobile services is also a driver for this change. The fact that at the upcoming years the demand for mobile data “will reach the sky” is making Wi-Fi an attractive low-cost option. In Europe mobile data services are no longer unlimited and there is a thought of embracing policy control in order to penalize users that exceed data limits by reducing data rates or charge them more than the regular charging rates. In order to respond the needs of the market, MNOs are seeing the adoption of Wi-Fi services and the convenience of automatic SIM-based authentication into the mobile services, as a major new business opportunity.

[1] According to Cisco VNI 2012

[2] Cisco ISBSG Consumer Research Report, May 2012

Another study[1] has found that Smartphone's using Apple's iOS in specific markets generate up to 10 times more data on Wi-Fi than on mobile networks. The ratio of 3G / LTE to Wi-Fi traffic on SIM-enabled devices across various markets ranges from about 1:1 to 1:4 or higher. The reason for seeing lower ratios in some regions is due to the fact of low penetration of fixed broadband services at Home Office and Small business, in such cases private access points can serve smartphone and tables. Below there is an illustrative graph that highlights this previously mentioned market trends.

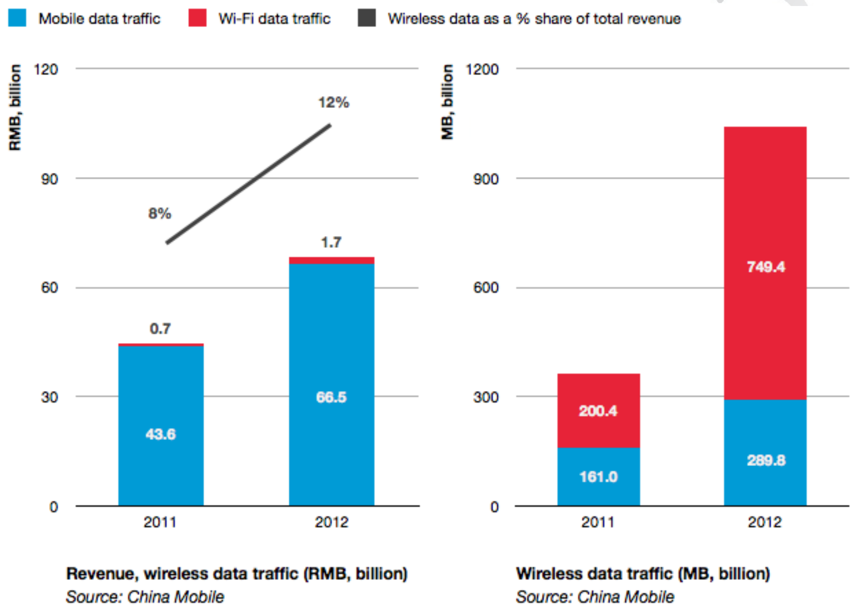

Figure 8 China Mobile, data traffic forecast

The average data consumption of Smartphone's in mature markets today, it reaches the amount of 3 to 4 GB per month per user, taken of course into consideration the Wi-Fi traffic. Meanwhile, the 3G portion regarding this traffic is in most cases less than 1 GB per month. Overall, one vendor[2] estimates that 36% of all Internet traffic will be delivered over Wi-Fi while mobile broadband will represent just 10%.

[1] Informa Media & Telecoms White Paper: "Understanding today's smartphone user: Demystifying data usage trends on cellular & Wi-Fi networks", February 2012 (commissioned by Mobidia)

[2] Ruckus Wireless, USA

3 Wi-Fi Offloading

As we have mentioned in detail during now at the above chapters, mobile network traffic is growing exponentially and MNO's are called to respond to this massive request by managing their networks efficiently. The evolution of radio access networks is limited by the laws of physics, and the growth of radio frequency efficiency has reached each limit. The LTE radio access is reaching the limits of Shannon's law and the available spectrum for mobile services is limited, and the only way to overcome this problem is to increase the overall mobile network capacity by increasing the carrier-to-interference ratio and implementing small cell technologies.

The best way to use small cells is to implement them in locations where the biggest need for data services exist (malls, large stadiums, universities, metro stations, etc.) and where users spend most of their time, which is translated to data consumption at these places(homes, offices, etc.). Wi-Fi, is one of the small cell technologies, is considered from many MNOs a *cost-effective mean* for offloading huge traffic load of mobile data traffic and in the mean time delivering a variety of new services. A sample of these features is:

- ✓ Widespread existing deployments
- ✓ Availability of user devices that support the technology
- ✓ Cost efficiency
- ✓ The ability to handle the capacity of new users and devices (without a subscriber identity module [SIM])
- ✓ Worldwide available spectrum capacity
- ✓ Standards availability for integration with mobile networks

During this chapter we will analyze the technical aspects of Wi-Fi offload architecture and its capabilities for integration with existing mobile networks, so as to provide an efficient way to offload subscriber's traffic from 3G and 4G nodes to Wi-Fi access nodes.

3.1 Wi-Fi offloading architecture

The Third-Generation Partnership Project (3GPP) standard differentiates two types of Wi-Fi access

- ✓ Untrusted: This category was introduced during the early stages of the Wi-Fi specification in 3GPP Release 6. Untrusted access category includes all types of Wi-Fi access that either is not under control of the operator (public open hotspot, user's home WLAN, etc.) or it does not provide the necessary security characteristics(authentication, encryption, etc.).
- ✓ Trusted: Trusted access from the MNOs perspective meant to build Carrier Wi-Fi access with encryption and standardized security authentication methods. Trusted non-3GPP IP access was introduced for the first time with the LTE standard in 3GPP Release 8.

> Regardless the fact that most of today's offload architectures are following the trusted model, 3GPP does not currently offer guidance for integration with the 3G or 2G packet core. However, as mentioned in this document, this type of access category is natively integrated into LTE's evolved packet core (EPC).

Today's most of the mobile operators are mainly using 3G network as a bases to deliver their data services, a significant part of this chapter is analyzing the possible methods of integrated trusted non-3GPP IP access into the 3G mobile packet core (MPC) combined with its associated policy and charging control (PCC) architecture.

Based on the 3GPP 24.302 we have the following definition: "For a trusted non-3GPP IP access network, the communication between the user equipment and the EPC is secure." So, by adopting the latest service provider Wi-Fi architectures we get the Extensible Authentication Protocol (EAP) and IEEE 802.1X-based authentication complementary to these we have IEEE 802.11i-based RF encryption and optionally the use of control and provisioning of wireless access points.

Having analyzed the 3G designs, the description of the evolution of the architectures toward EPC integration as specified in 3GPP standards, we follow. Moreover, session continuity/mobility and IP address persistence when moving between 3G, LTE, and Wi-Fi access nodes is also covered. Under this chapter we will also discuss the integration models for untrusted networks, although there is a small possibility to see such architecture being adopted from an MNO.

In the 3GPP specification, the Wi-Fi network is referred to as a Wi-Fi access network only. There is no reference to Wi-Fi network structure details. During the analysis within this document, there is a separation between the access network and the gateway components. Speaking of the Wi-Fi network infrastructure that is needed for mobile data offload, we have categorized it to three parts:

- ✓ Wi-Fi radio access network (Wi-Fi RAN)
- ✓ Wi-Fi access gateway (WAG) and Wi-Fi back-end systems
- ✓ Packet core integration elements (multiple options)

Below the user can see a figure that illustrates the architecture that we have mentioned until now. It includes the integration of elements with 3G as well as LTE networks.

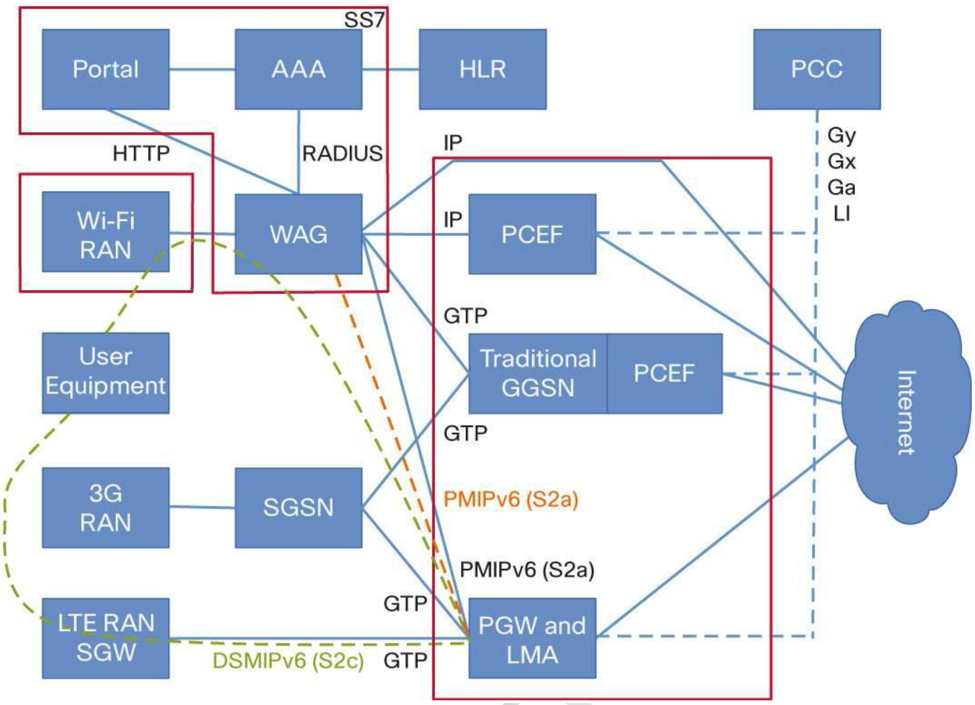

Figure 9 Cisco Wi-Fi offloading network architecture

If the only reason for using the Wi-Fi network is for offloading mobile data, the MNO needs to take care of these tasks:

- ✓ Authentication: It is crucial to ensure that only authorized subscribers can access the network
- ✓ PCC: In order MNO to be able to bill their services, to provide quality of service (QoS), and policy enforcement for the traffic generated through Wi-Fi access
- ✓ IP mobility/persistence: For service mobility between different access networks (3G to Wi-Fi, Wi-Fi to 3G, or across the Wi-Fi network)

During the following chapters we will analyze the details of each of the functions that were mentioned above.

3.2 Authentication

There are many authentication methods, in order to control the access of subscribers to Wi-Fi networks. Now the crucial part is the choice of this method, which will lead to the level of network usability. This authentication method should by transparent for the subscribers, the more transparent this method is the bigger the chance for the subscriber to connect to the network will be. It is crucial for the growth of such solution this kind of transaction to be transparent for the user.

Based on the way that users will be authenticated by the network, we will be able to choose which device and subscriber can “log into” our network, i.e. subscribers with or without SIM cards or visiting subscribers, etc.

In new generation Wi-Fi access networks, there are two possible ways for a subscriber to be authenticated and in parallel experience broadband access services based on the “pool” that he is assigned to. The first method, portal-based authentication which is known also as “captive portal authentication”, with this method MNO’s target customers without a permanent contract and this is done via vouchers, time-limited access and SMS payments. The second method for subscriber’s authentication is EAP authentication, this method provides transparent and easy access for the MNO’s own subscribers with SIM cards or certificates.

During the following pages, we will analyze in detail those two methods of authentication:

- ✓ Portal-based authentication
- ✓ EAP authentication

3.2.1 Portal-Based authentication

The “captive” portal base authentication is based on Layer 3 connectivity to the network and web communication (HTTP) before allowing access to the user. The Wireless Internet Service Provider Roaming (WISPr) standard also uses HTTPs communication with the captive portal so as to achieve automatic authentication, with the user’s mobile device initiating HTTP communication in the background without user participating to all this process. An illustrative architecture of a captive portal-based authentication can be seen at the below figure.

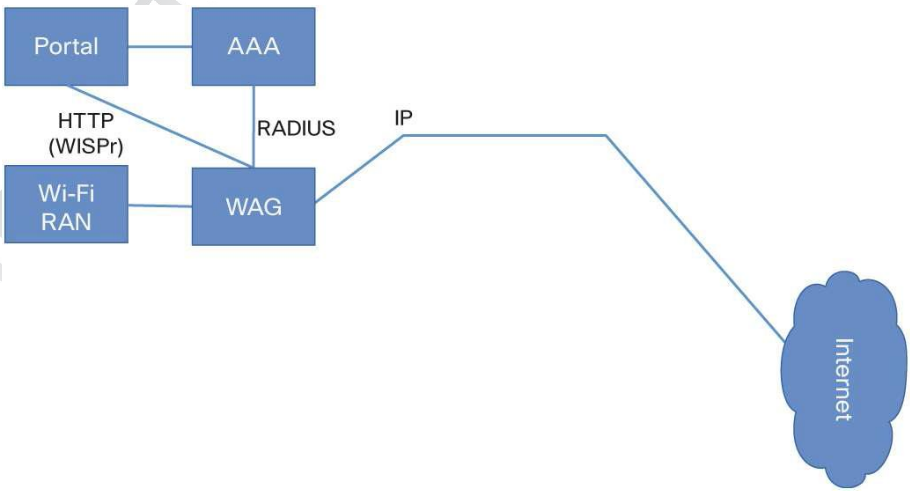

Figure 10 Cisco Captive Portal-Based Authentication Architecture

This method is based on the WAG node in the Wi-Fi network infrastructure, which is not allowing IP communication for new subscribers (unknown users to MNO's database) and redirects those HTTP sessions to a captive portal. By the time a user is consider as "new" the captive portal is responsible for requesting user credentials from the subscriber and start the authentication, authorization, with the accounting server (AAA) in order to authenticate and give access to the subscriber. After the user's successful login, the WAG will take the "green light" by the AAA server to proceed to serve the user. From this moment, the subscriber is considered as known for the AAA database, and the WAG will grand access to the subscriber to send and receive data.

Usually, the user's device MAC address is also saved in the AAA database, together with the user data and granted service. In case the subscriber leaves for a moment the Wi-Fi coverage area and then returns, in this case the subscriber's device will now be considered as "know" by the WAG due to the fact that his MAC address was cached and now it will be automatically authenticated against the saved AAA record, so at the end the user is not be redirected to the portal each time he loses Wi-Fi coverage. This authentication method of MAC address caching, is known as transparent automatic logon (TAL). The MAC address based authentication method is mainly used at public Wi-Fi hot spots, as a policy control method. Below there is an example of a typical TAL call flow for the case of a Layer 2 attached WAG.

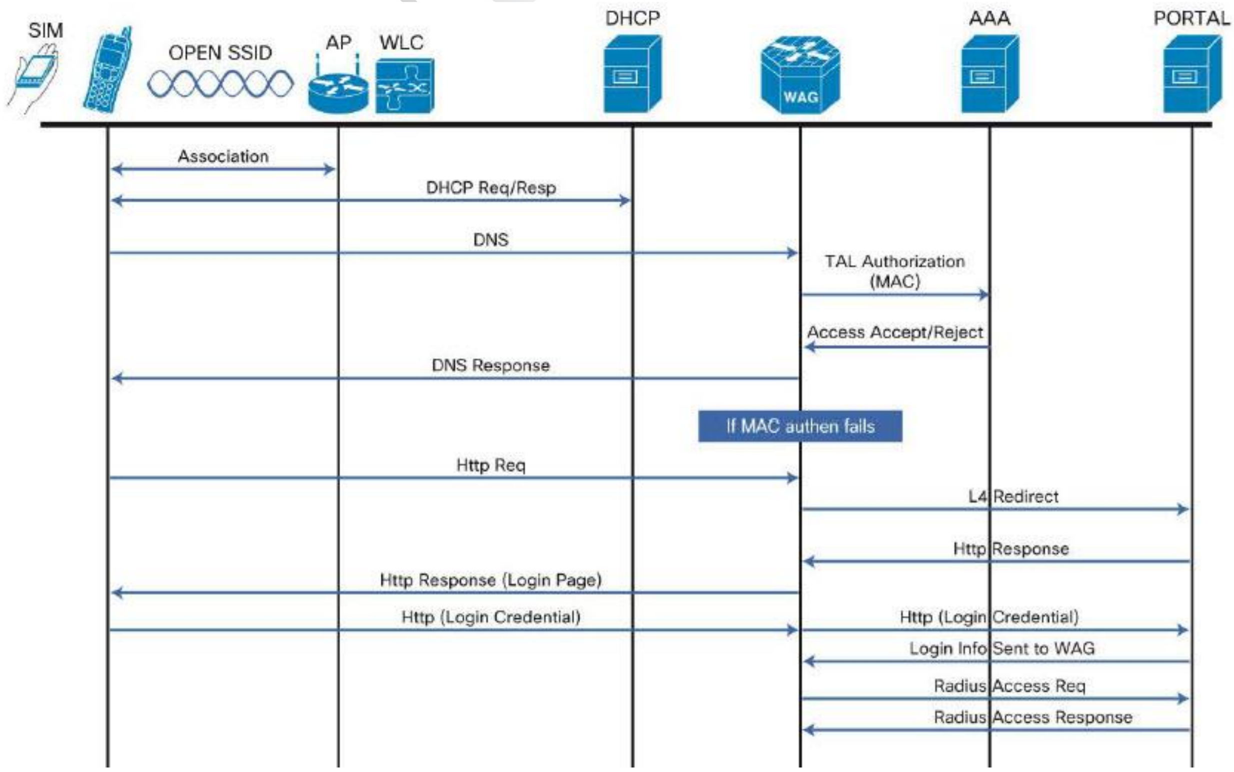

Figure 11 Cisco's Typical Transparent Automatic Logon Call Flow

3.2.2 EAP-Based authentication

During this chapter we will try to analyze EAP authentication, EAP-Based authentication uses EAP and IEEE 802.1x in order to provide Layer 2 authentication for users that are trying to access the Wi-Fi network with EAP-capable mobile devices. It is also possible for authentication reasons, multiple credentials to be used, depending on the capability of the device.

Devices that use SIM cards can encapsulate the SIM information that is exchanged with the EAP message, and all these data are transmitted by the AAA server to the home-location register (HLR) for subscriber authentication. Based on the “RFC 4186” and “RFC 4187” the standards EAP-SIM or the EAP-Authentication and Key Agreement, are the ones that used for the encapsulation, depending on the type of SIM card is used from the subscriber and the HLR capabilities. One of the prerequisites of this method is the need for interconnection between the AAA server and the home-subscriber server (HSS). The architecture of EAP-Based authentication can be show at the below figure.

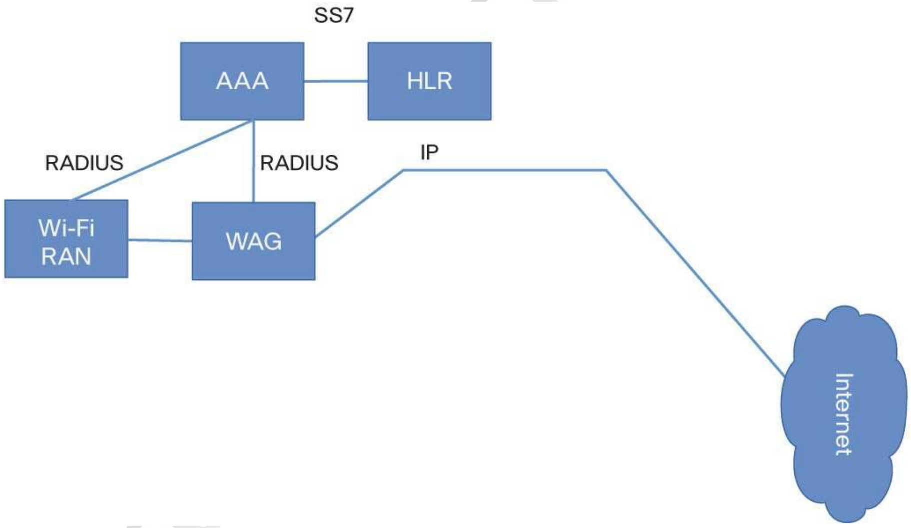

Figure 12 Cisco’s EAP-Based Authentication method Architecture

Regarding subscribers with non-SIM devices (i.e. tablets), the MNO can deliver certificates for the EAP-Transport Layer Security (EAP-TLS) or similar versions of EAP authentication. The usual call flow of EAP authentication (with HLR integration) can be seen at the below figure.

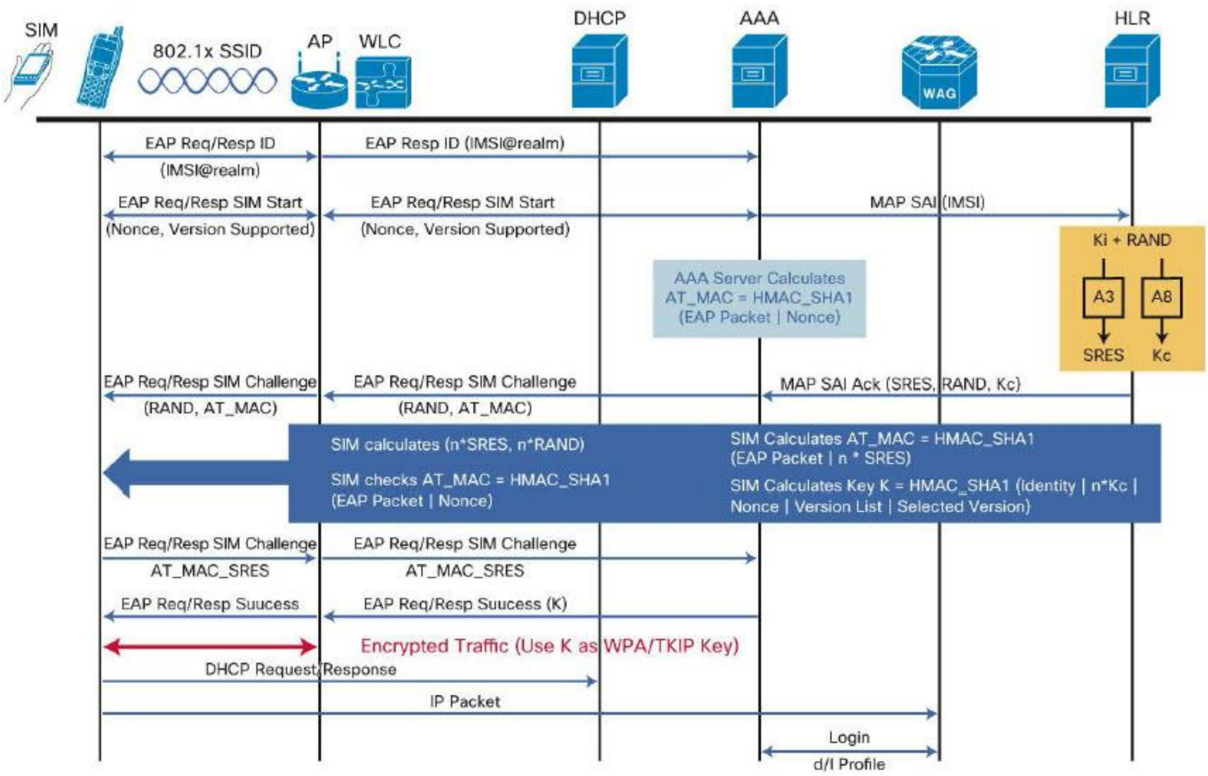

Figure 13 Cisco's Typical EAP Authentication Call Flow

At this point we must point that EAP-based authentication offers a radio security advantage. Because the whole authentication takes place at Layer 2, EAP messages can be used to exchange encryption keys for the IEEE 802.11i-based encryption of the radio interface. This method leads to a stronger and more robust security for radio communication compared to the unencrypted radio interface that captive portal-based authentication is using and is unlikely to prevent MAC address spoofing attacks.

3.2.3 Authentication Summary

Based on the previously discussed capabilities of both authentication methods, we can say that mobile operators which implement Wi-Fi access networks usually use both EAP and IEEE 802.1X authentication and captive portal-based authentication in their networks.

Captive Portal-based authentication is used to attract users that are visiting networks that don't yet have a relation to the MNO. So via this method, access to public Wi-Fi use cases such as *credit card payments, vouchers, and SMS passwords*, is allowed.

In general, with this approach we can have a generation of new revenue from Wi-Fi networks. With the EAP-based authentication we are targeting devices with the operator's SIM card. With this method we can have transparent authentication and secure collaboration without any participation from the subscriber, if we of course exclude the initial configuration of the SSID which is needed, during the first time that a Wi-Fi network is detected. Closing this summary chapter our conclusion is, that via the introduction of EAP-SIM or EAP-AKA authentication we can see much better utilization of the access network by subscribers and therefore enables even more savings from Wi-Fi offloading.

3.3 Policy and Charging Control (PCC)

One of the import parts of a MNO network is the policy enforcement and charging rules for the subscriber of his network, regardless of the RAN being used. So it is clear that the design of PCC integration is very important part of Wi-Fi offload.

The so far experience from the all the existing deployments shows that the best approach to PCC integration is the reuse of the elements deployed for the 3GPP services. The available integration options will depend on the existing PCC infrastructure that exists in the MNO's network. According to the literature of mobile industry "If the MNO uses a device with the standalone policy and charging enforcement function (PCEF), the WAG will be integrated as an additional gateway served by the PCEF". If the PCEF is integrated with the existing gateway General Packet Radio Service (GPRS) and support node (GGSN) of the MNO, the WAG may act as a serving GPRS support node (SGSN) and it will switch the Wi-Fi access node sessions to a GPRS Tunneling Protocol (GTP) tunnel to the traditional GGSN. At the following section of this document we will analyze the details of these two options.

At this part we would also like to mention that the trusted non-3GPP access integration into 2G and 3G PCC, will also be analyzed during this document. Later on, we will continue with the analysis of standardized LTE architecture integration and untrusted non-3GPP IP access integration. Either way, the fact that we analyzing the standards during this chapter is clearly for having a very good image of the needed actions that the providers have to follow, so Wi-Fi offloading could be a reality.

3.3.1 Standalone PCEF

At the standalone PCEF scenario, the WAG is configure to transmit the data traffic of the subscriber to the PCEF for PCC integration. In parallel, the traffic that does not need policy control (In general traffic from users that don't exist at the MNO's DB like visitor's traffic, wholesale traffic, traffic from Hotspot "coupon" users, etc.) is allowed to go directly to the Internet. Below there is an illustrative picture of standalone PCEF architecture.

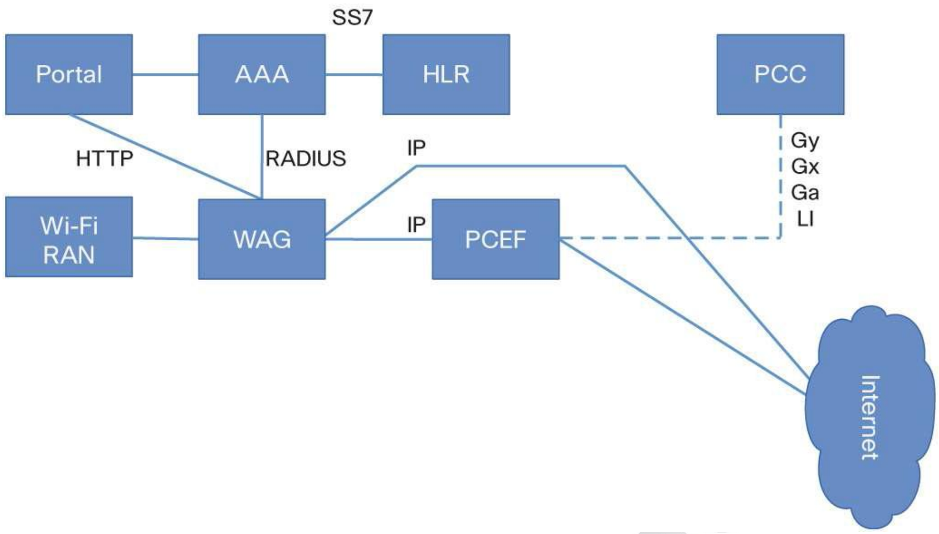

Figure 14 Cisco's Standalone PCEF Architecture

Due to the fact that the PCEF has to be able to associate the subscribers ID with the data flows passing the PCEF, a mechanism is needed that can synchronize and match the subscribers ID with the IP address of the subscriber. Likewise, the RADIUS proxy function on the PCEF is used to produce subscriber's session information according to the features that exist within the accounting messages that are initiated from the access gateway for a particular subscriber. Following the above architecture figure of PCEF, now we can see a typical call flow.

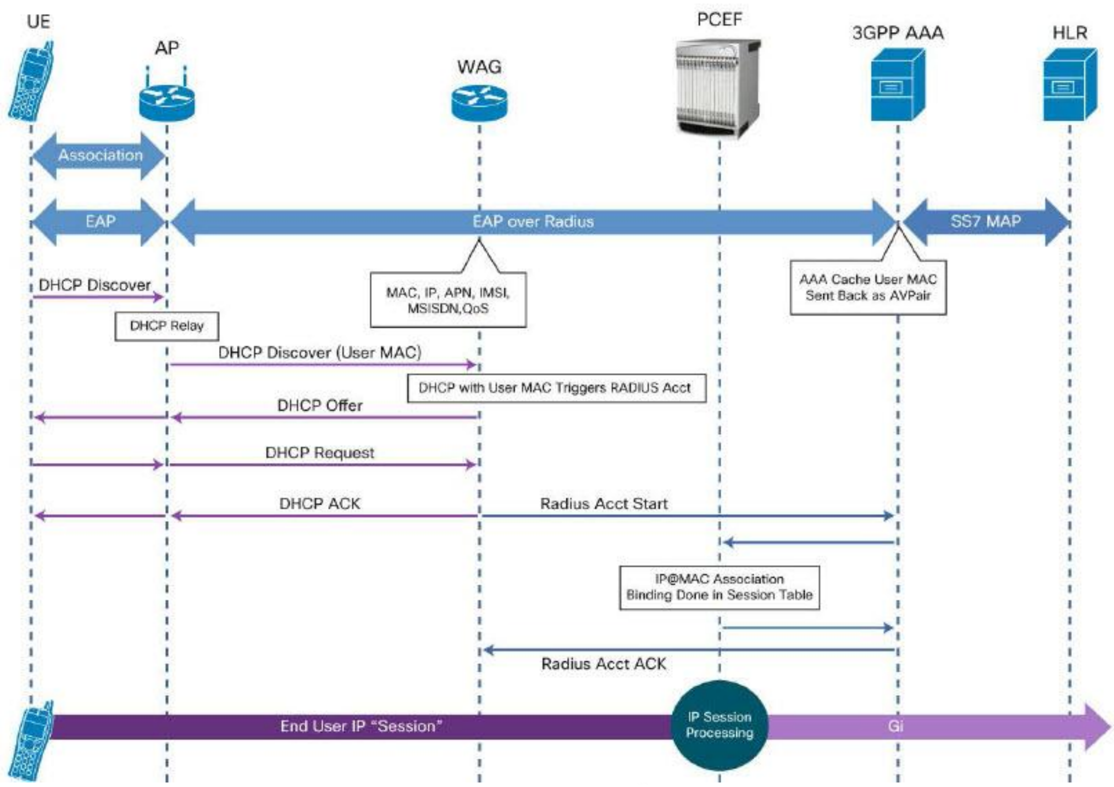

Figure 15 A Typical PCEF Authentication Call Flow according to Cisco

In case the above model is deployed, the MNO's needs to ensure that all the obligatory information needed by the PCEF is encapsulated in the RADIUS messages tha are received from the access gateway or proxied via the AAA server, where all the mandatory attributes are added to the message structure. Complementary to the IP address of the subscriber session ID there are some additional such as, the international mobile subscriber identity (IMSI) information, the mobile station international subscriber directory number (MSISDN) information, and the associated access point name (APN) information is usually needed.

3.3.2 PCC Integration Considerations

When performing PCC integration, there are some things that should be taken into consideration, such as:

- ✓ The integrity and validity of the listed options which are needed for 3G network. Later on we will analyze that, LTE provides native integration into the EPC and therefore into the PCC
- ✓ The crucial part is the capability of the WAG element to provide all the mandatory information for charging (i.e. the MSISDN, the QoS profile, and etc.)

- ✓ Usually, the PCEF does not process traffic from subscribers who are existing customers of the MNO (non-SIM subscribers). This traffic is sent directly to the Internet. But still If these particular subscribers sessions need policy or charging functions, these are usually handled by the WAG element and Wi-Fi back-end systems.

3.3.3 PCC in LTE scenario

At this chapter we will examine the integration of PCC in an LTE scenario. At the 3GPP TS 23.402 it is analyzed the native integration between the trusted and untrusted non-3GPP IP access networks into the EPC. Within the standard it is described that the Wi-Fi network is as valid an access network as any other 3GPP RAN. This acceptance enables MNO's to use the standards-based components of EPC for integration and due to that fact it helps ensure a good level of interoperability between different access types.

In order to forward the Wi-Fi traffic to the EPC, two interfaces are defined, both of them are terminating Wi-Fi sessions on the packet data network gateway (P-GW), such an illustrative example can be seen at the figure below.

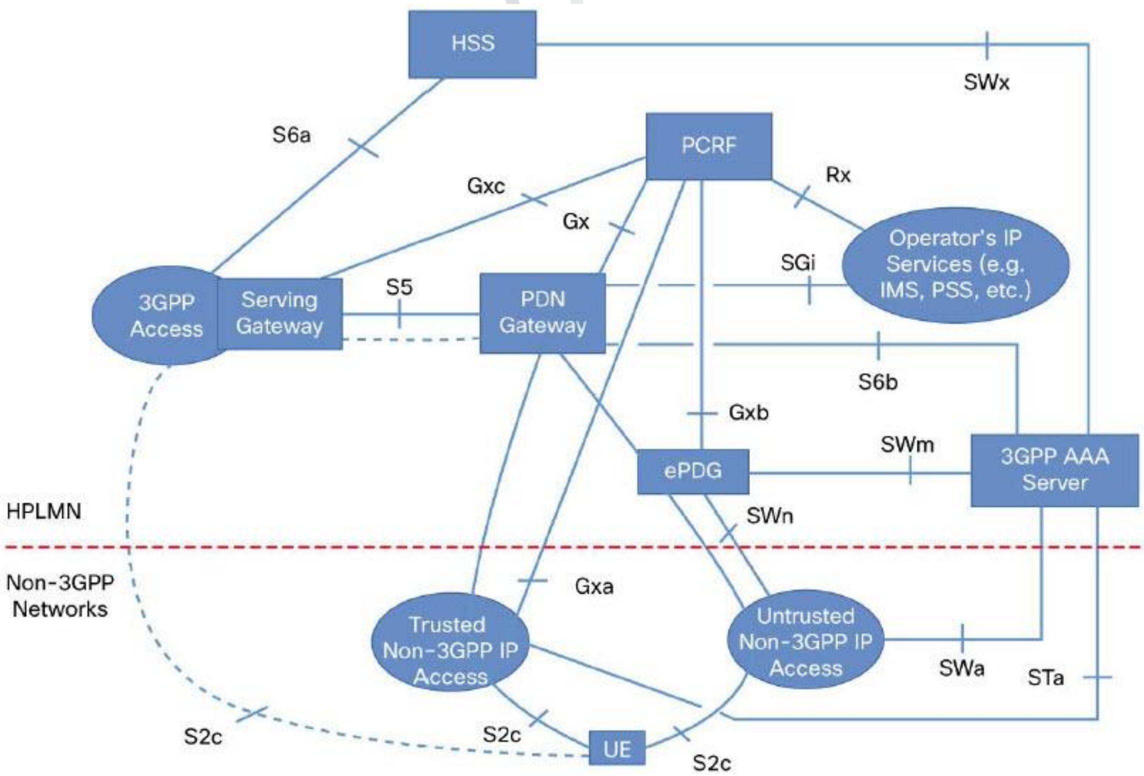

Figure 16 3GPP Architecture for Non-3GPP IP Access Integration into EPC, S2c Option

The S2c interface (that we can see at the above figure) is based on the Dual-Stack Mobile IP Version 6 protocol and requires user device to support it. DSMIPv6 creates a tunneled connection between the subscriber's device and the P-GW, which is used to forward all traffic to and from the user equipment. The P-GW element is the one that will assign a virtual IP address to the tunnel during the whole setup process. This previously mentioned IP address will be taken from the same IP pool that also exists for the LTE sessions. Due to the fact that all traffic initiated to and from the subscribers device is sent via the tunnel, the P-GW has complete control of the subscribers traffic and therefore can apply PCC and other needed functions to the traffic in the same way as it does to the LTE sessions, in order to have a better understanding we can see the illustrative figure below.

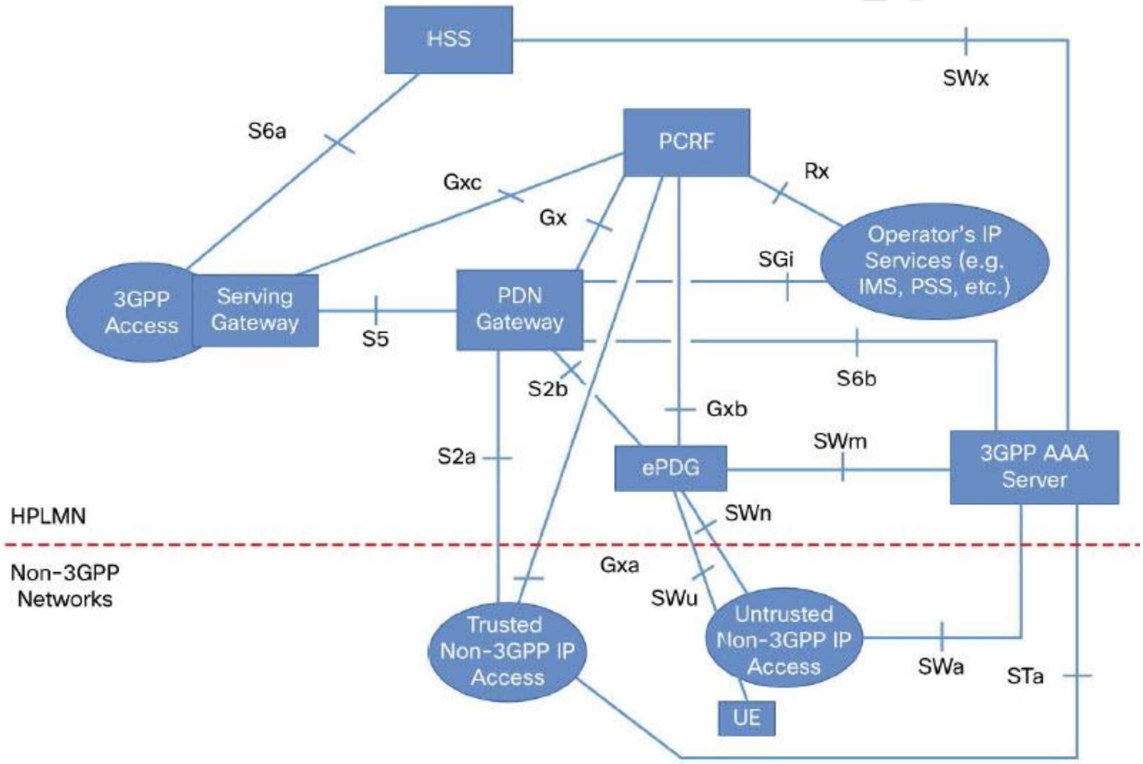

Figure 17 3GPP Architecture for Non-3GPP IP Access Integration into EPC, S2a Option

In order to have a reference about the architecture followed by LTE, there is a figure below. Again, except the traffic that is being tunneled towards to the EPC, users that are not considered subscribers of the MNO are directly routed to the internet via the WAG element.

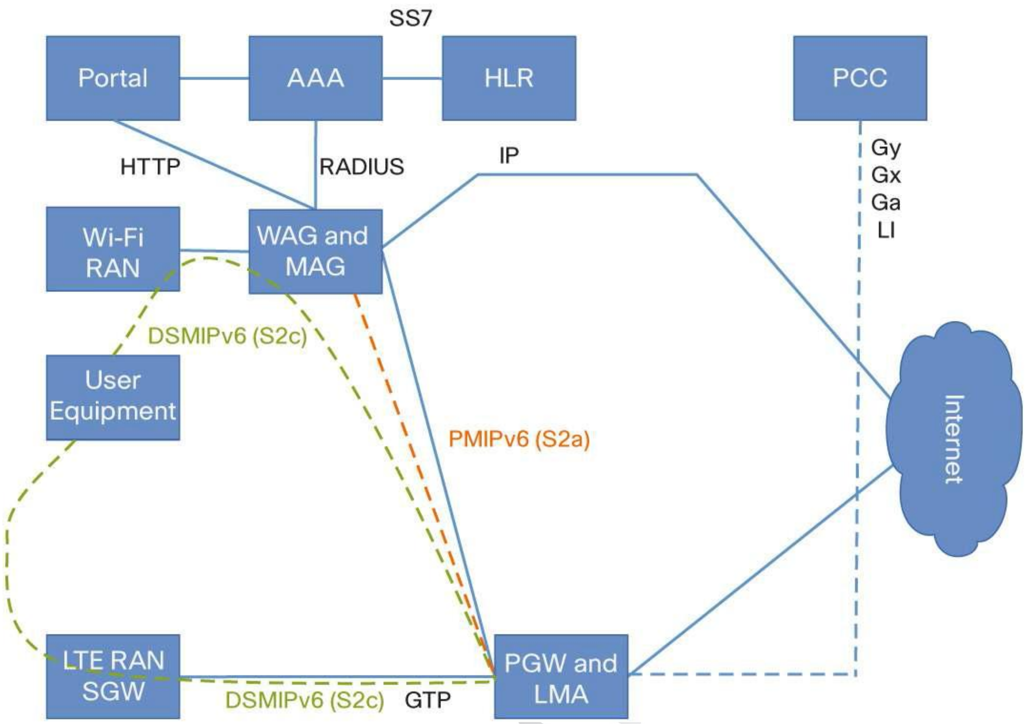

Figure 18 LTE Architecture

At the above figure it is obvious that two methods of integration (S2a and S2c) have been used, and each has different effects that should be taken into consideration during the deployment. In case the S2c approach is followed it requires changes on the subscriber equipment, therefore, it is considered client-based. This approach may not be trivial in a mobile network because of the need for an application at user part for functions. There is a very big challenge for the mobile operator, because with this solution he must ensure that an extensive pool of different handsets and operating systems can be addressed by the software, must provision the subscribers equipment with new versions of software, and in parallel the MNO must prompt the subscribers to use the client software.

3.4 Handover

There are several ways to achieve the handover from one technology to another, although it is crucial to understand the meaning of this term that is very often used. Specifically, it is important to understand what session handover is and which are the types of handover that can be implemented, of course always based on the requirements of the MNO.

One most important aspects/challenge of a mobile network is the handover process, by this term we refer to the move of a subscriber from one radio base station to another radio base station. In our case now the scope is to describe the behavior of the network when the user switches from one radio type to another (for example, from 3G to Wi-Fi).

Nowdays, several handover approaches can be used. What should be taken into consideration in case of mobile operator is the balance between the expectations of subscribers and the complexity of the architecture, according to the telecom industry this are the most common approaches:

- ✓ Handover without IP address persistency (connectivity handover): In this case the user connects to the Wi-Fi access network, then the user is authenticated transparently and a new IP address is given to him by the Wi-Fi network. All the applications of the user can now use the new IP address as the source. Although there is new IP assignment all the established TCP and UDP connections can still continue over the 3G network. But in case the user endpoint will logically turn off the 3G interface, then all these sockets should be reestablished over Wi-Fi, using the newly assigned IP address.
- ✓ Handover with IP persistency (IP handover): In case the IP is the key part for the handover, the user will connect to the Wi-Fi network, the user will have again the same IP address as he used on the 3G or LTE network. In case of the established TCP and UDP connections are bounded to a physical interface, then they will need to be reestablished using the new Wi-Fi air interface, regardless the fact that there was no change to the IP address.
- ✓ Session handover (transparent handover): This handover method is similar to IP handover, but the switch must take place in very small time gap, so as not to influence real-time media applications (voice over IP, streaming video, etc.) and in this way all the established UDP sockets for media and TCP sockets for the control-plane protocol - can continue without interruption and totally transparent for the users as the device is switching between Wi-Fi and 3G connectivity.

In any case we must mention that seamless handover can only be achieved with subscribers endpoint cooperation, this means that software updating should be also taken into consideration. As far as this software, there is a need of a virtual interface adapter behavior, in order to “map” the physical interface structure for TCP and UDP sockets.

At this point we should mention that, 3GPP defines handover mechanisms for trusted Wi-Fi in case only this Wi-Fi infrastructure is part of the LTE architecture. In case of an untrusted Wi-Fi, proposals exist for 3G and LTE.

Although handover is one of the most important procedures for mobile operator, we will not analyze the handover mechanisms in more depth due to the fact that, the handover mechanism is not the main subject of the Master Thesis and there is a danger to go outside the framework of this Thesis. The academic explanation of handover that followed is enough to go have a good understanding of the mechanism and proceed further.

3.5 Radio and load criteria for optimal network selection

During this section, we will discuss about network conditions that can be an important factor in making intelligent network selection decisions. In the existing solutions, there is no standard that could lead to, gathering all the needed information and distribute this amount of information to subscribers UE in order to take a decision. During this section we will analyze how we can use System Information Block messages (SIB messages) in order to transfer the necessary network status image to a UE, by enabling factors like real-time load and radio conditions to be used as inputs to network selection decisions[1]. These network conditions (Real-time and radio conditions) are two major KPIs that are used from the application of this Thesis in order to proceed to a decision. There is an illustrated example at the figure below regarding the described factors for optimal network selection.

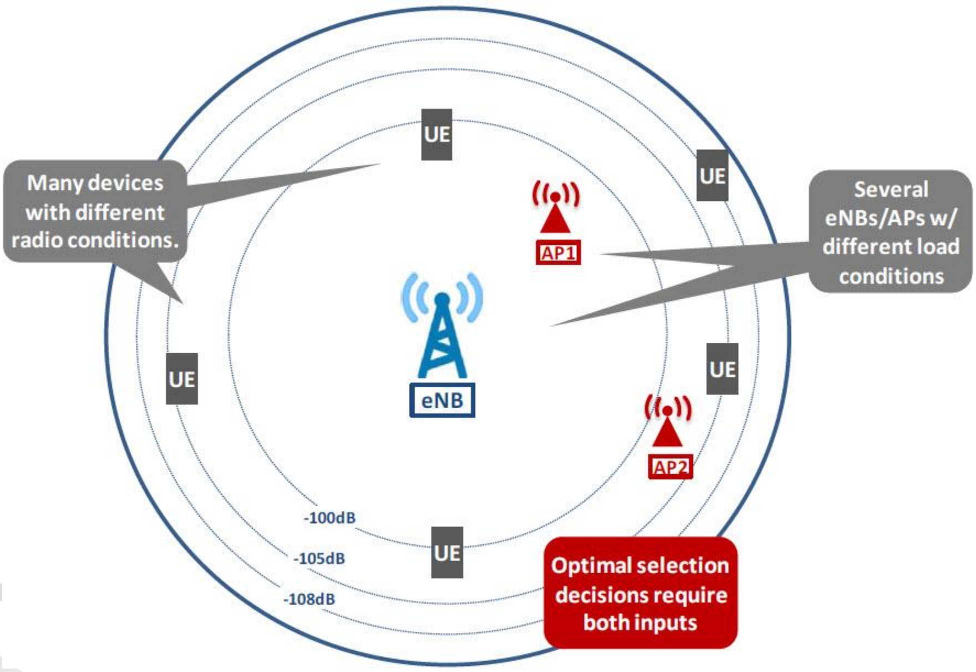

Figure 19 KPIs like real-time network load, radio conditions can play a key role in making optimal network selection decisions.

[1]Anna Cui, Don Zelmer, Vince Spatafora of AT&T and Cisco, "WLAN/Cellular Intelligent Network Selection" R2-131052, 3GPP TSG-RAN WG2 #81bis, Chicago, Illinois, April 15-19, 2013

It is a fact that System Information Block (SIB) messages are used in LTE to broadcast system information to UE's, SIB messaging is appropriate method to deliver the network condition information that is required in order to make the network selection decisions. But, there is a strong need for creation of a new SIB message or to enrich the existing one with new attributes that will transfer the following 3 types of information (which we will analyze in greater detail in the next sections):

1. Cellular network load level (both for 3G and 4G networks)
2. A KPI regarding the signal strength threshold parameter
3. A calculated value will force random subset of UE's to offload

In case of using the above information, UE's we be able to make access network selection and traffic routing decisions by combining the MNOs policy, access nodes load conditions, and RF signal strength information. At the following chapters we will analyze these 3 key points further.

3.5.1 Distributing network load

Imagine an MNO that can control both the cellular (3G or 4G) network and a WLAN network in any portion of his network. Now let's make three scenario with the below assumptions, 1st scenario the cellular network has a low congested (Load = Low), in that case the MNO may decide to serve their customers via the existing cellular network. Second scenario, as the subscriber's base of the MNO expands, it is logical to suppose that the congestion of that cellular network will increase (Load = Med) and we will probably start to see impact on customer experience, after this moment the MNO may want to start steering (offload) some of the user traffic towards the WLAN network in order to keep the customer experience in good level. Scenario 3, now let's suppose that at a future timing, the cellular network becomes even more congested (Load = High). In this case the MNO may want to steer even more users towards the WLAN network in order to keep the customer experience in the same good level. A good example of this scenario is depicted at the figure below.

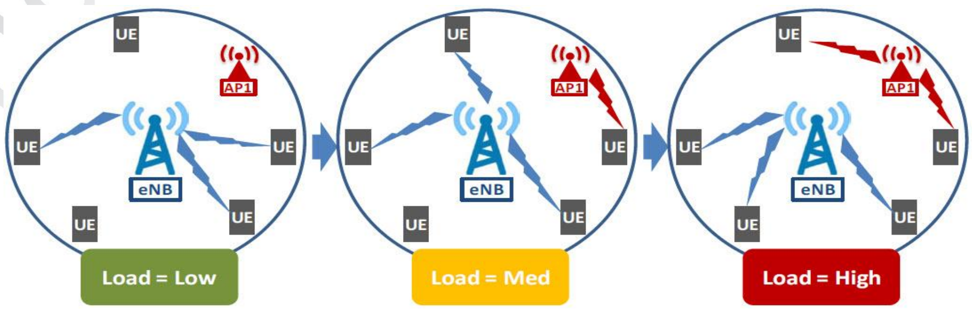

Figure 20 Illustrative example of steering users to WLAN network as the load on cellular network increases

In order to find the most suitable access network for steering (offloading) the user in case of network congestion, SIB messages can be very useful and may transmit the real-time load conditions (e.g., Low, Med, and High) to UE's in the area. When the UE has access to real-time access node load conditions, then it is easy to transfer the policies to the UE. Following this rule an MNO can assign this thresholds (Low, Med, High) to user group profiles in order to give better QoE to specific user groups, i.*e VIP users* could have the factor Med assigned to their *group* in order to be steered (offloaded) to a Wi-Fi AP and be served in a better way.

3.5.2 Distributing signal strength threshold KPI

In the above example, the MNO used cellular network load to steer (offload) traffic towards their WLAN network. This tactic could be even more effective, I case we start steering (offloading) those users that have the worst radio conditions on the cellular network (i.e. poor Received Signal Power) towards the WLAN network first.

For example, let's suppose that the cellular network is congested (Load = Med), the MNO who controls both the cellular and WLAN networks, could decide to steer the devices that have poor cellular RF conditions from cellular to WLAN (assuming of course that WLAN APs are available at that area and with enough free capacity and strong RF signal). Or if the operator has a Policy Control with which he calibrate application/services and he can steer users that are running application/services with better CoS (Class of Services) value. For example imagine two users with the same radio characteristics, assigned to the same node with the same congestion but running different applications, one is running YouTube (data consuming app) and other one is just browsing (low data app). The network could steer the user with the heaviest consuming app.

Now as the subscriber base of the MNO expands and the cellular network becomes more heavily congested (Load = High), the operator may want to steer (offload) more users towards WLAN. At the Figure below we can observe that as the cell site becomes congested, the UE's that have the worst radio conditions (i.e., those with the outer red ring) are steered (offloaded) to WLAN first, then UE's with better radio conditions are steered (offloaded) to WLAN right after that (i.e., those with the inner yellow ring).

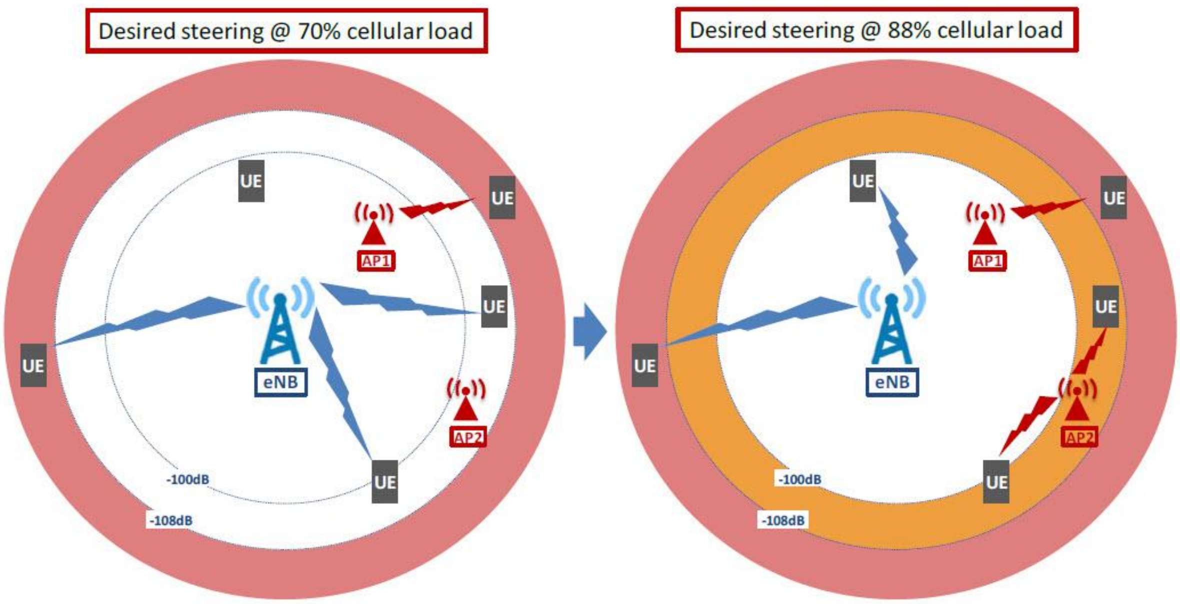

Figure 21 illustration of desired steering to WLAN based on cellular load and UE radio conditions.

There are some cases where, even when the cellular network is not so congested, there might be some users that will experience poor RF quality on the cellular network. These specific users can have access to a WLAN AP with acceptable quality and load, so instead of remaining to the mobile cell with poor service, the MNO may want to serve these users traffic by offloading them to WLAN network. At this point we must say that there is a strong correlation between the distribution of the UE's, the eNB's the AP's at the area and the exact thresholds at which certain users are steered (offloaded) to WLAN vs. cellular. It is clear now that the mapping between the access node's load level and signal strength level at which a user is steered (offloaded) to WLAN may not be static. The way that the steering mechanism is functioning (which was described above), is the main focus at this Master Thesis application and will be analyzed in detail in the chapters further down.

In order to ensure that the correct subscribers (i.e., subscribers with poor received signal level) are steered (offloaded) to WLAN while keeping other subscribers on cellular; we need a signal strength threshold KPI that can be transmitted to UE's via a SIB message. For example, when cellular load is 65%, the signal strength threshold may be -108dBm, but when cellular load reaches 90% then the signal strength threshold may be -95dBm.

By defining this signal strength threshold KPI, we decide the minimum received signal strength, below which UE's should offload to available and acceptable WLAN APs. More specifically the UE ,makes a comparison between their experienced signal strength so far to the SIB-distributed signal strength threshold, this comparison leads to a network selection decision based on that SIB information plus their operator distributed policy (e.g., via PCC or AAA). What a MNO achieves in this ways is the steering of the UEs with poor cellular radio quality to WLAN APs first, but in an intelligent way that also affects the cellular network congestion.

Furthermore, the signal strength threshold KPI may be coupled with thresholds broadcasted in the MNO's PCC or AAA policy, in this way the primary differentiation made in PCC or AAA policy will not be lost when load balancing occurs by using the SIB provided thresholds. As mentioned above at the YouTube example, let's say that it more cost efficient for a MNO to steer (offload) heavy video users to a Wi-Fi access point which costs less comparing to a 3G/4G access node, while preferentially keeping subscribers that complain about Wi-Fi service, to mobile access nodes.

3.5.3 Distributing a calculated value

When steering (offloading) large numbers of subscriber's devices between cellular and WLAN AP's, it is possible to dramatically affect the momentary network conditions. Let's have an example in order to understand better this case, suppose that a big amount of users are steered (offloaded) towards WLAN (e.g., due to high cellular network congestion), this will lead to a very big reduction of the load on the cellular network. This reduction in load could lead those same UE's to try and re-select the cellular network due to the fact that now is a much more desirable access network because its congestion level has been decreased dramatically. It is obvious that there is a big possibility to see all the users move back again to the cellular network, which will cause load increases and users to be steered back towards WLAN. By adopting this distribution of "a calculated value" we are facing a threat for the network, where there might be a ping-pong effect of UE's bouncing between the cellular and WLAN networks as long as this process continues.

To prevent this ping-pong effect, an additional calculated integer can be transmitted within the SIB message together with the load level and signal threshold KPI. This calculated integer is used to steer a portion of areas UE's at a time and not all targeted UE's of the area at once, with this approach we will not see the ping effect again. This solution works as follows: There is a calculated integer "X" (e.g., in the range of 0-10) that is published within each SIB message. Additionally, each UE generates a random number z, which is also in the range of 0-10. If z<X (and of course all the other conditions for offloading are met), then this UE will be steered towards the WLAN AP. Otherwise, the UE stays on the cellular network.

By adopting this solution, only a portion of the targeted population will be steered (offloaded) towards to WLAN each time. Another advantage is that this fraction could be tunable by the MNO. Additionally, there is the possibility of publishing more than one calculated value, so as to have one value per pool of users, where this users will the same service characteristics (It is something like a CoS where the MNO could steer different user categories for example VIP, in different ways).

3.5.4 Functionality of an INS mobile device

Mobile devices that are called to interwork mobile access network with carrier Wi-Fi networks and they are expected to have a set of key functions within their operating system. This should be done in order to achieve a more efficient way of offloading users and traffic between cellular and Wi-Fi networks and to select the best Wi-Fi network under some predefined conditions. In order to achieve this goal, a functionality called Intelligent Network Selection (INS) is expected to be adjusted to each device. During this section will analyze the INS functionality together with the technologies it uses.

The scope of the INS is to give the mobile device the flexibility and the ability to choose from which network it will be served, based on each user preferences, MNO policies, and network/device conditions. Particularly, INS will enable the mobile device to select from which of the available WLAN it will be served and in parallel whether it will choose to steer from WLAN to Cellular or vise versa in order achieve the best possible quality of experience.

One of the challenges of the INS functionality is whether it can leverage the existing technologies, such as the existing ANDSF Standard which is specified in 3GPP TS 24.302 and TS 24.312 documents, the IEEE 802.11u standards and/or its Wi-Fi certified version known as Hotspot2.0/Passpoint. An additional expectation regarding the INS is that, it will accomplish to leverage the technical solutions that are being developed in 3GPP for release 12 WLAN network selection, and all the other enhancements of ANDSF and HS2.0 that will lead into an intelligent network selection and traffic steering which will take into account of all the attributes of the network and the device in real time.

During the next generation of mobile devices we expect that the majority of them will embrace an INS platform, which will allow them to make network selection and traffic routing decisions between a pool of Wi-Fi APs and cellular networks. As a next step, INS could expand more in order to meet future MNO's needs, the enhanced mobile devices and network capabilities that are yet to come.

The adoption of INS functionality in the mobile device will allow:

- ✓ WLAN network selection based on user needs as well as MNO's policies and network requirements.
- ✓ Data traffic steering, by affecting the routing of IP flows to Wi-Fi or cellular networks. The routing path decision in the UE's will be performed based on inputs like operator-defined policies, user preferences, real-time conditions (for example load and RF condition), and

the local operating environment, such as user mobility state, etc. INS will be fully compatible with the PCC/AAA policy framework of the MNO, in order to be able to make network selection and traffic steering according to the MNO's definition.

In order to achieve WLAN network selection, the UE will use the parameters below, as input to the selection algorithm:

- ✓ Cellular network load (e.g. cellular real time condition)
- ✓ User subscription profile
- ✓ Wi-Fi conditions real time conditions based on information that the device can obtain through HS2.0 and Wi-Fi RSSI
- ✓ Time of Day (ToD)
- ✓ Device location (Geolocation functionalities)

Except the user steering UE's will be able for traffic steering decisions also, in order to achieve that the device will use additional parameters including:

- ✓ The type of the application that is generating data traffic
- ✓ Cellular network congestion
- ✓ RSRP/RSCP threshold
- ✓ The overall status of the Wi-Fi connectivity, defined in terms of Wi-Fi load, radio conditions, and backhaul conditions (all the previously mentioned attributes will be provided to the mobile device by the network or the mobile device will calculated/estimated those values)

Smartphone that will support INS, they should function both with the newest Wi-Fi networks that are HS2.0 capable and with much older, traditional Wi-Fi access points that cannot support HS2.0. In case our smartphone is connected to a traditional Wi-Fi access point non HS2.0 capable the QoE "calibration" of the access point will be done statically with specific KPI's like RTT delay, throughput and packet loss KPI's, due to the absence of HS2.0 information from the network. Regarding the Smartphone's capabilities, it is likely for the mobile device to use the calculated quality of the connection between the current AP, in order to trigger the switch to a different Wi-Fi AP if the experienced so far quality is below a threshold, defined by the ANDSF policies and user preferences values that are implemented in the device.

In case the connected Wi-Fi access point is not HS2.0 compliant, the mobile device could have cross reference tables, where the QoE for scenarios like this could be stored. This kind of data may be retrieved later from the device and with the combination of real-time information such as the radio conditions, the mobile device could proceed to the selection of the most suitable Wi-Fi AP.

By the time the selection of the access network has taken place by the mobile device, the INS can influence the traffic routing, over the selected Wi-Fi access point, in order to correspond to policies, e.g., on a per-application or a per-APN basis.

4 Problem Formulation

As we have stated at the beginning of this document, the scope of this Master Thesis is to build a program that could run “what if” scenarios for a specific network portion of the mobile provider. After the introduction that was preceded at the previous chapter regarding HetNets, it is clear that the main challenge of any mobile provider, is how he could manage to handle this massive need for data at the future and of course how he will manage to dimension his network in the best way in order to deliver these needs with maximum QoE to the customers.

In order to accomplish such a target with our program, we will Import all the data that are related to the providers network architecture and of course all this attributes, that together with the help of mathematical algorithms we lead us to, forecast the potential bottlenecks of the network and find also workarounds in order to handle this load with the minimum cost and the best possible QoE for the customers.

But referring to the target of this software and the challenges that we are going to face, many questions are born.

- ✓ What will be the inputs for this software, in order to forecast the bottlenecks of the access network.
- ✓ What is the objective function that we are called to solve with our mathematical algorithms.
- ✓ What are the constraints of this problem.
- ✓ And of course what will be the export of the application.
- ✓ After taking into account all the above considerations, we will explain below in detail the approach that we will follow in order to reach our goal.

4.1 Inputs of the Software

In order to accomplish finding the bottlenecks for a given data load of the mobile provider’s network, we must first of all feed our Software with respectively data. This will be the below:

- ➢ The network infrastructure of the mobile provider, i.e., the access 3G or 4G nodes that are used and their capacity. Depending on the access network technology that the provider uses to covered a specific area. Such information usually exists at Planning Tools DBs that Telecom providers use in order to document and dimension their network.

- The QoE that the users have for specific covered areas, i.e. TCP throughput, Packet Loss, Jitter, and coverage information. This kind of information can be found from mobile phone KPIs. Such KPIs are captured from QoE application, such application are echoed to mobile providers in order to measure the user experience of their customers.
- Of course, in order to be able to forecast the future data load and the impact that this will have to the existing network. You must first define some values, i.e. the projected % load growth of your network per semester or yearly. In this way, we will be able to run "what if" scenarios that will help us dimension more efficiently our network. Usually, this kind of projected load growth rates, come from the marketing department of each mobile provider.

For a better understanding of the application inputs we can refer to the graphical flow chart below.

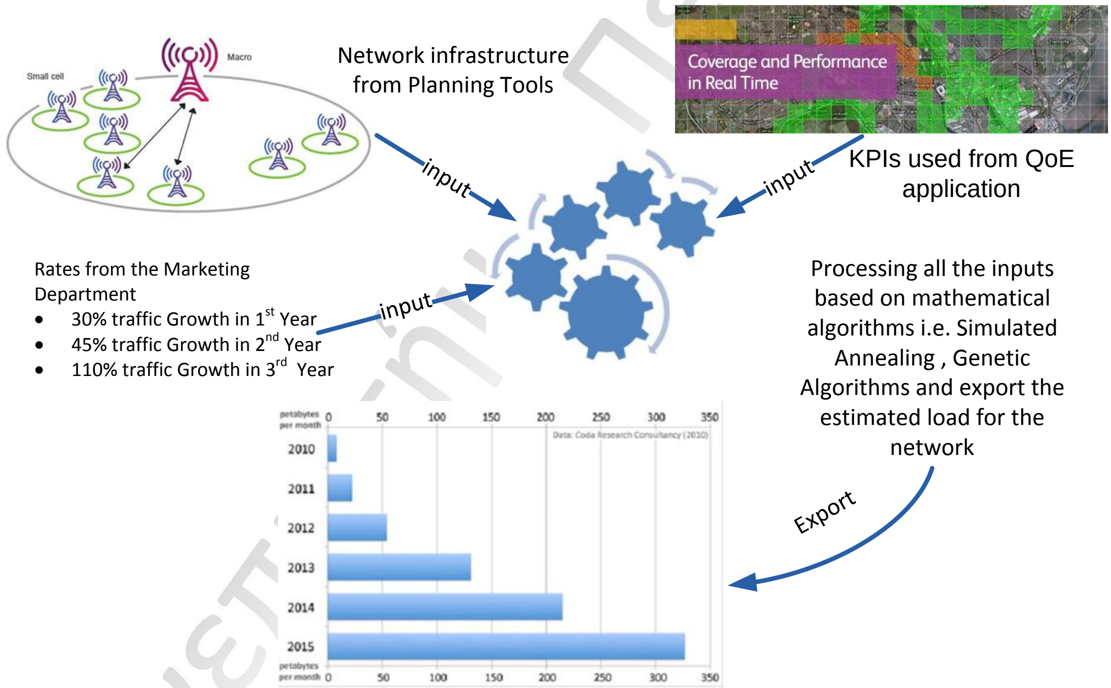

Figure 22 The applications Inputs & Exports

4.1.1 Network Architecture

In our case for the scope of the thesis, from all network architecture of the mobile provider we will depict only the access nodes. All the information of the access nodes is projected to maps. Such functionality is possible because we have a geographical location (longitude, latitude) of each access node of the network regardless the technology (3G, 4G). Except the geographical information, we also have the available CIR (Committed information rate) and PIR (Peak Information Rate). If we subtract the CIR from the PIR for each node, we will have the available capacity that each node of our network has.

Except the above attributes, we also have the information as far as the backhauling connections of the nodes, i.e. if it is an Ethernet backhauling connection or an Ethernet backhauling over SDH (EoSDH). These attributes play their role when we are going to differentiate the access technologies of the nodes and how much capacity could each technology/node to serve. For example, we all know that at the Ethernet world the oversubscription factor exists, but such a thing cannot exist at an SDH network. Considering the above and the fact that EoSDH is a hybrid backhauling type, it is difficult to see an oversubscription factor of 500% at an EoSDH backhauling when it is quite usually to see a factor like 1000% at pure Ethernet backhauling connections.

As the applications tend to evolve, their users are consuming even more data and the need for more capacity is increasing geometrically, unfortunately this enormous growth of the apps does not increase only the need for more capacity but it increases the variation of data traffic. According to Cisco, due to this variation that comes together with the increased demand for more capacity, we see quite often access networks to be designed with a 20:1 oversubscription in order to find a balance between the need for more capacity and these variances[1]. Depending of the application and subscriber's capacity usage patterns, we will see this oversubscription ratio to variance from low values 5:1 until very high values 20:1, for example a cell that is called to serve only subscribers that are using streaming service will drop the oversubscription factor compared to user that are just consuming data in order to exchange information over the social media.

4.1.2 Policy Management, QoS, and QoE metrics

Today's mobile broadband networks can support multiple services, in order to accomplish serving this variety of services, they must share resources from both access and core network. Until today, we have coupled the services mobile networks as best effort, but today this networks they are called to support delay-sensitive apps and real-time services. Each of this service has different QoS requirements in terms of packet delay tolerance, acceptable jitter rates, and required minimum bit rates.

The evolution of mobile networks resulted to IP infrastructure, which leads subscribers to experience high speeds in parallel with high quality services, additionally with the evolution of networks mobile providers can now implement QoS and policy control techniques in order to assure

[1] Design Guide for High Density Client Environments, Jim Florwick, Jim Whiteaker, Alan Cuellar, Amrod Jake Woodhams, November 2013

the level of services, without increasing their network capacity unreasonably. By following this strategy, mobile providers can ensure the quality of their services, they can have a distinction between their subscriber and the services that they are providing to them and cope network congestion in a more efficient way than before.

At these next generation networks, it is clear that Policy Control will play a key role to the implementation of QoS at the MNO's which will lead to more efficient use of network resources. With the use of Policy Control, the mobile operator will be able to define centrally the rules for specific user-groups that will lead to network resource pre-allocation. In addition, with the use of Policy Control MNO's will be able to control the data flow per network application by implementing QoS rules to each service data flows. Policy control can be crucial for the three areas below:

- ✓ Limiting network congestion
- ✓ Enhancing service quality
- ✓ Monetizing services

4.1.2.1 Policy Mgmt for Congestion Limitation

Until today in order to solve problems like network congestion, we have seen the implementation of additional transmission lines, the upgrade of backhaul links at key points and finally the improvement of network efficiency. But, such approach is more reasonable for wired networks than for mobile networks. Because the spectrum is expensive, increasing the capacity by improving spectrum efficiency is a one-way solution in order to handle the substantial growth of data demand. Nevertheless, only with data capacity improvements this challenge can't be solved.

MNO's cant spent unlimited resources and capital in order to solve network congestion problems. The radio spectrum is finite, and gains from further spectrum efficiency improvement are narrow. Even if mobile operators decide to increase their networks data capacity, bandwidth-hungry applications like video streaming services and on-line games will consume in the end any excess capacity. On the other hand the "delivery" of high service quality by over-provisioning network capacity, will lead to a competitive disadvantage against other mobile providers that offer equal or better QoS, at a lower cost. A strict policy strategy can guarantee network performance during peak traffic times and save the operator from additional costs by lowering the probability carrying excess data capacity.

By implementing proactive policies via QoS, combined with other strategies like excess traffic offloading and demand calibration, the existing network of the mobile provider with the current resources can serve better the consumers' demand for multiple services. Policy control distinct applications flow and subscriber types, and then controls the QoS of each type. At the figure below we can see, a classification of the mobile users QoE expectations by service. It is also obvious how each service has different attributes that influence the subscribers' QoE and their total perception of quality.

Services	QoE Expectation	Performance Attributes
Internet	Low – best effort	Variable bandwidth consumption Latency and loss tolerant
Enterprise/Business Services	High – critical data	High bandwidth consumption Highly sensitive to latency High security
Peer-To-Peer	Low – best effort	Very-high bandwidth consumption Latency and loss tolerant
Voice	High – Low latency and jitter	Low bandwidth – 21-320 Kbps per call One-way latency < 150ms One-way jitter < 30ms
Video	High – low jitter and extremely-low packet loss	Very-high bandwidth consumption Very sensitive to packet loss
Gaming and Interactive	Services High – low packet loss	Variable bandwidth consumption One-way latency < 150ms One-way jitter <30ms

Figure 23 Table of user services

The enforcement of Policy management from mobile operators allows them to control the availability and QoE of the delivered services. In general, policies can be launched in order to allocate network resources, for example, pre allocate bandwidth in the radio base station and core network in order to support Video conference services for VIP users. Additionally, policy rules can control the priority, packet delay, and the acceptable loss of video packets, so the network could treat video packets with priority and deliver high video quality services, during the video call.

4.1.2.2 Policy Control Role in Service Monetization

With the growth of smart devices (such as Smartphone's and tablets), mobile operators are facing the risk of becoming just bit transporters, due to the fact that OTT providers and device manufacturers are enjoying more market share from the subscribers revenue. Forcing policy control to the subscriber's services sessions is one method for mobile operators to make a turn and form new business models and maximize the service monetization.

According to AXIA**,** policy control enables operators to meet service expectations through network performance modulation, guaranteeing customer QoE and limiting subscriber churn[1]. With the enforcement of policy control, mobile operators can offer new scalar and more flexible services to their subscribers. Additionally, scalar services can guarantee higher levels of performance and quality to "VIP" subscribers (such as corporate customers) that usually are willing to pay more. With the use of policies, mobile operators are able to deliver better content quality for fixed time periods only, with this strategy we can impulse subscribers' buying premium services. For example, subscriber's can upgrade their service, for a specific period (during the Olympics) in order to watch the video from the games in high definition.

This new aspect for mobile networks, of end-to-end network flexibility and service quality control, can be a major opportunity for MNO's that could lead to revenue-sharing agreements with third-party content providers and application content vendors. The possibilities that policy management gives to the operators, creates new strategies that can lead strong relationships with content providers based on excellent network service delivery that could help subscribers to experience seamlessly the service of a content provider.

4.1.2.3 Service Quality Validation for Mobile Broadband and QoE

Service quality validation is a kind of benchmark for the mobile networks, it helps operators to evaluate networking devices, and optimize their QoS and policy management functions. In order to run a service-quality validation of a mobile network, a "flood" of mobile subscribers that will produce high load to the network is required and in parallel the KPIs that identify QoE should be measured. The scope of this service validation is to simulate as accurate as possible, the mobile data network with the traffic types and traffic mixes, which the mobile operator will have to serve. The stressing of Service quality and policy/QoS schemes can be achieved only when a network encounters congestion. By the time we achieve to fully load the network with a broad mix of real traffic services, extensive QoE measurements are made to quantify network performance.

Each time an operator wants to run a service quality validation scenario, a precise mobile subscriber model is needed in order to simulate real life. The most important detail for successful subscriber model is to, define subscriber types (i.e. corporate user"VIP users" vs. casual user), associating applications to a subscriber (i.e. Internet browsing, email, chat, video streaming, and social media apps), and finally define a dynamic modeling subscribers' usage of applications. During

[1]Ixia , Quality of Service (QoS) and Policy Management in Mobile Data Networks

the creation of the subscriber’s modeling, testers are replicating real traffic types and usage patterns, this kind of information from the network is necessary to fully understand the capacity limits of the network, reveal the bottlenecks, how network handles multiple services, and the network’s ability to differentiate services and subscriber types. The next step after modeling the behavior of a specific mobile subscriber is to make groups of specific subscriber types per group and model usage over time for each group. With this approach operator can emulate the behavior patterns of different categories of subscribers (subscriber group), such as business users, casual users, and telecommuters. Such test frame works should be dynamic and highly adaptive to future trends. An illustrative example of grouping and modeling a mobile subscriber can be seen below.

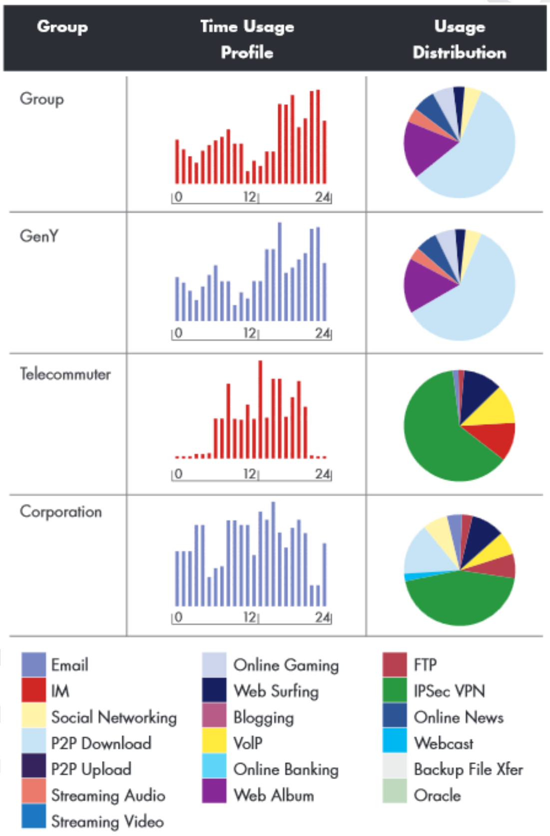

Figure 24 Applications usage & distribution

It is clear that traffic usage, is something stochastic and changes significantly within the duration of day. So based on this fact, subscriber’s behavior patterns should be flexible and be able to “mapped” at specific times of the day, allowing the emulation of peak usage times. For example, if we have a look at service usage statistics, we will see a big difference between working and non working hours, the traffic mixes of a variety of services types during working hours is much different than evening mixes. The use of application from subscribers’ is evolving rapidly, and the distribution varies greatly by subscriber type. With such subscriber allocation and with new applications emerging all the time, no one can predict exact network usage out into the distant future, the key point is flexibility.

Making successful tests in a close environment like a lab, using controlled and predefined small volumes of traffic does not guarantee success in the field. In order to “benchmark” Policy/QoS mechanisms, they must be measured by propelling high volumes of emulated subscriber traffic, and all this should be done when a network or node is near each capacity limits. This is translated to simulation/generation millions of concurrent web transactions and transactions per second.

QoE is a quality factor that enlightens the operator about the overall level of customer satisfaction with the service. The segmental measuring of QoE requires an overall knowledge of the KPIs that impact user’s perception of quality. KPIs are unique by service type. Each service type, such as conversational video, voice, and Internet browsing, has unique performance indicators that must be independently measured. Data applications are typically best-effort services, characterized by variable bit rates, and are tolerant to some loss and latency before the user perceives poor quality. Some of the KPIs for data services include:

- ✓ Transaction latency (including time-to-first-byte and time-to-last-byte of data)
- ✓ Transactions per second
- ✓ Concurrent transactions
- ✓ Page hits and object hits
- ✓ Uplink and downlink throughput
- ✓ Re-transmissions
- ✓ Failed-transactions

Voice applications are real-time services requiring a constant bit rate. Voice services are sensitive to latency and jitter, but tolerate of some packet loss. The main KPI for voice is the mean opinion score (MOS). MOS_V is a perceptual quality score that considers the effects of code/quantization level, the impact of IP impairments, and the effectiveness of loss concealment methods. At the figure below we can see the scale from 1 to 5 that indicate the transmission quality of video applications over a network.

MOS_V	What does it mean?
5	Excellent
4.5	Very Good
4	Good
3.5	Poor
3	Not Acceptable
2	Severe
1	Useless

Figure 25 MOS / QoE cross-reference table

In order to fully understand QoE KPIs we need time and that is because, big amount of KPIs should be examined and they should come from networks with various loads and application services mixes. Mechanisms like Policy control and QoS they should be judged only when the examined network is fully loaded, and there are competing demands buffered for network resources. Only when we follow this approach and these conditions can the effectiveness of rate limiting/policing, packet shaping and resource scheduling be thoroughly analyzed and tuned.

As opposed to MOS_V table we have a respectively way to imprint the QoE of the customer but at a map and regarding his SNR KPI. There is an illustrative image below which shows how we imprint this SNR KPI of the customer at the map.

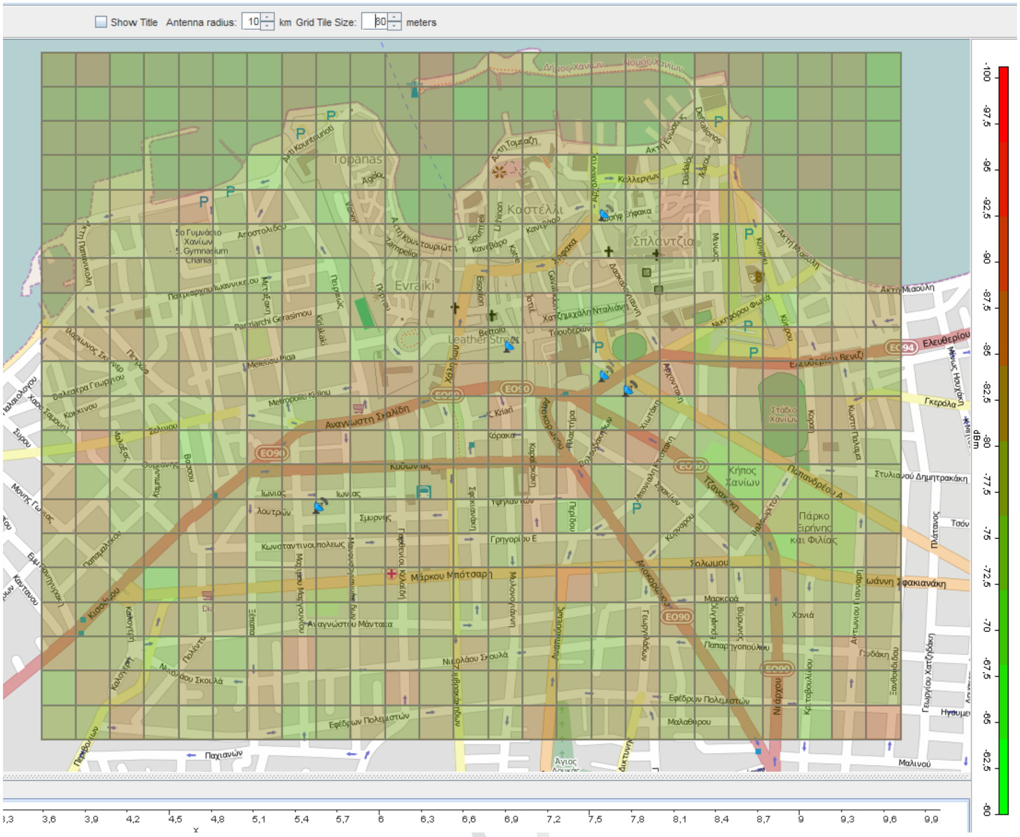

Figure 26 Network selection via the applications maps

As we can see at the above figure, we have selected an area from the map (the grid area) where we imprint the existing Cells in the area (each Node is displayed with a blue antenna). Except the antennas, at the highlighted area we imprint the SNR KPI, this could be taken from a QoE platform of the mobile provider or probes that are deployed at the network. For now such an interface with such platforms doesn't exist, so for that reason we produce this information with a random function. The SNR KPI as you can see also at the above figure, take values from -60 dbm to -100dbm.

4.1.2.4 Summary

Mobile operator and their subscribers are in the middle of the mobile data revolution. The level of growth and convenience that mobile applications have reached and how easy they can be approach via smart devices will be a lever for the future volume of data. Mobile operators are racing to add new services and more powerful devices to their networks. They are making huge investments to upgrade the capacity and in parallel increase the performance of their networks. Mobile operators have seen that revenues from voice traffic are relatively flat, and they count on new revenue streams that could possibly come from data services, in order to re-coup the money they have invested.

In case the forecast analysis about the growth of data traffic comes true, then the mobile operators should find a more efficient and intelligent way to manage the data traffic of their network. Of course there is limitation that do not let operator to upgrade the capacity of their networks, just as is easy as they would like and this limitation is a physical and financial limitation and his name is spectrum. Before operators start to upgrade massively their network they will try hard to maximize the revenue from their existing infrastructure. In order to accomplish this, new business models and more premium services will be adopted. As the technology gets more mature and the subscriber's needs are growing, mobile operators should start gradually to adopt policy control solutions like, congestion management. This strategy will lead to a better quality of the delivered services that will also lead to more successful business models. Today operator should think strategically more than even before and start planning the evolution of their network today, with the help of vendor's powerful tools like QoS management and control.

One of the most powerful tools, in order to measure the performance of the networks and benchmark the QoS, policy control and QoE mechanisms is, the mobile subscriber and service emulation. All these techniques are the most valuable feedback for a mobile provider, based on them he could adapt quickly to new demands and he could take faster his decisions about the evolvement of the network.

4.1.3 Traffic Forecast

Until now at the above paragraphs we have analyzed two of the three inputs of the software. We explained what QoE is for a mobile provider and how policy control is implemented, we clarified what network architecture means for our software and now during this paragraph we will analyze the way that we will forecast traffic for the mobile network and how this traffic will be pushed to the network.

As mentioned before to previous chapters, the scope of our software is to run "what if" scenario as far as future traffic and try to find how this traffic will be handle from the existing network infrastructure. This "what if" scenarios will be examined both from:

- ✓ The network side and this means, how many from the existing cells have to be upgraded and how many new small cells should we implement in order to handle this future traffic load. Also, we will see how the user policy control takes each place, i.e which user with what CoS (Class of Service) will be handle from which cell, in order to have the best possible QoE.
- ✓ From the financial part, this means that we will examine how much do this changes, which are to be made in order to handle the future traffic load, cost. This will be done with the help of an

algorithm, with which we will try to minimize this cost with maximum possible QoE given to the users.

4.1.3.1 Traffic forecast methodology

In order to be able to build reliable traffic forecast scenarios, we have to follow the same methodology each time in order to avoid inconsistency of our data results during the amount of scenarios that we will run. Below we can see an illustrative image of the methodology that we will follow.

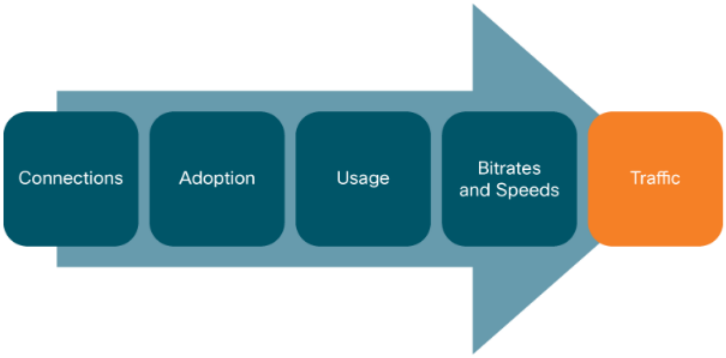

Figure 27 Cisco VNI Forecast Methodology

4.1.3.2 Connections

The forecast for mobile traffic begins with estimations of the number of data consumer users, this step represents the *Connections* at the above figure. Even such a basic measure as consumer data users can be difficult to assess, because few mobile operators analyst segment the number of users by both segment (consumer vs. business). Analysts usually forecast for consumer broadband connections, data on hotspot users from a variety of government sources, and population forecasts by age segment. The number of mobile data users will be given from a front-end form of the application, as we can see at the below figure we can select the amount of users.

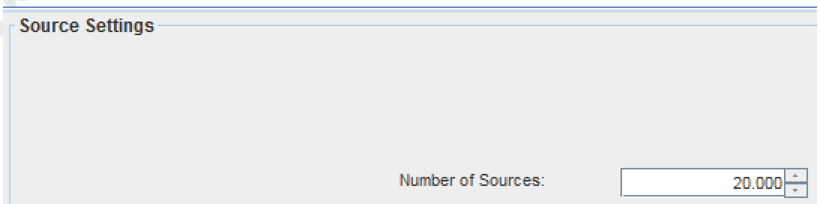

Figure 28 Sources definition via the applications front end

4.1.3.2 Application policy control adoption

After the number of mobile users has been established, now we have to assign profiles to each user that we have selected. The scope is to apply random applications to each user in order to classify them later to user groups. The reason that we are applying randomly applications to each user has to do with the fact that we don't have a real data from a AAA server, so we choose this way to assigned apps to each user. By assigning applications to users, you can differentiate users as VIP or non-VIP and this could help for the "what if" scenarios, i.e. depending of the target group for which we would like to run the scenario we can give respectively values to the applications. For example, in case the target group of our scenario will be young user between 15 years old and 30 we can give high priority to an application like, Social media, P2P & Games, Video streaming. By calibrating this way the applications priority and by playing with the VIP factor we can be sure that independently of the amount of data that the users will produce, the VIP user will be served 1st. Having a look at figure 27 below, we can have a good understanding of what we have described so far.

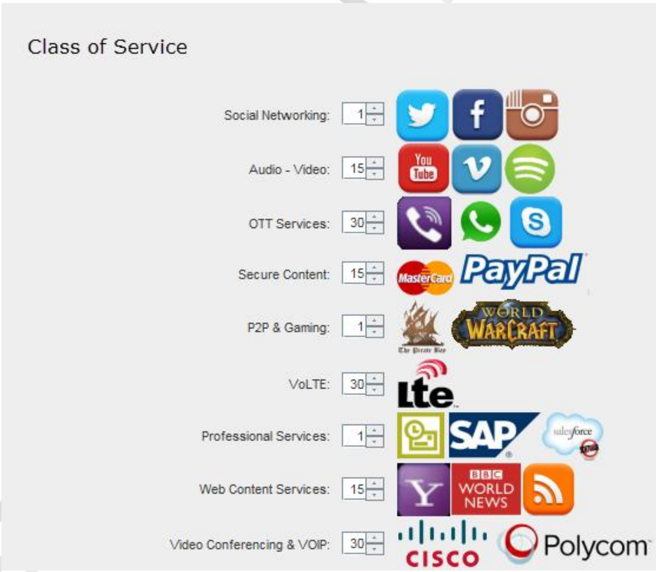

Figure 29 Class of Service definition per Application

As we can see below the lower the value that an application has, the bigger the priority is. Summing the values of app groups, Social Networking, P2P & Gaming, Audio- Video we will have an average of ~6, so by defining the VIP value threshold to 10 we can be sure that the user having the above CoS profile will have better priority and better QoE. Now can respectively change the CoS value in order to favor data consumer that use apps like, *Professional Services, Video Conferencing & VoIP and VoLTE*, by grouping this apps it easy to understand that the target group of our "what if" scenario will be user with age over 30 and probably company executives. The average of the three above applications makes a value of 4. At the figure below, we can have an image of how the CoS will be for this target group.

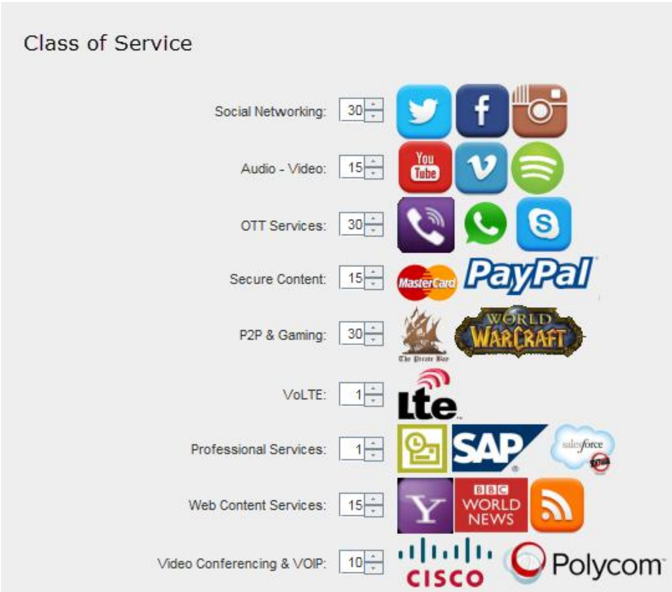

Figure 30 Class of Service modified example

Again with the above calibration of the values CoS and again having the value 10 for the VIP value threshold we can be sure that the VIP mobile user, in our case user of age bigger than 30, will be served with better priority and the will have better QoE.

4.1.3.3 Usage and Bit Rates

It is clear that when we want to run "what if" scenarios the scope is to force data traffic to the network, in order to understand where the bottlenecks of the network are. But the traffic model that we produce in order to push it to the network, is a snapshot of the network so we cannot discuss about the network usage. Considering also the fact that the usage we are referring to has to with the MBs/second.

Now regarding the Bit rates, as we have said in order to push traffic to the network massively we must do it via pattern that is near to the reality. In order to accomplish that, we create potentials future customers and each potential costumer has a random capacity from a range that is easily defined through our front end application. At the figure below, we can see an enlightening example of the user creation process.

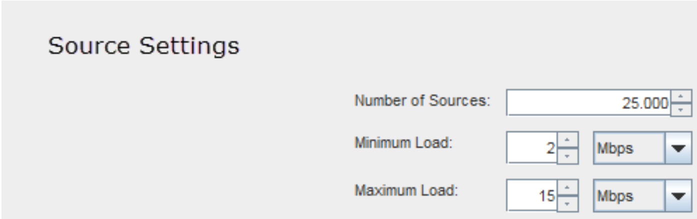

Figure 31 Sources setting attributes definition

At the above figure, we can see that for the new scenario we have chosen 25.000 users as potential new subscribers and these subscribers will be served from the existing network. Each user of the above will take a random value for his capacity from the defined range [Minimum Load…Maximum Load]. And in this way a total traffic will be created and it will be pushed, to the existing network portion that we have previously selected, in order to be served.

4.1.3.4 Rollup

Finally, after following all the above steps we can now run the scenario and push the traffic to the network. Below we can see the imprint of the above steps via print screens, one print screen for each step.

Step1. Selection of the network portion, for which we would like to run the scenario.

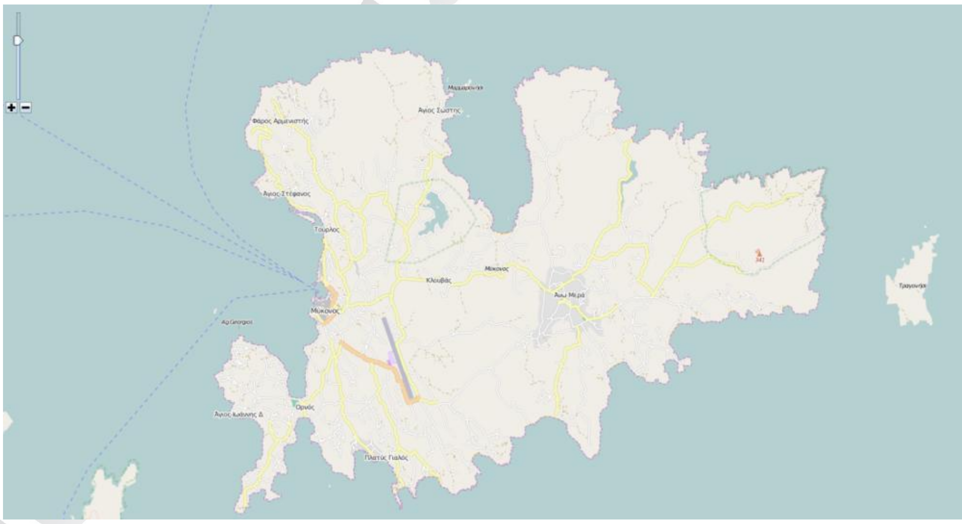

Figure 32 Map view via the application menu

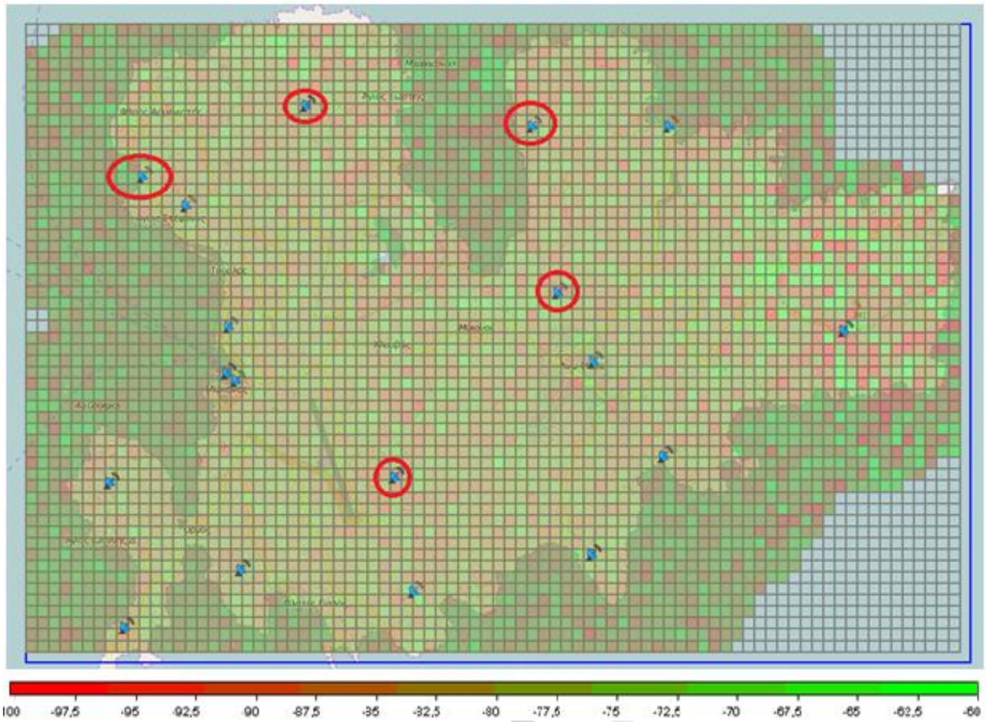

Figure 33 Selected area for the scenario, all the access nodes that exist in the area are projected

At the above print screen, figure 33, we can see that now the area of interest is under a grid. By selecting the area of our interest in order to run a scenario, two things can be seen:

1. The access nodes of the mobile provider's network are projected.
2. Automatically the QoE of the SNR KPI is produced. (see the colored bar above that has values from -60dBm to -100dBm)

   Step2 Define the Class of Service, for the amount of users that we will select in our scenario.

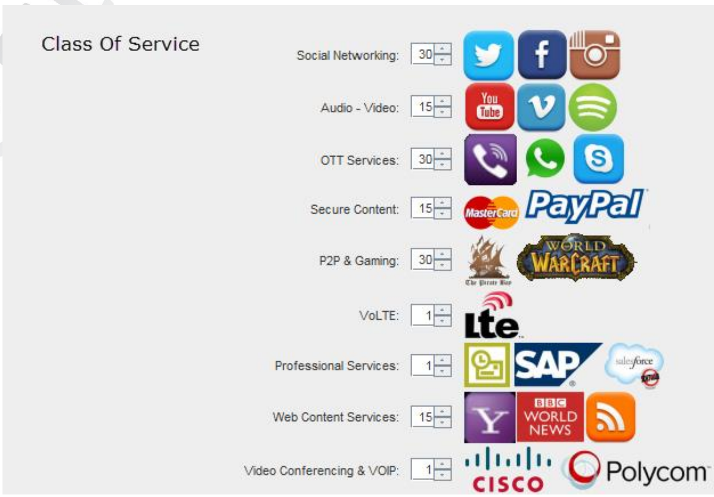

Figure 34 Class of Service for the scenario scope

Step3 Define the amount of Users and the Capacity range for them.

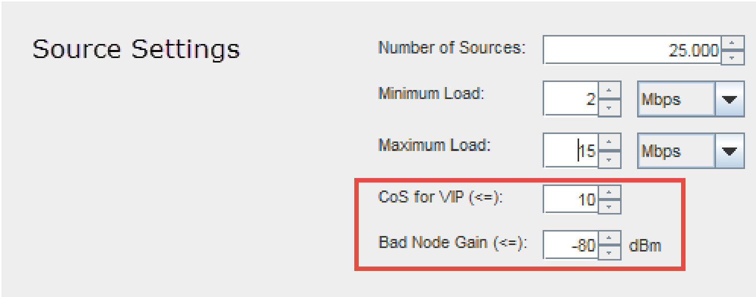

Figure 35 Source setting definition for the scenario scope

As we can see above we have selected 25.000 users with a range of capacity of [2…15] Mb/s for each user. Additionally at the highlighted red area we have defined the value for VIP threshold and the Bad Node SNR.

Step4: Finally, we just press "Run" and the generate traffic is pushed to the area of interest that we have previously selected.

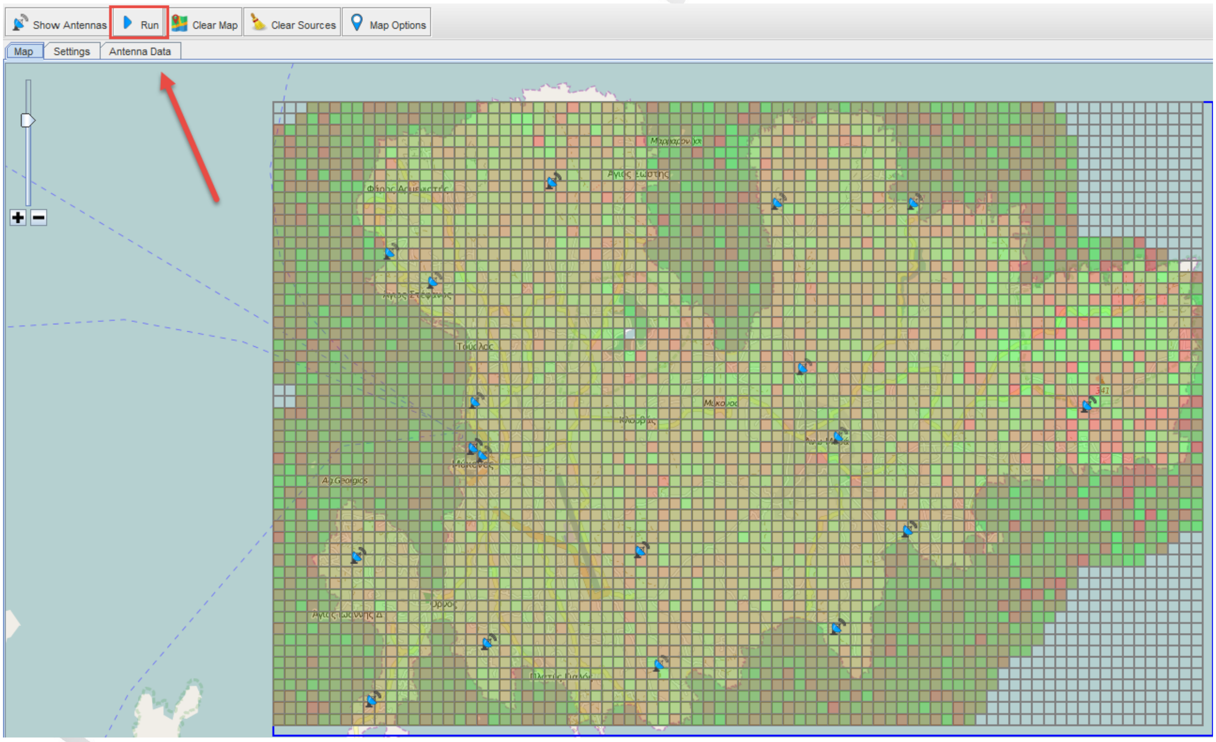

Figure 36 Selected area for the scenario, all the access nodes that exist in the area are projected

More information about the input and the outputs of the application, as well as the mathematical algorithm used by the application, will be analyzed at the following chapters.

4.2 Objective Function

Having analyzed extensively the inputs of the application at the above chapters, now we will emphasize at the objective function of the application. The objective function is the heart of the application. At the illustrative figure 21, it describes the workflow of the application.

As we mentioned many times at the previous chapters, our scope is to create an application in order to run "what if" scenarios for access network of a mobile operator and with the help of a simulated annealing algorithm we could find the *minimum cost* of the selected network portion that is to serve this future traffic load that we have produced during a "what if" scenario. The objective function that will be used in order to achieve this scope is the below:

MinCost=

$$\sum_{j=1}^{?}\left[\left(CAP_{j} / \sum_{i=1}^{i=?} C_{ij} * D_{i}\right) * NodeCost_{j}\right] + \sum_{z=1}^{?}\left[\left(CAP_{z} / \sum_{i=1}^{i=?} C_{zi} * D_{i}\right) * NodeCost_{z}\right] + \sum_{w=1}^{?}\left[\left(CAP_{w} / \sum_{i=1}^{i=?} C_{Wi} * D_{i}\right) * NodeCost_{W}\right]$$

Below you can see an explanation for all the above variables:

i: This index represents each user that is created for our scenario.

J: This index represents each 3G node of the mobile provider's access network.

z: This index represents each 4G node of the mobile providers access network.

w: This index represents each Wi-Fi access point of the mobile providers access network.

C_{ij} : This index represents the link between the mobile user and a 3G node. It takes value either (0,1)

C_{iz} : This index represents the link between the mobile user and a 4G node. It takes value either (0,1)

C_{wi} : This index represents the link between the mobile user and the Wi-Fi access point. It takes value either (0,1)

$NodeCost_{j}$: This index represents the cost of each 3G node that we insert to our application for our scenario.

$NodeCost_z$: This index represents the cost of each 4G node that we insert to our application for our scenario.

$NodeCost_W$: This index represents the cost of each Wi-Fi access point that is created in order to serve the traffic load that is pushed to the network.

D_i : This index represents the capacity that is randomly created for each user at the system. This variable takes values from the range that is given through the application.

CAP_j : This index represents the **free** capacity of each 3G node that we insert to our application for our scenario. These values are taken from the provider's database.

CAP_z : This index represents the **free** capacity of each 4G node that we insert to our application for our scenario. These values are taken from the provider's database.

CAP_w : This index represents the capacity of each Wi-Fi access point that is created. This variable takes values from the range that is given through the application.

In order to have a better understanding of the objective function, we are using it would better to have an example. So let's suppose that we select an area in the map and this area is served from 2 x 3G nodes and 1 x 4G node and respectively cost for each node are: 100.000 for the 3G nodes and 180.000 for the 4G nodes. Additionally let's assume that the 3G nodes have 100 Mb free capacity and the 4G 250 Mb free capacity and we would like to run a "what if" scenario of 500 additional new users to the network that they will produce almost 2.5 Gb of traffic to the network. So with the current selected nodes we have 450 Mb available capacity, and we are called to serve 2.5 Gb of traffic. In order to achieve that either, we will install new nodes 3G and 4G and we will sum the additional cost of those new nodes, or will try to implement Wi-Fi access point in to our network architecture in order to offload the 3G and 4G nodes and lower the total cost of solution that will serve the new traffic demand of 2,5 Gb. We will see at the next chapters that the 2nd option is far cheaper and serves the users with better QoE.

In order to emphasize to the essence of the objective function, we didn't take into consideration the aspects of Class of Service and the oversubscription factor. Those aspects give an image closer to the reality as far as the network but it would make our example too complicated for our 1st example.

Another important detail regarding the previously mentioned objective function, has to do with the "?" symbol at the upper part of the Taylor series i.e. $\sum_{j=1}^{?}$. The meaning of the "?" symbol has to do with the fact that we don't know the amount of 3G, 4G, nodes and Wi-Fi access points that will be used in advance to the scenarios that we are going to run. Below we can see an illustrative workflow chart imprinting the structure of the simulated annealing algorithm.

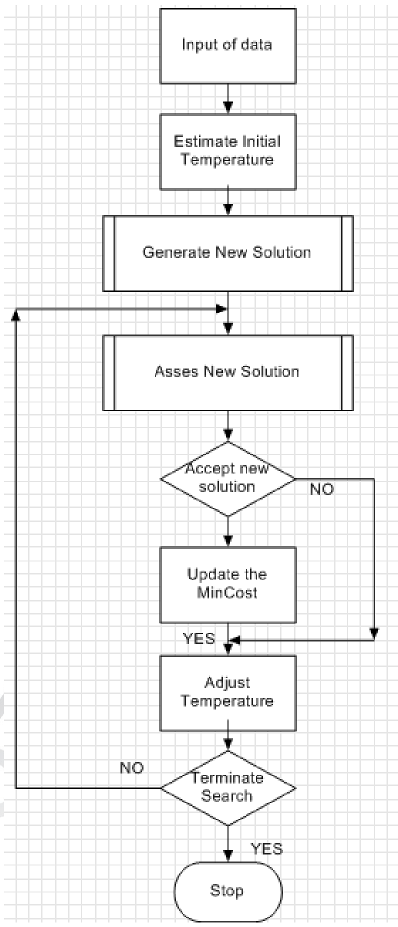

Figure 37 Simulated Annealing algorithm workflow

An interest trick that has been implemented during the design of the algorithm, in order to maximize the use of the selected mobile nodes and in parallel find the minimum cost of the architecture that will serve the traffic load, can be seen inside each Taylor series for example $\left(CAP_j / \sum_{i=1}^{i=?} C_{ij} * D_i\right)$ with this math, we are taking advantage of the oversubscription factor. At the numeric part of the fraction we have the free capacity of the access node (3G, 4G, or Wi-Fi) and at the denominator part of the fraction we have the sum of the produced user traffic the result of this fraction represents the oversubscription factor. Now after calculating this fraction, we multiple it with the cost of each respective node, with this way we achieve that the algorithm will be forced to utilize the access node to the fullest extent before proceed to the next access node. This is happening because during this calculation the cost of a node was used and the cost is main driver for our algorithm, so be trying to minimize the cost we achieve the maximum utilization of a node.

4.3 Constraints

One crucial part of the algorithm except the objective function is also the constraints. Below we will analyze one by one these constraints. First of all one cell phone (in our case one user) cannot be assigned to more than 1 access node in parallel. At the below table there is an example in case of 3G access nodes.

	mobile Users											
3G Access nodes	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12
3G node 1	0	1	0	0	1	0	0	0	0	1	0	0
3G node 2	1	0	0	0	0	0	0	1	0	0	0	0
3G node 3	0	0	0	1	0	1	0	0	1	0	0	0
3G node 4	0	0	1	0	0	0	1	0	0	0	1	1

Now from the above table we can easily see that each mobile user is connected only at one access node simultaneously. This is also a fact for all the access nodes that exist at the running scenario i.e. 3G, 4G, Wi-Fi. So the constraint as far as user assignment to access nodes is:

$$\sum_{j=1}^{j=?}\sum_{i=1}^{i=?} C_{ij} = 1, \quad \sum_{z=1}^{z=?}\sum_{i=1}^{i=?} C_{zi} = 1 \quad \text{and} \quad \sum_{w=1}^{w=?}\sum_{i=1}^{i=?} C_{wi} = 1$$

With the above mathematical relation, we are assure not only the each user will be assigned to only one node but also that all the user will be assigned to an access node, no user will left unassigned. Now another constrain of the algorithm has to do with the oversubscription factor, it can never be bigger than the defined value at application. Below you can see an illustrative image from the application and specifically from the point where we define the oversubscription threshold factor.

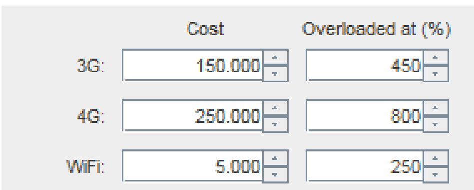

Figure 38 Oversubscription factor definition menu for all access technologies

So in the upper case the threshold for the oversubscription factor regarding 4G nodes is 800%. The constraint in this case will be like:

$$\sum_{j=1}^{?}\left(CAP_z \,/ \sum_{i=1}^{i=?} C_{zi} * D_i \right) \leq 800\%$$

In no case the total capacity of the users that are assigned to a node cannot be bigger than the defined threshold. In our case this threshold is the oversubscription factor as we can see it at the above example. Another constraint of the algorithm has to do with the CoS that the user belong. With the *CoS tagging* of the users we can differentiate the users into two categories VIP and non-VIP. So the constraint here is that a non-VIP user cannot have priority over VIP users. For example when we create a scenario of 10.000 users for a specific portion of the provider's network at the map and for a specific QoE as far as this are of the map, the VIP users will be served first and they will be handle from Wi-Fi access point and 4G access nodes and in some rare cases from 3G access nodes also. After all the VIP users will be served, then the assignment of non-VIP user starts to remaining nodes that have still free capacity. If there are no access nodes with free capacity, additional Wi-Fi nodes will be created in order to serve all the users of our scenario. One other constraint of the algorithm has again to do with a threshold of the application and this is the SNR KPI. Again as the VIP threshold, the SNR KPI has to with the filtering of the users, users that are experiencing worse SNR than the value defined at the SNR KPI will be moved into Wi-Fi access nodes in order to have better QoE, after all the target of the algorithm is not only to find the minimum cost for the traffic load that is to be served but also to combine this minimum cost with maximum possible QoE for the users. Such an example regarding the QoE constraint can be seen with the below Taylor series.

$\sum_{i=1}^{i=?} i \leq$ SNR KPI, which in our case will be custom and can be changed via the application each time we would like to run a scenario. An illustrative figure of a potential SNR KPI value can be seen at the figure below.

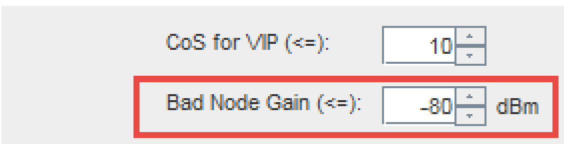

Figure 39 SNR KPI threshold definition menu

4.4 Proposed Algorithms

To solve the problem of finding the minimum cost we will use a simulated annealing algorithm. The structure of the algorithm will be analyzed further below but, before explaining how we solved our problem with the use of a simulated annealing algorithm we will make a small introduction of what is a simulated annealing algorithm.

4.4.1 What is Simulated Annealing

Simulated Annealing (SA) is a generic met heuristic algorithm that is used to find a very good result very close to the optimal solution of a given function. Usually simulated annealing it is used when the search space is discrete (for example, assign sources to concentrators), in cases like this simulated annealing is considered to be more efficient than exhaustive enumeration and that is because our goal is to find an acceptably good solution in reasonable time, rather than find the best possible solution in infinite time.

The inspiration for the algorithm naming, come from annealing in metallurgy. This technique involves heating followed by controlled cooling of the material, this process takes place in order to increase the size of its crystals and reduce its defects. Both of these attributes depend on the thermodynamic free energy of the material. Cooling and heating the materials affects both the thermodynamic free energy and temperature. The annealing algorithm simulates the cooling process of the material by lowering gradually the temperature until it converges to a steady and frozen state. This concept of slow cooling is the fundamental attribute of Simulated Annealing algorithm, while this cooling process there is a probability of accepting worse solutions. The acceptance of worse solutions is a fundamental property of met heuristics because it leads to more extensive search for the optimal solution.

In 1982, Kirkpatrick copied the main idea of the “Metropolis” algorithm and used it to optimization problems. The main idea was to use simulated annealing searching for acceptably good solutions, and then all this solutions converge to an optimal solution. Simulated annealing is described in many textbooks.

4.4.2 Analysis of the select SA algorithm

After having an academic introduction to what simulated annealing means, we will refer to how our algorithm was designed what are the decisions that are taken throughout the algorithm in order to get the desired result. To have a better understanding of the algorithm workflow we can have a look at the below figure.

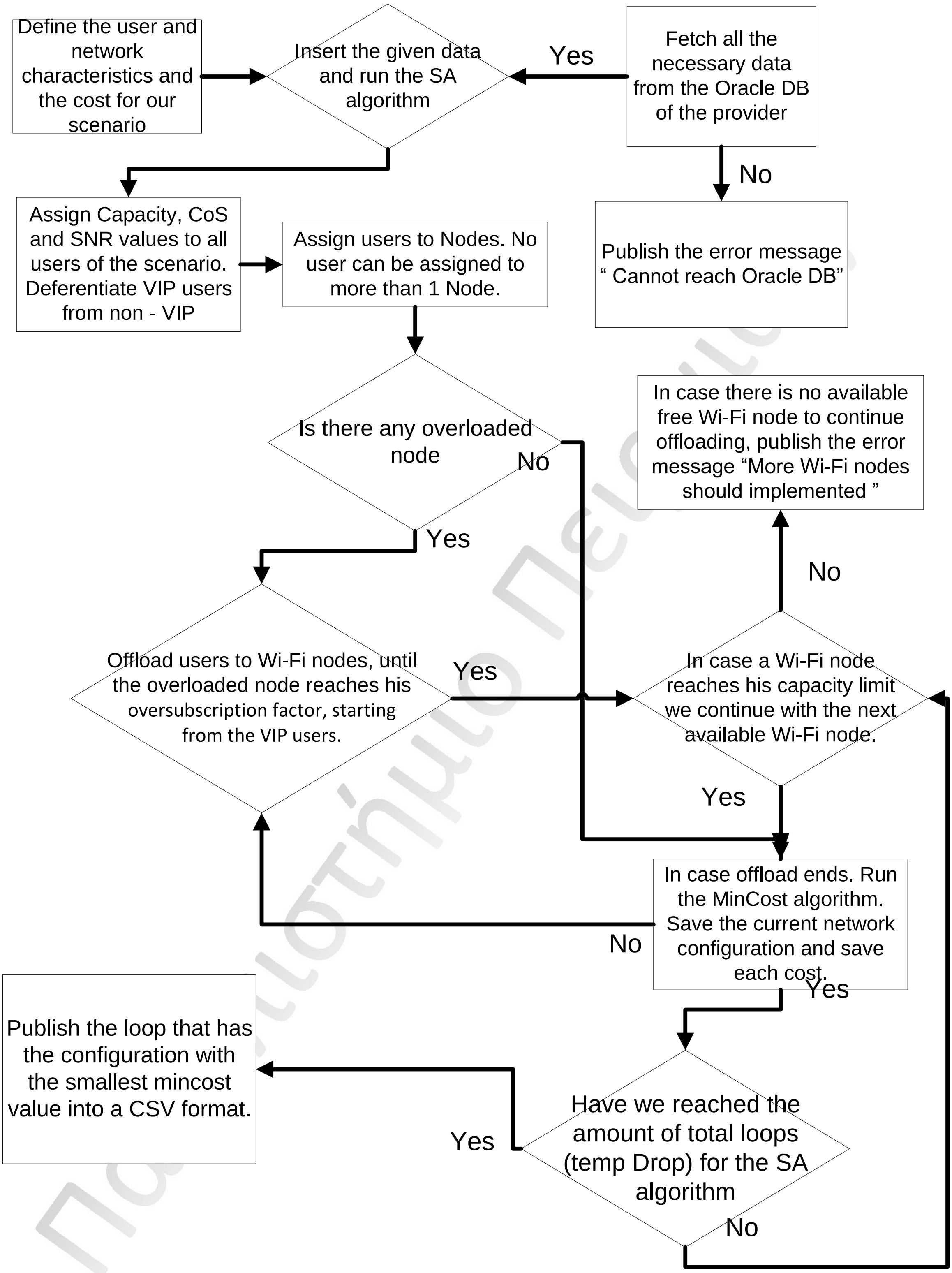

Figure 40 Simulated Annealing detailed workflow chart

It is quite obvious from the flowchart that the SA algorithm process is not so simple, but though its complexity we can ensure that we will achieve a more accurate result.

In order to simplify the SA workflow, it will be better to subdivide it to below steps:

- ✓ Data entry
- ✓ Data processing
- ✓ Data export (reporting)

Some of the data came from the provider's database. Such data are the total data capacity of the access nodes, the coordinates of the access nodes and the free capacity of each node. Except the data that come from the provider's database we also have the values that user's entry each time, depending the scenario that they want to run, for example:

- The CoS value definition for each Service.
- The value for the minimum and maximum capacity of the users that we will be created.
- The cost of the access nodes (3G, 4G and Wi-Fi).
- The minimum and maximum capacity that Wi-Fi access nodes could serve.
- The range of the Wi-Fi access nodes that will be created in order to serve all the users per scenario that we want to run.
- Threshold definition for VIP user and SNR value.
- Selection of the map area for which we would run the scenario.

Now after having all the data being entered to our application, the SA algorithm could start. During the execution of the algorithm, as you can see at the flow chart above, the algorithm is checking if the current network infrastructure could serve the "new" data demand. In order to that it is assigning all the users that where created to access nodes (3G, 4G) and it checks if the load of all the assigned users per access node exceeds the capacity of the access node multiplied with the oversubscription factor. In case an access node (3G, 4G) is "overloaded" the algorithm will create a Wi-Fi access node in order to offload it, the offload will start from the "VIP" users and after them the users that their QoE SNR KPI is over a predefined value. During the previous process in case the Wi-Fi access node is "filled" with users, i.e. the capacity of the Wi-Fi multiplied with the oversubscription factor is less than the load of the users that are offloaded, the algorithm will create and additional Wi-Fi access node and it will proceed with the offloading of the remaining users of the "overloaded" access node or it will proceed to the next "overloaded" access node (3G, 4G). The previous process will continue to run until all the access nodes (3G, 4G) stop to be "overloaded" and that will be only one loop of the SA algorithm. After the end of each loop the **MinCost** is executed and the value it is stored. After temp of the SA algorithm reaches zero (0) the algorithm stops and it comparers all the stored values of the **MinCost** and export the smallest and with the network configuration that gave us this result.

Having analyzed the data processing, now we will analyze the final step which is the data export. As we have mentioned from the beginning of this master Thesis, their scope is to help the provider to run "what if" scenarios of future data traffic and how this will be handled by HetNet architecture. After the end of SA algorithm the data that will be published have to do with the optimum solution of the network cost, considering the data traffic that will handle. In order to have a clearer image of the SA export, there is an illustrative figure below.

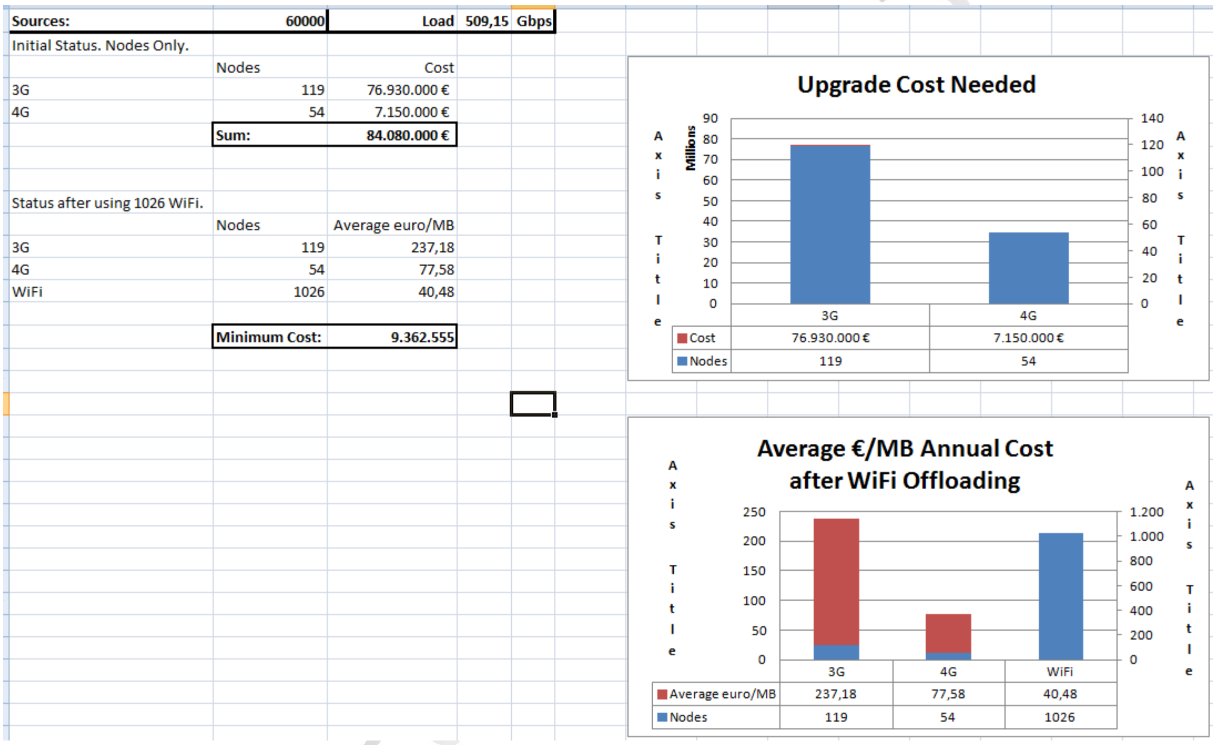

Figure 41 Software exported results

What we can see at the above figure 39, this is the export of a "what if" scenario where we forecasted an addition of 60.00 concurrent served users from a network portion, the users will produce a traffic load of 509,15 Gb/s. At the above export we see two costs, the first cost is reflects the cost that the provider has to pay in case it would not use a HetNet solution and decide to upgrade the existing 3G or 4G nodes in order to handle the traffic load and the second costs reflects the choice of offloading the additional traffic to Wi-Fi access nodes. There is a significant difference between the two solutions as far as the costs, the choice of adopting a HetNet solution means almost ten times less cost for the same served traffic load. More scenarios and analysis of forecast capabilities that the application has will follow at the above chapters.

5 Software Tool

During this chapter, we will cover all the aspect of our application:

- ✓ How it is structured
- ✓ Which is the programming language that is used?
- ✓ With which 3rd party applications it has interface in order to input data.

Our software is a java based application that has an interface with an Oracle 11g enterprise DB, in order to import the network architecture of the provider and all the access nodes attributes. Additionally all this network architecture is projected to integrated maps. Moreover, for a selected portion of the network at the map we simulate the users QoE KPI of SNR. Finally, one more important part of the application is the traffic generation, the network thresholds definition and Policy Control. Below you can see an illustrative example of the application inputs and the structure philosophy.

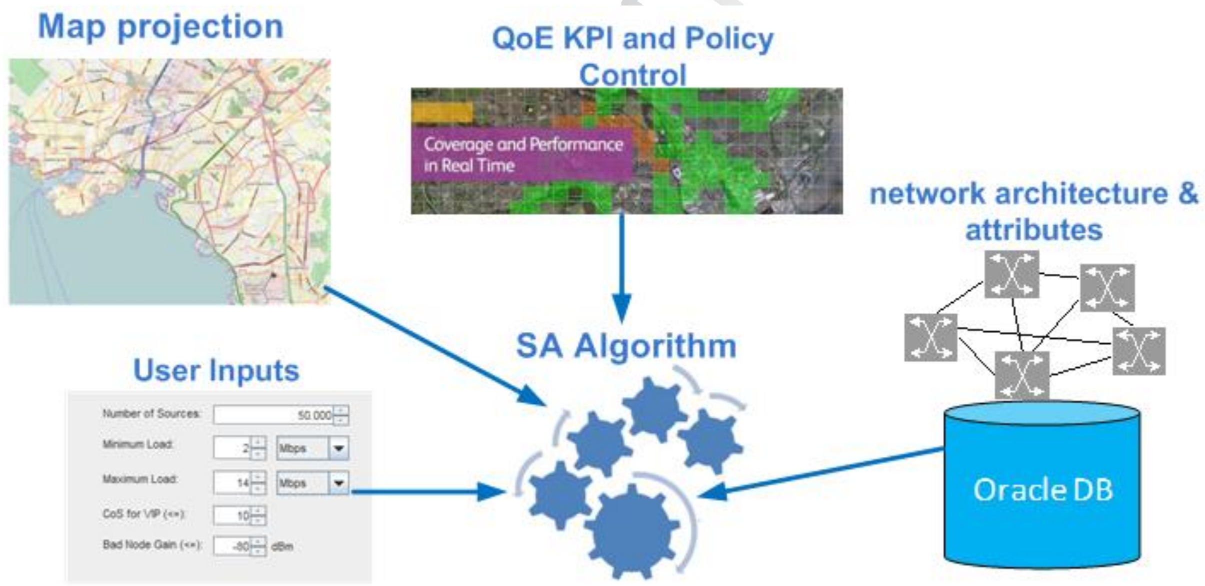

Figure 42 Software Inputs

5.1 Database Connection

As previously said our software was connected with an Oracle 11g DB in order to import all the necessary data so as to proceed with the traffic load forecast. In order to be able to connect to an Oracle 11g DB we had to build within the software a JDBC driver for 11g DB. Below you can see a part of the JDBC driver code:

```
connection = DriverManager.getConnection(
```

```
"jdbc:oracle:thin:@localhost:1521/xxxxx", "username",
                                "password");
```

Except connecting to the data, we also had to create a database View in order to collect all the data we need and are scattered in the database. Now all the access node attributes like, access technology (3G, 4G), free node capacity, node coordinates and node name, can exist at single point at the database. Having all the needed information at single point, a database view, there is no need for a complex SQL query statement, and so better application could be experienced.

5.2 QoE Feedback and User Inputs

Quality of Experience (**QoE**) is a measure of a customer's experiences with the service (web browsing, phone call, app performance, live streaming). QoE focuses on the entire service experience, and is a more holistic evaluation than the more narrowly focused user experience (focused on the software interface).

Nevertheless, QoE feedback is extremely helpful and efficient from network planning perspective. Let's assume that we are engineers and a providers planning department and we are going to choose an area, where we are going to upgrade massively all the access nodes to this area in terms of data capacity. But what meaning will this capacity upgrade will have when all the users that are served from the access nodes at this area, are experiencing poor browsing performance due to bad coverage, so in that case any additional capacity upgrade will not have any improvement for users. Based on the previous scenario it is obvious that QoE is extremely useful for network planning and can lead to better and more precise network planning.

In our case there is no interface with "robots/probes" or platforms that could feed our application with QoE information. So for that reason, we randomly are producing QoE information for the SNR KPI. We could have used more KPIs like:

- ✓ Loss packets
- ✓ Max and average jitter
- ✓ Average latency
- ✓ Connection latencies
- ✓ Time to the first byte of data received
- ✓ TCP retries
- ✓ TCP connection failures

But due to the fact that all the values for the above KPIs will be reproduce randomly and they will not be taken from a QoE platform, it is sure that it will add more complexity to our application and it is questionable how this complexity will lead to better conclusions. So for that reason only one KPI was chosen as QoE feedback and this is the SNR KPI. So every time we would like to run a forecast scenario for a network portion, by the time this area of study is selected at the map and all the access nodes of this area are fetched to the map, right after the SNR KPI is produced and projected at the map. For a better understanding of this description there is an illustrative example below.

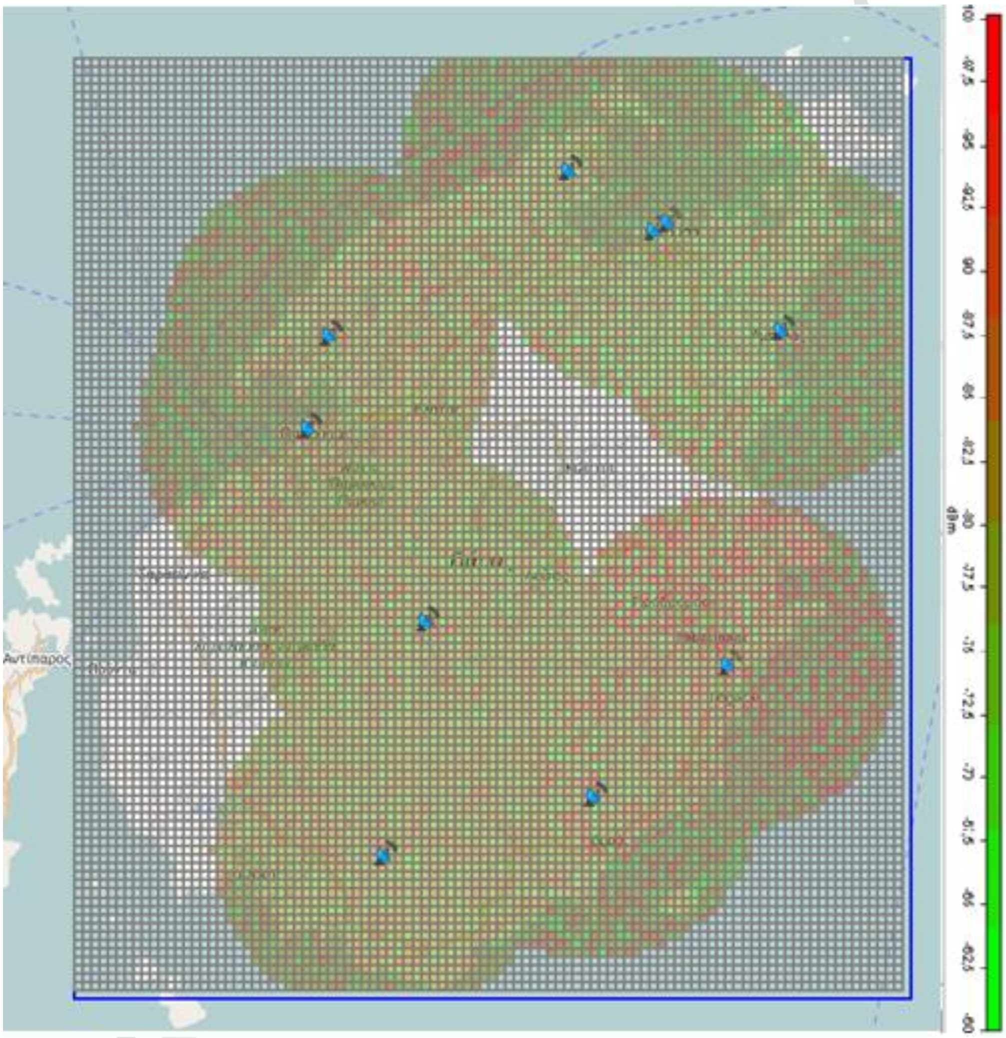

Figure 43 Network portion that is selected for the scenario.

At the above, figure we have selected an Island for our scenario and we can see the radius of the antennas and also the SNR KPI QoE for the selected area. So now the only thing that the engineer that want to run the scenario is to define the values for the Policy Control and also the values that will lead to future traffic load creation.

Now for the previously selected area let say that we would like to run a scenario for young people during the period of vacations. In order to achieve that will have to differentiate users into VIP and non-VIP. It is commonly known that young people are consuming their mobile data at social media, OTT services and video streaming applications, so we will give priority to those applications at our scenario and we will tag all the users of this group of applications as VIP users. Such user classification can take place at the CoS section of our application. An illustrative example of young people being classified as VIP for our scenario can be seen below.

Figure 44 Adapt application values in a way that suits our scenario

Now after defining the CoS value per application, the only thing is missing in order to differentiate VIP users from non-VIP, is to define the VIP threshold value over which users are considered as non-VIP users. Except the VIP threshold definition, the last step before the application starts to run the scenario is to define those values for the future data traffic creation as well as the values for Wi-Fi creation and the oversubscription factor of the current access nodes. An illustrative image of those data input described above can be seen below.

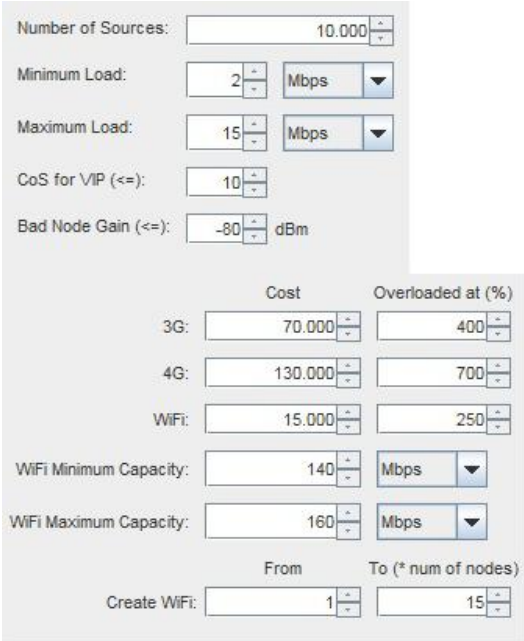

Figure 45 Network and User values definition according to scenario needs

At the above figure we have selected an addition of 10.000 concurrent network users for our scenario that will produce a traffic load, we have also define the minimum and maximum data capacity values of the carrier Wi-Fi access points and we can see also the oversubscription factor value for all the access nodes of the network being defined.

Finally, we have the “total” cost per access node technology defined. By mentioning “total” cost we mean, the access node cost, the installation/implementation cost and the backhauling cost. Again, one more important aspect of the application that is seen above is the range of the Wi-Fi nodes that will be created in order to offload the traffic from the mobile access nodes (3G, 4G). The exact amount of the Wi-Fi access nodes that will be created will be a result of the selected mobile access nodes for this scenario multiplied by a number from a given range. With this way, we give the opportunity to the algorithm to find the best solution for our scenario, instead of us giving a static value for the Wi-Fi access nodes that will be created. In this way we will have 2 solutions with our export, the snapshot of the access network that is capable of handling this future traffic load and of course the cost of this network, which in our case will be the least possible. More details about the applications export and forecast of network bottlenecks and costs will follow at the next chapter.

6 Use cases and Results

At this chapter we will try to run several scenarios in order to give a clear picture of the application forecasting capabilities as far as future traffic loads and also the ability not only to find the optimum solution for the network but at same time the minimum possible from cost perspective with the best possible QoE for the user.

6.1.1 First scenario

During this scenario will try to see a scenario that takes part at a Greek island, and we will try to see how the network could handle the traffic of 10.000 additional concurrent users. Statistically, this will mean that 100.000 new users will be added to the network, but a concurrent usage ratio will be close to 1:10. We could see below an illustrative image of the selected map area for which we would like to run our scenario.

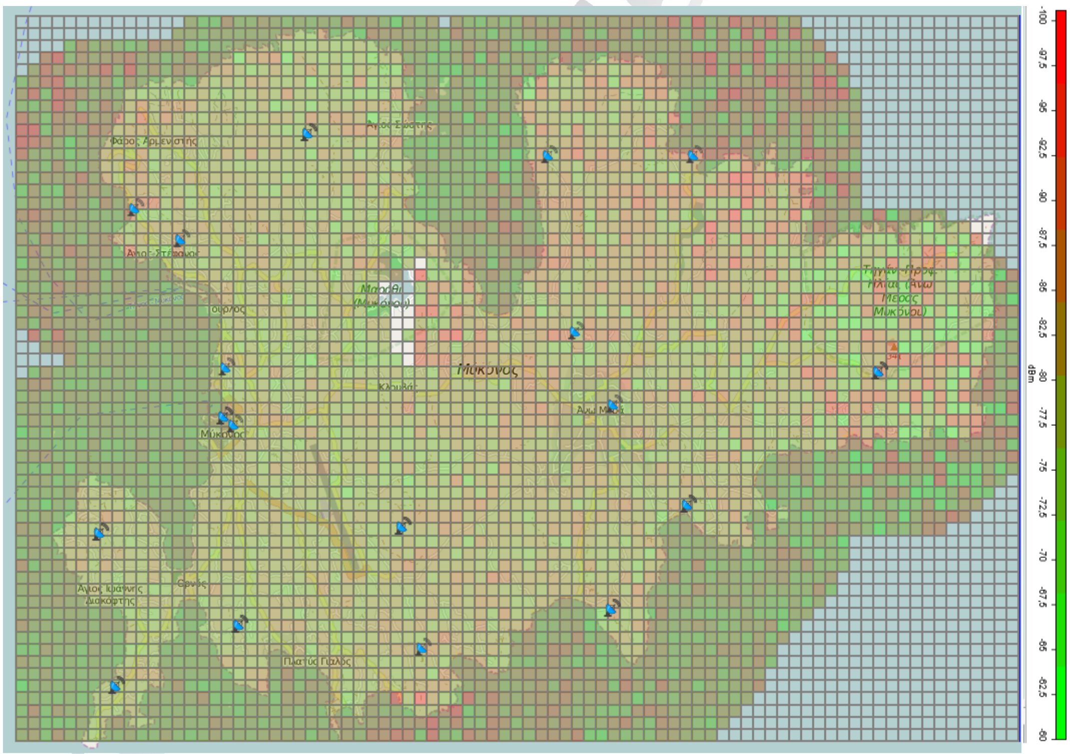

Figure 46 Network selection via the map for the scenario needs, KPI QoE projection on the map

The target user profile (VIP users) for our scenario will be the "professional" users, user that are using more professional application like SAP, Microsoft Outlook and Video Conferencing services instead of OTT and Social media apps. During this scenario will see several attempts with slight changes to the input values, which will lead us to understand better the potentials of the application. For our 1st attempt, the algorithm will run for 1.000 loops in order to find the optimal solution. Please see below and illustrative image of the configured CoS of our scenario.

Figure 47 Application values definition for the CoS based on the scenario scope

As far as the user and the traffic profile of each user, we have a given a range of capacity from 2Mb/s to 14 Mb/s for each user of our scenario. This traffic that each will produce is selected randomly within the previously mentioned range. Also for the 1st attempt of this scenario will define the value of 10 for the VIP threshold and -80 for the SNR KPI. We can have a look at figure 46 below.

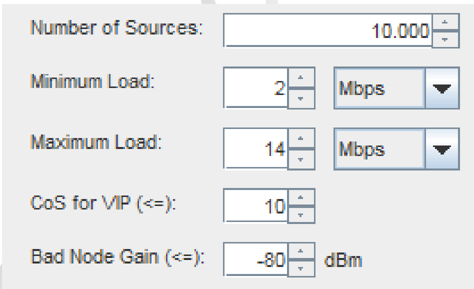

Figure 48 Resource value definition according to our scenario

Now after defining the user values, now we will define the cost values for our access nodes, the oversubscription factor per node technology and Wi-Fi the access nodes capacity for the current scenario. There is an illustrative figure below regarding the previously mentioned configuration.

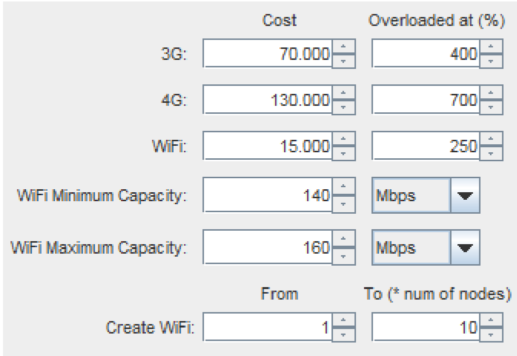

Figure 49 Network values definition according to our scenario

After defining all the input values for the network and the users we will run the algorithm for 1.000 loops. We can have a look below at the excel output of the algorithm.

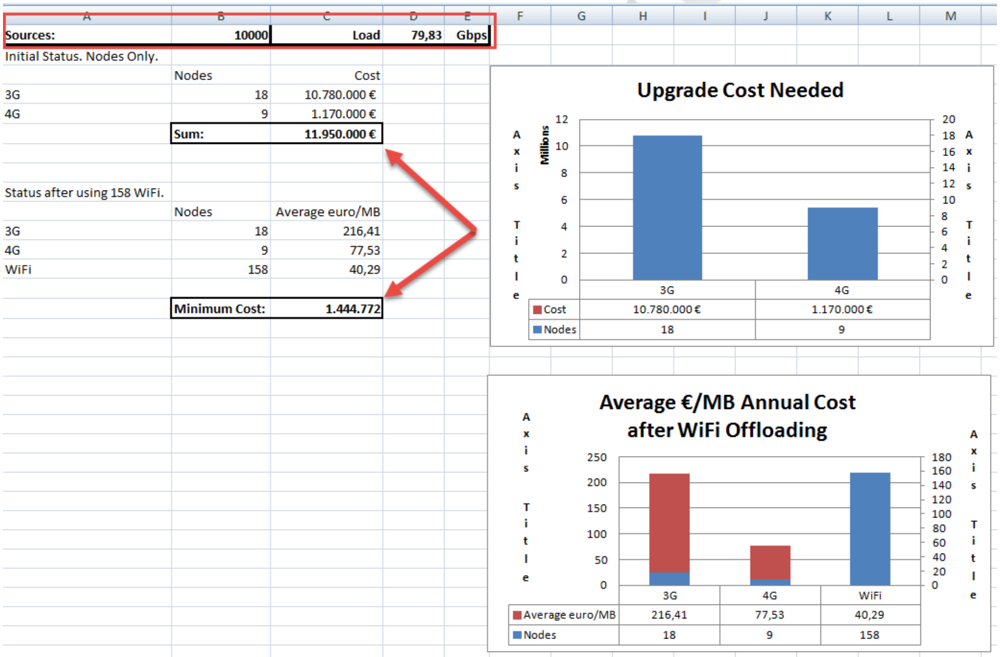

Figure 50 The export of the scenario

From the above export, mainly two conclusions may come out. One has to do with produced traffic load of the 10.000 concurrent network users, which is 79,83 Gb/s. It is far easier to understand that the network has to handle 79,83 Gb/s of traffic load, than saying that we should be able to serve 10.000 additional concurrent network users.

The second conclusion has to do with the cost of the solution. What we mean with the term "cost of the solution"? Considering the fact that the network has to handle 79,83 Gb/s there are two possible solutions, one is to upgrade the existing 3G and 4G access nodes with the cost of 11.950.000 euro and other one is to "offload" most of this traffic to the carrier Wi-Fi network of the provider with the cost of 1.444.722 euro. The difference of the two solutions is enormous and all this was accomplished with the maximum possible QoE for the users, since we differentiate the users to VIP and non-VIP, so the "professional" users which we are our target group for this scenario were served with the highest priority. Additionally all the users that were experiencing SNR values over -80dBm o their mobiles, were offloaded to the carrier Wi-Fi network of the provider.

Now in order to understand the potentials of the application and also test the efficiency of the algorithm, we will keep the same configuration as far as the user and CoS of the users but we will change amount of loops that the algorithm will run in order to find the optimal solution. At the next attempts we will also play with the oversubscription factor of the access nodes. We can see below the results of the application, for 2.000 loops and 4.000 loops and compare if there are better results as the loops are rising.

A	B	C	D	E
Sources:	10000	Load	79,24	Gbps
Initial Status. Nodes Only.				
	Nodes	Cost		
3G	18	10.570.000 €		
4G	9	1.170.000 €		
	Sum:	11.740.000 €		
Status after using 156 WiFi.				
	Nodes	Average euro/MB		
3G	18	217,73		
4G	9	77,53		
WiFi	156	40,16		
2.000 Loops	Minimum Cost:	1.437.014		

A	B	C	D	E
Sources:	10000	Load	79,16	Gbps
Initial Status. Nodes Only.				
	Nodes	Cost		
3G	18	10.570.000 €		
4G	9	1.170.000 €		
	Sum:	11.740.000 €		
Status after using 156 WiFi.				
	Nodes	Average euro/MB		
3G	18	217,48		
4G	9	77,57		
WiFi	156	40,21		
4.000 Loops	Minimum Cost:	1.434.487		

Figure 51 Comparing the difference between 2.000 and 4.000 algorithm loops performance

For the 2.000 loops attempt we see that there is a small reduction of the cost of the optimal network solution, rate of this reduction is approximately 0,5%. And for the 4.000 loops attempt we seen additional cost reduction for the optimal solution and the rate of this reduction compared with the 1st solution is approximately 0,7%. Those two new results show as two things:

- ✓ The efficiency of the algorithm as far as the optimal solution finding is very high considering that we have run two additional scenarios with 2 or 4 times more loops and the results we almost identical.
- ✓ Secondly, we can see the dynamic of the algorithm, which always improving himself when we let him run more loops.

After testing the dynamic of the algorithm by playing with the loops, now we will try to see the difference between the costs while changing the oversubscription factor. We are sure that the cost will fall, the scope is not see that, but to emphasize the flexibility of the algorithm to run several scenarios which will allow the provider to make strategic decisions about each own network, for example changing the oversubscription factor to his network from now and on in order to lower the "cost" of the network. We can see below the new values that are defined as far as the oversubscription factor.

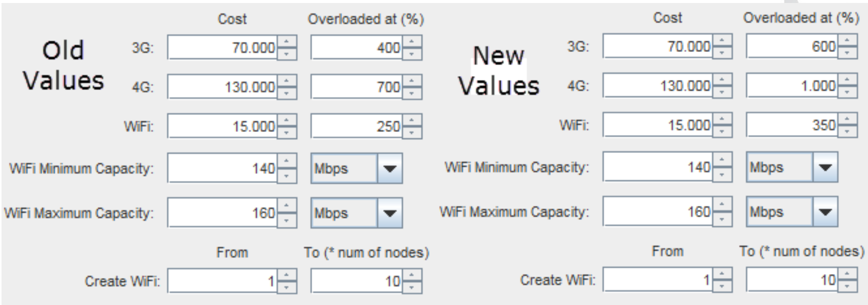

Figure 52 Projection of the Old and New value of the network values

Now after changing the oversubscription factor for all the access nodes of the network, we can see below the difference of the new cost compare to the 1st scenario that we had with the old values.

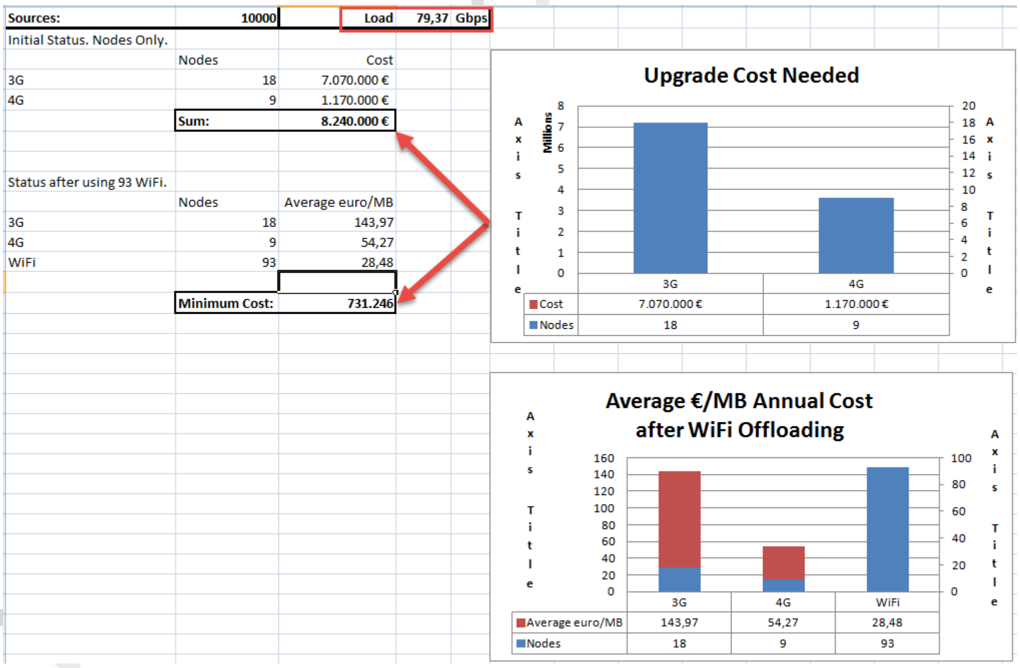

Figure 53 New application export based on the updated values

Comparing the above result with the 1st attempt that we had, we can see that by changing the oversubscription factor the “minimum cost” found from the algorithm when using a HetNet solution has dropped at half, from 1.444.722 euro to 731.246 euro.

Now let's play with the thresholds that we have defined for the VIP users. Let's see what we will mean for the network if we implement a more loose policy as far as the VIP users and we raise the VIP threshold from 10 to 15. In order to have a consistency at the tests, we will roll back the oversubscription factor values to the firstly defined values of our scenario.

A	B	C	D	E
Sources:	**10000**	**Load**	**79,87**	**Gbps**
Initial Status. Nodes Only.				
	Nodes	Cost		
3G	18	10.780.000 €		
4G	9	1.170.000 €		
	Sum:	**11.950.000 €**		
Status after using 158 WiFi.				
	Nodes	Average euro/MB		
3G	18	217,71		
4G	9	77,60		
WiFi	158	40,22		
	Minimum Cost:	**1.446.886**		

Figure 54 Export of the application with a new VIP threshold

There is no change the results are identical. The slight difference that can be seen is within the range of mathematical error.

6.1.2 First scenario conclusions

Here we will analyze the results that we have after the several small changes of some network and user attributes of the application.

For example, we have seen that although we have raised 2 and 4 times the loops that the application could run until it founds the optimal cost and network solution, there was no significant change at the network cost. We can see below a table where we compare the costs that we took from the application for 1000 loops, 2000 loops and 4000 loops.

		1000 loops	2000 loops	4000 loops
	Nodes	Cost	Cost	Cost
3G	18	10.780.000 €	10.570.000 €	10.570.000 €
4G	9	1.170.000 €	1.170.000 €	1.170.000 €
	Sum:	**11.950.000 €**	**11.740.000 €**	**11.740.000 €**
		After using 158 Wi-Fi		
	Nodes	Average euro/MB	Average euro/MB	Average euro/MB
3G	18	217,71	217,73	217,48
4G	9	77,60	77,53	77,57
WiFi	158	40,22	40,16	40,21
	Minimum	**1.446.886**	**1.437.014**	**1.434.487**

Figure 55 Pivot table of all the used loops

As we can see at the above figure55 and also mentioned at the chapter 6.1.1 we can see that the improvement of cost when stepping to more loops is very small.

Something that we didn't comment so far is the exported values except the costs. At the export of the application each time we run a scenario we can see the amount of the Wi-Fi that were used in order to offload the most of the forecasted traffic load from the 3G & 4G access nodes. Below we can see an illustrative picture of such information.

Sources:	**10000**	**Load**	**79,83**	**Gbps**
Initial Status. Nodes Only.				
	Nodes	Cost		
3G	18	10.780.000 €		
4G	9	1.170.000 €		
	Sum:	**11.950.000 €**		
Status after using 158 WiFi.				
	Nodes	Average euro/MB		
3G	18	216,41		
4G	9	77,53		
WiFi	158	40,29		
	Minimum Cost:	**1.444.772**		

Figure 56 The amount of the needed WiFi access nodes for the offload of the traffic.

The information that comes from the above figure regarding Wi-Fi nodes is very useful reference for a provider. Such information can be forwarded to the design department when the future dimensioning of the network will take place. It is a strategic advantage for the provider to have the knowledge of the network expansion and each cost in advance.

Additionally important information has to do with the Cost/MB for each access node technology (3G, 4G and Wi-Fi). Although the cost of the nodes is calculated at yearly base and the traffic load is a snapshot of real time transmitted traffic. However, it continues to be a very helpful index of the cost per served MB, which can be used as is to **other tools** that the provider has for **financial forecasting**. This is another indication of the application's power and usability.

6.2.1 Second Scenario

During this chapter, we will repeat the same tests as at the previous chapter, but as an area now we will not use an Island but the city of Athens. This decision was taken in order to stress the application as much as possible. And the enormous network of Athens is the best use case for that. Additionally now that the examined area is so big we will also raise the forecasted traffic load and in order to do that we will raise the potential users that the network will have to serve, of course again we are referring to concurrent users. We can see below for which portion of Athens network we will run the scenario.

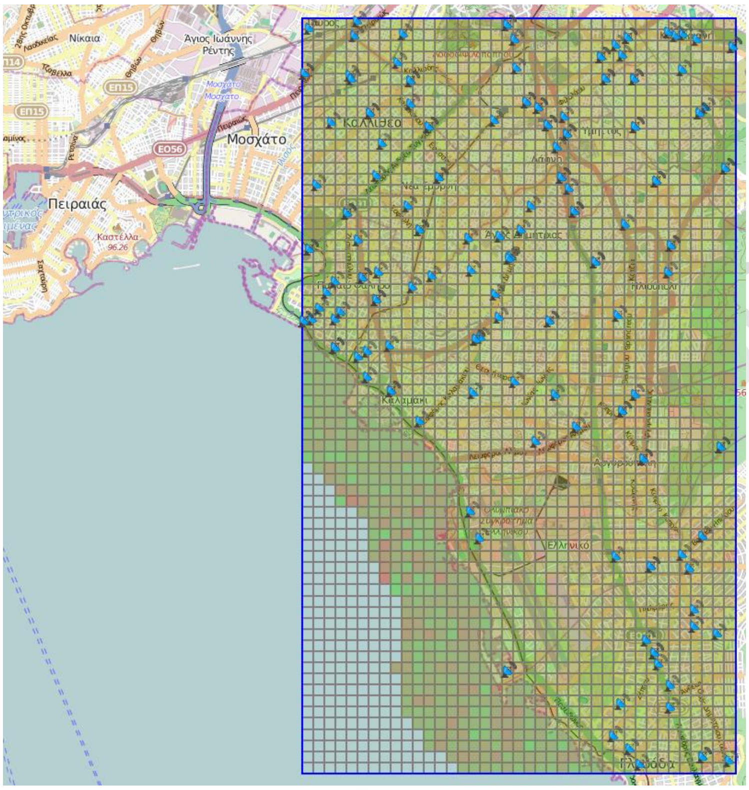

Figure 57 Network selection from Athens area.

At the above figure, we can see that the portion of the network for which we are going to run our scenarios for this chapter is the southern suburbs of Athens. Now our scenarios will be based on the need of serving additionally 150.000 concurrent new users. This amount of the users will produce a specific amount of traffic that the network is called upon to support. As 1st attempt of our scenario, we will follow the same approach as with the previous scenario and will assume as VIP users, the users that are using professional services. Below we can have a look at an illustrative image of the scenario users CoS (Class of Service) profile.

Figure 58 Definition of the CoS value for the scenario

And of course the network attributes and user values definition can be seen at the below figure.

Figure 59 User and network attributes definition for the scenario

After defining all the network and user values that are need so the application could, we can see the export of the algorithm for selected network portion of Athens. We can see below the export of the Simulated Annealing algorithm for our scenario.

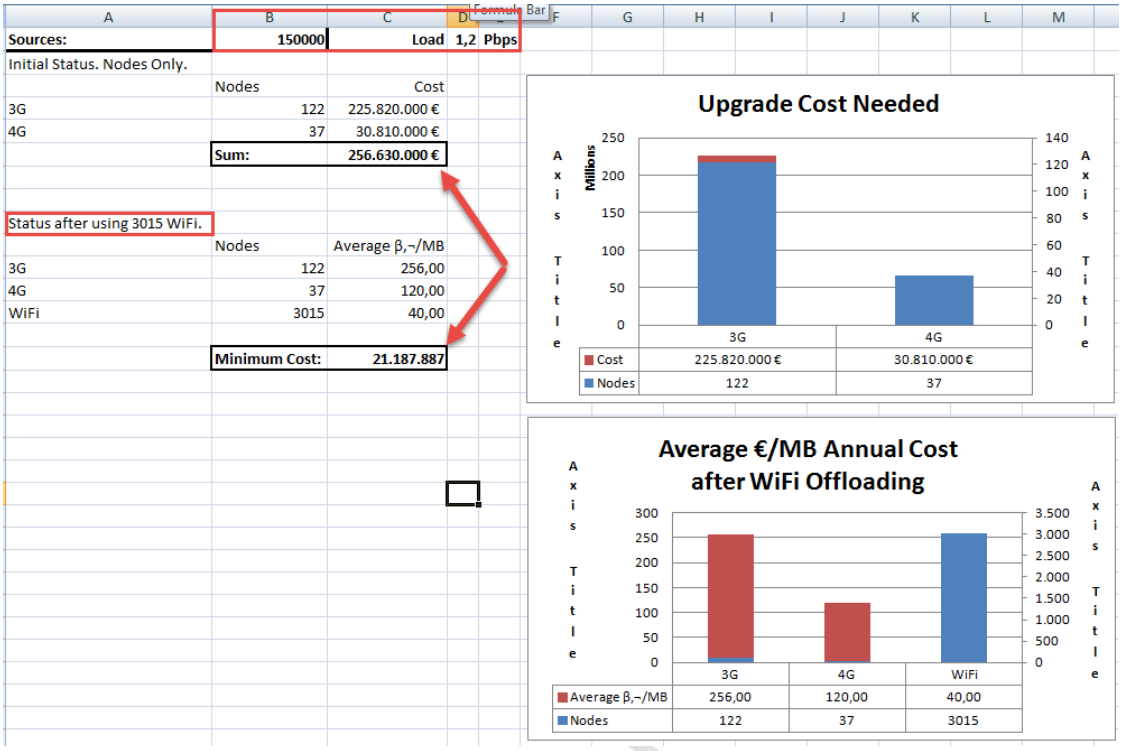

Figure 60 The export of the "Athens" scenario

The 1st conclusion that we can get from the export is the enormous amount of traffic load that the network is called to serve. The 2nd conclusion is that again we see a very big difference is the amount of many that the provider will have to pay in order his network to be able to handle this traffic load and we can see the cost in case the mobile provider doesn't adapt a HetNet solution and chooses to upgrade the existing cells is 256.630.000 euro and in case he chooses to step to a Wi-Fi offloading solution in order to handle traffic the cost is 21.187.887 euro. It quite obvious that the difference is over 10 times, with the Wi-Fi solution to be cheaper for the provider and of course better in terms of the user QoE.

Now that we had our 1st "run" for this 2nd scenario will try to see what will be the difference if the SA algorithm runs for 2000 and 4000 loops, as we have done with the 1st scenario at the previous chapter.

Sources:	150000	Load	1,2 Pbps
Initial Status. Nodes Only.			
	Nodes	Cost	
3G	122	226.730.000 €	
4G	37	30.680.000 €	
	Sum:	257.410.000 €	
Status after using 3015 WiFi.			
	Nodes	Average euro/MB	
3G	122	257,00	
4G	37	120,00	
WiFi	3015	40,00	
2.000 Loops	Minimum Cost:	21.177.833	

Sources:	150000	Load	1,2 Pbps
Initial Status. Nodes Only.			
	Nodes	Cost	
3G	122	226.730.000 €	
4G	37	30.810.000 €	
	Sum:	257.540.000 €	
Status after using 2988 WiFi.			
	Nodes	Average euro/MB	
3G	122	255,00	
4G	37	120,00	
WiFi	2988	40,00	
4.000 Loops	Minimum Cost:	20.800.456	

Figure 61 Comparing the values of the applications export after 2.000 & 4.000 loops

As we can see at figures 60 & 61, there is a difference between the first export that was a result of 1.000 loops and the above export that were running for 2.000 and 4.000 respectively. As far as the difference between the 1st and 2nd export is approximately 10.000 euro's which 0.0005% in terms of percentage is. Again we have the same conclusion with our 1st scenario, which is that the efficiency of the simulated annealing algorithm is so good, that although we have pushed the algorithm to run 2 times more, does not give us any significant change in terms of solution performance. Now the difference between the 1.000 loops and 4.000 loops is 387.431 euro's, which in terms of percentage is 0.018%., which is again is close to the limits of statistical error.

After "playing" with the algorithm so as to test each performance, we will now try to change the oversubscription factor in order see what difference we will have at the cost of the proposed solution of the algorithm, we know that the cost will be smaller but now the reference network is so big the result is more closer to reality. Below we can see an illustrative example of the newly defined values of the network regarding the oversubscription factor.

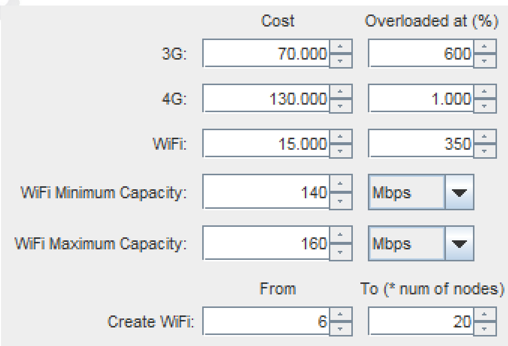

Figure 62 Increasing the oversubscription factor for all the network access nodes

After changing the oversubscription factor for all the access technologies (3G, 4G and Wi-Fi), now can run again the simulated annealing algorithm in order to see the new cost with the new oversubscription values. Below there we can see the figure of the new export.

Sources:	150000	Load	1,2	Pbps
Initial Status. Nodes Only.				
	Nodes	Cost		
3G	123	147.350.000 €		
4G	36	20.020.000 €		
	Sum:	167.370.000 €		
Status after using 2074 WiFi.				
	Nodes	Average euro/Mb		
3G	123	166,00		
4G	36	84,00		
WiFi	2074	29,00		
	Minimum Cost:	10.888.600		

Figure 63 The applications export with the new oversubscription factors

As we can see there is a significant cost drop for all the solutions, both for the upgrade of the existing network and the HetNet solution (Carrier Wi-Fi).

6.2.2 Second scenario conclusions

As we have said during this scenario, we will benchmark the performance of the simulated annealing algorithm due to the size of the selected network. The algorithm performed in the same as a way it had with the 1st scenario. Again regardless the increases of the loops that we have chosen either 1.000 or 4.000, the results were almost the same with the change being less than 1%.

Of course we have again to mention the data that were published to each export of the simulated annealing algorithm and we didn't mention them during the analysis of the results. For example during our scenario, at the selected network portion of Athens that we had selected through the map menu, we have pushed 1,2 Pbps of traffic to the selected network. In order to handle this future traffic load, the provider should implement or power on 3018 Wi-Fi hotspots at his carrier Wi-Fi network. Again with this kind of information we can see the power of the application, we do not just calculate the cost for statically given amount of access nodes, as a real forecast application we create the traffic load and with data load as a driver, we give a reference to the application user what will be the cost for each configuration which the provider wishes to choose. Below there is a figure, were the previously mentioned network attributes can be seen.

Sources:	**150000**	**Load**	**1,2**	**Pbps**
Initial Status. Nodes Only.				
	Nodes	Cost		
3G	123	222.390.000 €		
4G	36	29.510.000 €		
	Sum:	**251.900.000 €**		
Status after using 3018 WiFi.				
	Nodes	Average euro/MB		
3G	123	250,00		
4G	36	120,00		
WiFi	3018	41,00		
	Minimum Cost:	**21.199.510**		

Figure 64 The Wi-Fi access nodes that are need for the offload of the traffic from 3G and 4G cells

Additionally important information, which we have also mentioned at the 1st scenario has to do with the Cost/MB for each access node technology (3G, 4G and Wi-Fi). Although the cost of the nodes is calculated at yearly base and the traffic load is a snapshot of real time transmitted traffic. However, it continues to be a very helpful index of the cost per served MB, which can be used “as is” to **other tools** that the provider has for **financial forecasting**. This is another indication of the application's power and usability.

7 Conclusions

During the initial part of this master thesis we have seen the trend of the data traffic, which mobile providers are required to handle. We also saw several technological scenarios that provide a solution to this need, one of these is the HetNet solution of the carrier Wi-Fi, which can offload the biggest part of this future data storm.

Except the reference to the future traffic load and the candidate solutions that are capable to handle it, we have seen references highlighting the need of policy control for mobile users and the classification of them, as an additional measure in order to be able to handle this traffic load. Of course, the target is not only to handle this future data storm but to serve the users with the best possible QoE (Quality of Experience). And the fact that our application is capable of forecasting and planning with the user QoE and policy control taken in consideration gives an added value to our software.

As future extension of the software we think that it could endorse also the backhauling network and core network of the mobile provider, so as to give a more intergraded solution as far as the planning part of the software. Regarding the QoE aspect of the software, as a future extension we believe it would be better to adopt all the main QoE KPI that are used from the telecom industry such as Service availability, End to end delay, Delay variation, Packet loss ratio, RTT and etc, so as the algorithm to able to achieve more accurate results.

About the used algorithm of our software, there is always a margin for improvement, especially if all the above extension will take place. In such scenario the migration from a Simulated Annealing Algorithm to a Genetic Algorithm is compulsory, because Genetic Algorithms perform better to multilayer demands than Simulated Annealing algorithms can.

Closing this chapter will like to emphasize again to the added value of our software to the planning departments during the dimensioning of future network expansions. It gives the planners the ability to run "what if" scenarios for portions of their network massively and extremely quick and also it exports reports that can be used to high level meeting or it can be used as a reference during long and medium term network planning. This software can be considered an integrated planning & forecasting solution, since it takes into consideration both QoE and Policy Control aspects and financials values of the network and it does the job of a engineer in minutes instead of months, which this process could take for a human. Considering all the above, we can understand that this software could speed up the decisions for the future dimensioning of the network and this something that can be immediately translate to money, since 1-2 man working months is an extreme cost if you compare it to minutes or hours which the software wants in order to run a future scenario.

8 References

[1] Wireless LAN Design Guide for High Density Client Environments in Higher Education, Jim Florwick Jim Whiteaker, Alan Cuellar Amrod Jake Woodhams, Design Guide November, 2013

[2] Wi-Fi in heterogeneous networks, ericsson White paper June 2013

[3] Achieving carrier-grade Wi-Fi in the 3GPP world, Ericsson 2012

[4] Carrier Wi-Fi: the next generation, Ericsson 2013

[5] Cisco Visual Networking Index: Forecast and Methodology, 2013–2018, June 2014

[6] Quality Of Service Over LTE NETWORKS, LSI white paper

[7] A Practical Look at LTE Backhaul Capacity Requirements, Esteban Monturus Backhaul Market Analyst esteban@maravedis-bwa.com January 13th, 2011

[8] Over Subscription ratio as planning criterion, ATDI white paper 2008

[9] Quality of Service (QoS) and Policy Management in Mobile Data Networks, IXIA white paper

[10] Offloading Mobile Data Traffic to Wi-Fi Through MASS-ANET: Mobile Assisted Ad-hoc Networking, Delft University of Technology

[11] WLAN as an offloading solution for mobile cellular networks, AALBORG ANIVERSITY NOKIA SIEMENS NETWORKS

[12] Accelerating Wi-Fi Offload – Strategies for Solving Roaming Interworking Challenges, Dialogic White Paper

[13] Integration of Cellular and Wi-Fi Networks, 4G Americas

[14] LTE Design and Deployment Strategies, Cisco

[15] HetNet: Small Cells, Multi-mode & LTE-A, Radisys December 2012

[16] Wireless Broadband Alliance Industry Report 2013: Global Trends in Public Wi-Fi

[17] Architecture for Mobile Data Offload over Wi-Fi Access Networks, Cisco White Paper 2012

[18] Carrier Wi-Fi & Wi-Fi offload: New opportunities for service providers, Claus Hetting Senior consultant & analyst Maravedis-RethinkOctober 2013

[19] SEAMLESS WI-FI OFFLOAD: FROM VISION TO REALITY, APTILO NETWORKS WHITE PAPER

[20] SEAMLESS WI-FI OFFLOAD: A BUSINESS OPPORTUNITY TODAY, APTILO NETWORKS WHITE PAPER

[21] Holistic HetNet Hauling, Ceragon White Paper, February 2013

[22] HetNets by Ericsson: "Seamless User Experience Everywhere", 2012

[23] Smart Capacity Management in a HetNet World, February 2013

[24] HetNet: Where everything and everyone is connected, Ericsson 2011