IEEE Africon 2011 - The Falls Resort and Conference Centre, Livingstone, Zambia, 13 - 15 September 2011

Optimization of Resources for H.323 Endpoints and Terminals over VoIP Networks

Samuel Nlend, Theo G. Swart and Willem A. Clarke
Dept. of Electrical and Electronic Engineering Science
University of Johannesburg, P.O. Box 524,
Auckland Park, 2006, South Africa
Email: nlesam68@yahoo.com, tgswart@uj.ac.za, willemc@uj.ac.za

***Abstract*—We suggest a method of optimizing resource allocation for real time protocol traffic in general, and VoIP in particular, within an H.323 environment. There are two options in the packet network to allocate resources: aggregate peak demand and statistical multiplexing. Statistical multiplexing, our choice for this case, allows the efficient use of the network resources but however exhibits greater packet delay variation and packet transfer delay. These delays are often the result of correlations or time dependency experienced by the system's queue due to the variations observed in different point processes that occur at a point of time. To address these issues, we suggest a queuing method based on the diffusion process approximated by Orstein-Ulenbeck and the non-validated results of Ren and Kobayashi.**

I. Introduction

Taking into account the market behaviour, the trend and speed of the telecommunication industry, Africa seems to be the ideal place for investments in terms of telecommunications. Many of its biggest metropolis has a lack of cable pairs and the expansion is costly and needs a lot of work and time. As a result, we are experiencing poor service delivery in the PSTN networks. The advent of mobile communications has helped a lot in terms of alternative solutions but there is still a lot to explore for both technique and market.

It turns out that the solution to reach everyone is the New Generation Network suggesting the convergence of services into a common platform. This has been a dream for the users to have their telecommunication services consolidated on one unique device.

With the impact of the Internet, with most of today's traffic volume generated by applications running on top of IP protocol (such as voice and TV, amongst other services), it suggests that in the near future the Internet service providers will play a crucial role in that expansion. In fact, the new generation network allows the ISP to bridge the services to the end users while the main Telecommunications Companies terminate at the edges of the network as presented in Fig. 1.

If simply providing the services is good, delivering them with a required QoS is even better. In fact, when the ISP has to bridge different services to a large number of customers, it is always ensured that we have enough bandwidth to carry the traffic, but small details in the switching system is missed, wherein the multiplexers undergo heavy traffic, experiencing huge delays that result in packet loss and therefore degradation

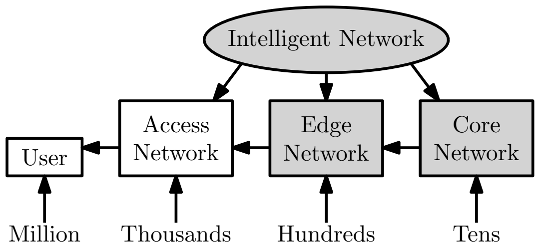

Fig. 1. Network by 2015

in the quality of service. Therefore, the need of resource allocation in an efficient way is required and the best way to do it is with statistical multiplexing.

The statistical multiplexer allocates resources from the average to the peak value and queues the overflow in the buffer. It is actually in the buffer that the problem lies. The aggregate traffic at the buffer's queue exhibits some correlation that discards the parametric from the Poisson process. In [1], Sriram and Whitt has proposed the rescale of time to tackle the issue by modulating the arrival rate by a factor of $1/n$ of the number of arrivals. In [2], Heffes and Lucantoni suggested the Markov Modulated Poisson process and reach the results that are widely used today. Anick, Mitra and Sondhi [3] in turn, have gone further and suggested a fluid queuing model to address the variable bit rate issue that the queues experience.

In this paper, we propose a diffusion approximation process, which is the Markov Gaussian process to extend the Markov process and the fluid model already suggested. The advantage of this method is that it presents a more general queue solution and can be fitted in any kind of system. Further, the method is equivalent [4] to the Markov Modulated Rate process which is based on the fluid method, which is the limit at the asymptotic level of the diffusion approximation and linked with the Markov Modulated Poisson Process by their moment generating function given by:

$$g_{\text{MMRP}}(z,t) = g_{\text{MMPP}}(1 + \log z, t)$$

and their first moments are identical [5].

The rest of the paper is organized as follows. In Section II, we present the ITU endpoints and terminal configuration according to H.323 recommendation and define the points of multiplexing. We will present the proposed model in

Section III through its representation and our motivation for the preference. In Section IV we derive the G/G/1 queue, leaning on the diffusion approximation, and finally conclude in Section V.

II. The ITU-H.323 Endpoints and Terminals Configuration

According to the ITU standard H.323, the H.323 architecture in Fig. 2 is proposed. We are interested on the endpoints equipped with the multipoint control Unit (MCU) to allow signal processing. We are therefore suggesting a two level multiplexing scheme for this environment, namely level 1 (MUX1) and level 2 (MUX2), both analyzed by the diffusion approximation process. In this paper, we will focus on level 1, with level 2 left for future work.

A. Description

The ITU Recommendation H.323 presents the following equipment:

- multipoint control Unit (MCU),
- the gateway (GW),
- terminal (T), and
- the gatekeeper (GK).

The above endpoints and terminals are characterized in terms of signal processing by the following functions:

- multipoint controller, and
- multipoint processor.

For their exact functions, we refer the reader to the related ITU text book [6], [7].

III. Diffusion Approximation Representation

Taking into account these results showing the relationship between the MMPP and the MMRP, instead of modulating the rate as Poisson in the aggregate process, it seems better to opt for the MMRP, which approximation is based on the fluid method. Since the limit of the fluid method gives rise to the diffusion model, we have chosen the latter as the main approximation model for both levels of multiplexing we are suggesting in the paper.

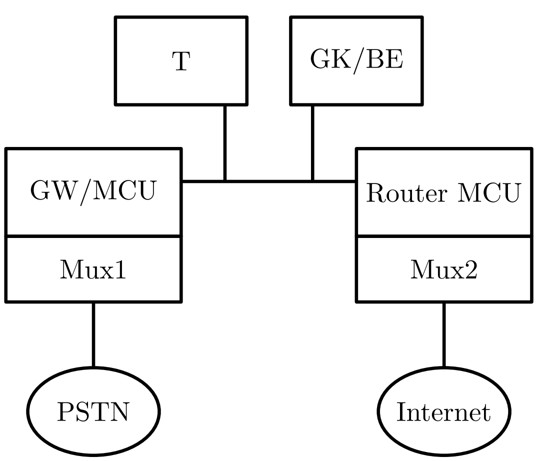

Fig. 2. ITU Standard H.323 Architecture

The diffusion approximation presented here, whether to aggregate the traffic of the On-Off sources or the one of the MMRP, tends to capture the variability of the buffer content.

Define $Q(t)$ as the queue size of the buffer and C packets/sec as the constant capacity of the transmission link.

By definition, the statistical multiplexer allocates capacity that lies between the average and the peak rates and buffers the traffic when the load exceeds the capacity. Therefore the changes in the multiplexer can be captured by the following differential equation:

$$\frac{dQ(t)}{dt} = \begin{cases} R(t) - C, & R(t) > C, Q(t) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

It turns out that neither $R(t)$ nor $Q(t)$ are Markovian. This drives us to step up towards the couple $(A_k(t), Q(t))$.

Then $(A_k(t), Q(t)) = (A_k, 1 \leq k \leq K; Q(t))$ is a Markov process.

IV. Statistic Multiplexing in the H.323 Network

The statistical multiplexer allocates resources from the average to the peak level of its capacity and buffers the traffic when the arrival rate is higher than its capacity. It is actually that arrival in the buffer that causes some dependence or correlation which in turn is the cause of packet transfer delays and packet variation delays, resulting in the congestion of the system. We are proposing here the diffusion approximation process in the sense of Ornstein-Uhlenbeck to derive the buffer's content.

The Ornstein-Uhlenbeck process approximation is a diffusion process characterized by the diffusion differential equation in the form:

$$\frac{\partial f(x,t)}{\partial t} = \beta \frac{\partial f(x,t)}{\partial x} + \sigma^2 \frac{\partial^2 f(x,t)}{\partial x^2} \quad (1)$$

where β is the mean and σ^2 is the variance of the process.

The solution of the conditional probability of this differential equation is a Gaussian distribution at the equilibrium state ($t \to \infty$):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\beta)^2}{2\sigma^2}\right\} \quad (2)$$

We are suggesting the optimization based on the lower bound and the upper bound of the above as described in the following.

Let $f(x)$ and $F(x)$ be the density and probability distribution functions defined by the Gaussian process function as above, taken at its standard values $N(0,1)$, by definition we have: $F(x) = \int_0^x f(u)du$. $F(x)$ is close to 1 for large x.

Its complementary is given by: $1 - F(x) = \int_x^\infty f(u)du$, which is close to 0 for the same condition.

The Upper Bound: Let $v = u^2/2$, then

$$\begin{aligned} 1 - F(x) &= \int_x^\infty f(u)du < \int_x^\infty \frac{u}{x} f(u)du \\ &= \int_{\frac{x^2}{2}}^\infty \frac{1}{x\sqrt{2\pi}} \exp\{-v\}dv = \frac{1}{x} f(x) \end{aligned}$$

978-1-61284-993-5/11/$26.00 ©2011 IEEE

Therefor $1 - F(x) < \frac{1}{x} f(x)$.

The Lower Bound: Similarly, using $f'(u) = -uf(u)$ and the quotient rule,

$$\begin{aligned}\left(1 + \frac{1}{x^2}\right)(1 - F(x)) &= \int_x^\infty \left(1 + \frac{1}{x^2}\right) f(u)du \\ &= \int_x^\infty \left(1 + \frac{1}{u^2}\right) f(u)du \\ &= -\frac{1}{x} f(x)dx\end{aligned}$$

It turns out therefore that:

$$\frac{x}{1 + x^2} f(x) < 1 - F(x) < \frac{1}{x} f(x) \quad (3)$$

giving the elementary bounds of the process.

Also, after having determined the lower bound and upper bound of the buffer, it seems reasonable to maximize the number of packet arrivals in the aggregate process. The joint distribution function $f(x_1, x_2, \ldots, x_n)$ for the N sources is given after normalization of the mean value for a fixed variance,

$$\begin{aligned}f(x_1, x_2, \ldots, x_n) &= f(x) \quad (4) \\ &= \frac{1}{\sqrt{(2\pi)^n \sigma^n}} \exp\left(-\frac{1}{2} \frac{\sum_{i=1}^n (x_i - \hat{x})}{\sigma^2}\right),\end{aligned}$$

with

$$\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

being the sample mean of the distribution.

For the multivariate Gaussian distribution as we find in the level 2 multiplexing, the maximum likelihood of the covariance matrix X allows us to optimize the distribution and is given by

$$\text{cov}(X) = \frac{1}{n} \sum_{i=1}^n [(X_i - \hat{X})(X_i - \hat{X})'], \quad (5)$$

which is the sample covariance matrix.

A. Level 1 Voice Multiplexer

Situated at the GW with MCU functionality, this multiplexing scheme targets the On-Off sources. The On-Off sources, being characterized by encoders with voice activity detectors (VAD), have voice packets being sent when the source is active and no packets during silence. As a result, the distribution process is a jump process with a Gaussian distribution solution as the number becomes large. Using the result of Kobayashi and Ren [8], who addressed the issue by using Ornstein-Uhlenbeck diffusion equation that has defined the complementary buffer content by:

$$1 - F(x) = \left[\frac{1}{y\sqrt{2\pi\sigma_r}} \exp\left\{-\frac{\sigma^2 y}{2}\right\}\right] \exp\{u_0^- x\} \quad (6)$$

where

$$y = \frac{c - m_R}{\sigma^2}.$$

According to Kobayashi and Ren, the resolution of the O-U diffusion differential equation yields a quadratic equation derived from Weber's equation written as:

$$\left(\sum_k \frac{R_k^2 \sigma_k^2}{\alpha_k + \beta_k}\right) u + (C - \sum_k R_k Y_k^*) u - \sum_k i_k(\alpha_k + \beta_k) = 0$$

where i_k is a non-negative required integer of Weber's equation at its asymptotic Weber's approximation. This equation admits two roots in which the dominant obtained at $i_k = 0$ is the one with negative value. That is $u_0^+ = 0$ and

$$u_0^- = \frac{C - \sum_k R_k y_k^*}{\sum_k \frac{R_k^2 \sigma_k^2}{\alpha_k + \beta_k}},$$

where m_R and σ^2 are respectively the mean rate and variance values of the superposed traffic of the number N of different sources of k type traffic in the diffusion process.

Also, u_0^- is the dominant root, the solution of the quadratic equation of the underlying development of the process.

B. Level 2 Voice Multiplexing

Situated at the router or any kind of bridge to the Internet, this multiplexing scheme targets any kind of traffic characterized by multiple sources with multiple rates.

Using the same results as precedent and expanding the variables to an $N \times K$ dimension, we also obtain from [9] at the equilibrium state, the complementary function:

$$1 - F(x) = \left[\frac{1}{y\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}\right] \exp\left\{-\frac{2\sigma y}{\alpha} x\right\} \quad (7)$$

where

$$y = \frac{c - u_R}{\sigma_R}$$

and u_R and σ_R^2 are respectively the mean rate and variance values of the superposed traffic of the number $N \times K$ different states of source traffic in the diffusion process, and

$$\alpha = \lim \frac{1}{t} \text{var}(R(t)), \quad t \to \infty$$

the variation of the rate in the diffusion process.

This function is an approximate solution of the multivariate Gaussian process distribution given by:

$$f(x, t) = \frac{1}{\sqrt{(2\pi)^M} \det|\Sigma|} \exp\left\{-\frac{1}{2}[(x - \hat{x})\Sigma(x - \hat{x})']\right\},$$

with Σ being the covariance matrix. However, we will focus on level 2 in the next part of the research.

V. Implementation

A. Level 1 Statistical Multiplexing

Let N_k be the total number of k-type sources with rate R_k; let a_k be the number of active k-type sources with mean active period β_k^{-1}; $N_k - a_k$ is the number of silence period of k-type sources with mean value α_k^{-1}. Since the aggregate rate depends on the number of active k-type sources, $R(t) = \sum R_k a_k(t)$. Let $E[a(t)]$ be the mean number of N_k sources, we can write: $\beta_k^{-1} = E[a_k(t)].T$, where T is the packetization duration of the voice algorithm.

B. Characterization of the Arrival Process and the Buffer Content

The distribution function of the number of k-type arrivals of the diffusion process at the equilibrium state is given by:

$$f_k(y) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{(Y - Y^*)^2}{2\sigma_r^2}\right\},$$

$$y^* = \frac{\alpha_k N_k}{\alpha_k + \beta_k}, \quad \sigma_r = \frac{\alpha_k \beta_k N_k}{(\alpha_k + \beta_k)^2}$$

The aggregate process is given by

$$f(y) = \prod_{k=1}^{n} f_k(y)$$

with mean rate $m_R = \sum R_k Y_k^*$ and the variance $\sigma_R^2 = \sum R_k^2 \sigma_k^2$.

The optimization of the number of arrival is obtained according (4). The resulting maximum likelihood is given by:

$$P(y) = \frac{1}{\sigma_R^2\sqrt{2\pi}} \exp\left\{-\frac{(y - \hat{y})^2}{2\sigma_r^2}\right\}$$

where

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

is the sample mean of the maximum likelihood.

The probability for a positive queue is given by the lower bound expression (6) and is defined by:

$$F(Q > 0) = \frac{1}{\mu\sqrt{2\pi\sigma_R}} \exp\left\{-\frac{\sigma_R^2\mu^2}{2}\right\}$$

with $\mu = \frac{c - u_r}{\sigma_R^2}$.

1) The Delay: According to Little's theorem the number of jobs in the queue is $N_Q = m_R W$.

The maximum arrival number of packets is given by the maximum likelihood which is combined with the positive queue probability to yield N_Q, so that the waiting time is given by:

$$W(y) = \frac{1}{m_R}\left[\frac{1}{\sigma_R^2\sqrt{2\pi}} \exp\left\{-\frac{(y - \hat{y})^2}{2\sigma_R^2}\right\}\right]. \left[\frac{1}{\mu\sqrt{2\pi\sigma_R^2}} \exp\left\{-\frac{\sigma_R^2\mu^2}{2}\right\}\right]$$

2) The Queue Length: Provided the positive probability of the queue $P(Q > 0)$ is also the probability of having the arrival rate R higher than C_1, $P(R > C_1)$ by combining this probability and the complementary probability function, we derive the lower bound using (3) and the following one for large values of x:

$$F(Q > x) = \frac{1}{\mu\sqrt{2\pi\sigma_R^2}} \exp\left\{-\frac{\sigma_R^2\mu^2}{2}\right\} \exp\{u_0^- x\}$$

We can then write $f(x) = \frac{1+x^2}{x} F(Q > x)$ is the maximum queue length distribution admissible.

VI. Application

Level 1 Multiplexing: We choose $k = 2$-type of voice sources:

- Source 1: G711A, packetization delay 1 ms, rate 64 kbps. Assuming mean number of active sources 22, packet size = 8 bytes.
- Source 2: ADPCM, packetization delay 16 ms, rate 32 kbps. Assuming the mean number of the active sources 22, packet size = 64 bytes.

The mean silence period = 650 ms and $C_1 = 100$ Mbps.

From Fig. 3 and Fig. 4 one can say that by introducing the maximum likelihood through its estimates we can suggest the result in the form of the asymptotical approximation of the Gaussian distribution of $N(0, \sigma^2)$.

From Fig. 5 we can say that the distribution approximation as shown by Ren and Kobayashi has a trend to underestimating the overflow of the queue as longer as $\mu > 0$.

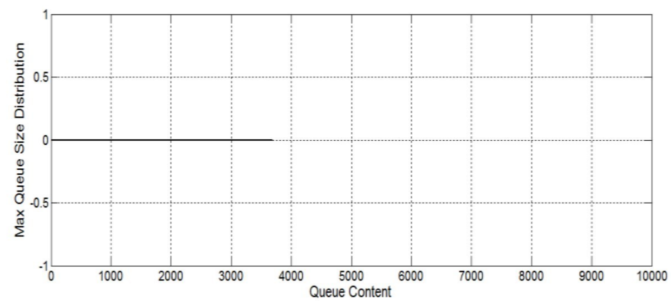

Fig. 3. Size distribution plot

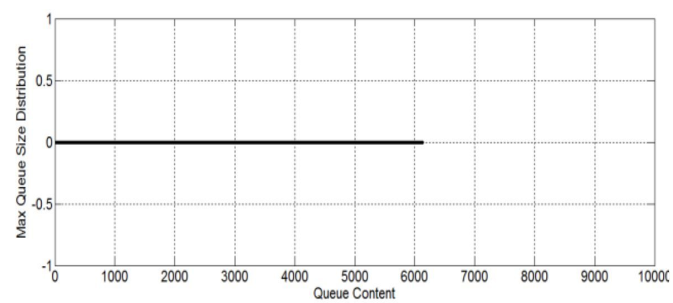

Fig. 4. Size distribution plot

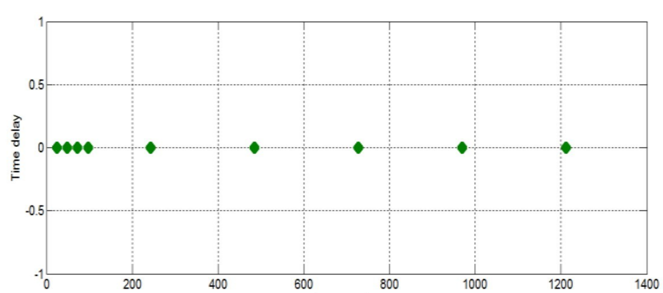

Fig. 5. Queue waiting time plot

978-1-61284-993-5/11/$26.00 ©2011 IEEE

VII. Conclusion

The results obtained give us a chance to considering the asymptotical approximation of the Gaussian distribution which is the central limit theorem. Therefore, the number of jobs in the queue without a perceptible delay is given by $\sqrt{n}(\hat{y} - y_k)$ as $n \rightarrow \infty$ in the identically distributed arrival normal distribution. This is in compliance with what has been stated earlier, as long as the intensity of traffic is less that 1 Erlang, that means $\rho = \frac{m_R}{C} < 1$: the diffusion approximation has a trend to underestimate the overflow of the buffer content. These results lead us to the requirements as imposed by Halfin-Whitt heavy traffic regime: if n the total arrival rate becomes large simultaneously, the amount of *excess capacity* required to achieve any given level of performance is of order of $\sqrt{n}$. As our intent is to carry heavy traffic from different source types on the PSTN, the diffusion approximation seems to be adequate compared to the Markov Modulated Poisson Process which allows only the jump to another rate. Also, the maximum likelihood as taken here, is worthy to mention, is more accurate when the sample number is large as is the diffusion approximation when the traffic intensity is close to one. The diffusion approximation also fits with the entire layer II schemes of TCP/IP used today.

A level 2 multiplexing scheme will allow us to have a better view of controller configuration for the Egress traffic, but we defer its analysis to our next paper as well as particular adjustments for different ubiquitous access layers of today.

References

[1] K. Sriram and W. Whitt, "Characterizing superposition arrival process in the packet multiplexer for voice and data," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 833–846, Sep. 1986.

[2] H. Heffes and D. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 856–868, Sep. 1986.

[3] D. Anick, D. Mitra and M. M. Sondhi, "Stochastic theory of a data handling system with multiple sources," *The Bell System Technical Journal*, vol. 61, no. 8, pp. 1871–1894, 1982.

[4] Q. Ren and H. Kobayashi, "Diffusion approximation modelling for Markov modulated bursty traffic and its application to bandwidth allocation in ATM networks," *IEEE Journal on Selected Area in Communications*, vol. 16, no. 5, pp. 679–691, Jun. 1998.

[5] J. M. Ferrardiz, "The relationship between a MMPP and an MMRP," *HP Labs Technical Report*, HPL-91-123, Nov. 1992.

[6] ITU-T Recommendation H.323, "Packet based Multimedia Communications Systems," version 1, 1996.

[7] ITU-T Recommendation H.323, "Packet based Multimedia Communications Systems," version 6, 2006.

[8] H. Kobayashi and Q. Ren, "A diffusion approximation analysis of an ATM statistical multiplexer with multiple types of traffic: Part I, Equilibrium state solution," in *Proceedings of the IEEE International Conference on Communications*, Geneva, Switzerland, 23–26 May, 1993, pp. 1047–1053.

[9] J. W. Roberts (ed.), *Performance Evaluation and Design of Multiservice Network*, Final Report COST 224, Commission of the European Communities, Brussels, Holland, 1992.