TRAFFIC ANALYSIS AND MODELING IN PMR SYSTEMS

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY
IN PARTIAL FULLFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Başak CAN
July 2003

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Hayrettin Köymen (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Nail Akar

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Billur Barshan

Approved for the Institute of Engineering and Sciences:

Prof. Dr. Mehmet B. Baray
Director of Institute of Engineering and Sciences

ABSTRACT

TRAFFIC ANALYSIS AND MODELING IN PMR SYSTEMS

Başak Can

M.S. in Electrical and Electronics Engineering

Supervisor: Prof. Dr. Hayrettin Köymen

July 2003

Reliable knowledge of traffic in PMR (Private Mobile Radio) systems is essential for assessing the issues in migration from analog to digital and trunked PMR systems. In this work, we investigated two concepts. First, we modeled the service time distribution of conventional PMR networks by using teletraffic data of a conventional PMR network. It is found that the density of the service time is a shifted exponential which is delayed by 0.7 second. The mean service time is about 2.5 seconds. We showed that voice call arrivals to a transmission trunked PMR network are not Poisson distributed. Analytical and simulation methods based on M/G/C (a C server queue with Poisson input and general service) models may not model the system as well as G/G/C (a C server queue with general input and general service) models. Several trunked PMR systems have been designed over the last decade, most of which have symmetric downlink and uplink channel capacities. These systems may not be spectrally efficient in case of group or broadcast-based voice and data calls, a common feature of PMR systems. Second, we studied a new asymmetric PMR system comprising of a wideband OFDM (Orthogonal Frequency Division Multiplexing)-based downlink, such as Digital Audio Broadcasting (DAB) system. We found that for 2 % GoS and a mean service time of 2.86 second, PMR users that can be supported by the proposed system is 315000.

Keywords: Private Mobile Radio (PMR), voice traffic modeling, asymmetric PMR system, voice call arrivals to a PMR system, waiting time in the queue.

ÖZET

PMR SİSTEMLERİNDE TRAFİK ANALİZİ VE MODELLENMESİ

Başak Can
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Prof. Dr. Hayrettin Köymen
Temmuz 2003

Analog PMR (Private Mobile Radio) istemlerinden dijital ve demetlenmiş PMR sistemlerine geçiş kriterlerini belirlemek için güvenilir trafik bilgisine ihtiyaç vardır. Bu çalışmada iki konu incelenmiştir. İlk olarak, analog PMR ağlarındaki kanal doluluk dağılımı analog bir PMR ağına ait olan teletrafik verisi kullanılarak modellenmiştir. Servis zamanlarının yoğunluğu 0.7 saniye gecikmiş bir üstel fonksiyon olarak bulunmuştur. Ortalama servis zamanı yaklaşık olarak 2.5 saniyedir. İletim demetli bir PMR sistemine gelen ses arama varışlarının dağılımının Poisson olmadığı gösterilmiştir. Bu sebeple, M/G/C (Poisson girişli ve genel servis dağılımına sahip C adet kanalı bulunan kuyruk) modeline dayalı olarak yapılan analitik ve simülasyon metodları, sistemi G/G/C (genel giriş ve servis dağılımına sahip C adet kanalı bulunan kuyruk) modeli kadar iyi modelleyemez. Son on yıl içerisinde, bitakım demetlenmiş PMR sistemleri dizayn edilmiştir. Bunların birçoğu simetrik yer-uydu bağı ve uydu-yer bağı kanalı kapasitesine sahiptir. Bu sistemler, grup veya yayın konuşmaları ve veri iletişimi göz önüne alındığında spektrum verimliliği açısından etkili olamayabilirler. İkinci bir çalışma olarak, geniş bantlı OFDM (Orthogonal Frequency Division Multiplexing)'e dayalı bir uydu-yer bağından oluşan bir asimetrik PMR sistemi inceledik. OFDM'e dayanan uydu-yer bağı olarak Dijital Ses Yayını sistemi göz önüne alınmıştır. 2% hizmet niteliği ve 2.86 saniyelik bir ortalama konuşma süresi için sistemin destekleyeceği kullanıcı sayısı 315000 olarak saptanmıştır.

Anahtar Kelimeler: PMR, ses trafiğinin modellenmesi, asimetrik PMR sistemi, ses arama varışları, kuyrukta bekleme süresi.

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Dr. Hayrettin Köymen for his instructive comments in the supervision of the thesis.

I would like to express my special thanks and gratitude to Assoc. Prof. Dr. Nail Akar, Prof. Y. Ziya İder, Assoc. Prof. Dr. Billur Barshan and my colleague Ersin Şengül.

I would like to express my thanks to my family for their endless love, trust, encouragement, and support throughout my life.

Contents

List of Figures

List of Tables

To My Family...

Chapter 1

Introduction

Today's Private Mobile Radio (PMR) networks must cope with ever increasing traffic inflation by the high capacity and reliable systems. The demand is increased by many different factors such as the rapid growth of the voice, data and private networking services. All of these factors increase the need for bandwidth in PMR networks. However the conventional analogue PMR networks have a very limited traffic capacity. Therefore migration from analogue to digital PMR systems is the major issue to be considered. The critical concept at this point is to accurately model the traffic parameters of PMR networks.

In this research, we investigated two major concepts. First, we modeled the channel holding time distribution of conventional PMR networks by using teletraffic data of a conventional PMR network. Second, we propose a new asymmetric PMR system comprising of a wideband OFDM (Orthogonal Frequency Division Multiplexing)-based downlink. Digital Audio Broadcasting (DAB) system is considered as OFDM- based downlink.

In section 1.1 we are discussing the reasons of migration from analogue PMR systems to Digital PMR systems. In sections 1.2 and 1.3 we are giving an

introduction to PMR networks and DAB System respectively. Section 1.4 outlines the work that has been carried out in this research.

1.1 Migration from Analogue to Digital PMR

A June 2001 study by the market research company IMS estimated the number of PMR users worldwide to be 33.1 million. Of these users, 18.9 million are in the Americas, 7.3 million Asia, 6.2 million in Europe and 0.8 million in the Middle East and Africa [1].

In terms of technology, approximately 77% of current users (or approximately 25.5 million users) rely on analogue systems, with the other 23% using one of the digital technologies. Outside America, only 7% of PMR users use digital PMR [1]. This confirms that PMR, particularly analogue PMR, offers a unique blend of cost effectiveness, reliability and features (such as group calls and press-to-talk), which will allow it to remain the preferred technology for many users in the foreseeable future. With analogue technology, users can retain their existing systems while providing basic added-value data services such as vehicle location and simple messaging. With increasing market demand for data applications the analogue service will be too limited. The data requirements of PMR users can be summarized as follows [1]:

- Multiple applications need to be supported;
- Applications should be a combination of symmetrical and asymmetrical up-and-down links in data transport demand;
- A minimum usable data rate of 50 to 150kbit/s is required at cell edges, with greater data rates being made available as signal levels increase;
- Terminals offering both voice and high speed data are preferred;
- Voice must have priority over high speed data during busy periods.

Digital technology is expected to capture all the PMR market in the future. However the pace of the migration from analogue to digital is proving hard to predict. The global market for digital is predicted to grow from 45% of total PMR use in 1999 to 87% in 2004, with the analogue (radio and infrastructure) share of the market declining by 10.7 % per year [1]. Overall migration to digital is taking place more slowly than manufacturers anticipated.

While the technical advances and supplementary services of digital PMR are clear, many users do not think that they are indispensable due to the following reasons [1]:

- Cost: An analogue PMR system currently costs substantially less to buy than a digital system.
- Data: Analogue systems' increasing ability to handle more complex mobile data applications could mean that users will not need to replace their systems with digital versions. Analogue radios have reached a maximum data rate of 12 kbit/sec. in a 12.5 kHz channel, which is unlikely to increase in the near future. In contrast, digital systems currently offer similar and higher data rates (up to TETRA's 28 kbit/sec.), alongside other desirable features.
- Reliability: Analogue PMR has a proven reliability record. This is crucial for some types of user (typically police forces and emergency services) who demand high reliability in "life and death" situations.
- Voice: The most important feature of PMR is voice quality. While, voice on analog PMR sounds natural, digital PMR makes voices sound more mechanical. This is due to the compression applied to voice. With digital PMR, voice clarity does not generally degrade but the voice will suddenly cut out when the user is out of the radio coverage area. However, with analogue PMR, the voice quality degrades gracefully with introduced static noise as the user moves out of the radio coverage area.

Figure 1.1 shows the digital PMR penetration on the world regions. While digital systems account for roughly a third of American user base, the number of digital users remains extremely low in Europe and Asia, as shown in Figure 1.5 [1].

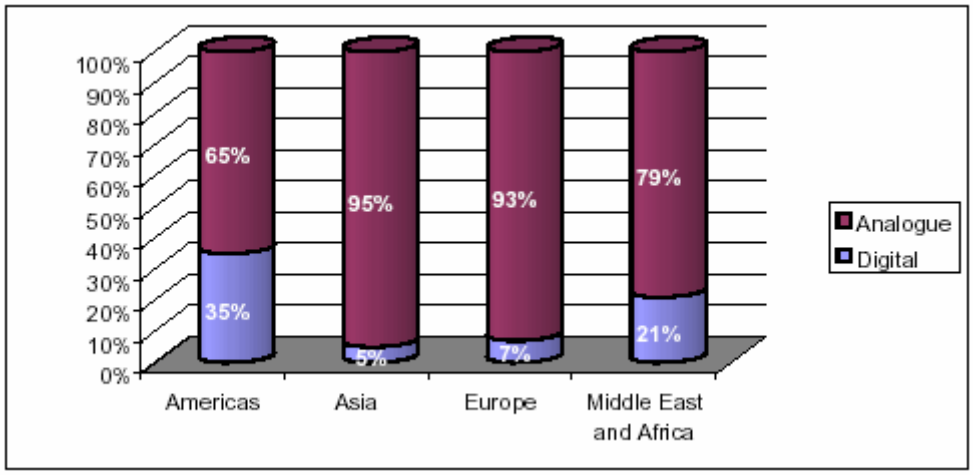


Figure 1.1 Breakdown of PMR users by technology [1].

Conventional PMR systems are non-trunked systems. However they can operate in trunked mode. Trunked analogue PMR systems are not commonly being used. In conventional PMR systems, there is only one channel for communication. Only voice communication and a limited data rate are supported. Digital PMR networks can provide data rates up to 28 kbit/sec and they can operate in trunked mode which further increases the traffic capacity. The critical point in terms of meeting the required GoS (Grade of Service) is the allocation of frequency channels to each base station. Careful dimensioning of the digital PMR network and the underlying teletraffic analysis plays a major rule in determining the various GoS parameters that service providers can provide at various network loads. Channel holding (occupancy) time is a very important parameter in analyzing mobile communication networks.

1.2 An Introduction to PMR Networks

PMR systems are systems set up by a company or group of users to provide mobile radio services for that group of users alone. In this way they differ from public cellular mobile systems. PMR user groups own their network of terminals and base stations. The PMR system may be restricted to a specific area, e.g. factory area or town. Users have their own channels –another PMR system uses a different set of frequency channels.

PMR users may group together to run joint systems, or have such systems run for them, in so called public access mobile radio (PAMR). PAMR systems or PMR systems with a common standard and interworking arrangement have the advantage of allowing users on different PMR systems to communicate with each other directly [2].

The simplest communication method achieved by PMR systems are "Walkie- Talkie" method where two parties can talk with each other without any base station or controlling network intervention. By this method they can talk with each other by simply pushing a Push To Talk (PTT) button. They must be within the coverage area of each other and calls to other networks are not possible.

There are a wide variety of users of PMR systems. Such users can be grouped as [2]:

- Public safety: emergency services: (police, fire, ambulance, mountain rescue, etc.)
- Non safety national government: Other governmental agencies, such as non-emergency health, customs, etc.
- Non-safety local government.

- Transport: railways, buses, taxi, etc.
- Other utilities: water, electricity, gas, coal.
- On-site PMR: general purpose businesses operating in local areas or within their own premises.
- Other PMR: operating over larger areas
- PAMR.

PMR systems provide their users a reliable communication. Some key requirements of PMR systems are [2]:

- Many PMR systems are used in safety critical systems. For users of such systems, the advantage of being a customer of a PMR system is that a reliable communication independent of other operators can be established.
- PMR systems should provide speech and data transmission capability. The examples of such data services are, transmitting medical telemetry, including still images, video, text messages and GPS data to assist in rescue operations.
- PMR systems should provide centralised and decentralised operation. In many businesses, PMR is used to organise users, and a central dispatch point is therefore required. However, users may want to communicate with each other in the absence of a centralised controller or any infrastructure at all.
- Group calls and point-to-point calls are essential for PMR users. Group calls where the call includes predefined users and broadcast calls where the call includes all terminals in addition to point to point calls are essential for PMR users.
- The most important feature of PMR systems is fast call setup. Rather than dialling a number to set up a call, with the called party answering a phone, PMR systems usually have a pressel or PTT button to activate a call to the dispatcher or user group, with the receiving terminal

annunciating the message without an answering procedure. Some calls therefore may consist of a couple of words. Fast call set-up is especially important in emergency services.

- Many PMR users have a requirement of high level of security. Protection of the transmitted information from tampering and interception and reliability of the system are such security requirements.
- PMR operators may wish to be able to differentiate between users to give different call priorities or quality of service to different user or call types. For example, an emergency call may pre-empt other call types to gain access to the network.
- PMR systems should also be able to communicate between other networks. This may be due to equipment equipment replacement cycles or regulatory restrictions. Also, communication with public telephony networks or data networks may be essential.

PMR Configurations

There are some basic PMR configurations which are commonly being used [2]. One of the most common PMR configurations is the dispatch operation which is shown in Figure 1.2. At least two channels are used, one for uplink and one for downlink. All terminals can receive the downlink transmissions by the dispatcher. Point to point communication is also possible. The messages from the mobile stations can only be received by the dispatcher.

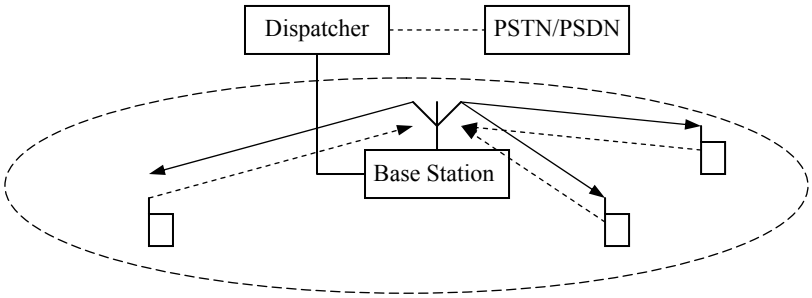


Figure 1.2 Dispatch mode PMR configuration

To extend the coverage area of mobile stations, a base station can be connected as a repeater. The base station retransmits the messages on the downlink. This extends the coverage area of mobile stations to that of base station. In Figure 1.3, mobile 1 can communicate with mobile 3 through the coverage area of base station.

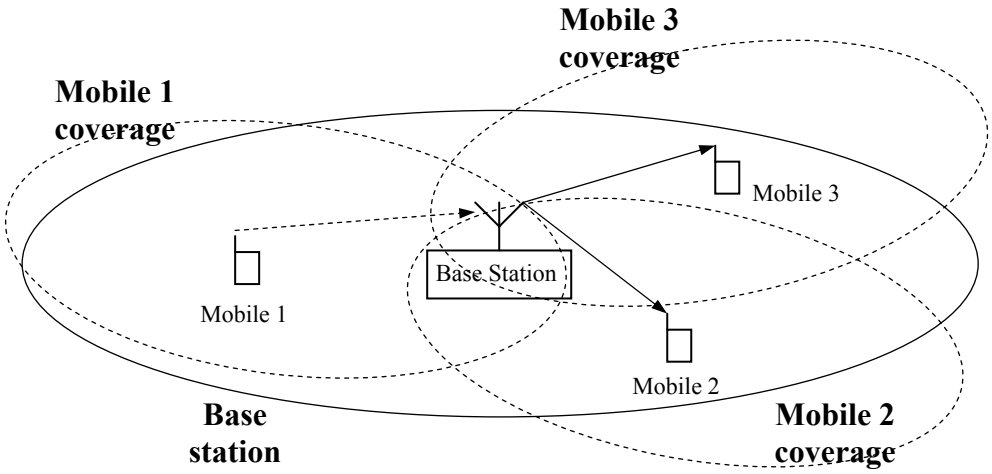


Figure 1.3 Talkthrough repeater operation

In some circumstances, a single base station may not be able to cover the entire service area, i.e. due to a shadow of a building. If the uncovered area is limited to a relatively small area, a radio port can be used to provide required coverage at this area (Figure 1.4).

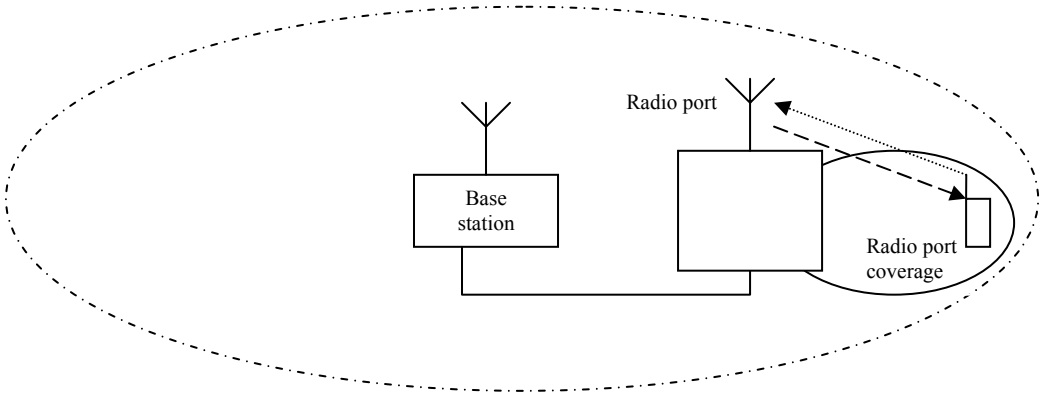


Figure 1.4 Using a radio port to illuminate uncovered service areas

Hand-held terminals usually have lower power than mobile terminals mounted in vehicles. Therefore mobiles have greater coverage areas than hand-helds. Portable vehicle-mounted repeaters can therefore be used to provide higher coverage area to users working near to their vehicles. This mode of operation is commonly used by the emergency services (Figure 1.5).

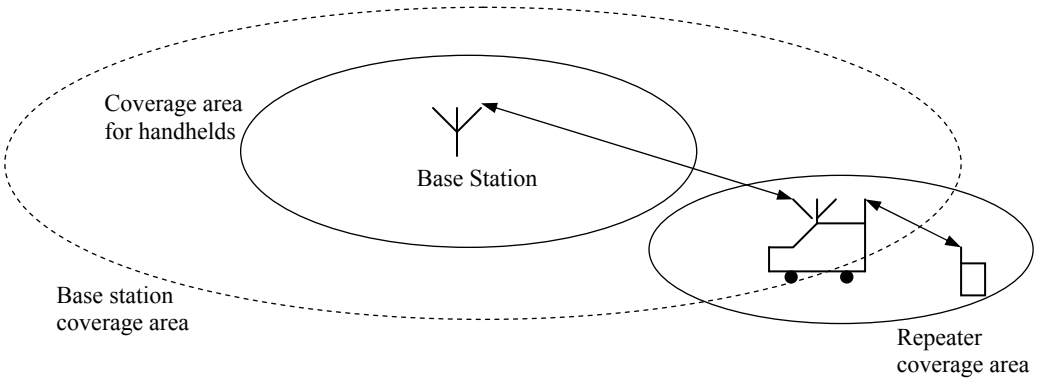


Figure 1.5 Vehicle mounted repeater for local hand-held coverage

There are seven well known PMR systems which are; Terrestrial Trunked Radio System (TETRA), Association of Public-Safety Communications Officials (APCO) 25, Integrated Dispatch Radio System (IDRA), Digital Integrated Mobile Radio System (DIMRS), TETRAPOL system, Enhanced Digital Access Communications System (EDACS),

Frequency Hopping Multiple Access System (FHMA). In this research , we consantrated on TETRA. The details of the TETRA system is given in the Appendix A.

1.3 A Brief Overview of the DAB System

The DAB system has been developed within the European Eureka 147 Project [3] and standardized by the European Telecommunications Standards Institute (ETSI), providing the means to broadcast high quality audio services to the users. Through the use of digital communications, the DAB system provides important potential benefits and exciting opportunities to the operators and users. Some of these benefits and opportunities are listed below [3]:

- *Reliable interference-free reception*: The DAB system is resistant to the effects of multipath propagation and interference, which would degrade reception of existing analogue services.

- *Efficient use of the limited RF spectrum*: A number of different services can be supported within the same RF bandwidth. The use of single-frequency networks (SFNs), where all the transmitters operate on the same radio frequency, allows a radio signal to convey a multiplex including audio, multimedia or voice call services, using a single frequency, throughout a large geographical area. SFNs are highly efficient in terms of the utilizing the scarce spectrum that is available for broadcasting. Since all the transmitters for a given multiplex operates on the same frequency, there is no need for a user re-tune its radio while on the move in order to follow a service within the service area of the broadcaster.

- *Flexibility and choice*: A wide range of services can be supported by the system, including audio based services; existing services with improved

features, such as text information, graphics or multimedia; entirely new services, such as independent data services, multimedia applications.

The DAB system contains three main signal-processing sub-systems [3]:

1. **Audio Coding:** Advanced audio compression techniques are applied.

2. **Transmission Coding & Multiplexing:** Data for a service, audio or independent data, must be combined into a single data stream for transmission. This is known as multiplexing. The configuration and content of the multiplex is totally under control of the broadcaster and re-configurations can occur when required. The frame based multiplex contains three distinct elements:

   I. The Synchronization Channel conveys the frequency and timing information to allow receivers to synchronize and to decode the DAB signals.
   II. The Fast Information Channel (FIC) carries the Multiplex Configuration Information (MCI) and Service Information (SI), which describes the configuration within a multiplex, and informs receivers how to extract and decode the information for the service selected by the user.
   III. The Main Service Channel (MSC) comprises audio frames or data packets of the services within the multiplex. This part of the multiplex carries essentially the useful payload of the DAB multiplex.

3. **COFDM modulation:** The Coded Orthogonal Frequency Division Multiplexing (COFDM) is a spectrally efficient multi-carrier digital modulation scheme, offering reliable reception under hostile reception conditions such as multipath propagation [4]. In addition, COFDM allows the intentional multipath, where all the transmitters within the SFN

operate the same frequency. These multipaths are constructive within the Guard Interval [4].

Before transmission, convolutional forward-error-correction, channel coding and bit interleaving are applied to each service to provide strong protection against bit-errors. This is important in the environment where interference and multipath propagation would lead to both flat and frequency selective fading of the received signal spectrum.

Especially MSC of DAB transmission frame in Figure 1.6 is of our concern. Since MSC carries the useful payload including the services supported by the system, the whole capacity of MSC is considered to be used for voice communications in the proposed PMR approach.

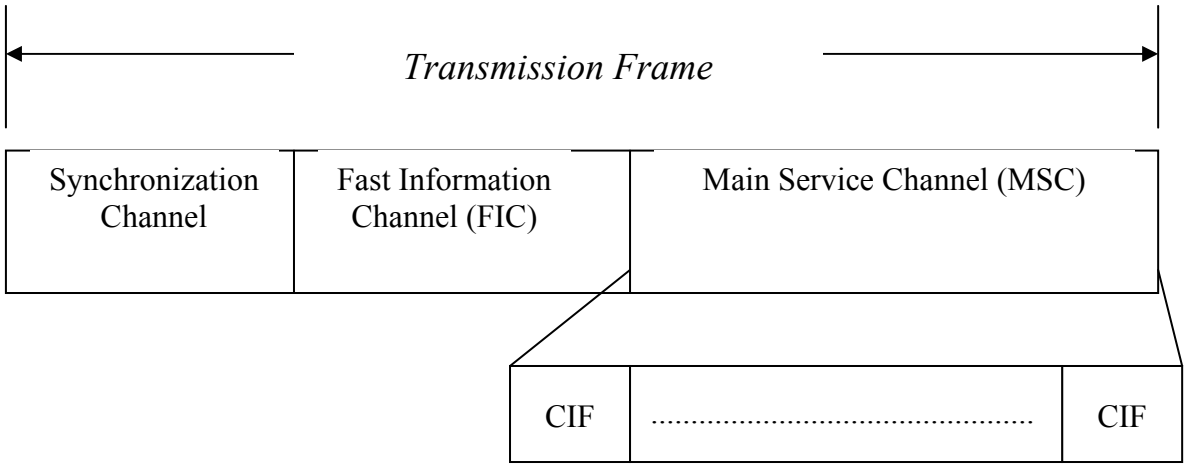


Figure 1.6 The structure of the DAB transmission frame

Totally 55296 bits are transmitted at every 24 ms [3] in the MSC. In proceeding sections, the word “slot” refers to duration of 24 ms.

1.4 Outline of this Work

We investigated two major concepts. First, we modeled the channel holding time distribution of conventional PMR networks by using teletraffic data of a conventional PMR network. Several trunked PMR systems had been designed over the last decade, most of which have symmetric downlink and uplink channel capacities. These systems are less efficient spectrally in case of group or broadcast-based voice and data calls. As a second study in this research, we propose a new asymmetric PMR system comprising a wideband OFDM-based downlink and a narrowband uplink, which not only achieves a better spectral efficiency but also can support high bit rate multimedia applications. The system is shown to have high trunking efficiency since all users are assumed to use the pool of channels available in the wideband downlink. We studied the performance and capacity of a PMR system using a DAB downlink. In particular, we studied the efficiency of such a system for voice calls using voice activity detection and statistical multiplexing. Moreover, we show that the efficiency of the system can significantly increase if the incoming calls, which can not find an available channel, are allowed to wait a certain amount of time before occupying a channel.

In Chapter 2, a model for the channel holding time distribution of a conventional PMR network is presented. Then an analytical model for interservice times for a single server queue of a transmission trunked PMR Network with Poisson input and general service pattern is created in Chapter 3. We analyzed a new asymmetric PMR system comprising a wideband OFDM-based downlink and a narrowband uplink in Chapter 4.

Chapter 2

Modeling Channel Holding Time Distribution for PMR networks Using Phase Type Service Distributions

Approximately 77 % of current PMR users rely on analogue systems [1]. Since the technical advances and supplementary services of digital PMR systems are better than that of analogue PMR systems, digital technology is expected to capture all the PMR market in the future. Some improvements to be achieved with the digital technology are:

- Technical improvements in the hardware of terminals to provide voice and high speed data
- Increasing ability to handle more complex mobile data applications with the digitalization of the system
- Increasing ability in confidentiality features
- Growth of the user population
- More efficient use of radio spectrum via resource sharing: Trunked systems

This last feature allows more traffic to be carried in a limited radio spectrum. Careful dimensioning of the network and the underlying teletraffic analysis

plays a major role in determining the various Grade of Service (GoS) parameters that trunked networks can provide at various network loads. The channel holding time of a cell is one of the major parameters that needs to be accurately modeled in the teletraffic analysis. To dimension the number of radio channels needed in a trunked system, the basic methodology is based on exponentially distributed channel holding times. The exponential distribution often used to model PMR systems is different than what is observed in our study. In this chapter a statistical model of the channel holding time in PMR systems is presented. We modeled channel holding times by phase type distributions. We made use of data acquired from a conventional PMR network.

Section 2.1 introduces phase type distributions. Section 2.2 describes the methods that we used to fit a phase type distribution to PMR teletraffic data. Section 2.3 describes the extraction of the teletraffic data that we used in our analysis carried out in this chapter. Finally, in section 2.4 we modeled channel holding time distribution of a conventional PMR network under different traffic loads.

2.1 Phase Type Distributions

Phase type distributions are statistical models that provide a compact statistical description of the empirical data. One important property of phase type distributions is that they are dense and can be used to approximate any kind of renewal process on $[0, \infty)$. In some applications, the phases have no physical interpretation and the phase type modeling is purely descriptive. Modeling channel holding time data via phase type distributions allows us to understand a service, which consists of a random sequence of tasks such that a phase corresponds to a task.

Phase type distributions represent a set of distributions that are combinations and convolutions of different exponential distributions [5]. Any continuous distribution, X, on [0 ;∞), which can be obtained as the distribution of time until absorption in a continuous time finite state Markov chain (which has a single absorbing state into which absorption is certain) is said to be of phase type. The initial state may be chosen randomly and all states are transient except the absorbing state. From the above definition we can define a Markov process, $\{J_u\}$ $u \geq 0$, on the states {1,..., p, p+1} with initial probability vector (**π,** π_{m+1}**)** with **π e** + π_{m+1}=1. **e** is a column vector with all elements equal to one. Furthermore, the states 1,..., p are transient, and consequently, state p + 1 is the one and only absorbing state, regardless of the initial probability vector. Figure 2.1 helps in visualizing the system [5].

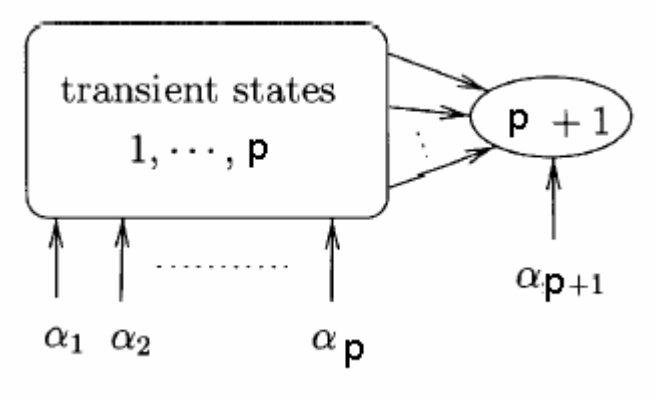


Figure 2. 1 Schematic view of a phase type distribution.

The infinitesimal generator **Q** of J_u can be written as [5]

$$\mathbf{Q}=\begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}$$

Where t_i (the ith element of **t**, the exit–rate vector) is the conditional intensity of absorption in p+1 from state i [5]. The (p x p)-dimensional matrix, **T** is called

the "phase type generator" [5]. The row sums of **Q** equal to zero, i.e., **t= - Te.** The pair (**π, T**) is referred to as a representation of the phase type distribution.

Some basic distributional characteristics of phase type distributions are [5]:

- distribution function: $F(y) = 1 - \boldsymbol{\pi} \exp\{\mathbf{T}y\}\mathbf{e}$
- density: $f(y) = \boldsymbol{\pi} \exp\{\mathbf{T}y\}\mathbf{t}$

Phase type distributions are used as statistical models due to several motivations. Since phase type distributions are combinations and convolutions of exponential distributions, problems which have an explicit solution assuming exponential distributions are also tractable with phase type distributions. Phase type modeling can be viewed as a semi-parametric density estimation procedure. In the next section, we describe the methods which are used to fit phase type distributions to channel holding time of a conventional PMR network.

2.2 Fitting Phase Type Distributions to Data from a PMR System

The parameters of phase type distributions are estimated via the *EM-algorithm* using the EMpht-program. S.Asmussen et al [6] present a general statistical approach to estimation theory for phase type distributions. The idea of this approach is: The class of phase type distributions may for a fixed p (the number of transient states) be viewed as a multi-parameter exponential family, provided the whole of the underlying absorbing Markov process is observed. The EM (expectation-maximization) algorithm is an iterative maximum likelihood method for estimating the elements of (***π, T***), the parameters of the phase type distribution [7]. The performance and the dynamics of the algorithm are illustrated in S.Asmussen et al [6] by a sequence of fits of phase type

distributions to three different theoretical distributions: Weibull, Log-normal and Erlang distribution with feedback.

EMpht calculates estimates of the elements in ($\boldsymbol{\pi}$, $\mathbf{T}$), the parameters of the chosen phase type distribution, for a fixed order p given by the user [7]. Starting with initial values ($\boldsymbol{\pi}^{(0)}$, $\mathbf{T}^{(0)}$) , which are either provided by the user or randomly generated in EMpht, the programme produces a sequence of parameter estimates ($\boldsymbol{\pi}^{(0)}$, $\mathbf{T}^{(0)}$), ($\boldsymbol{\pi}^{(1)}$, $\mathbf{T}^{(1)}$),……., ($\boldsymbol{\pi}^{(N)}$, $\mathbf{T}^{(N)}$). Each of these estimates corresponds to one iteration of the EM-algorithm, which implies that the likelihood function increases in every iteration. That is, if we use EMpht to fit a phase type distribution to a sample $\mathbf{y}= (y_1,\ldots, y_n)$, we can be sure that each new estimate is better than the previous one in the sense that

$$L(\boldsymbol{\pi}^{(k)}, \mathbf{T}^{(k)}; \mathbf{y}) \leq L(\boldsymbol{\pi}^{(k+1)}, \mathbf{T}^{(k+1)}; \mathbf{y}).$$

Where

$$L(\boldsymbol{\pi}, \mathbf{T}; \mathbf{y}) = \prod_{i=1}^{n} \boldsymbol{\pi} e^{\mathbf{T} y_i} \mathbf{t}$$

is the likelihood function [7].

In the EMpht program the user can choose among five different structures. We used the ones which are the most common phase type distributions. The phase type distributions that we used to model the channel holding time distribution with the EM-algorithm are [6]:

- Hyper exponential: Hyperexponential distribution consists of a finite mixture of exponentials. The Markov process may start in any state with probability π_i and is absorbed without visiting any other state.

- Sum of Exponentials (General Erlang): The Markov process starts in state one with probability 1 and makes a transition only from state *i* to *i*+1. It is absorbed from state p.

- Coxian: The Markov process starts in state one with probability 1 and it can be absorbed from any state.

In the next section we describe methods that we used to extract the channel holding times data that we modeled with phase type distributions.

2.3 Extraction of the Channel Holding Time Data

Recorded voice traffic of three individual channels of a conventional PMR network is investigated. Call durations and interservice times of calls which acquired services on three successive days are extracted. They are extracted from the carried traffic on the network. A total of 7348 wave files of voice recordings are processed via a voice activity detection algorithm working with Matlab. The voice records are imported to Matlab via "wavread" command. Therefore the amplitude values are in the range [-1, 1]. These records of three channels are separately studied. These three channels are for separate services.

2.3.1 Voice Activity Detection Procedure Applied to the Voice Records

Step 1) Hiss removal

Hiss noise is that "Shhhhh" sound in the background of every recording. It usually has an amplitude limited to the interval [-0.01, 0.01] (sometimes worse, sometimes better, but it is always there). For dealing with this type of noise,

voice amplitudes in the [-0.01, 0.01] interval are set to zero. However, if the hiss noise in a recording has a greater amplitude than [-0.01, 0.01] band, then it may seem to be part of voice. We illustrated this problem in Figure 2.2. In the (9, 11) sec. interval there are two FM noises of duration smaller than 0.3 second and there is hiss noise (with an amplitude in the [-0.02, 0.02] band) in between. The code can not remove this hiss noise and it seems to be part of a voice conversation. Code erroneously assumes that there is a voice call of duration 1.1 seconds in the (9, 11) sec. interval. To overcome this problem, amplitudes in a greater band, that is [-0.02, 0.02] interval can be set to zero. However, this is a more dangerous action than setting the voice amplitudes in the [-0.01, 0.01] interval to zero, because important part of voice records may be lost leading to a smaller or zero call duration.

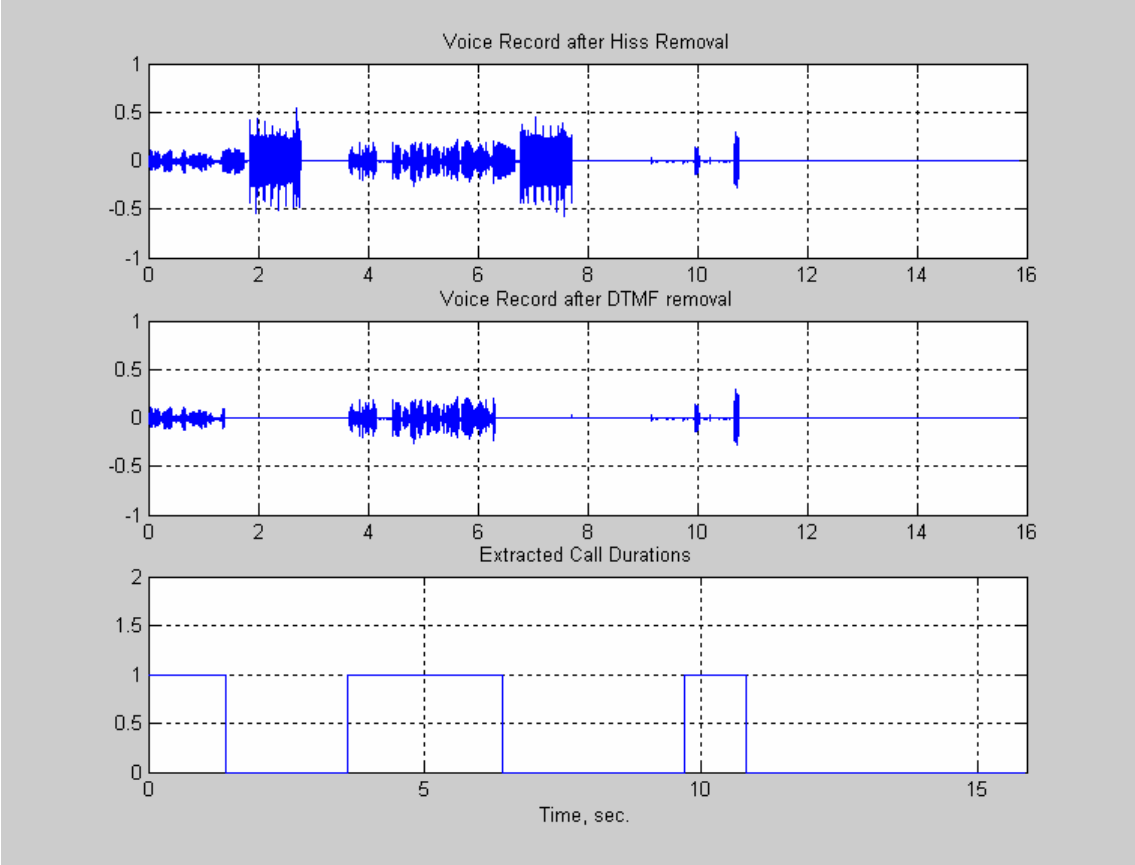


Figure 2. 2 Illustration of hiss noise with an amplitude greater than [-0.01, 0.01] band.

Step 2) DTMF (Dual Tone Multiple Frequency) removal

Some of the mobile stations have a DTMF code, some of them do not have. When a person with a mobile station which has a DTMF code completes his talk, a DTMF code appears at the end of the talk. As an example, see Figure 2.3. The wave plotted is the voice record obtained via wavread command of MATLAB. Therefore the amplitude of the wave is bounded in the interval [-1, 1].

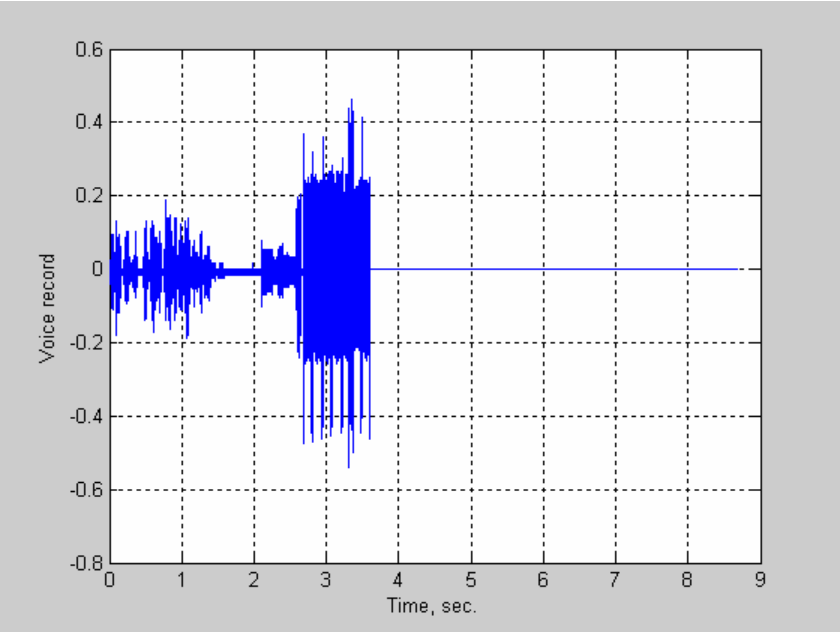


Figure 2. 3 DTMF code after talk

In Figure 2.3, the speech is within [0, 1.5] sec. interval and corresponding DTMF code is within (2, 4) sec. interval. The DTMF codes should be removed from the records to make call durations precise. Otherwise they seem to be part of voice. For this purpose, we filtered out the DTMF frequencies. After that operation, we recognized that there is also noise embedded with them (Figure 2.4).

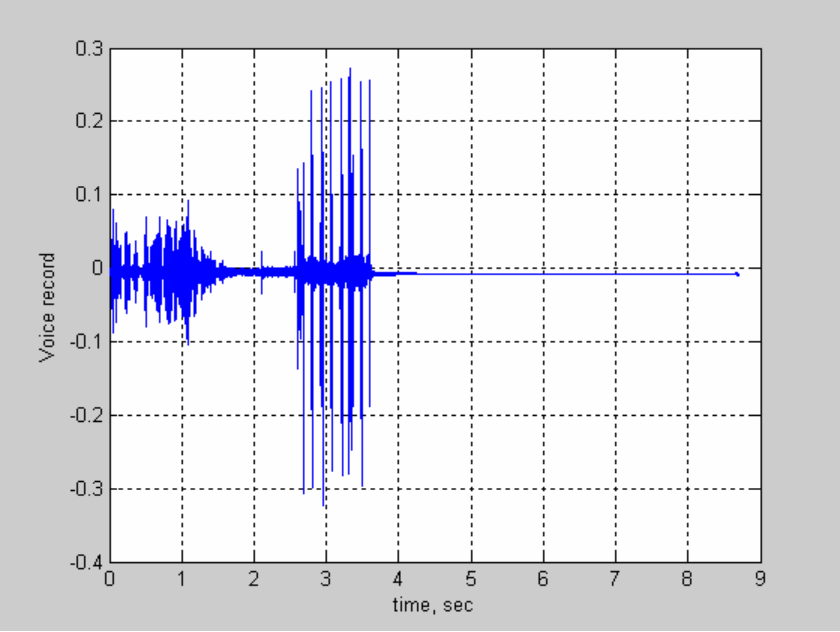


Figure 2. 4 Voice record after filtering DTMF frequencies

As seen in Figure 2.4, the wave in the interval (2, 4) sec. is not zero due to noise embedded with the DTMF code. Therefore filtering out DTMF frequencies did not satisfy our purpose. As another solution we used DTMF detection with threshold technique. As you see on Figure 2.3, the DTMF code which is within (2, 4) sec. interval is stronger than the voice and it starts with a strong peak. This peak is greater than 0.35. The observed DTMF codes have a duration of 0.91 second. The algorithm determines the first time instant where the record amplitude is greater than 0.35. Let us call it t_o. Then the voice record is set to zero within the interval [t_o – 0.47 sec, t_o + 0.91 sec]. The interval starts at t_o – 0.47 sec., because before each DTMF code a 5 tone selective call of duration 0.47 second is present. We should remove it as well.

There is a problem encountered with this technique of DTMF removal. A great majority of speech wave amplitudes are smaller than 0.35. However, some of them are not. Therefore some of the actual speech waves are erroneously set to zero. But probability of this event to occur is very low.

Step 3) Removal of the noise

The recorded wave files of each channel are obtained via FM demodulating the corresponding carrier. If there is noise present, FM receiver turns on an FM mute switch. However, it can detect the presence of noise after about 0.3 second Therefore there remain some noise records of duration about 0.3 second. As an example, see Figure 2.5. Such remained noises are in (0.2, 0.5) sec., (7, 7.5) sec. and (7.5, 8) sec. intervals.

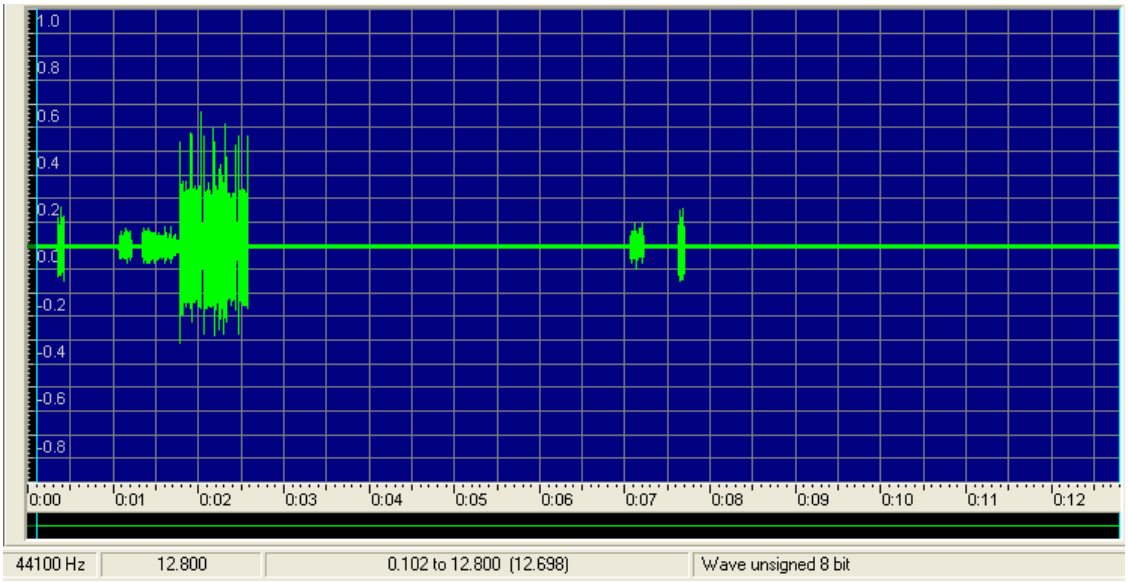


Figure 2. 5 Illustration of noise remained just before FM mute switch is turned on

There are also other noises of duration greater than 0.3 second. The largest duration of them is 0.6 second. According to our analysis carried over 567 messages taken from Channel 1 records, minimum duration of a call is 0.7 second (section 2.4). Therefore, if a detected call duration is smaller than 0.7 second, then it is ignored.

Step4) Extracting Call Durations and Interservice Times:

Interservice time is the time passed between the instants of time at which two successive customers enter service. Also, since the data belongs to a transmission trunked network, call duration refer to the duration of a talk which belongs to one user and each individual talk represents a call. Since the code which detects voice activity can not distinguish the voice of different users, there must be a heuristic for deciding whether the call is still ongoing or a new call has started. Let silence interval term represent an interval in which the voice wave is zero. For this purpose, if a silence interval which is larger than 0.3 second is detected, then it is assumed that a new call has started. If a silence period which is less than 0.3 second is detected, then it is assumed that the call is still ongoing. It is observed in the data that the next call can start only after a minimum of 0.3 second. Thus, minimum interservice time is observed to be 1 second. We make some errors if a user talks very slowly and stops speaking more than 0.3 second while holding the channel.

2.4 Phase Type Distribution Fits to Channel Holding Time of a Transmission Trunked PMR Network

In this section we fit phase type distributions to channel holding time data, which is extracted with the methods outlined in the previous section. The system analyzed is under its transmission trunking operation mode and users have their conversation time limited to 30 second. The statistical studies are based on fully empirical approach and make use of data acquired from actual working system that is considered to be sufficiently representative. Unlike analytical and simulation approaches, the empirical approach is environment dependent. The results presented would have been different if taken in a different place, time, etc.

Statistical Stopping Rule for Adding Phases

Usually, as the number of parameters in the fit is larger, the corresponding fitted model is better. In order to get a perfect fit of phase type distribution to the empirical cumulative distribution function (ECDF) we need infinitely many phases. The confidence interval for the ECDF provides a statistical stopping rule for adding phases to the fitted phase type distribution [8]. Based on the data, we have an ECDF, calculated with EMpht algorithm. We use Kolmogorov- Smirnov Goodness-of-Fit test to decide the accuracy of the estimator. The Kolmogorov-Smirnov (K-S) test is based on the ECDF. Given N *ordered* data points $Y_1, Y_2, ..., Y_N$, the ECDF is defined as

$$E_N = n(i)/N$$

Where $n(i)$ is the number of points less than Y_i. The K-S test is based on the maximum distance between the fit and ECDF. The Kolmogorov-Smirnov one sample test statistic is defined as

$$D = \max_{1 \le i \le N} | F(Y_i) - \frac{n(i)}{N} |$$

$$P\{D \ge D_{n,\gamma}\} \le \gamma$$

where F is the phase type fit to the data set $Y_1, Y_2, ..., Y_N$. The hypothesis regarding the distributional form is rejected if the test statistic, D, is greater than the critical value of $D_{n,\gamma}$ obtained from Table 2.1.

Significance level γ	0.20	0.10	0.05	0.02	0.01
$D_{n,\gamma}$ for $n > 40$	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Table 2.1 Critical values of the Kolmogorov-Smirnov One Sample Test statistics [8]

As the value of $D_{n,\gamma}$ gets smaller, the corresponding fit gets better. In this work we selected the smallest value of $D_{n,\gamma}$:

$$D_{n,0.20} = \frac{1.07}{\sqrt{n}} \approx \frac{1}{\sqrt{n}}$$

With these selected parameters, the true distribution is known to be within $1/\sqrt{n}$ neighborhood. Therefore, we keep adding phases until we get the fit under this resolution.

2.4.1 Statistical Results for the Busiest Hour Analysis (Manual Analysis)

The phase type model with the least number of phases is selected. This is also the selected phase type model with the least number of phases that fits into a confidence band $1/\sqrt{n}$ around the ECDF.

2.4.1.1 Statistical Results for Channel 3

The analysis is based on 556 messages taken over the busiest hour of Channel 3, which is between 20:15 PM and 21:15 PM. The data is taken by listening to

each voice call. The duration of each voice call is extracted and then its distribution is modeled via the EM algorithm.

i) *Coxian fit:*

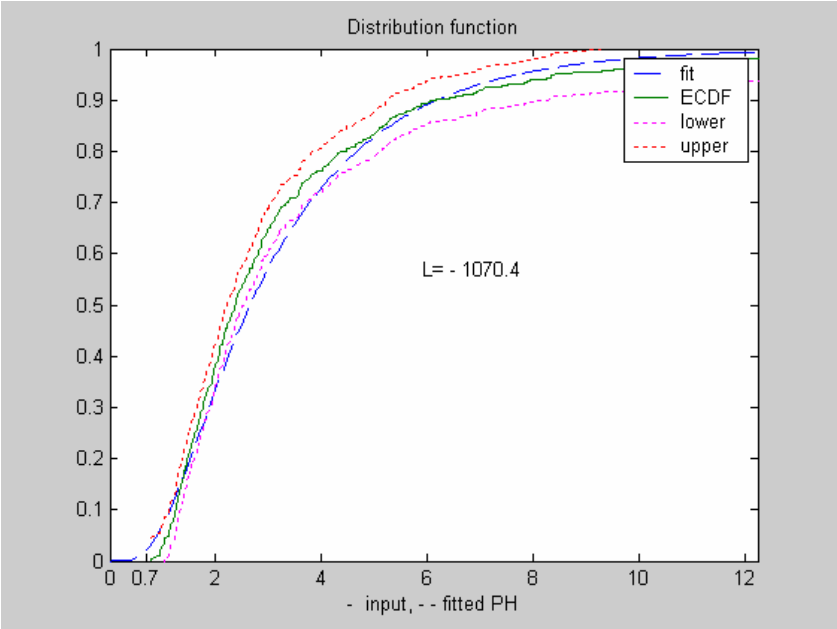


Figure 2.6 The confidence interval around the ECDF of resolution $1.36/\sqrt{n}$ with fitted phase type distribution of order 5.

As seen in Figure 2.6, the fitted CDF does not lie even in the confidence interval around the ECDF of resolution$1.36/\sqrt{n}$. Therefore we need to add more phases.

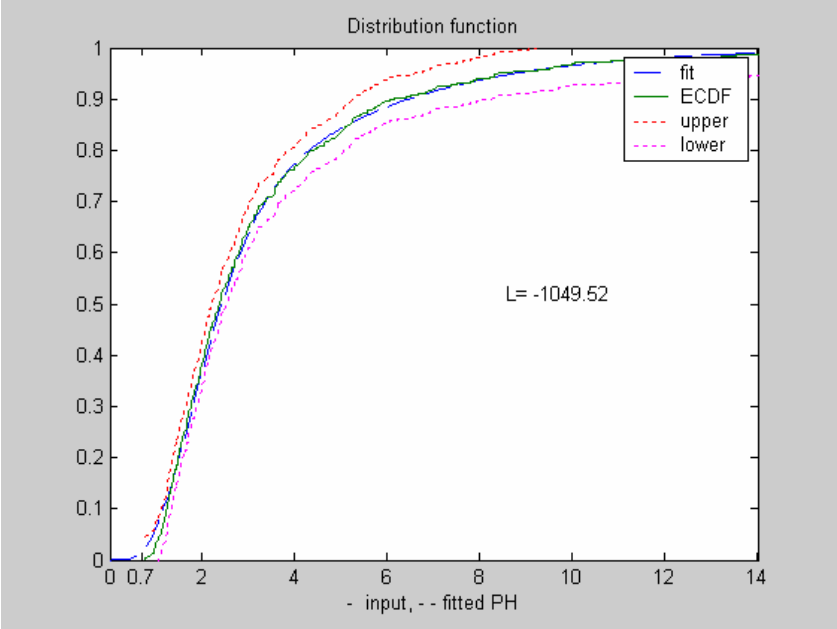


Figure 2. 7 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 7.

As seen in Figure 2.7, the fitted CDF lies in the confidence interval around the ECDF of resolution$1/\sqrt{n}$. Therefore we can stop adding more phases.

ii) *General Erlang fit:*

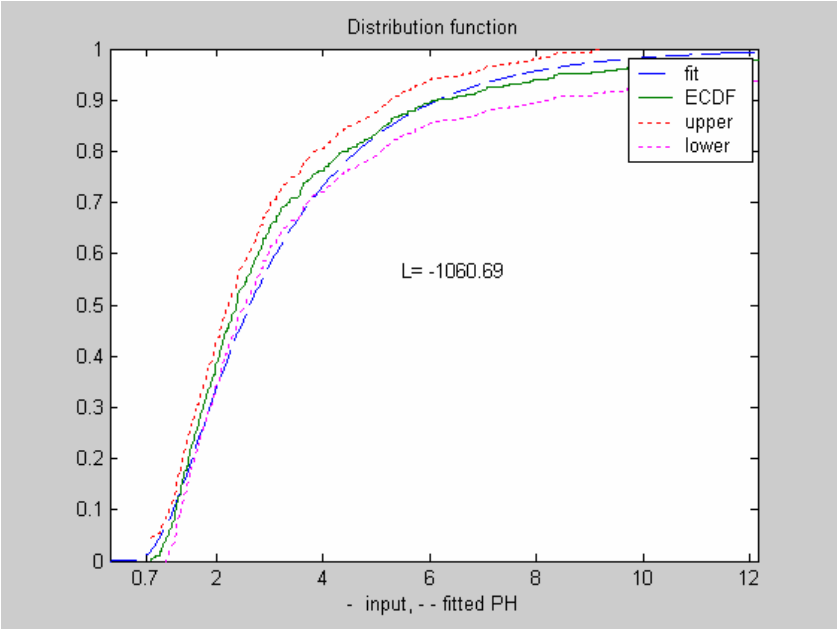


Figure 2. 8 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 7.

As seen in Figure 2.8, the fitted CDF does not lie in the confidence interval around the ECDF of resolution$1/\sqrt{n}$. Therefore we need to add more phases. As seen in Figures 2.7 and 2.8, the coxian fit has a higher likelihood function than that of General Erlang fit for the same order. This means that coxian fit is better than General Erlang fit. Therefore, we do not need to spend much time with General Erlang distribution by adding more phases.

iii) *Hyperexponential fit:*

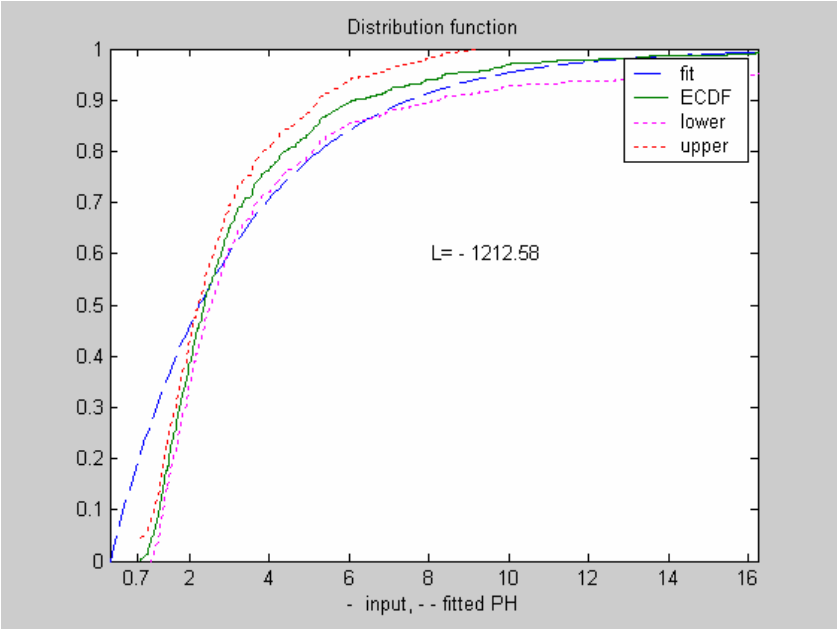


Figure 2. 9 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order > 5.

For hyperexponential fit, EMpht programme fits the same distribution with the same likelihood function even if the number of phases is increased beyond 5. As seen in Figure 2.9, the fitted CDF does not lie in the confidence interval around the ECDF of resolution $1/\sqrt{n}$. Therefore hyperexponential fit is not as good as the coxian fit for the same order.

Conclusion: As we see on Figures 2.6, 2.7, 2.8 and 2.9 we can say that the best fit is the coxian distribution with 7 phases. It also has the maximum likelihood function amongst others. The difference between the likelihood functions of coxian and hyperexponential fits is -1049.52+ 1212.58 = 163.06.
The estimated parameters of the best fit are:

$$\boldsymbol{\pi} = \begin{matrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix} \quad \mathbf{T} = \begin{matrix} -2.3092 & 2.3092 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2.3092 & 2.3092 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2.3092 & 2.3092 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2.3092 & 2.3092 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2.3092 & 0.611 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1.19 & 1.18 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.304 \end{matrix}$$

Mean =3.26 second and Standard-deviation =2.7 second.

2.4.1.2 Statistical Results for Channel 1

The analysis is based on 567 messages taken over the busiest hour of Channel 1, which is between 12:45 PM and 13:45 PM. The data is taken by listening to each voice call. The duration of each voice call is extracted and then its distribution is modeled via the EM algorithm.

i) *Coxian fit:*

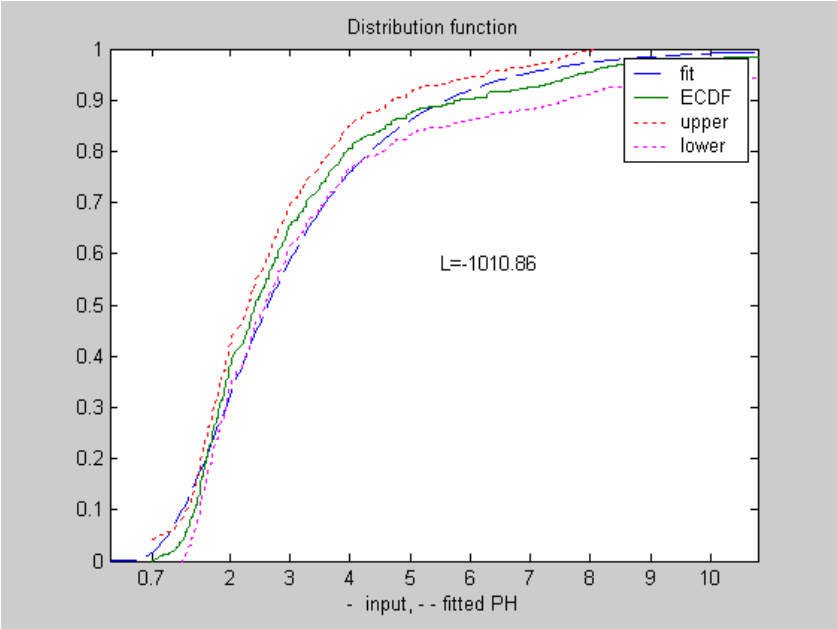


Figure 2. 10 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 5.

As seen in Figure 2.10, the fitted CDF does not lie in the confidence interval around the ECDF of resolution $1/\sqrt{n}$. Therefore we need to add more phases.

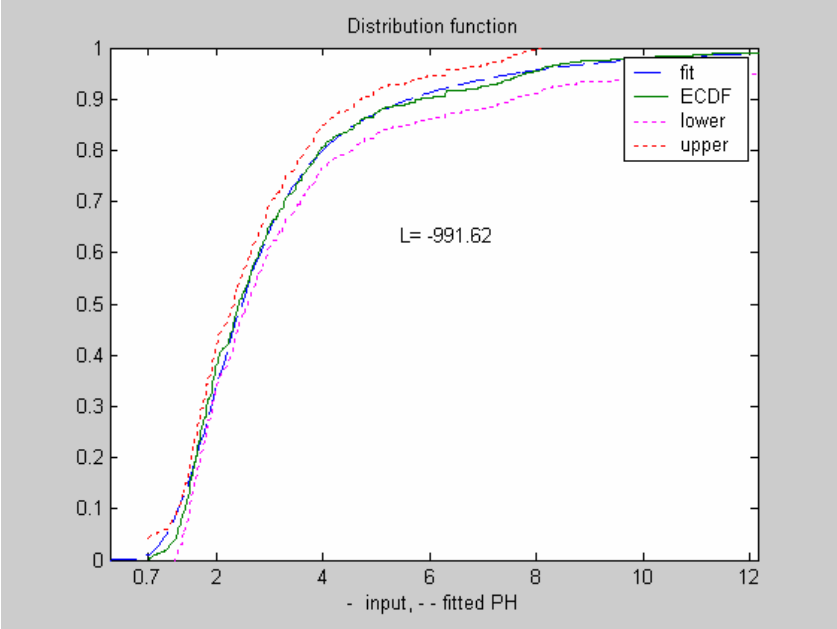


Figure 2. 11 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 7.

As seen in Figure 2.11, the fitted CDF lies in the confidence interval around the ECDF of resolution$1/\sqrt{n}$. Therefore we can stop adding more phases.

ii) *General Erlang fit:*

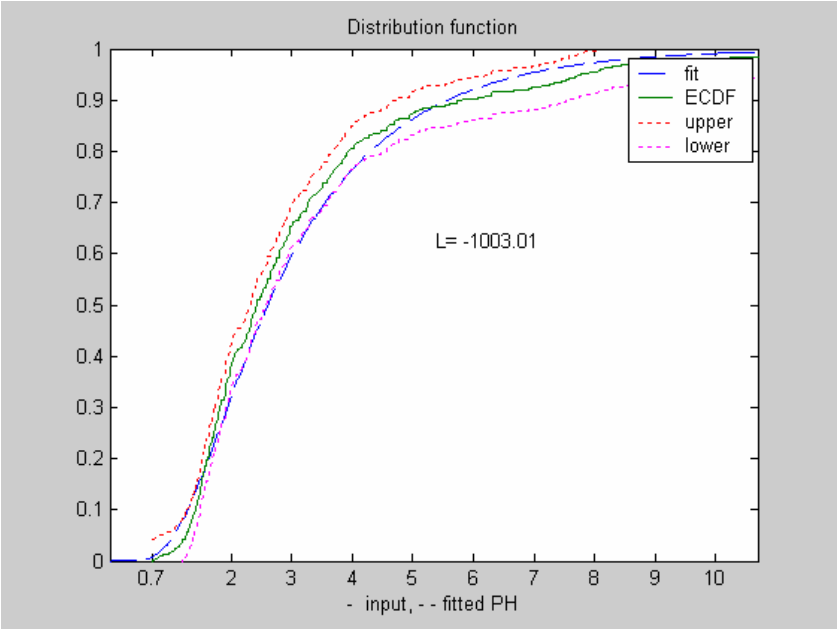


Figure 2. 12 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 7.

As seen in Figure 2.12 the fitted CDF does not lie in the confidence interval around the ECDF of resolution $1/\sqrt{n}$ at some points. Therefore we need to add more phases. Since coxian fit gives a better fit for the same order, say 7, we do not need to spend much time with the General Erlang fit by adding more phases.

iii) *Hyperexponential fit:*

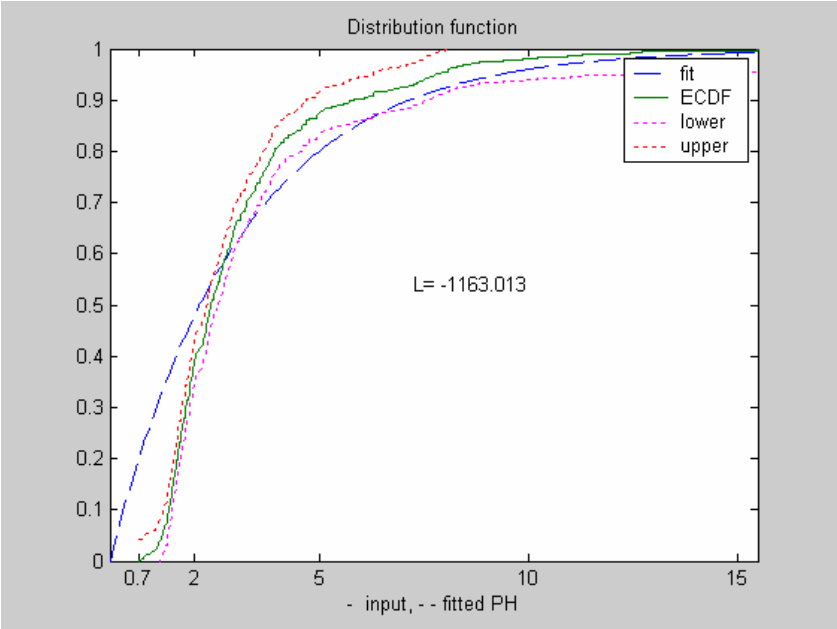


Figure 2. 13 The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order $\geq$ 5.

For hyperexponential fit, EMpht programme fits the same distribution with the same likelihood function even if the number of phases is increased beyond 5. As seen in Figure 2.13, the fitted CDF does not lie in the confidence interval around the ECDF of resolution $1/\sqrt{n}$. Therefore hyperexponential fit is not as good as the coxian fit for the same order.

Conclusion: As we see from Figures 2.10, 2.11, 2.12 and 2.13, we can say that the best fit is the coxian distribution with 7 phases. It also has the maximum likelihood function amongst others. The difference between the likelihood functions of coxian and hyperexponential fits is -991.62+ 1163.013 = 171.393.

The estimated parameters of the best fit are:

	1		-3.12	3.12	0.00	0.00	0.00	0.00	0.00
	0		0.00	-3.11	3.11	0.00	0.00	0.00	0.00
	0		0.00	0.00	-3.14	3.14	0.00	0.00	0.00
π =	0	**T=**	0.00	0.00	0.00	-3.14	3.14	0.00	0.00
	0		0.00	0.00	0.00	0.00	-3.16	3.16	0.00
	0		0.00	0.00	0.00	0.00	0.00	-3.6	1.34
	0		0.00	0.00	0.00	0.00	0.00	0.00	-0.37

Mean = 2.86 second and Standard-deviation = 2.2 second.

2.4.2 Statistical Results for the Three Successive Days Analysis

The phase type model with the least number of phases is selected. This is also the selected phase type model with the least number of phases that fits into a confidence band $1/\sqrt{n}$ around the ECDF. Cumulative Distribution Function plots present the confidence interval around the ECDF of resolution $1/\sqrt{n}$.

Statistical Results for Channel 2

The analysis is based on 3255 call durations extracted from three successive days' wave records which include all voice calls which acquired services by Channel 2 on those three successive days. Let μ_k be the detected duration of the kth call which acquired service by Channel 2. Let $\boldsymbol{\mu}$ be a *random data sequence* whose elements are $\{\mu_j, \text{j} = 1, 2, \ldots, 3255\}$.

As defined in section 2.2, minimum service time is 0.7 second. This leads to an increase in the order of the phase type distribution fitted to $\boldsymbol{\mu}$. In order to keep the order low, we tried to fit a phase type distribution to the random sequence $\boldsymbol{\mu} - 0.7$. This fit is plotted on Figure 2.14.

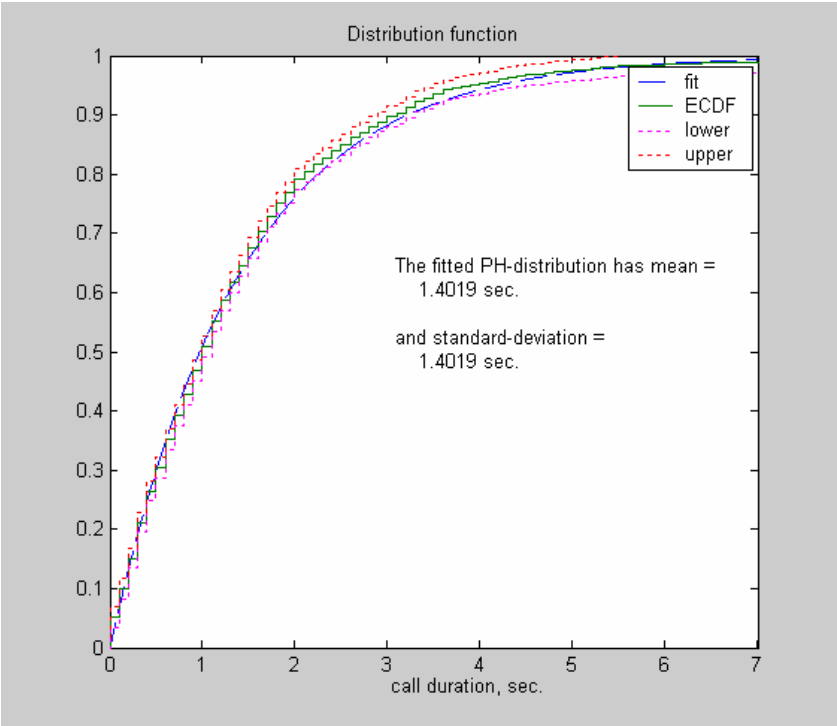


Figure 2.14 The fitted phase type distribution to the random data sequence $\boldsymbol{\mu} - 0.7$. The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 1.

Referring to this figure, we can say that the density of the random data sequence $\boldsymbol{\mu} - 0.7$ is given by,

$$0.71e^{-0.71t}\ ,\ t \geq 0 \qquad (2.1)$$

Let $f_s(s)$ be the density of $\boldsymbol{\mu}$. Using (2.1) we can write

$$f_s(s) = \begin{cases} 0.71e^{-0.71(s-0.7)} & , \ \mathrm{s} \geq 0.7\,\mathrm{sec}. \\ 0 & , \ \mathrm{s} < 0.7\,\mathrm{sec}. \end{cases} \tag{2.2}$$

Statistical Results for Channel 3

The analysis is based on 12756 call durations extracted from three successive days' wave records which include all voice calls which acquired services by Channel 3 on those three successive days. Let μ_k be the detected duration of the k th call which acquired service by Channel 3. Let $\boldsymbol{\mu}$ be a *random data sequence* whose elements are $\{\mu_j, \mathrm{j} = 1, 2, \ldots, 12756\}$.

In order to keep the order low, we tried to fit a phase type distribution to the random data sequence $\boldsymbol{\mu} - 0.7$. This fit is plotted on Figure 2.15.

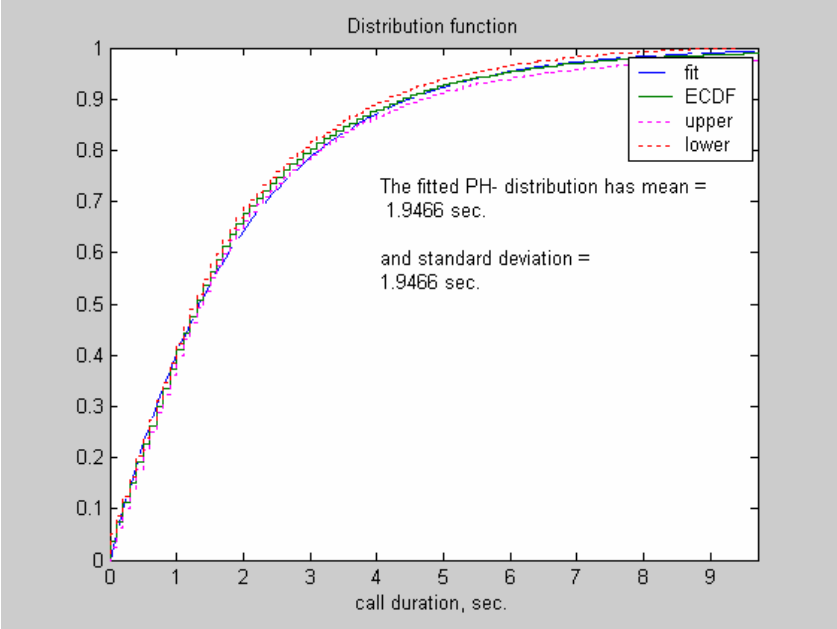


Figure 2.15 The fitted phase type distribution to the random data sequence $\boldsymbol{\mu} - 0.7$. The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 1.

Referring to this figure, we can say that the density of the random data sequence $\boldsymbol{\mu} - 0.7$ is given by,

$$0.51e^{-0.51t}, \quad t \geq 0 \tag{2.3}$$

Let $f_s(s)$ be the density of $\boldsymbol{\mu}$. Using (2.3) we can write

$$f_s(s) = \begin{cases} 0.51e^{-0.51(s-0.7)} & , \ s \geq 0.7 \text{sec}. \\ 0 & , \ s < 0.7 \text{sec}. \end{cases} \tag{2.4}$$

Statistical Results for Channel 1

The analysis is based on 15247 messages extracted from three successive days' wave records which include all voice calls which acquired service by Channel 1 on these days. Let μ_k be the detected duration of the k th call which acquired service by Channel 1. Let $\boldsymbol{\mu}$ be a *random data sequence* whose elements are $\{\mu_j, \mathrm{j} = 1, 2, \ldots, 15247\}$.

In order to keep the order low, we tried to fit a phase type distribution to the random data sequence $\boldsymbol{\mu} - 0.7$. The best fit is plotted on Figure 2.16.

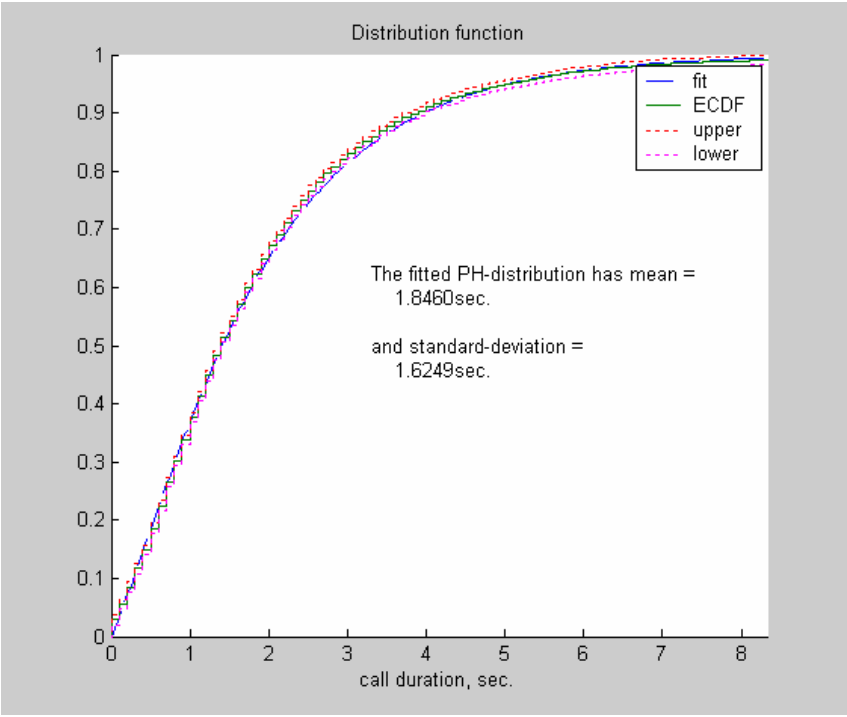


Figure 2.16 The fitted phase type distribution to the random data sequence $\boldsymbol{\mu} - 0.7$. The confidence interval around the ECDF of resolution $1/\sqrt{n}$ with fitted phase type distribution of order 2.

We tried to fit an exponential distribution to the random data sequence $\boldsymbol{\mu} - 0.7$. However it did not fit into the confidence interval of resolution $1/\sqrt{n}$. Therefore we tried another phase type distribution. On Figure 2.16 we plotted the coxian fit with two Phases. The General Phase type distribution and coxian distribution resulted in the same fit. We selected the coxian fit with 2 phases to represent the best fit. Using the parameters of the best fit, we determined the density of the random data sequence $\boldsymbol{\mu} - 0.7$ as,

$$0.87e^{-0.6512t} - 0.549e^{-1.64t}, t \geq 0 \tag{2.5}$$

Let $f_s(s)$ be the density of $\boldsymbol{\mu}$. Using (2.5) we can write

$$f_s(s) = \begin{cases} 0.87e^{-0.6512(s-0.7)} - 0.549e^{-1.64(s-0.7)}, & \text{s} \geq 0.7\,\text{sec}. \\ 0, & \text{s} < 0.7\,\text{sec}. \end{cases} \tag{2.6}$$

2.5 Discussion

In this chapter we modeled the distribution of channel holding time of a conventional PMR network under different traffic loads. Our analysis shows that channel holding time distribution of a conventional PMR network is different from exponential distribution. Minimum call duration is 0.7 second and not zero. This leads to a significant increase in the order of the phase type distribution fitted to the data. In order to model the CDF other than this fixed delay, we subtracted 0.7 second from each call duration before fitting a phase type distribution. The minimum channel holding time of 0.7 second can not be ignored since the mean channel holding time of Channel 1, 2 and 3 are 2.5 seconds, 2.1 seconds and 2.3 seconds respectively.

The channel holding time distribution of Channel 3 and 2 is a shifted exponential with a shift of 0.7 second. However the density of the duration of the calls served by Channel 1 consists of two exponentials which are each shifted by 0.7 second.

Chapter 3

A Model for Interservice Times for a Single Server Queue with Poisson Input and General Service

Call arrivals to a PMR network is one of the major parameters that need to be carefully modeled in these teletraffic analyses. A common assumption in telephony is that call arrivals are Poisson distributed and most analytical models are based on Poisson arrivals. However, for a transmission trunked PMR Network, where some calls may be correlated with each other, these models may not model the traffic accurately. Purpose of this study is to show whether the Poisson arrival assumption is true or not. The system analyzed is a conventional PMR network under transmission trunking mode. Transmission trunking is defined in Appendix A.

To characterize call arrivals that generates the offered traffic the time between call attempts is needed. This is difficult to measure since attempts are not seen when they really occur, but when the system allocates a radio channel to them. Therefore one can measure interservice times and channel holding times. Interservice time is the time passed between the instants of time at which two successive customers enter service. Since we do not have empirical observations of the time between call attempts, we could not characterize the call arrival statistics. It is a really challenging problem to characterize call

arrival statistics via interservice times. However, with interservice times data acquired from a conventional PMR network, we could be able to show whether call arrivals to a transmission trunked PMR network are Poisson distributed or not. To get closer to the statistics of the interarrival times to the system, we used the distribution and first two moments of interservice times. Let us assume first that call arrivals are Poisson distributed with rate λ. We created an analytical model for the distribution of interservice times for a single server queue. Since a conventional PMR network can be modeled with an M/G/1 queue, M/G/1 queueing model is used. M/G/1 queue is a single server queue with Poisson input and General service. Let Δ_n be a random variable which denotes the interservice times with these assumptions. Having created a model for the distribution of Δ_n, we compared it with that of empirical distribution of interservice times which are acquired from a conventional PMR network. If they are similar to each other then this shows that call arrivals are Poisson distributed. However, we observed that there is a significant difference between these two distributions. This shows that call arrivals are not Poisson distributed. We also compared the first and second moment of Δ_n with that of interservice times data obtained from a transmission trunked PMR network. If the first moments and second moments match at the same value of λ then this shows that call arrivals are Poisson distributed.

We used Kendall's notation in describing a queuing process since it is standard throughout the queuing literature. A queuing process is described by a series of symbols and slashes such as *A*/*B*/*X*/*Y*/*Z* where [9]

- *A:* Refers to interarrival time distribution (M (Markovian), *G (*General), etc.)
- *B:* Refers to service time distribution (*M* (Markovian), *G (*General), etc.)
- *X*: Refers to the number of parallel servers

- *Y*: Refers to the system capacity or queue size (finite of infinite). If *Y* is omitted then this shows that there is infinite capacity in the system.
- *Z:* Refers to the queue discipline (First in First Out (FIFO), Last Come First Served (LCFS), General Discipline (GD), etc.). If *Z* is dropped, then this shows that queue discipline is FIFO.

In section 3.1 we created an analytical model for the first two moments of interservice times. In section 3.2 we compared them with that of empirical data. To obtain information concerning the time an arrival must spend waiting in the queue, we derived the probability distribution of waiting time in the queue for M/G/1 and M/M/1 queues in section 3.3.

3.1 Interservice Time Distribution of an M/G/1 Queue

The service times are independently and identically distributed random variables with a probability distribution modeled in section 2.4.2. Any user who wishes to make a call continuously listens to the channel. He can make a call attempt when the channel is free. It takes approximately 250-300 msecond for the transmitter to reach its rated power after the PTT switch is pressed. For a conventional PMR network where a protocol channel is not available, this extra 0.3 second also occupies the channel. As discussed in section 2.3.1, service time data shows that minimum call duration is 0.7 second and not zero. Adding the 0.3 second to minimum service time modifies the service time distribution to an exponential distribution which is delayed by 1 second. Let b represent the random variable "service time". We denote cumulative distribution function (CDF) by $F_b(b)$ and the density function by $f_b(b)$. The general form of $F_b(b)$ and $f_b(b)$ can be written as

$$f_b(b) = \begin{cases} \mu e^{-\mu (b-1)}, & b \geq 1 \text{ sec.} \\ 0, & b < 1 \text{ sec.} \end{cases} \tag{3.1}$$

$$F_b(b) = \begin{cases} 1 - e^{-\mu (b-1)}, & b \geq 1 \text{ sec.} \\ 0, & b < 1 \text{ sec.} \end{cases} \tag{3.2}$$

Service time is also referred to as channel occupancy time.

In order to determine the distribution of the interservice times, we should consider two cases as shown in Figures 3.1.a and b.

1st case:

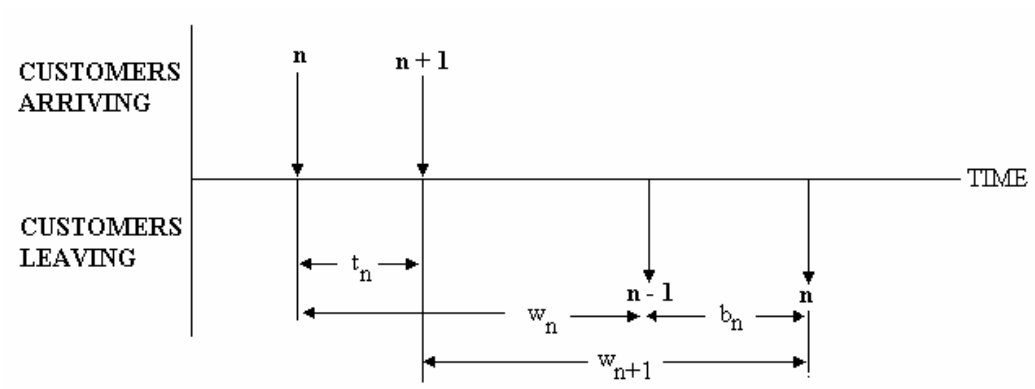


Figure 3.1.a Successive waiting times for the first case

2d case:

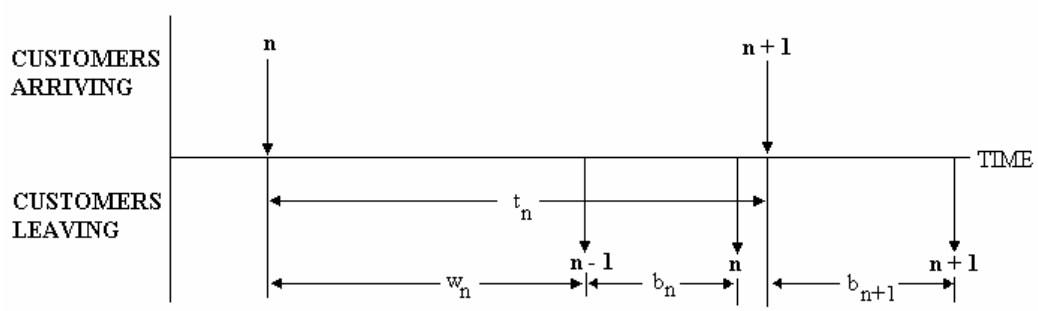


Figure 3.1.b Successive waiting times for the second case

The 1st case occurs if (n+1) st customer arrives before the service completion of the nth customer. This means $w_n + b_n - t_n > 0$, where the service time of the nth customer is b_n. t_n is the interarrival time between the nth and (n + 1)st customers and w_n is the waiting time of nth customer in the queue until his service begins. The nth user waiting in the queue immediately presses the PTT button when the $(n-1)$st customer leaves the system. The $(n+1)$st user waiting in the queue immediately presses the PTT button when the nth customer leaves the system. 2nd case occurs if (n + 1)st customer arrives after the service completion of the nth customer. This means $w_n + b_n - t_n \le 0$. The (n + 1)st customer finds the channel free when he arrives and immediately presses the PTT button. We are assuming that any user who presses the PTT button acquires service. Therefore, we observe that the waiting times, w_n and w_{n+1} of two successive customers in a single channel queue are related by the following recurrence relation [9]

$$w_{n+1} = \begin{cases} w_n + b_n - t_n & (w_n + b_n - t_n > 0) \\ 0 & (w_n + b_n - t_n < 0) \end{cases} \qquad (3.3)$$

We can rewrite (3.3) as

$$w_{n+1} = \max(w_n + b_n - t_n, 0) \tag{3.4}$$

The waiting time in the queue of the (n +1) st customer, w_{n+1} is dependent only on w_n, b_n and t_n. It is independent of b_{n+1}. Therefore, w_n and b_n are two independent random variables. Let Δ_n denote the interservice time between the nth and (n + 1) st customers. It can be verified from Figures 3.1.a and b Δ_n can be written as

$$\Delta_n = w_{n+1} + t_n - w_n \tag{3.5}$$

Substituting (3.4) into (3.5) gives

$$\begin{aligned} \Delta_n &= \max(w_n + b_n - t_n, 0) + t_n - w_n \\ &= \max(t_n - w_n, b_n) \end{aligned} \tag{3.6}$$

The interarrival times are assumed to be independently and identically distributed random variables. The random variable t_n can be denoted by t, and let $f_t(t)$ denote its density function. For a Poisson input, $f_t(t)$ can be written as

$$f_t(t) = \begin{cases} \lambda e^{-\lambda t}, t \geq 0 \\ 0 \quad , t<0 \end{cases} \tag{3.7}$$

Since the service times are assumed to be independently and identically distributed random variables, the random variable b_n can be denoted by b. Let ρ denote the utilization of the channel. It is given by $\rho = \lambda E[b]$. Where $E[b] = \int_0^\infty b dF_b(b)$. ρ <1 is a sufficient condition under which the M/G/1 has a steady state [9].

Waiting time is, for the most part, a continuous random variable except that there is a nonzero probability that the delay will be zero, that is, a customer entering service immediately upon arrival [9]. Let q denote the random variable "time spent waiting in the queue" with distribution $F_q(q)$ and density $f_q(q)$. $F_q(q)$ and $f_q(q)$ are both zero for $q < 0$. For an M/G/1 queue the Laplace-Stieltjes transform of $F_q(q)$, $F_q^*(s)$, is given by [9]

$$F_q^*(s) = \int_0^\infty e^{-sq} dF_q(q) = \frac{(1-\rho)s}{s - \lambda[1 - F_b^*(s)]} \tag{3.8}$$

where $F_b^*(s)$ is the Laplace-Stieltjes transform of $F_b(b)$ and it is given by

$$F_b^*(s) = \int_0^\infty e^{-sb} dF_b(b) \tag{3.9}$$

Using equation (3.2) in (3.9) we then have

$$F_b^*(s) = \frac{\mu e^{-s}}{(s+\mu)} \tag{3.10}$$

Substituting (3.10) into (3.8) gives

$$F_q^*(s) = \frac{(1-\rho)s(s+\mu)}{[s^2 + s(\mu - \lambda) - \lambda\mu + \lambda\mu e^{-s}]} \tag{3.11}$$

Let Δ_n denote the random variable "interservice times". Our purpose is to find the density and distribution of Δ_n. We should first find the density and distribution of the random variable

$$c_n = t_n - w_n \tag{3.12}$$

We assume that c_n's i.i.d. with distribution $F_c(c)$ and density $f_c(c)$. The probability density function of the random variable -q is simply $f_q(-q)$. t_n and w_n are two independent random variables. Then $f_c(c)$ can be written as [10]

$$f_c(c) = \int_{-\infty}^{\infty} f_t(\tau) f_q(-c+\tau) d\tau \tag{3.13}$$

The above integral is the convolution of the two functions $f_t(t)$ and $f_q(-q)$. Let $\phi_t(s)$ represent the Bilateral Laplace transform of $f_t(t)$. Using (3.13), the Bilateral Laplace transform of $f_c(c)$, $\phi_c(s)$, is then given by

$$\phi_c(s) = \phi_t(s) F_q^*(-s) = \frac{\lambda}{(s+\lambda)} F_q^*(-s) \tag{3.14}$$

Substituting (3.7) into (3.13) gives

$$f_c(c) = \lambda \int_0^{\infty} e^{-\lambda\tau} f_q(\tau - c) d\tau \tag{3.15}$$

Determination of $f_c(c)$ for $c \geq 0$:

Since $c \geq 0$ and $f_q(q)$ is zero for $q < 0$, (3.15) can be written as

$$f_c(c) = \lambda \int_c^{\infty} e^{-\lambda\tau} f_q(\tau - c) d\tau = \lambda e^{-\lambda c} \int_0^{\infty} e^{-\lambda u} f_q(u) du \tag{3.16}$$

Since $f_q(q)$ is the density and $F_q(q)$ is the distribution of the waiting time in the queue we can write,

$$f_q(q) = \frac{dF_q(q)}{dq} \Rightarrow f_q(q) dq = dF_q(q) \tag{3.17}$$

Substituting (3.17) into (3.16) we get

$$f_c(c) = \lambda e^{-\lambda c} \int_0^{\infty} e^{-\lambda u} dF_q(u) \quad \Rightarrow$$
$$f_c(c) = \lambda e^{-\lambda c} F_q^*(\lambda), \ for \ \ c \geq 0 \tag{3.18}$$

Determination of $f_c(c)$ for c < 0:

Determination of $f_c(c)$ for c<0 is cumbersome and it is unnecessary. Therefore we are leaving it as an expression. Let $f_c(c) = f(c)$, for $c < 0$.

Compact form of $f_c(c)$:

We can write $f_c(c)$ as

$$f_c(c) = \begin{cases} f(c) & , \ c < 0 \\ \lambda e^{-\lambda c} F_q^*(\lambda) & , \ c \geq 0 \end{cases} \tag{3.19}$$

In order to find the density of interarrival times, we should find $F_c(c)$ as well. It is given by

$$F_c(c) = \int_{-\infty}^{c} f_c(u) du \tag{3.20}$$

Using equation (3.19) in (3.20) we have

$$F_c(c) = \begin{cases} \int_{-\infty}^{c} f(u)du & , \ c < 0 \\ \int_{-\infty}^{0^-} f(u)du + \int_{0}^{c} \lambda e^{-\lambda\tau} F_q^*(\lambda)d\tau & , \ c \geq 0 \end{cases}$$
$$= \begin{cases} \int_{-\infty}^{c} f(u)du & , \ c < 0 \\ \int_{-\infty}^{0^-} f(u)du + F_q^*(\lambda)\,(1 - e^{-\lambda c}) & , \ c \geq 0 \end{cases} \qquad (3.21)$$

Let $p = \int_{-\infty}^{0^-} f(u)du$. It remains now only to obtain *p*. This can be accomplished by using equations (3.14) and (3.19). Using equation (3.19) to determine the Bilateral Laplace transform of $f_c(c)$, $\phi_c(s)$ can be written as

$$\phi_c(s) = F_q^*(\lambda)\frac{\lambda}{(s+\lambda)} + \int_{-\infty}^{0^-} f(u)e^{-su}du \qquad (3.22)$$

Equating (3.22) to (3.14) we obtain

$$F_q^*(\lambda)\frac{\lambda}{(s+\lambda)} + \int_{-\infty}^{0^-} f(u)e^{-su}du = \frac{\lambda}{(s+\lambda)}F_q^*(-s) \ \Rightarrow$$
$$\int_{-\infty}^{0^-} f(u)e^{-su}du = \frac{\lambda}{(s+\lambda)}F_q^*(-s) - F_q^*(\lambda)\frac{\lambda}{(s+\lambda)} \qquad (3.23)$$

p can be obtained by evaluating equation (3.23) at s=0:

$$p = F_q^*(0) - F_q^*(\lambda) \qquad (3.24)$$

Using equation (3.11) to find $F_q^*(0)$, we then have

$$p = 1 - F_q^*(\lambda) \tag{3.25}$$

Using (3.25) in (3.21) we can write

$$F_c(c) = \begin{cases} \int_{-\infty}^{c} f(u)du & , \ c < 0 \\ 1 - F_q^*(\lambda)e^{-\lambda c} & , \ c \geq 0 \end{cases} \tag{3.26}$$

Having found a closed-form expression for $f_c(c)$ and $F_c(c)$, we can now find the density and distribution of interservice times. We can rewrite equation (3.6) as

$$\Delta_n = \max(c_n, b_n) \tag{3.27}$$

Let $F_\Delta(\Delta)$ and $f_\Delta(\Delta)$ represent the cumulative distribution function and probability density function of the random variable Δ_n respectively. Since c_n and b_n are two independent random variables, using (3.27) we can write [10]

$$F_\Delta(\Delta) = F_c(\Delta)F_b(\Delta) \tag{3.28}$$

and

$$f_\Delta(\Delta) = \frac{dF_\Delta(\Delta)}{d\Delta} = f_c(\Delta)F_b(\Delta) + F_c(\Delta)f_b(\Delta) \tag{3.29}$$

Using equations (3.1), (3.2) and (3.29) we get

$$f_\Delta(\Delta) = \begin{cases} \lambda e^{-\lambda\Delta}F_q^*(\lambda)(1 - e^{-\mu(\Delta-1)}) + (1 - F_q^*(\lambda)e^{-\lambda\Delta})\mu e^{-\mu(\Delta-1)} & , \ \Delta \geq 1 \text{ sec.} \\ 0 & , \ \Delta < 1 \text{ sec.} \end{cases} \tag{3.30}$$

We then have,

$$E[\Delta_n] = \int_0^\infty u f_\Delta(u) du = f_1(\lambda, \mu)$$
$$E[\Delta_n{}^2] = \int_0^\infty u^2 f_\Delta(u) du = f_2(\lambda, \mu) \tag{3.31}$$

The first two moments of Δ_n are a function of λ and μ and $f_1(\lambda, \mu), f_2(\lambda, \mu)$ represents these functions. We can now use moment matching techniques to show whether the Poisson arrival assumption is true or not.

Comparison of First and Second Moments of Δ_n obtained with Poisson Input

$f_1(\lambda, \mu)$ and $f_2(\lambda, \mu)$ represent the first two moments of Δ_n if t_n is exponentially distributed. However we have no means to observe t_n directly and we can not find an empirical distribution for it. One way of checking the proximity of the distribution of t_n to exponential can be as follows. Since we can calculate the first and second moment of Δ_n for any λ of an exponentially distributed t_n, we can compare these moments with that of empirical. We plot the first moment of Δ_n and the first moment of measured data of interservice times versus λ. Matching of the first moments occurs when these two plots intersect with each other. Let λ_1 represent the value of λ where first moments match. We plot the second moment of Δ_n and the second moment of measured data of interservice times versus λ. Matching of the second moments occurs when these two plots intersect with each other. Let λ_2 represent the value of λ where second moments match. If $\lambda_1 = \lambda_2$, then call arrivals are Poisson distributed. If not, call arrivals are not Poisson distributed. In order to verify this, we carried out a simulation worked with Matlab. Let **t**, **t=** $\{t_1, t_2, ..., t_{15000}\}$, represent the random data sequence obtained by using an exponentially distributed random number

generator with mean $1/\lambda_o$ and length 15000. This sequence represents the interarrival times. λ_o is selected as $0.046\,\text{sec}^{-1}$. Let **b**, **b**= $\{b_1, b_2, ..., b_{15000}\}$, represent the random data sequence obtained by using an exponentially distributed random number generator with mean $2\,\text{sec}.$ and length 15000. **b** represents the service times. To represent the service time pattern defined in equation (3.1) correctly, we added 1 second to each element of the sequence **b**. Using equation (3.3), we generated the waiting times in the queue. Let **w,** **w**= $\{w_1, w_2, ..., w_{15000}\}$, represent the random data sequence obtained by using equation (3.3). w_1 is selected as zero for a starting point. Let $\mathbf{\Delta}$, $\mathbf{\Delta}$= $\{\Delta_1, \Delta_2, ..., \Delta_{15000}\}$, be the random data sequence representing interservice times. Where

$$\Delta_k = \max(t_k - w_k, b_k)$$

We compared first and second moments of $\mathbf{\Delta}$ with the one calculated by equation (3.31),

$$f_1(\lambda_1, 1/2) = E[\mathbf{\Delta}]$$
$$f_2(\lambda_2, 1/2) = E[\mathbf{\Delta}^2]$$
$$\lambda_1 = \lambda_2 = \lambda_o$$

From this result, we conclude that, if the first and second moments match at the same value of λ, then call arrival distribution is a Poisson process.

Interservice time data is taken depending on the traffic load of the corresponding channel, e.g. when it is low or when it is high. When traffic is low or high, we can assume that the variation of the call arrival rate is small. This traffic constraint has led to small data sizes. Therefore the data is comprised of many records ranging from 165 to 868 samples per channel records. Hence to simulate this situation we produced 25 records of 600

samples/record. We divided the **t** vector into 25 successive blocks of length 600. Let Ψ_i be the i th block. Then we plotted $E[\Psi_i]$ and the first moment interservice times which is given by equation (3.31) versus λ. We plotted $E[\Psi_i^2]$ and the second moment of interservice times which is given by (3.31) versus λ. Then we determined λ_1 and λ_2 for i=1,…, 25. The results are given in Table 3.1.

Table 3.1 λ_1 and λ_2 for i=1,…, 25

	λ_1	λ_2	λ_1 / λ_2
i=1	0.049	0.050	0.97
i=2	0.048	0.047	1.02
i=3	0.045	0.045	1.00
i=4	0.045	0.045	1.00
i=5	0.046	0.047	0.99
i=6	0.047	0.046	1.03
i=7	0.046	0.045	1.01
i=8	0.044	0.043	1.02
i=9	0.047	0.049	0.96
i=10	0.045	0.046	0.97
i=11	0.045	0.045	0.98
i=12	0.046	0.047	0.98
i=13	0.047	0.048	0.98
i=14	0.047	0.048	0.98
i=15	0.042	0.042	1.00
i=16	0.044	0.042	1.03
i=17	0.046	0.044	1.02
i=18	0.048	0.049	0.98
i=19	0.047	0.047	0.98
i=20	0.045	0.045	0.99
i=21	0.043	0.045	0.95

i=22	0.047	0.044	1.05
i=23	0.044	0.043	1.01
i=24	0.043	0.049	0.87
i=25	0.049	0.047	1.02

Let $E[\lambda_1]$ and $std(\lambda_1)$ be mean and standard deviation of λ_1 values given in Table 3.1 respectively. And Let $E[\lambda_2]$ and $std(\lambda_2)$ be mean and standard deviation of λ_2 values given in Table 3.1 respectively. We then have,

$$E[\lambda_1] = 0.046 \quad \text{std}(\lambda_1) = 0.0019$$
$$E[\lambda_2] = 0.046 \quad \text{std}(\lambda_2) = 0.0022$$
$$E[\lambda_1 / \lambda_2] = 0.99$$

This result shows that, even for small data sizes, if $\lambda_1 = \lambda_2$, then call arrival distribution is Poisson.

3.2. Comparison of the first Two Moments of both Δ_n and Interservice Times Data

In this section we compare the first two moments of Δ_n to that of interservice times data extracted with the algorithm presented in section 2.3. Data is taken depending on the traffic load of the corresponding channel, e.g. when it is low or when it is high. When traffic is low or high, we can assume that the variation of the call arrival rate is small. The traffic load of a channel is computed with the following procedure. Let A_i represent the traffic load in a time interval of T_i. T_i is given in units of minutes. A_i is calculated as follows

$$A_i = \frac{N_i}{T_i} H_i \tag{3.32}$$

where N_i is the total number of calls which acquired services during the interval T_i and H_i is the average call duration in the interval T_i. H_i is calculated as follows

$$H_i = \frac{\sum_{k=1}^{N_i} d_k}{N_i} \tag{3.33}$$

where d_k, {k=1, 2,…, N_i} is the duration of the k th call, in units of minutes, which acquired service in the interval T_i.

3.2.1 Statistical Results Based on Channel 2 Data

In this section, we used interservice times and call duration data of Channel 2. We plotted the traffic load of Channel 2 using equations (3.32) and (3.33). We plotted A_i's versus time to see the traffic carried on Channel 2 for three successive days. These plots are given in Figures (3.2), (3.3) and (3.4).

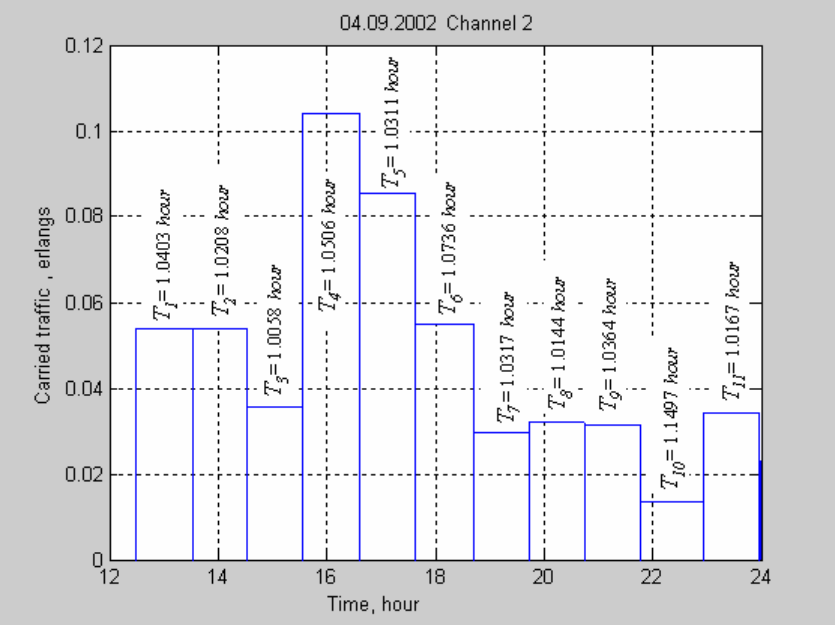


Figure 3.2 Traffic load of Channel 2 on 04.09.2002.

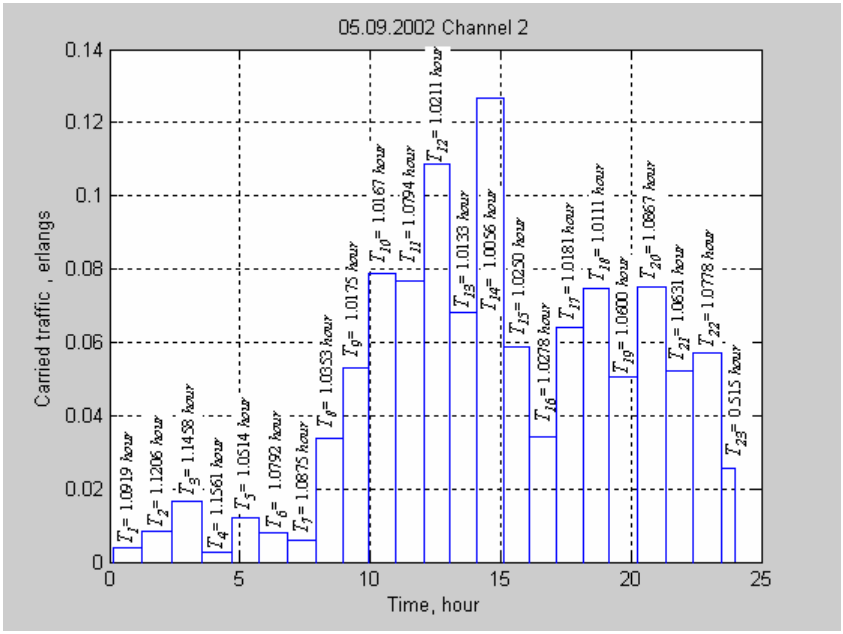


Figure 3.3 Traffic load of Channel 2 on 05.09.2002

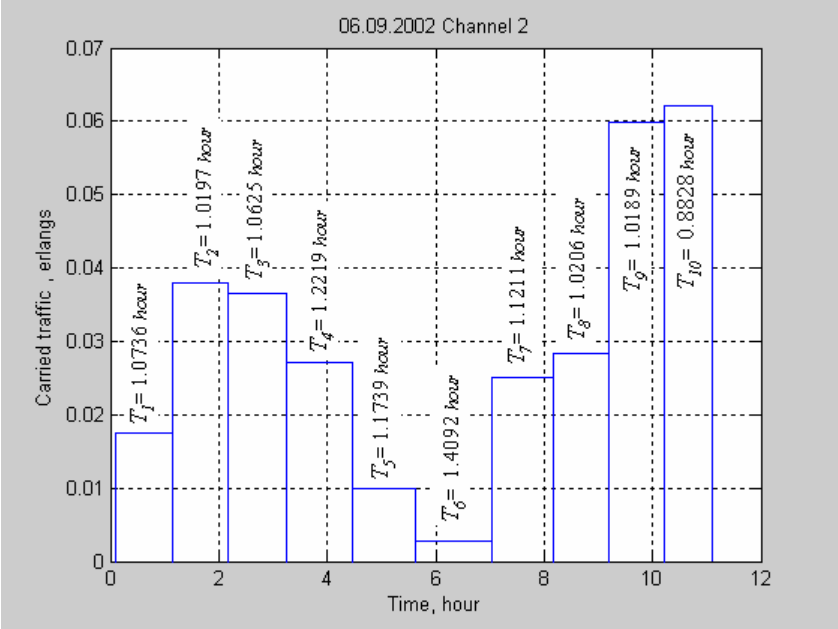


Figure 3.4 Traffic load of Channel 2 on 06.09.2002

3.2.1.1 Moment Comparison at the Time Interval when Traffic Load is Low

The data used in this section is taken when traffic load of Channel 2 is below 0.04 Erlang. For a single channel, a traffic load of 0.04 Erlang means a call blocking probability of 4%, which is sufficiently low for our analysis.

As you can see in figures (3.2), (3.3) and (3.4) there are some time intervals during which traffic load is below 0.04 Erlang. We extracted the interservice times and call durations detected during these intervals with the following algorithm:

1. For Channel 2 data
2. fid1=fopen (“file1.txt”, “w”)
3. fid2=fopen (“file2.txt”,”w”)

4. for (i =1: $i_{\max}$)

5. calculate A_i

6. if $A_i \le 0.04$

7. for (k=1: N_i)

8. write Δ_k into fid1

9. write μ_k into fid2

10. end

11. end

12. end

13. fclose (fid1); fclose (fid2)

where Δ_k is the measured interservice time between *k*th and (*k*+1) th call and μ_k is the duration of the k th call. There are 868 calls which acquired services when the traffic load of Channel 2 is below 0.04 Erlang. Let $\mathbf{\Delta}$ be a *measured data sequence* whose elements are $\{\Delta_j$, j =1, 2,…, 868$\}$ and let $\boldsymbol{\mu}$ be a *measured data sequence* whose elements are $\{\mu_j$, j =1, 2,…, 868$\}$. We found the first and second moments of $\mathbf{\Delta}$ as

$$E[\mathbf{\Delta}] = \frac{\sum_{k=1}^{868} \Delta_k}{868} = 105.35 \sec.$$

$$E[\mathbf{\Delta^2}] = \frac{\sum_{k=1}^{868} \Delta^2{}_k}{868} = 1.44 \ \ 10^5 \sec^2$$

We also modeled the distribution of $\boldsymbol{\mu}$ with the EM algorithm described in section 2.2. The distribution of this *measured data sequence* is a shifted exponential with parameter $\mu = 0.75 \sec^{-1}$ (Figure 3.5).

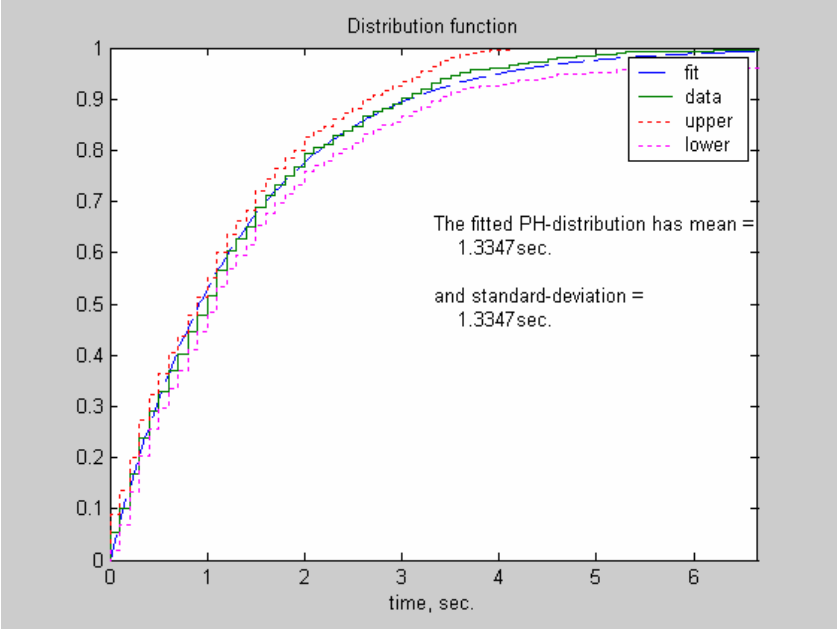


Figure 3.5 The fitted phase type distribution to the sequence $\boldsymbol{\mu}-0.7$. The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ (n =868) with fitted PH-distribution of order 1.

As defined in section 2.2, minimum duration of a voice call is 0.7 second. This leads to an increase in the order of the phase type distribution fitted to $\boldsymbol{\mu}$. Therefore, in order to keep the order low, we tried to fit a phase type distribution to the sequence $\boldsymbol{\mu}-0.7$. This fit is plotted in Figure 3.5, where the fitted density is

$$0.75e^{-0.75t}\ ,\ t \geq 0. \tag{3.34}$$

Let $f_{\mu}(\mu)$ be the density of service times which acquired services when the traffic load of Channel 2 is below 0.04 Erlang. Using (3.34) we can write

$$f_{\mu}(\mu) = \begin{cases} 0.75e^{-0.75(\mu-1)} & ,\ \mu \geq 1\text{sec.} \\ 0 & ,\ \mu<1\text{sec.} \end{cases} \tag{3.35}$$

We can now compare $E[\Delta_n]$ and $E[\Delta_n{}^2]$ with $E[\mathbf{\Delta}]$ and $E[\mathbf{\Delta}^2]$ respectively. Using equation (3.31) we plot $\varepsilon_1 = \dfrac{(f_1(\lambda, 0.75) - E[\mathbf{\Delta}])}{E[\mathbf{\Delta}]}.100$ and $\varepsilon_2 = \dfrac{f_2(\lambda, 0.75) - E[\mathbf{\Delta}^2]}{E[\mathbf{\Delta}^2]}.100$ on Figure 3.6. Let λ_1 and λ_2 denote the value of λ where ε_1 and ε_2 are zero respectively.

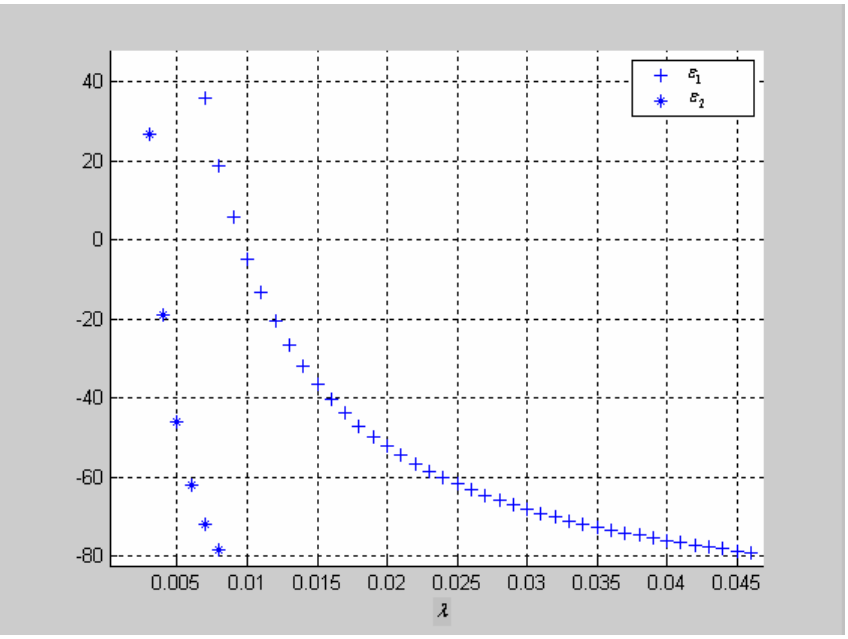


Figure 3.6 ε_1 and ε_2 versus λ

In Figure 3.6, it is obvious that $\lambda_1 \neq \lambda_2$. This shows that call arrivals are not Poisson distributed.

We plotted the cumulative distribution function of $\mathbf{\Delta}$ and $F_\Delta(\Delta)$ using equation (3.28). Let F denote the cumulative distribution function of the measured data sequence, $\mathbf{\Delta}$. $F_\Delta(\Delta)$ is a function of both λ and μ. This plot is given on Figure 3.7.

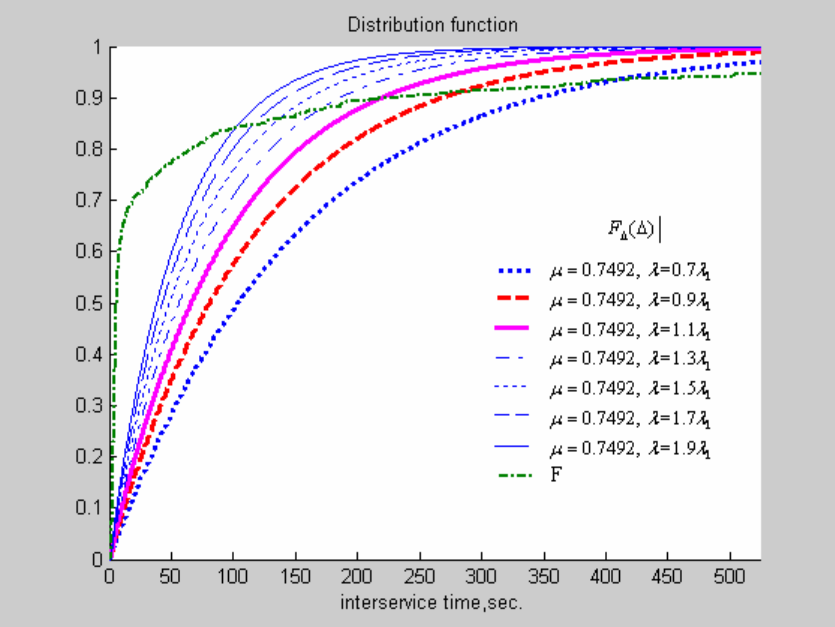


Figure 3.7 $F_\Delta(\Delta)$ and F, $\lambda_1 = 0.0095\,\text{sec}^{-1}$

As obvious in Figure 3.7, the difference between $F_\Delta(\Delta)$ and F is significant. This shows that, call arrivals are not Poisson distributed.

In the next section we further carry out the above procedure when traffic load of Channel 2 is highest.

3.2.1.2 Moment Comparison at the Time Interval when Traffic Load is Highest

In this section, we used interservice times and call duration data of Channel 2 which is taken when traffic load is maximum.

As you can see in figures (3.2), (3.3) and (3.4), the traffic load of Channel 2 is maximum on 05.09.2002 at the time interval 14:05:17PM to 15:05:37PM. We refer this interval as 'busiest interval'. Let Δ_k represent the

measured interservice time between *k*th and (*k*+1) th calls and let μ_k represent the duration of the k th call which acquired service during the busiest interval respectively. There are 165 calls which acquired services when the traffic load of Channel 2 is maximum. Let $\mathbf{\Delta}$ be a *random data sequence* whose elements are $\{\Delta_j, \text{j} = 1, 2, \ldots, 165\}$ and let $\boldsymbol{\mu}$ be a *random data sequence* whose elements are $\{\mu_j, \text{j} = 1, 2, \ldots, 165\}$. We found the first and second moments of $\mathbf{\Delta}$ as

$$E[\mathbf{\Delta}] = \frac{\sum_{k=1}^{165} \Delta_k}{165} = 21.94 \sec.$$

$$E[\mathbf{\Delta^2}] = \frac{\sum_{k=1}^{165} \Delta^2{}_k}{165} = 1.99 \ \ 10^3 \sec^2$$

Before moment matching procedure, we also modeled the distribution of $\boldsymbol{\mu}$ with the EM algorithm described in section 2.2. The distribution of this *random data sequence* is a shifted exponential with parameter $\mu = 0.5 \sec^{-1}$ (Figure 3.8).

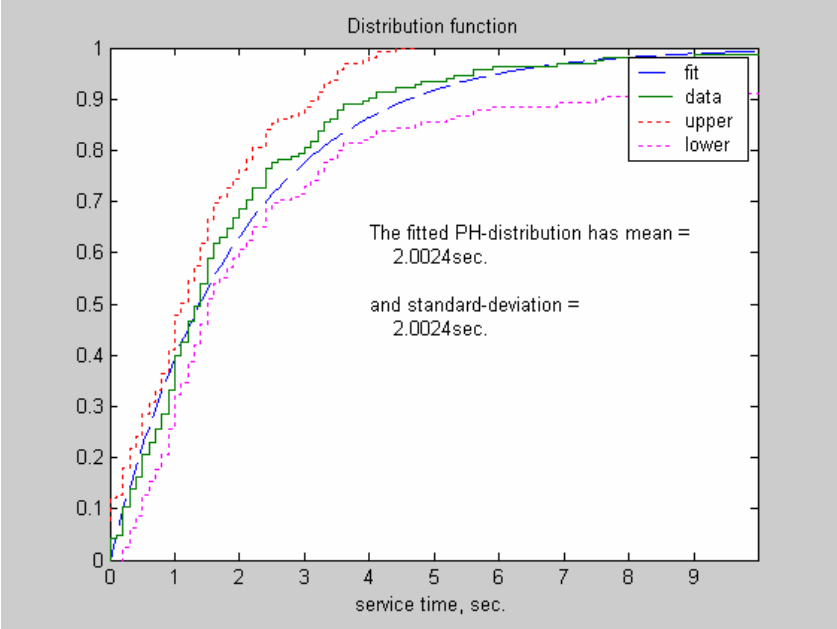


Figure 3.8 The fitted phase type distribution to the sequence $\boldsymbol{\mu}$ - 0.7. The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ (n = 165) with fitted PH-distribution of order 1.

The density that fits to $\boldsymbol{\mu}$ - 0.7 is given by,

$$0.5e^{-0.5t}\ ,\ t \geq 0. \tag{3.36}$$

Let $f_{\mu}(\mu)$ be the density of service times which acquired services when the traffic load of Channel 2 is maximum. Using (3.36) we can write

$$f_{\mu}(\mu) = \begin{cases} 0.5e^{-0.5(\mu-1)} & ,\ \mu \geq 1\text{sec.} \\ 0 & ,\ \mu < 1\text{sec.} \end{cases} \tag{3.37}$$

We can now compare $E[\Delta_n]$ and $E[\Delta_n^2]$ with $E[\boldsymbol{\Delta}]$ and $E[\boldsymbol{\Delta}^2]$ respectively. Using equation (3.31) we plot $\varepsilon_1 = \dfrac{(f_1(\lambda, 0.5) - E[\boldsymbol{\Delta}])}{E[\boldsymbol{\Delta}]}.100$ and

$\varepsilon_2 = \frac{f_2(\lambda, 0.5) - E[\Delta^2]}{E[\Delta^2]}.100$ on Figure 3.9. Let λ_1 and λ_2 denote the value of λ where ε_1 and ε_2 are zero respectively.

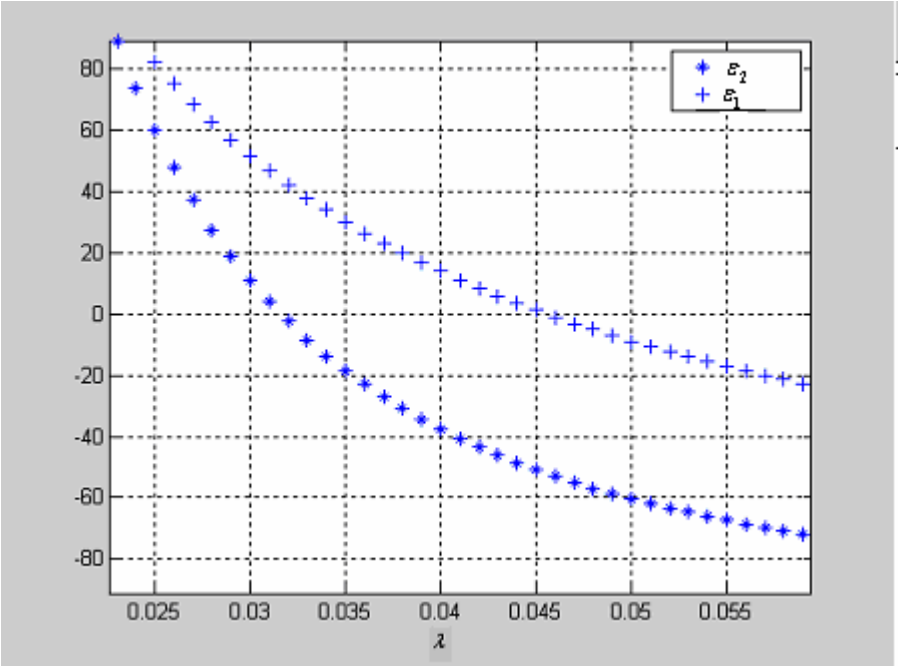


Figure 3.9 ε_1 and ε_2 versus λ

We see that $\lambda_1 \neq \lambda_2$. This again shows that call arrivals are not Poisson distributed.

3.2.2 Statistical Results Based on Channel 3 Data

In this section, we used interservice times and call duration data of Channel 3. We plotted the traffic load of Channel 3 using equations (3.32) and (3.33). We plotted A_i's versus time to see the traffic carried on Channel 3 for three successive days. These plots are given on Figures (3.10), (3.11) and (3.12).

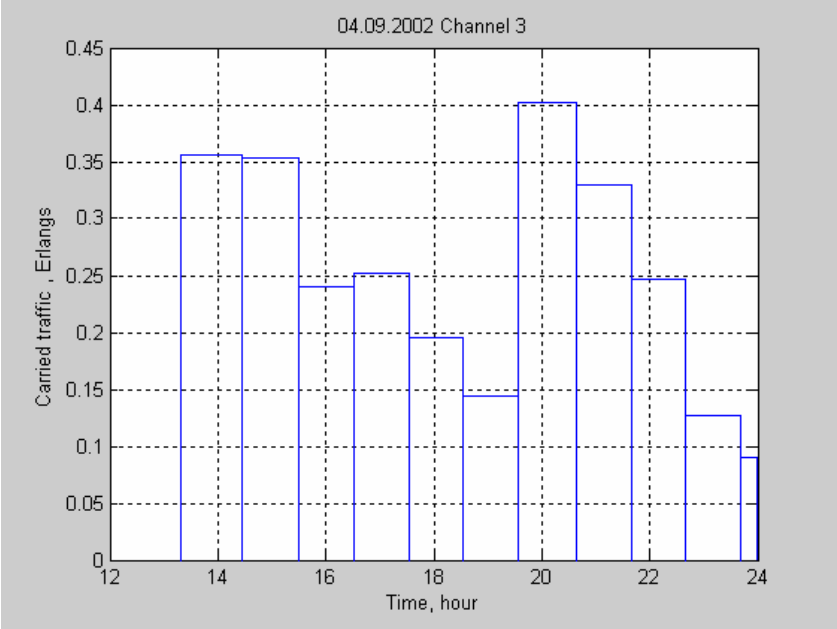


Figure 3.10 Traffic load of Channel 3 on 04.09.2002

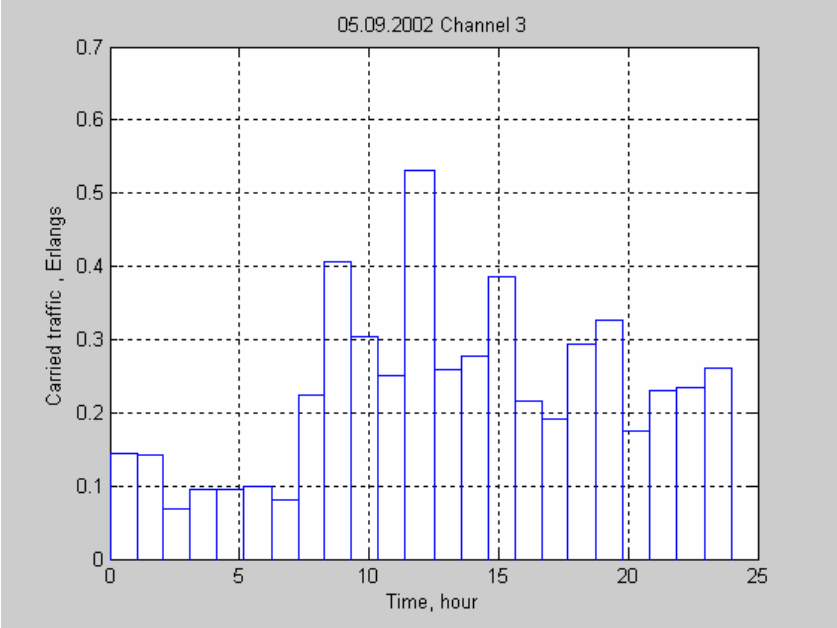


Figure 3.11 Traffic load of Channel 3 on 05.09.2002

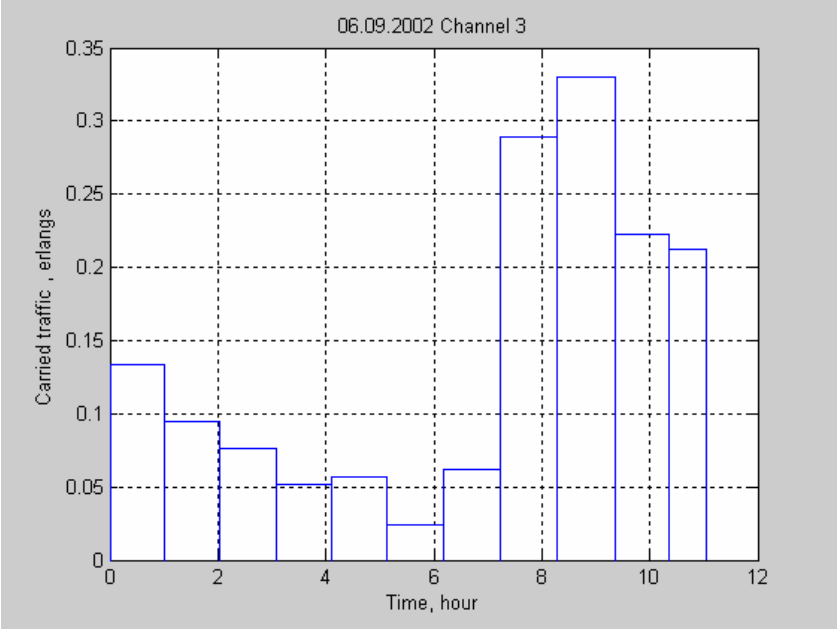


Figure 3.12 Traffic load of Channel 3 on 06.09.2002

3.2.2.1 Moment Comparison at the Time Interval when Traffic Load is Highest

The data used in this section is taken when traffic load of Channel 3 is maximum. As you can see in figures (3.10), (3.11) and (3.12), the traffic load of Channel 3 is maximum on 05.09.2002 at the time interval 11:22:56PM to 12:31:57PM. We refer this interval as 'busiest interval'. Let Δ_k represent the measured interservice time between *k*th and (*k*+1) th calls and let μ_k represent the duration of the k th call which acquired service during the busiest interval respectively. There are 544 calls which acquired services when the traffic load of Channel 3 is maximum. Let $\mathbf{\Delta}$ be a *random data sequence* whose elements are $\{\Delta_j, \text{j} = 1, 2, \ldots, 544\}$ and let $\boldsymbol{\mu}$ be a *random data sequence* whose elements are $\{\mu_j, \text{j} = 1, 2, \ldots, 544\}$. We found the first and second moments of $\mathbf{\Delta}$ as

$$E[\boldsymbol{\Delta}] = \frac{\sum_{k=1}^{544} \Delta_k}{544} = 7.7\,\text{sec}.$$

$$E[\boldsymbol{\Delta}^2] = \frac{\sum_{k=1}^{544} \Delta^2{}_k}{544} = 360.02\,\text{sec}^2$$

We also modeled the distribution of $\boldsymbol{\mu}$ with the EM algorithm described in section 2.2. The distribution of this *random data sequence* is a shifted exponential with parameter $\mu = 0.39\,\text{sec}^{-1}$ (Figure 3.13).

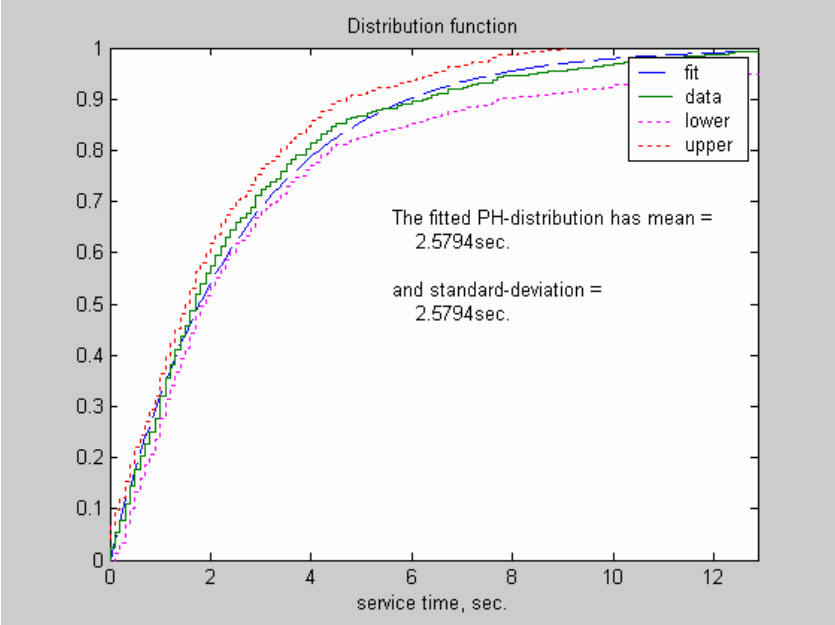


Figure 3.13 The fitted phase type distribution to the sequence $\boldsymbol{\mu}$ - 0.7. The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ (n= 544) with fitted PH-distribution of order 1.

The density that fits to $\boldsymbol{\mu}$ - 0.7 is given by,

$$0.39e^{-0.39t}\ , \text{t} \geq 0. \qquad (3.38)$$

Let $f_\mu(\mu)$ be the density of the service times which acquired services when the traffic load of Channel 3 is highest. Using (3.38) we can write

$$f_\mu(\mu) = \begin{cases} 0.39e^{-0.39(\mu-1)} & , \mu \geq 1\text{sec.} \\ 0 & , \mu < 1\text{sec.} \end{cases} \tag{3.39}$$

We can now compare $E[\Delta_n]$ and $E[\Delta_n{}^2]$ with $E[\mathbf{\Delta}]$ and $E[\mathbf{\Delta}^2]$ respectively. Using equation (3.31) we plot $\varepsilon_1 = \dfrac{(f_1(\lambda, 0.39) - E[\mathbf{\Delta}])}{E[\mathbf{\Delta}]}.100$ and $\varepsilon_2 = \dfrac{f_2(\lambda, 0.39) - E[\mathbf{\Delta}^2]}{E[\mathbf{\Delta}^2]}.100$ on Figure 3.14. Let λ_1 and λ_2 denote the value of λ where ε_1 and ε_2 are zero respectively.

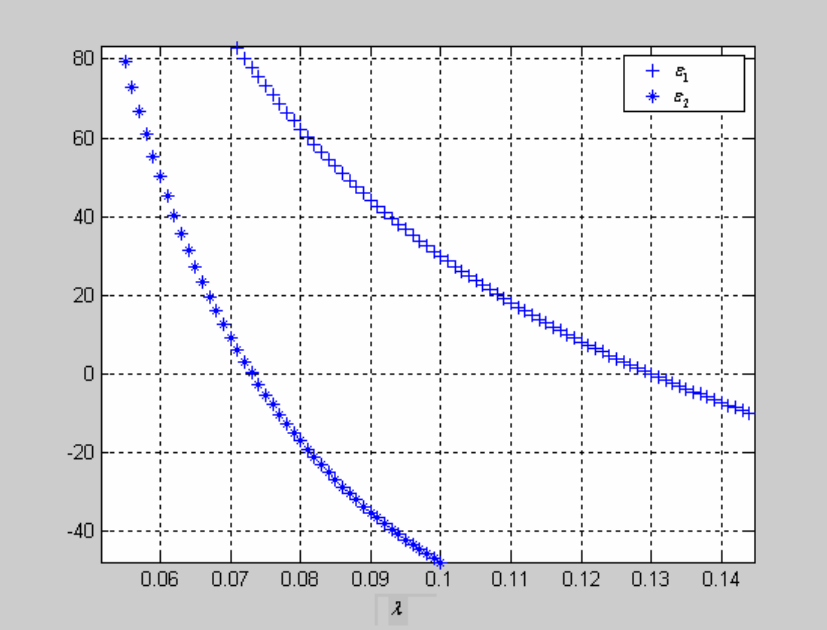


Figure 3.14 ε_1 and ε_2 versus λ

On Figure 3.14, we see that $\lambda_1 \neq \lambda_2$. This emphasizes that call arrivals are not Poisson distributed.

3.2.2.2 Moment Comparison at the Time Interval when Traffic Load is Low

There are only 37 calls which acquired services when the traffic load of Channel 3 is below 0.04 Erlang. Therefore, we tried the next higher measured traffic load, which is 0.07 Erlang. The data used in this section is taken when traffic load of Channel 3 is below or equal to 0.07 Erlang. For a single channel, a traffic load of 0.07 Erlang means a call blocking probability of 7%, which is sufficiently low for our analysis.

For determining the time interval at which traffic load is below 0.07 Erlang, we followed the same algorithm, except first and sixth line, which is presented in section 3.2.1.1. The first and sixth lines of this algorithm are changed to:

1. For Channel 3 data
6. if $A_i \leq 0.07$

As you can see on Figures (3.10), (3.11) and (3.12), the traffic load of Channel 3 is below or equal to 0.07 Erlang at the time intervals 02:04:17AM to 03:05:14AM on 05.09.2002 and 03:04:20AM to 07:13:44AM on 06.09.2002. There are 428 calls which acquired services when the traffic load of Channel 3 is below 0.07 Erlang. Let $\boldsymbol{\Delta}$ be a *measured data sequence* whose elements are $\{\Delta_j$, j =1, 2,…, 428$\}$ and let $\boldsymbol{\mu}$ be a *measured data sequence* whose elements are $\{\mu_j$, j =1, 2,…, 428$\}$. We found the first and second moments of $\boldsymbol{\Delta}$ as

$$E[\boldsymbol{\Delta}] = \frac{\sum_{k=1}^{428} \Delta_k}{428} = 43.14 \sec.$$

$$E[\boldsymbol{\Delta}^2] = \frac{\sum_{k=1}^{428} \Delta^2{}_k}{428} = 1.47\ 10^4\ \text{sec}^2$$

We also modeled the distribution of $\boldsymbol{\mu}$ with the EM algorithm described in section 2.2. The distribution of this *measured data sequence* is a shifted exponential with parameter $\mu = 0.65\,\text{sec}^{-1}$ (Figure 3.15).

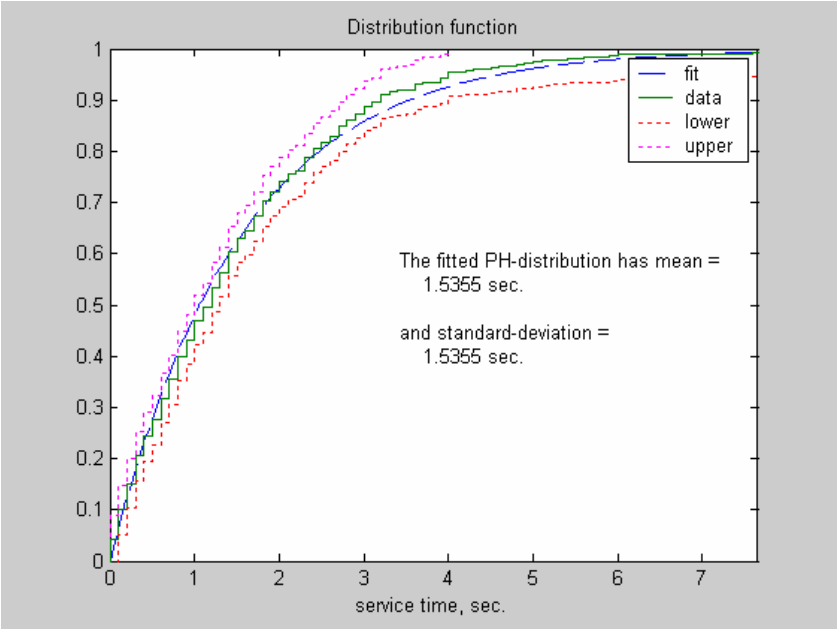


Figure 3.15 The fitted phase type distribution to the sequence $\boldsymbol{\mu}$ - 0.7. The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ (n=428) with fitted PH-distribution of order 1.

The density that fits to $\boldsymbol{\mu}$ - 0.7 is given by,

$$0.65e^{-0.65t}\ ,\ t \geq 0. \qquad (3.40)$$

Let $f_{\mu}(\mu)$ be the density service times which acquired services when the traffic load of Channel 3 is below or equal to 0.07 Erlang. Using (3.40) we can write

$$f_{\mu}(\mu) = \begin{cases} 0.65e^{-0.65(\mu-1)} & , \mu \geq 1\text{sec.} \\ 0 & , \mu < 1\text{sec.} \end{cases} \tag{3.41}$$

We can now compare $E[\Delta_n]$ and $E[\Delta_n{}^2]$ with $E[\mathbf{\Delta}]$ and $E[\mathbf{\Delta}^2]$ respectively. Using equation (3.31) we plot $\varepsilon_1 = \dfrac{(f_1(\lambda, 0.65) - E[\mathbf{\Delta}])}{E[\mathbf{\Delta}]}.100$ and $\varepsilon_2 = \dfrac{f_2(\lambda, 0.65) - E[\mathbf{\Delta}^2]}{E[\mathbf{\Delta}^2]}.100$ on Figure 3.16. Let λ_1 and λ_2 denote the value of λ where ε_1 and ε_2 are zero respectively.

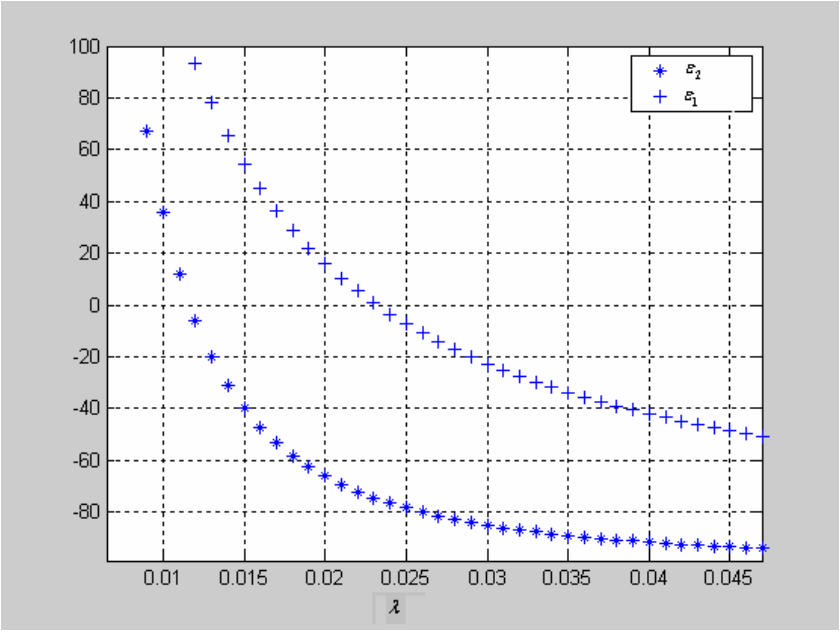


Figure 3.16 ε_1 and ε_2 versus λ

We see that $\lambda_1 \neq \lambda_2$. This again shows that call arrivals are not Poisson distributed.

3.2.3 Statistical Results Based on Channel 1 Data

In this section, we used interservice times and call duration data of Channel 1. We plotted the traffic load of Channel 1 using equations (3.32) and (3.33). We plotted A_i's versus time to see the traffic carried on Channel 1 for three successive days. These plots are on Figures (3.17), (3.18) and (3.19).

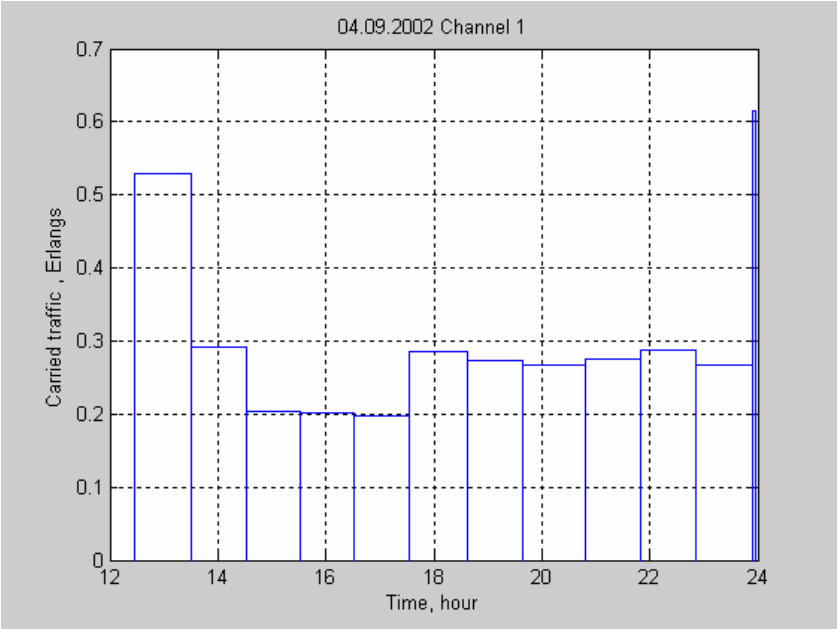


Figure 3.17 Traffic load of Channel 1 on 04.09.2002

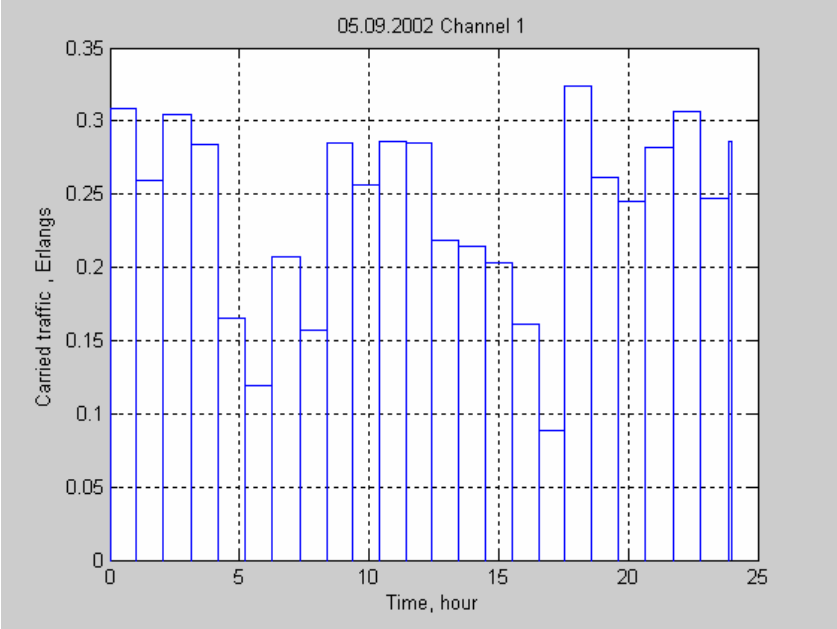


Figure 3.18 Traffic load of Channel 1 on 05.09.2002

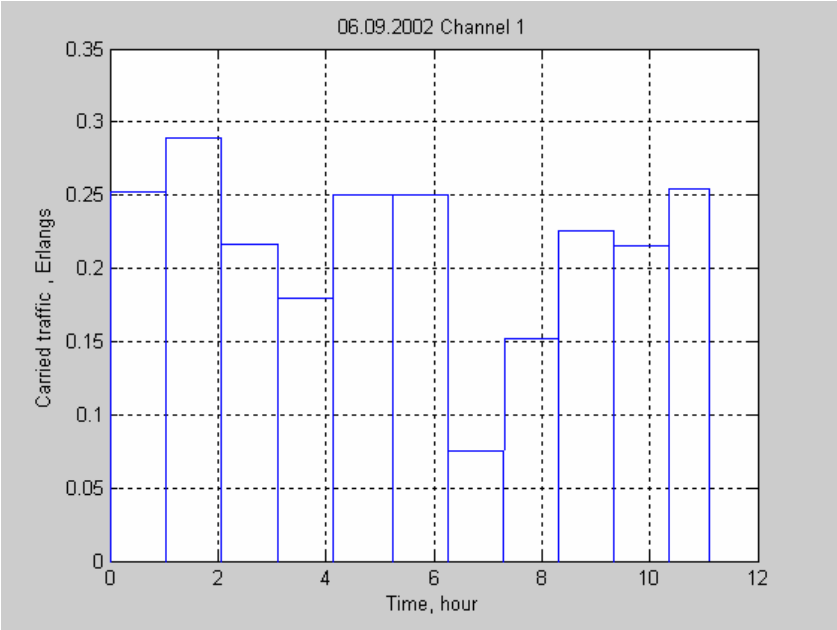


Figure 3.19 Traffic load of Channel 1 on 06.09.2002

As shown in section 2.4.2, the distribution and density of the service times of Channel 1 follow a different pattern than the one given in equations (3.1) and (3.2). Therefore, all the equations except (3.1), (3.2), (3.10), (3.11), (3.30) and (3.31) of section 3.1 are valid in this section. We should find a new expression

for the first two moments of interservice times. Let $F_b(b)$ and $f_b(b)$ denote the distribution and density of service times of calls served by Channel 1. As shown in section 2.4.2, the general form of $F_b(b)$ and $f_b(b)$ can be written as

$$f_b(b) = \begin{cases} p_1\mu_1\ e^{-\mu_1(b-1)} - p_2\mu_2 e^{-\mu_2(b-1)} & , \quad b \geq 1\text{sec.} \\ 0 & , \ b < 1\text{sec.} \end{cases} \tag{3.42}$$

$$F_b(b) = \begin{cases} 1 - p_1 e^{-\mu_1(b-1)} + p_2 e^{-\mu_2(b-1)} & , \quad b \geq 1\text{sec.} \\ 0 & , \quad b < 1\text{sec.} \end{cases} \tag{3.43}$$

where $p_1 - p_2 = 1$. The Laplace-Stieltjes transform of the service times, $F^*_b(s)$ is then given by

$$F_b^*(s) = \frac{p_1\mu_1 e^{-s}}{(s+\mu_1)} - \frac{p_2\mu_2 e^{-s}}{(s+\mu_2)} \tag{3.44}$$

Using (3.8) and (3.44) we can write

$$F_q^*(s) = \frac{(1-\rho)s(s+\mu_1)(s+\mu_2)}{(s-\lambda)(s+\mu_1)(s+\mu_2) + \lambda[p_1\mu_1(s+\mu_2)e^{-s} - p_2\mu_2(s+\mu_1)e^{-s}]} \tag{3.45}$$

We can now find the density of the random variable Δ_n using equations (3.42), (3.43), (3.19), (3.26), (3.45) and (3.29). We get

$$f_\Delta(\Delta) = \begin{cases} \lambda e^{-\lambda\Delta} F_q^*(\lambda)(1 - p_1 e^{-\mu_1(\Delta-1)} + p_2 e^{-\mu_2(\Delta-1)}) \\ + (1 - F_q^*(\lambda)e^{-\lambda\Delta})(p_1\mu_1\ e^{-\mu_1(\Delta-1)} - p_2\mu_2 e^{-\mu_2(\Delta-1)}) & , \Delta \geq 1\ \text{sec.} \\ 0 & , \Delta < 1\ \text{sec.} \end{cases} \tag{3.46}$$

Using (3.46) we then have

$$E[\mathbf{\Delta_n}] = \int_0^\infty u f_\Delta(u)du = g_1(\lambda, \mu_1, \mu_2, p_1, p_2)$$

$$E[\mathbf{\Delta_n}^2] = \int_0^\infty u^2 f_\Delta(u)du = g_2(\lambda, \mu_1, \mu_2, p_1, p_2) \tag{3.47}$$

First two moments of Δ_n are functions of $\lambda, \mu_1, \mu_2, p_1, p_2$ and $g_1(\lambda, \mu_1, \mu_2, p_1, p_2), g_2(\lambda, \mu_1, \mu_2, p_1, p_2)$ represents these functions.

3.2.3.1 Moment Comparison at the Time Interval when Traffic Load is Highest

As you can see on figures (3.17), (3.18) and (3.19), the traffic load of Channel 1 is maximum on 04.09.2002 at the time interval 23:54:18PM to 23:57:51PM. However, there are only 50 calls which acquired services within this interval. Therefore we skipped this interval and found the next busiest interval during which more calls are served. As obvious on Figure (3.17) this interval starts 12:26:33PM and ends at 13:31:03PM where the traffic load is above 0.5 Erlang. We refer this interval as 'busiest interval'. Let Δ_k represent the measured interservice time between *k*th and (*k*+1) th calls and let μ_k represent the duration of the k th call which acquired service during the busiest interval respectively. There are 528 calls which acquired services during the busiest interval. Let $\mathbf{\Delta}$ be a *random data sequence* whose elements are $\{\Delta_j$, j =1, 2,…, 528$\}$ and let $\boldsymbol{\mu}$ be a *random data sequence* whose elements are $\{\mu_j$, j =1, 2,…, 528$\}$. We found the first and second moments of $\mathbf{\Delta}$ as

$$E[\mathbf{\Delta}] = \frac{\sum_{k=1}^{528} \Delta_k}{528} = 7.36\,\text{sec}.$$

$$E[\Delta^2] = \frac{\sum_{k=1}^{528} \Delta^2{}_k}{528} = 410.05\,\text{sec}^2$$

We also modeled the distribution of $\boldsymbol{\mu}$ with the EM algorithm (Figures 3.20 and 3.21).

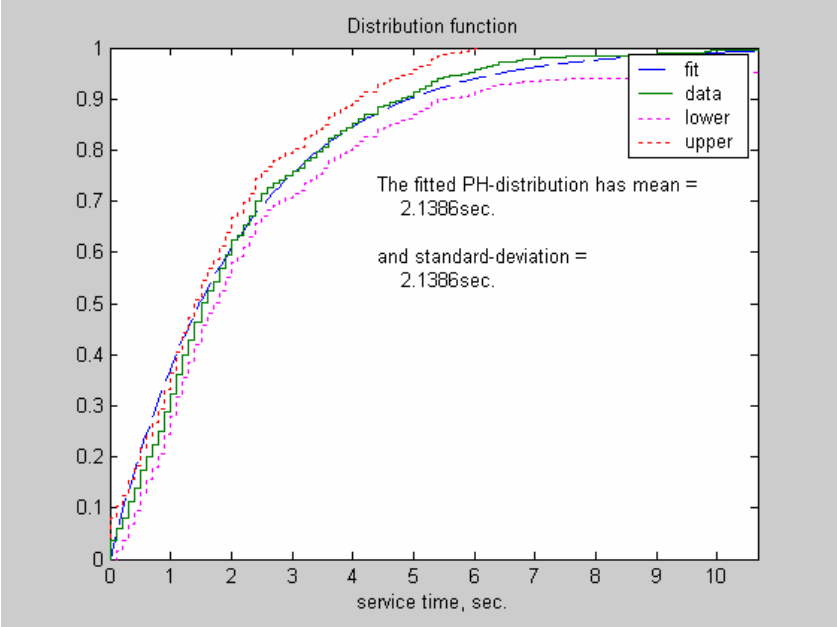


Figure 3.20 The fitted phase type distribution to the sequence $\boldsymbol{\mu}$ - 0.7 . The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ (n=528) with fitted hyperexponential-distribution of order 1.

As seen in Figure 3.20, exponential distribution does not fit into the confidence interval of $\pm 1/\sqrt{n}$ .The EM-algorithm fits the same distribution to $\boldsymbol{\mu}$ - 0.7 even if the order of the hyperexponential fit is increased beyond 1. Therefore we need to try another distribution.

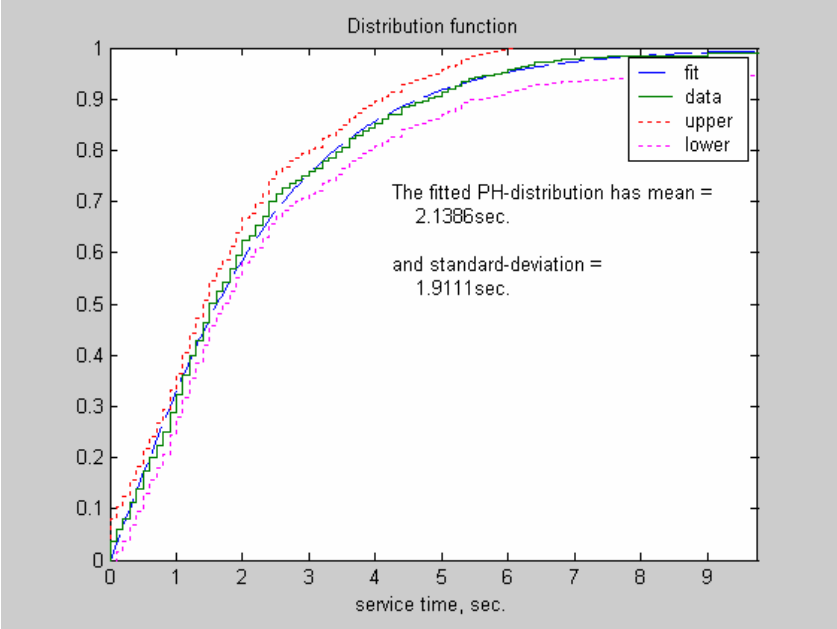


Figure 3.21 The fitted phase type distribution to the sequence $\boldsymbol{\mu}$ - 0.7. The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ with fitted General phase type Distribution of order 2.

As seen on Figure 3.21, General phase type distribution with two phases fits into the confidence interval of $\pm 1/\sqrt{n}$. Therefore we can stop adding more phases. The parameters of this fit are:

$$\boldsymbol{\pi} = \begin{pmatrix} 0.1012 \\ 0.8988 \end{pmatrix} \qquad \mathbf{T} = \begin{pmatrix} -0.696786 & 0.089420 \\ 0.776276 & -1.066619 \end{pmatrix} \qquad \mathbf{t} = \begin{pmatrix} 0.6 \\ 0.29 \end{pmatrix}$$

You can see the description of these parameters in section 2.1. Using these parameters we can write the density of the fit given in Figure 3.21 as:

$$f_m(m) = \begin{cases} 0.766\, e^{-0.5598m} - 0.4363 e^{-1.2036m} & , \quad m \geq 0 \\ 0 & , \quad m < 0 \end{cases} \tag{3.48}$$

Where $f_m(m)$ is the density of the PH type fit to the sequence $\boldsymbol{\mu}$ - 0.7. Let $f_\mu(\mu)$ and $F_\mu(\mu)$ be the density and distribution of service times which acquired services during the busiest interval of Channel 1. Using (3.48) we can write

$$f_\mu(\mu) = \begin{cases} 0.766\, e^{-0.55598(\mu-1)} - 0.4363 e^{-1.2036(\mu-1)} & , \quad \mu \geq 1\text{sec.} \\ 0 & , \quad \mu < 1\text{sec.} \end{cases} \tag{3.49}$$

$$F_\mu(\mu) = \begin{cases} 1 - 1.36 e^{-0.5598(\mu-1)} + 0.36 e^{-1.2036(\mu-1)} & , \quad \mu \geq 1\text{sec.} \\ 0 & , \quad \mu < 1\text{sec.} \end{cases} \tag{3.50}$$

We can now compare $E[\Delta_n]$ and $E[\Delta_n{}^2]$ with $E[\boldsymbol{\Delta}]$ and $E[\boldsymbol{\Delta}^2]$ respectively. Using equation (3.47) we plot

$$\varepsilon_1 = \frac{(g_1(\lambda, 0.5598, 1.2036, 1.36, 0.36) - E[\boldsymbol{\Delta}])}{E[\boldsymbol{\Delta}]}.100$$

and $\varepsilon_2 = \dfrac{(g_2(\lambda, 0.5598, 1.2036, 1.36, 0.36) - E[\boldsymbol{\Delta}^2])}{E[\boldsymbol{\Delta}^2]}$ on Figure 3.22.

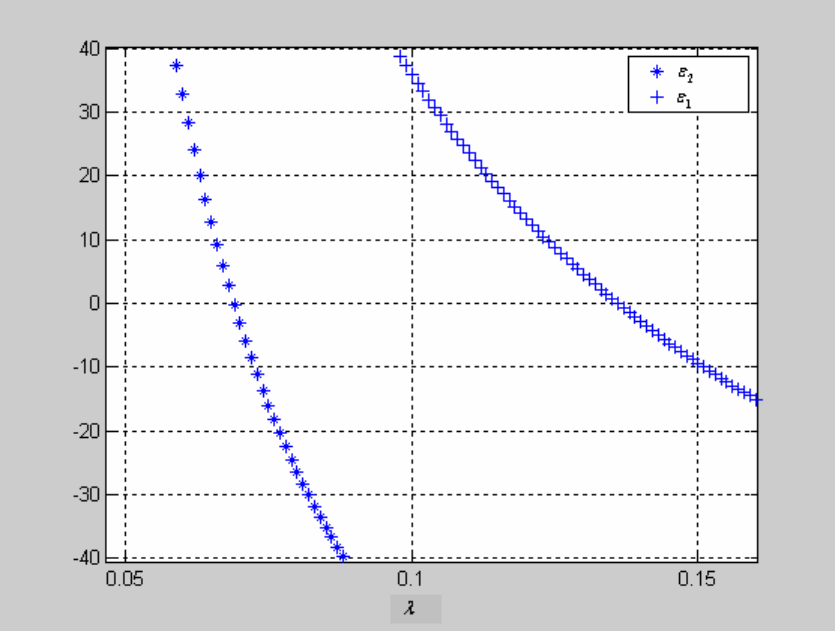


Figure 3.22 ε_1 and ε_2 versus λ

We see that $\lambda_1 \neq \lambda_2$. This again shows that call arrivals are not Poisson distributed.

3.2.3.2 Moment Comparison at the Time Interval when Traffic Load is Low

The data used in this section is taken when traffic load of Channel 1 is below or equal to 0.088 Erlang. For a single channel, a traffic load of 0.088 Erlang means a call blocking probability of 8.8 %, which is sufficiently low for our analysis. For determining the time interval at which traffic load is below 0.088 Erlang, we followed the same algorithm, except first and sixth lines, which is presented in section 3.2.1.1. The first and sixth lines of this algorithm are changed to:

1. For Channel 1 data
6. if $A_i \leq 0.088$

As you can see in Figures (3.17), (3.18) and (3.19), the traffic load of Channel 1 is below or equal to 0.088 Erlang at the time intervals 16:32:09PM to 17:32:56PM on 05.09.2002 and 06:16:11AM to 07:18:58AM on 06.09.2002. There are 251 calls which acquired services when the traffic load of Channel 1 is below 0.088 Erlang. Let $\mathbf{\Delta}$ be a *measured data sequence* whose elements are $\{\Delta_j, \text{j} = 1, 2, \ldots, 251\}$ and let $\boldsymbol{\mu}$ be a *measured data sequence* whose elements are $\{\mu_j, \text{j} = 1, 2, \ldots, 251\}$. We found the first and second moments of $\mathbf{\Delta}$ as

$$E[\mathbf{\Delta}] = \frac{\sum_{k=1}^{251} \Delta_k}{251} = 29.14 \sec.$$

$$E[\mathbf{\Delta^2}] = \frac{\sum_{k=1}^{251} \Delta^2{}_k}{251} = 9.38 \ \ 10^3 \sec^2$$

We also modeled the distribution of $\boldsymbol{\mu}$ with the EM algorithm. The distribution of this *measured data sequence* is a shifted exponential with parameter $\mu = 0.6 \sec^{-1}$ (Figure 3.23).

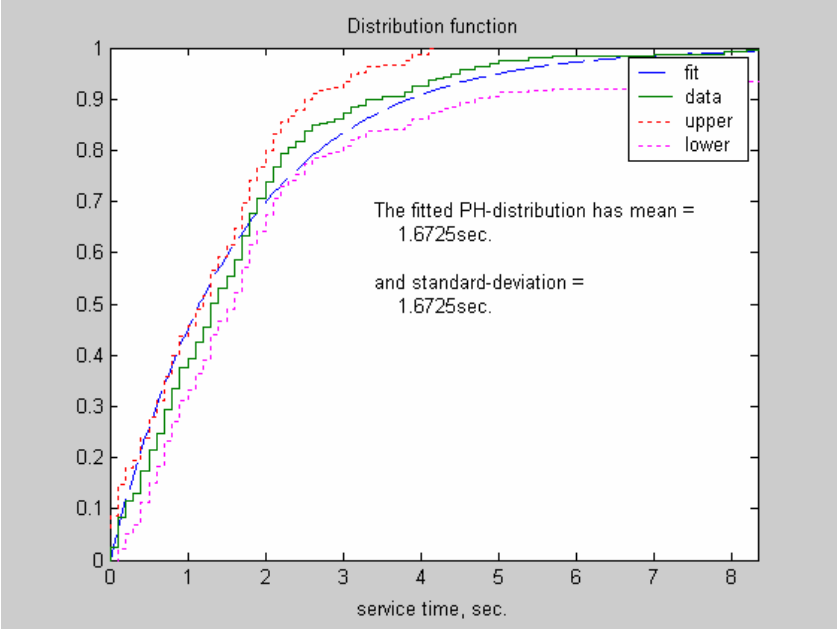


Figure 3.23 The fitted phase type distribution to the sequence $\boldsymbol{\mu}$ - 0.7. The confidence interval around the ECDF of resolution $\pm 1/\sqrt{n}$ (n=251) with fitted PH-distribution of order 1.

The density that fits to $\boldsymbol{\mu}$ - 0.7 is given by,

$$0.6e^{-0.6t}\ ,\ t \geq 0. \tag{3.51}$$

Let $f_{\mu}(\mu)$ be the density of service times which acquired services when the traffic load of Channel 1 is below 0.088 Erlang. Using (3.51) we can write

$$f_{\mu}(\mu) = \begin{cases} 0.6e^{-0.5979(\mu-1)} & ,\ \mu \geq 1\text{sec.} \\ 0 & ,\ \mu < 1\text{sec.} \end{cases} \tag{3.52}$$

We can now compare $E[\Delta_n]$ and $E[\Delta_n{}^2]$ with $E[\mathbf{\Delta}]$ and $E[\mathbf{\Delta}^2]$ respectively. Using equation (3.31) we plot $\varepsilon_1 = \dfrac{(f_1(\lambda, 0.6) - E[\mathbf{\Delta}])}{E[\mathbf{\Delta}]}.100$ and $\varepsilon_2 = \dfrac{f_2(\lambda, 0.6) - E[\Delta^2]}{E[\Delta^2]}.100$ on Figure 3.24.

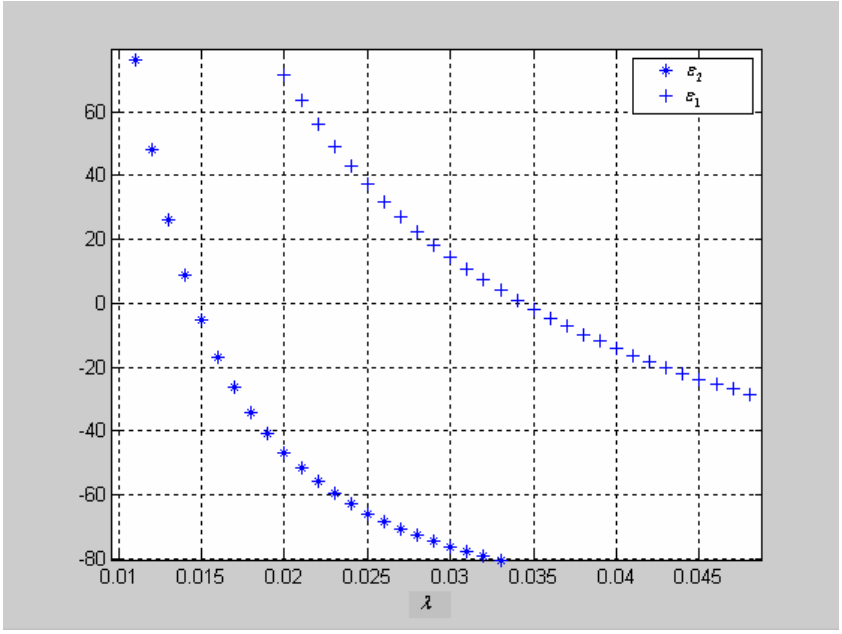


Figure 3.24 ε_1 and ε_2 versus λ

We see that $\lambda_1 \neq \lambda_2$. This again shows that call arrivals are not Poisson distributed.

3.3 Waiting Times in the Queue for M/M/1 and M/G/1 Queues

To obtain information concerning the time an arrival must spend waiting in the queue until entering service, we now proceed to derive the probability distribution for waiting time in the queue for M/G/1 and M/M/1 queues. When considering individual waiting time, however, queue discipline must be specified and we are assuming it is FIFO.

We derived the Laplace-Stieltjes transform of the waiting time in the queue by using equations (3.8) and (3.9) as

$$F_b^*(s) = \frac{\mu e^{-s}}{(s+\mu)} \tag{3.53}$$

$$F_q^*(s) = \frac{(1-\rho)s(s+\mu)}{[s^2 + s(\mu-\lambda) - \lambda\mu + \lambda\mu e^{-s}]} \tag{3.54}$$

The density and distribution of waiting time in the queue are denoted by $F_q(q)$ and $f_q(q)$ respectively. We obtained $f_q(q)$ by both inverse Fourier transform and Padé approximation techniques. These two techniques produced the same results. We found the Fourier transform of $f_q(q)$ by evaluating equation (3.54) at $s = j\omega$. The resulting $F_q^*(j\omega)$ has no pole for $-\infty < \omega < \infty$. Therefore $f_q(q)$ has a Fourier transform. We have taken the inverse Fourier transform of $F_q^*(j\omega)$ to obtain $f_q(q)$,

$$f_q(q) = \frac{1}{2\pi} \int_{\omega=-\infty}^{\omega=\infty} F_q^*(j\omega) e^{j\omega q} d\omega \tag{3.55}$$

For the M/M/1 queue, the density of the service times are given by

$$f_b(b) = \begin{cases} \mu e^{-\mu b}, & b \geq 0 \\ 0, & b<0 \end{cases} \tag{3.56}$$

Laplace-Stieltjes transform of the waiting time in the queue, with density $f_q(q)$, is then given by

$$F_q^*(s) = \frac{(1-\rho)s(s+\mu)}{[s^2 + s(\mu-\lambda)]} \tag{3.57}$$

To obtain $f_q(q)$, we found its Fourier transform by evaluating equation (3.57) at $s = j\omega$. The resulting $F_q^*(j\omega)$ has no pole for $-\infty < \omega < \infty$. Therefore $f_q(q)$ has a Fourier transform. We have then taken the inverse Fourier transform of $F_q^*(j\omega)$ to obtain $f_q(q)$.

Padé approximations are useful for representing time delays by rational LTI models [11]. We expressed the e^{-s} term in the denominator of equation (3.54) by a rational transfer function obtained by Padé approximation formulas. Using an 8'th order Padé approximation, the function e^{-s} can be represented by

$$e^{-s} = \frac{P_n(s)}{Q_d(s)} \tag{3.58}$$

where

$$P_n(s) = p_0 s^8 + p_1 s^7 + p_2 s^6 + ... + p_8$$

$$Q_d(s) = q_0 s^8 + q_1 s^7 + q_2 s^6 + ... + q_8$$

We substituted this expression in equation (3.54) and then plotted the impulse response of the resulting $F_q^*(s)$ to obtain $f_q(q)$. The resulting $f_q(q)$ obtained by using an 8'th order Padé approximation of e^{-s} has resulted in the same $f_q(q)$ obtained by inverse Fourier transform technique. Therefore, in sections 3.3.1, 3.3.2 and 3.3.3, the plots of $f_q(q)$ are obtained by inverse Fourier transform technique. High-order Padé approximations produce transfer functions with clustered poles. Because such pole configurations tend to be very sensitive to perturbations, Padé approximations with order N>10 should be avoided [11].

In the following sections, we present some numerical results based on the channel holding time statistics of Channel 1, 2 and 3 under different traffic loads.

3.3.1 Numerical Results Based on Statistical Behavior of Channel 2

In this section we plotted $f_q(q)$ with the traffic parameters of Channel 2. We adjusted λ and μ parameters according to the traffic load of Channel 2.

When the traffic load of Channel 2 is maximum, the service rate μ is equal to 0.5 and $\lambda_1 = 0.046$. Therefore to determine $f_q(q)$ for the busiest interval of Channel 2, we have set

$$\lambda = 0.046\text{sec}^{-1} \text{ and } \mu = 0.5 \text{ sec}^{-1}$$

Then we plotted $f_q(q)$ for both M/G/1 and M/M/1 queues on Figure (3.25).

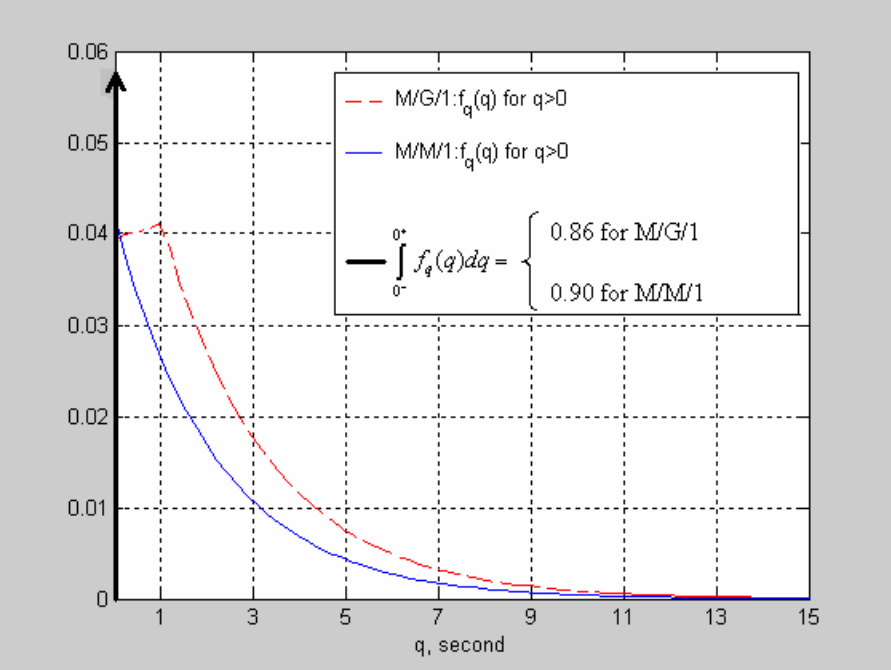


Figure 3.25 $f_q(q)$ for $\lambda = 0.046\text{sec}^{-1}$ and $\mu = 0.5\ \text{sec}^{-1}$.

When the traffic load of Channel 2 is below 0.04 Erlang, the service rate μ is equal to 0.75sec^{-1} and $\lambda_1 = 0.0095\ \text{sec}^{-1}$. Therefore to determine the waiting time in the queue for the nonbusy interval of Channel 2, we have set

$$\lambda = 0.0095\text{sec}^{-1} \text{ and } \mu = 0.75\ \text{sec}^{-1}$$

Then we plotted $f_q(q)$ for both M/G/1 and M/M/1 queues on Figure (3.26).

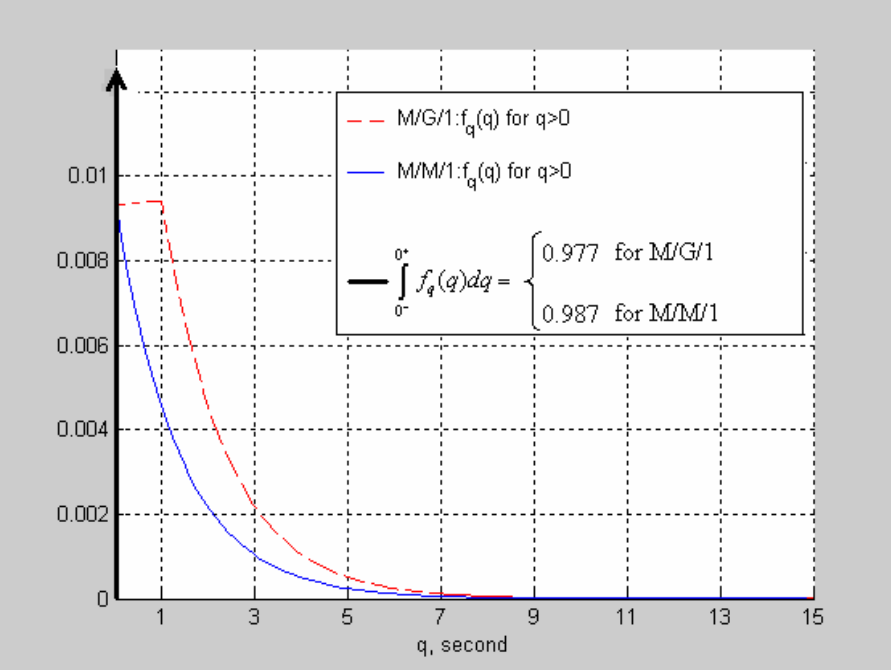


Figure 3.26 $f_q(q)$ for $\lambda = 0.0095\text{sec}^{-1}$ and $\mu = 0.75\ \text{sec}^{-1}$.

3.3.2 Numerical Results Based on Statistical Behavior of Channel 3

When the traffic load of Channel 3 is maximum, then the service rate μ is equal to $0.39\,\text{sec}^{-1}$ and $\lambda_1 = 0.13\,\text{sec}^{-1}$. Therefore to determine $f_q(q)$ for the busiest interval of Channel 3, we have set

$$\lambda = 0.13\text{sec}^{-1} \text{ and } \mu = 0.39\ \text{sec}^{-1}$$

Then we plotted $f_q(q)$ for both M/G/1 and M/M/1 queues on Figure (3.27).

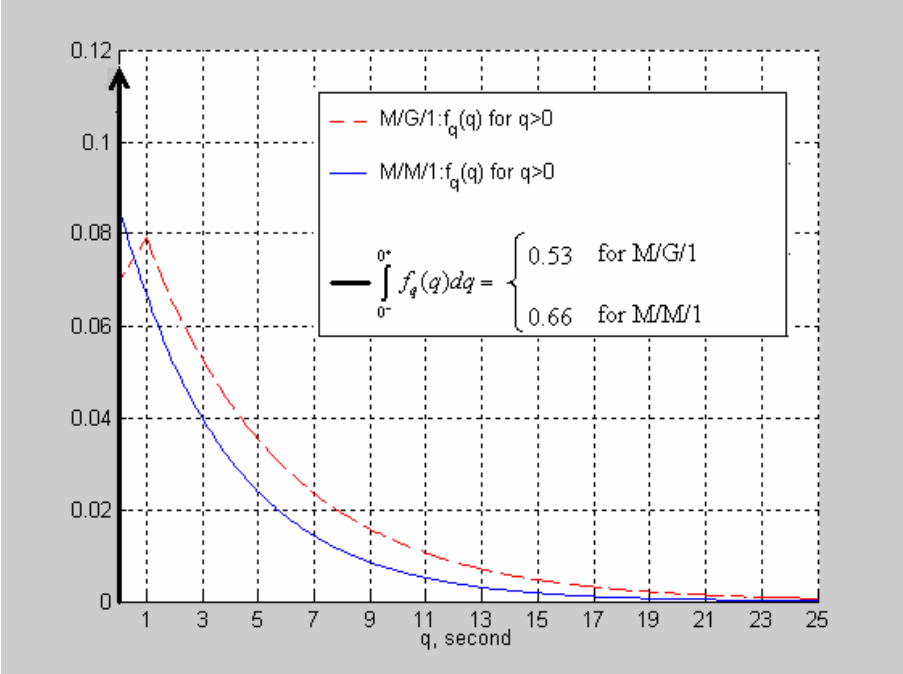


Figure 3.27 $f_q(q)$ for $\lambda = 0.13\text{sec}^{-1}$ and $\mu = 0.39\ \text{sec}^{-1}$.

When the traffic load of Channel 3 is below 0.07 Erlang, the service rate μ is equal to $0.65\,\text{sec}^{-1}$ and $\lambda_1 = 0.023\,\text{sec}^{-1}$. Therefore to determine $f_q(q)$ for the nonbusy interval of Channel 3, we have set

$$\lambda = 0.023\text{sec}^{-1} \text{ and } \mu = 0.65\ \text{sec}^{-1}$$

Then we plotted $f_q(q)$ for both M/G/1 and M/M/1 queues on Figure (3.28).

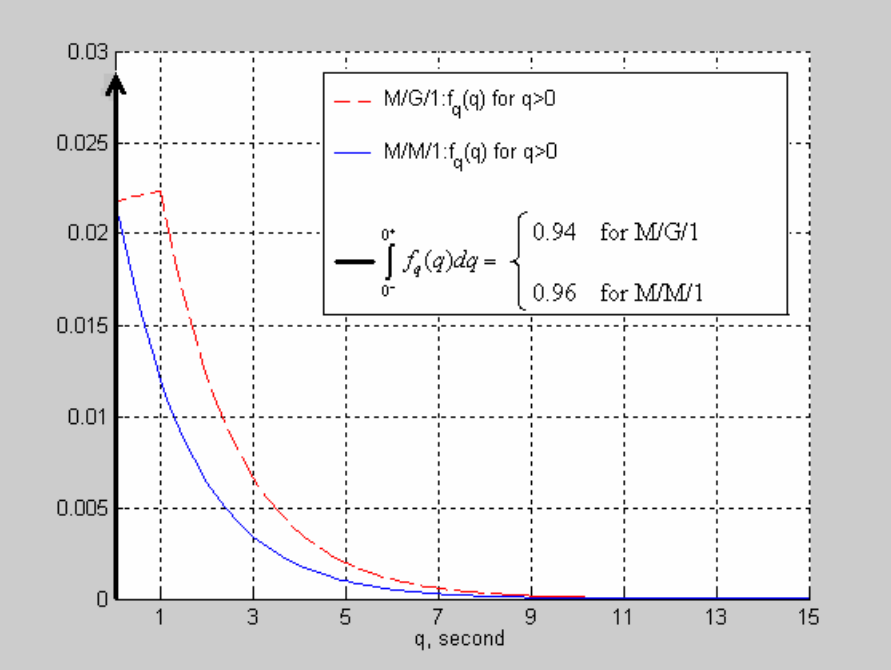


Figure 3.28 $f_q(q)$ for $\lambda = 0.023\text{sec}^{-1}$ and $\mu = 0.65\ \text{sec}^{-1}$.

3.3.3 Numerical Results Based on Statistical Behavior of Channel 1

$F_q^*(s)$ for Channel 1 traffic pattern is given by equation (3.45):

$$F_q^*(s) = \frac{(1-\rho)s(s+\mu_1)(s+\mu_2)}{(s-\lambda)(s+\mu_1)(s+\mu_2)+\lambda[p_1\mu_1(s+\mu_2)e^{-s} - p_2\mu_2(s+\mu_1)e^{-s}]} \quad (3.59)$$

We obtained $f_q(q)$ by both inverse Fourier transform and Padé approximation techniques. These two techniques produced the same results. We found the Fourier transform of $f_q(q)$ by evaluating equation (3.59) at $s = j\omega$. The resulting $F_q^*(j\omega)$ has no pole for $-\infty < \omega < \infty$. Therefore $f_q(q)$ has a Fourier transform. We have then taken the inverse Fourier transform to obtain $f_q(q)$. We expressed the e^{-s} term in the denominator of equation (3.59) by a rational

transfer function obtained by an 8'th order Padé approximation formula. This rational transfer function is given by equation (3.58). We put this expression in equation (3.59) and then plotted the impulse response of the resulting $F_q^*(s)$ to obtain $f_q(q)$.

When the traffic load of Channel 1 is maximum we have

$$\mu_1 = 0.5598\,\text{sec}^{-1},\ \mu_2 = 1.2036\,\text{sec}^{-1},\ p_1 = 1.36,\ p_2 = 0.36,\ \lambda_1 = 0.136\,\text{sec}^{-1}.$$

With these parameters we plotted $f_q(q)$ for the M/G/1 queue on Figure (3.29). We have set $\lambda = \lambda_1 = 0.136\,\text{sec}^{-1}$. Since the service times have a PH-Type distribution with two phases, the M/M/1 solution is not valid.

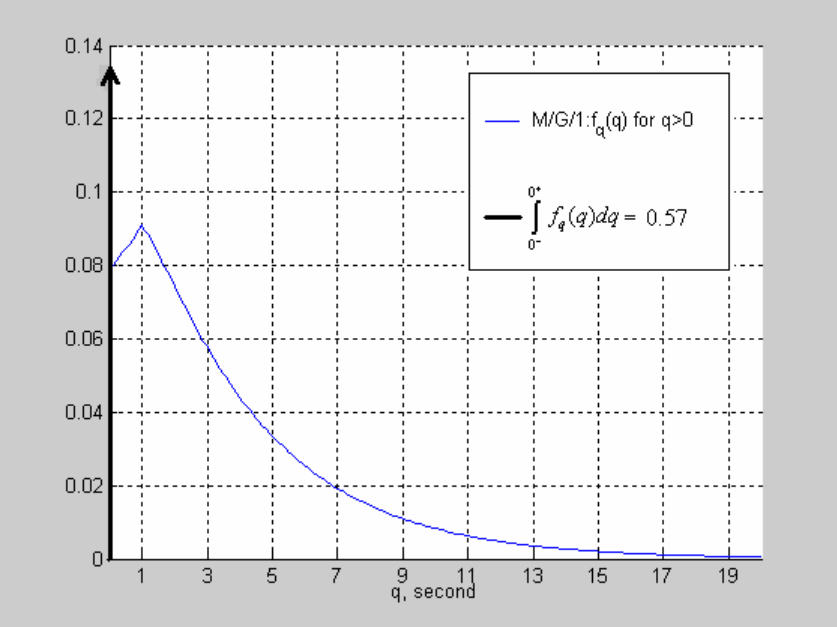


Figure 3.29 $f_q(q)$ versus q for M/G/1 queue

When the traffic load of Channel 1 is below or equal to 0.0881 Erlang we have

$$\mu_1 = 0.6\,\text{sec}^{-1},\ p_1 = 1,\ p_2 = 0,\ \lambda_1 = 0.034\,\text{sec}^{-1}.$$

With these parameters we plotted $f_q(q)$ for both M/G/1 and M/M/1 queues on Figure (3.30). We have set $\lambda = \lambda_1 = 0.034\,\text{sec}^{-1}$.

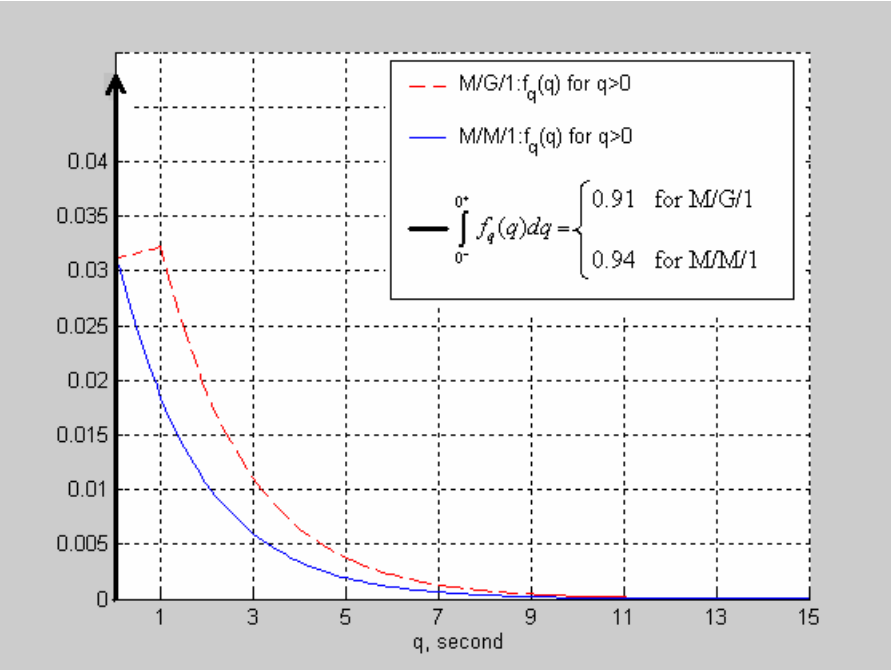


Figure 3.30 $f_q(q)$ for $\lambda = 0.034\text{sec}^{-1}$ and $\mu = 0.6\ \text{sec}^{-1}$.

3.4 Discussion

In this chapter we modeled the distribution of interservice times for a transmission trunked conventional PMR network. Having created a model for the interservice times' density and distribution, its first and second moments are compared to that of interservice time data. From the results presented in section 3.3, we see that the values of λ_1 and λ_2 for different traffic loads of each individual channel are different from each other. This result shows that call arrivals to a transmission trunked PMR network are not Poisson distributed. This

deviation from Poisson distribution must be expected in transmission trunking. Transmission trunking handles each voice activity in a conversation as a new call arrival. Since the voice activity start instants in a conversation are not independent, overall distribution of call arrivals may deviate from Poisson distribution.

For the M/G/1 queuing model, as discussed in section 3.3, the minimum service time is 1 second and the service times are exponentially distributed which is delayed by 1 second. Any user who wishes to make a call continuously listens to the channel. He can push the PTT switch when the channel is free. The waiting time depends on the phase of the previous call in the channel. If the new call attempt arrives within the first 1 second of the ongoing call, then he should wait in the queue at least 1 second. Therefore, the waiting time in the queue is uniformly distributed within the first 1 second. If the new call attempt arrives after the first 1 second of the ongoing call, then he should wait in the queue a random amount of time which is exponentially distributed. This means that the waiting time in the queue becomes exponentially distributed after 1 second. For the M/M/1 queuing model, the minimum service time is 0 second. The waiting time in the queue is exponentially distributed. These effects are depicted on the $f_q(q)$ figures given in Section 3.3. We observed that the distribution of the channel holding time affects the waiting time in the queue.

Chapter 4

Capacity Analysis of a PMR System with DAB Downlink

Several trunked PMR systems have been designed over the last decade, most of which have symmetric downlink and uplink channel capacities. These systems may not be spectrally efficient in case of group or broadcast-based voice and data calls, a common feature of PMR systems. In this chapter, we propose a new asymmetric PMR system comprising a wideband OFDM-based downlink and a narrowband uplink, which not only achieves a better spectral efficiency but also can support high bit rate multimedia applications. The system is shown to have high trunking efficiency since all users are assumed to use the pool of channels available in the wideband downlink. In this chapter, we study the performance and capacity of a PMR system using a DAB downlink. In particular, we study the efficiency of such a system for voice calls using voice activity detection and statistical multiplexing. Moreover, we show that, the efficiency of the system can significantly increase, if the incoming calls, which can not find an available channel, are allowed to wait a certain amount of time before occupying a channel.

A number of digital trunked mobile radio systems have been recently developed in Europe, North America, Japan, and Israel. Although these systems have been developed for either general-purpose applications or more specific users, they share some basic properties and objectives. There are seven well known Professional or PMR systems; Terrestrial Trunked Radio System

(TETRA), Association of Public-Safety Communications Officials (APCO) 25, Integrated Dispatch Radio System (IDRA), Digital Integrated Mobile Radio System (DIMRS), TETRAPOL system, Enhanced Digital Access Communications System (EDACS), Frequency Hopping Multiple Access System (FHMA). Three of the PMR systems, APCO25, TETRAPOL and EDACS, are based on FDMA technology; however the other three, TETRA, IDRA and DIMRS, are based on TDMA [12]. All of the above PMR systems have symmetric downlink and uplink channel capacities. For example in APCO25 both uplink and downlink channel for a particular user has a bandwidth of 12.5 KHz. In TETRA each channel is one fourth of a 25 kHz TDMA channel [2].

Recently, there has been an increasing interest for high data rate multimedia applications and internet services incorporated. An asymmetric system with a wideband downlink for higher data rates is desirable, because it is generally accepted that wideband is needed more for the downlink. Studies for such asymmetric systems have been undertaken in several projects in which the public access communications system, GSM, is complemented with DAB [13] or Digital Video Broadcasting (DVB) [14]. These proposals consider the wideband downlink only for multimedia applications, and voice calls are still carried by the narrowband channels. However if a wideband downlink is available then it may be possible to achieve spectrum efficiency for the speech signals as well if they are carried by the wideband downlink channel via a suitable trunking protocol. Asymmetric PMR systems have not yet been proposed.

In PMR systems there are a number of voice and data applications which are not supported by Public Access Communications Systems such as GSM [2]. These are group (acknowledged or unacknowledged) and broadcast calls for both voice and data. In a cellular symmetric PMR system, in regions where offered traffic is high, cell sizes must be decreased. This in turn, increases the number of calls between users in different cells. If the members of a group call are in different cells, each one of them must be assigned uplink and downlink channels

individually. However if the group members are in the same cell then they use the same uplink and downlink channels. This spectrum efficiency in a PMR system is therefore lost if the group members are in different cells. However in an asymmetric system where the downlink has a wide area coverage, then spectrum efficiency in a PMR system is maintained even with group calls. In PMR systems group calls constitute a major proportion of all calls (50%) [15]. Therefore an asymmetric system is also advantegous when group and broadcast calls in a PMR system are considered.

In this study a new PMR system comprising an OFDM based wideband downlink and a narrowband uplink, based on FDMA or TDMA (TETRA system) technology, is proposed. DAB system is considered as the OFDM based wideband downlink. This study focuses on system capacity and GoS associated with the downlink channel for voice communications. In particular, we study the efficiency of such a system for voice that would be obtained using silence detection. Moreover we show that, if the incoming calls are allowed to wait a certain amonut of deterministic time before occupying a channel, the efficiency can significantly increase.

4.1 Methods

In the following sections methods for calculating the maximum number of users the system can accommodate are developed for a given average frame loss rate (FLR), which is the ratio of the average number of the lost frames to the total number of the frames and GoS, which is the probability of blocking.

It is well known that voice contains an alternating sequence of ON and OFF states (talkspurts and silence gaps) [16]. It is assumed in this study that a voice user generates typically 9.6 kbit/s data during the ON time, including the overhead coming from error correction coding and packet headers. This data rate

corresponds to ≈ 230 bits/slot where one slot corresponds to 24 ms of DAB transmission frame duration. This amount of information per slot per person is called a "frame" hereafter. The whole data transmission capacity, 55296 bits/slot, of DAB is assumed to be used for voice transmission in this study. Therefore 55296 / 230 ≈ 240 users can be in the ON state simultaneously if no loss of information is desired. However, $n_{oc} > 240$ ongoing calls can be admitted at the same time if a certain amount of BER is allowed. In the proposed system, the maximum number of ongoing calls, N_{oc}, is limited by the amount of allowable FLR. At any slot, if n_{on}, which is the number ongoing calls in the ON state, is more than 240, then n_{on}-240 frames are assumed to be lost.

Once N_{oc} is determined for a specified (given) value of FLR, the total number of users which can be accommodated (subscribed) by the system is calculated by considering call-level statistics of the users. The whole system is modeled as a continuous time Markov chain. In the following, three approaches are used for obtaining numerical results. The first approach is based on the analytical solution of the Continuous Time Markov Chain model (section 4.2). The second one uses numerical simulations (section 4.3) and the third one is based on the Discrete Time Markov Chain Model of the system (section 4.5).

4.2 Continuous Time Markov Model of the System and Analytical Solution

The system is modelled as two interacting Markov chains, one for call-level characterization and the other for ON-OFF level characterization. It is assumed that calls arrive at a rate λ (calls/hour) and the average call duration is $1/\mu$ (hour). Arrivals are assumed to be Poisson distributed and the call duration is assumed to be exponentially distributed. Moreover, both ON and OFF durations are assumed to be exponentially distributed with mean $1/\alpha$ (hour) and $1/\beta$ (hour) respectively. The state of the system can be described by the number of ongoing calls in the

system, *j*, and the number of calls in that are in talkspurts $i \le j$ [17]. The state (*j, i*) forms a two-dimensional Markov Chain, as shown in Figure 4.1. The upper Markov chain in Figure 4.1 is called the call-level model, and the lower one is called the ON-OFF level model.

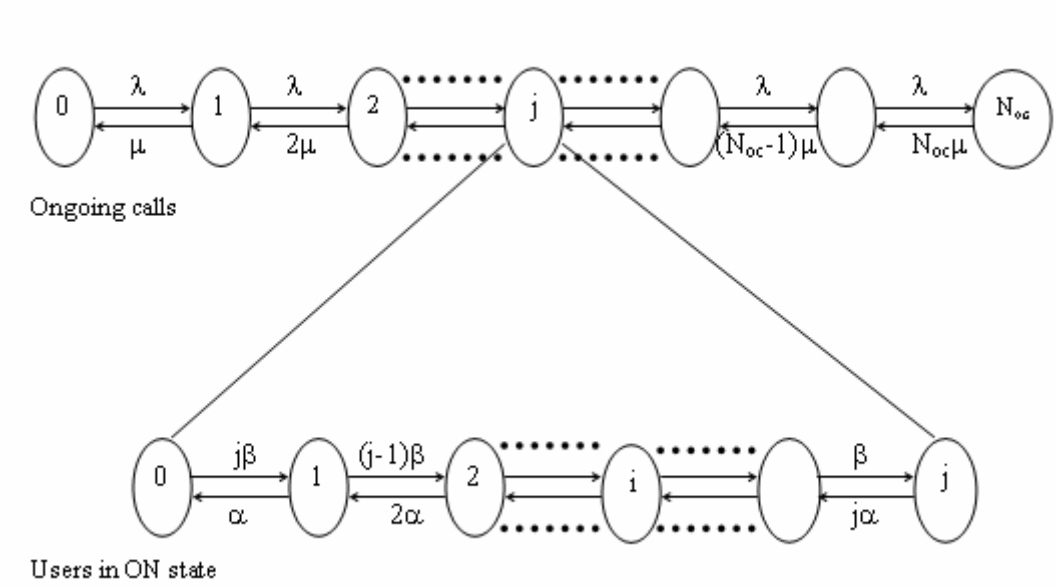


Figure 4. 1 Markov chain model of the proposed system

It is straightforward to show that the steady-state probability of having *i* calls out of *j* being in the ON state is given by [18];

$$B_i = \frac{\binom{j}{i}\left(\beta/\alpha\right)^i}{\left(1+\beta/\alpha\right)^j}, \quad 0 \le i \le j \tag{4.1}$$

FLR is calculated by the ratio of the expected number of lost frames to the expected number of total frames using

$$FLR = \frac{\sum_{i>240}^{N_{oc}} (i-240)B_i}{\sum_{i=0}^{N_{oc}} iB_i} \qquad (4.2)$$

where N_{oc} is the maximum number of calls that can be supported simultaneously.

With respect to call level characterization the system can be described by an M/M/c/c queueing model in which the state of the system is the number of ongoing calls. To find GoS we must now analyze the call level Markov chain. The probability of having *j* ongoing calls at any observation time is:

$$P_j^{N_{oc}} = \frac{\left(\lambda/\mu\right)^j / j!}{\sum_{k=0}^{N_{oc}} \left(\lambda/\mu\right)^k / k!} \quad , 0 \le j \le N_{oc} \qquad (4.3)$$

The call blocking probability can be derived using the P_j values from equation 4.3 which is the well known Erlang-B formula [19],

$$P_B = \frac{\left(\lambda/\mu\right)^{N_{oc}} / N_{oc}!}{\sum_{k=0}^{N_{oc}} \left(\lambda/\mu\right)^k / k!} \quad , 0 \le j \le N_{oc} \qquad (4.4)$$

For a given value of FLR, the maximum number of users that can be supported by the system simultaneously, N_{oc}, is calculated from equation (4.2) by a binary search algorithm. Later, λ is calculated from equation (4.4) for a given values of P_B, N_{oc} and μ. Since $\lambda = \lambda_u N_{pop}$ where λ_u is number of calls per hour per user, N_{pop}, which is the total number of users accommodated by the system, i.e. the number of subscribers, can be calculated by

$$N_{pop} = \lambda / \lambda_u \quad (4.5)$$

We also study a variant M/M/c/∞ queueing model with reneging. In this model, if a user is not directly admitted to the system, that user is willing to wait a certain amount of deterministic time. The user waits until this maximum waiting time, and then he reneges if service has not yet been provided. We analyzed such a system by simulations, since there exist little work on queueing models with deterministic waiting time in literature.

4.3 Solution by Simulation

The simulation method includes the following three steps:

4.3.1 Finding N_{oc} Using ON-OFF Simulation

For each of n_{oc} ongoing calls, 3 hour-long ON-OFF patterns are generated. These ON and OFF periods are exponentially distributed with mean $1/\alpha$ and $1/\beta$.

The simulation period, 3 hours, is divided into slots; lasting 24 ms each. During the steady state region, for each slot, n_{on} is determined. If $n_{on} > 240$, then n_{on} -240 frames are assumed to be lost since a maximum of 240 calls in ON state can be supported by the system. If $n_{on} \leq 240$, then all voice information is sent successfully. FLR is calculated as the ratio of total lost frames to total frames sent during the steady state region. N_{oc} is taken to be the value of n_{oc} for which desired FLR is obtained.

For any given value of n_{oc}, the 3 hour-simulation is repeated 20 times to find 20 FLR's which are averaged to find the FLR corresponding to that $n_{oc.}$ For n_{oc}=500, FLR was found to be 1.67e-04 with a 95 % confidence interval of ±5.8e-06. Therefore 20 repetitions of 3-hour simulations were found to be sufficient.

4.3.2 Finding N_{pop} Using Call-level Simulation without Queueing

N_{pop} is the maximum number of users i.e. subscribers that can be supported by the system. This step of the simulation is performed for a longer period, typically 150 hours. Call arrivals are Poisson distributed; therefore interarrival time is exponentially distributed. The steps in the call level simulation without queueing are:

1) For the 150-hour simulation period, random call arrival times are generated using an exponentially distributed random number generator with mean $1/\lambda$.

2) For each arrival point, i.e., for each call, call duration is assigned using the same random number generator with mean $1/\mu$.

3) Total simulation time is divided into 24ms slots. When a new call arrives in a certain slot, if there are already N_{oc} ongoing calls in that slot, then that call is blocked. Otherwise the system can support that new call and the call is admitted. In other words "blocked calls cleared" strategy is employed.

4) Call blocking probability, i.e. GoS, is determined as the ratio of the number total blocked calls to the number total calls in the 150-hour period.

For a given value of GoS, the above procedure is repeated using a simple search algorithm to find the corresponding λ. N_{pop} is then found using equation (4.5).

4.3.3 Finding N_{pop} Using Call-Level Simulation with Queueing

This method is similar to the previous one except for the blocking mechanism. Again call arrivals are Poisson distributed. In this simulation, blocked calls are put into a FIFO queue and stay in that queue for a maximum of WT (waiting time) slots before they enter the system. If they cannot enter the system within WT time slots they are blocked.

The steps of this simulation are the same as the simulation without queueing except for the third step. In the third step, for a new call arrival, if there are already N_{oc} ongoing calls, then that call is placed in a FIFO queue where that call can maximally stay for WT time slots. If there are less than N_{oc} ongoing calls, the system can support a new call and that call is admitted. The system accepts the calls waiting in the queue in each time slot if there are lower than N_{oc} ongoing calls. This strategy is called "reneging".

In the following section, we give the numerical results that we obtained with the simulation methods described in this section.

4.4 Results

The analytical and simulation methods explained in sections 4.2 and 4.3 are applied for two scenarios. In the first scenario, the system is assumed to be used for GSM type voice calls while in the second scenario the system is assumed to be used for PMR type voice calls. The simulation input parameters are shown for each scenario in Table 4.1.

Table 4.1 System parameters for GSM and PMR scenarios

	GSM	PMR
Average call arrival rate (λ calls/hour/user)	2	10
Average call duration (1/μ sec)	180	20
Average duration in ON state (1/α sec)	1	1
Average duration in OFF state (1/β sec)	1.35	1.35

4.4.1 Number of Ongoing Calls, N_{oc}:

The variation of FLR with N_{oc} as calculated by both the analytical and simulation method is given in Figure 4.2. If a frame loss rate of 10^{-4} is accepted, the maximum number of users that can be simultaneously supported by the system is found to be approximately 500 from simulations for both GSM and PMR scenario, since ON-OFF patterns are assumed to be the same for both scenarios.

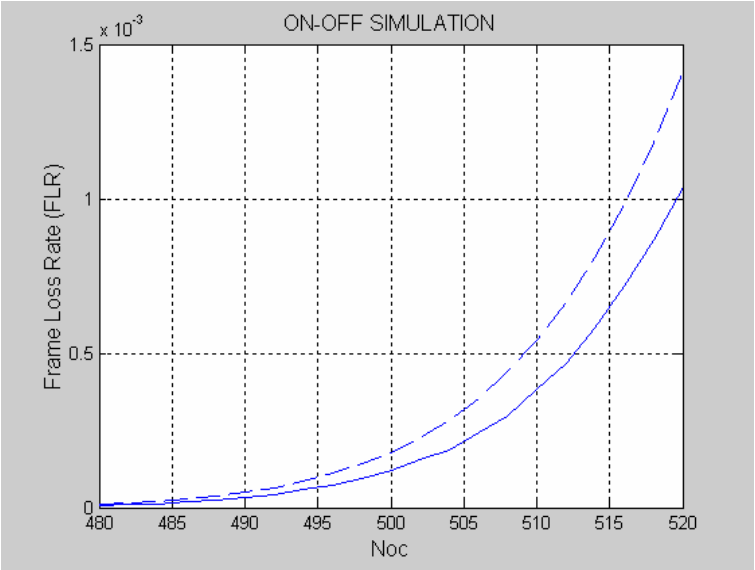


Figure 4. 2 Frame loss rate as a function of number of ongoing calls. The dash-line stands for the analytical results and the solid line stands for the simulation results.

As it is obvious in Figure 4.2, the results of the simulation and analytical methods do not agree. Thus, the analytic method that we used to determine FLR is not quite representative for our system. In order to model our system correctly, we created a simple discrete-time Markov model for the Frame Loss Rate of the sytem with homogeneous ON-OFF sources. This method is presented in section 4.5.

4.4.2 Number of the Users Supported by the System, N_{pop} , without Queueing

Using the call level simulation without queueing, N_{pop} is found for both GSM and PMR scenarios. In Figures 4.3 and 4.4, call-blocking probability vs. N_{pop} is plotted for both scenarios with analytical and simulation results. Simulation results are very close to analytical results. For a typical GoS of 2 %, and frame loss rate of 10^{-4}, population that is supported by the system is found to be 4865 for GSM and 8765 for PMR scenario.

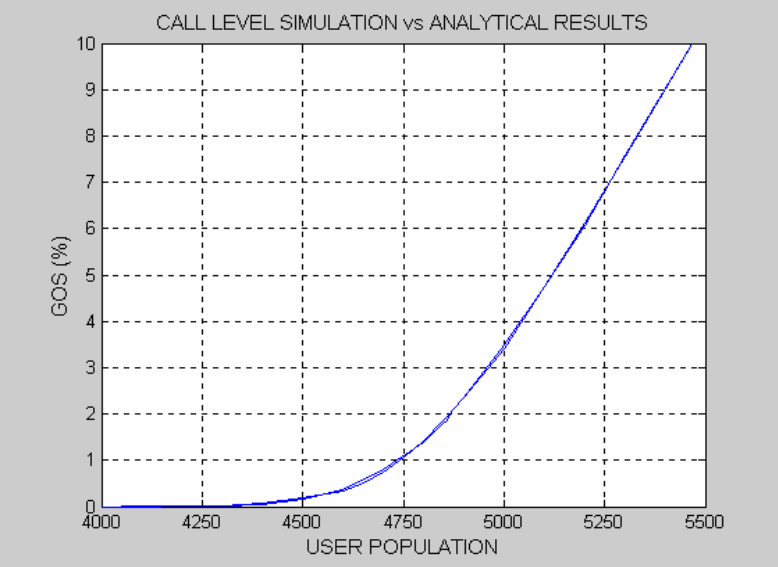


Figure 4. 3 GoS:Grade of Service providing on the average 10^{-4} FLR for GSM scenario. For GoS of % 2, optimum N_{pop} is found to be 4865.

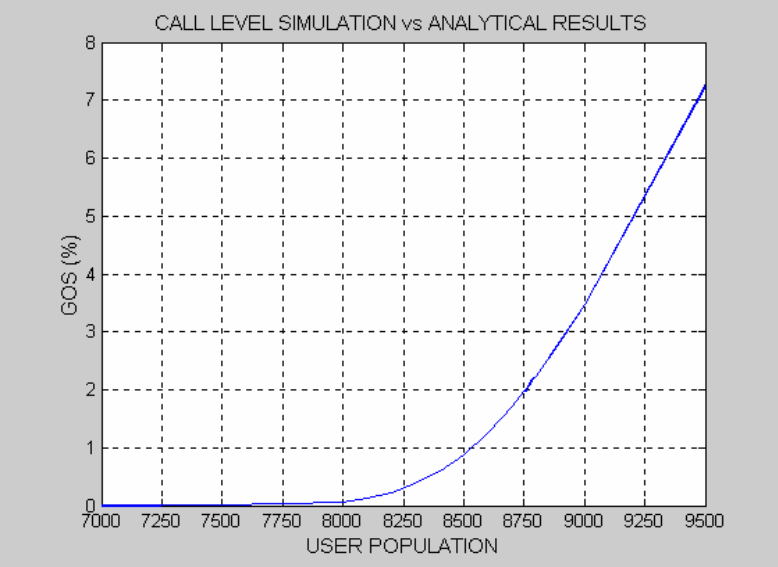


Figure 4. 4 GoS: Grade of service providing on the average 10^{-4} FLR for PMR scenario. For GoS of % 2, optimum N_{pop} is found to be 8765.

4.4.3 N_{pop} with Queueing

In the simulations, which also consider a maximum waiting time, user population is taken as 4865 for GSM scenario, and 8765 for PMR scenario. In Figures 4.5 and 4.6, call-blocking probability is plotted as a function of waiting time. As expected, call blocking probability decreases as the waiting time increases. While the total number of users that generate traffic does not change, for GSM scenario, a maximum waiting time of 2.16 second (90 slots) provides nearly zero call-blocking probability. Since for the PMR scenario call durations are much smaller than GSM call durations, the decreasing rate of call blocking probability is higher than that of GSM as seen in Figure 4.6. A maximum waiting time of 0.48 second (20 slots) supports nearly zero call-blocking probability in PMR scenario.

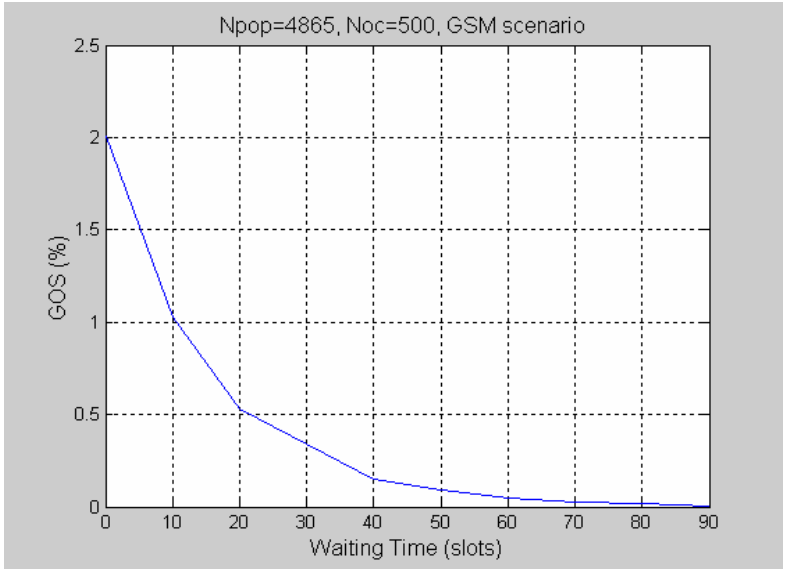


Figure 4. 5 GoS as a function of maximum waiting time for GSM scenario where N_{pop}= 4865, FLR = 10^{-4}.

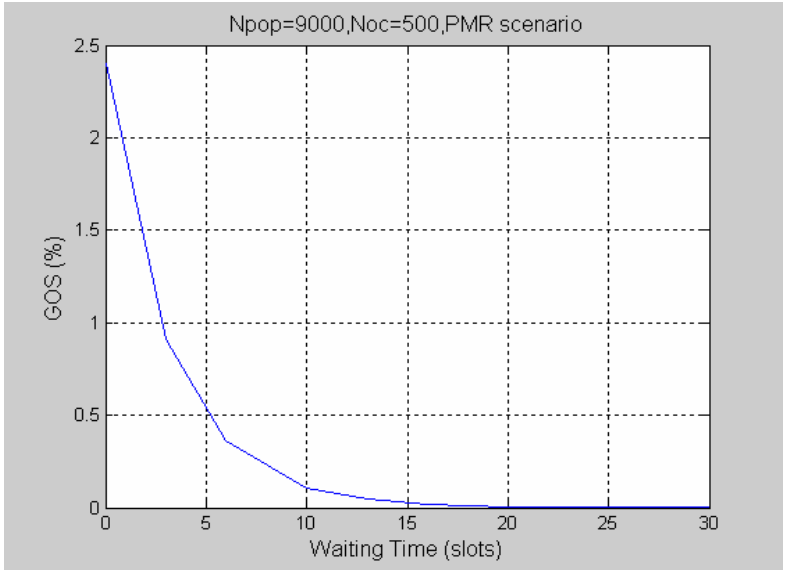


Figure 4. 6 GoS as a function of maximum waiting time for PMR scenario where N_{pop} = 9000.

For a given fixed call blocking probability, the system user capacity increases as the waiting time increases. As seen in the Figure 4.7, for GSM scenario of %2 GoS, system user capacity is 4865 when maximum waiting time

is zero. When approximately 2.2 second maximum waiting time is allowed, the system capacity, in terms of population, increases as much as three times.

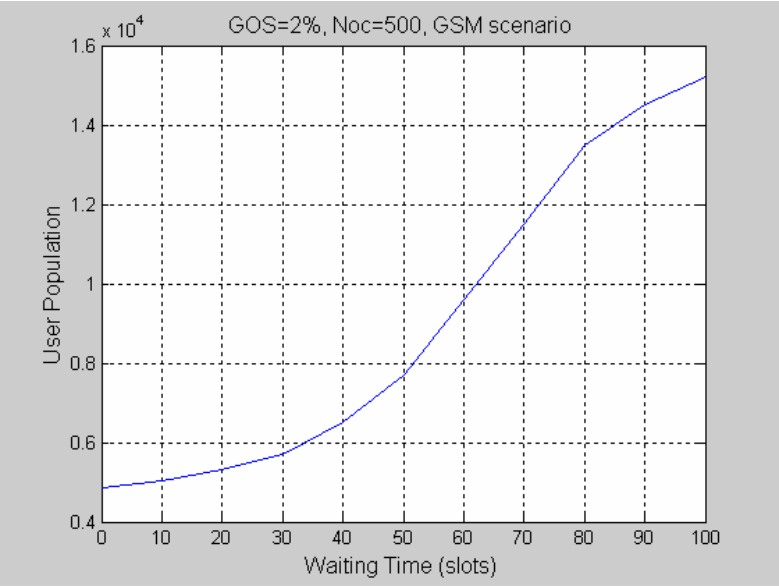


Figure 4.7 The maximum number of users that can be supported by the system as a function of maximum waiting time for GSM scenario for 2% GoS and 10^{-4} frame loss rate.

For PMR scenario with %2 GoS, system user capacity is 8765 when maximum waiting time is zero. As shown in Figure 4.8 if 2.88 second maximum waiting time is allowed, the system capacity increases to approximately 20000.

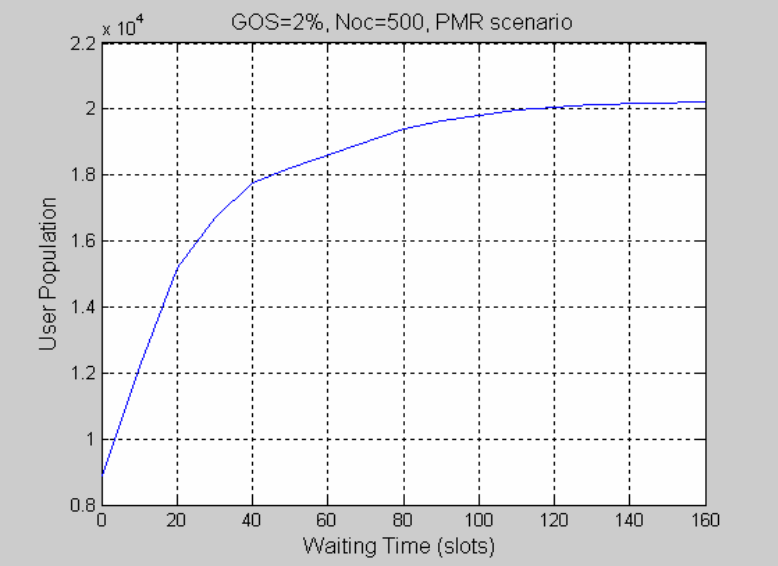


Figure 4. 8 The maximum number of users that can be supported by the system as a function of maximum waiting time for PMR scenario for 2% GoS and 10^{-4} frame loss rate.

4.5 Simple Discrete Time Markov Model for the Frame Loss Rate of the Sytem with Homogeneous ON-OFF Sources

As it is obvious in Figure 4.2, the variation of FLR with N_{oc} as calculated by both the analytical and simulation methods do not agree. The reason for this difference is as follows: The Markov chain (ON-OFF level model) given in Figure 4.1 assumes that only transition between neighboring states are possible. However in our system because of the 24ms slot period, we observed in the simulations that n_{oc} can change more than ±1 in each slot. This shows that the analytical method that we used to determine FLR is not quite representative for our system. In this section, we created a simple discrete-time Markov model for the Frame Loss Rate of the sytem with homogeneous ON-OFF sources.

4.5.1 Model Assumptions

- N_{oc} identical sources with two states (ON, OFF).
- Sources in the ON state generate packets with rate 9.6 kbps.
- Sources in the on state generate 9.6 kbps x 24 msec.=230 bits/slot.
- Sourcess in the off state do not generate any bits.
- There is one output transmission link with the transmission capacity of 55296 bits/slot.
- There are 55296/230=240 channels available in one slot.
- If more packets arrive than the total number of channels available then those extra packets are lost.

4.5.2 Source Process

Assume that the behaviour of a voice source can be described by a discrete-time Markov chain (DTMC) with two states (ON-OFF). We consider a source that alternates between active and idle states according to a Markov chain (Figure 4.9). The distribution of the length of the ON periods is assumed to be geometrical with parameter 1- α, while the OFF periods are also geometrical with parameter 1-β .

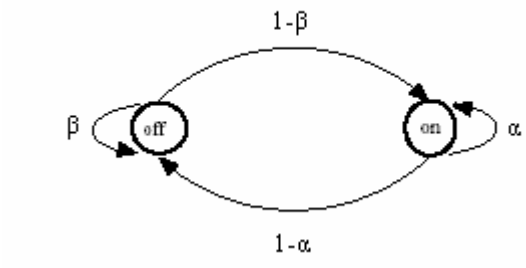


Figure 4. 9 The interrupted Bernoulli Process

The transition probabilities of this DTMC are:

$$\Pr[\text{OFF} \longrightarrow \text{ON}] = 1-\beta$$
$$\Pr[\text{ON} \longrightarrow \text{OFF}] = 1-\alpha$$

Let X_n denote the number of sources in ON state at time epoch n. It is obvious that this process is also DTMC with state space $S=\{0,1,...,N_{oc}\}$ and the state transition probabilities can be written as [20]:

$$p_{ij} = \sum_{l=0}^{i} [\binom{i}{l}(1-\alpha)^l \alpha^{(i-l)} \binom{N_{oc}-i}{j-i+l}(1-\beta)^{j-i+l} \beta^{N_{oc}-j-l} I(j-i+l>0) I(N_{oc}-j-l>0)] \tag{4.6}$$

where I(x)=1 iff logical expression x is true and 0 otherwise. The probability transition matrix is then given by :

$$\mathbf{P}=\{p_{ij}\},\ 0 \le i \le N_{oc},\ 0 \le j \le N_{oc} \tag{4.7}$$

This expression of the transition probabilities takes into account that the transition from state i to state j may occur if *l* of the i sources in the ON state make transition to OFF state and that m, $0 \le m \le N_{oc}$-i, of the sources in the OFF state make a transition to ON state, and that i-*l*+m=j [20]. The probability of the first event to occur is given by binomial probability density function :

$$\binom{i}{l}(1-\alpha)^l \alpha^{i-l}$$

The probability of the second event to occur is given by :

$$\binom{N_{oc}-i}{j-i+l}(1-\beta)^{j-i+l} \beta^{N_{oc}-j-l}$$

4.5.3 Long-run Behaviour of the Discrete Time Markov Chain

Let $\boldsymbol{\Pi}^{(n)}=\{\pi_i^{(n)}\}$ ($\pi_i^{(n)}=\Pr\{X_n=i\}$) denote the state probability vector of process X_n at time epoch n and $\boldsymbol{\Pi} = \{\pi_i , i=0,1,...,N_{oc}\}$ denote the steady-state probability vector of X_n, we then have

$$\lim_{n\to\infty} \boldsymbol{\Pi}^{(n)} = \lim_{n\to\infty} \boldsymbol{\Pi}^{(n+1)} = \boldsymbol{\Pi}$$

so that

$$\boldsymbol{\Pi} = \boldsymbol{\Pi}\,\mathbf{P} \tag{4.8}$$

$$\boldsymbol{\Pi}\,\mathbf{e} = 1 \tag{4.9}$$

or equivalently,

$$\boldsymbol{\Pi}\,(\mathbf{I} - \mathbf{P}) = \mathbf{0} \tag{4.10}$$

These well-known equations are called the *stationary equations* of the Markov chain, and their solution is called the *stationary distribution* [9]. Note that **e** is a column vector with all elements equal to one, **P** is given by equations (4.6) and (4.7) , **I** is $(N_{oc}+1)$-by-$(N_{oc}+1)$ identity matrix.

4.5.4 Numerical Results

The input parameters of the system are:

Average duration in ON state ($1/\eta$ sec)	1
Average duration in OFF state ($1/\tau$ sec)	1.35

Since we are looking at the system at time epochs of 24msec, we can find 1-β and 1-α by (1-β)/0.024= τ =1/1.35 and (1-α)/0.024= η = 1. For a given N_{oc}, the probability transition matrix given by equations (4.6) and (4.7) is constructed. Then equations (4.9) and (4.10) are solved to find $\boldsymbol{\pi}$. FLR is calculated by the ratio of the expected number of lost frames to the expected number of total frames using:

$$FLR = \frac{\sum_{i>240}^{N_{oc}} (i-240)\pi_i}{\sum_{i=0}^{N_{oc}} i\pi_i} \tag{4.11}$$

The variation of FLR is depicted on Figure 4.10. The solid line represents the CTMC solution which is given by equation (4.2), the dashed line represents the the results obtained by simulation method presented in section 4.3.1 and the plus signs show DTMC solution which is given by equation (4.11).

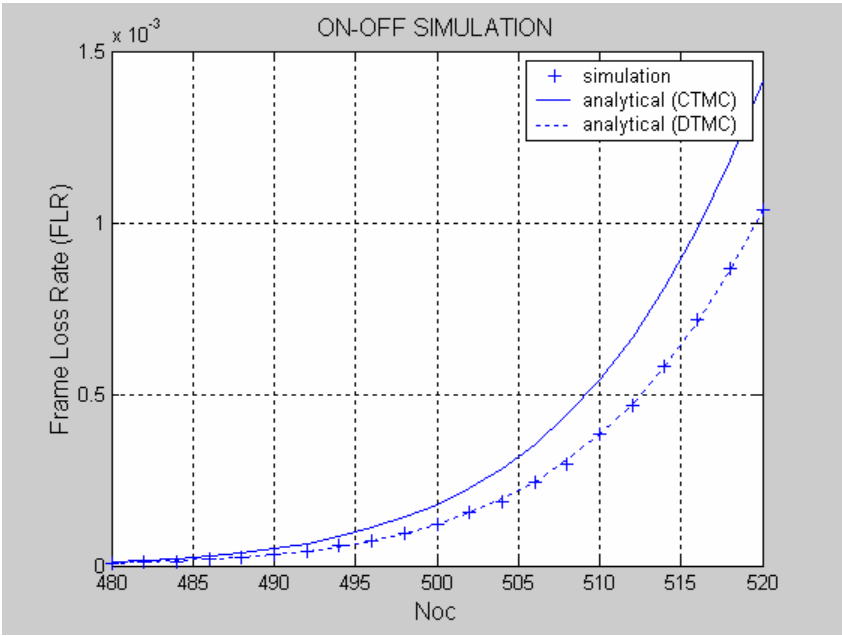


Figure 4. 10 The variation of FLR with N_{oc}.

Figure 4.10 shows that the discrete-time Markov model developed in this section accurately models the FLR variation of our system.

4.6 Capacity Analysis of the System with Coxian Channel Holding Time Distribution

In this section, methods for calculating the maximum number of users the system can accommodate are developed. For this purpose we assumed that the channel holding time distribution is coxian instead of exponential. We used the call duration statistics obtained in section 2.4.1.

4.6.1 Number of the Users Supported by the System, N_{pop}, without Queueing

N_{pop} is the maximum number of users, i.e. subscribers that can be supported by the system. This step of the simulation is performed for a 15 hour period. Call arrivals are Poisson; therefore interarrival time is exponentially distributed. The steps in the call level simulation without queueing are:

1) For the 15-hour simulation period, random call arrival times are generated using an exponentially distributed random number generator with mean $1/\lambda$.

2) For each arrival point, i.e., for each call, call duration is assigned using a coxian distributed random number generator with mean $1/\mu$.

3) We assume that each user making a call constantly generates 230 bits/slot irrespective of being in the ON or OFF state. Therefore the total number of downlink channels is set to 240. Total simulation time is divided into 24ms

slots. When a new call arrives in a certain slot, if there are already 240 ongoing calls in that slot, then that call is blocked. Otherwise the system can support that new call and the call is admitted. In other words "blocked calls cleared" strategy is employed.

4) Call blocking probability, i.e. GoS, is determined as the ratio of the total number of blocked calls to the total number of calls in the 15-hour period.

This is the simulation of an M/G/240/240 system, which is a classical Erlang loss system. For a given value of GoS, the above procedure is repeated using a simple search algorithm to find the corresponding λ. N_{pop} is then found using equation (4.5).

4.6.2 Results

Using the call level simulation without queueing, N_{pop} is found. The simulation input parameters are shown in Table 4.2.

Table 4.2: System parameters

Average call arrival rate (λ calls/hour/user)	0.9 for Ch 1 2.2 for Ch 3
Average call duration ($1/\mu$ sec)	2.86 for Ch 1 3.25 for Ch 3

In Figures 4.11 and 12, call-blocking probability vs. N_{pop} is plotted by using exponentially distributed service times and coxian distributed service times both having the same mean. The variation of GoS in both cases are similar. In a classical Erlang Loss system, the GoS depends on the service-time distribution only through its mean [22].

For a typical GoS of 2%, population that is supported by the system is found to be 315 000 for Channel 1 and 114 000 for Channel 3.

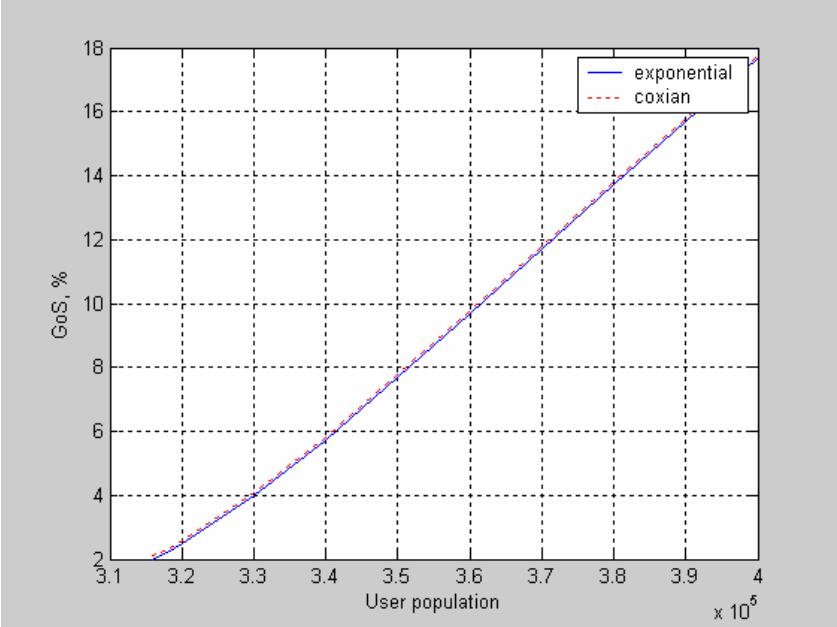


Figure 4. 11 Results for Channel 1

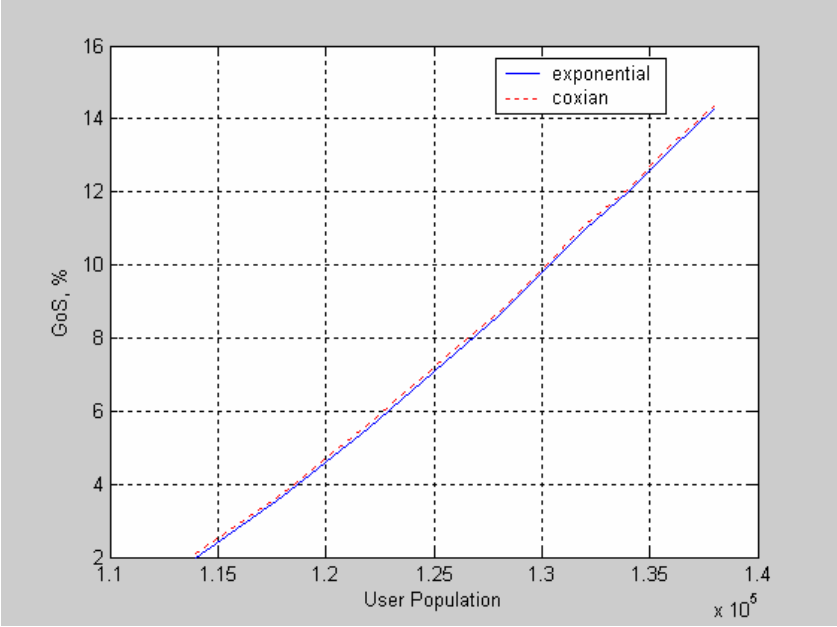


Figure 4. 12 Results for Channel 3

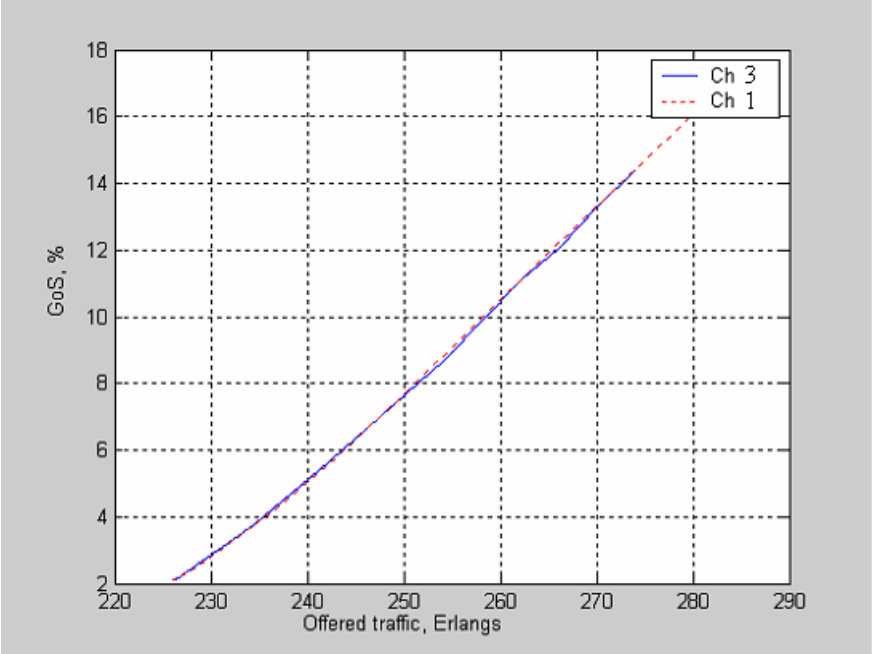


Figure 4. 13 GoS versus offered traffic for Channel 1 and 3.

Offered traffic per user is calculated by multiplying call arrival rate (per user) by mean connection holding time:

$$\text{Channel 3:}\frac{2.2\text{x}3.25}{3600} = 0.002\text{Erlangs}$$

$$\text{Channel 1:}\frac{0.9\text{x}2.86}{3600} = 7.15\ 10^{-4}\text{Erlangs}$$

Figure 4.13 is a plot of GoS versus offered traffic, where offered traffic is calculated by multiplying user population with offered traffic per user. For a GoS of 2%, user population for Channel 1 and 3 are 315 000 and 114 000, respectively. In this model, if a user who is not directly admitted to the system is allowed to wait in a FIFO queue a finite amount of time instead of indefinite waiting time, then the user population supported by the system would increase considerably. We can conclude from these results that only a small proportion of an existing DAB system's capacity needs to be allocated to the downlink of a transmission trunked PMR system.

4.7 Discussion

In section 4.4.3, we found that for 2 % GoS, 10^{-4} FLR and 2.88 maximum waiting time approximately 20000 PMR users can be supported by the proposed system. It is obvious that the specifications of PMR system can be relaxed to for examle 5 % GoS, 10^{-3} FLR and 4 second maximum waiting time. In such a case a considerable increase in the number of users will be achieved. In general the number of subscribers of a PMR system is low. For example in a big metropol like Istanbul there are about 8000 users of the police PMR system. We can conclude from these observations that only a small proportion of an existing DAB system's capacity needs to be allocated to the downlink of a PMR system. Therefore the proposed system offers new service opportunities for existing DAB operators.

A PMR system as proposed in this study also has the capability of supporting high data rate downlink applications such as multimedia services.

In this study an acceptable value for FLR is taken to be 10^{-4} . In speech communications typically a BER of 10^{-3} is considered to be appropriate [2]. We have taken a smaller FLR because from the point of view of transient behaviour (the frames lost have 230 bits each) a more conservative loss rate would be appropriate.

Chapter 5

Results and Conclusions

The evolution of digital PMR networks and increasing demand for data applications by PMR users increase the importance of the teletraffic analysis of PMR networks. The results presented in this study are useful to evaluate the performance of PMR networks with a better degree of accuracy.

A statistical model of the channel holding time for PMR systems is presented. The exponential distribution often used to model trunked systems with the M/M/C queue is different than what is observed. We modeled the channel holding time distribution of three individual channels of a conventional PMR network. For Channel 2 and 3, the density of the channel holding time is a shifted exponential which is delayed by 0.7 second. The mean channel holding times for Channel 2 and 3 are 2.1 seconds and 2.64 seconds respectively. The channel holding time distribution of Channel 1 consists of two exponentials which are both delayed by 0.7 second. The rates of them are $0.65\,\mathrm{sec}^{-1}$ and $1.64\,\mathrm{sec}^{-1}$ respectively. These results show that a transmission trunked PMR network can be designed for short call holding times which enables a large number of users to be accommodated within a particular frequency allocation. Since we do not have empirical observations of data traffic, we could not model the data traffic load of PMR networks. Methods for estimating the data traffic load can be developed.

We developed an analytical model for interservice times for a single server queue of the transmission trunked PMR network. M/G/1 queuing model is used. The model is developed using the channel holding time distribution of Channel 1, 2 and 3. We showed that voice call arrivals to a transmission trunked PMR network are not Poisson distributed. However we could not model the distribution of call arrivals. Analytical and simulation methods based on M/G/C models may not model the system as well as G/G/C models. We do not know the distribution of G regarding the call arrivals. As a future study, methods for modeling the call arrival statistics of voice calls can be determined and simulations based on G/G/C and M/G/C models can be used. We also modeled the waiting time in M/M/1 and M/G/1queues. We observed that the distribution of service times has an effect on the distribution of waiting time in a single server queue. The waiting time distribution of an M/M/C and M/G/C system can be derived to obtain further information concerning waiting time in a trunked system. As a future study, various GoS parameters such as the allocation of frequency channels to each base station of a trunked PMR network can be estimated by simulating an M/G/C system.

The design of a PMR system with DAB downlink proposed in this work can be used to improve the efficiency and cost in case of group or broadcast-based voice and data calls, a common feature of new PMR systems. The system is assumed to be used for voice calls. We assume that call arrivals are Poisson. We decoupled the design process into two major assumption sets. First major assumption set is as follows. Channel holding time distribution is assumed to be exponential with a mean of 20 seconds. We are assuming that the uplink channel is operating in the message trunking mode. Average call arrival rate per user is assumed to be 10calls/hour. Users in the ON state generate 230 bits/slot, and users in the OFF state do not generate any bits. Under these assumptions, we found that for 2 % GoS, 10^{-4} FLR and 2.88 maximum waiting approximately 20000 PMR users can be supported by the proposed system. Second major assumption set is as follows. We determined the capacity of the system by using

the channel holding time statistics of Channel 1 and 3. The mean connection holding time is 3.25 seconds for Channel 3 and 2.86 seconds for Channel 1 (Section 2.4.1). The system used for uplink is a conventional PMR network which operates under transmission trunking. Users holding a channel constantly generate 230bits/slot. Thus we have 240 available channels for the downlink. Under these assumptions, we found that for 2 % GoS, PMR users that can be supported by the proposed system is 315000 if Channel 1 statistics is considered, 114000 if Channel 3 statistics is considered. If waiting time is allowed, a considerable increase in the number of users can be achieved. In general the number of subscribers of a PMR system is low. For example in a big metropol like Istanbul there are about 8000 users of the police PMR system. We can conclude from these observations that only a small proportion of an existing DAB system's capacity needs to be allocated to the downlink of a PMR system. Therefore the proposed system offers new service opportunuties for existing DAB operators. A PMR system as proposed in this study also has the capability of supporting high data rate downlink applications such as multimedia services.

Appendix A

The Terrestrial Trunked Radio – TETRA System

The main focus of this appendix is the TETRA (TErrestrial Trunked RAdio) standard. TETRA was developed from scratch as a digital PMR standard within ETSI. TETRA provides lots of services from specialized safety services to cellular operating modes. It also has a wide selection of data services. The trunked TDMA access technique allows more efficient use of radio spectrum, but means that an operator must be assigned a minimum of at least four voice channels. However, the system has a number of operating modes, which allow wide area coverage with a single radio carrier without resorting to cellular frequency reuse schemes that would increase radio carrier demands still further. Of more potential concern is that the complexity of the system will make the infrastructure and terminals relatively expensive, which should be offset by the economies of scale if the system becomes popular [2].

To put it very briefly, TETRA is a wireless communication standard for PMR and Private Access Mobile Radio (PAMR) applications. PAMR systems, or PMR systems with a common standard and interworking arrangement, have the advantage of allowing users on different PMR systems to communicate with each other directly. That is, a PAMR network operator offers trunked system services to many different organizations. The operator administers a “trunk” of

channels which are made available individually for the period of a call (individual or group call).This way, the frequency resources are shared much more efficiently. The PMR user groups mentioned above could share their networks thus becoming PAMR users. Typical examples of PAMR users include maintenance fleets, courier and delivery services, as well as construction and taxi companies.

TETRA is a digital format, i.e. speech is transmitted as binary data, which makes it far more difficult to monitor or eavesdrop.

Table A.1 summarizes the main parameters of the TETRA system.

Parameter	Value
Carrier spacing	25kHz
Modulation	π/4-DQPSK
Carrier data rate	36 kbps
Voice coder rate	ACELP (4.56 kbps net, 7.2 kbps gross)
Access method	TDMA with 4 time slots per carrier
User data rate	7.2 kbps per time slot
Maximum data rate	28.8 kbps
Protected data rate	Up to 19.2 kbps

Table A.1 Main parameters of TETRA system [2]

The following sections of this appendix gives an overview of the TETRA system and serves as an introduction to the layering concepts upon which TETRA is based.

A.1 TETRA Frame Structure

Each cell of TETRA is allocated one or more pairs (uplink, downlink) of carriers. Each carrier provides four physical channels by use of TDMA, which divides the carrier into four slots of duration 14.167 msec. The TETRA TDMA frame has a period of 56.67 msec. This frame is repeated 18 times in order to produce one multiframe of duration 1.02 second. TETRA frame structure is depicted in Figure A.1. The multiframe is repeated 60 times in order to produce a hyper frame of duration 61.2 seconds which is related to encryption and synchronization [2].

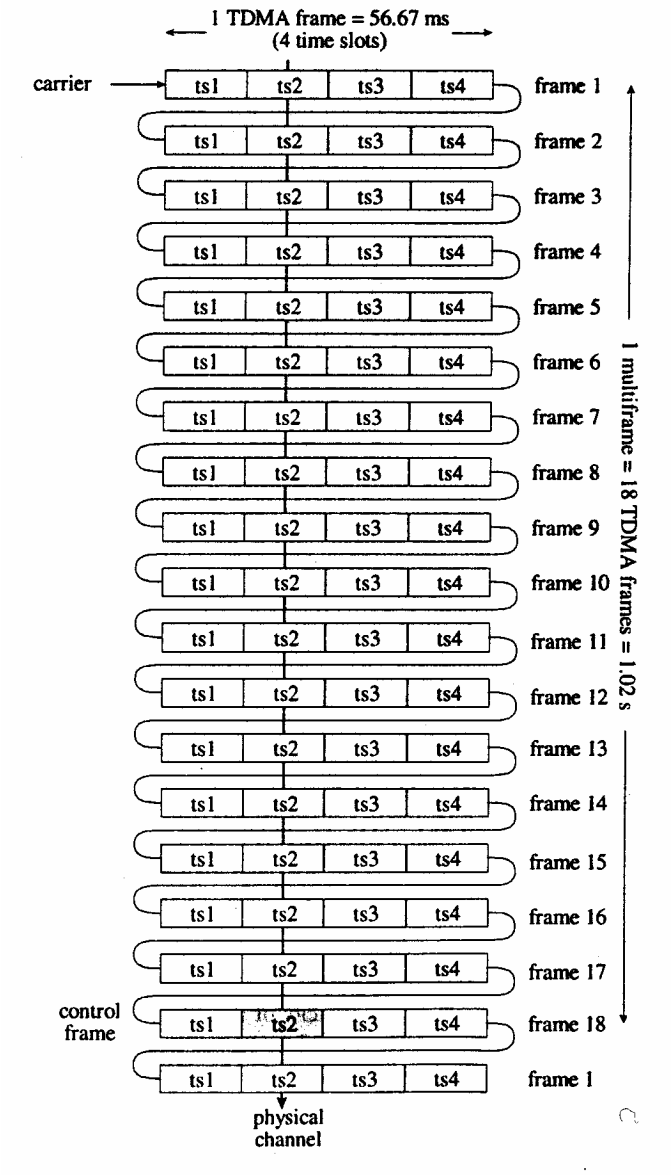


Figure A.1 TETRA frame structure [2]

Considering slot 2, Figure A.1 indicates that traffic may be transmitted in the first 17 occurrences of this slot and that the 18th occurrence is used for signaling purposes. Thus the Traffic Physical Channel may be regarded as 17 consecutive TDMA frames followed by a control frame [2].

The uplink and downlink transmission channels are offset in time by two slots to allow half duplex operation to be supported by low cost mobile

terminals which do not require a duplexer. In order to avoid interference, base station and mobile station use different frequencies with a fixed duplex offset, e.g. 10 MHz.

A.2 Logical Channels and Their Mapping on Physical Channels

Each carrier in TETRA system accommodates four time slots, which represent the physical channels available. These physical channels are shared between a number of logical channels, which carry both traffic and signaling information. The transmission mechanism for logical channels is provided by a physical channel (e.g. a specific slot on a carrier frequency) [2]. A particular physical channel may be used for several logical channels on a shared basis and therefore involves the concept of multiplexing. Physical channels are identified in TETRA as Control Physical Channels (CP) and Traffic Physical (TP) channels. The logical channels are mapped onto both Control Physical Channels and Traffic Physical Channels depending upon the mode of operation.

Within each slot data is transmitted in the form of bursts. Examples of these bursts are given in Figure A.2.

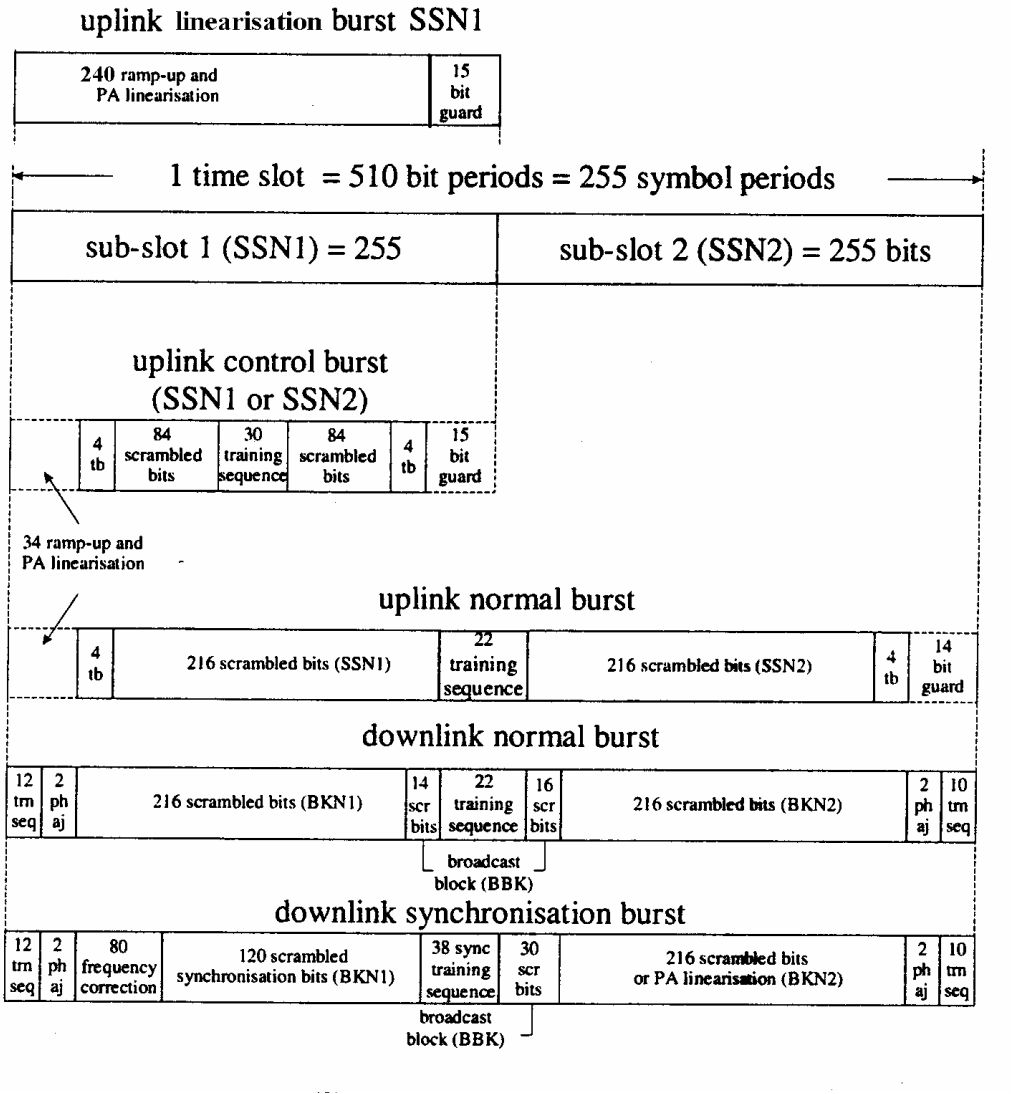


Figure A.2 TETRA burst types

In order to understand the way in which logical channels in TETRA are mapped onto the available physical channels, it is appropriate to consider the logical channels in more detail.

A.2.1 Logical Channels Hierarchy

There are seven logical channels defined in TETRA. These are Broadcast Control Channel (2), Common Control Channel (2), Associated Control Channel (3), Access Assignment Channel (1), Common Linearization Channel (1), Traffic Channels (4) and the Signaling Channel(3) .

The positions of the logical channels on physical channels are shown on Tables A.2 and A.3. While proceeding, please refer to Figure A.2 and these tables together to understand the mapping in more detail.

Frame Number	DOWNLINK		UPLINK	
	BKN1	BKN2	Subslot SSN1	Subslot SSN2
1 to 17	SCH/F SCH/HD SCH/HD SCH/HD BNCH		SCH/F SCH/HU SCH/HU CLCH SCH/HU	
18	BSCH	SCH/HD		

Table A.2 Mapping of logical channels on a CP channel

Frame Number	DOWNLINK		UPLINK	
	BKN1	BKN2	Subslot SSN1	Subslot SSN2
1 to 17	TCH STCH+TCH STCH+STCH		TCH STCH+TCH STCH+STCH	
18	SCH/F SCH/HD SCH/HD BSCH SCH/HD SCH/HD BNCH		SCH/F SCH/HU SCH/HU CLCH SCH/HU	

Table A.3 Mapping of logical channels on a TP channel

A.2.1.1 Common Control CHannel (CCCH)

The CCCH is a bi-directional channel for transmitting control information to or receiving control information from mobiles not actively

engaged in a circuit mode call. CCCH is generally divided in two categories, the Main Control CHannel (MCCH) and the Extended Control CHannel (ECCH). The CCCHs are always present. One carrier at each base station is designated the main carrier and slot1 on this carrier is used for MCCH. MCCH is carried on the control physical channel, which is shown in Table A.2. The MCCH is used among other things, to provide opportunities for mobiles to access the system. Mobiles which are not involved in a call listen to the downlink transmissions on the MCCH. For a system with a large number of carriers a single slot on a single carrier may not provide sufficient capacity. If this is the case, TETRA can use an extended mode where additional common control channels called Extended Control Channels are provided on other carriers, again in slot 1 [2].

A.2.1.2 Broadcast Control Channel (BCCH)

The BCCH is a uni-directional channel (downlink only) for common use by all mobiles [2]. BCCH is generally divided in two categories, the Broadcast Network CHannel (BNCH) and the Broadcast Synchronization CHannel (BSCH). The BSCH is carried on the downlink synchronization burst, which is shown in Figure A.2.The BNCH broadcasts network information to mobile stations. TETRA system channel acquisition is performed automatically when the mobile is powered up. The relevant channel is contained within the mobile station memory or a search is performed to find a channel. When a mobile is activated it must first acquire synchronization before it can decode any of the messages broadcast by the BS. This is done by synchronizing with the training sequence of the SYNC burst of any frequency in the cell. The SYNC burst appears in the BSCH, which is always transmitted in subslot BKN1 of frame 18 of the Traffic Physical channel. Once the synchronization is achieved the mobile decodes the rest of the SYNC burst, which identifies the cell, slot and frame, and the mode of operation. This gives the mobile full frame synchronization. The mobile then searches for the BNCH on the current frequency and decodes information on the frequency of the main carrier, power control information and

some random access parameters. Having decoded this information the mobile then locates the MCCH on slot1 of the main carrier. The mobile at this point has all the information needed to communicate with the system [2].

A.2.1.3 Associated Control CHannel (ACCH)

The ACCH is a uni-directional channel (downlink only) for common use by all mobiles. ACCH is generally divided in three categories, the Fast Associated Control CHannel (FACCH), the Slow Associated Control CHannel (SACCH) and the Stealing CHannel (STCH).

FACCH uses frames 1 to 17 when they are not used for traffic and the Slow Associated Control Channel always uses frame18. When a mobile is first assigned a physical channel, the channel is provided as the fast associated signaling channel for the control signaling which occurs at the start of a call (a typical example is the Base Station (BS) allocating transmit permission to a particular mobile). When this signaling phase is complete the assigned channel reverts to a Traffic CHannel (TCH). At the end of the call the assigned channel reverts to a FACCH to pass control signaling (for clearing down the call) to the system.

It is necessary for signaling information to pass between the MS (Mobile Station) and the BS during the progress of a call. There are two mechanisms provided for this. One of them is SACCH and the other is STCH. SACCH always uses frame 18. Capacity is stolen on the uplink or the downlink for urgent messages, such as handover signaling. In this case part of the assigned channel is allocated to the Stealing CHannel (STCH) and thus the TCH and STCH can normally exist at the same time [2].

A.2.1.4 Access Assignment CHannel (AACH)

This is a uni-directional channel (downlink only). Its purpose is to indicate access rights on control channels and to indicate the assignment of the uplink and downlink slots on each physical channel. AACH is carried in every BBK of downlink normal bust, which is shown in Figure A.2.

AACH carries the number of the mobile using the slot. This notification prevents crossed calls which might occur if the signal was lost for a short time due to shadowing and the slot used by the mobile undergoing shadowing was allocated to another mobile. Mobiles must monitor the AACH to ensure that they are still permitted to use the slot. The slot may have been reallocated due to pre-emption by a higher priority call, or if contact was lost for a time due to fading. If the slot is reallocated, the mobile cannot continue to use it and goes back to the CCCH to request a traffic channel again.

A.2.1.5 Common Linearization CHannel (CLCH)

This channel is transmitted on uplink linearization burst, which is shown in Figure A.2. Although the Linearization Channel is regarded as a logical channel no useful information is actually transmitted over the air. The CLCH may be regarded as a time interval during which mobile stations may transmit on air for the purpose of monitoring and linearising their own power amplifiers after switching frequency (e.g. at a channel assignment). It is necessary to maintain a linear characteristic in order to avoid degrading the advantages of pulse shaping. Please refer to Tables A.2 and A.3 and Figure A.2 in order to understand where this channel is carried on the physical channels [2].

A.2.1.6 Traffic CHannels (TCH)

These channels are bi-directional channels, which carry user voice or data. TCH is generally divided in four categories:

TCH/S: Carries digitized voice information produced by an ACELP coder at a net rate of 4.56 kbps. This rate is increased to 7.2 kbps by the addition of error protection bits. TCH/S is carried on BKN1 and BKN2 of the downlink normal burst and on the SSN1 and SSN2 of uplink normal burst which is shown in Figure A.2. As shown in the figure, BKN1, BKN2, SSN1 and SSN2 convey 216 bits of protected and digitized voice information produced within 30 msec. (216/30 msec. =7.2 kbps). Therefore one downlink slot and one uplink slot can carry 60 msec. of voice information of one user.

TCH/7.2: Carries unprotected user data at 7.2 kbps net rate.
TCH/4.8: Carries low protected user data at 4.8 kbps net rate.
TCH/7.2: Carries high-protected user data at 2.4 kbps net rate.

Higher net rates up to 28.8 kbps, 19.2 kbps or 9.6 kbps may be provided by allocating up to four physical channels to the same communication, which must use consecutive slots on the same frequency [2].

A.2.1.7 Signaling CHannel (SCH)

Signaling Channel is shared by all mobiles, but may carry messages specific to one mobile or one group of mobiles. System operation requires the establishment of at least one SCH per BS. SCH may be divided into three categories, depending upon the size of the message:

SCH/F: Uses full slot
SCH/HD: Uses only a half slot on downlink
SCH/HU: Uses only a half slot on uplink

SCH represents MCCH, FACCH, SACCH and BNCH [2].

A.3 Modes of Operation

There are three modes of operation defined in TETRA. They are normal mode, extended mode and minimum mode [2]. The details of these operation modes are described briefly in the following subsections.

A.3.1 Normal Mode

One pair of carriers per cell is designated to carry the main control channel (MCCH). This carrier is called the *main carrier*. In normal operation time slot 1 of every frame on main carrier (both uplink and downlink) is allocated for control purposes. This is known as Control Physical Channel (CP) and MCCH is carried on this channel. The remaining slots can be used for traffic. If this is the case, they represent the traffic physical Channels (TP) [2].

A.3.2 Extended Mode

A single MCCH will be satisfactory for most installations. However, with installations with large number of carriers, or with high levels of packet data, or a large number of short transactions (as may occur with short messages and transmission trunking), the access capacity available from a single slot MCCH may be insufficient. In such cases, additional control channels,

Secondary control channels (SCCH), can be provided, whereupon the system is said to be in extended mode. There are two types of Secondary Control Channels. They are introduced in the following subsections [2].

A.3.2.1 Common SCCH

Common SCCH have the same functionality as the MCCH, but apply to a subset of the mobile population, effectively reducing the loading on each channel. Like the MCCH, a Common SCCH is a single slot on the main carrier. One, two or three Common SCCHs can be defined, which occupy slots 2, 3 and 4 of the main carrier respectively. The maximum configuration therefore has common signaling on all four slots of the main carrier. The mobile population is shared between the MCCH and the common SCCHs, which then all act similarly [2].

A.3.2.2 Assigned SCCH

The base station may also operate Assigned SCCHs after the mobile has made an initial access on the MCCH (or appropriate common SCCH). This increases signaling capacity. Assigned SCCHs may be used for particular purposes and may be multi-slot (up to four), although the same number of slots must be used for uplink and downlink. This increased signaling capacity could be used to support a general packet data channel of a mobile or group of mobiles [2].

A.3.3 Minimum Mode

Minimum mode is where the MCCH is replaced by a traffic or assigned control channel. It is intended for use in low traffic density areas. Such areas will normally have a single carrier, and to reserve a slot in every frame would mean that more than 25% of the available capacity would be used for signaling.

Minimum mode operation allows a base station to allocate all time slots on the main carrier for traffic or dedicated control purpose. Therefore in minimum mode, only frame 18 would be available for common control, although stealing is still possible and can be used to contact waiting mobile stations [2].

A.4 Call Setup Procedure

Rather than dialing a number to set up a call, with the called party answering a phone, PMR systems usually have a pressel or PTT button to activate a call to the dispatcher or user group, with the receiving terminal annunciating the message without an answering procedure. Calls may therefore consist of a sentence or two, and users expect to be connected to the called terminal without delay. This is particularly important in the emergency services where the radio may be used to give urgent commands.

The call setup procedure is shown schematically in Figure A.3.The message sequence starts by the calling mobile M1 making a random access (known as a u-setup) in slot 1. This message is sent in a single burst and contains sufficient information to establish the service. The BS acknowledges with a message (known as d-call proceeding) and at the same time pages the called mobile M2 with a message (known as d-setup) in its next slot1. The

called mobile M2 responds with a u-connect message in the subsequent slot1. The BS then assigns the calling mobile a channel using the d-connect message and the called mobile using the d-connect ack message. In principle the assigned channel can be slot 2, 3 or 4 on the same carrier or any slot on another frequency duplex pair. It should be noted that BS gives immediate CLCH permission, on the assigned channel, by means of the CLCH permission flag. The BS then announces to the initiating mobile on the assigned channel that a connection has been established and grants transmit permission etc. (the assigned channel effectively exists as the FACCH at this stage). The initiating MS then proceeds to transfer coded voice frames [2].

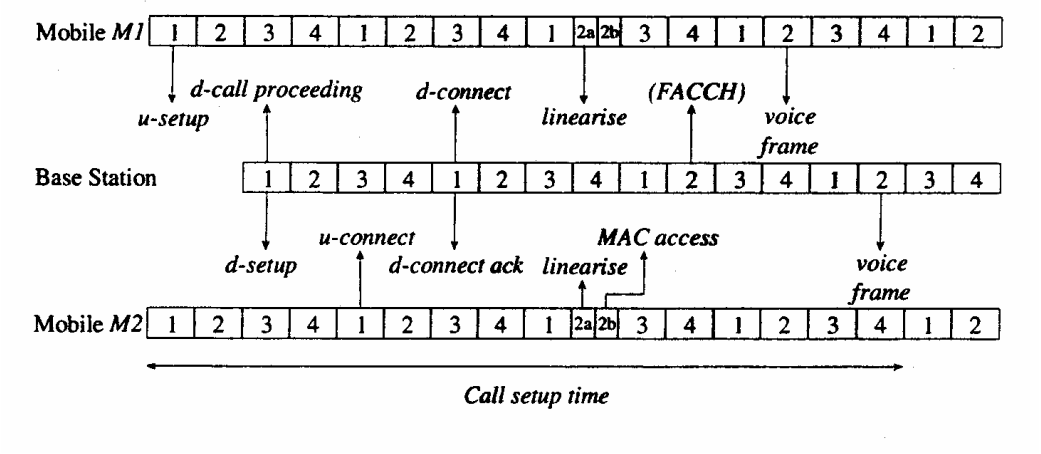


Figure A.3 Call set-up procedure

The call-setup time is defined as the time between the initial access request (u-setup) sent by mobile M1 and the first voice frame transferred to the called party. Under ideal conditions this is around 230ms depending on which channel has been allocated (e.g. 16 x 14167ms = 226.67 ms), but may be longer if several attempts are required due to propagation conditions. The full call set-up time for TETRA is specified as less than 300 ms [2].

A.5 Random Access

When a mobile wishes to make contact with the system this is done by making a random access, generally on a subslot on the MCCH, using the Slotted ALOHA mechanism.

Instability caused by heavy offered load is avoided in TETRA by allowing a mobile to attempt an access only if it receives an access code on the AACH. There are four different access groups designated A, B, C, D and a mobile will belong to one of these groups by a process called binding.

When a mobile does have permission to make a random access it randomizes its response within a specified number of access frames. This process is necessary to reduce the probability of collision [2].

A.6 Trunking Methods

In a trunked radio system such as TETRA, the radio channels are in a common pool and the TETRA system automatically allocates them to radio users at the start of each call – the user does not need to know what channel they are using. This automatic channel allocation from a common pool is called trunking and systems using this method are called trunked mobile radio systems.

The TETRA standard supports three different methods – message trunking, transmission trunking and quasi-transmission trunking [2]. It is up to the network operator or manufacturer to decide which trunking method to use, as different methods will provide solutions depending on whether it is desired to optimize call dropping, or some other characteristic.

A.6.1 Message Trunking

Message trunking is a traffic channel allocation strategy in which the same traffic channel is continuously allocated for the duration of a call. This may include several separate call transactions i.e. pressel activations by separate terminals. The traffic channel is only deallocated when the call is explicitly cleared by the call owner in the case of a group call, either party hanging up during an individual call or if an activity timer expires.

The main advantage of message trunking is that once a traffic channel has been allocated the users will not experience any delay at each new call transaction since there is no queuing for the allocation of channel resources. This absence of interruption ensures that a conversation can proceed without any interruption. Message trunking also minimizes processing and signaling overheads in the infrastructure, and fits well with telephone calls to external networks, which will almost always use message trunking. However message trunking makes inefficient use of the radio resource, as the channel is still allocated even between pressel activations when there is no speech to transmit. This means that there are fewer traffic channels available for other conversations, and initial call set-up times will be longer, particularly when the system is heavily loaded [2].

A.6.2 Transmission Trunking

With transmission trunking a traffic channel is allocated only for the duration of each individual transaction (i.e. for each activation of the pressel). PMR transmission trunking is usually based on pressel activations.

Transmission trunking makes efficient use of the available traffic channels since a traffic channel is only allocated when the users are actually speaking. The problem is that if the channel needs to be allocated on each pressel activation, there is a delay while the channel is set up. If the available traffic channels are heavily loaded, there may be a noticeable delay before the channel can be re-established, and this may prove disconcerting to users and result in a loss of the flow of conversation [2].

A.6.3 Quasi-Transmission Trunking

With quasi-transmission trunking a traffic channel is allocated for each call transaction but the channel de allocation is delayed for a short period (called the channel hang time) at the end of each transaction (i.e. after each pressel (PTT) release). If the hang time expires the traffic channel is released and the mobile returns to the common control channels for signaling. Quasi-transmission trunking therefore offers a compromise between message trunking and transmission trunking, offering some of the improved traffic throughput of the latter without as great risk of breaking up the flow of conversations [2].

A.7 Teleservices for Voice Transmission

The teleservices for voice transmission offer five different types of connection [21]:

Individual call: Point to point connection between calling and called subscribers.

Group call: Point to multipoint connection between calling subscriber and a group called through a common group number. The call is set up quickly

because no confirmation is required. The communication takes place in half-duplex mode through the activation of a *push-to-talk switch.*

Direct call (Direct Mode, DM): Point to point connection between two mobile devices with no use of infrastructure. A mobile station establishes a connection with another mobile station without the services of a base station, maintains the connection and takes over all the functions needed for local communication normally handled by the base station.

Acknowledged Group Call: Point to multipoint connection between a calling subscriber and a group called through a shared group number in which the presence of the group members is confirmed to the calling subscriber through acknowledgement. If one of the group members is not present or is on another call, the TETRA infrastructure informs the calling subscriber. If not enough members in the group can be reached, the caller can decide whether to discontinue or to maintain the call. An option is to have the group members who initially were not available to switch into the conversation later.

Broadcast Call: Point to multipoint connection in which the subscriber group dialed through a broadcast number can only hear the calling subscriber, i.e. simplex operation.

A.8 Independent Allocation of Slots on the Uplink and Downlink

Unlike traditional systems where uplink and downlink radio channels are paired to form a single logical channel, TETRA allows uplink and downlink channels on the same slot of the same frequency pair to be allocated for different purposes. If for example, a group call originates in one cell and has receiving mobiles in another cell, only a downlink is required in that second cell (see

Figure A.4). The corresponding uplink channel may be allocated to call which only requires an uplink channel in that cell, or for signalling [2].

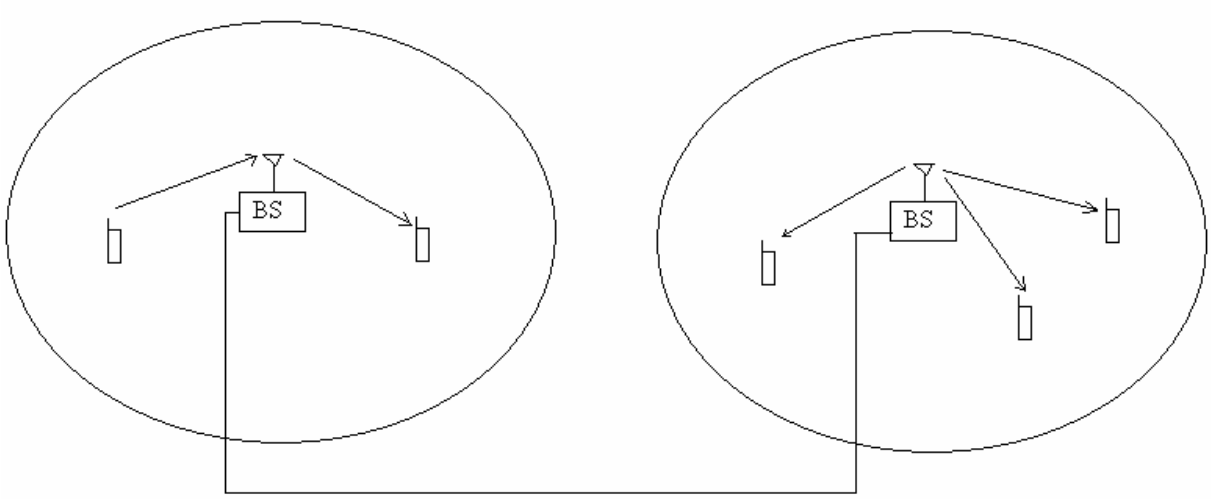


Figure A.4 Group call spanning two cells

The ACCESS-ASSIGN message in the AACH gives access rights to uplink and downlink slots. Permitted combinations on normal channels are as follows:

1. Uplink and downlink assigned to the same circuit mode call.
2. Uplink and downlink assigned to different circuit mode calls.
3. Uplink assigned to a circuit mode call, with the downlink assigned to a Secondary Common Control Channel (SCCH).
4. Uplink assigned to a SCCH, with the downlink assigned to a circuit mode call.

In addition, in minimum mode, it is possible to allocate only one direction of the MCCH to a circuit mode call, rather than both directions. The unused direction is available for normal MCCH use [2].

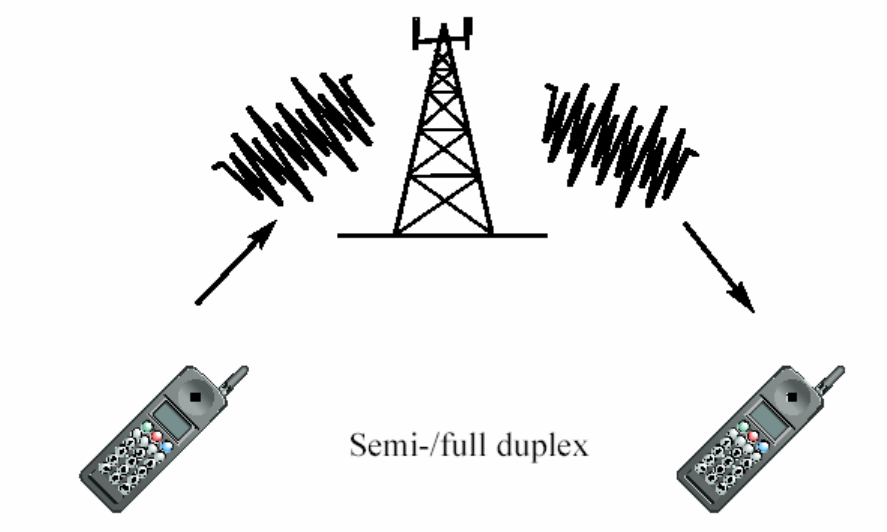
A.9 Trunking Capacity Estimation of TETRA system

TETRA system allow queuing of set-up calls and the Erlang C formula is applicable for trunking capacity estimation under the call queuing strategy [2,15]. Lets give an example: Assume that the total mean offered speech traffic is 12.5Erlangs/cell. To reach a call blocking probability of 5% (i.e. probability that delay>0 is 5%) at least 20 traffic channels (from Erlang C table) and one control channel are required. Therefore, in a TETRA system at least six carrier frequencies with four time slots, accordingly 150kHz spectrum per radio cell, are required.

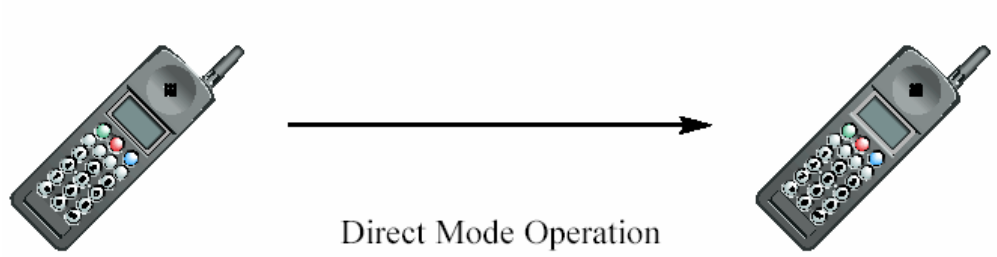
A.10 Simplex, Semi-Duplex, Full Duplex

In trunked mode operation, there is always a base station and network to switch information between one or several users. Any type of communication can be:

• Full duplex (both users can speak and listen at the same time)
• Semi-duplex, where only one party can speak at a time, usually a Press-To-Talk
button is used and the end of speech is announced by “Over”
• Simplex, which can be a broadcast message.



There is also a direct mode operation (DMO) being used when there is no base station available. In this case, only one party can speak at a time (simplex). Receive and transmit frequencies are identical (no duplex spacing) [2].



A.11 Quasi Synchronous Transmission

Quasi-synchronous transmission can be used to emulate wide area coverage from a single base station over a larger area than would be possible from a single transmitter alone. In quasi-synchronous transmission, transmitters on different sites are used to simultaneously transmit the signal at the same time. This is illustrated in Figure A.5.

With the Quasi- syncronous approach, while several base sites are used as in the cellular approach, different frequencies are not required since the same

carrier is transmitted on all of them. This is an advantage when the number of available frequencies is low, since a large area can be covered. Quasi-synchronous operation is therefore useful in areas where traffic density is low such as rural areas, but is wasteful of resources in areas of high user density [2].

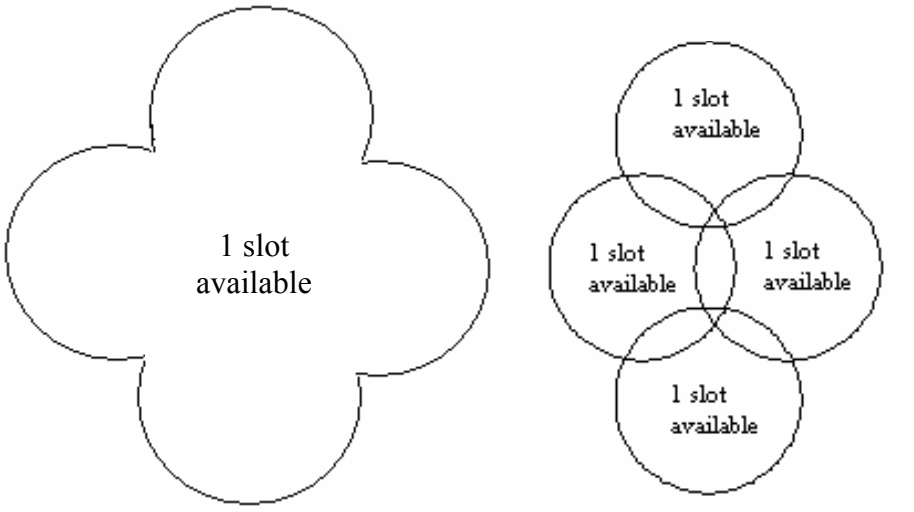


Figure A.5 Cellular and Quasi-synchronous approaches [2]

As an example lets suppose an area is covered with a regular arrangement of cells. In the first case, suppose up to 12 users may have to use the system simultaneously. Using quasi-synchronous operation, 3 carriers would be required (using minimum mode). Using a cellular arrangement, the number of carriers would depend on the frequency re-use pattern, but would be 21 carriers with a reuse of 7, or 36 carriers with a re-use of 12, since a minimum of three carriers would be required in every cell. Quasi-synchronous operation therefore requires fewer carriers. In the second case, suppose 12 users have to be supported in each cell at the same time. Quasi-synchronous operation on its own does not employ re-use. This means that 3 carriers would be required for each cell, or a total of (3 x #of cells) carriers. The cellular system, since it employs re-use, would only require 21 carriers for a reuse pattern of 7 or 36 carriers for a re-use of 12. The cellular system therefore requires fewer carriers than the quasi-synchronous

system if the number of cells in the system exceeds the number of cells in the cellular re-use pattern [2].

Quasi-synchronous operation has the advantage of simplifying system control, as the area can be considered to be one large cell, and so no handover is required. Mobiles can communicate as long as they can transmit to or receive from any of the base stations. The problem occurs in areas where coverage from several base sites overlap and the mobile can receive several signals. These signals need to be transmitted at exactly the same time to ensure that they do not interfere with each other, which implies complex synchronisation between base stations [2].

BIBLIOGRAPHY

[1] RA and FCS, "PMR Market Appendix," *Radio Communication Agency*, 2002.

[2] J.Dunlop, D. Girma and J. Irvine, *Digital Mobile Communications and the TETRA System*. Wiley, 1999.

[3] ETS 300 401, "Radio broadcasting systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers," *European Telecommunications Standards Institute (ETSI),* 1997.

[4] J. H. Stott, "The how and why of COFDM," *EBU Tech. Review*, 1998.

[5] Boudewijn R. Haverkort*, Performance of Computer Communication Systems: A Model-Based Approach.* John Wiley & Sons, 1998.

[6] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distributions via the EM algorithm," *Scandinavian Journal of Statistics*, pp.419– 441, 1996.

[7] O. Haggstrom, S. Asmussen, and O. Nerman, "EMPHT – a program for fitting phase-type distributions," *Department of Mathematics, Chalmers University of Technology, Goteborg,* 1992.

[8] Eva Ishay, "Fitting Phase-Type Distributions to Data from a Telephone Call Center," Master's thesis, Israel Institute of Technology, 2002.

[9] Donald Gross, Carl M. Harris, *Fundamentals of Queueing Theory*. John Wiley & Sons, 1985.

[10] Athanasios Papoulis**,** *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, Inc., 1991.

[11] Golub, G. H. and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1989, pp. 557-558.

[12] Report ITU-R M.2014, "Spectrum efficient digital land mobile systems for dispatch traffic," 1998, pp. 4-8.

[13] T. Lauterbach and M. Unbehaun, "Multimedia Environment for Mobiles (MEMO) – Interactive Multimedia Services to Portable and Mobile Terminals," *ACTS Mobile Commun. Summit, vol. 2, Oct 1997*, pp. 581-586

[14] SABINA (System for Asymmetric Broadband Internet Access) project page: http://www.teracom.se

[15] ETSI (RES 6.1), "Scenarios for Comparison of Technical Proposals for MDRTS," *European Telecommunications Standards Institute (ETSI),* France, June 1991.

[16] H. H. Lee and C. K. Un, "A study of on-off characteristics of conversational speech," *IEEE Trans. Commun*., vol. 34, June 1986, pp. 630-637.

[17] K. Samaras, C. Demetrescu, C. Luschi and R. Yan, "Capacity calculation of a packet switched voice cellular network," *IEEE Vehicular Technology Conf. Proc*., 2000, pp. 26-30.

[18] L. Kleinrock, *Queueing Systems*, *Volume 1*. John Wiley & Sons, 1975.

[19] T. S. Rappaport, *Wireless Communications; Principles and Practice*. Prentice Hall, 2nd. Edition, 2002.

[20] Khaled M. Elsayed, Harry G. Perros, "Analysis of a Statistical Multiplexer with Hterogenous Markovian On/Off Sources and Applications to Call Admission Control in ATM Networks," *Department of Computer Science and Center for Communications and Signal Processing North Carolina State University*, 1997.

[21] Bernhard H. Walke, *Mobile Radio Networks*. John Wiley & Sons, 1999.

[22] Jimmie L. Davis, William A. Massey, Ward Whitt, "Sensitivity to the Service Time Distribution in the Nonstationary Erlang Loss Model," *AT & T Bell Laboratories*, October 1992.