SELF-SIMILARITY AND MODELING OF LTE / LTE-A DATA TRAFFIC

by

ROOPESH KUMAR POLAGANGA

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2015

DEDICATION

To Mom & Dad.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr.Qilian Liang, for supervising me on this thesis and helping me in every step of my work. His broad knowledge and expertise in the field of wireless communication will be very useful to me all through my career.

I wish to thank Dr.Saibun Tjuatja and Dr.Yuze Alice Sun for taking time to serve on my thesis committee.

I would like to thank Dr. Alan Davis and Dr. William E. Dillon, my graduate advisors for their advice and guidance throughout my M.S.

I would also like to thank Jasen Virasawmy, Sr.Manager at T-Mobile USA, for guiding me during my Internship. I thank entire Tier-II RAN Nokia team at T-Mobile USA for supporting me.

March 3, 2015

ABSTRACT

SELF-SIMILARITY AND MODELING OF LTE / LTE-A DATA TRAFFIC

Roopesh Kumar Polaganga, M.S.

The University of Texas at Arlington, 2015

Supervising Professor: Qilian Liang

The mobile wireless communication system has evolved into its 4th generation with ubiquitous availability assuring higher data rates and reliable communication. To achieve a robust and reliable network, it is necessary to understand the traffic characteristics of the network. In this work, network traffic characteristics like Self-Similarity property and modeling of user arrivals are estimated for real-world LTE & LTE-Advanced (LTE-A) traffic.

Lot of work has been done on Self-Similarity of Ethernet and Ad-Hoc Network's data traffic. But Self-Similarity in LTE and LTE-A data traffic has been left un-explored. Real world LTE and LTE-A data traffic collected from a live network has been studied and Self-Similarity property is also evaluated. This work emphasizes the existence of Self-Similarity in LTE and LTE-A data traffic which can facilitate the future research on forecasting data traffic. A degree of Self-Similarity is observed in live LTE and LTE-A networks and a comparison is made between LTE, LTE-A and previously researched Ethernet data traffics.

Modeling and understanding the user traffic is of utmost importance to optimize the network. It is important to model and evaluate the incoming user traffic in our current generation of mobile communication systems which dominated data traffic more than the traditional voice traffic. This work will address the validity of considering Poisson Arrival for modeling user arrival traffic in live LTE network.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Introduction

Wireless Communication has been one of the fast growing industries since mid-1990s. It has evolved from a cutting edge technology owned by military, to an essential part of the cellular phone system used by billions of customers. Over the last few decades, due to the increasing demand for higher speed data and widespread network access in mobile communication, there has been a tremendous ongoing research in the field of cellular communication which has resulted in achieving significant developments.

Long Term Evolution-Advanced (LTE-A) and Long Term Evolution (LTE) are the current trending technologies aimed at providing substantial performance enhancements at reduced cost. The 3rd Generation Partnership Project's (3GPP) Long Term Evolution (LTE) represents a major advancement in cellular technology and marks the evolutionary move from third generation of mobile communication (UMTS) to fourth generation mobile technology. LTE is the first cellular communication system supporting packet optimization radio access technology with high data rates and low latencies. Thus, data constitutes to be the pre-dominant traffic when compared to other traditional modes of traffic like the voice.

The statistical characteristics of network traffic have been of interest to researchers for many years, not least to obtain a better understanding of the factors that affects the performance and scalability of large systems such as the internet [5]. This

Tele-traffic knowledge of statistics including queuing theory, the nature of traffic, their practical models, their measurements and simulations to make predictions and to plan telecommunication networks helps in providing reliable services at lower cost.

Self-Similarity is an important concept that, in a way, it is surprising that only recently has it been applied to data communications traffic analysis. This concept helps in forecasting the data traffic which in turn helps in optimizing the network. Lot of work has been done on Self-Similarity of Ethernet and Ad-Hoc network's data traffic. But no considerable work has been done on LTE and LTE-A data traffics as they are the latest evolution technologies in mobile communication system. This work mainly focuses on the Self-Similarity property of LTE and LTE-A network's live data traffic.

Good network traffic modeling is necessary to better understand and utilize the network resources. In Tele-traffic theory, Often the arrival process of customers can be modeled by a Poisson Process [6]. In this present generation of telecommunication system with multiple types of traffic, it is necessary to evaluate the holding of this basic assumption for better understanding of the network. This work highlights the importance of user traffic evaluation in LTE network, which has high data traffic when compared to traditional voice traffics.

The most demanded feature of LTE-Advanced is Carrier Aggregation. This feature allows the aggregation of multiple carriers and increases the bandwidth which in turn increases the data rates. The users are served with extra resources as Secondary Cells. This feature is very popular in theoretical developments, but very novel in real world implementation. It is very important to analyze the real world performance metrics

of this new LTE-A feature which helps in knowing the feature's implementation issues that need some more research to make it more effectual.

1.2 Thesis Outline

This thesis is organized into four chapters. Chapter 1 introduces the current state of wireless system with the importance of this piece of work. Chapter 2 explains the Self-Similarity property observed in the data traffic and the evidence to show the existence of Self-Similarity in LTE data traffic. Chapter 3 talks about the modeling of LTE user traffic and its evaluation in live LTE network. It also introduces the concept of Poisson Arrival Process. Chapter 4 introduces the concept of Carrier Aggregation in LTE-Advanced and Self-Similarity is analyzed in LTE-Advanced data traffic. Trending of LTE-A Carrier Aggregation metrics are analyzed and degree of Self-Similarity is compared to that of LTE. Chapter 5, 'Conclusion and Future Wok' summarizes about the results achieved in the previous chapters and discuss about what possible work can be done to extend the work presented in this thesis.

Chapter 2

SELF-SIMILARITY

2.1 Introduction

Understanding the nature of network traffic is important to properly design and implement communication networks and network services. Recent examinations of LAN traffic, Wide area network traffic and Ad-Hoc networks have challenged the commonly assumed models for network traffic, e.g., the Poisson process [21]. Where the traffic to follow a Poisson or Markovian Arrival process, it would have a characteristic burst length which would tend to be smoothed by averaging over a long enough time scale. Rather, measurements of real traffic indicate that significant traffic variance (burstiness) is present on a wide range of time scales.

Traffic that is bursty on many or all time scales can be described statistically using the notion of Self-Similarity. Self-Similarity is the property we associate with one type of fractal – an object whose appearance is unchanged regardless of the scale at which it is viewed. In that case of stochastic objects like time series, Self-Similarity is used in the distributional sense: when viewed at varying scales, the object's correlational structure remains unchanged. As a result, such a time series exhibits bursts – extended periods above the mean- at a wide range of time scales.

Since a self-similar process has observable bursts at a wide range of timescales, it can exhibit long-range dependence; values at any instant are typically non-negligibly positively correlated with values at all future instants. Surprisingly (given the counterintuitive aspects of long-range dependence) the Self-Similarity of Ethernet traffic

has been established as per [11] and Ad-Hoc networks in [1]. The importance of long-range dependence in network traffic is beginning to be observed, which show that packet loss and delay behavior is radically different when simulations use either real traffic data or synthetic data that incorporates long-range dependences.

LTE is the recent development in telecommunication systems whose traffic nature is of utmost interest to researches in this field of study. As LTE technology is quickly growing in terms of coverage and number of users, there is a need to study its network traffic. Since data traffic is predominant in LTE, this work concentrates on the evidence to show that LTE data traffic exhibits consistent Self-Similarity property. Intensity of this Self-Similarity property is evaluated in LTE networks based on the defined parameters.

A phenomenon that is self-similar looks the same or behaves the same when viewed at different degrees of 'magnification' or different scales on a dimension. The dimension can be space (length, width) or time. Self-Similarity is where a certain property of an object is preserved with respect to scaling in time and/or space. "If an object is self-similar…, its parts, when magnified, resemble – in a suitable sense – the shape of the whole." The following diagram provides a pictorial view of this phenomenon.

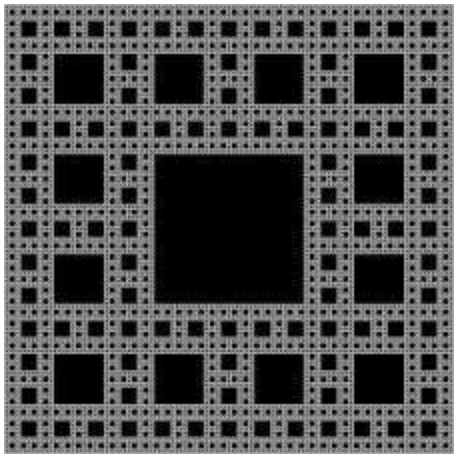

Figure 2-1 Pictorial view of Self-Similarity

Will Leland and *Daniel Wilson* put together a comprehensive study of the self-similar nature of Ethernet traffic. They were able to formulate their results via data collected from the network environment at the Bellcore Morris Research and Engineering Center.

A common continuous time definition of self-similar stochastic processes is based on a direct scaling of the continuous time variable, as follows. A stochastic process $X(t)$ is statistically self-similar with parameter H if for any real a>0, the process $a^{-H}X(at)$ has the same statistical properties as $X(t)$. This relation can be expresses by the following three conditions.

1. Mean $$E[X(t)] = \frac{E[X(at)]}{a^H} \tag{2.1}$$

2. Variance $$Var[X(t)] = \frac{Var[X(at)]}{a^{2H}} \tag{2.2}$$

3. Autocorrelation $$R_X(t,s) = \frac{R_X(at,as)}{a^{2H}} \tag{2.3}$$

2.2 Self-Similarity of LTE Network traffic

Given a zero mean, stationary time series $X = (X_t; t = 1, 2, 3, ...)$, we define the *m*-aggregated series $X^{(m)} = (X_k^{(m)}; k = 1, 2, 3, ..)$ by summing the original series X over non-overlapping blocks of size *m*. Then it's said that X is *H*-self-similar, if, for all positive *m*, $X^{(m)}$ has the same distribution as X rescaled by m^H. That is

$$X_t \square m^{-H} \sum_{i=(t-1)m+1}^{tm} X_i \forall m \in N \tag{2.4}$$

If X is *H*-self-similar, it has the same autocorrelation function $r(k) = E[(X_t - \mu)(X_{t+k} - \mu)] / \sigma^2$ as the series $X^{(m)}$ for all *m*. which means that the series is the same as that if the original.

Self-Similar processes can show long-range dependence. A process with long-range dependence has an autocorrelation function $r(k) \square k^{-\beta}$ as $k \to \infty$, where 0< β <1. The degree of Self-Similarity can be expresses using *Hurst* parameter $H = 1 - \beta / 2$. For a self-similar series with long-range dependence, H value varies from 0.5 to 1 (1/2<H<1). As $H \to 1$, the degree of both Self-Similarity and long-range dependence increases.

2.3 Existence Reason of Self-Similarity

A plausible physical explanation for the occurrence of Self-Similarity in high speed network traffic is explained in [12] is based on convergence results for processes that exhibit high variability.

Early studies of Internet traffic proved particularly interesting as they exposed self-similar characteristics that were not previously commonplace. There are many papers

debating the reason for the apparent Self-Similarity in modern communication traffic. These range from ON/OFF models of heavily tailed distribution, file size distribution in file systems and web servers, user behavior, network protocols, back-off algorithms in the Ethernet, buffer in routers and the TCP congestion avoidance algorithms.

Superposition of many ON/OFF sources (also known as packet trains) whose ON-periods and OFF-periods exhibits the Noah Effect (i.e., have high variability or infinite variance) produces aggregate network traffic that features the Joseph Effect (i.e., is self-similar or long-range dependent). There is moreover, a simple relation between the parameters describing the intensities of the Noah Effect (high variability) and the Joseph Effect (Self-Similarity). An extensive statistical analysis confirms that the data at the level of individual sources or source-destination pairs are consistent with the Noah Effect. The implication of this simple physical explanation for the presence of self-similar traffic patterns in modern high-speed network traffic for (i) parsimonious traffic modeling, (ii) efficient synthetic generation of realistic traffic patterns, and (iii) relevant network performance and protocol analysis.

2.4 The Effect of Self-Similarity

The effect of Self-Similarity on networks boils down to packet loss. When traffic increases to the point where neither the bandwidth nor the router buffer sizes are suffice to handle the burst, packets are lost. This is where the financial loss comes into play. When packets are lost, in certain situations they must be resent. Resending the packets wastes bandwidth and could further congest the network. In situations where the packets

don't have to be resent, the quality of service is degraded. These two aspects can lead to millions of dollars of loss.

Two possible solutions include the dynamic control of traffic flow and structural resource allocation. Predictive feedback control is a method of dynamically controlling the traffic flow. In this method, the on-set of concentrated periods of either high or low traffic activity are identified and the mode of congestion control is adjusted appropriately. Another method of dynamic control of traffic flow is adaptive forward error correction. This method deals with real-time constraints where retransmission of data is not viable like streaming audio or video. When congestion is high, the level of redundancy is increased. In other words, several copies of the same packet are sent in hopes of at least one of them reaching the destination. Duplicate packets received at the destination are discarded. When congestion is low, the level of redundancy is reduced. This method can be dangerous in that increasing the redundancy too much can backfire and only serve to increase congestion.

The other solution used to curtail the effects of Self-Similarity is structural resource allocation. Bandwidth and buffer size are the two structural resources that are of importance. The bandwidth could be increased such that bursts of traffic could be "swallowed." This accounts for the resending of packets as well. However, in times of low network use, the extra bandwidth would be wasted.

The second option in structural resource allocation is the buffer sizes in routers etc. Increasing their capacity would decrease the likelihood of packets being dropped

because the queues are at maximum capacity. The optimal solution to curtailing the effects of Self-Similarity is a mix between these solutions.

2.5 Hurst Parameter

The Hurst Parameter or the Self-Similarity parameter is used as a measure of long-term memory of time series. It relates to the autocorrelation of the time series, and the rate at which these decreases as the lag between pairs of values increases.

It is a key measure of Self-Similarity. The Hurst parameter is referred to as the "index of dependence" or "index of long-range dependence". It qualifies the relative tendency of a time series either to regress strongly to the mean or the cluster in a direction. A value H in the range of 0.5-1 indicates a time series with long-term positive autocorrelation, meaning both that a high value in the series will probably be followed by another high value and that the values a long time into the future will also tend to be high. A value in the range of 0 to 0.5 indicates a time series with long-term switching between high and low values in adjacent pairs, meaning that a single high value will probably be followed by a low value and that the value after that will tend to be high, with this tendency to switch between high and low values lasting a long time into the future. A value of H=0.5 can indicate a completely uncorrelated series, but in fact it is the value applicable to series for which the autocorrelation at small time lags can be positive or negative but where the absolute values of the autocorrelation decay exponentially quickly to zero. This is in contrast to the typically 'power law' decay for the $0.5<H<1$ and $0<H<0.5$ cases.

More precisely, H is a measure of the persistence of a statistical process. Given a stationary Long range dependent sequence, the partial sum if viewed as a process indexed by the number of terms after a proper scaling, is a self-similar process with stationary increments asymptotically. In the converse, given a self-similar process with stationary increments with Hurst index H>0.5, its increments is a stationary long range dependent sequence. A value of H=0.5 indicates the absence of Self-Similarity. The closer H is 1, the grater the degree of persistence or long-range dependence.

2.6 Modeling and Estimation of Self-Similar Data Traffic

A number of approaches have been taken to determine whether a given time series of actual data is self-similar and, if so, to estimate the Self-Similarity parameter H. Some approaches to determine Self-Similarity property are:

1. Variance – Time Plot
2. R/S Plot
3. Whittle's Estimator

Variance-Time Plot:

For the aggregated time series $X^{(m)}$ of a self-similar process, the variance obeys the following for large *m*:

$$Var(X^{(m)}) \approx \frac{Var(X)}{m^{\beta}} \tag{2.5}$$

And the Self-Similarity parameter is defined as, $H = 1 - \beta / 2$. This can be rewritten as

$$\log[\mathrm{Var}(\mathbf{X}^{(m)})] \approx \log[\mathrm{Var}(\mathbf{X})] - \beta \log(m) \tag{2.6}$$

Because $\log[\mathrm{Var}(\mathbf{X})]$ is a constant independent of m, if we plot $\mathrm{Var}(\mathbf{X}^{(m)})$ versus m on a log-log graph, the result should be a straight line with a slope of $-\beta$. The plot is easily generated from the data series $\mathbf{X}(t)$ by generating the aggregate process at different levels of aggregation m and then computing the variance. A number of researches have been done and found that the experimental results do fall on a negatively sloping straight line. It is then a straightforward matter to estimate *H*. Slope values between -1 and 0 suggest Self-Similarity [8].

2.7 Contribution of this work

Lot of work has been done for the Self-Similarity of Ethernet, World Wide Web traffic and Ad Hoc Wireless network traffic. No considerable work has been done on Self-Similarity in the fast growing LTE and LTE-Advanced technology. This piece of work observes the Self-Similarity of real world live network's LTE traffic.

1. Self-Similarity of LTE data traffic is demonstrated using the live network's collected data traffic sets.

2. Hurst parameter values are calculated for each data sets collected.

3. Difference between the degrees of Self-Similarity is observed between different collected data sets.

4. The variation of H-parameter with respect to the slope of Variance-Time Plot is observed.

5. Self-Similarity of LTE data traffic is compared to the Ethernet traffic based on H-Parameters.

Such self-similar time-series can be forecasted [1] and none of the commonly used traffic models is able to capture this fractal behavior, and that such behavior has serious implications for the design, control and analysis of high-speed, cell-based networks [9] and helps in further optimizing the fast growing LTE network.

2.8 Importance of this work

1. This study is very important as the LTE technology is quickly growing in terms of coverage and number of users.

2. Knowledge on the traffic helps in optimizing the network with efficient utilization of resources.

3. Design, Control and Analysis of high-speed, cell-based networks are made more efficient.

4. Knowledge on the degree of Self-Similarity can help in effective forecasting of network traffic as the Self-Similar time-series are forecasted.

5. Self-Similarity is very important as none of the commonly used traffic models are able to capture this fractal behavior, and that such behavior has serious implications for the design, control and analysis of high-speed, cell-based networks.

6. Knowledge of traffic helps in saving the resources like the UE battery life by exploiting intelligent DRX cycles.

2.9 Data Collection & Constraints

Data sets used for this work are collected from a US based cellular operator's live LTE network. The data is captured at eNodeB level which contributes to S1/X2 interfaces in LTE system architecture. There is no separation of data sets based on cell sectors. The eNodeBs from which the data sets are collected typically has 3 cell sectors which are of least relevance to this piece of work. All the data collection is done remotely using an IP traffic collection tool. The captured data include both the user and control traffic logs. In particular, data set's capture include all the message header details, type of message, source and destination details of the message and a timestamp associated with the message. The granularity of the data sets is in the resolution of 10 milliseconds. Each traffic packet can be observed with its header details. Each packet has an average size of around 870 ~ 880 bytes. Overall, the data consists of various data traffic log captures with minimum of 10 min duration till 30 minutes with 10 millisecond resolution and with a minimum of 60,000 entries for each data set. All the logs are collected on multiple eNodeBs at various instances of time with normal load traffic conditions on weekdays. Results of 3 data sets are presented in this work. The importance of these 3 data sets is that they are collected at different instances of a day for same eNodeB at 4AM, 11AM and 6PM where the traffic intensity is expected to vary based on the usual 24-hour traffic pattern.

Constraints:

As with any measurement-based study, these datasets has certain limitations.

1. Since all the data is collected remotely, only a limited duration of data is collected based on the traffic intensity and the buffer size of the traffic collection tool.

2. No traffic data is collected from other technologies like UMTS and 2G for comparison purpose because of the lack of traffic collection tools to perform similar data collection with such granularity.

3. There is no differentiation of Downlink and Uplink traffic in the collected data sets. But it is observed that Uplink constitutes only to a small fraction of total traffic when compared to Downlink.

2.10 Simulation Results

2.10.1 Traffic Representation

A sample real world LTE data traffic is shown in different granularity indexes as below. This traffic is nothing but the data set-I that is collected from a live network. The data set shown has a maximum duration of 10 minutes and includes all the incoming and outgoing data traffic from an eNodeB. The traffic load of 1.23GB is captured in 10 minutes duration with an average of 13.3Mbps data rates in the downlink with nearly 2164 active incoming users. Each packet has an average size of about 800~850 Bytes. The Y-axis represents the traffic size in Bytes per time resolution. As the time resolution increases from minutes to seconds, the Y-axis (bytes/time resolution) decrease correspondingly making a constant count of data traffic collected in the data set.

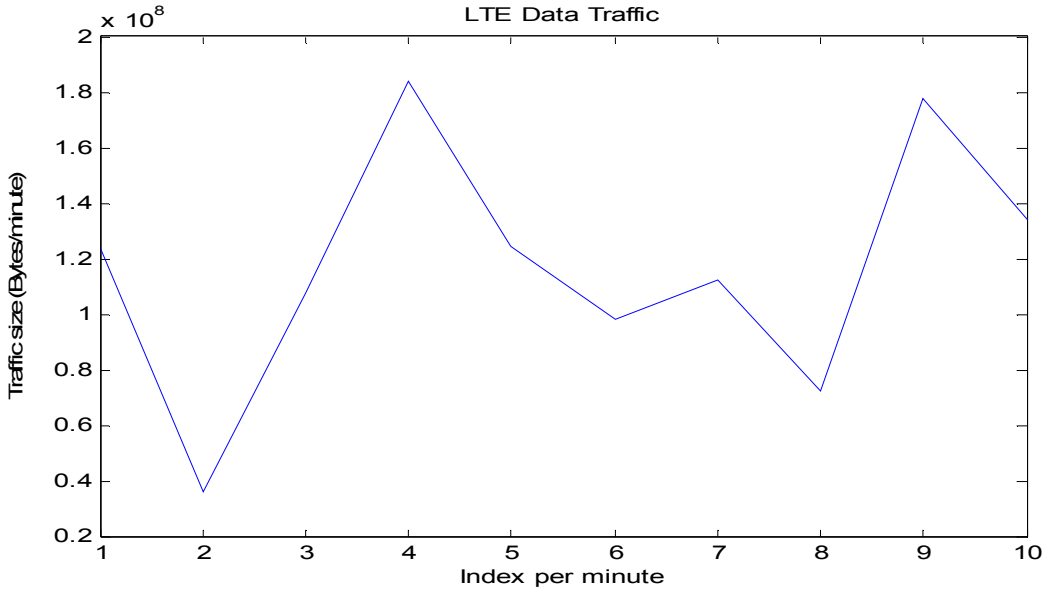

Figure 2-2 Traffic of 10 min shown in minute level Index

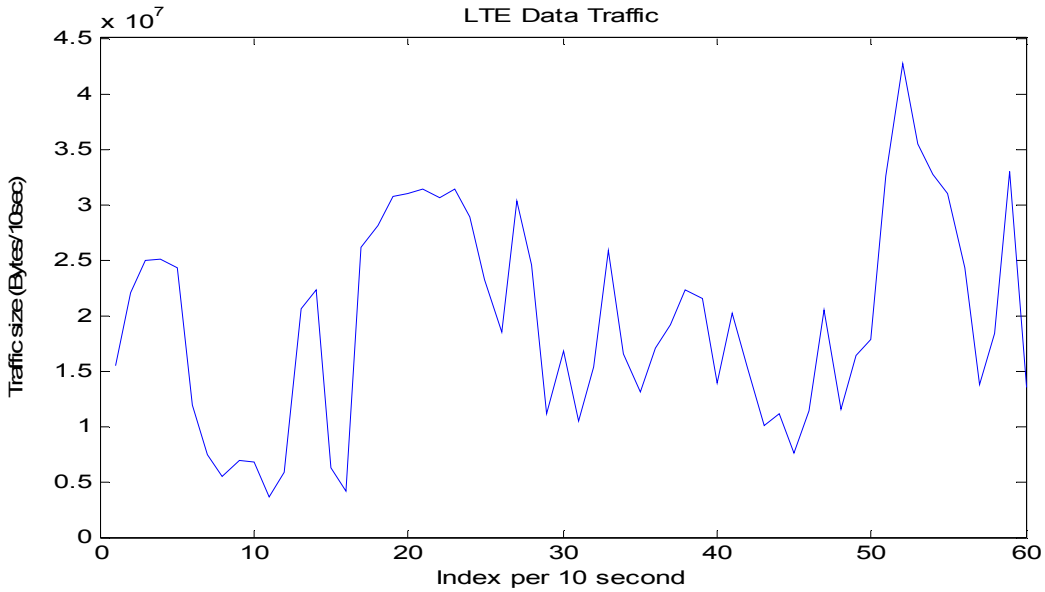

Figure 2-3 Traffic of 10min (600sec) shown in 10sec Index

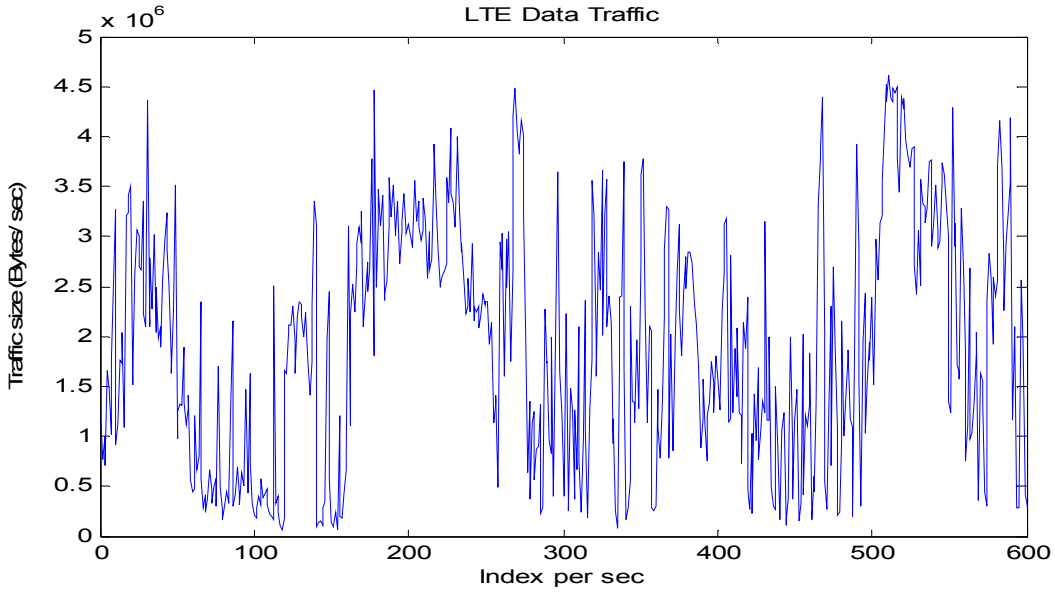

Figure 2-4 Traffic of 10min shown in seconds Index

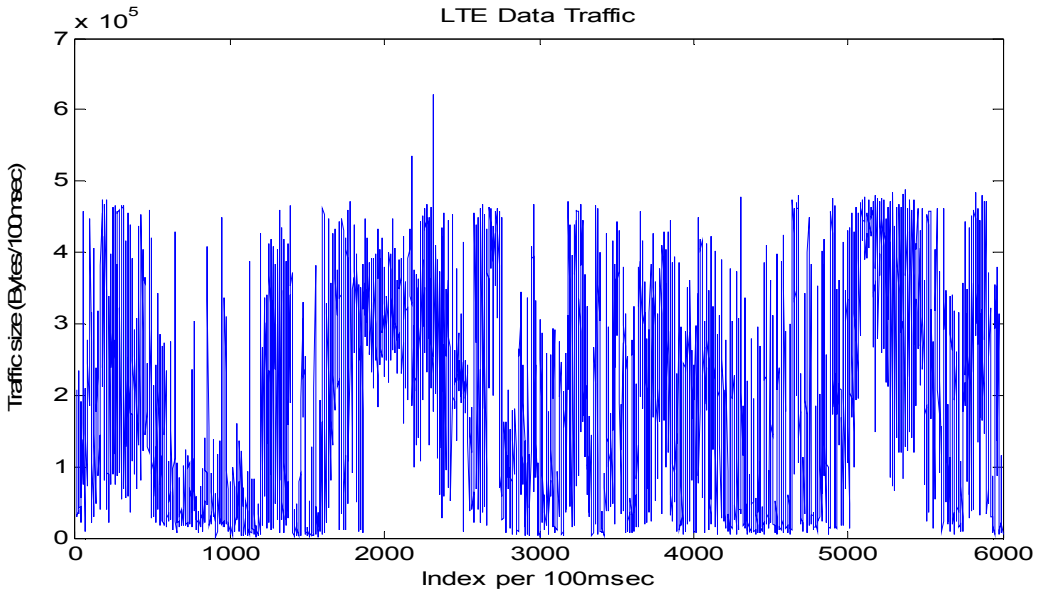

Figure 2-5 Traffic of 10min shown in 100milliseconds Index

The above 4 figures (Figure 2-2 to Figure 2-5) are just the sample of how the collected LTE data traffic looks like in different granularity indexes. It can be observed

that as the granularity of time series is increasing, burstiness of the traffic increases. If the traffic is observed in large indices of minutes, it is almost continuous. But if observed on high granularity of milliseconds, burstiness can be realized with number of Bytes for each millisecond index.

If these traffic series are zoomed in by selecting only interested duration in every index representation, the burstiness can still be observed in detail. The below figures Fig., 2.6 to Fig., 2.9 shows the zoom-in of traffic segment in different granularity indices. Figure 2.9 represents the lowest granularity of 10msec considered for this work. Thin black lines are used to represent the zoom-in portion of traffic in next figures. 10 minutes duration and 1 second granularity traffic is zoomed in till 10millisecinds granularity of 2seconds duration traffic.

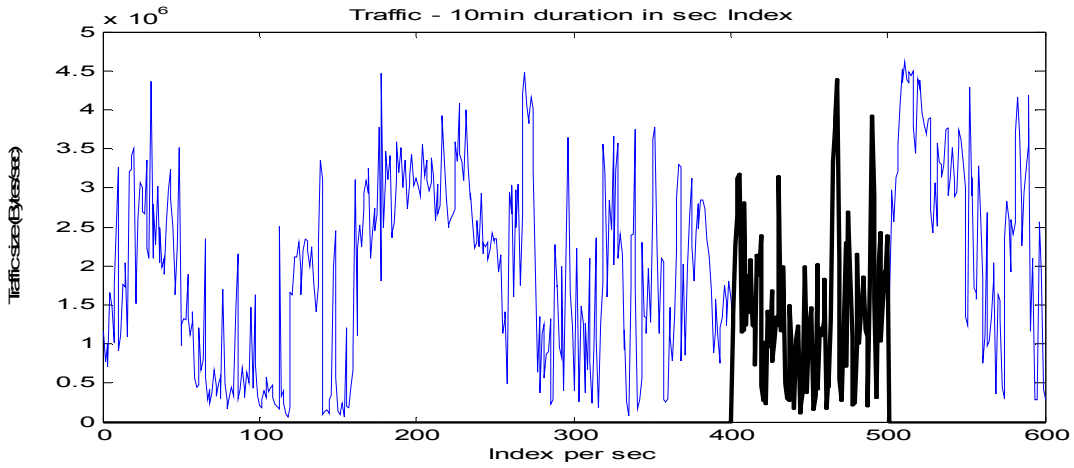

Figure 2-6 Total Data Set – I Traffic Representation

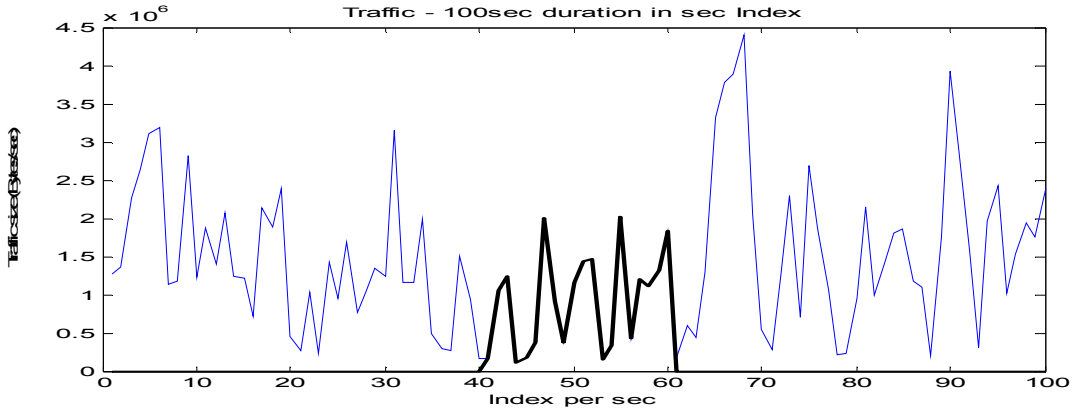

Figure 2-7 Zoom-in of Traffic on second Index

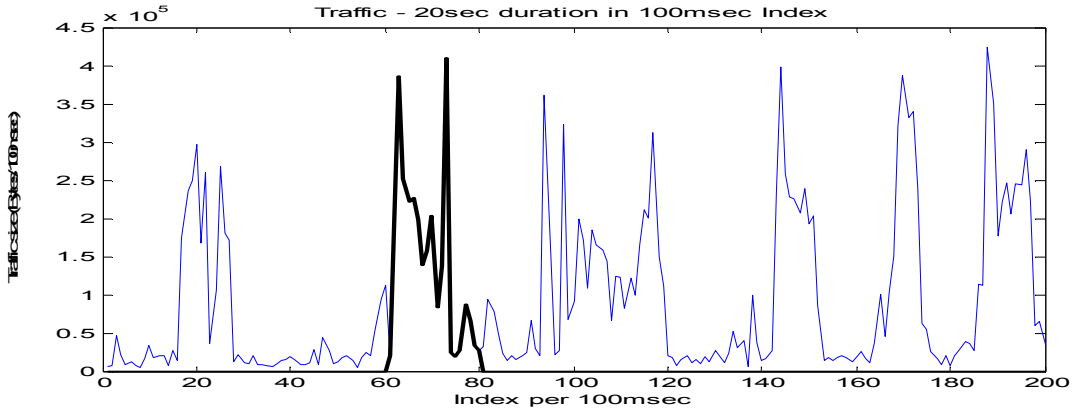

Figure 2-8 Zoom-in of Traffic on 100msec Index

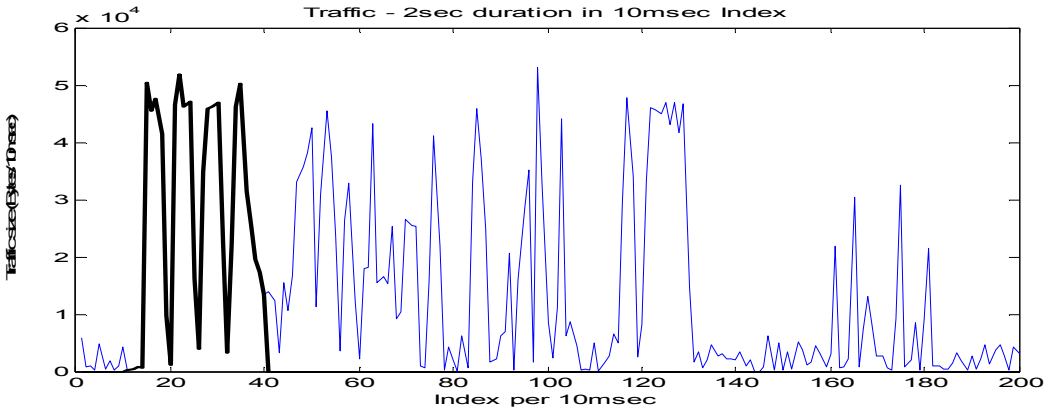

Figure 2-9 Zoom-in of traffic on 10msec Index

2.10.2 Variance – Time Plots

2.10.2.1 V-T Plot for Data Set-I

For the above shown data set-I, V-T graph is plotted to evaluate Self-Similarity property.

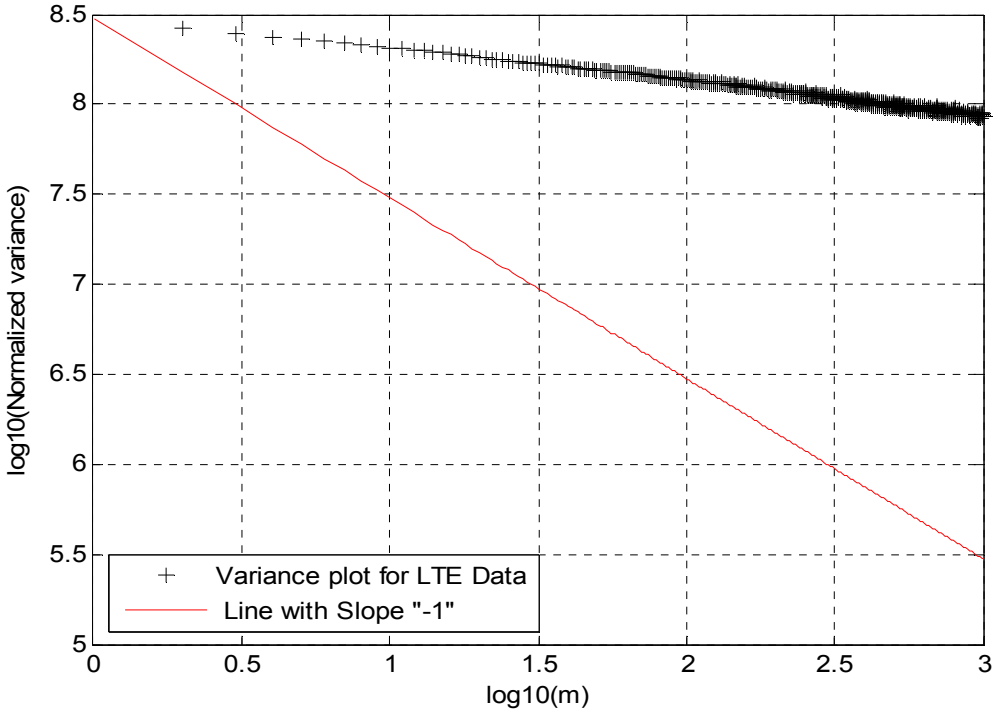

Figure 2-10 V-T Plot for LTE traffic data set -I

V-T Plot for data set -1 is collected at around 6PM on a normal weekday and observed to have a slope greater than -1. So this particular data traffic is concluded to exhibit Self-Similarity property. The Degree of Self-Similarity (H – Parameter) is obtained to be 0.9116. The goodness of fit for the slope of this V-T curve is observed to be -0.1954.

Goodness of Fit:

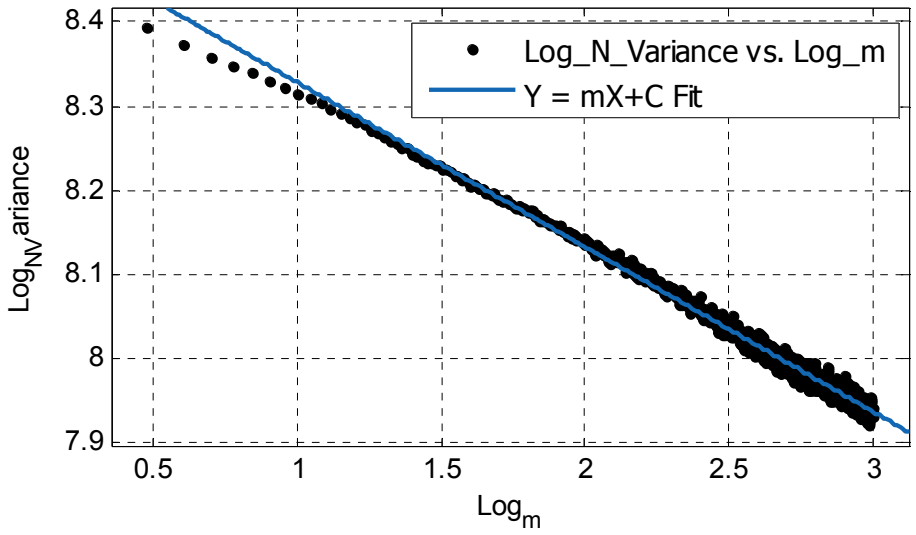

Figure 2-11 Goodness of fit for V-T curve with slope -0.1954

Table 2-1 Goodness of fit for Data Set – I V-T curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.1954 (-0.1966, -0.1942) p2 = 8.523 (8.52, 8.526)	SSE: 0.06147 R-square: 0.9907 Adjusted R-square: 0.9907 RMSE: 0.007856

The plotted V-T curve is fitted with a Y=mX+C curve with best fit slope of -0.1954 and y-intercept of 8.523.

The variation of H-parameter for values m-valued aggregation levels can be observed in the below figure. The variation in H-parameter varies constantly over all aggregation levels holding the fact that Self-Similarity is all about having the same characteristics at different aggregated levels.

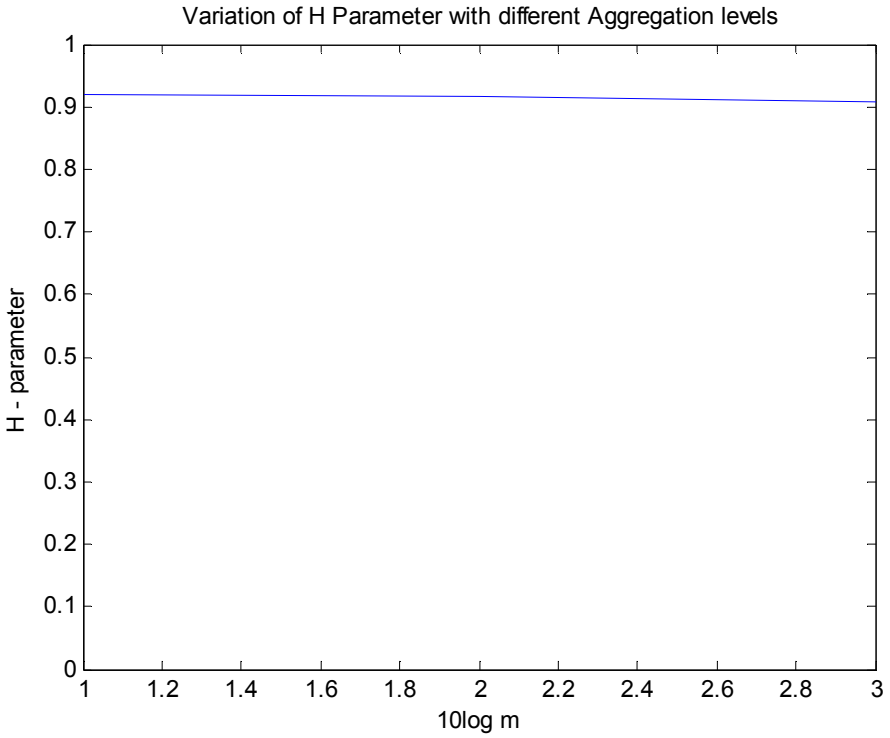

Figure 2-12 Variation of H-Parameter for different m values

2.10.2.2 V-T Plot for Data Set-II

Data Set – II represents the LTE data traffic collected at around 11AM of a normal weekday. The traffic load follows the usual 24-hour pattern and has less load when compared to data set-I (6PM). A traffic load of size 1.02GB is captured for 10minutes duration with average data rates of 12.3Mbps with about 3373 active incoming users. Each packet has an average size of about 800~850 Bytes. A sample traffic of about 20s duration is shown in below figure with 20ms granularity. The actual data duration of data set used for this result is about 10minutes long.

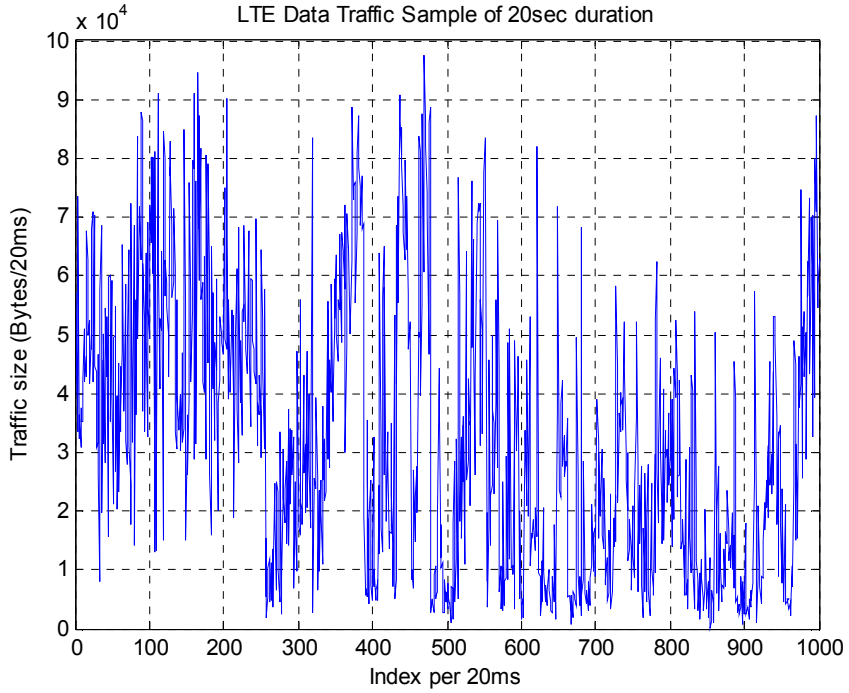

Figure 2-13 Sample data set for 20sec duration at 20msec index

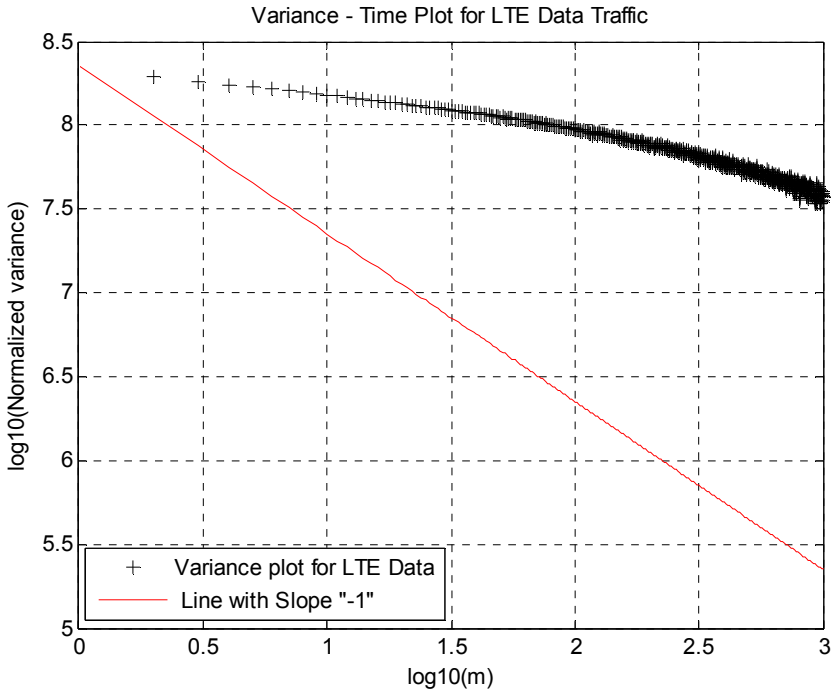

Figure 2-14 V-T Plot for data set-II

V-T Plot for data set -II is collected at around 6PM on a normal weekday and observed to have a slope greater than -1. So this particular data traffic is concluded to exhibit Self-Similarity property. The Degree of Self-Similarity (H – Parameter) is obtained to be 0.8873. The goodness of fit for the slope of this V-T curve is observed to be -0.3402.

Goodness of Fit:

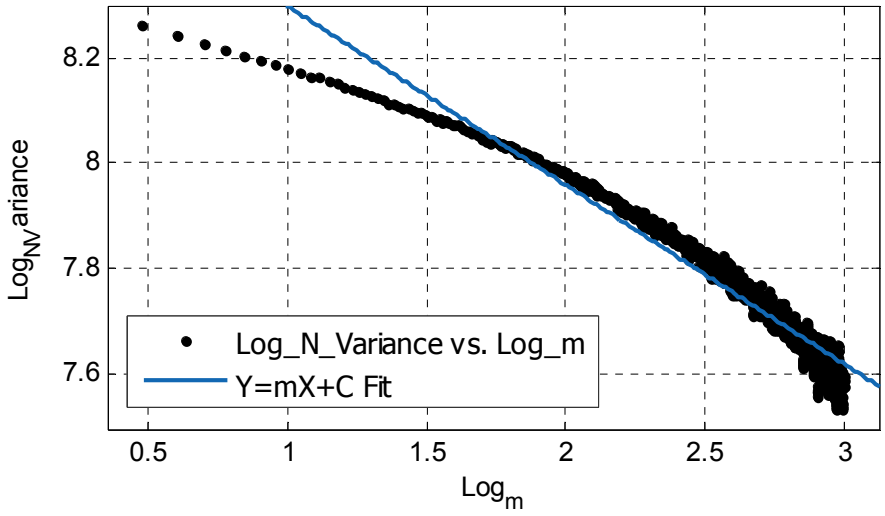

Figure 2-15 Goodness of Fit for V-T Curve with slope -0.3402

Table 2-2 Goodness of fit for Data Set – II V-T curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.3402 (-0.3453, -0.3351) p2 = 8.639 (8.625, 8.652)	SSE: 1.156 R-square: 0.9449 Adjusted R-square: 0.9448 RMSE: 0.03407

The plotted V-T curve is fitted with a Y=mX+C curve with best fit slope of -0.3402 and y-intercept of 8.639.

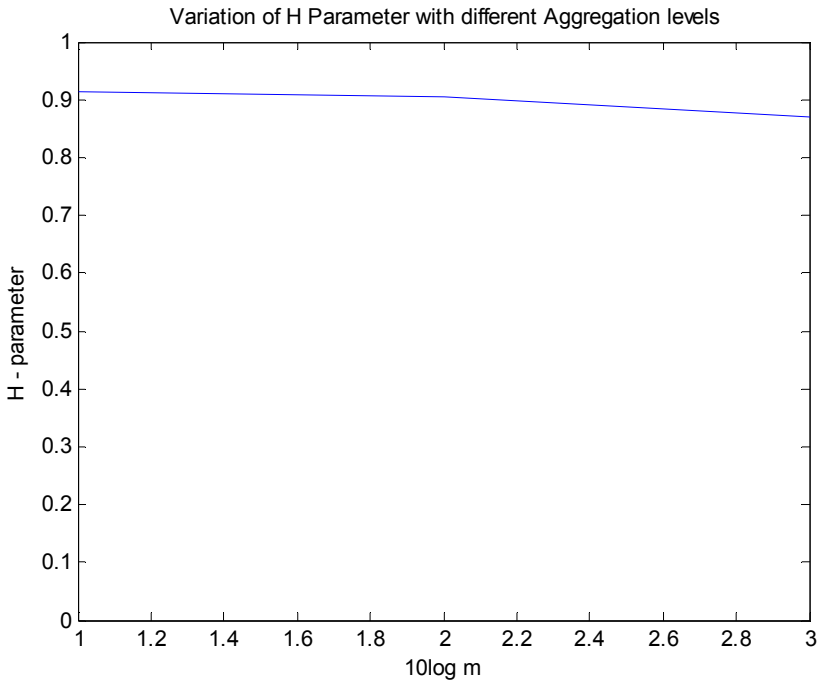

Figure 2-16 H-Parameter Variation for different aggregation levels

The variation of H-parameter for values m-valued aggregation levels can be observed in the above figure. The variation in H-parameter varies constantly over all aggregation levels holding the fact that Self-Similarity is all about having the same characteristics at different aggregated levels.

2.10.2.3 V-T Plot for Data Set-III

Data Set – III represents the LTE data traffic collected at around 4AM of a normal weekday. The traffic load follows the usual 24-hour pattern and has fewer loads when compared to data set-II (11AM). A traffic load of size 0.65MB is captured for 10minutes duration with average data rates of 8Kbps with about 35 active incoming users. Each packet has an average size of about 800~850 Bytes. A sample traffic of about 20s duration is shown in below figure with 20ms granularity.

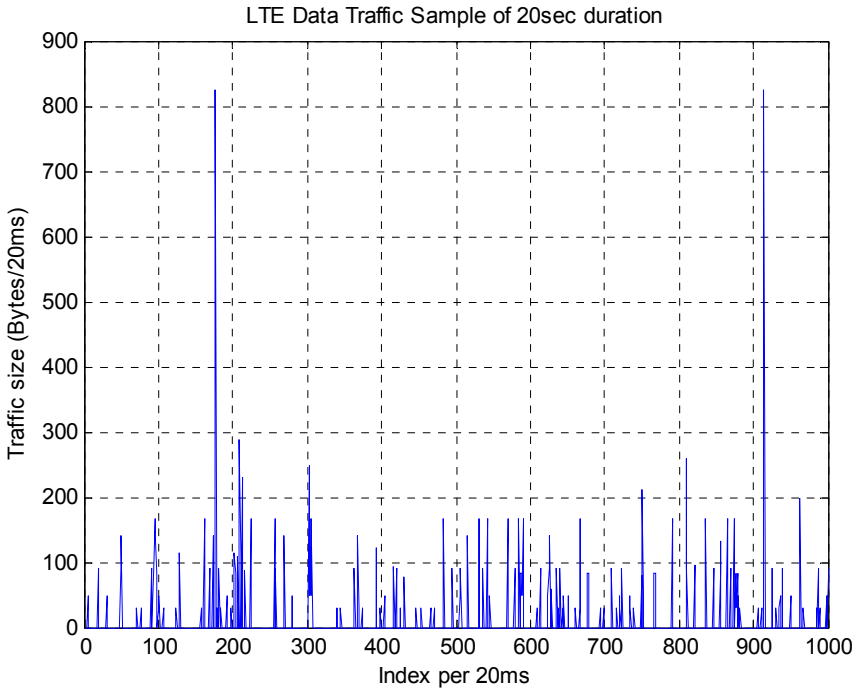

Figure 2-17 Sample data set for 20sec duration at 20msec index

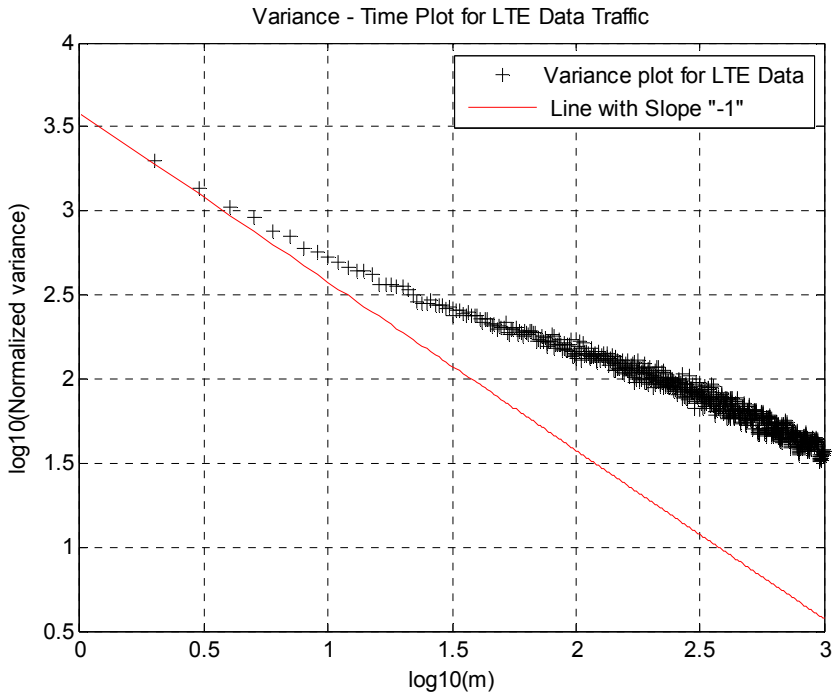

Figure 2-18 V-T Plot for data set-III

V-T Plot for data set -III is collected at around 4AM on a normal weekday and observed to have a slope greater than -1. So this particular data traffic is concluded to exhibit Self-Similarity property. The Degree of Self-Similarity (H – Parameter) is obtained to be 0.6609. The goodness of fit for the slope of this V-T curve is observed to be -0.5791.

Goodness of Fit:

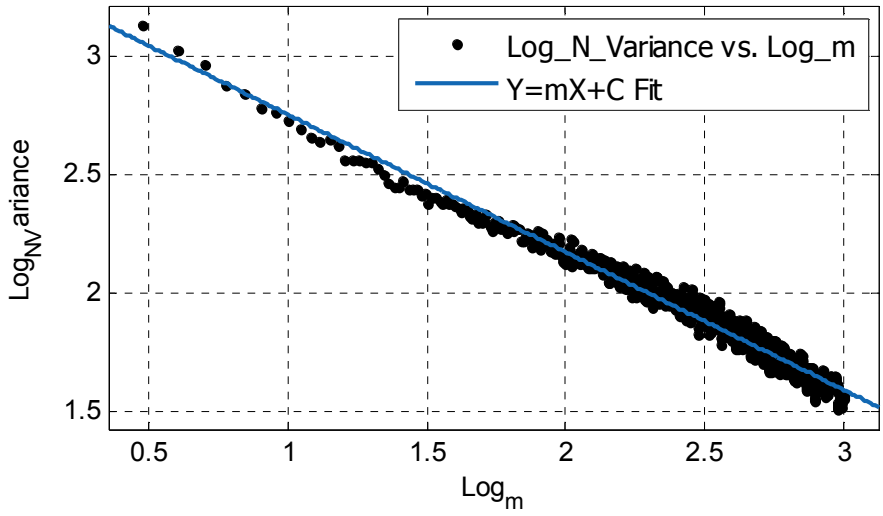

Figure 2-19 Curve fit for V-T plot with slope -0.5791

Table 2-3 Goodness of fit for Data Set – III V-T curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.5791 (-0.5851, -0.573) p2 = 3.332 (3.316, 3.348)	SSE: 1.615 R-square: 0.9726 Adjusted R-square: 0.9726 RMSE: 0.04027

The plotted V-T curve is fitted with a Y=mX+C curve with best fit slope of -0.5791and y-intercept of 3.332.

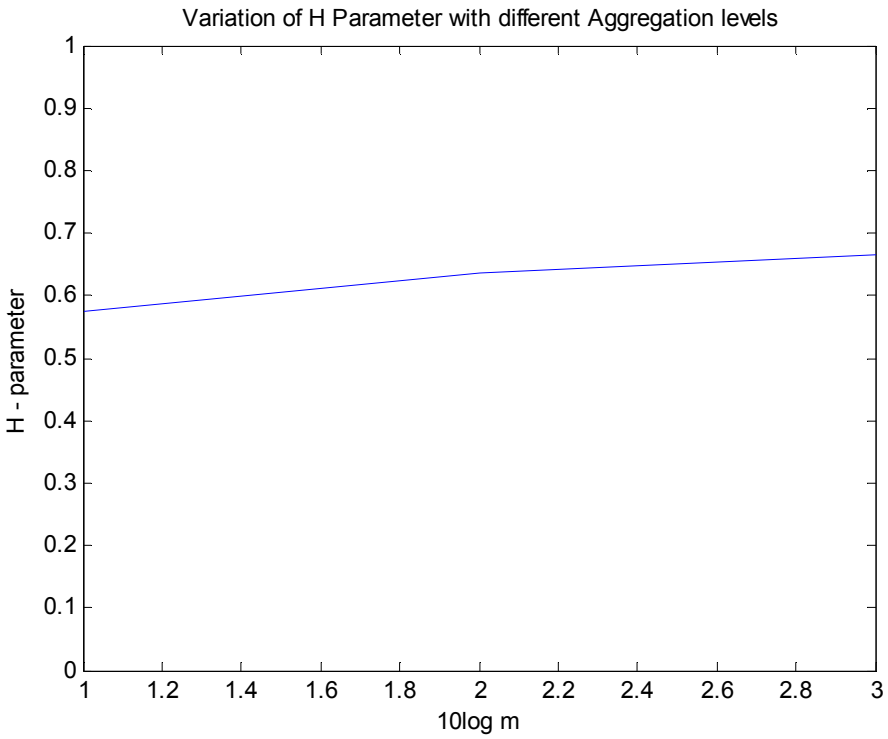

Figure 2-20 H-Parameter Variation at different aggregation levels

The variation of H-parameter for values m-valued aggregation levels can be observed in the above figure. The variation in H-parameter varies constantly over all aggregation levels holding the fact that Self-Similarity is all about having the same characteristics at different aggregated levels.

2.10.3 Comparison of Data Sets

Comparison of all the 3 data sets can be done on the basis of the typical load corresponding to the time of the day at which data is collected. The below table summarizes the results of 3 data sets with the slope of V-T Curve and the degree of Self-

Similarity (H-Parameter). It can be observed that as the time of the day increases, traffic load increases in terms of the data captured and so the H-parameter value. The slope of V-T plot also increased along with load and H-parameter value.

Table 2-4 Comparison of H – Parameters for different LTE data sets

Data Sets	**Time of the Day**	**Traffic Load**	**Slope**	**H-Value**
I	18.00	1.23 GB	-0.1954	0.9116
II	11.00	1.02GB	-0.3402	0.8873
III	04.00	0.65MB	-0.5791	0.6609

From [9]'s vast Ethernet data collection and observation of Self-Similarity for about 4 years of duration with over 100 million packets, the highest H-Parameter value observed is only 0.9 and average H-Parameter is only around 0.85. But with only limited duration data in LTE network, the average H-parameter is around 0.9. This suggests that LTE data traffic shows high degree of Self-Similarity when compared to that of Ethernet data traffic.

H-Parameter variation for different m-level aggregations is also observed with the comparison of all the data sets. H-parameter variation of all the data sets falls only

between 0.5 and 1 as described. The variation trending of each individual data set is almost the same over various aggregation levels.

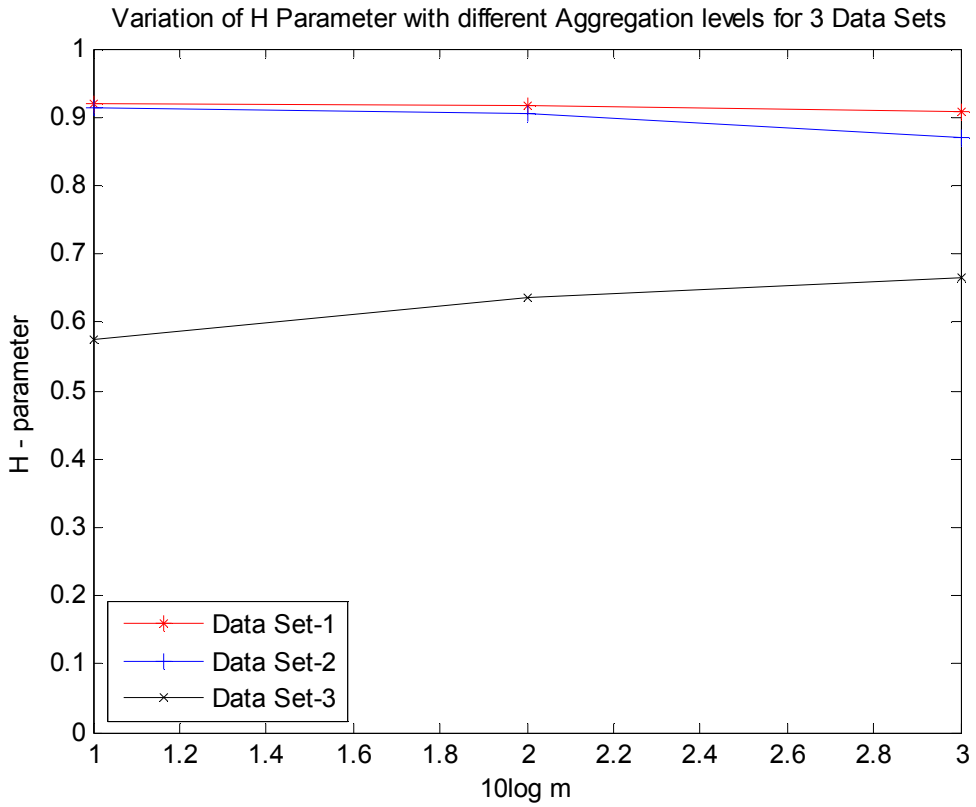

Figure 2-21 Comparison of H values for different Data Sets

2.11 Conclusions

1. Self-Similarity of real world LTE data traffic is demonstrated using the collected data set's V-T plots.

2. Hurst parameter values are calculated for each data sets collected.

3. Variation of H-Parameter remains almost the same with different aggregation levels which depicts the meaning of Self-Similarity.

4. As the traffic load varies, H-Parameter and the slope of V-T plot vary. This suggests that high traffic has more degree of Self-Similarity.

5. It is observed that H-Parameter is in direct relation to the slope of Variance-Time Plot. It increases as the V-T curve slope increases and vice versa.

6. LTE data traffic shows higher degree of Self-Similarity when compared to the Ethernet traffic based on the previous work.

Chapter 3

LTE USER TRAFFIC MODELING

3.1 Introduction

The growth of telecommunication system is more than ever stimulating the need for mathematical modeling and evaluation techniques to predict the performance of these systems. These techniques have become an indispensable tool for assessment of the performance of a system, evaluation of design alternatives, resource dimensioning and system configuration [13].

Poisson distribution is a good representative for most of the telecommunication events. The arrival process of calls to a call center is usually associated with a Poisson process. The Poisson Process provides simple solution to many different problems related to call center operations management. However, the latest telecommunication technology i.e., 4G framework is establishing new levels of user experience and multi-service capacity by integrating all the previous mobile communication technologies. This multiple services accounts for the arrival of user requesting for specific service. In this scenario, a user arrival for service is not just equivalent to be for a traditional voice call but also can be for data browsing session or a video call etc.

In the present scenario of multiple services, it is necessary to model the arrival of users requesting for a service. This piece of work is focused on modeling the user arrival to an eNodeB requesting for any kind of data services. Since this work is based on the data collected on live eNodeBs in the network, most of the user arrival is expected to be requesting for a data session and also some voice sessions like the voice over LTE.

The main assumption of the Poisson arrival model are the number of incoming calls per interval follows the Poisson distribution, and that the numbers of arrivals in the disjoint intervals are independent. We are interested in studying the arrivals to the eNodeB requesting for any kind of service either directly or indirectly. This work is based on a real world live network similar to [19] which is based on real traffic.

3.2 Arrival Process

A Poisson process is a simple and widely used stochastic process for modeling the times at which arrivals enter a system. It is in many ways the continuous-time version of the Bernoulli process. Poisson process could also be characterized by a sequence of geometrically distributed inter-arrival times.

For the Poisson process, arrivals may occur at arbitrary positive times, and the probability of an arrival at any particular instant is 0. This means that there is no very clean way of describing a Poisson process in terms of the probability of an arrival at any given instant. It is more convenient to define a Poisson process in terms of the sequence of inter-arrivals times, $X_1, X_2, \ldots$, which are defined to be independent and identically distributed (IID).

An arrival process is a sequence of increasing random variables $0 < S_1 < S_2 < \ldots$, where $S_i < S_{i+1}$ means that $S_{i+1}-S_i$ is a positive random variable X such that $F_X(0) = 0$. The random variables S_1, S2, are called arrival epochs and represent the successive times at which some random repeating phenomenon occurs. Usually this process starts at time 0 and that multiple arrivals can't occur simultaneously (the phenomenon of bulk arrivals

can be handled by the simple extension of associating a positive integer random variable to each arrival) [14].

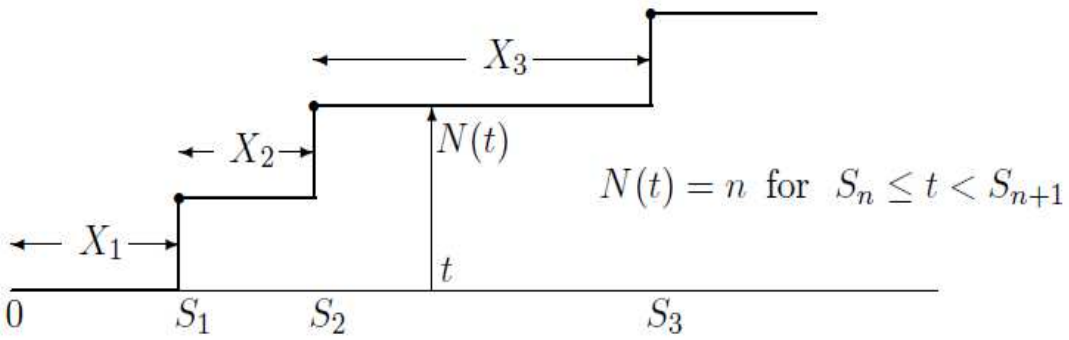

Figure 3-1 Poisson Arrival representation

Any arrival process can also be specified by either of two alternative stochastic processes. The first alternative is the sequence of inter arrival times { X_1, $X_{2,...}$}. The Second is to specify an arrival process in the counting process N (t). In inter arrival times, $X_1 = S_1$ and $X_t = S_t - S_{t-1}$ for i>1. Similarly, given the X_t, the arrival epochs S_t are specified as

$$S_n = \sum_{t=1}^{N} X_t \tag{3.1}$$

For the second alternative of counting process N(t), where for each t>0, the random variable N(t) is the number of arrivals up to and including time t.

The counting process {N(t); t>0} represents an uncountable infinite family set of random variables where N(t), for each t>0, is the number of arrivals in the interval (0,t]. N(0) is always defined to be 0 with probability 1, which means, as before, that we considering only arrivals at strictly positive times.

The counting process $\{N(t), t \geq 0\}$ for any arrival process has the properties that $N(\tau) \geq N(t)$ for all $\tau \geq t > 0$ (i.e., $N(\tau) - N(t)$ is a non-negative random variable. For any given integer $n \geq 1$ and time $t \geq 0$, the arrival epoch, S_n, and the counting random variable, N(t), are related by

$$\{S_n \leq t\} = \{N(t) \geq n\} \tag{3.2}$$

$\{S_n \leq t\}$ is the event that the nth arrival occurs by tie t. This event implies that N(t), the number of arrivals by the time t, must be at least n; i.e., it implies the event $\{N(t) \geq n\}$. Similarly $\{N(t) \geq n\}$ implies $\{S_n \leq t\}$, yielding the equality in above equation.

So, an arrival process can be specified by the joint distribution of the arrival epochs, the inter-arrivals, or the counting random variables. In principle, specifying any one of these specifies the other also [14].

Arrival Rate:

In modeling arrival process in terms of exponential inter-arrivals, Average Inter-arrival time is represented as

$$\overline{T} = \frac{\sum_{t=1}^{N} X_t}{N} \tag{3.3}$$

So, the expected arrival rate = $\hat{\lambda} = \frac{1}{\overline{T}}$

For a Poisson arrival process, if we plot (t, N(t)) it will follow y=m*x line with slope as λ. From page 167 of [14], for a Poisson process of rate λ and for any t>0, the PMF for N(t) (i.e., the number of arrivals in (0,t] is given by the Poisson PMF

$$p_{N(t)}(n) = \frac{(\lambda t)^n \exp(-\lambda t)}{n!} \tag{3.4}$$

Properties:

To describe calls occurring in communication systems, the so-called Poisson stream is most frequently used, which has the following properties.

1. N(0)=0; Initial event as time zero is always zero.
2. Stationary – A stream is stationary if its intensity does not depend on time: $\lambda(t) = \lambda = const$; this means that the average number of arrival calls within a time interval remains unchanged.
3. Memorylessness – A stream has a memoryless property if the number of calls within any chosen time interval t_1 does not have any effect upon the number of calls within any other, randomly chosen, interval t_2; this means that successively arriving calls are not mutually interdependent.
4. Orderliness (singularity) – A stream that is singular if within an infinitely small time interval Δt , one call at most can arrive; the probability of the arrival of more than one call is shown below.

A Poisson call stream that is characterized by the above properties is often called the simple stream.

3.3 Contribution of this Work

Traffic modeling is very important to better understand and utilize network resources. In this present generation of telecommunication system with multiple types of traffic, it is necessary to model the traffic for better understanding of the network. This work

highlights the importance of evaluation in LTE network, which has high data traffic when compared to traditional traffics.

A Poisson Process model provides simple solutions to many different problems related to call center operation management. But there is no considerable work done in LTE network based on real world data. This work adds quality in this angle.

1. User arrival in LTE network requesting for a service is modeled to follow Poisson Process.

2. User's service request Inter-Arrivals is evaluated to follow exponential distribution

3. Poisson arrival process is modeled for these inter-arrivals.

4. Mean arrival rate parameters are measured for the collected real world data sets.

5. Goodness of fit for all the results is observed with error deviation.

3.4 Importance of this Study

Knowledge on the traffic arrival helps in further optimizing the network with efficient utilization of resources.

1. Design, Control and analysis of high-speed, cell-based netowrks are made further efficient.

2. With the knowledge of arrival rate parametrs, traffic prediction algorithms at network level can be made more robust.

3. This work models the traffic in LTE network which is fast growing and expected to be ubiquitous.

4. Since this work is based on live network's data, it can serve as a base reference for future traffic analysis on LTE networks.

3.5 Data Collection & Constraints

Data sets used for this work are collected from a US based cellular operator's live LTE network. The data is captured at eNodeB level which contributes to S1/X2 interfaces in LTE system architecture. There is no separation of data sets based on cell sectors. The eNodeBs from which the data sets are collected typically has 3 cell sectors which are of least relevance to this piece of work. All the data collection is done remotely using an IP traffic collection tool. The captured data include both the user and control traffic logs. In particular, data set's capture include all the message header details, type of message, source and destination details of the message, user service request details and a timestamp associated with the message. The granularity of the data sets is in the resolution of 1 millisecond. Overall, the data consists of various data traffic log captures with 30 minutes duration and 1millisecond granularity of analysis collected on multiple eNodeBs at various instances of time with normal load traffic conditions on weekdays.

Constraints:

As with any measurement-based study, these datasets has certain limitations.

1. Since all the data is collected remotely, only a limited duration (30 min) of data is collected based on the traffic intensity and the buffer size of the traffic collection tool.
2. No traffic data is collected from other technologies like UMTS and 2G for comparison purpose because of the lack of robust tools to perform similar data collection with such granularity.

3.6 User Service Request Consideration

As this work is on modelling the user arrival requesting for a service to network, it is important to clarify the 'User Arrival' consideration. A costumer/user requests for service in two ways as, Mobile Origination of service and Mobile termination of service. (i) Mobile origination refers to the service requests made by the user like, User trying to make a phone call, User starting a browsing session etc. (ii) Mobile termination refers to the service requests that should be provided to the user. Like, User getting an incoming call, paging request, scheduled downlink data etc. (iii) Apart from these two arrivals, one more scenario of X2 handover needs is to be considered as the new user coming to the target eNodeB all of a sudden with request for the support of an on-going service from the old eNodeB. These 3 cases should include all the normal service requests from the serving users. In collected data logs, we can seperate these three scenarios in 'service request' and 'extended service request' messages as mo-data, mt-access & path switch requests.

3.7 Simulation Results

3.7.1 Data Set-I

Data Set – I is a live network's data traffic collected from one of the site in Kansas City. The data set is captured at around 1PM on a normal weekday with normal load following 24 hour usual trend. The duration of data set is around 30 minutes and includes all the header details of user plane and control plane messages with details of user arrival requests. The user plane traffic includes around 1.08GB data with an average data rate of 21.9Mbps for 30 minutes duration. About 3436 user requests are observed for this 30 minutes duration of log collection.

3.7.1.1 Inter- Arrival Exponential Distribution

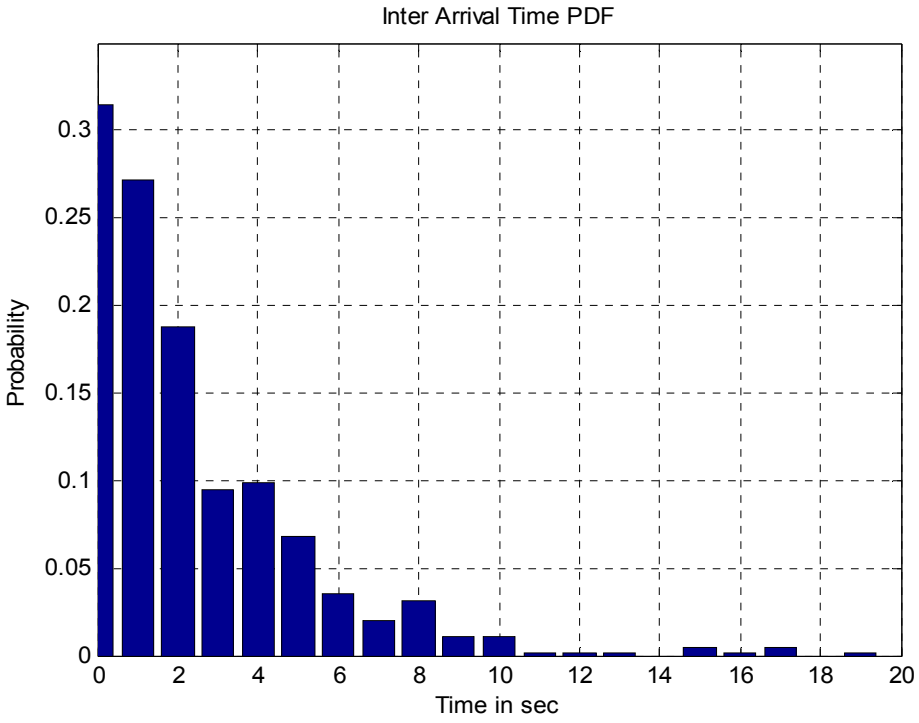

Figure 3-2 Inter Arrival Time PDF of User Arrivals

The data set provides the details of time difference between each user arrival which is nothing but the Inter Arrival Time. The probability distribution of Inter-Arrival times is plotted with Inter-Arrival time on X-axis and Probability function on Y-axis and observed to follow exponential distribution as shown.This exponential distribution is observed to have lambda = 0.33 as Arrival Rate.

Goodness of Fit:

The Inter-Arrival distribution is fitted with Exponential curve with an exponential power of 0.33. The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares due to Errors (SSE), (ii) R-Square, (iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE).

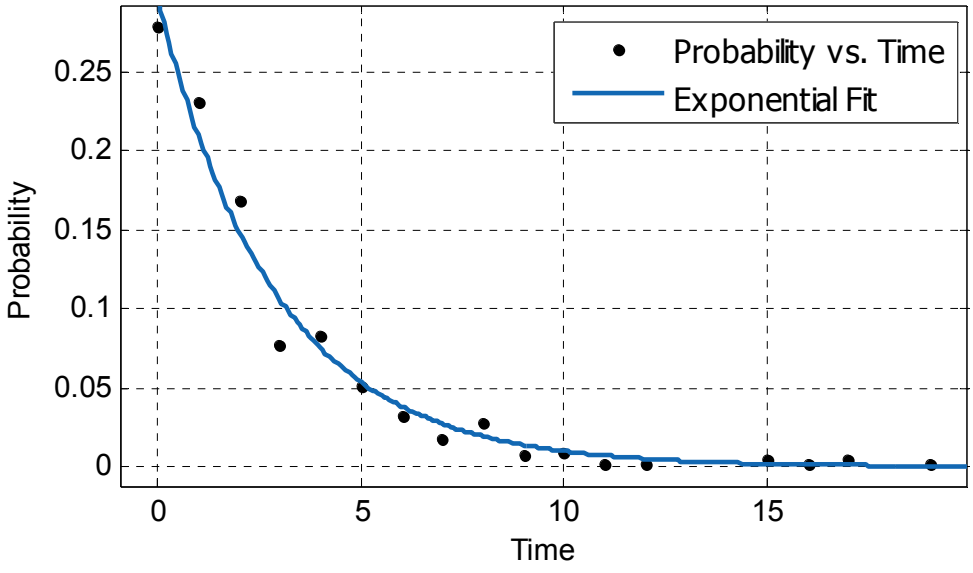

Figure 3-3 Goodness of fit for Exponential Distribution

Table 3-1 Goodness of fit for Data Set – I Exponential curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = a*exp(b*x)	a = 0.3345 (0.3095, 0.3595) b = -0.3307 (-0.3717, -0.2898)	SSE: 0.003032 R-square: 0.981 Adjusted R-square: 0.9798 RMSE: 0.01377

Q-Q Plot:

Q-Q Plot can help in comparing the theoretical and practical distributions. The goodness of fit can be observed in graphical form from these plots. Q-Q Plot is implemented for the theoretical exponential and above plotted Inter-Arrival distribution and it is observed to have fair goodness of fit in concurrence with the numerical goodness of fit. Average Inter-Arrival (T) is observed to be 3.02 Sec/User.

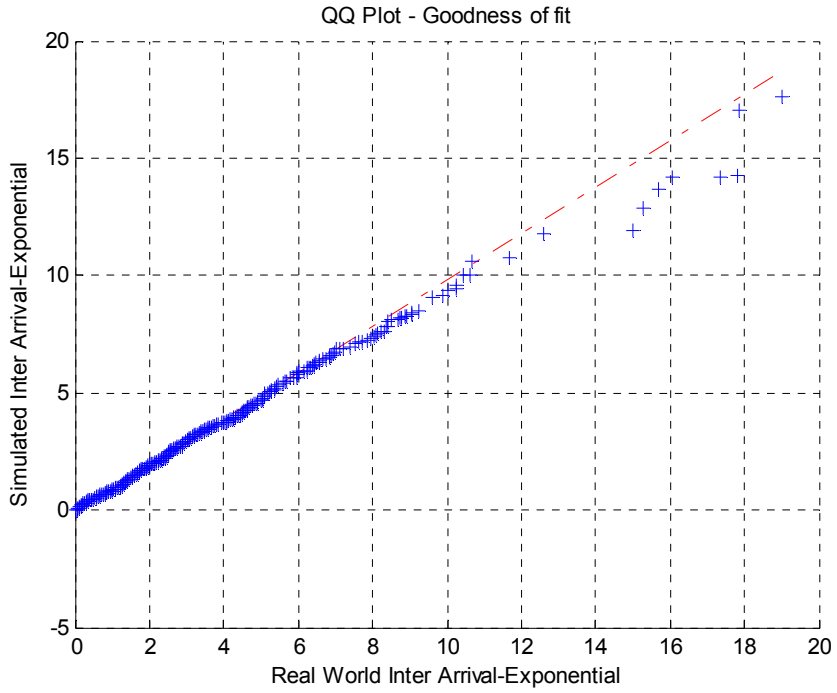

Figure 3-4 QQ-Plot for Exponential distribution Fit

3.7.1.2 Poisson Arrival

An arrival process can also be specified in counting process with the number of arrivals in the specified time duration. As the counting process can include infinite family set of random user arrivals, the count should always be an increasing function as the time progresses. For a Poisson Arrival process, if the arrivals are plotted against time, it will follow a straight line. The data set's Poisson modeling is observed to follow the same as shown below. The user arrival counting process is modeled as poisson with arrival rate (λ) of 0.3604.

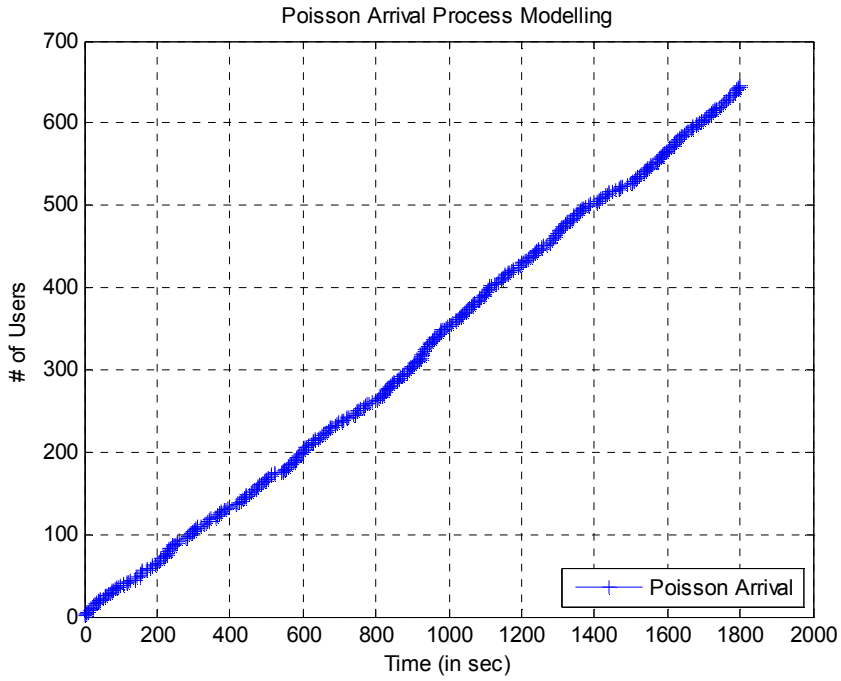

Figure 3-5 Poisson Arrival of Users is plotted across time

Goodness of Fit:

The plotted Poisson counting process is fitted to a Y=mX+C straight line with good slope fit of 0.3604 which is the Arrival Rate.

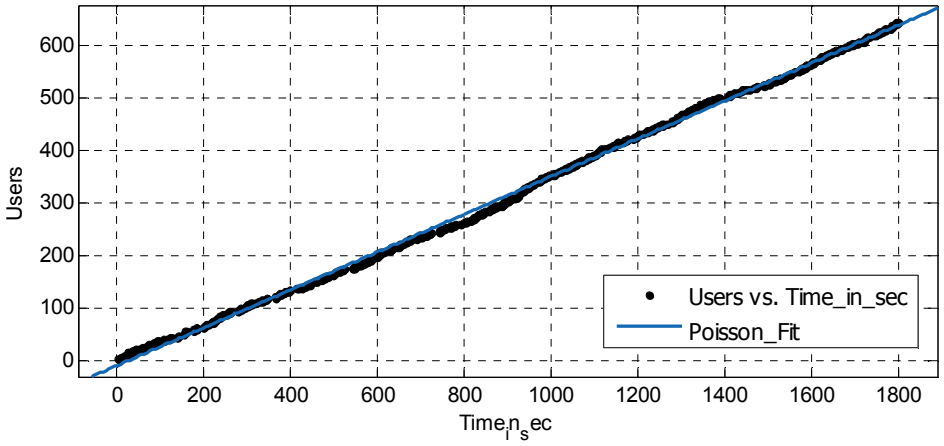

Figure 3-6 Goodness of Curve fit for Poisson Arrival Process

Table 3-2 Goodness of fit for Data Set – I Poisson Arrival

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = 0.3604 (0.3593, 0.3615) p2 = -9.166 (-10.33, -8.005)	SSE: 3.444e+04 R-square: 0.9985 Adjusted R-square: 0.9985 RMSE: 7.319

The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares due to Errors (SSE), (ii) R-Square, (iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE).

QQ-Plot:

Q-Q Plot is implemented for the theoretical straight line and above plotted Inter-Arrival distribution and it is observed to have fair goodness of fit in concurrence with the numerical goodness of fit. Average Inter-Arrival (T) is observed to be 3.02 Sec/User.

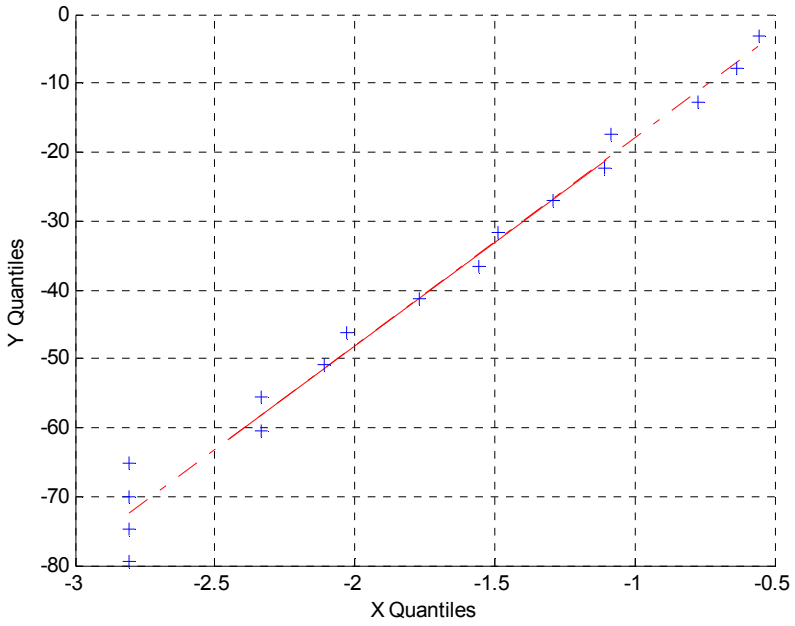

Figure 3-7 QQ Plot for Poisson Arrival Fit

3.7.1.3 CDF of Instantaneous Active Incoming Users

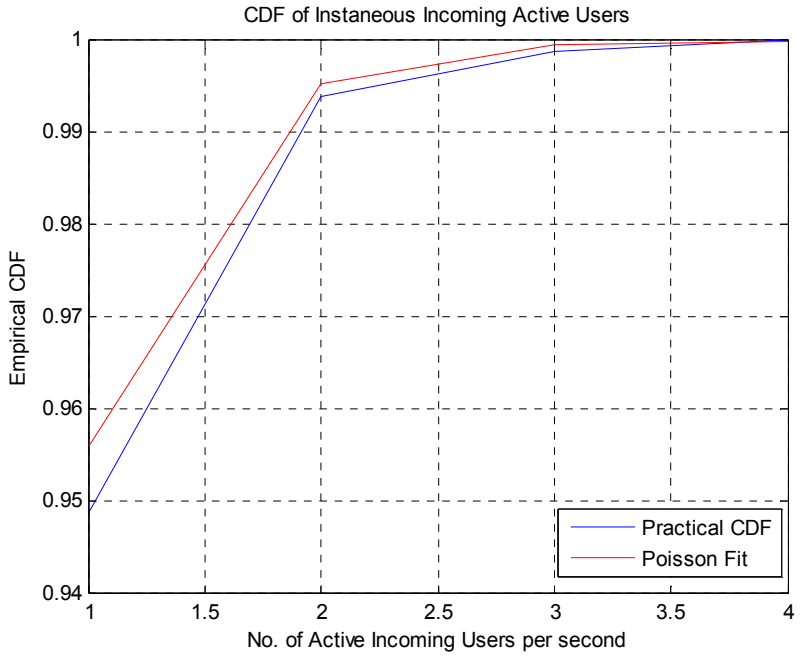

Figure 3-8 CDF of instentanous Active Users

The CDF is discontinuous at the integers of X-axis and flat everywhere else because a variable that is Poisson distributed only takes on integer values. A Poisson CDF fit is also shown in the above CDF plot for comparison [20].

3.7.1.4 Logarithm of Exponential Inter Arrival

The Log plot of Exponential Inter-Arrival is expected to be a Straight Line. This servs as a verification of the exponential Inter-Arrival apart from the numerical goodness of fit and pictorial Q-Q Plot fit. This is observed to a close straight line but not a perfect one.

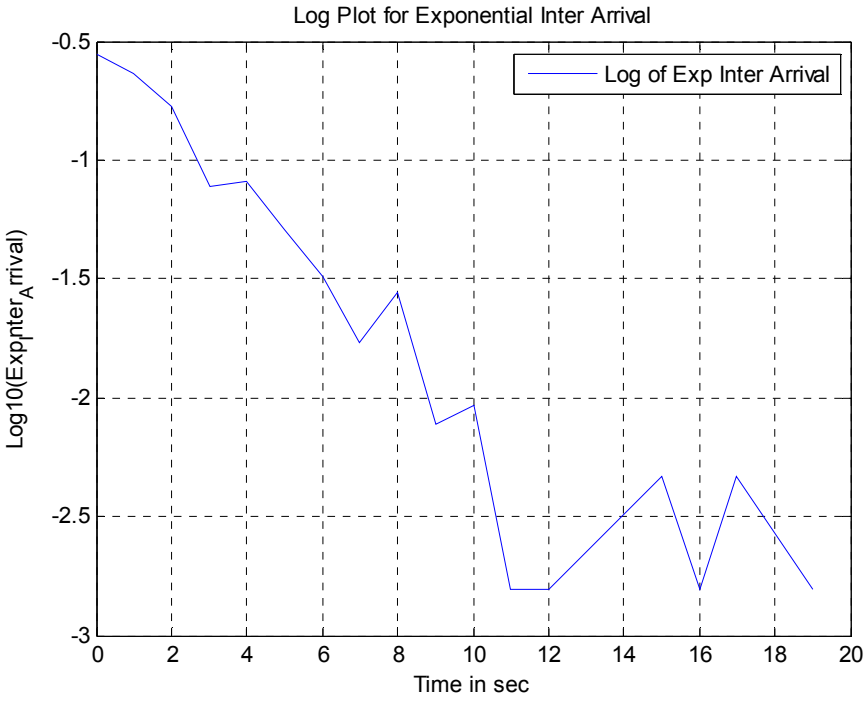

Figure 3-9 Log of Exponential Poisson Inter Arrival distribution

Goodness of Fit:

The Logarithm of Exponential Inter Arrival Curve is fitted with Y-mX+C straight line.

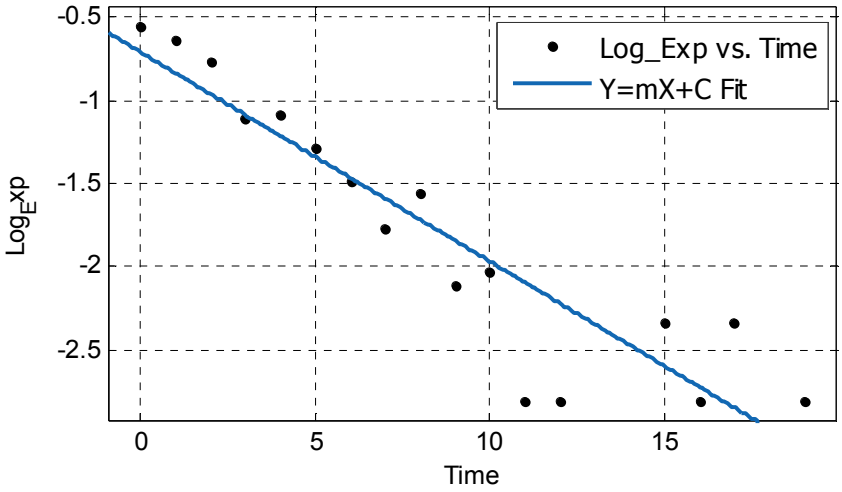

Figure 3-10 Curve Fitting for Log- Poisson Inter Arrival Distribution

The plotted Logarithm of Exponential distribution curve is fitted with a Y=mX+C curve with best fit slope of -0.1258 and y-intercept of -0.7096.

Table 3-3 Goodness of fit for Data Set – I Log of Exponential

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.1258 (-0.1551, -0.0965) p2 = -0.7096 (-1.009, -0.4097)	SSE: 1.542 R-square: 0.8481 Adjusted R-square: 0.838 RMSE: 0.3207

Q-Q Plot:

Q-Q Plot is implemented for the theoretical straight line and plotted logarithm of Exponential Inter-Arrival distribution and it is observed to have fair goodness of fit in concurrence with the numerical goodness of fit.

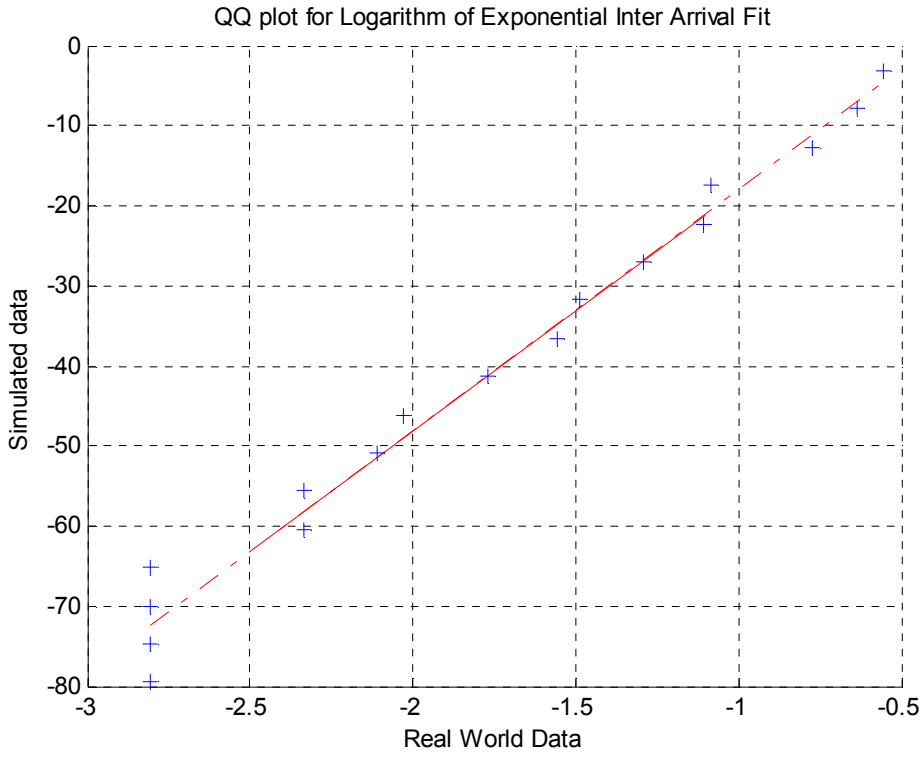

Figure 3-11 QQ-Plot for Log of Poisson Inter Arrival Distribution

3.7.2 Data set – II

Data Set – II is a live network's data traffic collected from one of the site in Kansas City. The data set is captured at around 4PM on a normal weekday with normal load following 24 hour usual trend. The duration of data set is around 30 minutes and includes all the header details of user plane and control plane messages with details of user arrival requests. The user plane traffic includes around 1GB data with an average data rate of 4Mbps for 30 minutes duration. About 1285 user requests are observed for this 30 minutes duration of log collection.

3.7.2.1 Inter- Arrival Exponential Distribution

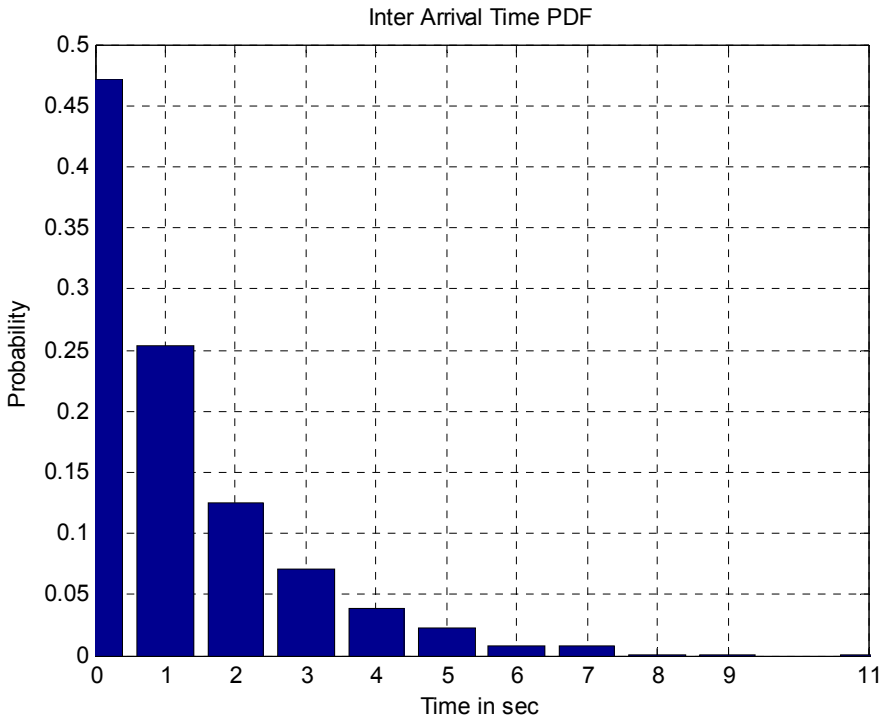

Figure 3-12 Inter Arrival Time PDF of User Arrivals

The data set provides the details of time difference between each user arrival which is nothing but the Inter Arrival Time. The probability distribution of Inter-Arrival times is plotted on Y-axis with Inter-Arrival time on X-axis, we observed it to follow exponential distribution as shown with lambda = 0.6389 as Arrival Rate

Goodness of Fit:

The Inter-Arrival distribution is fitted with Exponential curve with an exponential power of 0.6389. The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares due to Errors (SSE), (ii) R-Square, (iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE).

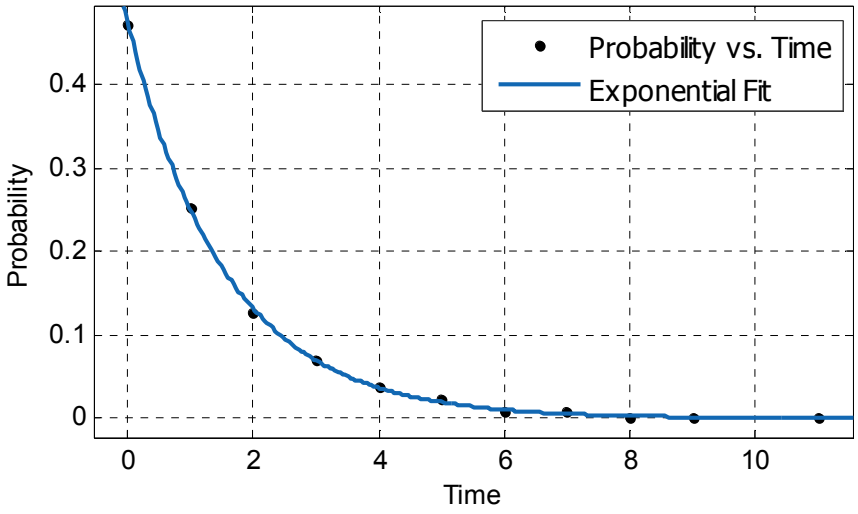

Figure 3-13 Goodness of fit for Exponential Distribution

Table 3-4 Goodness of fit for Data Set – II Exponential curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = a*exp(b*x)	a = 0.4721 (0.4655, 0.4787) b = -0.6389 (-0.6559, -0.6219)	SSE: 8.384e-05 R-square: 0.9996 Adjusted R-square: 0.9996 RMSE: 0.003052

QQ-Plot:

Q-Q Plot can help in comparing the theoretical and practical distributions. The goodness of fit can be observed in graphical form from these plots. Q-Q Plot is implemented for the theoretical exponential and above plotted Inter-Arrival distribution and it is observed to have fair goodness of fit in concurrence with the numerical goodness of fit. Average Inter-Arrival (T) is observed to be 1.56 Sec/User.

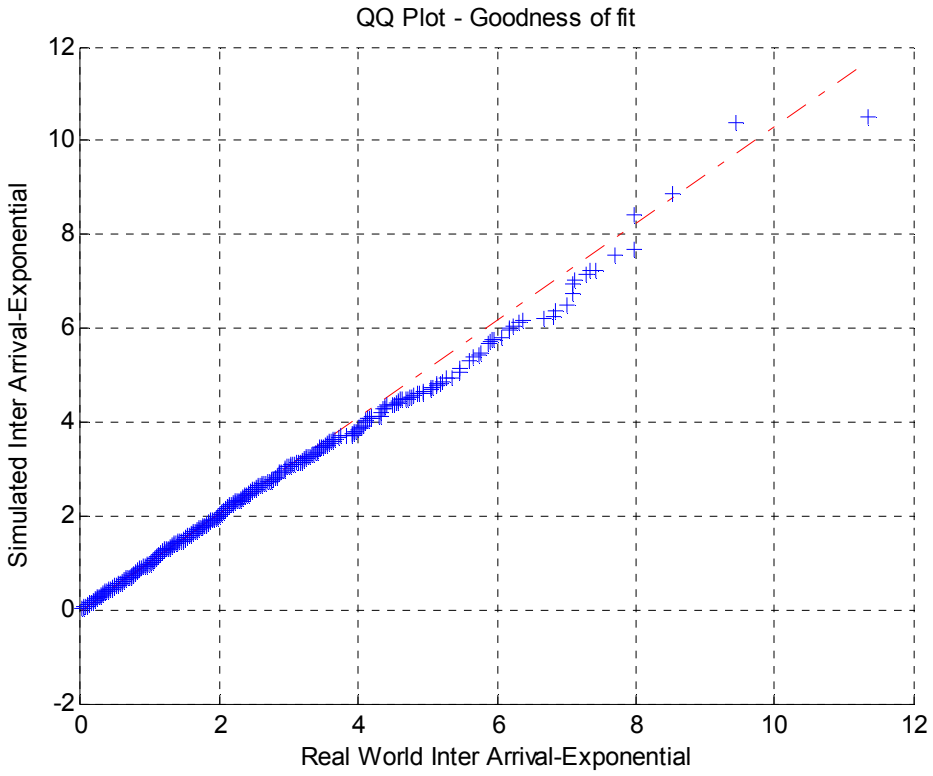

Figure 3-14 QQ-Plot for Exponential distribution Fit

3.7.2.2 Poisson Arrival

An arrival process can also be specified in counting process with the number of arrivals in the specified time duration. As the counting process can include infinite family set of random user arrivals, the count should always be an increasing function as the time progresses. For a Poisson Arrival process, if the arrivals are plotted against time, it will follow a straight line. The data set's Poisson modeling is observed to follow the same as shown below. The user arrival counting process is modeled as poisson with arrival rate (λ) of 0.6389.

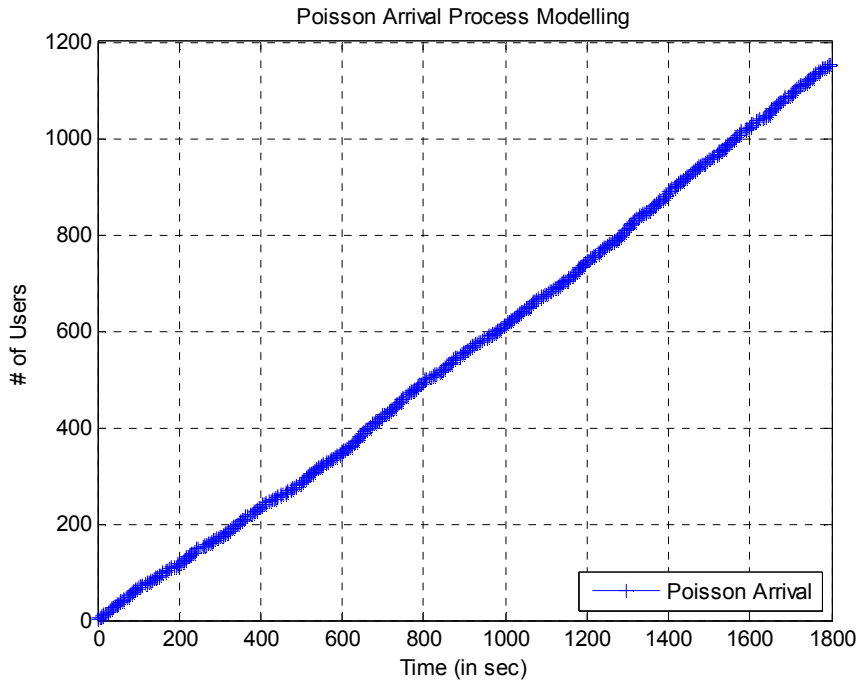

Figure 3-15 Poisson Arrival Process Modelling

Goodness of Fit:

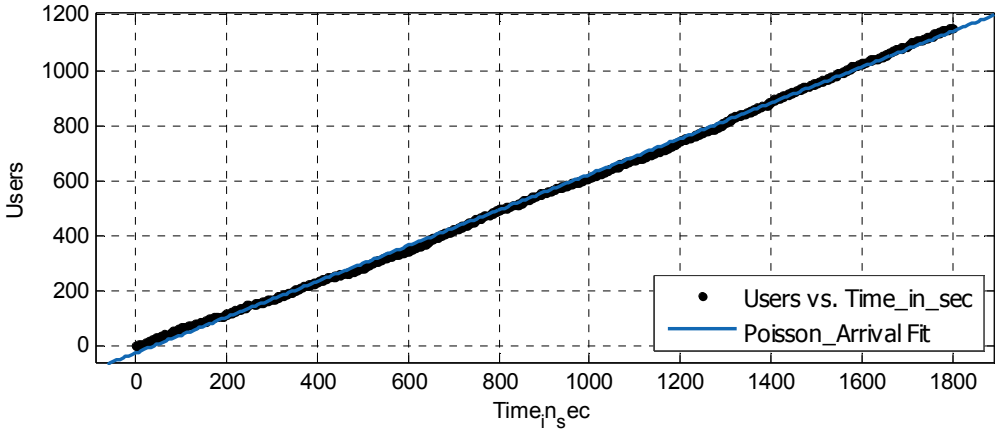

Figure 3-16 Goodness of curve fit for Poisson Arrivals

The plotted Poisson counting process is fitted to a Y=mX+C straight line with best fit slope of 0.6468 which is the Arrival Rate.

Table 3-5 Goodness of fit for Data Set – II Poisson Arrival

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = 0.6468 (0.6454, 0.6481) p2 = -22.68 (-24.09, -21.27)	SSE: 1.611e+05 R-square: 0.9987 Adjusted R-square: 0.9987 RMSE: 11.83

The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares due to Errors (SSE), (ii) R-Square, (iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE).

QQ-Plot:

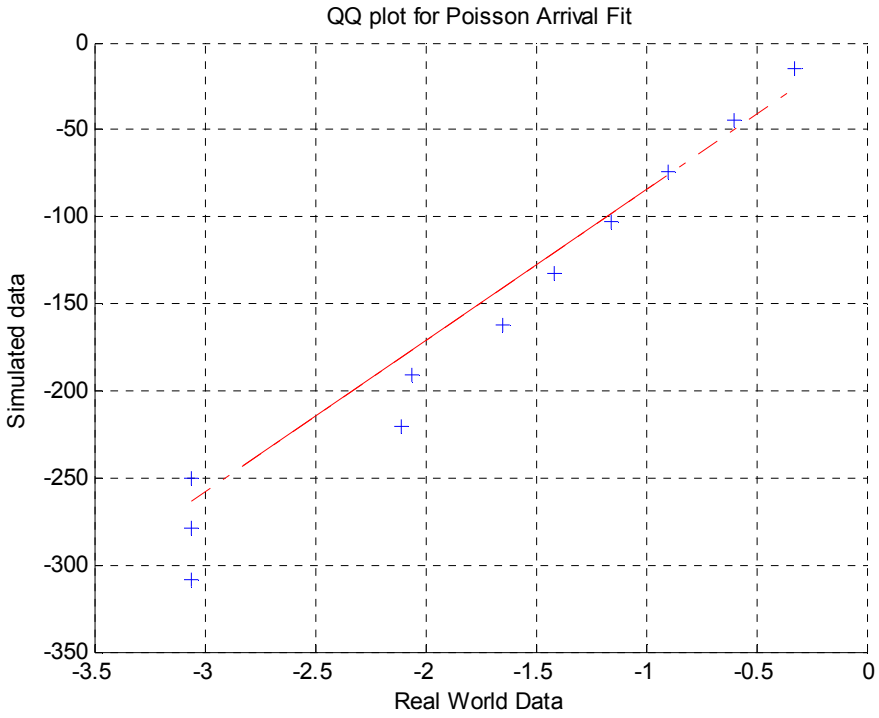

Figure 3-17 QQ Plot for Poisson Arrival Fit

Q-Q Plot is implemented for the theoretical straight line and above plotted Inter-Arrival distribution and it is observed to have fair goodness of fit in concurrence with the numerical goodness of fit. Average Inter-Arrival (T) is observed to be 1.54 Sec/User.

3.7.2.3 CDF of Instantaneous Active Incoming Users

The CDF is discontinuous at the integers of X-axis and flat everywhere else because a variable that is Poisson distributed only takes on integer values. A Poisson CDF fit is also shown in the above CDF plot for comparison [20]. The below figure is the empirical CDF of number of users arrived in per second. In the data sets, maximum number of incoming new users requesting for a service is observed to be 4 in 1 second time duration.

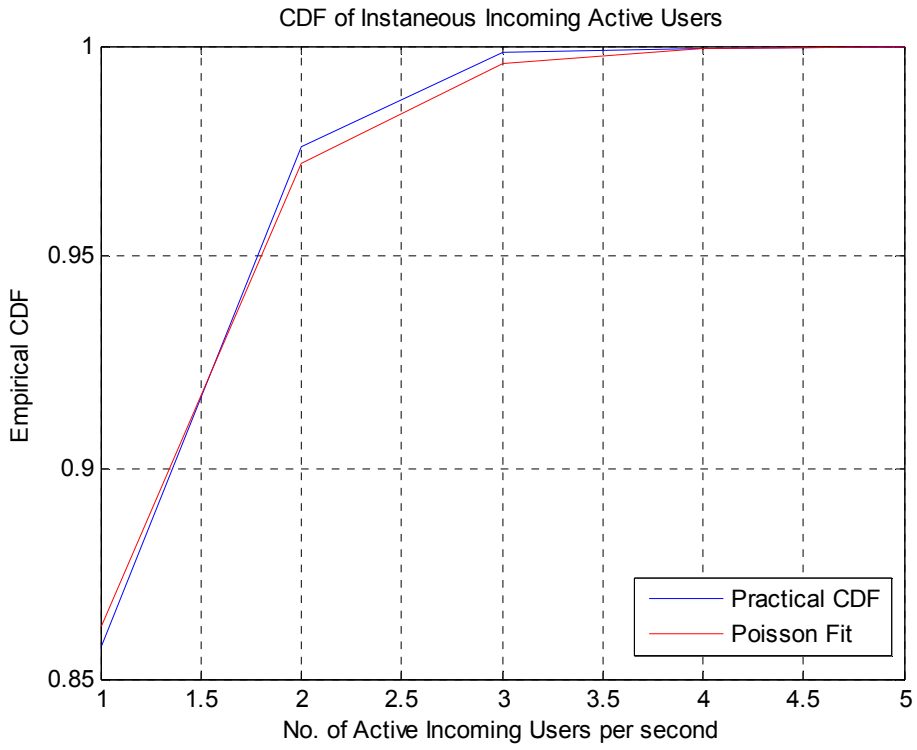

Figure 3-18 CDF of instentanous Active Users

3.7.2.4 Logarithm Plot for Exponential Inter Arrival

The Log plot of Exponential Inter-Arrival is expected to be a Straight Line. This servs as a verification of the exponential Inter-Arrival apart from the numerical goodness of fit and pictorial Q-Q Plot fit. This is observed to a close straight line but not a perfect one. The deviation from the stright line after 8seconds on x-axis can be reasoned by the type of data plotted and also by the truncation error.

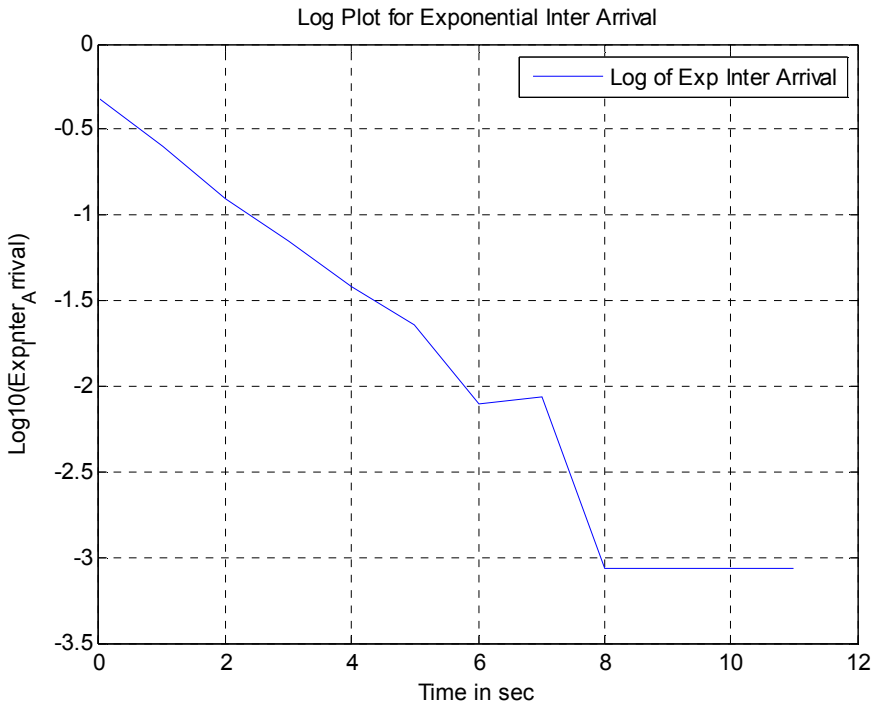

Figure 3-19 Plot of Logarithm of Exponential Inter Arrival

Goodness of Fit:

The plotted Logarithm of Exponential distribution curve is fitted with a Y=mX+C curve with best fit slope of -0.2798 and y-intercept of -0.3394. The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares

due to Errors (SSE), (ii) R-Square, (iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE).

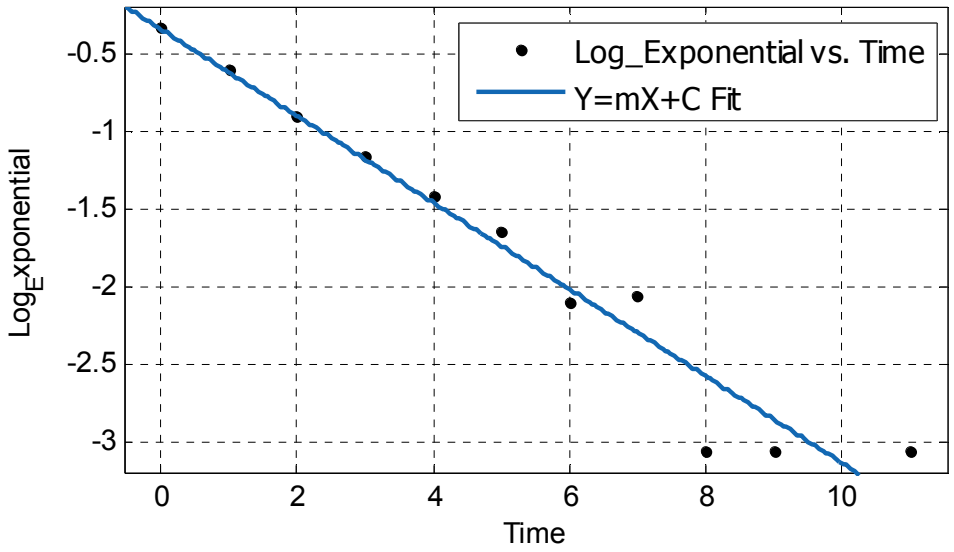

Figure 3-20 Goodness of Curve Fit for Log of Exponential Inter Arrival

Table 3-6 Goodness of fit for Data Set – II Log of Exponential

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.2798 (-0.3272, -0.2324) p2 = -0.3394 (-0.6272, -0.0515)	SSE: 0.4774 R-square: 0.952 Adjusted R-square: 0.9466 RMSE: 0.2303

The plotted Logarithm of Exponential distribution curve is fitted with a Y=mX+C curve with best fit slope of -0.2798 and y-intercept of -0.3394. This fitness fit to Logarithm of Exponential curve to straight line concludes that the Inter-Arrival distribution of a Poisson Arrival to follow Exponential distribution which is expected for a Poisson Arrival Process.

QQ-Plot:

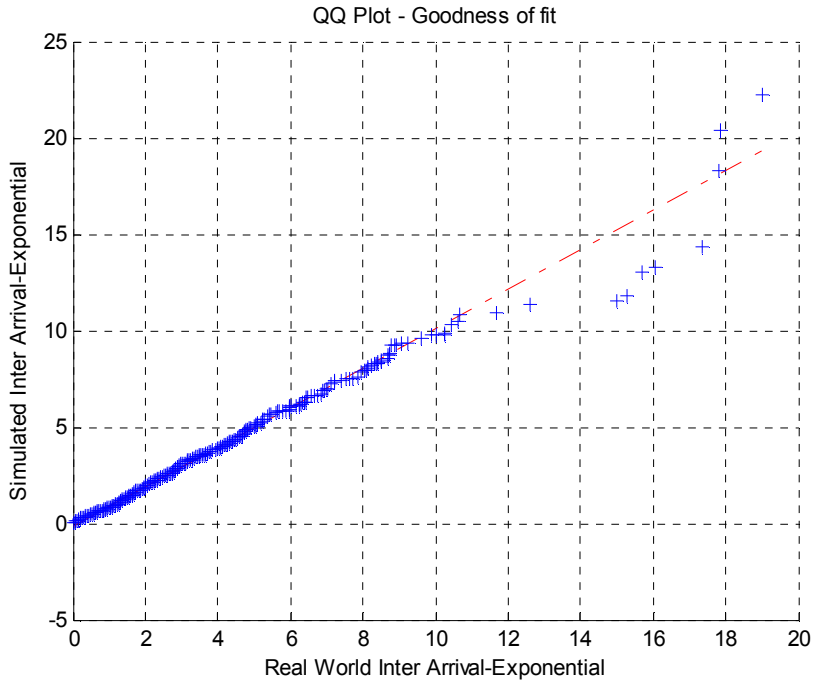

Figure 3-21 QQ-Plot for Log-Exponential Inter Arrivals

3.8 Conclusion

- User Arrival to an eNodeB requesting for a service is modeled to follow Poisson.
- Inter-Arrival time of Poisson Process is to follow Exponential distribution.
- Curve fitting is performed and Poisson Arrival rate is measured for each data set.
- Rate of arrival for Poisson process is verified with Exponential Inter-Arrivals.
- QQ-plots for real world data and simulated data are plotted.
- CDF of Instantaneous User Arrival is plotted and compared to that of theoretical values.

Chapter 4

LTE-A SELF-SIMILARITY & TRAFFIC ANALYSIS

4.1 Introduction

In order to meet the demands of mobile communication services with increasing number of users, mobile communication has developed rapidly. The next generation mobile communication systems, entitled fourth-generation mobile communication systems are currently in use. It is observed that, by 2015 the later release of LTE which is termed as LTE-Advanced will be completely available to customers. Both LTE & LTE-Advanced are of same technologies. In specific, LTE-Advanced primarily supports the LTE release 10 and ITU/IMT-Advanced. In 3GPP LTE-Advanced, there are promising technologies, like enhanced MIMO, CoMP, eICIC, CA that can provide a higher spectral efficiency and reliable transmission [16]. Carrier Aggregation (CA) has been considered to be the most important technology among other technologies, where more than one component carriers are aggregated to support a higher data rate. The CA technology increases the affinity of the operators since it is difficult to ensure a single continuous wide frequency band.

4.2 Carrier Aggregation

Carrier Aggregation is one of the key features for LTE-Advanced in Rel-10 for meeting the peak data rate requirement of IMS-Advanced, 1Gbps and 500Mbps for the downlink and uplink, respectively. It enables operators to create large 'Virtual' carrier bandwidths for LTE services by combining separate spectrum allocations. As of

December 2014, CA is the primary feature deployed by operators with commercial LTE-Advanced services. The need for CA in LTE-Advanced arises from the requirement to support bandwidths larger than those currently supported in LTE (up to 20 MHz) while at the same time ensuring backward compatibility with LTE [18]. CA is supported in uplink & downlink in both Time-division-duplex (TDD) LTE and Frequency-division-duplex (FDD) LTE systems. This work is mainly focused on FDD LTE implementation of CA.

In non-CA Scenario, A UE is served by an eNodeB with single Down-link Carrier and Single Uplink Carrier. It has one physical Cell Area. The down-link usually serves GBR and non-GBR data.

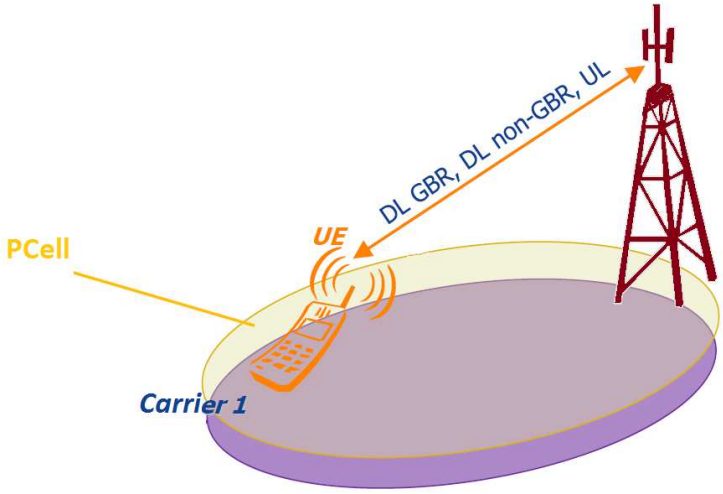

Figure 4-1 General non-CA scenario

In CA Scenario, the Down-link is supported by one more additional Down-link Carrier to send all the Non-GBR Data. The additional Downlink carrier is dynamic and can be activated/de-activated based on the requirement as represented below [22].

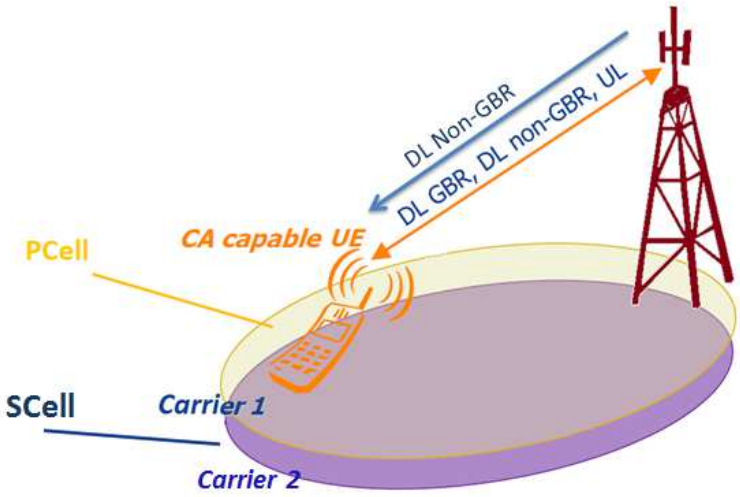

Figure 4-2 Typical CA Scenario

Each aggregated carrier is referred to as component carrier, CC. Each component carrier corresponds to a serving cell. The different serving cells may have different coverage. Primary Cell corresponds to Primary Component Carrier and Secondary Cell corresponds to Secondary Component Carrier.

4.3 Carrier Aggregation Key Features

- The component carrier can have a bandwidth that supports aggregation of 1.4, 3, 5, 10, 15, 20 MHz and a maximum of five component carriers can be aggregated and hence the maximum aggregated bandwidth can be 100MHz.
- The number of component carriers is always equal to or lower than the number of downlink component carriers. The individual component carriers can also be of different bandwidths.
- Primary cell supports both Uplink and Downlink

- Secondary Cell always supports Downlink (while CA is activated) but may or may not supported in Uplink
- Primary Cell is always activated whereas Secondary Cell can be activated and de-activated based on the requirement.
- In Primary Cell, control and shared physical channels (PUCCH and PUSCH) are available in both uplink and downlink
- In Secondary Cell, Physical downlink control channel is optional in downlink and Physical shared channel is mandatory in both uplink and downlink [22].

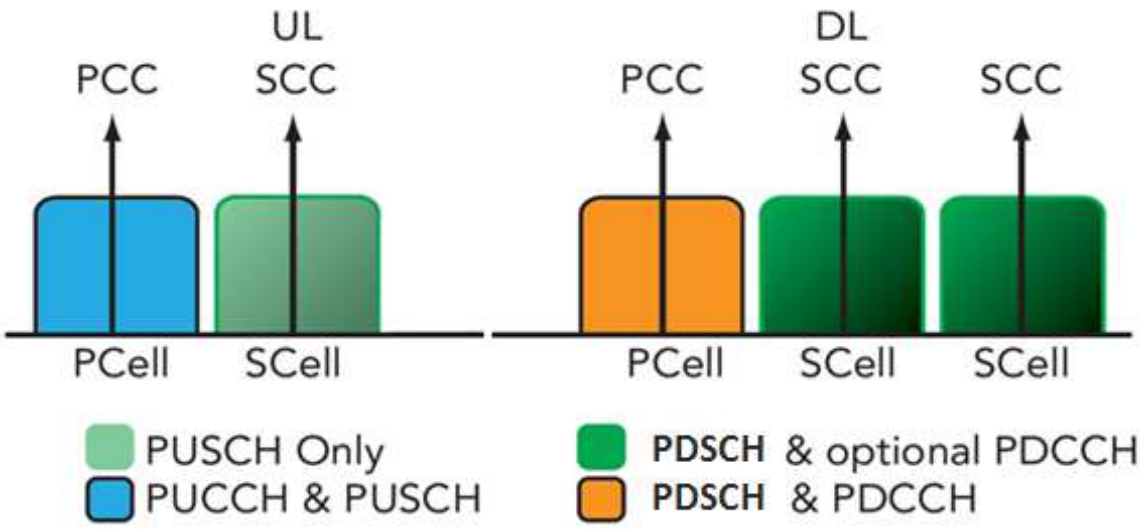

Figure 4-3 Physical channel allocations in Primary and Secondary Cells

- UE doesn't acquire any system information from secondary cell. Only non-GBR data is send/received using Secondary Component Carriers.
- CA supports cross component carrier scheduling, where the control channel at one carrier can be used to allocate resources at another carrier for user data transmission.

- Carrier Aggregation can be implemented in both continuous and non-continuous bands based on the spectrum availability [22].
    - For Contiguous component carriers aggregation,

      Intra-Band Contiguous Type: The aggregated carriers are contiguous and within the same operating frequency bands.
    - For Non-Contiguous component carriers aggregation,

      Intra-band Non-Contiguous Type: The aggregation carriers are within the same operating frequency band but are non-contiguous.

      Inter-band Non-Contiguous Type: The aggregation carriers are in different operating frequency bands and are non-contiguous.

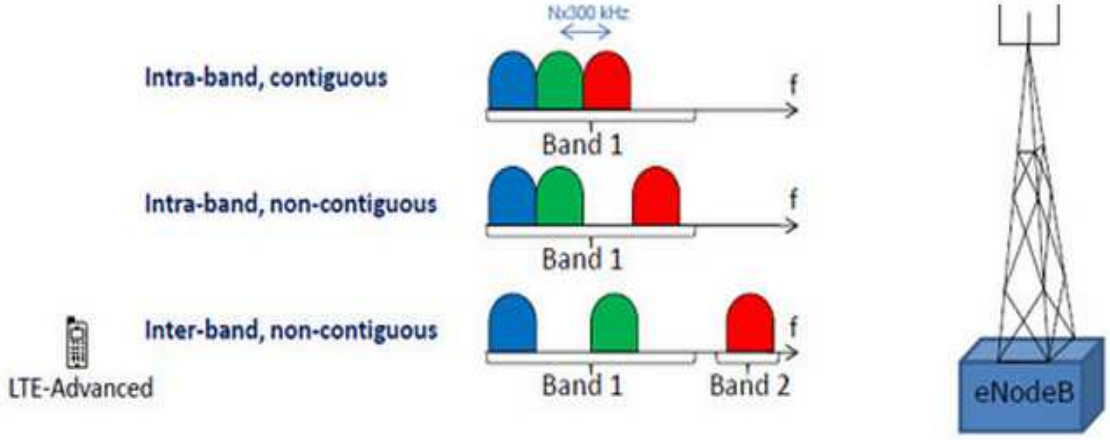

Figure 4-4 Different CA configurations

- CA is designed to be backward compatible, meaning that legacy Rel-8 and Rel-9 users should still be able to coexist with LTE-Advanced on at least part of the total bandwidth. Thus, each individual spectrum chunk, denoted component carrier (CC), inherits the core physical layer design and numerology from LTE Rel-8.

- CA naming convention can be understood as shown below as per [22].

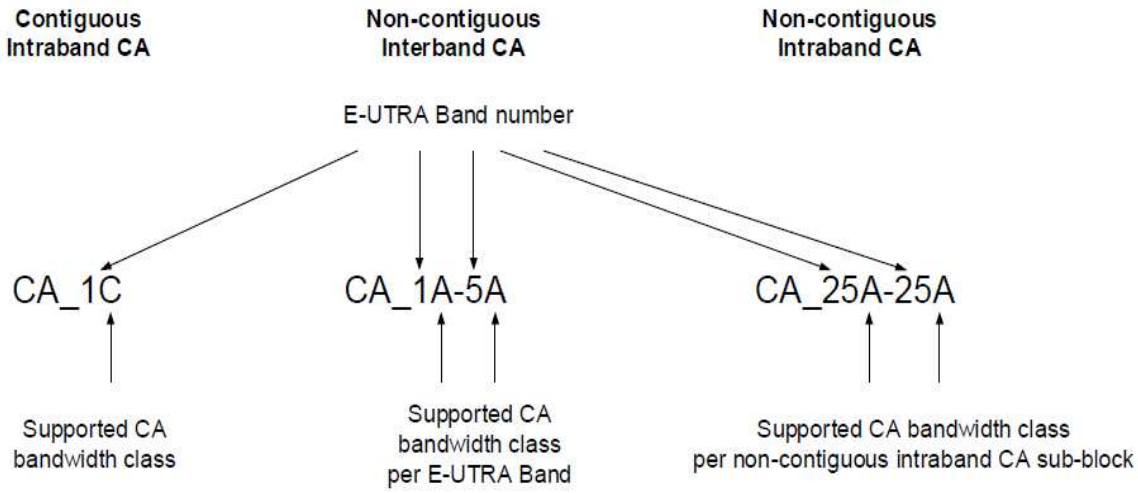

Figure 4-5 CA naming Convention

4.4 3GPP Specifications for CA

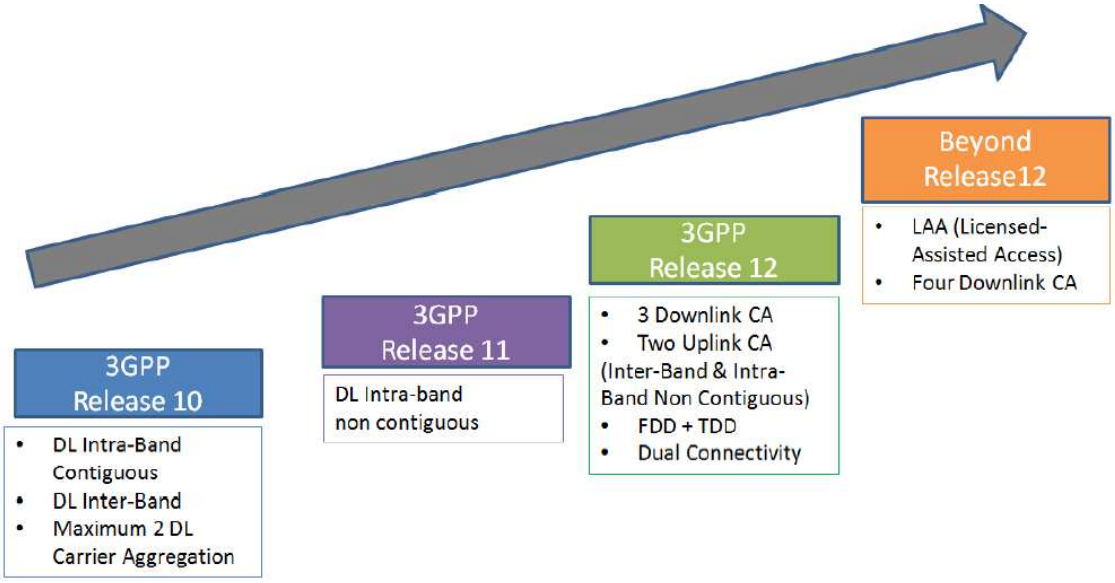

Figure 4-6 3GPP releases for CA

3GPP Rel-10 is the first release to support CA. It defines a limited variety of CA configurations, including: contiguous intra-band CA for bands 1 (FDD) and band 40

(TDD) named CA_1C and CA_40C respectively and an inter-band CA between bands 1 and 5, named CA_1A-5A.

3GPP Rel-11 offers many more CA configurations, including non-contiguous intra-band CA and Band 29 for inter-band CA, which is also referred as supplemental DL. Band 29 is a special FDD band in Rel-11 that has only a DL component and no UL component. The intention is that this band is available for bandwidth expansion only and holds the functionality of DL Secondary cell (SCell). This serving cell is referred to as the primary cell (PCell). In the downlink, the carrier corresponding to the PCell is the Downlink Primary Component Carrier (DL-PCC), while in the uplink it is the Uplink Primary Component Carrier (UL-PCC). Other serving cells are referred to as secondary cells (SCells) and are used for bandwidth expansion for the particular UE. It is to be noted that there are additional differences in the functionality.

Release 12 will include CA of FDD and TDD frequency bands, as well as support for aggregating two UL CCs and three DL CCs. Note that even though certain CA combinations are specified in later releases, these are release independent and can be supported from previous releases' equipment due to backward-compatible signaling. 3GPP Rel-12 offers aggregation of three DL CCs, support for CA configurations for UL inter-band and intra-band non-contiguous CA and band 32 support (1.5GHz L Band), which has a DL component only and further enhanced capabilities that are still under study.

Carrier Aggregation has much more to offer and it continues to be a significant area of work for 3GPP, equipment manufacturers and network operators. Over the coming years there will be a number of important developments, including:

- Increasing the number of component carriers and the total bandwidth supported in both the downlink and the uplink.
- Supporting a greater number of frequency bands and combinations of frequency bands.
- Using Carrier Aggregation between cells to enhance the support of small cells and Heterogeneous Networks (HetNets).
- Enabling flexible aggregation of FDD and TDD LTE carriers.
- Supporting LTE Carrier Aggregation between licensed and unlicensed spectrum.

4.5 Carrier Aggregation Performance & Benefits

CA is an effective tool to combine together fragmented operator spectrum and deliver higher throughputs with a bigger spectrum pipe. However, CA has several additional benefits:

- *Higher speeds:* Aggregation of carriers increases spectrum resources, which provides higher speeds across the cell coverage.
- *Capacity gain:* Aggregating multiple carriers increases spectrum but also includes trunking gains from dynamically scheduling traffic across the entire spectrum. This in turn increases cell capacity and network efficiency and improves the experience for

all users. A user previously facing congestion on one carrier can be seamlessly scheduled on a carrier with more capacity and maintain a consistent experience.

- *Optimum utilization of an operator's spectrum resources:* The majority of operators have fragmented spectrum covering different bands and bandwidths, Carrier Aggregation helps combine these into more valuable spectrum resource.
- *Aggregation of spectrum resources in HetNet or LTE-Unlicensed scenarios generates new spectrum sources for operators:* Licensed spectrum is in perpetually short supply, and even when it is available, it is sometimes an expensive investment. CA enables operators to maximize the return on those investments, as well as use unlicensed spectrum to supplement their portfolios.
- The CA feature increases the bandwidth for a CA-capable UE by aggregating several LTE carriers, thereby increasing the UE's bit rate. Each of the aggregated carriers is referred to as a CC. In order to keep backward compatibility, the aggregation is based on Rel-8/Rel-9 carrier structure and so legacy UEs should still be able to coexist with LTE-Advanced on at least part of the total bandwidth. Therefore, the component carrier can have a bandwidth of 1.4, 3, 5, 10, 15 or 20 MHz
- Even though LTE Rel-8 can support bandwidths up to 20 MHz, most American wireless operators do not have that much contiguous spectrum. In spectrum below 2 GHz, most operators have 5-15 MHz of contiguous spectrum in a single frequency band. Also, many operators own the rights to use spectrum in many different bands. So from a practical perspective, CA offers operators a path to combine spectrum

assets within the bands they operate in and to combine assets across multiple frequency bands.

4.6 Impact of CA on Protocol layer level

Implementation of CA led to some changes for LTE Protocol Architecture at layer levels.

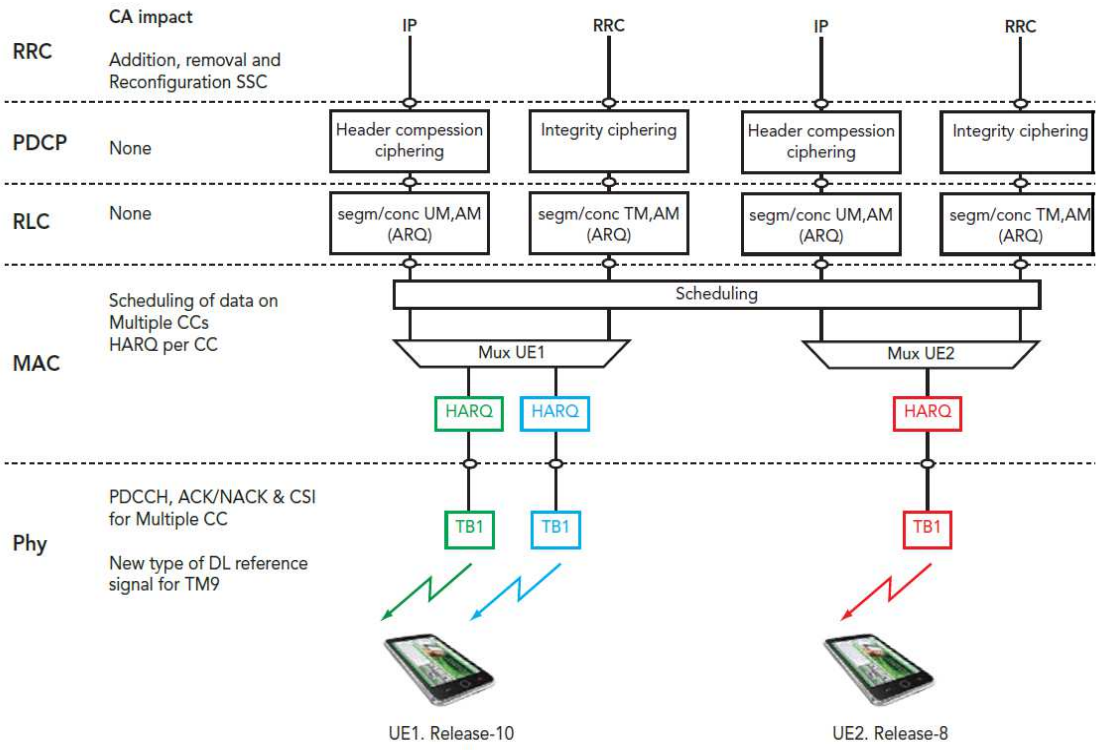

Figure 4-7 LTE Architecture with changes due to CA

- In Physical Layer, number of Transport Blocks is increased corresponding to the number of simultaneous Down-link Component Carriers. Separate ACK/NACK is implemented for each of those Transport Blocks.

- In Mac Layer, Scheduling of Multiple Component Carriers and Hybrid ARQ per Component Carriers are implemented.
- There is no change in RLC and PDCP layers
- At RRC Layer, Addition/Removal and Reconfiguration of Secondary Cells is taken care.

4.7 Data Collection & Constraints

Data used for this piece of work is collected from a US based cellular operator's live LTE-A network in Kansas City, Missouri. The data is captured at eNodeB level which contributes to S1/X2 interfaces in LTE system architecture. There is no separation of data sets based on cell sectors. The eNodeBs from which the data sets are collected typically has 3 cell sectors which are of least relevance to this piece of work. All the eNodeBs where data is collected are implemented with band 2 and 4 carrier aggregation with Band 2 (PCS – 1900MHz) serving as Primary Component Carrier and Band 4 (AWS – 2100MHz) serving as Secondary Component Carrier. Band 2 has 10MHz bandwidth and band 4 with 5MHz bandwidth.

Data Traffic logs collected for Self-Similarity analysis are similar to that of the data collection described in chapter-2 for LTE data traffic. User and Control plane Data traffic is collected from eNodeB level with LTE-Advanced Carrier Aggregation feature is turned ON with 2 band aggregation. All data collection is about 10 minute's duration with minimum of 60000 entries in .pcap format. The 3 data sets are collected at 3 different times of a day similar to that of LTE data collection for reference.

Data sets collected for the analysis of LTE-Advanced feature in live network are collected form 4 different eNodeB's for a period of 1 week (5 weekdays) with hour granularity. Data pertaining to Primary and Secondary Cells can be distinguished from the data sets. Secondary Component Carrier is considered to carry only non-GBR data and Primary component carrier to carry both GBR and non-GBR data.

Constraints:

- With the limited implementation of carrier aggregation in the live network, Only 4 enodeB data is considered. With the collected data sets, only hour granularity is observed.
- Since the penetration and availability of CA is very limited, all the collected data represented less frequent numbers depicting the reality.

4.8 Simulation Results

4.8.1 LTE-Advanced Self-Similarity Results

4.8.1.1 V-T Plot for Data Set - I

This data set represents the LTE-Advanced data traffic collected at around 6PM of a normal weekday. A traffic load of size 1.32GB is captured for 10minutes duration with average data rates of 15Mbps with about 1987 active incoming users. Each packet has an average size of about 800~850 Bytes. A sample traffic of about 20s duration is shown in below figure with 20ms granularity. Collected data's sample of 20 seconds is plotted below. The Y-axis represents the traffic size in Bytes per time resolution.

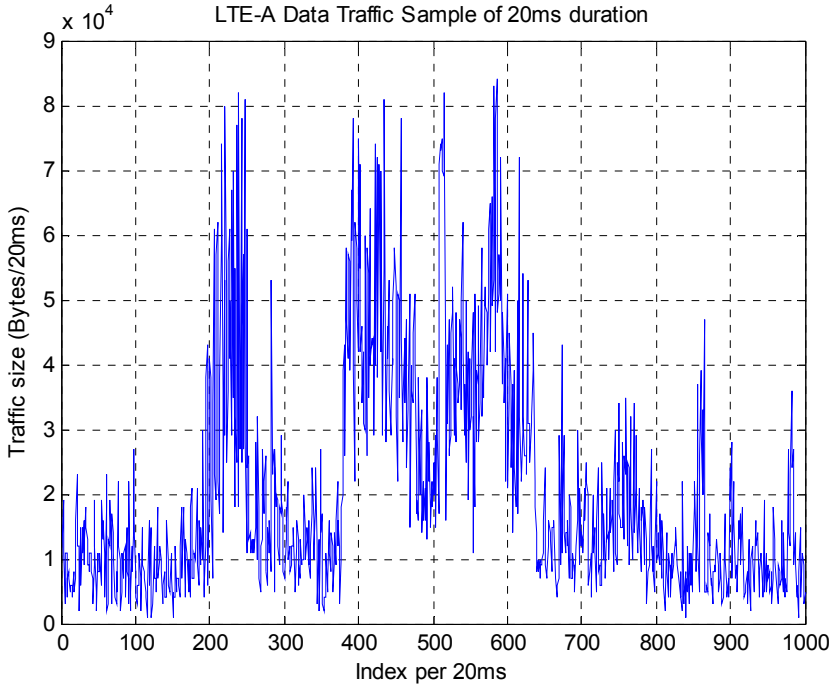

Figure 4-8 Sample data set – I plot for 20ms resolution

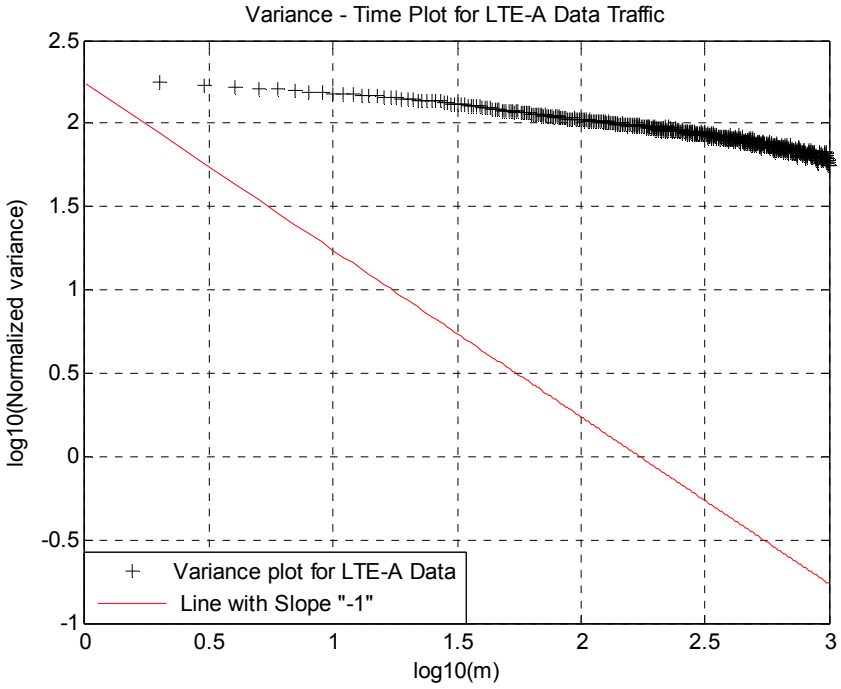

Figure 4-9 V-T Plot for data set -I

From the data set –I, V-T Plot is plotted to evaluate the Self-Similarity property. V-T Plot concludes that this particular data traffic exhibits Self-Similarity property as the slope of V-T plot is observed to be greater than -1. The Degree of Self-Similarity (H – Parameter) is obtained to be 0.9361. The goodness of fit for the slope of this V-T curve is observed to be -0.2114.

Goodness of Fit:

The plotted V-T curve is fitted with a Y=mX+C curve with best fit slope of -0.2114 and y-intercept of 2.445. The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares due to Errors (SSE), (ii) R-Square, (iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE). The above V-T plot is fitted with a straight line with slope -0.2114.

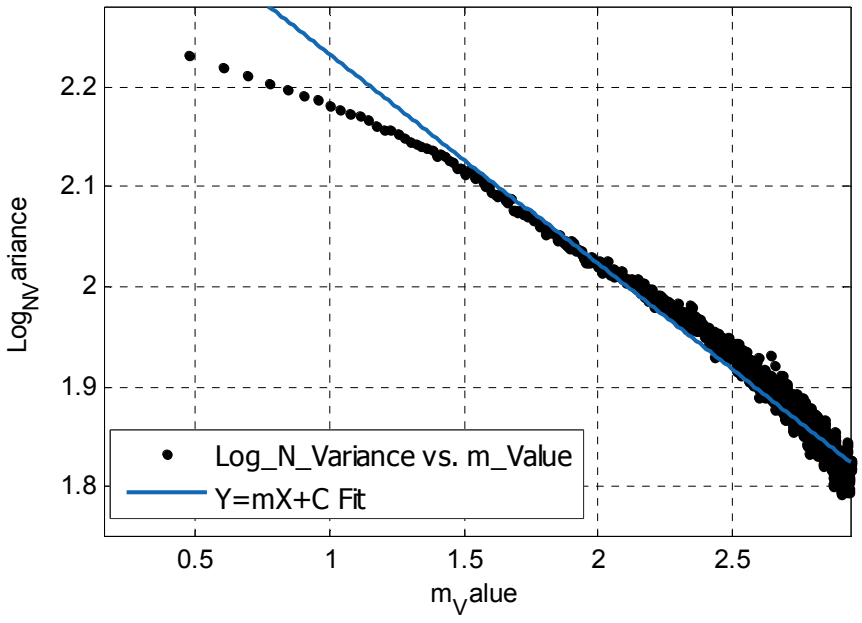

Figure 4-10 Curve Fitting for V-T Plot

Table 4-1 Goodness of fit for Data Set – I V-T curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.2114 (-0.2142, -0.2086) p2 = 2.445 (2.438, 2.453)	SSE: 0.3408 R-square: 0.9574 Adjusted R-square: 0.9573 RMSE: 0.0185

The variation of H-parameter for values m-valued aggregation levels can be observed in the above figure. The variation in H-parameter varies constantly over all aggregation levels holding the fact that Self-Similarity is all about having the same characteristics at different aggregated levels.

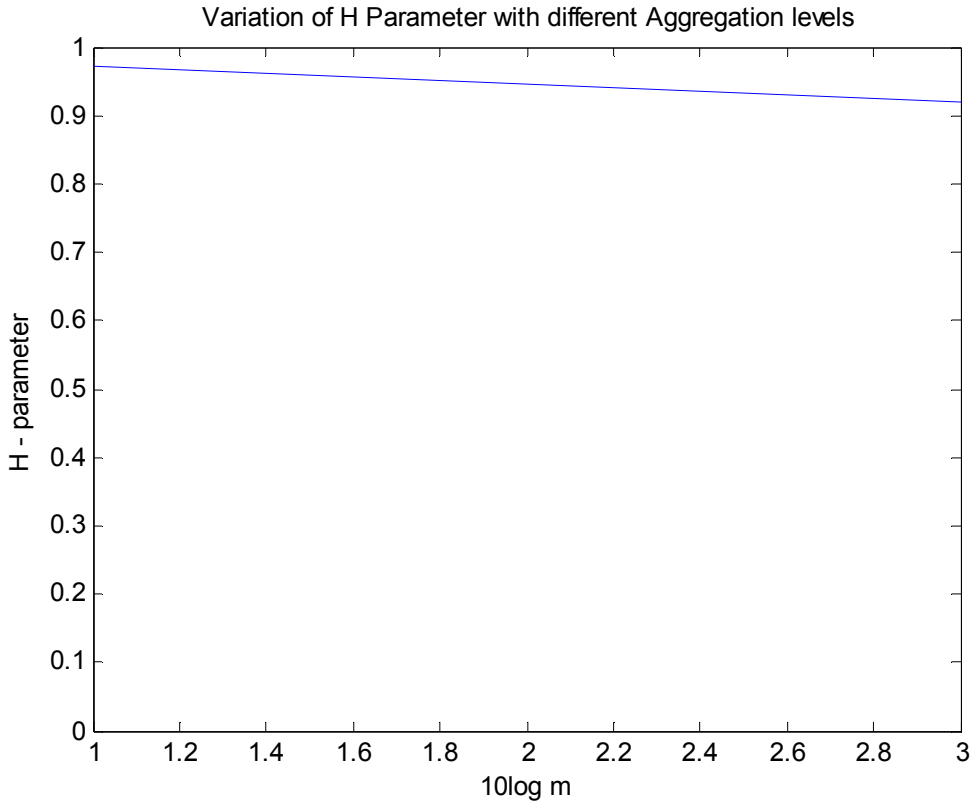

Figure 4-11 Variation of H-Parameter for different m values

4.8.1.2 V-T Plot for Data Set – II

This data set represents the LTE-Advanced data traffic collected at around 11AM of a normal weekday. A traffic load of size 1.07GB is captured for 10minutes duration with average data rates of 13Mbps with about 1894 active incoming users. Each packet has an average size of about 800~850 Bytes. A sample traffic of about 20s duration is shown in below figure with 20ms granularity. Collected data's sample of 20 seconds is plotted below. The Y-axis represents the traffic size in Bytes per time resolution. Collected data's sample of 20 seconds is plotted as below.

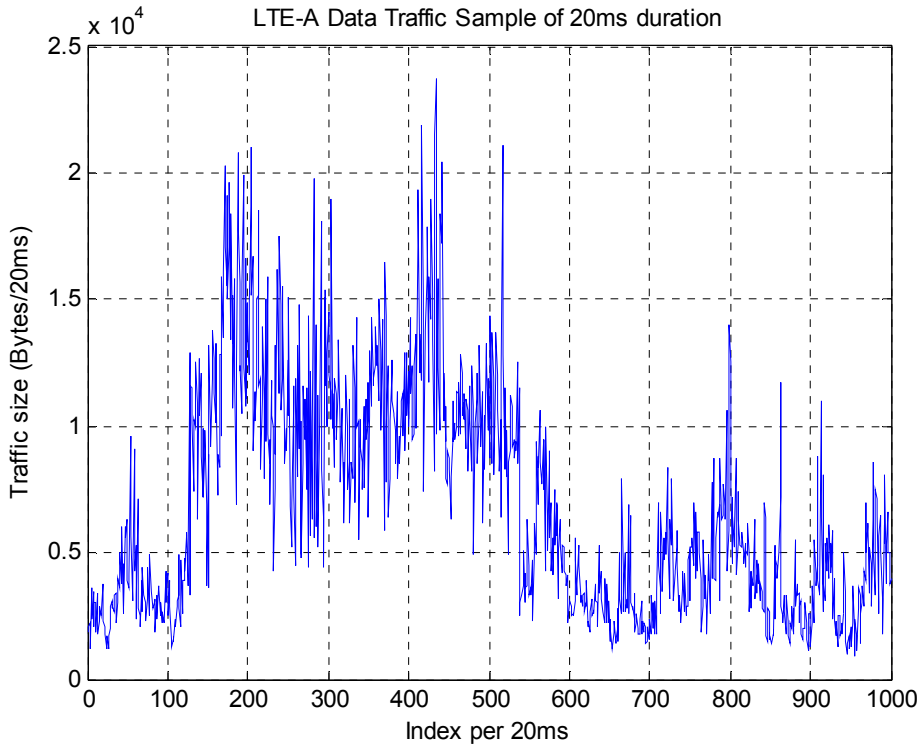

Figure 4-12 Sample LTE-Advanced data traffic for data set – II

From the data set –I, V-T Plot is plotted as below to evaluate the Self-Similarity property.

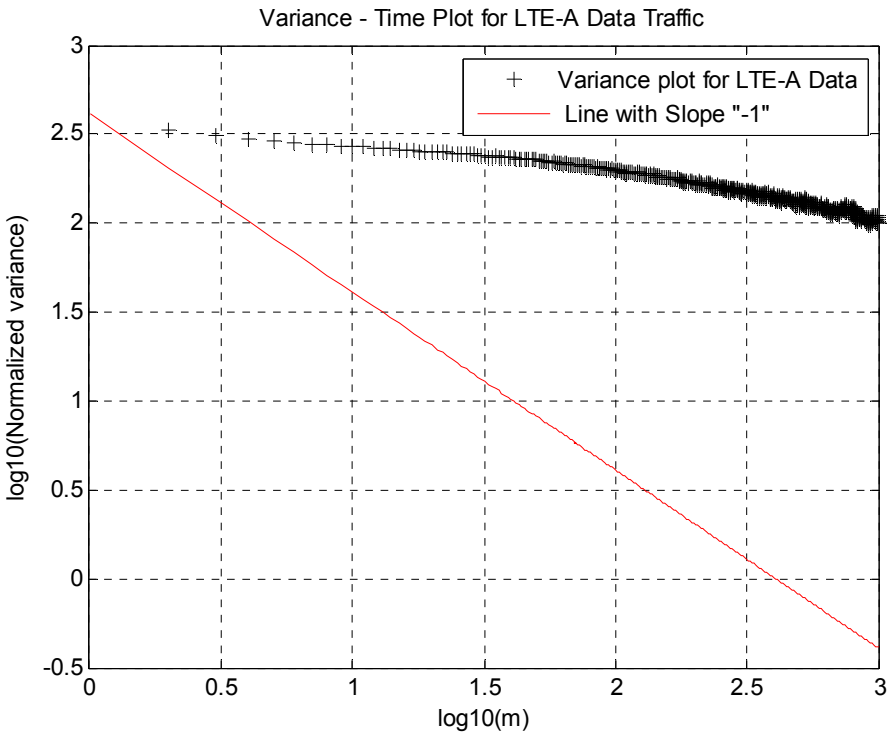

Figure 4-13 V-T Plot for Data Set -II

For the above shown data set-II, V-T graph is plotted to evaluate Self-Similarity property. V-T Plot is observed to have a slope greater than -1. So this particular data traffic is concluded to exhibit Self-Similarity property. The Degree of Self-Similarity (H – Parameter) is obtained to be 0.9095. The goodness of fit for the slope of this V-T curve is observed to be -0.2407.

Goodness of Fit:

The plotted V-T curve is fitted with a Y=mX+C curve with best fit slope of -0.2407 and y-intercept of 2.762. The deviation of difference between the predicted and observed values is measured by (i) Sum of Squares due to Errors (SSE), (ii) R-Square,

(iii) Adjusted R-Square (iv) Root Mean Square Error (RMSE). The above V-T plot is fitted with a straight line with slope -0.2407.

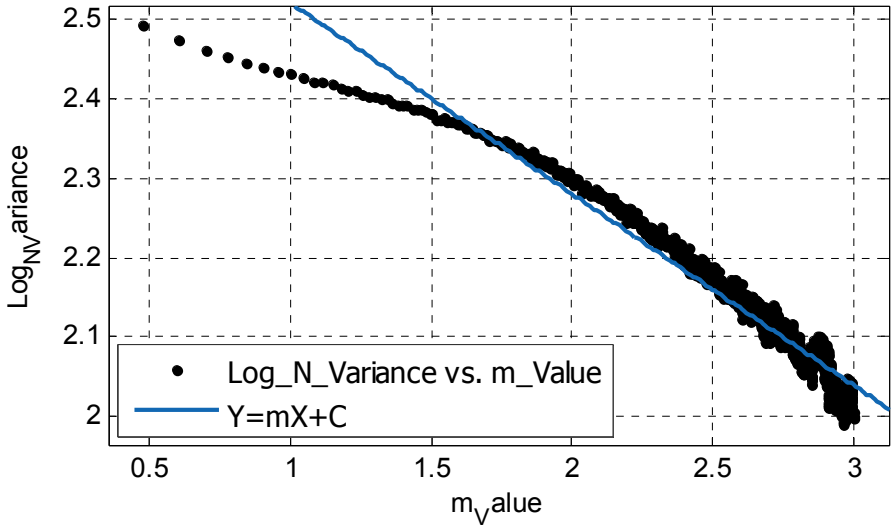

Figure 4-14 Goodness of fit for V-T curve with slope -0.2407

Table 4-2 Goodness of fit for Data Set – II V-T curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.2407 (-0.2441, -0.2373) p2 = 2.762 (2.753, 2.771)	SSE: 0.5068 R-square: 0.9514 Adjusted R-square: 0.9514 RMSE: 0.02256

The variation of H-parameter for values m-valued aggregation levels can be observed in the above figure. The variation in H-parameter varies constantly over all aggregation levels holding the fact that Self-Similarity is all about having the same characteristics at different aggregated levels.

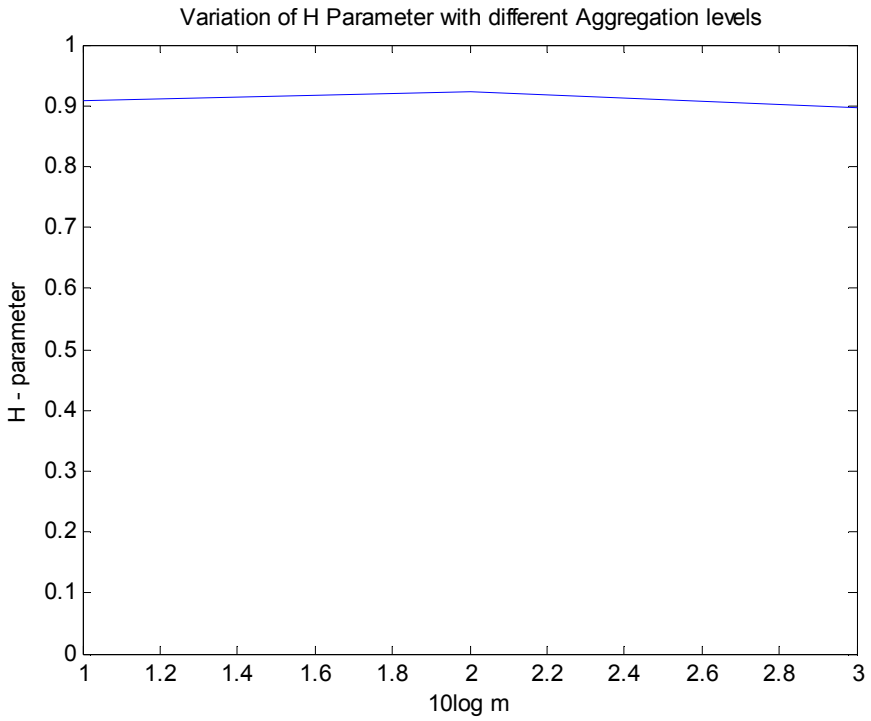

Figure 4-15 Variation of H-Parameter for different m values

4.8.1.3 V-T Plot for Data Set - III

This data set represents the LTE-Advanced data traffic collected at around 4AM of a normal weekday. A traffic load of size 0.41GB is captured for 10minutes duration with average data rates of 15Kbps with about 30 active incoming users. Each packet has an average size of about 800~850 Bytes. A sample traffic of about 20s duration is shown in below figure with 20ms granularity. Collected data's sample of 20 seconds is plotted below. Collected data's sample of 20 seconds is plotted as below.

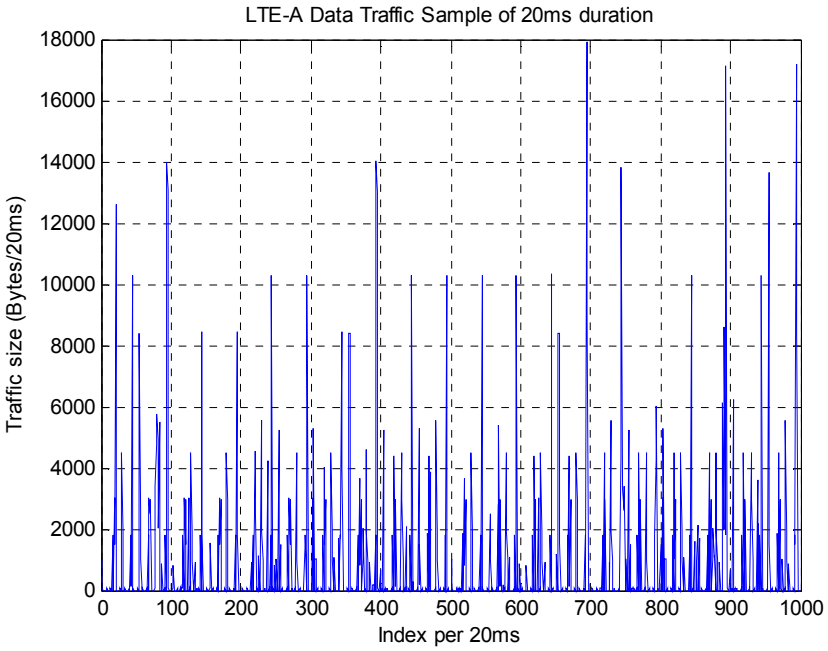

Figure 4-16 Sample data plot for Data Set – III for 20 seconds

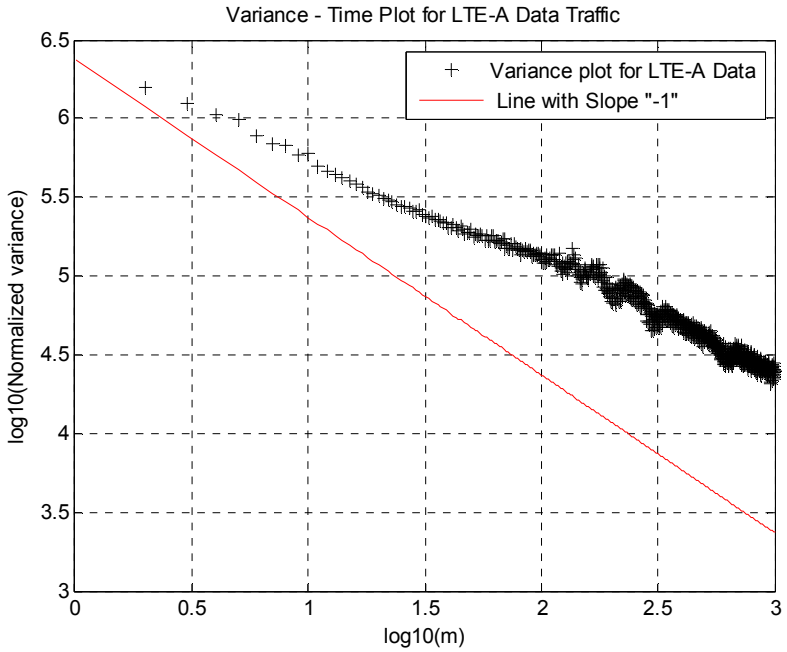

Figure 4-17 V-T Plot for Data Set - III

For this shown data set, V-T graph is plotted to evaluate Self-Similarity property. V-T Plot for data set -1 is collected at around 6PM on a normal weekday and observed to have a slope greater than -1. So this particular data traffic is concluded to exhibit Self-Similarity property. The Degree of Self-Similarity (H – Parameter) is obtained to be 0.6761. The goodness of fit for the slope of this V-T curve is observed to be -0.7049.

Goodness of Fit:

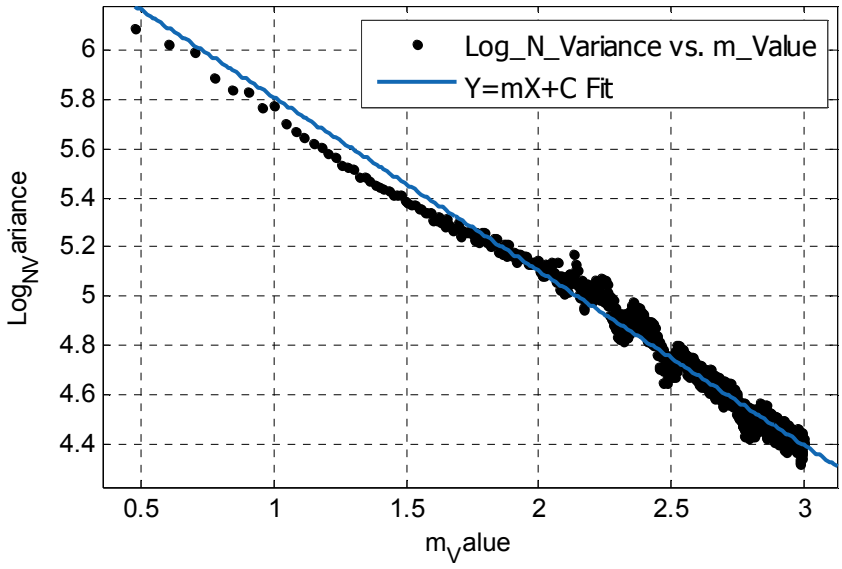

Figure 4-18 Goodness of fit for Data Set – III V-T Plot

Table 4-3 Goodness of fit for Data Set – III V-T curve

Equation	Coefficients (with95% Fit)	Goodness of Fit
f(x) = p1*x + p2	p1 = -0.7049 (-0.7118, -0.6979) p2 = 6.513 (6.495, 6.531)	SSE: 2.125 R-square: 0.9756 Adjusted R-square: 0.9756 RMSE: 0.04619

The plotted V-T curve is fitted with a Y=mX+C curve with best fit slope of -0.7049 and y-intercept of 6.513.This goodness of fit also concludes that the V-T plot is a straight line with negative slope greater than -1. So this data can be said to exhibit Self-Similarity.

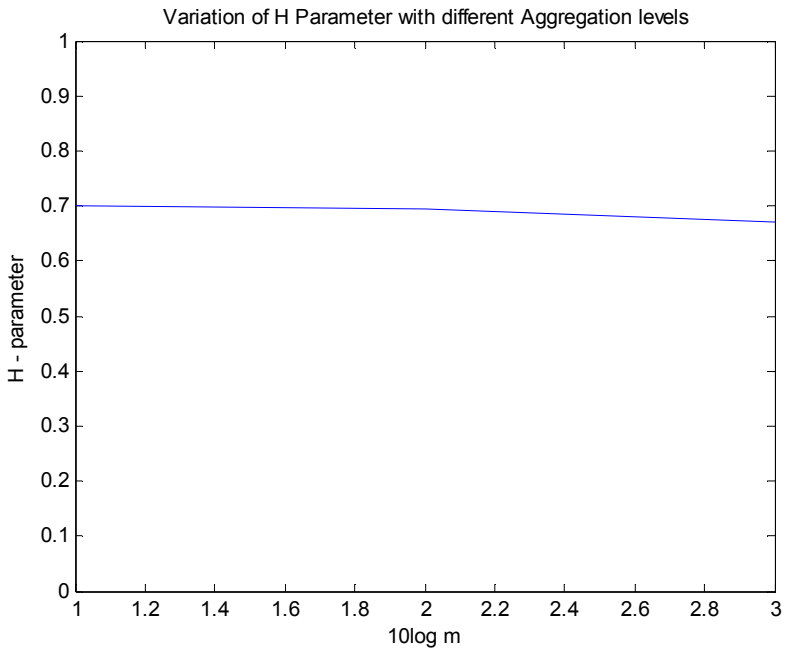

Figure 4-19 Variation of H-Parameter with different aggregation levels

The variation in H-parameter varies constantly over all aggregation levels holding the fact that Self-Similarity is all about having the same characteristics at different aggregated levels.

4.8.1.4 Comparison of H-Parameter Variations

H-Parameter of all the data sets are compared to each other over different aggregation levels as shown below. This shows that the variation of H-Parameter value is more or less constant over the entire aggregation levels following the Self-Similarity

definition. All the H-Parameter variations for 3 data sets vary constantly without any overlapping's which clearly demonstrates the difference in the degree of Self-Similarity for all data sets.

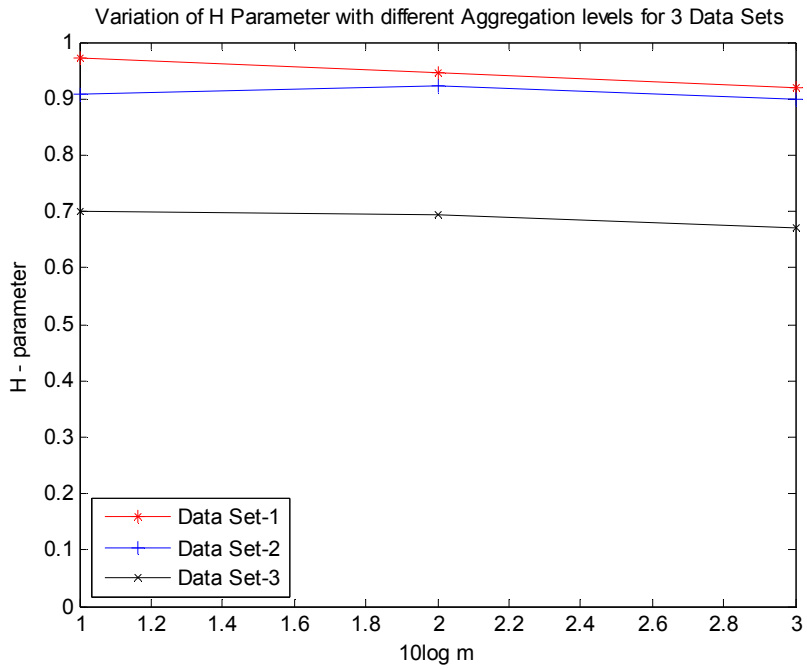

Figure 4-20 Comparison of H values for different Data Sets

Table 4-4 Comparison of H-Parameters for different LTE-A data sets

Data Sets	**Time of the Day**	**Traffic Load**	**Slope**	**H-Value**
I	18.00	1.32GB	-0.2114	0.9361
II	11.00	1.07GB	-0.2407	0.9095
III	04.00	0.41GB	-0.7049	0.6761

Above table concludes that as the traffic load increases, H-Parameter also increases along with the slope of V-T curve. The degree of Self-Similarity is varies in accordance to the traffic load and the time of the day. Higher the traffic load, higher is the degree of Self-Similarity.

4.8.2 Comparison of LTE and LTE-Advanced Self-Similarity

From the Self-Similarity analysis of LTE and LTE-Advanced data traffic, a comparison between these two technologies can give us some more insight on this property. Since all the 3 data sets correspond to 3 data sets in other technologies, we can compare them accordingly under the assumption that sites with these technologies experiences same traffic load as per the usual 24-hour pattern.

Table 4-5 Comparison of H-Parameters for LTE & LTE-A data sets

Data Set	**Time of day**	**LTE H-Value** (10MHz BW)	**LTE-A H-Value** (10MHz+5MHz BW)
1	18.00	0.9116	0.9361
2	11.00	0.8873	0.9095
3	04.00	0.6609	0.6761

In concurrence to the comparison, the duration of each data is set is almost the same with varying traffic load and number of users. The main difference of bandwidth can be observed between two technologies as LTE has only 10MHz bandwidth and LTE-Advanced Carrier Aggregation has aggregation of 10MHz and 5MHz making it 15MHz

as a whole. Table 4-2 concludes that under similar conditions and only with bandwidth increase, LTE-Advanced shows more degree of Self-Similarity when compared to that of LTE. From the previous comparison of Self-Similarity between LTE and Ethernet networks, it is evident that LTE-A exhibits highest degree of Self-Similarity when compared to that of LTE and Ethernet networks.

4.8.3 LTE-A Carrier Aggregation Performance Analysis

4.8.3.1 Trending of Users vs DL Throughput

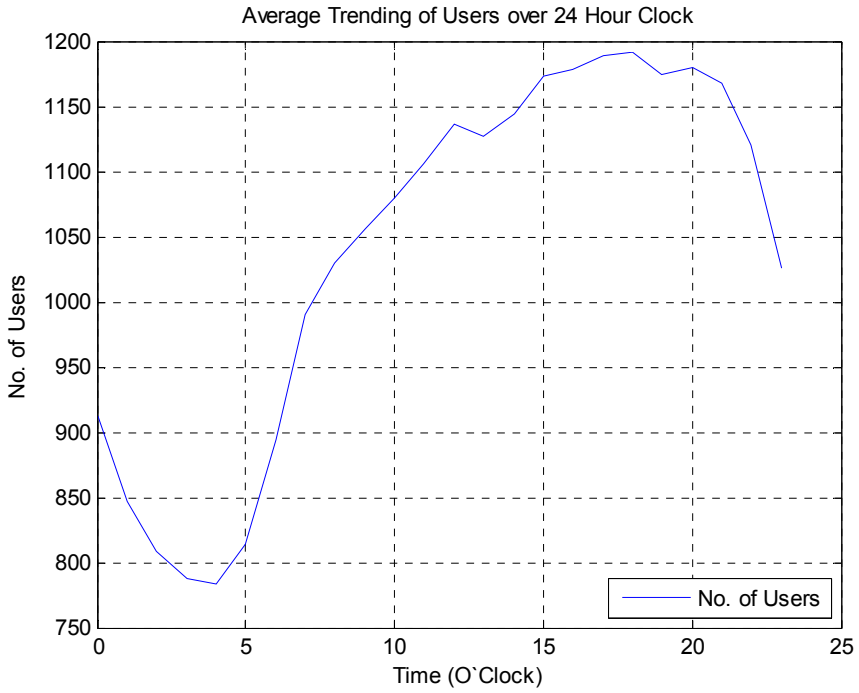

Figure 4-21 Trending of Users on 24 hour Clock in LTE

The trending of real world traffic is observed over 24 hour's average. The above figure is based on one normal load eNodeB data which is an average of Users and DL throughput on hourly basis on all weekdays in October'14. This trend is similar to what

has been observed in UTRAN network in [7]. Trend observed in [2] and [3] are considered as base for this analysis. Comparing 3G data in [2] shows that the average throughput in LTE much higher when compared to other earlier technologies.

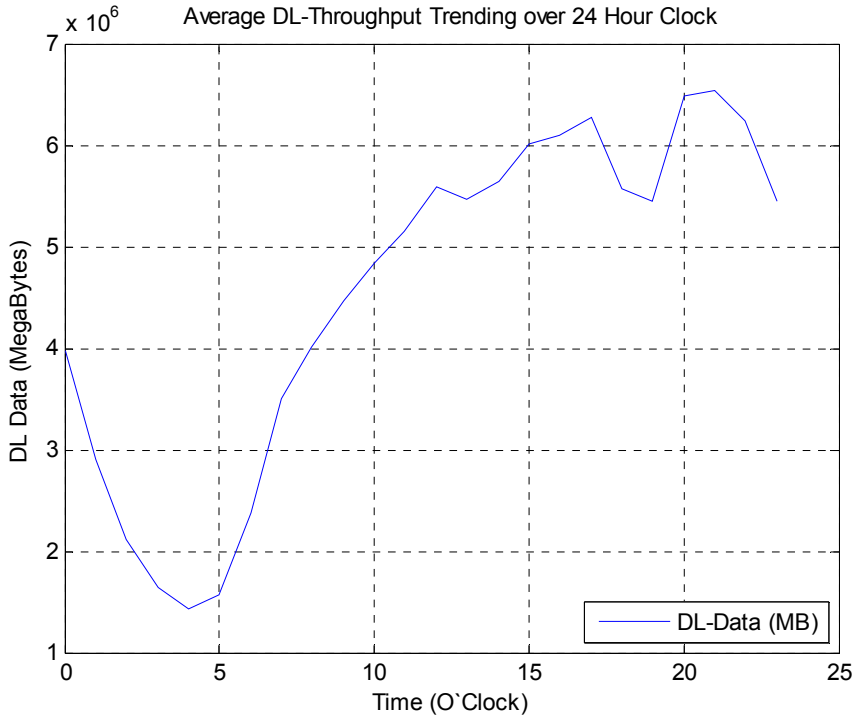

Figure 4-22 Trending of DL-Throughput over 24 Hour Clock in LTE

The above trending of LTE 'users' and 'data usage' on 24 hours clock is considered to be the base of LTE-A daily trending. The significance of this trending pattern is that, this is observed to be the same for all generations in mobile communication systems. The most inactive hour of a day is observed to be at 4AM with highest active number of users observed during evening times. The same pattern hold good for the Downlink data usage by the users. It can be expected that the LTE-A trending pattern should follow more or less the same. The below sections of LTE-A penetration analysis concludes in this regard.

4.8.3.2 Trend of DL throughput via SCell during CA

With the deployment of LTE-A feature - Carrier Aggregation in live network, Downlink Traffic trending on aggregated Secondary Cell is observed to evaluate the usage of CA feature in current scenario. Below shown trend of downlink traffic via SCell can help in getting an idea of how the Secondary Cell resources are exploited by the users.

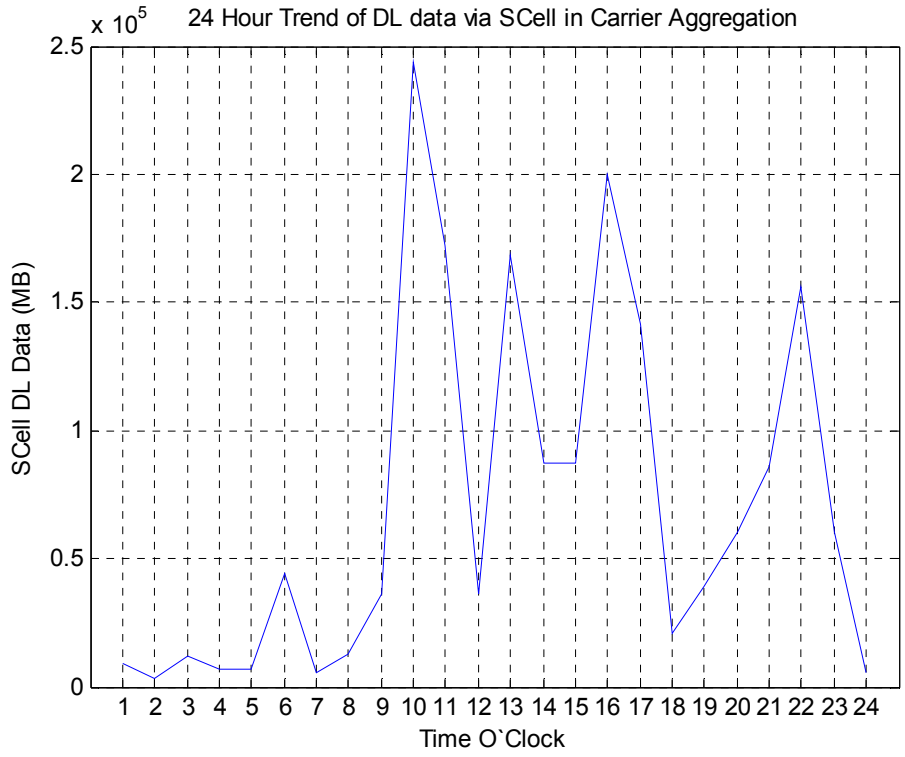

Figure 4-23 Trend of DL Data via SCell in CA

This SCell usage is expected to follow normal 24 hours trend of downlink throughput as shown in Figure 4-22. But this is not following the exact same trend. One reason to accommodate this is because of very less penetration of CA capable UEs in the network. There are only very few UEs (in single digit count) available for the users to

use. This initial real time analysis at the very starting of this new technology can help researchers to focus their work on interesting findings.

It can be expected that, once after the acceptable count of UEs is reached, the Secondary Cell Downlink pattern is to follow the typical DL-Throughput trending of a 24 hour clock.

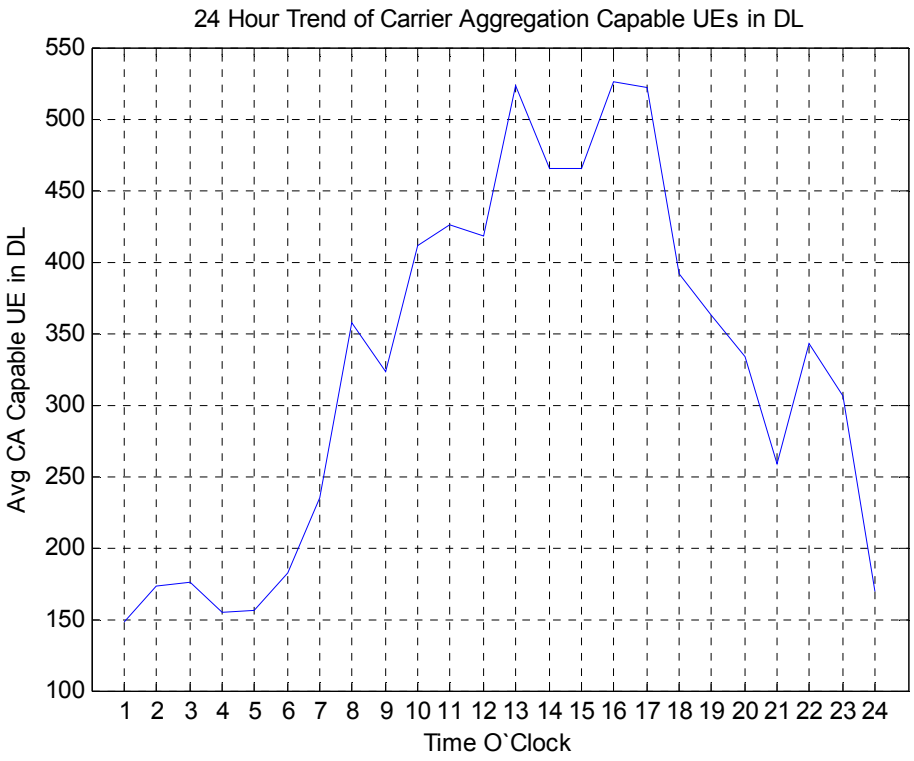

Figure 4-24 Trend of CA capable UEs in the Downlink

Figure 4-24 plots the trending of Average Carrier Aggregation capable UEs in the downlink. A UE is CA capable or not is meant form the eNodeB point of view meaning that it is enough that UE capability Information message indicates at least one DL SCell is supported by UE along with the band combination supported by the eNodeB. It can be observed from the plot that, for any given hour of a day, highest CA capable UEs

availability in Downlink is only 550 which is very less when compared to that of normal LTE users which is about 2000.

4.8.3.3 SCell Configured vs Activated UEs

In Carrier Aggregation, a SCell is first configured to the UE when eNodeB knows that it is CA capable. Once the UE knows the details about the secondary cell, if there is any non-GBR threshold met, SCell is activated to send the DL data via SCell.

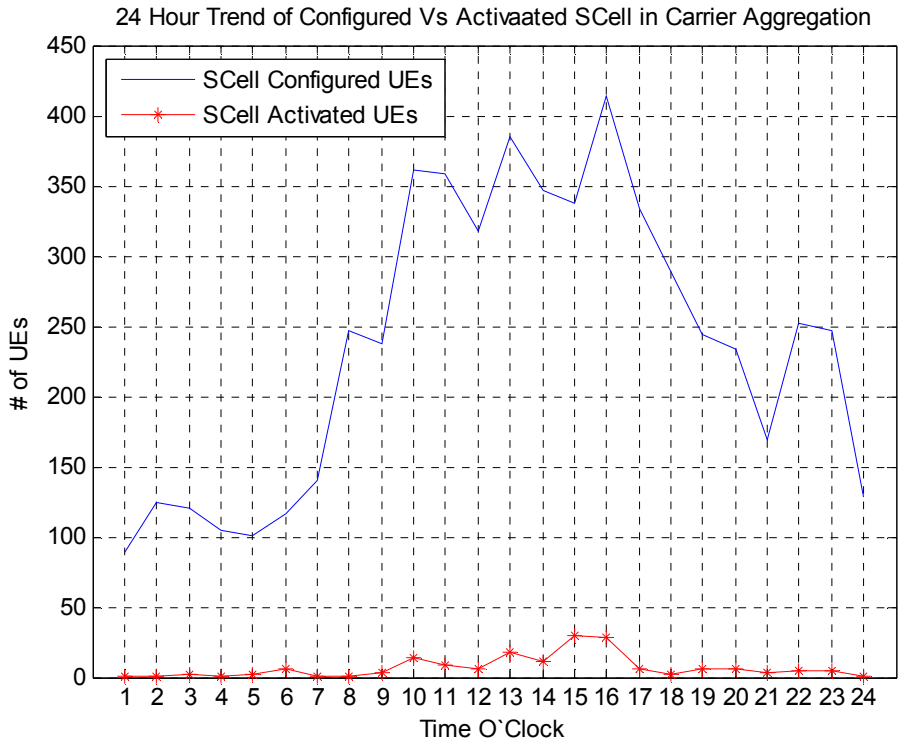

Figure 4-25 Configured & Activated UEs in a SCell

UEs with CA feature configured but without any active non-GBR Secondary link to UE is represented by 'SCell Confgured UEs' and all the UEs with Secondary Cell configured and also with an active non-GBR link to UE is considered as 'SCell Active UEs'. This analysis shows the comparison of configured and activated SCells showing

that only a maximum of 8.95% of configured UEs are activated and served using SCell. This reviles that there is a need of efficient utilization of SCell resources.

4.9 Conclusion

- LTE-Advanced data traffic is observed to exhibit Self-Similarity property.
- Degree of Self-Similarity increases with the increase in data traffic load.
- Higher the traffic load on usual 24 hours traffic trend, higher is the H-parameter value.
- With comparison of LTE and LTE-A Self-Similarity, LTE-A exhibits higher degree of Self-Similarity when compared to that of LTE data traffic.
- LTE-A exhibits higher degree of Self-Similarity when compared to that of LTE and Ethernet networks.
- Trending patterns for LTE-Advanced networks are observed based on the real world data.
- 24 hours trending pattern for the usage of Secondary Cell in DL throughput is analyzed.
- 24 hours trending pattern of Secondary Cell resource utilization is observed and need further modeling analysis to optimize the LTE-Advanced networks.

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Conclusions

This thesis work has contributed by evaluating the properties of data traffic in live LTE & LTE-A networks which are of utmost significance in further design considerations and deployments. Key analysis and observations are presented on live deployments of LTE-A which is shaping out for general availability in near future.

This work demonstrated the existence of Self-Similarity in real world LTE data traffic and the degree of Self-Similarity is calculated for various collected data sets. It is concluded from the results that higher the data traffic, higher is the degree of Self-Similarity. Degree of Self-Similarity is observed to be more in LTE networks when compared to that of Ethernet networks.

User arrival pattern in live LTE networks is affirmed to follow Poisson process similar to that of legacy networks. Corresponding Inter- Arrival time is verified to follow Exponential distribution with curve fitting and QQ-plots.

In the best known knowledge, the analysis of LTE-Advanced done in this work can be claimed to be the only available real world analysis in LTE-Advanced networks so far in the scope of this work. Self-Similarity is observed in LTE-A networks and the comparison of degree of Self-Similarity with LTE data traffic concludes that LTE-A exhibits higher Self-Similarity to that of LTE and Ethernet networks. Analysis on the daily trending pattern for Secondary Cell resource utilization in CA implementation suggests that more robust scheduling techniques are needed in order to use the resources

efficiently. The current penetration trend of CA capable Release 10 UEs in the network is also observed to be low.

5.2 Future Work

- Enhanced traffic forecasting algorithms can be developed to better estimate the drastically growing data traffic in rapidly expanding LTE & LTE-A networks.
- Poisson user arrival analysis can be used to develop robust tools to estimate the traffic patterns and proper utilization of network resources.
- The end user resources like the battery power are enhanced with the knowledge of traffic patterns exploiting the DRX concept in LTE networks [10].
- Secondary Cell resource trending suggests that robust resource allocations algorithms are to be developed which can efficiently allocate the resources in forthcoming technologies like 5G, HetNets.
- Trending of the penetration of CA capable UEs in the existing network is to be further analyzed for better network planning and deployments.

APPENDIX A

ACRONYMS

3GPP	Third Generation Partnership Project
CA	Carrier Aggregation
CDF	Cumulative Distributive Function
DL	Downlink
DL-PCC	Downlink Primary Component Carrier
DL-SCC	Uplink Secondary Component Carrier
FDD	Frequency Division Duplexing
GBR	Guaranteed Bit Rate
LTE	Long Term Evolution
PCELL	Primary Cell
PDCCH	Physical Downlink Shared Channel
PDF	Probability Distribution Function
PDSCH	Physical Downlink Shared Channel
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
SCELL	Secondary Cell
SS	Self-Similarity
TDD	Time Division Duplexing
UE	User Equipment
UL	Uplink
UL-PCC	Uplink Primary Component Carrier
UL-SCC	Uplink Secondary Component Carrier

REFERENCES

[1] Qilian Liang, "Ad-Hoc Wireless Network Traffic - Self-Similarity and Forecasting," IEEE Communication Letters, vol. 6, no. 7, July 2002.

[2] Nan E, Xiaoli Chu, Weisi Guo and Jie Zhang, "User Data Traffic Analysis for 3G Cellular Networks", in 8th International ICST Conference on Communication and Networking, China, pp.14-16, Aug. 2013.

[3] Andrea Goldsmith, "Wireless Communications," Cambridge: Cambridge University Press, 2005, Print.

[4] Theodore Rappaport, "Wireless Communications: Principles and Practices," Upper Saddle River, NJ, Prentice Hall PTR, 1996, Print.

[5] A.J. Field, U. Harder and P.G. Harrison, "Measurement and modeling of Self-Similarity traffic in computer networks," IEEE Proc.-Communications, vol. 151, no. 4, August 2004.

[6]Daniel WillKomm, Sridhar Machiraju, Jean Bolot and Adam Wolisz, "Primary Users in Cellular Networks: A Large-scale Measurement Study," in New Frontiers in Dynamic Spectrum Access Networks, pp. 1-11, Oct. 2008.

[7] Elias Jailani Muhammad Ibrahim and Ruhani Ab Rahman, "LTE Speech Traffic Estimation for Network Dimensioning," in IEEE Symposium on Wireless Technology and Applications (ISWTA), pp.315-320, 2012.

[8] W.Stallings, *High-Speed Networks: TCP/IP and ATM Design Principles*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[9] Will E. Leland, Murad S. Taqqui, Walter Willinger and Daniel V. Wilson, "On the Self-Similarity nature of Ethernet traffic (extended version)," in IEEE/ACM Trans., vol. 2, Issue 1, Aug. 2002.

[10] Ke Wang, Xi Li, Hong Ji and Xiaojiang Du, "Modeling and Optimizing the LTE Discontinuous Reception Mechanism under Self-Similar Traffic," in IEEE Trans. on Vehicular Technology, Issue 99, June 2014.

[11] Mark E. Crovella and Azeer Bestavrus, "Self-Similarity in World Wide Traffic: Evidence and Possible Causes," in IEEE/ACM Transactions on Networking, vol. 5, no. 6, pp. 835-846, Dec. 1997.

[12] Walter Willinger, Murad S. Taqqui, Robert Sherman and Daniel V. Wilson, "Self-Similarity through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," in IEEE/ACM Transactions on Networking, vol. 5, Issue 1, Aug. 2002.

[13] M.A. Farahani and M. Guizani, "Markov Modulated Poisson Model for Hand-off Calls in Cellular Systems," in Wireless Communications and Networking Conference, vol. 3, pp. 1113-1118, Sept. 2000.

[14] Anurag Kumar, D. Manjunath and Joy Kuri, *Wireless Netorking.* Morgan Kaufmann Publishers Inc., CA, 2008.

[15] Eric Dahlman, Stefan Parkvall and johan Skold, "4G: LTE/LTE-Advanced for Mobile Broadband," 2nd edition, Academic Press, 2013.

[16] Chang Min Park, Hae Beom Jung, Song Hee Kim and Duk Kyung Kim, "System Level Performance of Various Carrier Aggregation Scenarios in LTE-Advanced," in International Conference on Advanced Communication Technology (ICACT), pp. 814-817, Jan. 2013.

[17] Klaus Ingemann Pedersen, Frank Frederlksen, Claudlo Rosa, Hung Nguyen, Luls Gullheme Uzeda and Yuanye Wang, "Carrier Aggregation for LTE-Advanced: Functionality and Performance Aspects," in IEEE Communication Magazine, pp. 89-95, June 2011.

[18] Members of 4G Americas, "LTE Carrier Aggregation Technology Development and Deployment Worldwide," Oct. 2014.

[19] Foddis G, Garroppo R.G, Giordano S, Procissi G, Roma S and Topazzi S, "LTE Traffic Analysis and Application Behaviour Characterization," in European Conference on Networks and Communications at Bologna, pp.1-5, 2014.

[20] Laner Markus, Svoboda P, Schwarz S and Rupp M, "Users in Cells: A Data Traffic Analysis," in IEEE Wireless Communications and Networking Conference (WCNC), Shangai, pp.3063 - 3068, 2012.

[21] Paxson V and Floyd Sally, "Wide Area Traffic: The Failure of Poisson Modeling," in IEEE/ACM Transactions on Networking, vol. 3, Issue 3, pp.226-244, 2002.

[22] Group of Telecommunication Associates, 3RD Generation Partnership Project - The Mobile Broadband Standards [Online]. Available: www.3gpp.org. (URL)

BIOGRAPHICAL INFORMATION

Roopesh Kumar Polaganga was born in India in 1992. In May of 2013, he earned a Bachelor's Degree in Electronics and Communication Engineering (ECE) from Pondicherry Engineering College. Since August of 2013, he has been pursuing his Master of Science Degree in Electrical Engineering from the University of Texas at Arlington (UTA). While at UTA, he served as Graduate Research Assistant in the Communication and Networking Lab under the guidance of Dr. Liang where his research was focused on Ultra Wide Band and LTE technologies. He also worked as a Graduate Intern at T-Mobile USA during Summer 2014 and Fall 2014 furthering his knowledge on LTE and LTE-Advanced technologies.

Roopesh currently holds a position as an Associate Engineer in Tier 2 National Operations at T-Mobile USA. His technical areas of interests include Wireless Telecommunications and Computer Networks.