# Adaptive sampling technique for computer network traffic parameters using a combination of fuzzy system and regression model

SALAMA, A, SAATCHI, Reza <http://orcid.org/0000-0002-2266-0187> and BURKE, Derek

Available from Sheffield Hallam University Research Archive (SHURA) at:

https://shura.shu.ac.uk/15933/

---

This document is the Accepted Version [AM]

## Citation:

SALAMA, A, SAATCHI, Reza and BURKE, Derek (2018). Adaptive sampling technique for computer network traffic parameters using a combination of fuzzy system and regression model. In: 2017 4th International Conference on Mathematics and Computers in Science in Industry. MCSI 2017. Island, Greece, August 24-26, 2017. IEEE, 206-211. [Book Section]

---

## Copyright and re-use policy

# Adaptive Sampling Technique for Computer Network Traffic Parameters Using a Combination of Fuzzy System and Regression Model

A. Salama and R. Saatchi
Department of Engineering and Mathematics
Sheffield Hallam University
Sheffield, United Kingdom

D. Burke
Sheffield Children's Hospital
Sheffield, United Kingdom

*Abstract*— **In order to evaluate the effectiveness of wired and wireless networks for multimedia communication, suitable mechanisms to analyse their traffic are needed. Sampling is one such mechanism that allows a subset of packets that accurately represents the overall traffic to be formed thus reducing the processing resources and time. In adaptive sampling, unlike fixed rate sampling, the sample rate changes in accordance with transmission rate or traffic behavior and thus can be more optimal. In this study an adaptive sampling technique that combines regression modelling and a fuzzy inference system has been developed. It adjusts the sampling according to the variations in the traffic characteristics. The method's operation was assessed using a computer network simulated in the NS-2 package. The adaptive sampling evaluated against a number of non-adaptive sampling gave an improved performance.**

*Keywords— adaptive sampling; computer network quality of service; regression model; fuzzy logic.*

## I. INTRODUCTION

Determining the effectiveness of computer networks for communicating various applications is important for both the network users and service providers. The evaluation entails analysing the traffic parameters such as delay, jitter and packet loss ratio that need to be gathered by monitoring information packets [1][2]. However, performing this monitoring in real-time is computationally intensive as very large a number of packets are involved. Sampling is a well-established technique that allows a subset, containing smaller number of packets to be formed. An important feature of this subset is that it still accurately represents the traffic. Sampling can be adaptive or non-adaptive [3]. In adaptive sampling, unlike non-adaptive sampling, the number of selected packets to produce the subset changes in accordance with the extent of traffic variation and thus may provide a more optimal way of analysing the traffic.

Sampling is important in determining quality of service (QoS) for networks. The purpose of QoS is to either provide priority to certain time-sensitive applications (like video conferencing) or to guarantee an agreed level of service through resource (such as bandwidth) reservation [4] [5].

In this study an adaptive sampling method based on regression modelling and fuzzy logic was developed for wired and wireless networks. Fuzzy logic uses linguistic rather than numeric values for information processing and has an ability to model some complex problems more conveniently than mathematical modelling. As a result they have found applications in scenarios such as decision making, control and data modelling and analysis.

In conventional (crisp) logic membership is binary, i.e. either full membership (i.e. binary 1) or no membership (i.e. binary 0) while in fuzzy logic memberships is a continuum that ranges from 0 to 1 as shown in Fig.1. The degree of membership in fuzzy logic indicates the extent of membership. The performance of a network based on crisp logic can either be good or bad. However, in fuzzy logic, a network can simultaneously be a member of good and bad performance sets with different degrees of memberships [6].
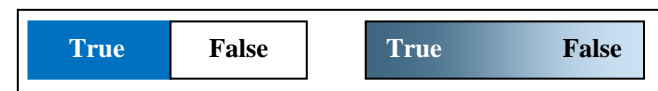


Fig. 1 Conventional crisp logic and fuzzy logic

Fuzzy Inference System (FIS) is an implementation of fuzzy logic where numeric inputs are first fuzzified into linguistic terms then through inferencing decisions are made by comparing the fuzzified inputs with the data that are coded (for example using a number of IF-THEN rules) in the knowledge base and finally the outputs are determined by defizzifying the linguistic results into numeric form [7][8]. This is illustrated in Fig.2.
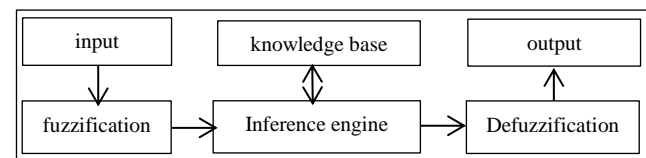


Fig. 2. A block diagram of fuzzy inference system.

In this study, the traffic was first modeled using regression analysis. Several regression models are available [9]. Regression analysis is a way to explore a relationship between dependent and explanatory variables. Regression can be linear or nonlinear but linear regression is most commonly used for predictive analysis and is the type used in this study. Regression models has been used in [10] for future sensors network readings, allowing network component to be predicted

based on current captured data or based on nearest network node. This led to a reduction in the amount of transmitted data packets. In the following sections, an overview of non-adaptive sampling approaches is provided, the study's methodology is explained and results are discussed.

## II. NON-ADAPTIVE SAMPLING APPROACHES

In general non-adaptive sampling can be divided into systematic, random and stratified as illustrated in Fig.3.
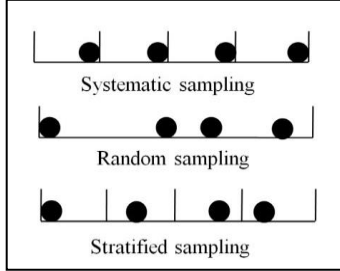


Fig.3. Illustrations of non adaptive sampling techniques [8][11].

Systematic sampling is a widely used method where a packet is selected in fixed predefined time periods or in some cases based on packet counts. In random sampling, packets are selected based on predefined time intervals like systemic method but the packet chosen randomly within that interval or based on random counts. Stratified method is a combination of systematic and random techniques where a packet is chosen randomly with predefined time intervals [8][11].

## III. METHODS

A modular and scalable network was designed using a network simulation package called NS-2. The network (shown in Fig.4) followed the recommended hierarchical network design that divides the network in into three tiers called core, distribution and access. This design approach improves network management by ensuring its modularity [12]. The design was compliant with the Open Source Interconnection (OSI) network model [13].
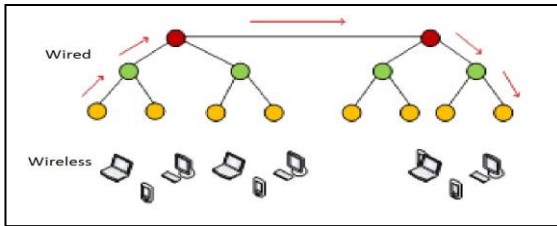


Fig.4. Network topology

The wired part of the network contained the core layer and had a capacity of 10 Mbps. The wireless parts contained the distribution and access tiers and were configured in the IEEE 802.11e protocol with Enhanced Distributed Channel Access (EDCA). The wireless channel capacity was 2 Mbps. The routing protocol in this scenario was Destination-Sequenced Distance Vector (DSDV) and the queuing mechanism for all scenarios was First-In-First-Out (FIFO). The queue size was 50 packets.

The traffic types were: video streaming, VoIP, HTTP and FTP. The packet size for VoIP was 160 bytes. G711 protocol was used as audio coding with 64 kbps transmission rate. The packet size for video streaming was 512 bytes. The video streaming frames were configured with maximum length of 1024 bytes and MPEG-4 coding scheme. The NS-2 scenarios ran for 800 seconds. Following each simulation, a trace file was generated by NS-2 that contained the network and traffic transmission details such as the packet types, transmitted and received times and packet sizes and delivery status. A Perl language based tool was developed to read the information from the trace file and determine the traffic parameters: delay, jitter, and packet loss ratio. These measurements were performed using equations explained below.

Delay ($D_i$) for the $i^{th}$ packet was determined as in equation (1) where $R_i$ and $S_i$ are the times a packet was received and sent respectively.

$$D_i = R_i - S_i \qquad (1)$$

Jitter ($J_i$) was determined using equation (2) where $D_i$ and $D_{i-1}$ are the delays associated with the current and previous packets respectively. The absolute parameter ensures jitter values remain positive.

$$J_i = absolute\ (D_i - D_{i-1}) \qquad (2)$$

The percentage packet loss ratio ($\%PL_i$) was determined by using equation (3) where $R_i$ and $S_i$ are $i^{th}$ packets received and sent respectively.

$$\%PL_i = \left(1 - \frac{\sum R_i}{\sum S_i}\right) \times 100 \qquad (3)$$

## IV. DESCRIPTION OF ADAPTIVE SAMPLING METHOD

The algorithm used regression to model traffic by considering delay, jitter and percentage packet loss ratio. The output of the model was then processed by the fuzzy inference system to adapt traffic sampling. The operation of the algorithm is shown in Fig.5 and its key parameters and elements are explained below. Fig.6 complements the flowchart in illustrating the sampling operation.

- *Pre and post-sampling sections*: These intervals contain the traffic that needs to be sampled. These intervals are kept fixed (predefined) and do not changed during sampling.

- *Inter-section Interval of data packets (isi)*: This interval is between pre- and post-sampling sections. Its duration is adaptively determined by considering the output of the fuzzy inference system.

- *Traffic matrix:* The traffic parameters were represented by an $n \times n$ traffic matrix to allow regression analysis, where $n$ is the number of sub-sections in the pre- and post-sampling sections. Each sub-section contains $n$ packets. The data modelling was carried out for the measured traffic parameters, i.e. delay, jitter and percentage packet loss ratio.

- *Traffic difference calculation using Euclidean distance (ED)*: ED measure was used to determine the amount of

traffic difference (*td*) between pre- and post-sampling sections for all traffic parameters.

- *Fuzzy inference system*: FIS was used to determine updated (*isi*) based on the current *(isi)* and the three traffic difference *(td)* values of delay, jitter and percentage packet loss ratio.
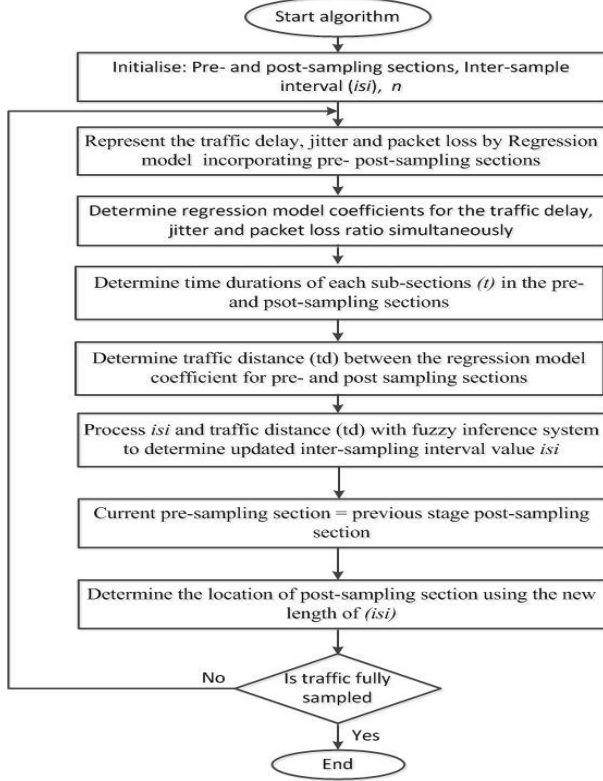


Fig. 5 Flow chart for the adaptive sampling method.

The traffic parameters delay, jitter and packet loss ratio were considered as the independent variables representing *p* values in regression equation (4). The sampling section was divided to sub-sections ($s_1$, $s_2$,...$s_n$). Each sub-section contained *n* packets as shown in Fig.5, where the traffic parameter values for each sub-section were entered as a row of the traffic matrix *P* and the associated time period of every sub-section were represented by the vector *T*. The vector $[E]' = [e_1, e_2,...,e_n]'$ is the error vector (the symbol ' signifies transpose).

$$T = PC + E = \begin{bmatrix} t_1 \\ t_2 \\ : \\ t_n \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & .. & p_{1n} \\ p_{21} & p_{22} & .. & p_{2n} \\ : & : & & \\ p_{n1} & p_{n2} & .. & p_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ : \\ c_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ : \\ e_n \end{bmatrix} \quad (4)$$

In this study, *n* was 3 which resulted in 3 sub-sections: $S_{1pre}$, $S_{2pre}$, and $S_{3pre}$ for pre-sampling section and $S_{1post}$, $S_{2post}$ and $S_{3post}$ for post-sampling sections as illustrated in Fig.6. Each

sub-section contained 3 data packets. For both post and pre-sampling sections a $3 \times 3$ traffic matrix was formed where each of its rows contained the traffic information of each sub-sections. This was repeated for the pre and post-sampling sections.
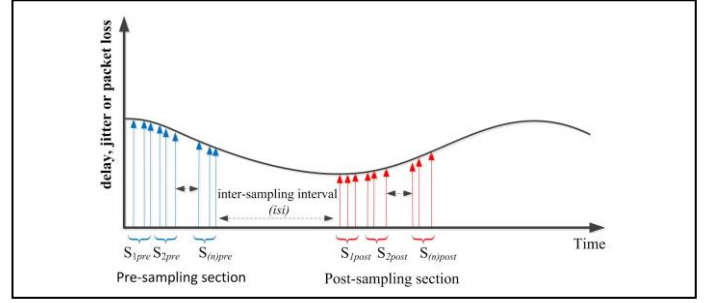


Fig. 6 Traffic representation for the algorithm

The time durations of the sub-sections were represented by $t_1$, $t_2$ ... $t_n$. These intervals were measured by subtracting the arrival time of the last packet for a section from the arrival time of the first packet for the same section. The error vector (*E*) in tested scenarios was set to zero. The regression coefficients; $c_1$, $c_2$ ... $c_n$ were determined by equation 5.

$$C = P^{-1}T \quad (5)$$

The magnitude of traffic difference (*td*) between the pre- and post- sampling sections was determined by comparing their respective regression model coefficients using the Euclidean distance measure as shown in equation 6.

$$traffic\ difference\ (td) = \sqrt{\left(c_{1pre} - c_{1post}\right)^2 + \left(c_{2pre} - c_{2post}\right)^2 + .... + \left(c_{npre} - c_{npost}\right)^2} \quad (6)$$

The fuzzy inference system received the current value of inter-sampling interval (*isi*) and the traffic difference (*td*) for traffic parameters delay, jitter and percentage of packet loss ratio and determined the updated value for inter-sampling interval (*isi*) as shown in Fig.7.
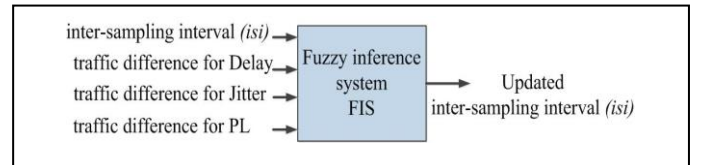


Fig. 7 FIS system to update inter-sample interval

The inputs to the fuzzy inference system, the values of traffic difference for delay, jitter and percentage packet loss

ratio and the inter-sampling interval (isi) were individually fuzzified by five membership functions. The traffic difference for delay, jitter and packet loss were represented by *VLow*, *Medium*, *High* and *Vhigh* fuzzy sets. The input inter-sampling interval (*isi*) was represented by *Vsmall, small, Medium, Large* and *Vlarge* fuzzy sets. The output was defizzzified by four membership functions, represented by *IL* (Low Increase), *NC* (no change), *DL* (Low Decrease), and *DH* (High decrease). These membership functions are shown in Fig.8.
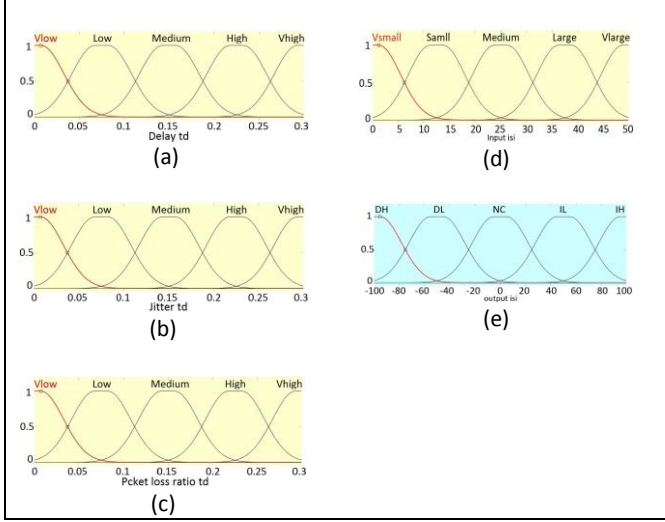


Fig. 8. Membership functions for (a-c) traffic difference sets for delay, jitter and percentage packet loss ratio. (d) inter-sampling interval (e) the updated inter-sampling interval

Fuzzy rules processed the values of inter-sampling interval (*isi*), traffic differences for delay, jitter and percentage packet loss ratio to update the inter-sampling interval (*isi*). Table I shows examples of the fuzzy rules.

To assess the efficiency of the developed sampling technique, the relative error was used to assess the closeness between the mean values of the original traffic parameters and its sampled version by Equation (7).

$$Relative\ error = absolute\ (M_1 - M_2) / M_1 \qquad (7)$$

Where $M_1$ and $M_2$ are the mean values of the original traffic parameters, i.e. (delay, jitter and percentage packet loss ratio) and sampled packets. Another evaluation measure was Bias that was determined by equation 8.

$$Bias = absolute\ (M_1 - M_2) \qquad (8)$$

The network was simulated using NS-2 package. Multiple traffic applications were transmitted over simulated network.

Table I Examples of fuzzy rules

| no | current isi | TD delay | TD jitter | TD packet loss ratio | updated isi |
|---|---|---|---|---|---|
| 1 | Very small | Very low | Very low | None | Increase high (IH) |
| 2 | Very small | Very low | None | Very low | Increase high (IH) |
| 3 | Very small | None | Very low | Very low | Increase high (IH) |
| 4 | None | Very low | Very low | Very low | Increase high (IH) |
| 5 | None | Low | Low | Low | Increase low (IL) |
| 6 | Small | None | Low | Low | Increase low (IL) |
| 7 | Small | Low | None | Low | Increase low (IL) |
| 8 | Small | Low | Low | None | Increase low (IL) |
| 9 | Medium | Medium | Medium | None | No change (NC) |
| 10 | Medium | Medium | None | Medium | No change (NC) |
| 11 | Medium | None | Medium | Medium | No change (NC) |
| 12 | None | Medium | Medium | Medium | No change (NC) |
| 13 | None | High | High | High | Decrease low (DL) |
| 14 | Large | None | High | High | Decrease low (DL) |
| 15 | Large | High | None | High | Decrease low (DL) |
| 16 | Large | High | High | None | Decrease low (DL) |
| 17 | None | Very high | Very high | Very high | Decrease low (DH) |
| 18 | Very large | None | Very high | Very high | Decrease low (DH) |
| 19 | Very large | Very high | None | Very high | Decrease High (DH) |
| 20 | Very large | Very high | Very high | None | Decrease High (DH) |

*td:* traffic difference, measured by Euclidean distance

QoS parameters delay, jitter and packet loss ratio were measured by equations 1-3. The simulation ran for 800 seconds. The linear regression model shown in equation (4) was used to model traffic parameters for pre- and post-sections of the inter-sampling interval (isi). The traffic parameter differences were measured using equation (6) to determine the magnitude of the traffic changes. This was performed for the three traffic parameters simultaneously. Fuzzy inference system updated the inter-sampling interval based on the current value of inter-sampling interval, the extent of traffic change and fuzzy rules for each update. The results are shown in Figs.9-10.

## V.   RESULTS

Fig.9(a) indicates the changes in the updated inter-sampling interval based on the changes in the values of delay, jitter and percentage packet loss ratio. The traffic difference for delay is shown in Fig. 9(b). The sampled delay in Fig.9(d) shows that it is closely following the original delay shown in Fig.9(c). Figs.10(d) and 10(f) show the behavior of the sampled jitter and percentage packet loss ratio to be similar to the original jitter and percentage packet loss ratio shown in

Figs.10(c) and 10(e). Traffic differences for jitter and percentage packet loss ratio are shown in Figs.10(a) and 10(b) respectively.
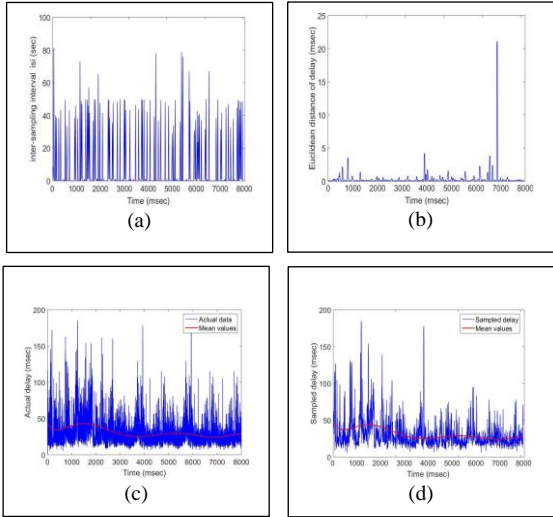


Fig. 9. Typical results obtained from the developed adaptive technique (a) FIS output for the inter-sampling interval (isi) (b) traffic difference for delay (c) original traffic delay (d) sampled traffic delay.
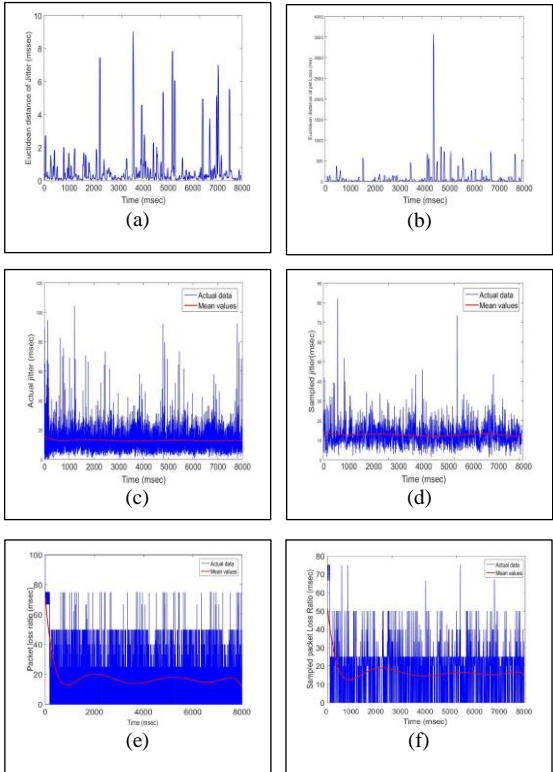


Fig. 10. Typical results obtained from the developed adaptive technique: (a) measured traffic difference for jitter, (b) measured traffic difference for packet loss (c) original traffic jitter, (d) sampled traffic percentage jitter (e) original traffic packet loss ratio (f) sampled traffic packet loss ratio.

Table II provides a summary of the mean values of traffic parameters; delay, jitter and percentage packet loss ratio and the original traffic parameters. For comparison the mean values for the traffic parameters; delay, jitter and percentage packet loss ratio obtained by non-adaptive sampling methods of systematic random, and stratified sampling are also included in Table II.

To compare the developed adaptive sampling and non-adaptive sampling methods, the bias and relative errors were determined. These values are shown in Table III. They indicate that the developed adaptive method has the lowest relative error and bias values as compared as compared with the non-adaptive methods, signifying an improved performance. The results indicate that the developed adaptive sampling represented the original traffic parameters more accurately than the non-adaptive methods.

Table II Mean values of the traffic parameters: delay, jitter and percentage packet loss

| Sampling Methods | Delay (ms) | Jitter (ms) | PL (ms) |
|---|---|---|---|
| Original traffic | 31.00 | 12.84 | 17.87 |
| adaptive method | 31.20 | 12.67 | 17.81 |
| Non-adaptive: systematic | 31.27 | 12.62 | 17.45 |
| Non-adaptive: random | 30.03 | 12.34 | 17.32 |
| Non-adaptive: stratified | 30.52 | 12.51 | 17.61 |

Table III Comparisons of bias and relative error between developed technique and non-adaptive methods

| Sampling method | | Delay (ms) | Jitter (ms) | % Packet loss ratio |
|---|---|---|---|---|
| Developed adaptive method | % Relative error | 0.006 | 0.0125 | 0.0034 |
| | Bias | 0.20 | 0.16 | 0.06 |
| Non-adaptive Systematic | % Relative error | 0.0087 | 0.0171 | 0.0235 |
| | Bias | 0.27 | 0.22 | 0.42 |
| Non-adaptive random | % Relative error | 0.0313 | 0.0389 | 0.0308 |
| | Bias | 0.97 | 0.50 | 0.55 |
| Non-adaptive stratified | % Relative error | 0.0155 | 0.0257 | 0.0145 |
| | Bias | 0.48 | 0.33 | 0.26 |

VI.   CONCLUSIONS

A novel adaptive method that samples multimedia network traffic has been developed. It incorporates the traffic parameters delay, jitter and percentage packet loss ratio simultaneously in its analysis. Its performance was compared with the non-adaptive sampling techniques of systematic, random, and stratified. The developed method adaptively increased the inter-sampling interval section resulting in an increase in the number of packets sampled when the traffic

variations increased and vice versus. The adaptive sampling method represented the original traffic more accurately than the non-adaptive methods.

## VII. REFERENCES

[1] Al-Sbou, Y. A., Saatchi, R., Al-Khayatt, S., Strachan, R., Ayyash, M., and Saraireh, M. (2008). A novel quality of service assessment of multimedia traffic over wireless ad hoc networks. In Next Generation Mobile Applications, Services and Technologies, NGMAST'08, IEEE, The Second International Conference on, 479-484.

[2] Saraireh, M., Saatchi, R., Al-Khayatt, S. and Strachan, R. (2007). Assessment and improvement of quality of service in wireless networks using fuzzy and hybrid genetic-fuzzy approaches. Artificial intelligence review, 27(2-3), 95-111.

[3] Alippi, C., Anastasi, G., Di Francesco, M. and Roveri, M. (2010). An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors. IEEE Transactions on Instrumentation and Measurement, 59(2), 335-344.

[4] Anjum, B. and Perros, H. (2015). Bandwidth estimation for video streaming under percentile delay, jitter, and packet loss rate constraints using traces. Computer Communications, 57, 73-84.

[5] Gokhale, V., Nair, J. and Chaudhuri, S. (2016). Opportunistic adaptive haptic sampling on forward channel in telehaptic communication. In Haptics Symposium (HAPTICS), IEEE, 217-222.

[6] Egaji, O. A., Griffiths, A., Hasan, M. S. and Yu, H. (2015). A comparison of mamdani and sugeno fuzzy based packet scheduler for MANET with a realistic wireless propagation model. International Journal of Automation and Computing, 12(1), 1-13.

[7] Egaji Jafari, A., Al-Khayatt, S. and Dogman, A. (2012). Performance evaluation of IEEE 802.11 p for vehicular communication networks. Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2012 8th International Symposium on, 1-5.

[8] Dogman, A., Saatchi, R., Al-Khayatt, S. and Nwaizu, H. (2011). Adaptive statistical sampling of voip traffic in wlan and wired networks using fuzzy inference system. 2011 7th International Wireless Communications and Mobile Computing Conference, 1731-1736.

[9] Fan, J., Liao, Y. and Liu, H. (2016). An overview of the estimation of large covariance and precision matrices. The Econometrics Journal, 19(1), C1-C32 [4].

[10] Zhang, B., Liu, Y., He, J. and Zou, Z. (2013). An energy efficient sampling method through joint linear regression and compressive sensing. In Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference on, IEEE, 447-450.

[11] Dogman, A., Saatchi, R. and Al-Khayatt, S. (2010). An adaptive statistical sampling technique for computer network traffic. Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on, 479-483.

[12] de Souza Couto, R., Secci, S., Campista, M. E. M. and Costa, L. H. M. K. (2016). Reliability and survivability analysis of data center network topologies. Journal of Network and Systems Management, 24(2), 346-392.

[13] Díaz Santos, J. R. (2016). Design and Implementation of a Communication Protocol to Improve Multimedia QoS and QoE in Wireless Ad Hoc Network.