Carrier Utilization Strategy for Traffic Load Balancing in Mobile Radio Networks

Osama W. Ata
Elect. Eng. Dept., Faculty of Engineering, Palestine Polytechnic University, Hebron, Palestine
Email: oata@ppu.edu; oata@fulbrightmail.org

***Abstract*—Ineffective deployment of carriers and sector load balancing are causing a significant poor carrier utilization in vendor markets. While identified performance carriers need to be retained, underutilized non-performance carriers need to be shut down and credited into the carrier bank. We propose a derived threshold for good carrier utilization and an empirical model that effectively describes a relationship between sector-carrier utilization and average sectors utilized in a market. Several millions of US dollars could be saved in capital and operational expenditure with implementation of proper guidelines and tools.**

***Index Terms*—carrier utilization; Erlang threshold; empirical model; load balancing; mobile networks; underutilized sectors**

I. Introduction

Carrier service providers have traditionally considered multi-carrier deployment in Base Transceiver Stations (BTS) as a means for enhancing network capacity, improving block and drop performance, in addition to investing in the network in the longer term. This is not surprising since global data traffic grew 63% by the end of 2016 and was predicted it would increase sevenfold between 2016 and 2021 [1], [2]. While this approach continues to be a realistic implementation of a network capacity resolution, carrier utilization optimization becomes an important technique for efficient use of deployed carriers, without degrading performance, which would consequently delay the need for carrier deployment [3]. A year's delay in carrier deployment could provide a net 15% saving in capital and operational expenditures.

Forecasting network capacity [4]-[6] is a critical mission that impacts the voice and data business of any carrier service provider. Such forecasting is based on statistical capacity models and historical traffic performance which bears a margin of uncertainty that depends on time span and unpredictable factors such as market share, customer churn, special events and possibly an overbuilt network. Karimi [7] presented solutions for effective utilization of wireless resources to keep up with the ever-increasing user demands for mobile content. Liu *et al*. [8] evaluated the cell wireless utilization rate of a studied network by introducing the concept of cell best wireless utilization rate. They concluded that the network did not meet the proposed configuration and accordingly recommended that, with effective control of blocking rate, a higher wireless utilization rate would result. Sharma *et al*. [9] reviewed various techniques and presented comparison tables of load balancing in cellular network. Benezit *et al*. [10] proposed a traffic load balancing scheme which allowed distributing the traffic of non-carrier aggregation users and another for balancing the traffic of carrier aggregation users. The models achieved quasi–ideal load balancing over the carriers.

In our work, we propose that monitoring the utilization of the last carrier, at the sector level, would enable the carriers truly needed for traffic to be verified and the carriers that would be redeployed to be identified. The metric is a pure capacity one as it does not consider any performance issues. At the sector level, the metric verifies how balanced sectors are and how efficiently the carriers are used by all sectors. A sector is defined as under-utilized when the sector's traffic is less than the Erlang threshold of one less carrier, as demonstrated in the next carrier. We demonstrate that if on a three-carrier (F3) site, the traffic sector is less than the two-carrier (F2) Erlang threshold, then this indicates that the sector would probably be fine without the third carrier. The definition, however, overlooks performance consideration.

Correlation between average sectors utilized /site and sector carrier utilization could be strong in some vendor markets, but weak in other markets. Reasons would be analyzed.

We report monitored observations [11] in real vendor markets in the United States and proposed guidelines and tools for improving carrier utilization overlay. Four vendor market regions are accordingly analyzed. Two markets representing US national extremes; San Francisco and Philadelphia, are compared in carrier and average sectors utilized. The data is not current and may have significantly changed, but at least represents real data that is worth analyzing, in pursuit of an action plan on options, implementation and tools required.

II. Carrier Utilization Strategy

Our concern, in this study, is multi-carrier sites, hence it makes sense to exclude single carrier sites, towards achieving our objective. The Erlang threshold is a voice capacity metric that is traditionally used by network

Manuscript received November 11, 2018; revised December 20, 2018; accepted January 7, 2019.
Corresponding author: Osama W. Ata (email: oata@ppu.edu).

doi: 10.18178/ijeetc.8.3.152-157

service providers. Since voice communication predominantly remains the "bread and butter" of the business; at least until the data applications and usage dramatically pick up, data communication models would not be addressed here. The maximum number of radio channels per cell is closely related to an average calling time *T* (in minutes) in the system. If the maximum calls per hour per cell is Q, then the traffic figure can be described as:

$$A_e = \frac{QT}{60} \quad \text{Erlang} \tag{1}$$

For example, if a group of users made 30 calls in one hour, and each call had an average call duration of 5 minutes, then the number of Erlangs this represents is worked out as follows:

Minutes of traffic per hour = number of calls × duration

Minutes of traffic per hour= 30 ×5=150

Hours of traffic per hour=150/60=2.5 Erlangs

With a standard 2% of blocking probability, the required number of calls, during busiest hour per cell can be determined from one of the most commonly used traffic model tables; the Erlang-B table, described in [12]. To determine the Erlang threshold for each sector, monthly bouncing busy hour primary Erlangs (the traffic Erlangs) are calculated using the "Six-Peak" method; known by the service provider. Sector Erlang threshold for each carrier is calculated using current handset penetration. Traffic Erlangs are compared to Erlang threshold of one less carrier in the cell sector.

The sector can hence be determined whether is utilized or under-utilized. The number of utilized sectors, A, could then be counted, as well as, the total number of multi-carrier sectors, B. Market utilization would then be expressed as A/B.

III. Engineering Considerations

The engineering significance of the sector-carrier utilization of a mobile network should derive from a strategy characterized by an objective for sustaining improvement of asset utilization. Its principal purpose would hence be focused on improving the utilization of current and future carrier overlay. While traffic balancing remains an important metric for improving network traffic distribution, mobile operators traditionally deploy carriers for meeting capacity demands without following engineering considerations and guidelines to determine when carriers should be actually deployed and when to shut them down and credit them to the carrier bank.

Shutting down carriers demands an engineering process that should identify metric parameters and thresholds below which a number of carriers could be turned down; in order to sustain improvement of asset utilization, while save on operational running costs.

The general process would involve a number of steps:

- Use a Cap Plan approach to estimate the number of under-utilized sector-carriers by the end of the intended year.
- Use:
  - The latest subscriber forecast
  - The latest handset penetration forecast
  - The latest call model forecast
- Forecast traffic for the end of the intended year.
- Calculate utilization by then.
- Estimate carriers that can be turned down.
- Recalculate utilization after most turn-downs are executed.

The two parameters, necessary to carry out the process, that we developed and consequently calculated, are namely, the 'average sectors utilized' and the 'sector-carrier utilization percentage'. The metric definition and assumptions, as well as their utilization in building customized models will be fully explained, in the following sections.

The developed models would serve the following purposes:

- The modeled total average value of the 'average sectors utilized' parameter within the vendor market region would set a threshold value, below which carriers could be candidates for removal, their parameters recalculated and consequently shut down while the carrier utilization recalculated.
- The modeled curve would serve as an engineered baseline for comparing future trends and taking consequent appropriate actions, for sustaining optimized carrier utilization.
- The developed models could serve other future network carrier deployment since the three vendor markets, in this study are highly correlated with a minimum average sectors utilized of 1.6 and sector-carrier utilization of 30%.

IV. Metric Definitions and Assumptions

A site is carrier utilized when at least one of its sector's *N*th carriers is triggered by the traffic Erlangs exceeding the *N* minus 1 carrier threshold. If not, the site is labeled as under-utilized. The analysis only takes multi-carrier sites into consideration and excludes the single carrier (F1) sites.

By reasoning, all F1 sites are utilized. The analysis assumes that the predominant sites in the markets are three-sector based. Furthermore, the derived statistical values for achieving higher carrier utilization are not meant to set fixed objectives that all markets should adopt.

Consider Fig. 1 which displays a possible scenario, representing Sector Traffic Erlangs.

Suppose that each of the sites (sectors) have 4 carriers with the following Erlang thresholds:

- 1 carrier: 20 Erlangs
- 2 carriers: 40 Erlangs
- 3 carriers: 70 Erlangs
- 4 carriers: 95 Erlangs

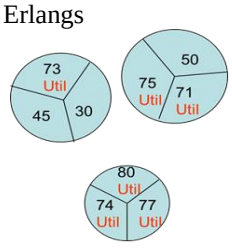

Fig. 1. Site traffic Erlangs scenario

There is a total of 6 utilized sectors, each of which exceeds the three (4 minus 1) carrier capacity threshold. There are 3 under-utilized sectors because they fall below the three carrier capacity threshold of 70 Erlangs.

Hence the average sectors utilized would be the total number of utilized sectors/total number of sites = 6/3 = 2 sectors per site on average. This metric would hence monitor the utilization of the last installed carrier at a sector level. It is a pure capacity metric as it does not consider any performance issues. The sector-carrier utilization, on the other hand, works out as 6/9 ×100%= 66.7%. Therefore, only the last carrier is considered in order (a) to verify that carriers are needed for traffic (b) to identify carriers that could be redeployed. In addition, the sector level is considered in order (c) to verify that sectors are balanced as much as possible (d) to verify that carriers are efficiently used by all sectors.

V. Methodology for Calculating Market Utilization

In calculating market utilization, only multi-carrier sites are considered, whereas single carrier sites are excluded.

- For each sector:
  - Collect monthly bouncing busy hour primary Erlangs for each sector ("six-peak" method).
  - Calculate sector Erlang threshold for each carrier using current handset penetration.
  - Compare traffic Erlangs to Erlang threshold of one less carrier.
- Determine if sector is utilized or under-utilized.
- Count the number of utilized sectors (A).
- Count the total number of multi-carrier sectors (B).
- Calculate market utilization (A/B).
- Calculate sector-carrier utilization (A/3B ×100%). This assumes 3 sectors per site.

There are mainly two parameters that affect utilization:

- 3G/4G handset penetration:
  - For the same traffic load, when 3G/4G handset penetration increases, Erlang thresholds would increase too, showing more under-utilized sectors.
- Traffic growth:
  - For the same Erlang thresholds, when traffic load increases sectors become more utilized.

Consequently, there are two methods to improve this utilization; both of which would lead to a better utilization of the carrier:

- Remove carriers on sites, where all sectors are under-utilized.
- Load balance sectors within each site so that more sectors become utilized.

A site which is typically 3-sectored is considered utilized if at least one sector is utilized. This is so because the other two sectors would not be deprived of similar carriers that would become necessary for HHO before the time comes when they also become utilized sectors. Accordingly, we consider:

- Average sector utilization in a market = sum of utilized sectors divided by sum of utilized sites.
- Sector-carrier utilization = sum of utilized sectors divided by the sum of all sectors.

VI. Modeling

From previous discussion, the relationship between average sector utilization and sector-carrier (site) utilization is obviously nonlinear. We, accordingly, propose a general formula:

$$y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \qquad (2)$$

where x is the average sectors utilized in a market, y is the sector-carrier utilization, n is the order of the equation, and a_n, a_{n-1},⋯, a_0 are constants.

Fig. 2 to Fig. 5 and Table I describe actual data, obtained from several vendor market regions in the USA.

Fig. 2 shows average sector utilization on the primary y-axis and the sector carrier utilization on the secondary y-axis for 27 city markets of vendor market region VMR-1. The figure shows a moderate statistical correlation of 75.6% between the two metrics, sorted in an ascending order. Depending on the network design and performance for different regions, market penetration, traffic conditions and customer growth and churn rate, the correlations vary from a market to market and hence from a region to another.

Fig. 3 to Fig. 5 apply for vendor market regions VMR-2 (12 markets), VMR-3 (28 markets) and VMR-4 (1 market), with 36%, 83% and 100%, respective correlations

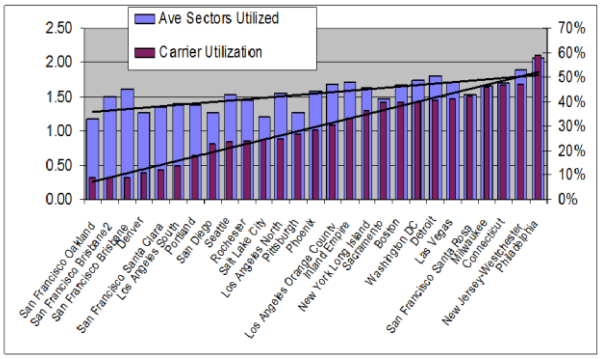

Fig. 2. Average sectors utilized and sector-carrier utilization% for vendor market region VMR-1.

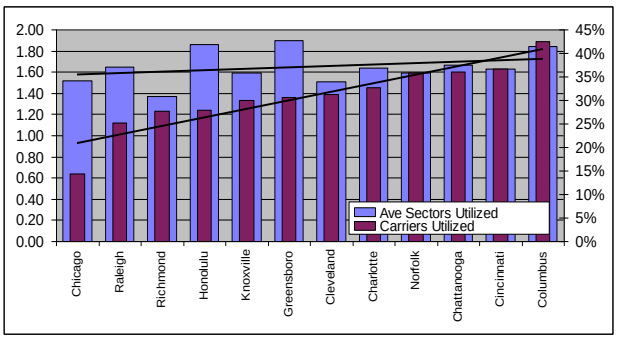

Fig. 3. Average sectors utilized and sector-carrier utilization% for vendor market region VMR-2.

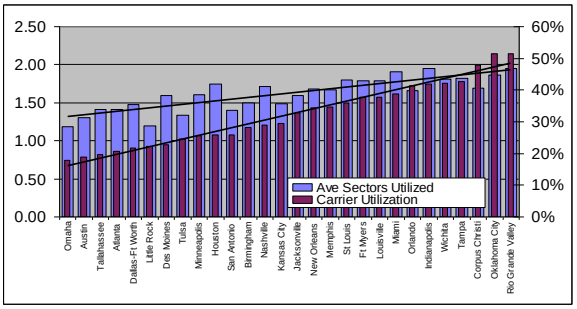

Fig. 4. Average sectors utilized and sector-carrier utilization% for vendor market region VMR-3

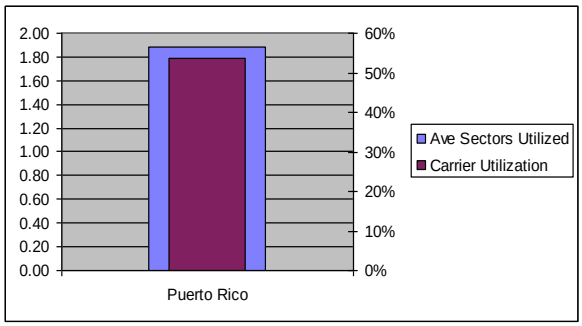

Fig. 5. Average sectors utilized and sector-carrier utilization% for vendor market region VMR-4

TABLE I: CARRIER UTILIZED AND SECTOR LOAD BALANCED VENDOR MARKET REGIONS

Vendor's Market Region (VMR)	No. markets that are carrier utilized & sector load balanced	Total No. of Markets	% utilized & sector load balanced
VMR-1E	10	15	66.7 %
VMR-1W	1	12	8.3 %
VMR-2	4	12	33.3 %
VMR-3	13	28	46.4 %
VMR-4	1	1	100 %

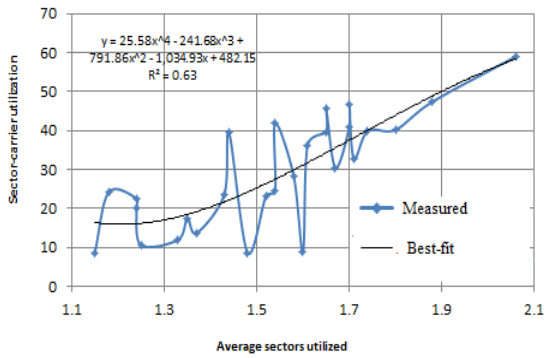

Fig. 6. Measured and modeled sector-carrier utilization% versus average sectors utilized for vendor market region VMR-1.

Table I shows the percentage of markets that are carrier utilized and sector load balanced in each vender market region. In particular, vendor market region VMR-1 with its two sub-regions; 1E and 1W show extreme respective percentages of 66.1% and 8.3%.

Polynomial curve fitting was performed on measured sector-carrier utilization versus average sectors utilized for the three regional markets; VMR-1, VMR-2 and VMR-3.

Fig. 6 is a sample measurement that describes VMR-1 with a fourth degree power series and a moderate correlation as determined by the coefficient of determination [13], [14]; R-squared, which amounted to 0.63. VMR-1, is composed of two sub-regions; VMR-1E with 15 markets and VMR-1W with 12 markets. Accordingly, R-squared measured a strong 0.79 value for VMR-1E and a weak 0.4 value for VMR-1W.

Fig. 7, on the other hand, compared sector-carrier utilization variation of another vendor, VMR-3, of a comparable market size to VMR-1. R-squared measured a strong 0.72 value, with the 4th degree polynomial.

The equations in Fig. 6 and Fig. 7 depict trends in VMR-1 and VMR-3, respectively. A general equation could be derived that would describe all vendor market regions.

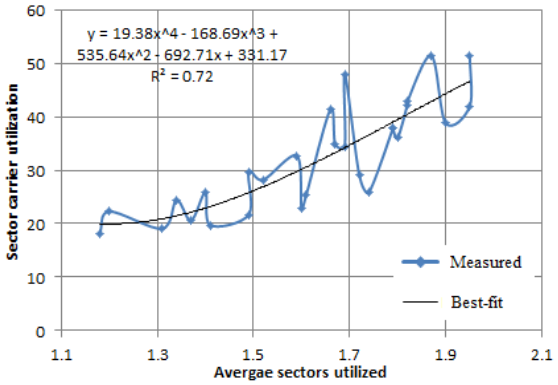

Fig. 7. Measured and modeled sector-carrier utilization% versus average sectors utilized for vendor market region VMR-3.

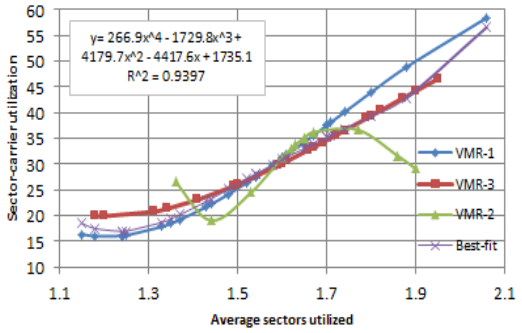

Fig. 8. Modeled sector-carrier utilization% versus average sectors utilized and their total best-fit curve for all vendor market regions VMR-1, VMR-2 and VMR-3

VII. RESULTS AND DISCUSSION

Obtained carrier utilization formulas that described fitted curves for vendor region markets, VMR-1, VMR-2 and VMR-3 were combined and compared, as shown in Fig. 8. A general average model in (3) was accordingly derived with a very strong R-squared value of 0.94.

$$y = 266.9x^4 - 1729.8x^3 + 4179.7x^2 - 4417.6x + 1735.1 \quad (3)$$

VMR-1 and VMR-3 showed close comparative variation in sector-carrier utilization versus average sectors utilized, with moderate to strong *R*-squared values. Sub-region VMR-1E showed a strong R-squared, while VMR-1W showed a weak value. The low score in VMR-1W may have to do with the fact that several sites were

either performance sites or underutilized non-performance sites. Another reason is that competing service providers in the sub-region may have lowered the minutes of use (MOU) per subscriber per month.

VMR-2 is almost half the size of VMR-1 and VMR-3, with 12 markets with an average fitted curve of R^2= 0.56 Although, average sectors utilized per site, as a result of better load balancing and tearing unnecessary carriers, are key for achieving higher utilization, VMR-2 can be considered as an exception. Consequently, pilot beacon HHO technology with pushed away carrier boarders necessitated a large number of performance (hand-down) carriers that impacted relatively low carrier utilization.

The two curves in Fig. 9 that pertain to VMR-1 and VMR-3, with the exclusion of VMR-2, show a relatively slightly higher R^2 value of 0.98, than the three curves in Fig. 8. The general average model can be described as:

$$y = 199x^4 - 1304x^3 + 3193.4x^2 - 3416.1x + 1358.7 \quad (4)$$

It is interesting to know that all curves in Fig. 8 meet at approximately x=1.6 and y=30%. This point could be thought of as a threshold point, above which none-performance sites are considered utilized. This would meet expectations since average sectors utilized is in the 1-2 range and is practically in the 1.15-2.06 range. The average sectors utilized is hence theoretically 1.5 or practically 1.6. If, on the other hand, the average market performance carrier % worked out, say 30%, then the net site-carrier utilization would actually be 70%. Consequently, the expected sector-carrier utilization =70%×1.5/3=35%. Actual range of sector carrier utilization measured 8.75%~59.04% with an average of 34% and an overall average of the three vendor markets amounting to 30%.

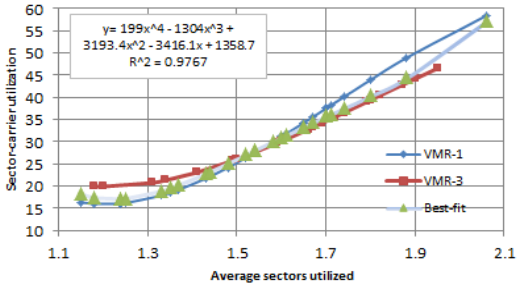

Fig. 9. Modeled sector-carrier utilization% versus average sector utilized and their total best-fir curve for vendor market regions VMR-1 and VMR-3.

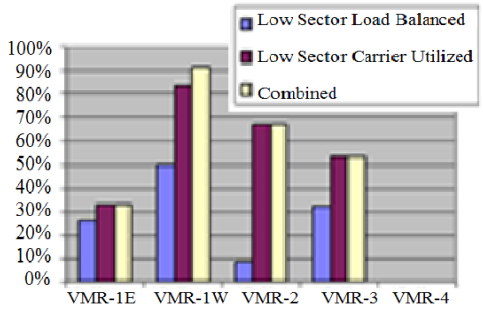

Fig. 10. Comparison of low-sector-carrier utilized vs. low-sector-load balanced vendor markets.

Fig. 10 shows low-sector-carrier utilized versus low-sector-balanced vendor markets. Here, low-sector-carrier utilized are below 30% and low-sector-load balanced are below 1.6 sectors/site. It is interesting to note that sub regions 1E and 1W in VMR-1 show extreme combined percentages between approximately 30% and 90%, respectively.

VIII. Observations

It is difficult to dictate specific market guidelines and accordingly set objectives. However, in general, one may say that a market should attain a minimum of 30% for sector-carrier utilization. Moreover, an average sector utilization of 1.6 should be the minimum acceptable for achieving minimum sector carrier utilization.

Accordingly, four scenarios were observed. Some markets were sector carrier utilized but low sector load-balanced (Sacramento), hence sectors would require traffic load optimization. Other markets were sector load-balanced but low sector carrier utilized (Minneapolis and Honolulu). Hence, unnecessary carriers, pending performance and traffic load optimization should be shut down. Some other markets were under load-balanced and under-utilized (Omaha, Richmond and San Francisco-Oakland). Hence sectors would require traffic load optimization and unnecessary carriers need to be shut down. Finally, some markets were sector load-balanced and sector carrier utilized (Philadelphia and Puerto Rico). Hence, continued monitoring of sites for possibly degraded sector load balancing or unnecessary carriers would be required.

IX. Conclusions

Average sectors utilized per site as a result of better load balancing and tearing unnecessary carriers are key for achieving higher utilization. Markets that were considered sector carrier utilized, scored above 30%. Additionally, markets that were considered sector load-balanced, scored above 1.6 average sectors per site.

A model in the form of a fourth degree polynomial, that described the relationship between sector carrier utilization versus average sectors utilized was developed for all studied vendor market regions, which would enable and guide operators to identify performance carriers that needed be retained, as well as, underutilized non-performance carriers that needed be shut down and credited into the carrier bank.

VMR-1E had the highest score of average sectors utilized and carrier utilization % while VMR-1W had the lowest score. Consequently, Philadelphia scored highest market carrier utilization 59.04% and 2.06 average sectors per site. On the other hand, San Francisco-Oakland scored lowest score of 8.75% and 1.18 sectors per site. The 1.6 average sectors threshold might be raised once markets start using appropriate optimization tools for sector balancing.

VMR-2 showed an exception of 14.36% and 1.52 average sectors per site. This is so because pilot beacon hard handoff (HHO) technology with pushed away carrier boarders, necessitated a large number of performance

(hand-down) carriers that impacted relatively low carrier utilization.

Finally, the literature, in network carrier utilization, is not broad. More extensive work is required to determine other metric parameters that could affect carrier utilization for better refinement. Meanwhile the derived models serve as a very good baseline for other mobile operators to either predict or compare their network carrier utilization performance.

X. Recommendations

- Implement sector load balancing
  - Achieve a minimum 1.6 sectors utilized per site on average per market to affect increasing carrier utilization.
  - Improve on 1.6 sectors for relatively load – balanced sectors to continue improving carrier utilization.
- How? What are the options?
  Need an effective tool that would allow the user of the market to:
  - Pull market data (actual utilization, objective, number of sectors, sites, etc.).
  - Look at current sector balancing and site level carrier utilization.
  - Modify the sector balancing forecast and/or the site-carrier utilization forecast.
- Tear down under-utilized non-performance carriers
  - Any carrier that is under-utilized and is not required as a performance carrier should be torn down and stored in the carrier bank.
- Identify skill set gaps in optimization engineers and develop curriculum courses and on the job training to minimize and eliminate those gaps

References

[1] *Global Mobile Data Traffic Forecast Update, 2016–2021*. Cisco, San Jose, CA, USA. (Feb. 2017). [Online]. Available: https://www.cisco.com/c/en/us/ solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] *Ericsson Mobility Report*. Ericsson, Kista, Stockholm, Sweden. (Nov. 2016). [Online]. Available: https://www.ericsson.com/assets/ local/mobility-report/documents/2016/ericsson-mobility-report-november-2016.pdf

[3] T. M. de Almeida Pedro, "Capacity management and load balancing for 3G wireless access networks," MSc thesis, Tecnico Lisboa, Lisboa, Portugal, 2016.

[4] K. Clattenburg, "A view on subscriber forcast and network capacity analysis," Former Manager Network Forecasting & Analysis, Sprint, Overland Park, USA, Private Communication.

[5] *Forcast:Number of 3G mobile subscribers in North America from 2008 to 2020 (in 1,000)*. Satista-The Statistics Portal, Hamburg, Germany. (Sep. 2011). [Online]. Available: https://www.statista.com/statistics/226369/3g-mobile-subscriber-forecast-in-northern-america/

[6] R. Ibrahim, H. Ye, P. L'Ecuyer, and H. Shen, "Modeling and forecasting call center arrivals: A literature survey and a case study," *Int. Journal of Forecasting*, vol. 32, no. 3, pp. 865-874, Sep. 2016.

[7] O. B. Karimi, "Efficient resource utilization in advanced wireless networks," Ph.D. dissertation, Faculty Appl. Scien., Simon Fraser University, Vancouver, Canada, 2013.

[8] Z. Liu, Z. Jia, X. Qin, C. Chang, G. Xu, and X. Xia, "Best traffic carrier frequency number and cell wireless utilization research," *Procedia Engineering*, vol. 24, pp. 303-307, Jan. 2011.

[9] A. Sharma, A. Roy, S. Ghosal, R. Chaki, and U. Bhattacharya, "Load balancing in cellular network: A review," in *Proc. IEEE Third Int. Conf. on Computing Communication & Networking Technologies*, 2012, pp. 1-5.

[10] F. Bénézit, S. Elayoubi, R. M. Indre, and A. Simonian, "Modelling load balancing and carrier aggregation in mobile networks," in *Proc. 12th IEEE Int. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, 2014, pp. 78-85.

[11] O. W. Ata, "Carrier utilization strategy," presented at Technical University of Cluj-Napoka meeting, Cluj-Napoka, Romania, Jun. 2017.

[12] W. C. Y. Lee, *Mobile Cellular Telecommunications*, 2nd ed., New York, USA: McGraw-Hill, 1995.

[13] W. I. Notz, M. A. Flinger, and D. S. Moore, *The Basic Practice of Statsitics*, 6th ed., New York: W.H. Freeman and Company, 2013, p. 138.

[14] W. G. Zikmund, *Business Research Methods*, Fort Worth, USA: Harcourt College Publishers, 2000, p. 513.

Osama W. Ata received his B.Sc. (Hon) degree in Electronic Engineering from Hull University in 1980, an MSc in Electronic Communications and a Ph.D. in Microwave Antenna Systems from Cardiff University in UK, in 1981 and 1989, respectively. He was granted a Postdoctoral Fellowship at Bangor University, in UK, for three years, a Fulbright Senior Fellowship at Stanford University in USA, and Senior Visiting Fellowships at Villanova University in USA and Victoria University of Technology in Australia, respectively. Dr. Ata is an Associate Professor at Palestine Polytechnic University in Hebron, working in the electrical engineering department and a key member in the telecommunications program. He teaches courses in Antennas and Propagation, Microwave Communications, Electromagnetics, Transmission Line Theory, Optical Communications, Radar Systems, Applied Engineering Mathematics and Research Methodology. His research interests are in Microwave Horn Antennas, Microstrip Antennas, Radio Frequency Propagation, In-building Penetration Loss Modeling and Measurement, Electromagnetic Radiation, RF Cellular Planning and Mobile Radio Networks. He recently won the University prestigious award for Abundance in Quality Scientific Research production.