Forward-link Throughput Optimization for Wireless Cellular Packet Data Networks

Moses E. Ekpenyong
Department of Computer Science
University of Uyo, PMB. 1017, 520003, Uyo, NIGERIA
(ekpenyong_moses@yahoo.com, mosesekpenyong@gmail.com)

ABSTRACT

An attempt has been made in this paper to access the control layer of the General Packet Radio Service (GPRS) network base station (BS), using throughput optimization. We focus on the effect of packet transmissions on the throughput of forward-link (downlink) power balancing, with specific interest on the relative throughput of the four GPRS coding schemes for a realistic GPRS network operating in Nigeria. We derive a model capable of improving the channel quality and simulate the throughput and choice of optimum coding scheme over diverse packet sizes using the MATrix LABoratory toolkit. Simulation results show that the existing system requires effective network optimization to improve the system's throughput performance.

Keywords: GSM/GPRS coding, payload, SINR, payload length, optimization

1. INTRODUCTION

Throughput is defined as the average rate of successful message delivery over a communication channel and is therefore a good measure of the channel capacity of a communication link. The message or information delivery may be over a physical or logical link, or over a wireless channel passing through the cellular network. The throughput of a cellular network depends on a number of variables namely, the packet size, transmission rate, number of overhead bit in each packet, received signal power, received noise power spectral density, modulation technique and channel conditions. The key to maximizing throughput is maintaining the signal-to-interference-and-noise ratio (SINR) at an optimum level [1]. Cellular networks use a scheme called handoff to minimize signal breakup as a mobile user migrates from one cell to another. They also employ spread-spectrum technology and special coding schemes (where each base station is assigned a code) to enable the multiplexing of multiple users over the same physical channel.

The effect of additive noise and interference from other cells is common to all cellular networks. The total power of these impairments is σ^2 watts. To be precise, we refer to the impairments (collectively) as noise. Noise appears at the receiver as an additional signal that does not contribute to the overall throughput. The system has to use some of its power and bandwidth resources to overcome the effect of noise.

In cellular networks, transmission from a base station say i, to a node j, is seen as interference by transmission from any node n to any other node k, such that $(n,k) \neq (i,j)$, and consequently, reducing the bit service rates for link (n,k). This interaction introduces a negative feedback (into the system) in the form of temporary congestion in the network, caused by fluctuations in the wireless channel quality, thus reducing the network throughput around the congested area and increasing further the congestion and fading rates, thereby obstructing the transmission of packets and hindering the propagation of network signals.

This paper emulates a realistic network by using empirical data from the field. Firstly, we obtain realistic data measurements from Airtel Nigeria (formally, Zain communications limited), using a dashboard Motorola cell monitor. The data were obtained from base station controllers within the southeast region of Nigeria during busy hours. In studying the network, we observe the behaviour (of important parameters) of the system, operating under ideal conditions and derive a system model that depicts the system.

2. GPRS: BACKGROUND ISSUES

Due to growing interest in mobile data applications, the global system for mobile communications (GSM) has been extended by the general packet radio service (GPRS). GPRS allows the transmission of packet data at moderate bit rates using random access channels within the existing GSM infrastructure. GPRS is a packet data overlay to GSM, which has the largest number of subscribers in the world among all wireless systems [2]. It provides the necessary functionalities to handle packet data in GSM networks. Mobile terminals such as notebooks, personal digital assistants (PDAs), mobile phones and pagers can give access to a wide range of new mobile data services. There are three main parameters that determine the link layer performance of a GPRS network. They include frequency reuse, error coding scheme (CS) and data packet delay [3].

(i)   Frequency Reuse

Frequency reuse capability is a concept in which, with precise power control, different cells in a cellular network can share same frequencies (from

one cell). GPRS systems typically use universal frequency reuse, as oppose to earlier networks, where the mobile stations and base stations use the whole bandwidth to transfer and receive information. Since GSM voice services usually require a frequency reuse of 3 or 7, as separate overlaying networks, GPRS provides better resource utilization and maximizes the number of users fulfilling the specified QoS requirements. The cellular network has a predetermined range of frequencies to exchange information between the base stations and mobile stations. Therefore, the smaller the frequency reuse factor which a cellular network can employ, the higher the network capacity.

(ii)   Error Coding Scheme (CS)
Table 1 shows the four coding schemes employed by GPRS systems:

Table 1. GPRS Coding Schemes

Coding Scheme	Code Rate	PLC data block payload (bytes)	Service data rate kb/s
CS – 1	½	22	8.8
CS – 2	$\approx \frac{2}{3}$	32	12.8
CS – 3	$\approx \frac{3}{4}$	38	15.2
CS – 4	1	52	20.8

The payload of each Radio Link Control (RLC) block is added to the Radio Link Control/Media Access Control (RLC/MAC) header and a block check sequence. For CS-1, CS-2 and CS-3, the radio block is convolutionally coded with code rate ½ and then punctured to the desired code rate. CS-4 does not contain forward error correction. The service data rate is calculated from the 240ms duration of a GPRS super-frame. The super-frame consists of 52 GSM frames including 48 traffic frames and 4 control frames. The 48 traffic frames carry 12 RLC blocks. Therefore the data rate is computed as follows:

$$\frac{payload(bytes/block) \times 12(block/\sup er-frame) \times 8(bits/byte)}{240(ms/\sup erframe)}$$

$$= 0.4 \times payload(bytes(kb/s))$$

Among the 4 coding schemes, CS-1 has the least coding rate. It is also the most robust coding scheme. Therefore, CS-1 is used for all control messages. CS-2 and CS-3 have lower coding rates and more information bits. CS-4 is the most data efficient coding scheme, but is most vulnerable to channel impairments. The base station sub-system can select any of the coding schemes to optimize the throughput and delay performance.

The GSM system uses a combined FDMA/TDMA air interface, where one TDMA frame is sent at 4.615ms, in a 200KHz wide frequency band. Each TDMA frame is divided into eight slots and a mobile terminal is often assigned to a slot. A sequence of these slots is called a channel. In GPRS, four TDMA frames are combined with additional control information to form one Logical

Link Control (LLC) frame. Each of the eight channels in a TDMA frame can be used for GPRS.
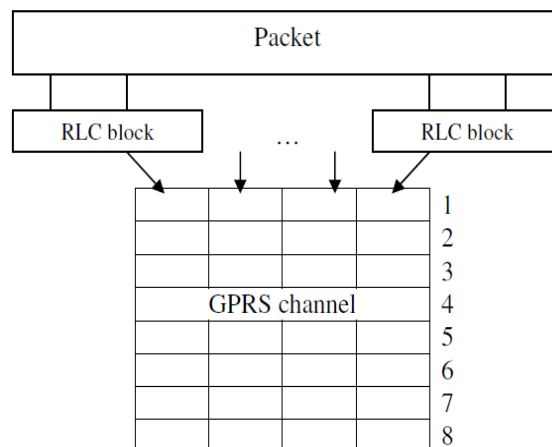

Figure 1. A simplified GSM frame structure with GPRS (packet data) channels.

At the link layer, there are a maximum of eight GPRS packet data channels (PDCHs) per allocated frequency. Each of them capable of transferring one Radio Link Control (RLC) block every 20ms, providing a raw bit rate of 23Kbps per channel. All mobile stations requiring data services in a particular cell share the available PDCHs.

3.   RELATED LITERATURE

Several research publications exist on throughput analysis of cellular networks [4-5]. Also, many models have been developed to suit the resource issues investigated. These include choosing an optimal power level to maximize throughput [6]. In [7], coding of GPRS networks by forward error correction, using punctured convolutional codes at different coding rates and retransmission of RLC blocks containing unrecoverable errors, usually referred to as Automatic Repeat Request (ARR) is used to strengthen the throughput of cellular networks. Sung and Wong [8] have presented forward-link throughput maximization, where the interference is modeled by the aid of the orthogonality factor and determined by the transmission strategy for a range of the orthogonality factor values and user's channel gains, so as to maximize the system throughput, subject to a total power constraint. Most of the previous works found have taken specific look at throughput in different cellular network technologies (such as TDMA, FDMA, CDMA, etc.) and consider various system parameters in their analysis.

4.   DOWNLINK QUALITY MEASUREMENT

Due to the success of cellular telephone services in the 1990's, the global communication system has shifted attention to data transmission [9]. Throughput and soft handoff provides a good measure of the quality of the wireless data link:

Maximum Throughput with No Additive Noise

Before carrying out a performance analysis for arbitrary number of terminals, N, let us examine a two terminal case (N=2), to provide a clue on the effect of power levels on base station throughput.

(i)   Two terminals case

According to [8], the simplest nontrivial throughput maximization of

$$\gamma_i = G \frac{Q_i}{\sum\limits_{\substack{j=1 \\ j\neq i}} Q_j + \sigma^2} \qquad (1)$$

occurs when N=2 and $\sigma^2 = 0$. In this case, the normalized throughput U, is a function of just one variable, $Z = Q_1 / Q_2$ and $\gamma_1 = G/Z$. However, $\gamma_2 = G_2$. By adopting the notation $U_2(Z) = f(GZ) + f(G/Z)$ as the normalized throughput when N=2 and $\sigma^2 = 0$, it is found that $\frac{dU_2}{dZ} = 0 \, at \, Z = 1$. This suggests that $U_2(1) = 2f(G)$, could be a local maximum, depending on the sign of the second derivative at Z=1. Examining the second derivative for arbitrary S-shaped $f(\gamma)$, we find that $U_2(1)$ is a local minimum at low values of G, and a local maximum at high values. Specifically,

$$\frac{d^2 U_2}{dZ^2}\Big|_{Z=1} = 2G[Gf''(G) + f'(G)]$$
$$= 2G \frac{d}{dx}(df'(x))\big|_{x=G} \qquad (2)$$

The sign of the second derivative of U_2 is the same as that of the derivative of $xf'(x)$. For the class of functions $f(x)$, $xf'(x)$ has a bell shape reaching a single maximum at some point $x = G*$. Therefore, the derivative of $xf'(x)$ is positive for any x, to the left of G*, and is negative otherwise. This implies that $U_2(1)$ is a local maximum when G is large enough to exceed G*, and is a local minimum otherwise. It is determined whether there is a global maximum at the boundaries of the feasible region when Z=1 yields a local minimum; i.e. at Z=0 and Z=∞. Even when Z=1 is a local maximum, the probability that the global maximum lies at the boundary of the feasible region must also be considered. This suggests that we compare the equal-received-power condition (Z=1) with the condition that terminal two, does not transmit at all (Z=0). The comparison reveals that Z=1 produces higher throughput when $2f(G) > 1$, while Z=0 produces higher throughput when $2f(G) > 1$. Accordingly, the critical processing gain G_c, is defined as the value of G for which $f(G) = 0.5$. G_c has the property that single terminal transmission (Z=0 or Z=∞) is better than equal-received-power transmission when $G<G_c$ and conversely, when $G>G_c$. Since G is proportional to the bandwidth, we can state that the system requires a bandwidth corresponding to at least $G=G_c$ to support two data terminals.

(ii)   Arbitrary Number of Terminals

Let's consider

$$\frac{\partial u}{\partial Q_i}\Big|_{Q_i=0} = 0; \sigma^2 = 0 \qquad (3)$$

and

$$\frac{\partial u}{\partial Q_i}\Big|_{Q_i=0} > 0; \sigma^2 = 0 \qquad (4)$$

Equations (3) and (4) are used to explore power balancing to determine whether they provide maximum or minimum throughput. We extend the approach adopted in the case of N=2. To do so, we assume that signals f from k transmitters arrive at the base station with equal power and that the other N-k terminals turn off their transmitters when $Q_i = Q$ for i = 1, 2, …, k and $\sigma^2 = 0$

and

$$U(k) = kf\left[\frac{G}{k-1}\right] \qquad (5)$$

where

$U(k)$ is the power balancing throughput and a function of k.

k are the transmitters

f are the signals transmitted by k transmitters

G is the processing gain

For values of $G>G_c$, U(k) has a maximum value at $k = N* \geq 2$. Throughput is maximum when N* signals are received with equal power and the other N-N* signals are not transmitted at all.

5. THROUGHPUT AND DELAY MODELING: PARAMETERS DEFINITION

The expected throughput of the system can be represented mathematically as:

$$E[throughput] = \lim_{t\to\infty} \frac{N(t) \times packet\_length}{t}(kb/s) \qquad (6)$$

where

$N(t)$ is the number of packets transmitted in the system

$packet\_length$ is the length of packets (assuming equal length of packets)

t is the expected service time

Let $\frac{N(t)}{t} = 1$, represent the inverse of the expected packet service time. As t approaches infinity, we have

$$E[throughput] = \frac{packet\_legnth}{\overline{T}}(kb/s) \qquad (7)$$

where $\overline{T}$ is the average transfer delay excluding the random access phase that takes place before the uplink channel assignment.

To compute the packet transmission delay, we segment the total delay into several components starting with channel assignment till the final acknowledgement:

$$\overline{T} = T_{assign} + T_{tx} + T_{error-correction} + T_{final-ack} \qquad (8)$$

where

T_{assign} is the duration of the channel assignment procedure for transmission and processing time.

T_{tx} is the data block's first time transmission period based on the packet length and coding scheme.

$T_{error-correction}$ is the handling time for backward error control including retransmission.

$T_{final-ack}$ is the final acknowledgement (ACK) message including control acknowledgement message processing and transmission time.

For our model, T_{assign}= 2 (blocks) and $T_{final-ack}$ = 4 (blocks)

$$T_{tx} = n (blocks) \qquad (9)$$

where

$$n = \left\lceil \frac{packet\_length}{block\_length} \right\rceil, \quad block\_length = \begin{bmatrix} 22\ bytes\ for\ CS-1 \\ 32\ bytes\ for\ CS-2 \\ 38\ bytes\ for\ CS-3 \\ 52\ bytes\ for\ CS-4 \end{bmatrix}$$

$T_{error-correction}$ can be split into two:

$$T_{error-correction} = T_{retx} + T_{nack} \qquad (10)$$

where

T_{retx} is the retransmission time of erroneous data blocks

T_{nack} is the negative acknowledgement (NACK) message processing and transmission times.

$$T_{rectx} = \frac{bler(cs,ci) \times n}{(1 - bler(cs,ci))} (blocks) \qquad (11)$$

where $bler$ is the block error rate

Let m(n) be a random variable representing the number of NACK messages successfully sent during the transmission of a packet consisting of n RLC blocks, which is equivalent to the number of retransmission cycles for the frame. Let $T_{one-nack}$ denote the transmission time for one NACK message (Turina and Ludiving, 1997),

$$T_{nack} = T_{one-nack} \times E[m(n)] \qquad (12)$$

To account for errors in the downlink ACK/NACK messages, we note that the process of sending one NACK message consists of 2 parts: initial transmission and retransmission of the message. Here $T_{initial-nack}$ = 3. The initial after-frame NACK message is transmitted with CS-1. It is correctly received with probability $(1 - bler(CS-1, Ci))$. If the NACK is not received by the Mobile Station (MS), the MS will resend the oldest block with errors or the oldest unacknowledged block, expecting the base station (BS) to re-send the

ACK/NACK message. Our model assumes that if the data block is again in error, the whole process is repeated. Only if both the resent data block and the NACK are successful, will the NACK transmission process complete. The round trip delay is 4 blocks. Assuming there is no limit on the number of transmissions allowed for one NACK message,

$$T_{one-nack} = T_{initial-nack} + \frac{4\,bler(CS-1,Ci)}{(1-bler(CS-1,Ci)) \times (1-bler(Cs,Ci))} \qquad (13)$$

6. THE SYSTEM MODEL

Recall from equation (6) that the throughput is given by

$$Thr = \lim_{t \to \infty} \frac{N(t) \times P_L}{t}$$

$$\therefore Thr = \frac{P_L}{\overline{T}} \qquad (14)$$

where the delay $\overline{T} = \frac{N(t)}{t}$ and P_L is the packet length. Also recall from equation (8) that

$$\overline{T} = T_{assign} + T_{tx} + T_{error-correction} + T_{final-ack}$$

where

T_{assign} is taken as 2 (blocks)

$T_{final-ack}$ is taken as (4 blocks)

$$T_{tx} = n = \frac{P_L}{B_L}$$

$$B_L = \begin{cases} 22\ bytes\ for\ CS-1 \\ 32\ bytes\ for\ CS-2 \\ 38\ bytes\ for\ CS-3 \\ 52\ bytes\ for\ CS-4 \end{cases},$$

which represents the block length or payload.

Now, substituting the respective parameters into equation (10) gives:

$$T_{error-correction} = \frac{bler \times n}{1 - bler} + T_{one-nack} \times E[m(n)]$$

$$= T_{initial-nack} + \frac{4\,bler}{(1-bler)(1-bler)} \times E[m(n)] \qquad (15)$$

Hence, equation (14) can be rewritten as

$$\qquad (16)$$

$$Thr = \frac{P_L}{T_{assign} + \left(\frac{P_L}{B_L} \right) + \left(\frac{bler \times n}{1-bler} \right) + T_{initial-nack} + \left[\frac{4\,bler}{(1-bler)} \times \sum_{i=1}^{n} n \binom{n}{i} bler^i (1-bler)^{n-1} (1-bler)^n (m-1) \right]}$$

where

m is the number of messages sent during transmission

The throughput can also be related to $bler$ and the data rate as:

$$Thr = (1 - bler) \times R$$

where R is the data rate

$$\therefore \frac{P_L}{\overline{T}} = (1 - bler) \times R$$

$$\overline{T} = \frac{P_L}{(1 - bler)R} \tag{17}$$

But $R = 0.4 \times payload$

$$\therefore \overline{T} = \frac{P_L}{(1 - bler)(0.4 \ x \ payload)} \tag{18}$$

7. EXPERIMENTAL DATA ANALYSIS

Figure 2 reveals that for the duration under study, the system throughput fluctuates at the various base station locations. This could be attributed to network instability caused by propagation impairments that distort the received data packets, resulting in poor delivery at the receiver's end. In Figure 3, we plot the average throughput as a function of the base stations over the study period. We observe that the throughput performance is not significantly influenced by the base stations, as evident from the obtained correlation coefficient (R^2) value (i.e. R^2=0.0013). However, other factors such as increased number of users, fading, interference due to noise, equipment malfunctioning or downtime, etc., could contribute to a great extent and explain better the instability that has threatened the system's performance. Airtel Nigeria therefore requires an optimized technique to deal with these impairments.
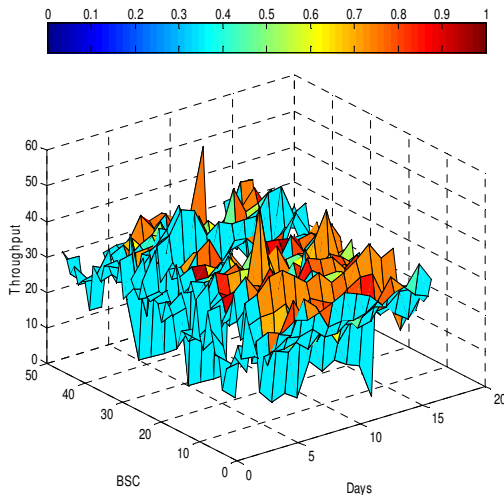

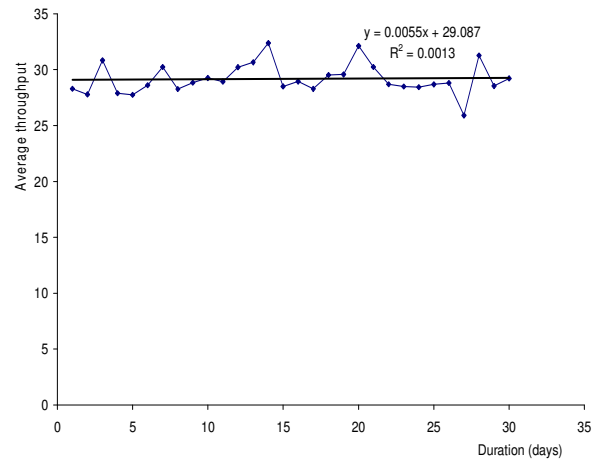Figure 2. Forward-link Throughput analysis at various BSs during the study period

Figure 3. Graph of average throughput vs. number of days

8. SIMULATION INPUT

To evaluate the proposed system model, sample inputs were used to simulate the model. Some of the input data were drawn from the field, under ideal conditions. Table 2 shows the parameters used for the research.

Table 2. Sample simulation parameters and their values

Parameter	Value
The packet length (PL)	200 – 1,600 kb/s
The block length or payload (BL)	CS-1 =22 bytes CS-2=32 bytes CS-3=38 bytes CS-4=52 bytes
Duration of the channel assignment procedure for transmission and processing time (T_{assign})	2 blocks
Average forward-link throughput from the field (Thr)	29.114 kb/s
Block error rate (bler)	0.001-0.009
Number of packet transmitted (n)	40 bytes
Initial after-frame negative acknowledgment message(s) transmitted ($T_{initial\ nack}$)	3 blocks
Number of message(s) sent during transmission(m)	1-10

Sample outputs were produced in the form of graphs and studied. A discussion of the outputs obtained is done in the following section.

9. DISCUSSION OF RESULTS

Figure 4 shows the effect of messages transmitted on the system throughput using diverse block lengths or payload sizes. We observed that as the coding scheme increases, the throughput also increases, which validates our optimization technique. As more messages are sent, throughput tends to slowdown for the different coding schemes, except for CS-4, which shows improved system performance.

Shown in Figure 5, is a relationship between the throughput and packet length. Comparing the throughput using varied packet lengths, we observed that the packet length had dominant effect on the system throughput. The throughputs for shorter packets are substantially lower than longer packets. Again, the throughput seems inhibited for other coding schemes except CS-4, which shows improved system performance.
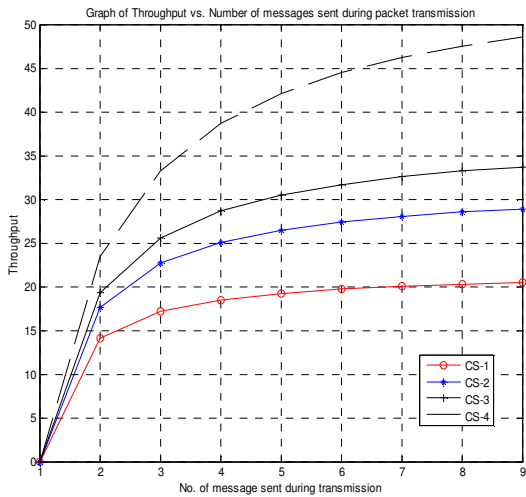

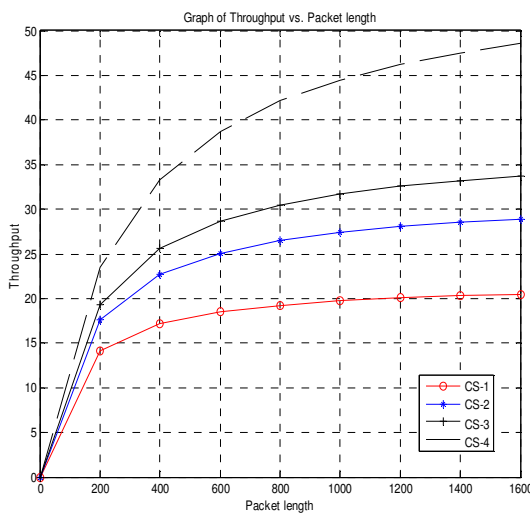Figure 4. Graph of throughput vs. number of massages sent

Figure 5. Graph of throughput vs. packet length

As expected in Figure 6, the throughput increases with the block error rate (bler). At lower

bler, the throughput increases rapidly but as the bler exceeds 0.02, the throughput slows down, indicating network congestion as a result of more users accessing the system. This may lead to the possibility of poor message delivery at the receiver's end.
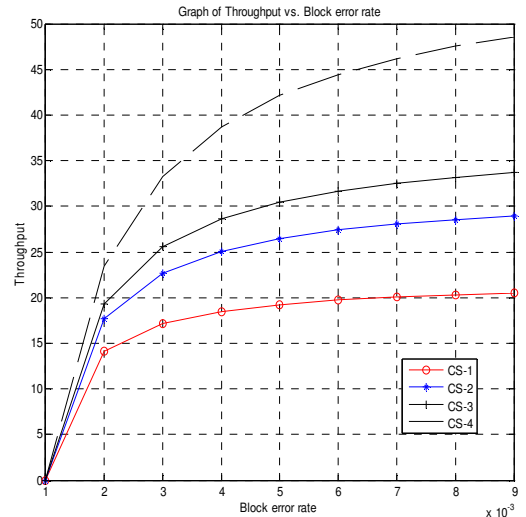

Figure 6. Graph of throughput vs. block error rate

In Figure 7, we examine the performance of the system by studying the effect of packet length on system's delay using the four coding schemes. We observed that the longer the packet length, the longer (time) it takes to deliver the desired message. However, as the coding scheme is increased (from CS-1 to CS-4), the delay in transmission shortens. This explains the robustness score of the various coding schemes in dealing with channel impairments. Therefore, network operators should implement efficient coding schemes that will optimize the system's performance and guarantee effective message deliveries.
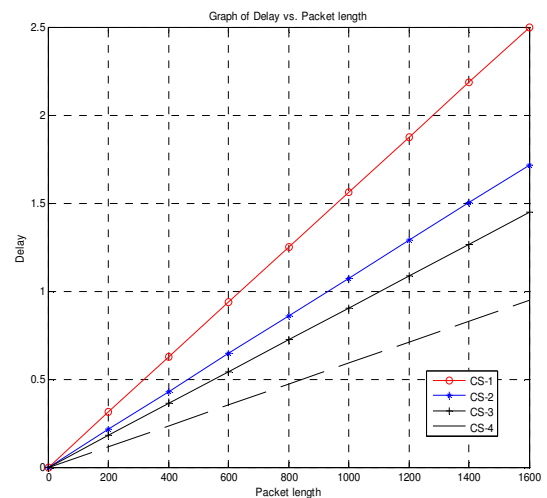

Figure 7. Graph of delay vs. packet length

10. CONCLUSION

Due to sharing of common communication channel by users, interference in form of noise is always generated and is the main limitation for determining the optimum throughput. A general overview of GPRS technology has been presented and a summary of those attractive features, which demonstrates GPRS technologies superiority over 2G cellular network technologies and other 3G technologies given.

This paper has been able to optimize the throughput in GPRS cellular network through forward-link (downlink) transmit power balancing by simulating a model with relevant parameters. Some practical issues regarding the application of this research have been considered and series of investigations identifying the key features and parameters, as well as the functionality of the GPRS coding schemes have been presented. Simulation results show that the system's throughput has been optimized by the proposed model. This technique will aid in transmitting signals from the GPRS base station with equal possible power.

REFERENCES

[1] Rodriguez V. and Goodman D. J. (2003). Prioritized Throughput Maximization via Rate and Power Control for 3G CDMA: The Two Terminal Scenario. In Proceedings of 40th Allerton Conference Communication, Control and Computing.

[2] Goodman D. (2005). Performance Analysis of Radio Link/Media Access Control of General Packet Radio Service System, Ph.D. Thesis Report. WINLAB TR 209.

[3] Hamalainen D. and Kari P. (1995). Handover and Dynamic Channel Allocation Techniques in Mobile Cellular Network. IEEE Transaction on Vehicular Technology. 44(20):229-237.

[4] Bali S. A. (2001). Study of the Throughput of General Packet Radio Service (GPRS) for Internet Applications. M.Sc. Thesis, University of Jordan, Jordan.

[5] Lindemann C. and Thümmler A. (2001). Evaluating the GPRS Radio Interface for Different Quality of Service Profiles. In Proceedings of 12th GI/ITG Fachtagung Kommunikation in Verteilten Systemen (KiVS), Hamburg, Germany: 291-301.

[6] Larson R. and Olofsson D. (1998). Handover and Dynamic Channel Allocation Techniques in Mobile Cellular Network. IEEE Transaction on Vehicular Technology. 44(2):120-143.

[7] Cai J. and Goodman D. (1997). General Packet Radio Service in GSM. IEEE Communication Magazine. 35: 122-131.

[8] Sung C. W. and Wong W. S. (2002). Power Control and Rate Management for Wireless Multi-media CDMA system. IEEE Transactions on Communication. 49(7): 1215-1226.

[9] Goodman D. and Marantz Z. (2003). Maximizing the throughput of CDMA Data Communications. In Proceedings of 58th IEEE Conference on Vehicular Technology. 5: 3468-3472.