Multiple Access Protocols for Mixed Services Wireless Packet Communications

Thesis submitted to the University of Surrey for the degree of Doctor of Philosophy

By
Pouya Taaghol

Mobile Communications Research Group, Centre for Communication Systems Research, University of Surrey, Guildford, Surrey GU2 5XH, United Kingdom.

April, 1999
To my late father, mother, brothers, and friends
Acknowledgements

I am indebted to many people for their stimulating interactions and discussions throughout this study.

Firstly, I would like to sincerely thank Dr. Rahim Tafazolli and Prof. Barry G. Evans for providing me with the opportunity of doing a PhD in the Centre for Communications Systems Research (CCSR), University of Surrey. Again, many thanks to my supervisor Dr. Rahim Tafazolli for his constructive guidance, discussions, and help throughout this study.

I would also like to comment my family for their support and help; my late father (Morteza), mother (Keyhan) and brothers (Amir and Payam Taaghol).

Many thanks to members of the centre and the mobile group.

Last but not least, I would like to thank the support staff of the CCSR, namely, Moira Macmillan, Stephanie Evans, and Lakshmi Chennell.
Abstract

Packet communications have increasingly become popular since the launch of the first ever packet-switched network, the ARPANET. Evolution of such networks resulted in creation of the Internet, the global packet-switched network. Amongst the well-known advantages of packet-switching are service transparency, cost-effectiveness, and high system capacity. A natural pathway will incorporate tetherless mobility features to packet services. The key component to realisation of wireless packet communication is the multiple access technology (the subject of this thesis), capable of optimum sharing of scarce radio resources amongst a diverse set of packetised information sources.

Based on core features of packet multiple access protocols, a classification is presented by which most proposed protocols could be mapped into.

For mixed services, two equivalent offered loads, namely “access load” and “traffic load”, are derived. The former can be used for design, optimisation, and stabilisation of the access mechanism, whilst the latter is to evaluate the overall system performance. The proposed normalised offered loads are found accurate and instrumental in predicting and designing packet access protocols for mixed services.

“Statistical upper-bounds” for packet access protocols are derived (both for random access and general access mechanisms), based on which an “efficiency factor” is defined to evaluate and compare merits of protocols against each other.

Conventionally, analysis of resource-sharing systems with mixed services (such as multiple access protocols) involve multidimensional Markov chains which become quite intractable for more than two service types. In order to avoid multidimensional Markov analysis, an analytical method called “Aggregate Source Modelling” is proposed by means of which various service types are aggregated into a representative service type. The proposed method is deployed to analyse various queueing and multiple access scenarios.

Finally, a series of multiple access protocols are presented for adapting real-time and non-real-time packet services to existing cellular and satellite systems. For evaluation purposes, the proposed statistical models of packet services (such as FTP, E-mail, WWW, packet voice, and so forth) are used throughout the thesis. These models have components of various statistical distributions, namely, Poisson (Exponential), log-normal, Pareto, Cauchy, Gaussian, Geometric, and Hyper-Exponential. Extensive discrete-time simulation models have been developed for evaluation purposes.
# Table of Contents

ACKNOWLEDGEMENTS ..................................................................................................................... III

ABSTRACT ........................................................................................................................................ IV

TABLE OF CONTENTS ....................................................................................................................... V

LIST OF FIGURES ............................................................................................................................. X

LIST OF TABLES ............................................................................................................................... XIV

ACRONYMS ......................................................................................................................................... XV

LIST OF PUBLICATIONS .................................................................................................................. XXII

CHAPTER 1: INTRODUCTION ............................................................................................................. 1
  1.1 PROLOGUE ........................................................................................................................................ 1
  1.2 OBJECTIVES AND MOTIVATIONS ............................................................................................... 2
  1.3 STRUCTURE OF THESIS .............................................................................................................. 3
  1.4 ORIGINAL CONTRIBUTIONS ....................................................................................................... 6

CHAPTER 2: AN OVERVIEW OF PACKET-SWITCHING ..................................................................... 8
  2.1 INTRODUCTION ................................................................................................................................... 8
  2.2 CIRCUIT-SWITCHING CONCEPT ..................................................................................................... 8
  2.3 PACKET-SWITCHING CONCEPT ...................................................................................................... 10
    2.3.1 Multiplexing Advantages (Averaging) ...................................................................................... 11
    2.3.2 Cost-Effectiveness ................................................................................................................... 11
    2.3.3 Service Transparent and Multimedia Support ......................................................................... 11
    2.3.4 Future-Safe ............................................................................................................................. 11
  2.4 BURST-SWITCHING CONCEPT ...................................................................................................... 11
  2.5 PACKET-SWITCHED NETWORKS ................................................................................................... 12
    2.5.1 Concepts ....................................................................................................................................... 12
      2.5.1.1 Packetisation ......................................................................................................................... 12
      5.1.1.2 Addressing ........................................................................................................................... 13
      5.1.1.3 Routing ...................................................................................................................................... 13
    2.5.2 Existing Packet-Switched Networks .......................................................................................... 14
      2.5.2.1 The ARPANET and the Internet ............................................................................................ 14
      2.5.2.2 Asynchronous Transfer Mode (ATM) .................................................................................... 15
  2.6 CELLULAR PACKET SYSTEMS ...................................................................................................... 19
    2.6.1 Cellular Digital Packet Data (CDPD) ......................................................................................... 19
    2.6.2 General Packet Radio Services (GPRS) .................................................................................... 22
      2.6.2.1 Applications and Service Characteristics ............................................................................ 23
      2.6.2.2 Packet Routing ..................................................................................................................... 24
      2.6.2.3 Mobility Management .......................................................................................................... 25
      2.6.2.4 Physical and Logical Channels ............................................................................................. 26
      2.6.2.5 Mobile-Originated Transfer ................................................................................................. 28
      2.6.2.6 Mobile-terminated Transfer ................................................................................................. 29
      2.6.2.7 GPRS Mobile Terminal Model ........................................................................................... 29
      2.6.2.8 Data Encoding ....................................................................................................................... 30
  2.7 SATELLITE PACKET SYSTEMS ..................................................................................................... 30
  2.8 SUMMARY ........................................................................................................................................ 31
CHAPTER 3: PACKET MULTIMEDIA SERVICES

3.1 INTRODUCTION ................................................................. 32

3.2 SERVICES CLASSIFICATION ................................................... 32

3.2.1 Interactive Services ......................................................... 33

3.2.1.1 Conversational Services .............................................. 33

3.2.1.2 Messaging Services .................................................... 33

3.2.1.3 Retrieval Services ...................................................... 34

3.2.2 Distribution Services ....................................................... 34

3.2.2.1 Distribution Services Without User Individual Presentation Control .............................................................. 34

3.2.2.2 Distribution Services With User Individual Presentation Control ........................................................... 34

3.3 SERVICE DEFINITION PARAMETERS ...................................... 35

3.4 PACKET VOICE ................................................................. 36

3.4.1 PV Source Model with Slow Voice Activity Detection .......... 38

3.4.2 PV Source Model with Fast Voice Activity Detection .......... 39

3.4.3 Hangover ................................................................. 41

3.4.4 Packet Dropping .......................................................... 41

3.5 INTERNET PACKET SERVICES ............................................ 42

3.5.1 Electronic Mail (E-MAIL) .............................................. 42

3.5.2 File Transfer (FTP) ....................................................... 43

3.5.3 World-Wide-Web (WWW) ............................................. 44

3.5 OTHER PACKET DATA SERVICES .................................... 47

3.5.1 Railway Control Data Model ........................................ 47

3.5.2 Mobilex Data Model .................................................... 47

3.5.2 Poisson Arrival Packet Data Model ................................ 47

3.6 SUMMARY ................................................................. 48

CHAPTER 4: PACKET MULTIPLE ACCESS PROTOCOLS (THEORY) .... 50

4.1 INTRODUCTION ................................................................. 50

4.2 CLASSIFICATIONS OF MULTIPLE ACCESS PROTOCOLS ........ 51

4.3 PROPOSED RADIO PACKET MULTIPLE ACCESS PROTOCOLS .... 53

4.3.1 Reservation-ALOHA (R-ALOHA) .................................. 53

4.3.2 ALOHA Reservation (ALOHA-R) ................................... 53

4.3.3 Packet Reservation Multiple Access (PRMA) .................... 54

4.3.4 Integrated PRMA (IPRMA) ........................................... 55

4.3.5 Frame Reservation Multiple Access (FRMA) ..................... 56

4.3.6 Burst Reservation Multiple Access (BuRMA) ..................... 56

4.3.7 PRMA++ ................................................................. 56

4.3.8 Mitrou’s Protocol ......................................................... 57

4.3.9 Time Multiplexed-Base Controlled Multiple Access/Collision Detection (TM-BCMA/CD) ........ 57

4.3.10 Block Reservation Multiple Access (BRMA) .................... 58

4.4 RANDOM ACCESS MECHANISMS ....................................... 58

4.4.1 Performance Analysis .................................................... 59

4.4.2 Non-Prioritised Schemes ................................................ 61

4.4.3 Prioritised Schemes ..................................................... 62

4.4.4 Adaptive Schemes ....................................................... 65

4.4.5 Dynamic Retransmission Control Schemes ....................... 66

4.4.5.1 Optimum Variable Permission Probability .................... 66

4.4.5.2 Two-Level Variable Permission Probability Scheme ........ 67

4.4.5.3 Bayesian Broadcast Control Scheme ......................... 67

4.4.5.4 Exponential Back-off Scheme ..................................... 68

4.4.5.5 Linear Back-off Scheme ........................................... 68

4.4.5.6 µ-Law Back-off Scheme ........................................... 68

4.4.5.7 Step-Function Back-off Scheme ................................. 69

4.4.5.8 Stack Based Scheme .............................................. 69

4.4.5.9 Comparison of Retransmission Schemes ...................... 69

4.5 GENERAL ANALYSIS FOR MIXED SERVICES ...................... 70

4.5.1 Formulation ............................................................. 70

4.5.1.1 Protocol Parameters ............................................... 71

4.5.1.2 Performance Metrics ................................................ 71
6.2.1 Description ......................................................................................................................... 135
6.2.2 Performance Evaluations ....................................................................................................... 139
6.2.2.1 Effect of Offered Traffic Load .................................................................................. 139
6.2.2.2 Effect of Frame Size ..................................................................................................... 144
6.3 Burst Reservation Multiple Access Protocol ............................................................................ 148
6.3.1 Description and Modelling .................................................................................................... 149
6.3.2 BuRMA for GSM .................................................................................................................. 151
6.3.3 BuRMA for DECT ................................................................................................................ 152
6.3.4 Performance Evaluation ....................................................................................................... 153
6.3.4.1 Full-Rate GSM Case .............................................................................................. 153
6.3.4.2 Half-Rate GSM Case .............................................................................................. 154
6.3.4.3 DECT Case ................................................................................................................ 156
6.4 CONCLUSIONS ......................................................................................................................... 157

CHAPTER 7: SATELLITE PACKET COMMUNICATIONS .............................................................. 159

7.1 INTRODUCTION ....................................................................................................................... 159
7.2 NON-GEOSTATIONARY SATELLITE CONSTELLATIONS ......................................................... 160
7.3 SATELLITE GPRS ..................................................................................................................... 160
7.3.1 Proposed MAC Layer for S-GPRS ...................................................................................... 162
7.3.2 Performance Evaluations .................................................................................................... 167
7.4 Burst Reservation CDMA for Mixed Services (BR-CDMA) .................................................... 174
7.4.1 Hybrid TD-CDMA ............................................................................................................. 174
7.4.2 Burst Reservation CDMA Protocol ................................................................................... 176
7.4.3 Performance Evaluation ..................................................................................................... 179
7.4.3.1 Simulation Model ....................................................................................................... 179
7.4.3.2 Numerical Results ........................................................................................................ 179
7.5 CONCLUSIONS ........................................................................................................................ 183

CHAPTER 8: CONCLUSIONS AND FURTHER RESEARCH ...................................................... 185

8.1 SUMMARY AND CONCLUSIONS ............................................................................................ 185
8.2 FURTHER RESEARCH TOPICS .............................................................................................. 187

REFERENCES ............................................................................................................................ 189

APPENDIX A: STATISTICAL DISTRIBUTIONS ........................................................................ 206
A.1 Uniform Distribution............................................................................................................... 206
A.2 Negative Exponential Distribution ........................................................................................ 207
A.3 Hyper-Exponential ................................................................................................................ 207
A.4 Extreme Distribution ............................................................................................................ 208
A.5 Normal (Gaussian) Distribution ........................................................................................... 208
A.6 Pareto Distribution ................................................................................................................ 208
A.7 Pareto Distribution with Cut-Off .......................................................................................... 209
A.8 Log-Normal Distribution ...................................................................................................... 210
A.9 Log-Extreme Distribution .................................................................................................... 210
A.10 Geometric Distribution ....................................................................................................... 210
A.11 Cauchy Distribution ............................................................................................................ 211
A.12 Cauchy Distribution with Cut-Off ...................................................................................... 211

APPENDIX B: NON-INTERLEAVED GILBERT-ELLIOT STATISTICAL CHANNEL MODEL ... 212

APPENDIX C: FINITE POPULATION MONOMEDIA M/G/N//FCFS QUEUE .......................... 214
List of Figures

Figure 1.1 Structure and Chapter’s Relationships of the Thesis ........................................... 6
Figure 2.1 Circuit-Switched Terminal Model ................................................................. 9
Figure 2.2 Packet-Switched Terminal ............................................................................. 10
Figure 2.3 Burst-Switched Terminal Model ................................................................. 12
Figure 2.4 Fundamental Differences of Circuits, Datagrams, & Virtual Circuits ........ 17
Figure 2.5 Virtual Circuit Concept ................................................................................ 17
Figure 2.6 Standard ATM Cell ..................................................................................... 19
Figure 2.7 A Network Level View of the CDPD Network ............................................ 20
Figure 2.8 Development of GSM Radio Technology .................................................... 22
Figure 2.9 Architecture of GPRS Network .................................................................. 25
Figure 2.10 Mobility Management State Model .......................................................... 26
Figure 2.11 GPRS Logical Channel Structure ............................................................. 26
Figure 2.12 GSM Channel Organisation & Normal Burst Structure ........................... 28
Figure 2.13 Data Flow in GPRS ................................................................................... 30

Figure 3.1 ITU L.211 Classification of Services .......................................................... 33
Figure 3.2 Two-Way Communication of On-Off Sources ........................................... 36
Figure 3.3 Voice Activity Detection .............................................................................. 37
Figure 3.4 Schematic Model of a Typical Wireless Packet Voice Terminal ................. 38
Figure 3.5 Packet Stream Produced by a Packet Voice Source .................................... 38
Figure 3.6 Markov Model for Packet Voice Source with Slow VAD ......................... 39
Figure 3.7 Talkspurt-Silence Pattern Generated by Fast VAD .................................... 40
Figure 3.8 Markov Model for Packet Voice Source with Fast VAD ............................ 41
Figure 3.9 Traffic Pattern Generated by an FTP Session ............................................. 43
Figure 3.10 FTP Activity Variation with Respect to Channel Capacity ...................... 44
Figure 3.11 WWW Traffic Pattern Generated by a Browsing Session ....................... 45
Figure 3.12 WWW Activity Variation with Respect to Channel Capacity .................. 47
Figure 3.13 Packet Arrival Pattern .............................................................................. 48

Figure 4.1 Classification of Multiple Access Protocols ................................................. 51
Figure 4.2 General Channels’ Structure for Hybrid Multiple Access Protocols .......... 52
Figure 4.3 PRMA Uplink Frame Structure .................................................................. 55
Figure 4.4 PRMA++ Uplink Frame Structure .............................................................. 57
Figure 4.5 Contention Associated with Random Access Protocols ......................... 59
Figure 4.6 Random Access Protocol for Low Population, No-Priority Multimedia Systems ......................................................................................................................... 61
Figure 4.7 Random Access Protocol for High Population, No-Priority Multimedia Systems ......................................................................................................................... 62
Figure 4.8 Random Access Protocol Low Population, Prioritised Multimedia Systems .......................................................................................................................... 63
Figure 4.9 Random Access Protocol for High Population, Prioritised Multimedia Systems ......................................................................................................................... 64
Figure 4.10 Random Access Protocol for Higher Population, Prioritised Multimedia Systems ......................................................................................................................... 64
Figure 4.11 Adaptive Random Access Mechanism ...................................................... 65
Figure 4.12 State Transitions of an ON/OFF Source .................................................... 76
Figure 4.13 A Finite Population Non-Prioritised Multimedia Queueing System ....... 82
Figure 4.14 Aggregate Source Modelling Procedure .................................................... 83
FIGURE 5.36 Voice Sub-System Efficiency of Packet Multiple Access Protocols
Using Contention-Free Upper-Bound (See Table 5.1) ........................................... 124
FIGURE 5.37 Voice-Only Efficiency of Packet Multiple Access Protocols
Using Upper-Bound with Contention (See Table 5.1) ........................................... 124
FIGURE 5.38 Data Subsystem Efficiency for Mixed Services Using Contention-Free
Upper-Bound (See Table 5.2) ................................................................................. 125
FIGURE 5.39 Data Subsystem Efficiency for Mixed Services Using Upper-Bound with
Contention (See Table 5.2) ................................................................................. 125
FIGURE 5.40 Performance of FCFS Queue with Internet Traffic Inputs
A) Throughput and B) Average Access Delay ..................................................... 128
FIGURE 5.41 Service Distribution Independence of PRMA
A) Throughput and B) Average Access Delay ..................................................... 130
FIGURE 5.42 Accuracy Verification of ASM Method for Mixed Services RRAP
A) Throughput and B) Average Access Delay ..................................................... 131

FIGURE 6.1 Virtual Serving Queue Organisation (in this case, N=10 and N=5) .......... 135
FIGURE 6.2 Data Transfer Procedure in RRMA with Access Queue ......................... 137
FIGURE 6.3 Data Transfer in RRMA with no Access Queue ..................................... 138
FIGURE 6.4 Offered Load Versus No. Terminals .................................................... 140
FIGURE 6.5 Normalised Offered Load Versus No. Terminals ..................................... 140
FIGURE 6.6 Blocking Probability Versus Normalised Offered Traffic Load ............... 141
FIGURE 6.7 Average Frame Size Versus Normalised Offered Traffic Load ............... 141
FIGURE 6.8 Transmission Delay Versus Normalised Offered Traffic Load ............... 142
FIGURE 6.9 Average Access Delay Versus Normalised Offered Traffic Load ............ 142
FIGURE 6.10 Average Queueing Delay Versus Normalised Offered Traffic Load ..... 143
FIGURE 6.11 Average Queue Length Versus Normalised Offered Traffic Load ......... 143
FIGURE 6.12 Throughput Versus Normalised Offered Traffic Load ......................... 143
FIGURE 6.13 Normalised Throughput Versus Normalised Offered Traffic Load ....... 144
FIGURE 6.14 No. Contention Attempts Versus Maximum Frame Size .................... 145
FIGURE 6.15 Average Contention Delay Versus Maximum Frame Size .................... 145
FIGURE 6.16 Average Frame Size Versus Maximum Frame Size ............................ 146
FIGURE 6.17 Average Transmission Time Versus Maximum Frame Size ................. 146
FIGURE 6.18 Average Access Delay Versus Maximum Frame Size .......................... 147
FIGURE 6.19 Blocking Probability Versus Maximum Frame Size ............................ 147
FIGURE 6.20 Throughput Versus Maximum Frame Size ....................................... 148
FIGURE 6.21 Normalised Throughput Versus Maximum Frame Size .......................... 148
FIGURE 6.22 BURMA Terminal Model with Two-Level Variable Permission Prob. .... 149
FIGURE 6.23 Example of the Header Structure for GSM ........................................ 151
FIGURE 6.24 DECT Channel Organisation and Packet Structure .............................. 152
FIGURE 6.25 BURMA Performance for Full-Rate GSM ......................................... 154
FIGURE 6.26 Normalised Channel Throughput of BURMA for Full-Rate GSM ........ 154
FIGURE 6.27 BURMA Performance for Half-Rate GSM ......................................... 155
FIGURE 6.28 Channel Throughput of BURMA for Half-Rate GSM ........................ 155
FIGURE 6.29 BURMA Performance for DECT ..................................................... 156
FIGURE 6.30 Channel Throughput of BURMA for DECT ......................................... 156

FIGURE 7.1 First Generation Satellite Constellations .............................................. 160
FIGURE 7.2 Propagation Delay of the ICO Constellation Vs. Elevation Angle ............ 163
FIGURE 7.3 Propagation Delay of the Iridium Constellation Vs. Elevation Angle163
FIGURE 7.4 Maximum Differential Delay Vs. Spotbeam Number for ICO ............... 164
FIGURE 7.5 Maximum Differential Delay Vs. Spotbeam Number for Iridium ......... 164
FIGURE 7.6 Elevation Angle Variation for ICO Spotbeams .................................... 165
FIGURE 7.7 163-Spotbeam Pattern of ICO .......................................................... 165
FIGURE 7.8 Mobile-Initiated Data Transfer Procedure .......................................... 167
FIGURE 7.9 Offered Traffic Load Versus Total No. Terminals ................................. 169
List of Tables

TABLE 2.1  GPRS QoS Profiles .................................................................................................................. 24
TABLE 3.1  Default Mean Values for Distributions of a Typical WWW Service .... 45

TABLE 4.1  Definitions of Typical Protocol Parameters (S_p) ................................................................. 71
TABLE 4.2  Definitions of Typical Traffic Parameters of Service Class i (S_t) ............................... 72
TABLE 4.3  Definitions of Performance Metrics for Service Class i .................................................... 73

TABLE 5.1  Parameter Sets of Proposed Reservation Protocols (Voice Subsystem) 101
TABLE 5.2  Parameter Sets of Proposed Reservation Protocols (Data Subsystem) 101
TABLE 5.3  Traffic Parameters of Example 1 .......................................................................................... 126
TABLE 5.4  Measures of Effectiveness Comparison (Example 1) .......................................................... 127
TABLE 5.5  Traffic Parameters of Example 2 .......................................................................................... 127
TABLE 5.6  Measures of Effectiveness Comparison (Example 2) .......................................................... 127

TABLE 6.1  System Design Spreadsheet ................................................................................................. 139
TABLE 6.2  System Design Spreadsheet ................................................................................................. 153

TABLE 7.1  LEO66 and ICO10 System Parameters ............................................................................... 168
TABLE 7.2  System Design Spreadsheet ................................................................................................. 180
Acronyms

ABR: Available Bit Rate
ACK: Acknowledgement
ACTS: Advanced Communications Technologies and Services
ADPCM: Adaptive Differential Pulse Code Modulation
AMPS: Advanced Mobile Phone System
AQ: Access Queue
ARDIS: Advanced Radio Data Information Service
ARPA: Advanced Research Projects Agency (of the U.S.)
ARPANET: Advanced Research Projects Agency Network
ARQ: Automatic Repeat Query
ASM: Aggregate Source Modelling
ATDMA: Advanced Time Division Multiple Access
ATM: Asynchronous Transfer Mode
BCCH: Broadcast Control Channel
BCH: Broadcast Channel
BER: Bit Error Ratio
B-ISDN: Broadband - Integrated Services Digital Network
BR-CDMA: Burst Reservation - Code Division Multiple Access
BuRMA: Burst Reservation Multiple Access
CAC: Call Admission Control
CAMEL: Customised Application Mobility Enhanced Logic
CBR: Constant Bit Rate
CC: Call Control
CCH: Common Control Channel
CCI: Constellation Communications Inc.
CCITT: International Consultative Committee for Telecommunications and Telegraphy
CCS: Control Channel
CDF: Cumulative Distribution Function
CDMA: Code Division Multiple Access
CDPD: Cellular Digital Packet Data
CF-PRMA: Contention Free-Packet Reservation Multiple Access
CH: Call Handling
CLNP: Connectionless Network Protocol
CONP: Connection Oriented Network Protocol
CRC: Cyclic Redundancy Check
CS: Coding Scheme
D-AMPS: Digital-Advanced Mobile Phone System
DARPA: Defence Advanced Research Projects Agency (of the U.S.)
DCOM: Distributed Component Object Model
DCS: Digital Communications System
DECT: Digital Enhanced (European) Cordless Telephone
DNS: Domain Name System
DPE: Distributed Processing Environment
DSI: Digital Speech Interpolation
DSMA/CD: Digital Sense Multiple Access / Collision Detection
DSMA: Digital Sense Multiple Access
DSP: Digital Signal Processing
D-TDMA: Dynamic-Time Division Multiple Access
DTX: Discontinuous Transmission
E-mail: Electronic-mail
EPA: Equilibrium Point Analysis
ETCS: European Train Control System
E-TDMA: Enhanced-Time Division Multiple Access
ETSI: European Telecommunications Standards Institute
FBR: Fixed Bit Rate
FCCH: Frequency Correction Channel
FCH: Frequency Correction Channel
FDD: Frequency Division Duplexing
FDMA: Frequency Division Multiple Access
FEC: Forward Error Correction
F-ES: Fixed-End System
FH: Frequency Hopping
FIFO: First-In-First-Out
FMA: FRAMES Multiple Access
FPLMTS: Future Public Land Mobile Telecommunications System
FRAMES: Future Radio wideband Multiple Access
FSFC: First Come First Served
FTP: File Transfer Protocol
GEO: Geostationary Orbit
GFR: Guaranteed Frame Rate
GGSN: Gateway GPRS Support Node
GP: Guard Period
GPRS: General Packet Radio Service
GPS: Global Positioning Systems
GR: GPRS Register
GRAN: Generic Radio Access Network
GSM: Global System for Mobile communications
GSN: GPRS Support Node
GUT: GPRS User Terminal
HLR: Home Location Register
HSCSD: High Speed Circuit Switched Data
ICO: Intermediate Circular Orbit
IETF: Internet Engineering Task Force
IMSI: International Mobile Subscriber Identity
IMT-2000: International Mobile Telecommunications after the year 2000
IN: Intelligent Network
IP: Internet Protocol
IPv4: Internet Protocol version 4
IPv6: Internet Protocol version 6
IS: Interim Standard
ISDN: Integrated Service Digital Network
IT: Information Technology
ITU: International Telecommunications Union
LA: Location Area
LAN: Local Area Network
LEO: Low Earth Orbit
LES: Land Earth Station
LLC: Logical Link Control
MAC: Medium Access Control
MAI: Multiple Access Interference
MAN: Metropolitan Area Network
MDBS: Mobile Data Base Station
MD-IS: Mobile Data Intermediate System
MEO: Medium Earth Orbit
M-ES: Mobile-End System
MHF: Mobile Home Function
MM: Mobility Management
MPEG: Moving Picture Experts Group
MRCS: Multi-Rate Circuit-Switching
MSC: Mobile Switching Centre
MSF: Mobile Serving Function
MSS: Mobile Satellite System
MT: Mobile Terminal
NACK: Negative Acknowledgement
NDP: Number of Dropped Packets (from a front-end)
NRT: Non-Real-Time
OBP: On-Board Processing
ODE: Open Distributed Environment
ODMA: Opportunity Driven Multiple Access
ODP: Open Distributed Platform
OMA: Open Management Architecture
OO: Object-Oriented (programming)
OS: Operating System
OSI: Open Systems Interconnections
PACCH: Packet Associated Control Channel
PAGCH: Packet Access Grant Channel
PBCCH: Packet Broadcast Control Channel
PCCCH: Packet Common Control Channel
PCM: Pulse Code Modulation
PCN: Personal Communications Network
PCS: Personal Communication System
PDBCH: Packet Data Broadcast Channel
PDCH: Packet Data Channel
PDF: Probability Density Function
PDN: Public Data Network
PDP: Packet Data Protocol
PDU: Protocol Data Unit
PER: Packet Error Ratio
PLMN: Public Land Mobile Network
PNCH: Packet Notification Channel
PP: Protocol Parameters
PPCH: Packet Paging Channel
PPN: Public Packet Network
PPRCH: Packet Paging Response Channel
PPSN: Public Packet Switched Network
PRACH: Packet Random Access Channel
PRMA: Packet Reservation Multiple Access Protocol
PRNET: Packet Radio Network
PSPDN: Packet Switched Public Data Network
PSTN: Public Switched Telephone Network
PTCH: Packet Traffic Channel
PTM: Point-To-Multipoint
PTM-M: Point-To-Point - Multicast
PTP: Point-To-Point
PTP-G: Point-To-Point Group Call
PTP-M: Point-To-Point Multicast
PV: Packet Voice
QoS: Quality of Service
RA: Routing Area
RAC: Random Access Channel
RACE: Research / development in Advanced Communications technologies in Europe
RACH: Random Access Channel
R-ALOHA: Reservation - ALOHA
RAQ: Random Access Queue
RAR: Random Access Request
RASM: Representative Aggregate Source Model
RF: Radio Frequency
RFC: Request For Comments
RFP: Request For Proposal
Rlogin: Remote Login
RM: Resource Management
RR: Round Robin
RRA: Random Reservation Access
RRMA: Round Robin Multiple Access
RSVP: Resource Reservation Protocol for IPv4 and IPv6
RT: Real-Time
RTTI: Road Traffic and Transport Informatics
RX: Reception
S-ALOHA: Slotted - ALOHA
S-ARQ: Selective Automatic Repeat Query
SCH: Synchronisation Channel
S-GPRS: Satellite-General Packet Radio Service
SGSN: Serving GPRS Support Node
S-GUT: Satellite GPRS User Terminal
SMS: Short Message Service
SMTP: Simple Mail Transfer Protocol
S-PCN: Satellite-Personal Communication Network
S-PCS: Satellite-Personal Communication System
SQ: Serving Queue
SRI: Stanford Research Institute
STM: Synchronous Transfer Mode
TASI: Time Assignment by Speech Interpolation
TCH: Traffic Channel
TCP: Transfer Control Mechanism
TDD: Time Division Dulexing
TDMA: Time Division Multiple Access
TETRA: Trans European Tranked RAdio
TFI: Temporary Frame Identity
TINA-C: Telecommunication Information Network Architecture Consortium
TIPHONE: Tele-Internet PHONE
TLLI: Temporary Logical Link Identity
TSP: Time Slot Period
TX: Transmission
UBR: Unspecified Bit Rate
UCLA: University of California in Los Angles
UIC: Union Inernationales des Chemins de Fer
UMTS: Universal Mobile Telecommunications System
USF: Uplink Status Flag
VBR: Variable Bit Rate
VC: Virtual Circuit
VLR: Visiting Location Area
VLSI: Very Large Scale Integration
WAN: Wide Area Network
WATM: Wireless Asynchronous Transfer Mode
W-CDMA: Wideband Code Division Multiple Access
WLAN: Wireless Local Area Network
WWW: World Wide Web
List of Publications


CHAPTER

Introduction

1.1 Prologue

Digitisation techniques have allowed the expression of various information contents in a uniform binary format. Researchers developed various digital encoders for information sources such as speech and video. Growth in computer-related applications further boosted and justified the movement towards a digitised world.

Likewise, packetisation in the communications world, has had a similar effect to digitisation in information technology; packetised services looked transparent to packet-switched networks. Soon after the implementation of the first ever packet-switched network, ARPANET, numbers of (other but similar) networks came into existence. In order to marry such networks (that used different communications protocols), a novel higher layer was proposed which came to be known as TCP/IP (Transfer Control Protocol/Internet Protocol). The resultant virtual network was named "The Internet", the network of networks. The Internet has become the communication means of choice and its growth has been phenomenal in commercial, scientific, and personal sectors. Communication services offered now over the Internet range from non-real-time file transfers to real-time services such as voice and video (both point-to-point and broadcast). Although packet-switched networks have demonstrated reliability and cost-efficiency for non-real-time services, they do not yet satisfy the required quality of service of real-time users. The undergoing research and development on resource reservation protocols (RSVP), fast packet routing and the like will resolve the packet transfer delay experienced in packet-switched networks.
The market phenomenon created by services of the Internet (and corporate Intranets) such as World Wide Web (WWW), Electronic-mail (E-mail), File Transfer Protocol (FTP), and other IP-based applications, has further fuelled the demand for data access on the move; a natural evolution of packet multimedia communications is to expand to the wireless, i.e. indoor, cellular and satellite systems. The application of packet switching to cellular mobile communications has become a reality due to an ever increasing demand for flexible and efficient multimedia services. There are currently a number of cellular packet-switched networks operational, such as Advanced Radio Data Information Service (ARDIS), and Cellular Digital Packet Data (CDPD). Similarly, General Packet Radio Service (GPRS) is also expected to be shortly introduced for GSM networks.

The key component to realisation of efficient wireless packet communications is the multiple access technique. Amongst the design parameters, not only does the choice of multiple access technique greatly influence the spectrum efficiency, it also dictates the total system design. The design of the medium access control (MAC) sub-layer in wireless packet communications is different from that in the wired network. The differences are rooted in the radio channel (unpredictability, spectrum expenses, propagation delays, and differential delays). A great number of packet multiple access protocols have been proposed to fulfil air-interface requirements of future bandwidth-capacity hungry wireless packet communication systems [GOOD89, WONG92, BAIE96, WU94, DEVI93, NARA96, MITR93, WONG91, STER95, DUNL93, TAN96, BRAN96, PRAS95, PICH96, STER95, HÄMÅ95a, HÄMÅ95b, TAAG96b, TAAG97b, TAAG98a, TAAG98b].

1.2 Objectives and Motivations

Design, evaluation, comparison, and analysis of the optimum wireless protocol for mixed packet traffic and operational environment (i.e. indoor, cellular, and satellite) are the ultimate objectives in this study.

Despite the recent phenomenal growth in the number of proposed multiple access protocols for the future generation packet multimedia communication systems, general methods for analysis and evaluation of such protocols are lacking in the literature. Almost every proposed multiple access protocol is claimed to perform better than
1 Introduction

others. Although one protocol may demonstrate better performance than others for a given system configuration, it does not indicate that it necessarily outperforms all other protocols for all possible system configurations. Therefore, a mathematical tool is required to ease the fair comparison of protocols. This tool could be based on the statistical upper-bound of such protocols, imposing an upper limitation on the systems capacity. With help of the upper-bound analysis, performance comparison of packet-based multiple access protocols will be possible which has not been carried out in the literature so far. Performance comparisons will lead to obtaining the efficient operational region for each protocol and optimum design.

Most of evaluations carried out in the literature assume Poisson (or equivalently exponential) distributions as inputs to protocols, which simplifies the system analysis. However, recent observations show that packet inter-arrival times and message lengths are far from exponential distributions and demonstrate some form of long-range-dependent behaviour. For example, it has been shown that the file sizes (e.g. in FTP) and IP packet sizes (e.g. in WWW) are best approximated by a Pareto distribution, and interarrival times could follow log-normal distributions (as in FTP). Assumption of Poisson arrival could result in incorrect design of the MAC and other system layers.

Packet multiple access protocols mainly rely on the feedback acknowledgements from the base (whether terrestrial or satellite). As the propagation delay associated with medium increases (specially in satellite systems), the conventional access mechanisms fail to perform efficiently. Therefore, new sets of access mechanisms are required to rely less on the feedback acknowledgement in order to cope with long round-trip delays.

1.3 Structure of Thesis

The thesis is divided into eight chapters. Chapter 2 introduces the history, concepts, and applications of packet-switching technology followed by introductions on the well-known existing packet-based networks (such as the ARPANET, ATM, and the Internet). In that chapter, the evolution of the packet-switching technology from the wired to the wireless networks is presented. The wireless applications range from cellular, such as CDPD (Cellular Digital Packet Date) and GPRS (General Packet Radio Service), to satellite systems.
In Chapter 3, packet-based service components, such as voice, Internet applications (FTP, WWW, and E-mail), and other packet data services (either provided in the ETSI GPRS evaluation guidelines or the literature) are presented with their corresponding packet-based statistical models. Statistical characteristics and models of services to be supported are essential to designing and evaluating MAC protocols. The statistical distributions used in this chapter are fully defined in Appendix A. These stochastic service models will be used in all the consecutive chapters.

In Chapter 4, the theoretical background on packet multiple access protocols is established. Packet-based multiple access protocols are classified according to their core features. The well-known packet access protocols are presented. A normalised offered load for mixed services is derived by using which comparison, evaluation, study, and design of multiple access protocols become tractable. Various random access mechanisms are proposed for different traffic configuration. Most famous retransmission control schemes are presented followed by a novel retransmission scheme based on two-level variable permission probabilities. The statistical upper-bound (both with and without contention) of multiplexing protocols are derived based on which a mathematical tool has been proposed to ease the fair comparison between different protocols [TAAG95, TAAG97a]. This mathematical tool has been analysed using the finite source $M/G/N//FCFS$ queueing analysis detailed in Appendix B. It then proceeds to present a novel analytical method called "Aggregate Source Model" (ASM) which can be deployed in analysing the multi-class resource-sharing (queueing or multiple access) systems [TAAG98c, TAAG98f, TAAG98g]. Impacts of physical layer on the protocol performance and design are also discussed [TAAG98d].

In Chapter 5, the theories developed in Chapter 4 are extended further and verified. Accuracy of the offer load estimation as well as the upper-bound analysis are verified by simulation for mixed services. Performances of a broad range of packet-based protocols (15 well-known MAC protocols) are thoroughly examined and compared with each other. Optimisation through the proposed dynamic retransmission scheme is evaluated [TAAG96a]. Furthermore, mixed services analysis of packet-based protocols, both in the uplink and the downlink, is carried out by means of the ASM method introduced in Chapter 4.
Chapter 6 presents packet techniques to enhance the capacity of cellular (similarly indoor) systems for non-real-time (data) and real-time packetised services (voice and data). To realise cellular packet services, two multiple access protocols called "Burst Reservation Multiple Access" (BuRMA) [TAAG96b] and "Round-Robin Multiple Access" (RRMA) [TAAG97b] for real-time packet communication and General Packet Radio Service (GPRS), respectively, by dedicating a portion of the channel resources to packet techniques for cellular packet communications. Evaluations presented in this Chapter are carried out using a mixture of packet-voice, Internet-applications, and ETSI defined GPRS services.

In Chapter 7, two MAC protocols for real-time and non-real-time packet communications in mediums with long propagation latency (such as satellite systems) have been proposed. These protocols rely less on the long-delayed feedback. For the non-real-time case, similar to the terrestrial GPRS, an equivalent systems for satellite, named "Satellite-GPRS" (S-GPRS), is studied [TAAG98a, TAAG98c]. For the real-time case, a hybrid-switched protocol, called "Burst Reservation-Code Division Multiple Access" (BR-CDMA) [TAAG98b], has been proposed. The concept of "temporary reservation" is used in order to realise prompt access for the delay sensitive services. Again, in this chapter, evaluations are carried out for mixed services scenarios presented in Chapter 3.

Finally, Chapter 8 concludes this thesis by highlighting, summarising, and proposing further research directions and topics.

The chapters explained above are followed by four appendixes (Appendix A-D). Throughout the thesis, various statistical distributions are used which are listed and briefly described in Appendix A. And finally, the list of references used in this study ends this thesis.

The sequential and non-sequential relationships between the chapters are illustrated in Figure 1.1. The thickness of a border line shows the amount of original contributions in that chapter.
1.4 Original Contributions

The novel achievements can be outlined as follows:

- Derivation and confirmation of normalised traffic and access load for mixed services,
- Derivation of statistical upper-bound for random reservation protocols,
- Development of a novel analytical technique for performance and efficiency
• Simultaneous performance comparison of 15 different packet-based multiple access protocols,

• A method of designing random access sub-systems for mixed services

• An analytical method called "Aggregate Source Modelling" (ASM) by which analysis of mixed services systems can be confined to a simple monomedia analysis,

• Analysis of a various queueing and multiple access systems for mixed Internet-type and other packet services using ASM,

• A dynamic random access mechanism using two-level permission probabilities,

• A novel MAC protocol called "Burst Reservation Multiple Access" (BuRMA) for real-time cellular packet communications,

• A new MAC protocol called "Round-Robin Multiple Access" (RRMA) for General Packet Radio Service (GPRS) systems,

• Evaluation of RRMA for a mixture of Internet-type and other packet-based services,

• An original MAC protocol for satellite data packet communications "Satellite GPRS" (S-GPRS),

• A novel hybrid MAC protocol called "Burst Reservation-Code Division Multiple Access" (BR-CDMA) for satellite real-time packet communications.
An Overview of Packet-Switching

2.1 Introduction
In this chapter, the required introductory background together with the relevant packet communication concepts are initially established. Three different switching schemes known as circuit-, packet-, and burst-switching are hence reviewed. Each of these switching schemes forms a bearer service which is suitable for supporting specific service types. It is also argued that the above mentioned switching schemes may cause different implications in wired and wireless networks. The main advantages and benefits brought about through deployment of such systems are highlighted. This is then followed by a brief review of packet communication evolution from the early days of the ARPANET (the first ever packet-switching network) to mobile cellular packet systems, such as CDPD (Cellular Digital Packet Data) and GPRS (General Packet Radio Service) systems. These packet-based networks could act as stand-alone or complementary to that of the circuit-switched. Furthermore, the network of the networks, the Internet phenomenon, has been abstractly described and its impact on this evolution is identified.

2.2 Circuit-Switching Concept
The “circuit switching” transfer mode was introduced at the end of 19th century to be deployed for telephony. In this transfer mode, a “circuit” is established without interruption for the entire duration of the call. This type of switching has been employed in the PSTN (Public Switched Telephone Network). The communication path remains
allocated to a number of users and, during dialogue time, no other potential user can use that path, in any way; even when the link is under-utilised or not utilised at all.

With circuit switching, a connection has well-defined requirements in terms of network resources. If the necessary resources are available, then request for a new connection will succeed, otherwise the attempt will fail. However, once a connection has been established, the user has sole use of network resources dedicated to the connection. End-to-end delay is usually fixed and is caused by transmission delays and delays at the switching centres.

The circuit-switching concept has also been extensively deployed in wireless communication systems such as GSM and DECT. In such systems, a terminal could be in three different modes of OFF, IDLE, and TRANSMISSION (see Figure 2.1).

When the wireless terminal is turned on, it will attach itself to the network (IDLE). Upon reception or origination of a circuit call, the call set-up procedure is invoked resulting in allocation of dedicated traffic and control channels (TRANSMISSION). The dedicated channels will be reserved for the whole duration of the call after which the terminal moves to the IDLE state again. Depending on the required data rate, the terminal could be allocated multiple traffic channels. This concept is been known as Multi-Rate Circuit-Switching (MRCS) [PRYC95].

Circuit-switched connections offers simple however rigid resource management. Circuit-switching restricts services to fixed bandwidths regardless of their instantaneous bit-rate requirements, resulting in inefficient usage of scarce radio resources. This will waste the channel resources, as they are retained even during their inactive periods.
2.3 Packet-Switching Concept

In the 19th century, the “packet-switching” concept was introduced in telegraphy as the first “transfer mode” in the telecommunications world. A “packet” (i.e. the telegraph message in this case) was transported from one relay station to another. This packet consisted of the source and destination addresses along with the actual message.

In packet-switching, a message (higher layer PDU) is split into smaller segments and each segment (or packet) is sent one at a time. Each packet is stored and forwarded from a router to the next. A packet includes not only information bits, but additional bits, referred to as overhead, may have to be added. These bits can contain the source or destination address, the user ID for billing purposes, synchronisation bits, start and end delimiters or other information. There are two distinct types of packet services; connection-oriented and connectionless. In connectionless mode, there is no need for setting up a connection between the communicating parties. Each packet is treated as a self-contained entity, which is routed through the switches (routers) to the final destination, according to a routing algorithm. Different packets of a message may follow different routes and therefore arrive to the destination in different order. In connection oriented services, a ‘connection’ is set up between communicating parts. This is not a continuous physical connection as in circuit switching, it is only a pre-determined route that all packets will follow during the session. This kind of connection-oriented packet switched connection is known as ‘virtual circuit’ to distinguish from circuit switched systems’ connection. As an example, the Asynchronous Transfer Mode (ATM) technology deploys the packet-switching concept.

A packet-switched terminal could be in either of the four states (see Figure 2.2) of OFF, IDLE (on but not connected to the network), STANDBY (connected but not active), ACCESS/TRANSMISSION.

![Figure 2.2 Packet-Switched Terminal](image)
The major characteristics of packet-switching networks can be highlighted as follows: (1) statistical multiplexing, (2) cost effectiveness, (3) service transparent, and (4) future-safe.

2.3.1 Multiplexing Advantages (Averaging)
In packet-switching techniques, resources are only allocated and tailored according to a particular service in active periods. Therefore, packet communications take advantage of the statistical variance in the communication requirements resulting in high bandwidth efficiency.

2.3.2 Cost-Effectiveness
Bursty applications, such as WWW-surfing and FTP-connection (see Chapter 3), can have a connection (session) open for a long time while the connection is active only for small portion of the connection duration. From the user's prospective, charging is one of the main driving forces. In packet switching, billing could be based on the amount of transferred data (byte-charging) as opposed to the second-charging of circuit-switching.

2.3.3 Service Transparent and Multimedia Support
Packet communication is service independent as packets from different sources are generated and emitted in the same manner, which is essential for mixed services communications. Despite variable delay in packet communications, it is been shown that they can still be applied to time-sensitive applications such as speech [WEIN83, GOOD91].

2.3.4 Future-Safe
It can also cope with future technological advances such as low bit rate encoders, mixed voice/data terminals, and new advanced services (e-mail, paging, telex, image, video, video-phone, and so forth) with varying bit rate requirements [DUNL88, GOOD90, PRAS95a, PRAS95b].

2.4 Burst-Switching Concept
Burst (or message, packet stream, flow) switching, as opposed to circuit switching, does not require a dedicated end-to-end connection before transmission is commenced.
Instead, in a communication from A to B, the transmitter A will send a message to the switching centre X where it will be stored in a buffer. The message will only be forwarded to switching centre Y when the link from X to Y is not utilised by another connection. It will be buffered again at Y, until the line from Y to B is free.

This switching method is not very practical in wired networks due to buffering requirements. However, in wireless packet communications, burst-switching is quite necessary for bursty services. For example, if a wireless terminal has a stream of packets to send in a packet switched mode, then each packet should be transmitted separately. This will impose delays between packets and cause excessive load on access channels. In contrary, a burst-switched terminal will only need to access channel once and its steam of packets will be transmitted periodic at equal intervals until the end of the burst. Figure 2.3 demonstrates the modes of a burst-switched terminal.

![Figure 2.3 Burst-Switched Terminal Model](image_url)

Similar to circuit-switching, multiple burst-switched channels can be dedicated to specific services [THOM96, TAAG97b, BRAS96, BRAS97].

### 2.5 Packet-Switched Networks

#### 2.5.1 Concepts

##### 2.5.1.1 Packetisation

As aforementioned, packet-switched networks carry information *packets*. A packet has two parts: the actual information content, called the *payload*, and information about the payload, called the *meta-data* or *header*. The header can consist of fields such as the source and destination addresses, data length, sequence number, and data type. A packet resembles a postcard where the payload is the message written on a postcard and the
2 An Overview of Packet-Switching

header is the receiver’s address. To appreciate the innovation of header, recall that the telephone network carries voice using digital samples, which are not self-descriptive. Thus, the network cannot determine where samples originate, or where they are destined, without additional context information. Headers make information self-descriptive, allowing the network to interpret the data without additional context information. In particular, if the header contains a source and destination address, no matter where in the network the packet is, the network knows where it came from and where it wants to go. The network can store the packet, then “unfreeze” it, and still know what has to be done to deliver the data. In contrast, in the telephone network, the destination of a sample is derived from the time slot in which it arrives at a switch. If a preceding switch stores the sample, this timing information is lost. Thus, unlike packet networks, telephone networks cannot store and then forward samples.

2.5.1.2 Addressing

Addresses in packet-based networks are structured as a two-part hierarchy. The first part is the network number and the second is the interface number (also called the host number). Since network numbers are globally unique, and interface numbers within a network are also unique, each interface in the network is uniquely identified. Once a central authority assigns a network operator a unique network number, the operator can allocate a globally unique address with that prefix, allowing decentralised control of the address space. If addresses were “flat” or non-hierarchical, a central authority would need to check every new address for uniqueness.

2.5.1.3 Routing

A packet-switched network forwards packets from a source to a destination using the destination address field in the packet header. A router is defined as a host that has an interface on more than one network. Every router along the path has a routing table with at least two fields: a network number and the interface on which to send packets with that network number. The router reads the destination address from an incoming packet’s header and uses the routing table to forward it on the appropriate interface.
2.5.2 Existing Packet-Switched Networks

2.5.2.1 The ARPANET and the Internet

In 1967, an experimental computer network was proposed which was later to become the US Department of Defence Advanced Research Projects Agency (DARPA) Network—the ARPANET [REBE67]. Being a defence project, one main goal of the ARPANET was to devise a network that could still be operational if part of the network failed. For a number of years before 1967, DARPA had been funding the growth and development of many multi-access time-shared computer systems at a number of university and industrial centres. A cost analysis performed at the time indicated that use of packet-switching for the ARPANET would lead to more economical communications and better overall availability of utilisation of resources than many other methods [ROBE67]. This started a serious effort to define the functional details of packet switching. A specification was created for a packet-switched network, and in early 1969 a contract was awarded for the implementation of the ARPANET to Bolt, Beranek, and Newman (BBN), a Massachusetts-based engineering firm. In September 1969, the embryonic one-node network came to life when the first packet-switching computer was connected to the Sigma 7 computer at UCLA [KLEI76]. By 1972, there were 37 host computers connected to the ARPANET. Soon after in 1973, the ARPANET went beyond the boundaries of the United States by making its first international connections to England and Norway [KLEI76].

The ARPANET can be viewed as a point-to-point store-and-forward network. It pioneered the practical use of wide-area packet-switching, decentralised routing, flow control, and many applications still in use today, such as TELNET and FTP.

Based on the ARPANET, other similar packet-switched networks, such as CSNET, NEARnet [KESH97], the Cigale sub-network [POUZ74], TELNET [OPDE76], and DATAPAC [CLIP76] were built. However, such networks could not inter-connect, as they used their own independent communication protocols (languages). The research in this area resulted in higher layer adaptation by means of a set of networking protocols called TCP/IP (Transmission Control Protocol/Internet Protocol). Thus was born the Internet, the network of networks. The ARPANET was the first network in the Internet, fondly called “net ten” because of its network number. The
design of the Internet was also motivated by a packet-radio network that was also sponsored by DARPA and built at the Stanford Research Institute (SRI). Researchers using the pack-radio network wanted to communicate over the leased-line-based ARPANET. This led to a shared packet format, routing, and addressing schemes. Nevertheless, care was taken to decentralise administration to allow independent expansion of the separate networks. The notion of a “gateway”, a computer whose job was to route packets between networks, was developed. Also, it was necessary to decide that IP would not make any assumptions about the underlying transmission medium. These key decisions—decentralisation of administration, design for scaling, and bare-bones assumptions made by IP—were the results of the experiences with building this two-network Internet and have contributed greatly to the Internet’s rapid growth [KESH97].

2.5.2.2 Asynchronous Transfer Mode (ATM)

Earlier packet-switched networks, such as X.25, were designed in the sixties, at the time when only poor to medium quality transmission links were available; a BER of $10^{-6}$ was considered excellent at that time. In order to offer an acceptable end-to-end performance on each link of the network, complex protocols were therefore necessary basically performing error and flow control on every link of the connection. This link-by-link error control was essential because of the low quality of the links to ensure that the traffic increases was not too large to guarantee the required semantic transparency. The higher complexity protocols substantially increase the processing requirements and switching delay inside the network. With such packet-switched networks, it becomes very difficult to support real-time services (transfer delay is too long because of retransmission) and for high rate services (processing requirements are too high).

Asynchronous Transfer Mode (ATM) technology came into existence in response to combining the flexibility of packet-switched networks with per-user Quality-of-Service (QoS) guarantees of the circuit-switching networks. They are designed for high bandwidth, scalability, and manageability. Thus, they have the potential to subsume both the Internet and the telephone network, creating a unified infrastructure that carries voice, video, and data. Although ATM networks are likely to play an important role in the future, the research community is still debating how best to build them. Fast packet-
switching used in ATM networks, is a concept that covers several alternatives, all with the same basic characteristics, i.e. packet-switching with minimal functionality. Names other than ATM (official name used by ITU-T) were also used for alternative solutions as proposed by several organisations. The most famous of those are: ATD (Asynchronous Time Division) which was the name originally used by CNET and later taken over in Europe and FPS (Fast Packet Switching) in the United States.

ATM technology is based on some important concepts: (1) virtual circuits, (2) fixed-size packets or cells, and (3) integrated services. Together, these aspects enables integration of multiple traffic classes (unlike the telephone networks) with guaranteed QoS to individual streams (unlike the Internet which is based on best-effort QoS). They also enable large, parallel switches and provide a uniform framework for network management. These concepts are studied next.

2.5.2.2.1 Virtual Circuits
There are two ways to build a packet-switched network. The first is for each packet to carry a header with the full destination address (much like a postal address on an envelope). This is the datagram method, and each packet containing the full destination address is called a datagram. Addresses can be large, so if the average datagram length is small, this wastes bandwidth.

In the second way to build a packet-switched network, packet headers carry identifier instead of addresses, and each switch maintains a translation from the identifier to a destination. This saves header space because identifiers are small than destination address (see Figure 2.4). However, the mapping from an identifier to a destination must be set up at each switch along the path before data transmission begins. In other words, a call set-up phase must precede a data transmission phase. This variant is called virtual circuit (VC) switching which is identified by the identifiers.
An Overview of Packet-Switching

<table>
<thead>
<tr>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit: no header</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Destination Address</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet: datagram- header contains the destination address</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Virtual circuit identifier</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet: virtual circuit- header contains identifier</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.4 Fundamental Differences of Circuits, Datagrams, and Virtual Circuits

Consider a switch S1 connected to three endpoints (hosts) as in Figure 2.5. Host H1 can send a message to host H2 in one of two ways. In the datagram method, it creates a header with H2's address and hands it to switch S1. The switch reads H2's address and a routing table directs it to pass on the packet to S2. S2 in turn forwards the packet to H2.

In the virtual circuit method, H1 sends a set-up message to S1 requesting a connection to H2. Using the same routing table as before, S1 decides that the next switch along the path is S2. It forwards the set-up message to S2 and returns an identifier called the Virtual Circuit (or Channel in the ATM Forum terminology) Identifier (VCI) to the host (in this example, the VCI “1” is assigned to H1). It also creates a per-VC record that stores the output interface corresponding to the assigned VCI. After the set-up completes, if H1 wants to send a message to H2, it only needs to put the VCI in the header instead of the full destination address. By examining the identifier, S1 knows that this message is destined to S2, and simply forwards it on the correct output interface.

Figure 2.5 Virtual Circuit Concept
Now suppose host H3 also wants to communicate with H2. It must send a set-up message to S1 and get a VCI in return. Consider the link from S1 to S2. Packets from H1 on this link have "1" in the header. If packets from H3 also have the same identifier on the link, then packets from H1 and H3 will be indistinguishable. Clearly, packet from H3 must use a different VCI. There are two ways to ensure this. In one solution, any two virtual circuits sharing a link must be guaranteed to use different VCIs. A host that wants to set up a new connection must know which VCIs are being used by all other hosts sharing any link along its path. Although this solution is theoretically possible, it does not scale to large networks. A better solution is for switch S1 to resolve conflicts by VCI swapping. When S1 receives packets from H3, it modifies the VCI in the header from 1 to 2. Thus, packets on link S1-S2 have unique VCIs, as required. VCI swapping requires every switch to maintain a translation table mapping input VCIs to output VCIs. When a switch accepts a call, it chooses a unique VCI on the output trunk, and enters the translation from the input VCI to the output VCI in the translation table.

Virtual circuit is connection-oriented, that is, a source must establish a circuit or connection before it can transfer any data. In contrast, we call datagram networks connectionless networks, because a source does not need to establish a connection before communication (for example, we can send a letter without having to inform the recipient beforehand). We now make some observations about VC switching.

2.5.2.2.2 Fixed-Size Packets

ATM networks use fixed-size packets called cells instead of variable-length packets for three reasons: (1) simple buffer hardware, (2) simpler line scheduling, and (3) large parallel switches that are easier to build.

Fixed-size packets have some advantages over variable-size packets, but they also have some disadvantages. A source that wants to send data larger than the chosen size must segment the data into fixed-size cells, and the destination may need to reassemble these packets. This processing can be computationally expensive, particularly if the chosen size is smaller than the mean application-level data unit. Even in variable-sized-packet networks, endpoints must segment and reassemble data units larger than the mean application data size. This maximum size, however, is typically larger than the mean application data size. Conversely, if a source wants to send less data than the
chosen size, bandwidth is wasted. A second problem is that in fragmenting a long message into smaller fixed-size cells, the last cell may not be fully filled. This also wastes bandwidth. The designers of ATM technology believes that the advantages of a fixed-size cell size outweighs other problems [KESH97].

Figure 2.6 demonstrates the standard ATM cell defined by the ITU-T. The payload consists of 48 bytes and there is an additional 5-byte header, so that a full cell is 53 bytes long.

![5-byte header 48 bytes of data]

**Figure 2.6 Standard ATM Cell**

### 2.5.2.2.3 Integrated Services

Voice traffic has traditionally been carried in telephone networks, video traffic in broadcast television networks, and data traffic in data communication networks.

The ATM technology allows users to access one or more media simultaneously resulting in applications that exploits multiple media to provide new and interesting services such as video-conferencing and inter-active game-playing. Moreover, the network becomes easier to manage. From the consumer's prospective, a single plug in the home can provide voice, video, and data service.

### 2.6 Cellular Packet Systems

#### 2.6.1 Cellular Digital Packet Data (CDPD)

Cellular Digital Packet Data Network (CDPD) [AGOS97, SAHA97, SVEE96] is a data network designed to provide data services to Mobile-End Systems (M-ES). CDPD is used by law enforcing officers to gain access to large data-bases, for credit card verification, for financial transactions and remote telemetry. CDPD, however, can be seen to have limited capabilities and not be well suited for real time services and mass markets. Figure 2.7 shows the network-level architecture of the CDPD network.
Although CDPD can stand alone, it usually overlays on the cellular voice system by drawing radio resources from the pool of unused or free voice channels. CDPD resides outside the circuit-switched cellular network. In its simplest form, the CDPD network can be represented by a packet switching "cloud" consisting of a set of IP (RFC 791) or the CLNP (Connectionless Network Protocol, ISO 8473) capable routers, known as Intermediate Systems (IS). This packet cloud is the vehicle by which connected end systems can communicate. End systems are network addressable entities that support user and/or network application services. End system entities are categorised as either Fixed-End Systems (F-ES) or Mobile-End Systems (M-ES).

In a CDPD network, multiple M-ESs share a channel with a single Mobile Data Base Station (MDBS). M-ESs can communicate directly with the MDBS. The MDBS is the CDPD radio control system which can be found at each cell site. Mobile Data Intermediate System (MD-IS) defines the logical interface between the mobile and the fixed backbone environment. The MD-IS is concerned with CDPD specific activities, such as authentication, connection management, and mobility management. The MD-IS is the only network element that has knowledge of mobility, and because of its characteristic it can be considered the heart of the CDPD network. Resident within the MD-IS are two functions: the Mobile Home Function (MHF) and the Mobile Serving Function (MSF). Each individual M-ES is logically associated with a given MD-IS.
acting as the mobile system's "home". All datagrams destined for a mobile-end system are routed to that mobile's MHF, which is resident within the mobile's home MD-IS. The MHF maintains a table of the locations (MD-IS/MSF) from which each of its home clients is being serviced at the moment. In the forward direction, the MHF then forwards datagrams to the appropriate MD-IS serving a target client. This forwarding function is facilitated by simply encapsulating the original datagram, which is addressed to the MD-IS currently providing the mobile service.

Packets are delivered to a cell site via terrestrial land lines, microwave transport systems, or packet switched network services. Once they are at the cell site, data packets are relayed onto a cellular radio channel toward the mobile user population. The radio segment consists of forward (downlink) and reverse (uplink) channel streams which together provide a full-duplex transmission path over the air. The access method used over the air interface is Digital Sense Multiple Access Protocol with Collision Detection (DSMA/CD). Contention resolution is controlled by information broadcast by the MDBS on the reverse channel. The M-ES with a message to transmit checks the busy/idle flag. If the channel is idle then it waits for approximately 8 bit intervals and initiates transmission. If the channel is busy, the M-ES goes into defer mode and waits for a random interval to sense the channel again. Once the M-ES seizes a channel, then it transmits until it receives a decoding error. Then, it goes into back-off mode and waits for a random interval till it senses the channel again. If a channel is marked idle and two M-ESs decide to initiate transmission, a decoding error occurs and both terminals enter back-off mode. When the MDBS detects the presence of a signal, it sets the busy/idle flag to busy. If the received block is correctly decoded, the decode status flag is set to success. The M-ES continues transmission, until a block with uncorrectable errors is received by the MDBS. Then the decode status flag is set to failure. The forward and the reverse channel are always paired and they have a bandwidth of 30 kHz and a transmission rate of 19.2 kb/s.

Network layer packets undergo packet-header compression, packet encryption, segmentation and framing before transmission, in both directions. The basic unit of information transfer is a variable length ordered sequence of octets called a frame. The length of a frame may vary from 2 to 136 octets. Frames are delimited by frame flag sequences and are transmitted over the radio channel as a burst of an integral number of
blocks, interleaved with control flags and synchronisation words. Each block contains 378 bits, as it is encoded with Reed-Solomon (63,47) error-correcting code using 6-b symbols as code letters. Any frame that contains uncorrectable errors is discarded and re-transmitted.

2.6.2 General Packet Radio Services (GPRS)

The existing mobile cellular systems, such GSM/DCS/PCS, only offer low-rate circuit-switched data communications, i.e. 9.6 kb/s. This service uses the GSM standard physical layer which provides BER of less than $10^{-3}$ [MOUL92].

There are two main data items being standardised in the ETSI GSM Phase 2+ at present (1998). These are the High Speed Circuit-Switched Data (HSCSD), the General Packet Radio Service (GPRS), and the Enhanced Data Rates for GSM Evolution (EDGE). In the HSCSD, the increased user rate is achieved by assigning multiple TDMA time slots for a connection. The physical layer of the HSCSD service is again based on that of the current GSM data services. The standardisation of the GPRS, on the other hand, will fulfil the need for bursty packet-oriented services over GSM. GPRS is being developed as part of GSM Phase 2+ Specification efforts [GSM03.60, HÄMA95c]. The maximum offered rates of such services are shown in Figure 2.8.

![Figure 2.8 Development of GSM Radio Technology](image-url)
GPRS will use the physical medium provided by GSM and will offer multi-rate packet-oriented services to GSM users. It will also aim at attracting new users, as it will support new services and will provide billing according to the amount of data carried and the quality of service negotiated and not according to distance or duration of communication. GPRS will realise packet communications as a complementary service in which part of the current GSM cellular system resources will be dedicated to packet communications. This dedication could also be dynamic, meaning that, circuit- and packet-switching channel resources are shared in order to respond to the instantaneous traffic demand in the cell.

### 2.6.2.1 Applications and Service Characteristics

GPRS is designed to support services up to almost 115 kb/s, that can be flexibly allocated, according to the user's demand [GSM03.60]. Any conventional Internet-based application, such as file transfer, e-mail transmission and reception, or the World Wide Web (WWW) can be supported. Video transfer is a key element of multimedia services and is also part of the GPRS range of services. Another important application area is Road Traffic and Transport Informatics (RTTI) applications. GPS AFC (Global Positioning System Automatic Fee Collection) use GSM SMS (Short Message Service) for information exchange. GPRS is likely to replace or provide a bearer for SMS. It may also be used by the Union Internationales des Chemins de Fer (UIC) to replace the incompatible national train control system by a GPRS based European Train Control System (ETCS). Financial transactions are other possible applications. It can be seen that the range of possible applications for GPRS is very wide and covers simple as well as very advanced services.

GPRS will support two type of services, to cope with the potentially offered ones: Point to Point (PTP) as well as Point to Multipoint (PTM). The PTP service will be offered either connectionless, based on CLNP or IP, or connection-oriented, based on the ISO 8348 Connection Oriented Network Protocol (CONP). The PTM service is divided into multicast (PTP-M) or group call (PTP-G) service. A PTP-M message is received by all the subscribers in a geographical area, while a PTP-G message is only received by a group of subscribers, controlled by a multi-cast server. The offered QoS profiles are divided into five classes, presented in Table 2.1 [GSM03.60].
An Overview of Packet-Switching

### Table 2.1 GPRS QoS Profiles

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedence</td>
<td>This profile indicates the relative importance of maintaining the service commitments under abnormal conditions. Three precedence levels are proposed, namely “high priority”, “normal priority”, and “low priority”.</td>
</tr>
<tr>
<td>Delay</td>
<td>The network operator provides adequate transmission resources on the radio and network communication channels. Four delay classes are considered (class 4 represents the best effort case).</td>
</tr>
<tr>
<td>Reliability</td>
<td>Data reliability is defined in terms of the residual error rates. This class specifies the requirements on various network protocol layers. Four reliability classes are defined.</td>
</tr>
<tr>
<td>Peak Throughput</td>
<td>It specifies the maximum rate at which data is expected to be transferred across the network. However, there is no guarantee that this peak rate can be achieved. The network may limit the subscriber to the negotiated peak data rate, even if additional transmission capacity is available. Nine peak throughput classes are defined.</td>
</tr>
<tr>
<td>Mean Throughput</td>
<td>It specifies the average rate at which data is expected to be transferred across the GPRS network. This class includes 19 subclasses ranging from best effort to 0.22 b/s to 111kb/s.</td>
</tr>
</tbody>
</table>

#### 2.6.2.2 Packet Routing

Packet routing and transfer within the Public Land Mobile Network (PLMN) is supported by a new logical node, referred to as GPRS support Node (GSN). An association between the Mobile Switching Centre (MSC) and the GSN is created (Gs interface), so that routing and location area information are kept up to date in both the GSM and the GPRS network. Exchange of short messages via GPRS is considered by connecting the GSM SMS-MSC with the serving GSN (SGSN) [GSM03.60].

The GPRS network is connected to the outside world by the Gateway GSN (GGSN). GGSN acts as a logical interface to external Public Data Networks (PDNs) and transports Protocol Data Units (PDUs) to the SGSN, which routes them to the GPRS User Terminals (GUTs) within its service area. The GPRS user related data needed to perform the routing and data transfer functionality is stored within the GPRS register (GR), which is part of the GSM Home Location Register (HLR). The GR stores the routing information and maps the International Mobile Subscriber Identity (IMSI) to one
or more Packet Data Protocol (PDP) addresses, as well as mapping each PDP address to one or more GGSNs. The resulting logical architecture is depicted in Figure 2.9.

![Figure 2.9 Architecture of GPRS Network](image)

2.6.2.3 Mobility Management

Two different encapsulation schemes are used within GPRS. All packets are encapsulated with a common tunnelling protocol, between GSNs, in order to enable use of different PDPs. Encapsulation between GUTs and the SGSN is performed to decouple the logical link management from the network-layer PDPs. Before a GUT uses the GPRS network, it has to attach with a GPRS. With the attachment procedure, a logical link context is established and a Temporary Logical Link Identity (TLLI) is assigned to the GUT. After the attachment, one or more PDPs can be negotiated with the SGSN. For each negotiated PDP, the SGSN has to confirm that the GUT is allowed to use it, by interrogating the GR. If the authentication is successful, the SGSN asks the GGSN to update its routing context accordingly. Figure 2.10 presents a state model of the GUT. The GUT informs the network about every cell change in the ready state. When it is in the standby state, location information is updated only if the Routing Area (RA) is changed. RA is a subset of the GSM’s Location Area (LA). When the GUT changes RA, a location update request is sent to the SGSN, containing the identity of the new cell as well as the old and new RA’s identity. If the two RAs are serviced by the same SGSN there is no need to update the HLR/GR because the routing context remains the same. If the new RA is served by a different SGSN, however, the new SGSN
inquires the old SGSN to send the MM and PDP context of the GUT. After that, the HLR/GR and the GGSN are informed about the new routing context and PDP contexts are removed from the old SGSN.

2.6.2.4 Physical and Logical Channels

In the proposed radio link architecture, the GPRS radio channels (Packet Data Channels, PDCHs) consist of three types of logical channels: The Packet Common Control Channels (PCCCH), the Packet Traffic Channels (PTCH), and the Packet Broadcast Control Channels (PBCCH). Common control signalling information are performed between a GUT and the Base Station (BS) are performed on the PCCCHs. PBCCH broadcasts packet data specific system information. The user data and dedicated signalling are essentially transmitted on the PTCHs [GSM03.64]. The GPRS logical channels are classified in Figure 2.11.

![Figure 2.10 Mobility Management State Model](image)

**Figure 2.10 Mobility Management State Model**

![Figure 2.11 GPRS Logical Channel Structure](image)

**Figure 2.11 GPRS Logical Channel Structure**
An Overview of Packet-Switching

The PCCCH channels are as follows:

- The Packet Random Access Channel (PRACH) is used by the GUT to initiate uplink transfer (for sending data or paging response),

- The Packet Paging Channel (PPCH) is used to page the GUT prior to packet transfer,

- The Packet Access Grant Channel (PAGCH) is used in the packet transfer establishment phase to send resource assignment to a GUT prior to packet transfer,

- The Packet Notification Channel (PNCH) is used to send a PTM-M (Point To Multipoint-Multicast) notification to a group of GUTs prior to a PTM-M packet transfer. The notification has the form of a resource assignment for the packet transfer.

The logical channels of the PCCCH are mapped on different physical resources than the logical channels of the CCCH. The PCCCH does not have to be allocated permanently in the cell. If the PCCCH is not allocated, the CCCH will be used to initiate a packet transfer. The PCCCH, when it exists, is mapped onto the physical channels according to a 51-multiframe or a 52-multiframe [GSM03.64].

The PTCH channels are as follows:

- Packet Data Traffic Channels (PDTCH) is a channel allocated for data transfer. It is temporarily dedicated to one GUT or to a group of GUTs in the PTM-M case. In the multislot operation, one GUT may use multiple PDTCHs in parallel for individual packet transfer,

- Packet-Associated Control Channels (PACCH) conveys signalling information transport signalling information such as acknowledgements, timing advance, and power control information. PACCH also carries resource assignment and reassignment messages.

Each MAC PDU is transmitted over four consecutive time slots. The GPRS uses the GSM RF and time slot and structure shown in Figure 2.12.
Selective ARQ (S-ARQ) is used for error correction, based on Temporary Frame Identity (TFI). TFI also contains a job identification in order to allow for multiplexing several jobs into one PTCH.

Organisation of slot assignment to the different GUTs is done by the BSC. The channel reservation includes a time-slot number and an Uplink Status Flag (USF), coded by 3 bits. This allows for up to eight GUTs to share one PTCH. If the GUT is multislot capable, it can transmit on more than one PDCHs, thus reducing packet delay, and allowing for high-priority packets to be transmitted first. The medium access uses a Random Reservation Access (RRA) MAC protocol (which will be explained thoroughly in Chapter 4).

2.6.2.5 Mobile-Originated Transfer

Mobile originated transfer is initiated by a random access request on the PRACH, that is determined by USFs sent on corresponding downlink MPDCH. The request indicates the number of time slots needed. If the base station correctly receives the access request, it sends a reservation command, including TFI and timing advance, on the PAGCH. Because the capacity of this channel is limited, the BTS may concatenate an access grant notification for a GUT at the end of a reservation command to another GUT. If no response is received by the GUT, a re-transmission occurs for a number of times, according to a back-off algorithm. After transmission in the reserved time slots is completed, an acknowledgement is sent by the BTS. In case of an error, a Negative
Acknowledgement (NACK) is transmitted, as the S-ARQ dictates. The NACK contains the time slots the GUT should attempt re-transmission. If an ACK is not received, the GUT requests a single time-slot through the random access procedure, and when it is granted it transmits the first block of the frame. The BTS then recognises that the last ACK to this frame was not received and should be re-transmitted.

### 2.6.2.6 Mobile-terminated Transfer

For mobile terminated transfer, a page is sent from the BTS to the GUT on the PPCH. If the position of the GUT is known at cell-level, then a reservation of slots for uplink is contained within that page or an indication for downlink transmission. The GUT will either start data transmission/reception on reserved time slots or answer with a random access request. If a page without a reservation is made, the GUT answers by initiating the random access request procedure and asks for a reservation of one block, to be able to identify itself after access is granted.

### 2.6.2.7 GPRS Mobile Terminal Model

The transmission of a bursty GUT can generally be modelled with a five-state state transition model (see Figure 2.3). A data user enters a log-on state after accomplishing a call set-up or a hand-over process. When registering with the network, the GUT also specifies its individual need, i.e., maximum required bandwidth. A TLLI is assigned to the GUT in the call set-up phase. It stays in the stand-by state until a packet or a sequence of packets (message) is generated for transmission. After succeeding in contention in the PRACH, it starts reserving assigned channel(s) for the duration of the message. A GUT not only listens to PCCCHs but also to other control channels such as synchronisation (SCH) and broadcast control channels (BCCH) for monitoring of GSM system related parameters [MOUL92, GSM03.64]. An acknowledgement follows from the BSS after the frame has been received to request re-transmission of erroneous blocks and arrangement of the TDMA frame.
2.6.2.8 Data Encoding

Four different coding schemes (CS) are defined for data coding in GPRS (i.e., CS-1, CS-2, CS-3, and CS-4). The CS-1 applies $1/2$ rate convolutional coding with a 40-bit FIRE code and is used for all signalling messages. The CS-2 and the CS-3 use half rate convolutional coding with code punctuation of $2/3$ and $3/4$. The CS-4 does not apply coding, and hence, the output data rate will not change. In CS-2 to CS-4, the USF is coded separately, in order to allow faster decoding. The data flow is presented in Figure 2.13. The structure of the bursts used is similar to the one used in GSM (see Figure 2.12).

![Data Flow in GPRS Diagram](image)

**Figure 2.13 Data Flow in GPRS**

2.7 Satellite Packet Systems

As wired packet switched networks grow in size and coverage, the need to provide inexpensive, long-haul, high-capacity communication channels becomes more pressing. The satellite communication technology offers promise as a solution for such data transmissions.

Few packet satellite networks are in operation for experimental purposes [TOBA84]. The first to be implemented was SATNET, which uses a 64-kb/s channel on
an INTELSAT satellite [JACO78, CHU79] and provides multiaccess/broadcast services. A similar network to SATNET, called TACNET (or MATNET), consists of a ship-to-ship and ship-to-shore network using a 19.2-kb/s channel on the FLTSAT satellite [JACO78]. The TACNET employs link encryption and error-correction techniques appropriate to the bursty shipboard RF environment [EVEN82]. Another implementation utilises a 3-Mb/s satellite channel on WESTAR, which has an experimental emphasis on the sharing of a wide-bandwidth channel by many packet voice users [FALK83, HEGG83].

2.8 Summary

Packet switched networks, such as the ARPANET and the Internet, came into existence in response to the need for rapid, efficient, and economical data communications. In this chapter, the fundamentals of packet-switching techniques have been reviewed. A variation of packet-switching called burst-switching was presented which will be used in most of our proposed protocols in this thesis. Protocols that use burst-switching are referred to as reservation-based protocol (see Chapter 4). Then a brief history of packet-switching from the ARPANET time to the Internet and the ATM era was presented. Evolution of packet communication from the wired to the wireless world such as cellular and satellite was also discussed. Thus was presented CDPD, GPRS, and packet satellite systems. Such systems (in particular GPRS) will be used in the following chapters for design and evaluation purposes.
3.1 Introduction

This chapter provides an overview of packet-based statistical models of multimedia service components. Efforts have been made to classify and to define packet-based models of multimedia service components [ADAS97, ATMF98, CLAF93, ITU93]. These service components range from data block transfers, such as non-real-time (NRT) Internet services, to real-time (RT) delay-sensitive services, such as voice. Here, the focus is only on packet model of the variable rate representation of services in order to take advantage of statistical multiplexing gain. The packet-based service models developed here will provide a background for evaluation analysis in the subsequent chapters. Explanations of statistical distributions used in this chapter can be found in Appendix A.

3.2 Services Classification

Services can be classified in many ways. Here, we present a classical approach presented in ITU-T Recommendation I.211 [ITU93].

This recommendation identifies two main classes of services; interactive services and distribution services. The former is divided into sub-classes of conversational services, messaging services and retrieval services, while the latter distinguishes between distribution services with and without user individual presentation control (see Figure 3.1).
3.2.1 Interactive Services

3.2.1.1 Conversational Services

Conversational services offer bi-directional communication with real-time (RT) constraints. Examples of conversational services include:

- telephony;
- video-telephony;
- video-conferencing;
- remote video surveillance/monitoring;
- video/audio information transmission service;
- multiple sound programme signals;
- voice-band data services;
- high speed unrestricted digital information transmission service;

3.2.1.2 Messaging Services

Messaging services provide communications via store and forward transportation of data, mailboxes and/or message handling functions (e.g. conversation, editing and processing). Examples of messaging services include:

- paging and short messaging;
- document mailbox (electronic mail, voice mail, fax mail);
- video mailbox (electronic mail including moving pictures and sound).
3.2.1.3 Retrieval Services

Retrieval services provide retrieval of stored information at database centres. Information is sent on demand only and delivered at the time of the request. Examples of retrieval services include:

- teletex
- videotex;
- database access (tele-software, directory services);
- audio, video, image and document retrieval service (library, archives);

3.2.2 Distribution Services

3.2.2.1 Distribution Services Without User Individual Presentation Control

Services which provide distribution without user individual presentation control form a continuous flow of information, which is distributed from a central source to an unlimited number of customers. The order and time of presentation are predetermined; i.e., they are not user controlled. Potential services include:

- TV programme distribution at various quality modes (PAL, SECAM, NTSC, EDTV, HQTV, HDTV);
- high quality video broadcast;
- high quality audio broadcast;
- document distribution service (electronic newspaper, electronic publishing);
- unrestricted digital information transmission service (data broadcast, data multicast);
- message broadcast/multicast (emergency/public announcement, driver information service).

3.2.2.2 Distribution Services With User Individual Presentation Control

Services which provide distribution with user individual control present a sequence of information entities (e.g. frames) which are repeated in cycles. By selecting the start of the sequence, the user can observe the presentation from the beginning and thus control the start and the order of presentation. Typical service examples include:
• videography (remote education and training, tele-advertising, news retrieval);
• restricted digital information transmission service (geolocation, vehicle navigation support, traffic congestion management).

3.3 Service Definition Parameters

Ideally we would like to have a finite set of parameters by which any service can be described. We call these parameters “Service Description Parameters” (SDP). SDPs are derived based on statistical and stochastic features of the traffic stream generated by a source.

One important class of services is that of on-off services. This service class includes services such as the Internet applications, packet voice, packet video, bulk data transfer and the like. For example, when a WWW session is established, requests for documents are triggered by the user from time to time. The generated traffic of each on-off source can be shown by occasional bursts of data separated by inactive periods. Apart from non-real-time data services, real-time services (such as voice and video) can be expressed with on-off models [BRAD69, MAGL88].

An important SDP for on-off sources is the activity factor ($\Gamma$) defined as

$$\text{Activity Factor } \Gamma = \frac{\text{Average Busy Period}}{\text{Average Busy Period} + \text{Average Idle Period}} = \frac{T_{on}}{T_{on} + T_{off}}$$

(3.1)

The Burstiness Factor ($\beta$) is another well-known descriptor used for the worst-case analysis

$$\text{Burstiness Factor } \beta = \frac{\text{Maximum Busy Period}}{\text{Average Busy Period}} = \frac{T_{on,\text{MAX}}}{T_{on}}$$

(3.2)

Interaction of on-off sources have some specific features. The on-off pattern generated by two conversing sources is shown in Figure 3.2.
The main feature of two-way (conversational) communications is the amount of inter-activity. For example, if the activity factor of one source is far less than the other source, it implies that the communications has low inter-activity (at the extreme case where one of the activity factors is zero, the communications become one way). On the other hand, if the activity factors are similar, the inter-activity will reach its peak. As a result, a SDP called $I_{12}$ can be defined as

$$I_{12} = \frac{\text{MIN}(F_{Ai}, F_{A2})}{\text{MAX}(F_{Ai}, F_{A2})}$$

where $I_{12}$ is the inter-activity factor of sources 1 and 2. $I_{12}$ can vary between 0 (representing no inter-activity) and 1 (representing the maximum inter-activity). For conversation speech, the inter-activity factor has shown to be around 6% [BRAD69].

There are times when both of the parties activities overlap. In order to quantify this overlapping, a SDP called Mutual Inter-activity Factor between two on-off sources is defined as the result of dividing the average period of time that the two source are mutually active by the average activity of both of them together. Or

$$\text{Mutual Interactivity Factor} = MI_{12} = \frac{\text{Average Mutual Busy Periods of } S_1 \text{ and } S_2}{\text{Average Busy Periods of } S_1 \text{ and } S_2}$$

where $MI_{12}$ represents the mutual activity factor between the source 1 and 2.

### 3.4 Packet Voice

A conversational speech can be characterised with a sequence of talkspurts (service times, messages) separated by silent spurts (idle times) [BRAD69]. Norwine first introduced the idea of exploiting the silence periods in a conversation in order to
improve the channel efficiency and overall system capacity [NORW38]. Bursty voice transmission was commercially pioneered in the “Time-Assigned Speech Interpolation” (TASI) system [BULL62]. After the advent of digital communications, an improved version of TASI, known as Digital Speech Interpolation (DSI), was then produced in which VBR speech encoding is employed to tackle temporary overload plights in the network [LYGH74].

The concept of packet voice (PV) transmission had been proposed and investigated at the beginning of ISDN research and development [FUBN79]. The primary purpose of packet voice transmission was to optimise the utilisation of finite communications capacity by sending only significant voice signals in the form of packets. The active periods of voice (talkspurts) can be extracted by means of a Voice Activity Detector (VAD). Figure 3.3 shows the operation of voice activity detection on a speech signal. Empirical results show that the speech activity factor is typically around 40% in quiet background, whereas, in noisy background it could increase to 80% [BRAD69].

![Figure 3.3 Voice Activity Detection](image)

Figure 3.4 shows the operation of a typical wireless PV terminal using VAD. As shown, the activity of the voice source is detected after source-encoding. Then the header (containing the identifiers or addresses) is appended to the digital stream for packetisation purposes. The packet size is defined by the network.
3.4.1 PV Source Model with Slow Voice Activity Detection

A Slow Voice Activity Detector (S-VAD) is only sensitive to long silences and talkspurts. Figure 3.5 demonstrates packet generation pattern of a PV source using such a VAD.

As shown in [BRAN69], talkspurt and gap periods are of negative exponential distributions with means denoted by $T_{on}=1/\mu_{PV}$ and $T_{off}=1/\lambda_{PV}$, respectively. These means are functions of voice activity threshold and degrading factors such as background noise, echo or delay [BRAN69]. Typical values for $T_{on}$ are 1000 ms and 1410 ms and for $T_{off}$ are 1350 ms and 1740 ms [GOOD91, DEVI93]. Voice activity factor, $\Gamma_{PV}$, is defined by (3.1). Therefore, a S-VAD PV terminal can be modelled as a two-state Markov model as shown in Figure 3.6 [BRAD69, GOOD89].
Figure 3.6 Markov Model for Packet Voice Source with Slow VAD

When in “SIL” (silence), the terminal generates no packet and when in “TLK” (talkspurt), it generates packet at a constant rate [GOOD91].

As active and inactive periods both follow negative exponential distributions, the transition probabilities of $\sigma$ and $\gamma$ are calculated by [GOOD91]

$$\sigma = 1 - e^{-\frac{\tau}{T_{PV,m}}}$$ (3.5)

and

$$\gamma = 1 - e^{-\frac{\tau}{T_{PV,sg}}}$$ (3.6)

where $\tau$ is the time (simulation) step after which the Markov chain is updated.

3.4.2 PV Source Model with Fast Voice Activity Detection

A Fast-Voice Activity Detector (F-VAD) is very sensitive and can even detect short silence gaps, "pauses", occurring between syllables and between words, "minispurts". Hence, the activity factor of a speech source can be further reduced. Figure 3.7 demonstrates the talkspurt-silence pattern created by an F-VAD.
We denote the mean periods of talkspurt, minispurt, silence, and pause with $T_{on}^{(1)}$, $T_{on}^{(2)}$, $T_{off}^{(1)}$, and $T_{off}^{(2)}$, respectively. Typical values are $T_{on}^{(2)} = 158.2$ ms, $T_{off}^{(1)} = 726.2$ ms and $T_{off}^{(2)} = 36.2$ ms [STER95].

Experimental results have shown that minispurts and pauses still follow negative exponential distributions [GOOD91]. Therefore, the distribution of combined silence and pause periods forms a two-stage hyper-exponential distribution, $H_2$ (see Appendix A) [OREI87, KLEI97] where $\lambda_1 = (1/T_{off}^{(1)})$ and $\lambda_2 = (1/T_{off}^{(2)})$ are parameters of silence and pause distributions and $\alpha_1$ (probability that a minispurt is followed by a pause) and $\alpha_2$ (probability that a silence period is followed by a talkspurt). The parameters $\alpha_1$ and $\alpha_2$ are given approximately by [GOOD91, OREI87]

$$\alpha_1 = 1 - \frac{T_{on}^{(2)} + T_{off}^{(2)}}{T_{on}^{(1)}}$$

(3.7)

and

$$\alpha_2 = \frac{T_{on}^{(2)} + T_{off}^{(2)}}{T_{on}^{(1)}}$$

(3.8)

Since the duration of events for the case of an F-VAD source follow negative exponential distributions, similar to S-VAD, we can model the source with a three-state Markov model as shown in Figure 3.8 [GOOD91].
The state transition probabilities are calculated as

\[ \sigma_M = 1 - e^{-\frac{1}{T_{\text{off}}}} \]  \hspace{1cm} (3.9)

\[ \gamma_M = 1 - e^{-\frac{1}{T_{\text{on}} + T_{\text{off}}}} \]  \hspace{1cm} (3.10)

\[ \gamma_I = \frac{T_{\text{on}}(2) + T_{\text{off}}(2)}{T_{\text{on}}(1)} \]  \hspace{1cm} (3.11)

### 3.4.3 Hangover

Even though the energy of speech signal is low at the end of a talkspurt, the signal may contain important information for transmission. The hangover technique is deployed to eliminate sharp cut-off at the end of talkspurts, and hence to minimise the degradation of speech quality. Another advantage of such a technique is reduction in number of access attempts (see Chapter 5).

The hangover mechanism is parameterised by hangover frame duration, \( T_h \), defined as the vulnerable time after the end of a talkspurt which will be still considered as talkspurt [DUNL88, TAAG96b].

### 3.4.4 Packet Dropping

Voice is a real-time service that is very sensitive to delay. If the end-to-end delay builds up to around 300 ms, conversation starts to become annoying, and at around 500 ms it is unacceptable.
The access mechanism used for PV burst transmission will impose an extra delay. Considering other delays associated in packet-switched networks, the maximum delay budget assumed for the PV wireless access, $D_{v,\text{max}}$, is somewhere between 10-50 ms [GOOD91, WAIS88]. Any voice packets delay over $D_{v,\text{max}}$ will be discarded at the PV terminal. For an acceptable QoS, front-end clipping of less than 1% is desirable [GOOD91, WEIN83].

### 3.5 Internet Packet Services

In the following, statistical models for traffic loads of typical Internet-based services are presented. These services are electronic mail (e-mail), file transfer (FTP), and World Wide Web (WWW).

#### 3.5.1 Electronic Mail (E-MAIL)

Electronic mail (e-mail, E-MAIL) allows a user to compose memos and send them to individuals or groups.

The E-MAIL source is an on-off non-interactive service. The OFF-STATE represents the state in which the source (terminal) is thinking (not engaged with E-MAIL) or composing a new message. When in the ON-STATE, the terminal has composed a message and is in process of transferring it.

E-MAIL messages could be either a small memo of several text lines, or lengthy encoded\(^1\) file transfer over E-MAIL (namely known as “FTP-by-mail” [MOGU92]). The FTP-by-mail has gained great popularity recently for transferring documents, data files, and computer programs amongst working groups or individuals. The message size distribution for the case of FTP-by-mail depends on the file size distribution (to be discussed in the FTP Service modelling of Section 3.6.3) and the encoding scheme used.

A statistical model called FUNET has been proposed for conventional E-mail messages. The PDF of message sizes has been approximated by a Cauchy distribution with location parameter of $k = 800$ bytes and scaling factor $\alpha = 1$ and a maximum size

\(^1\) The standard E-MAIL transfer protocol known as SMTP (Simple Mail Transfer Protocol) use ASCII text for communicating between a client and server. Therefore, binary files should be encoded into ASCII files before transferring.
of 10 Kbytes [BRAS97, UMTS30.03]. The Cauchy distribution with cut-off of \( m=10 \) Kbytes yields a mean of \( \bar{X}_m = 830 \) bytes per E-mail message (see Appendix A). In this study, the arrival of E-mails are assumed to follow Poisson distribution.

### 3.5.2 File Transfer (FTP)

The File Transfer Protocol (FTP) is an application that allows users to send ("put") or receive ("get") arbitrary large files. Modelling FTP is particularly important [CLAF93]. As with WWW session arrivals, user-generated FTP session arrivals are well-modelled as Poisson with fixed hourly rates.

Figure 3.9 depicts the traffic pattern generated by a session. As seen the session is divided into two parts: "connections" and " spacings". Connections refer to commands or back data transfer and spacings represent the amount of time between the end of one connection within a session and the beginning of the next. The FTP activity model is a special case of WWW model in which a packet call only carries one packet.

![Figure 3.9 Traffic Pattern Generated by an FTP Session](image)

**Profile of Spacing:** The empirical research shows that the spacings follow a log-normal distribution [PAXS95]. From the CDF graph of spacing times provided in [PAXS95], it can be observed that the inflection point (i.e., the point at which the second derivative of a function becomes zero) is around 4 seconds thus

\[
e^{\mu - \sigma^2} = 4
\]

And assuming that the mean spacing time is \( T_{off} = 10 \) seconds

\[
e^{\mu + \sigma^2/2} = 10
\]

hence

\[
\mu = 2, \sigma^2 = 0.6
\]
Profile of transferred file sizes: As shown in [PAXS95], the file sizes transferred by FTP have a Pareto distribution. Reference [PAXS95] reports a mean file size of $\bar{X} = 22370$ bytes and the shape parameter ranging $0.9 \leq \alpha \leq 1.4$. In this study, the mean $\alpha$ (i.e. $\alpha = 1.15$) has been used. Hence, the position parameter can be calculated as $k = 2917.8$ bytes (Appendix A). The activity factor of an FTP session ($\Gamma_{FTP}$) for a given traffic channel rate $R_c$ can be calculated as

$$\Gamma_{FTP} = \frac{\bar{X}}{R_c} \left( \frac{X}{R_c} + T_{off} \right)$$

Figure 3.10 verifies the accuracy of equation (16) for different channel capacity (average taken over 100 FTP open sessions). As seen, the activity factor of FTP sources decreases considerably for higher channel rates. That is because with higher transmission rate the bulk data (file) will be transferred faster and hence the session will be active for less amount of time.

![Figure 3.10 FTP Activity Variation with respect to Channel Capacity](image)

3.5.3 World-Wide-Web (WWW)

The traffic generated by a WWW session is different with that of a typical on-off source like PV. Figure 3.11 demonstrate a typical WWW browsing session, which consists of a sequence of packet calls. We only consider packets from a source which may be at either end of the link but not simultaneously. The user initiates a packet call when requesting an information entity. During a packet call several packets may be generated,
which means that the packet call consists of a bursty sequence of packets [UMTS30.03, VICA97].

![Packet Call Diagram]

**Figure 3.11** WWW Traffic Pattern Generated by a Browsing Session

A packet service session contains one or several packet calls depending on the application. In a WWW browsing session a packet call corresponds to the downloading of a WWW document. After the document is entirely arrived to the terminal, the user is consuming certain amount of time for studying the information. This time interval is called *reading time*.

**Table 3.1** Default Mean Values for Distributions of a Typical WWW Service

<table>
<thead>
<tr>
<th>Net Channel Rate (Kb/s)</th>
<th>$N_{pc}$</th>
<th>$T_{off}^{(1)}$ (s)</th>
<th>$N_d$</th>
<th>$T_{off}^{(2)}$ (s)</th>
<th>PacketSize Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>5</td>
<td>412</td>
<td>25</td>
<td>0.5</td>
<td>$k = 81.5, \alpha = 1.1$</td>
</tr>
<tr>
<td>32</td>
<td>5</td>
<td>412</td>
<td>25</td>
<td>0.125</td>
<td>$k = 81.5, \alpha = 1.1$</td>
</tr>
<tr>
<td>64</td>
<td>5</td>
<td>412</td>
<td>25</td>
<td>0.0625</td>
<td>$k = 81.5, \alpha = 1.1$</td>
</tr>
<tr>
<td>144</td>
<td>5</td>
<td>412</td>
<td>25</td>
<td>0.0277</td>
<td>$k = 81.5, \alpha = 1.1$</td>
</tr>
<tr>
<td>384</td>
<td>5</td>
<td>412</td>
<td>25</td>
<td>0.0104</td>
<td>$k = 81.5, \alpha = 1.1$</td>
</tr>
<tr>
<td>2048</td>
<td>5</td>
<td>412</td>
<td>25</td>
<td>0.00195</td>
<td>$k = 81.5, \alpha = 1.1$</td>
</tr>
</tbody>
</table>

The following statistical model has been developed to catch the typical behaviour described in Figure 3.11 [UMTS30.03, VICA97, KARL98, PAXS95]:

**$N_{pc}$**: *Number of packet call requests per session*: This a geometrically distributed random variable with a mean $\mu_{N_{pc}}$ (packet calls).
$T_{\text{off}}^{(1)}$: Reading time between two consecutive packet call requests in a session:

This is a geometrically distributed random variable with a mean $\mu_{\text{Dpc}}$ (model time steps).

$N_d$: Number of packet in a packet call: It is again modelled with a geometrically distributed random variable with a mean $\mu_{Nd}$ (packet).

$T_{\text{off}}^{(2)}$: Time interval between two consecutive packets inside a packet call: It is a geometrically distributed random variable with a mean $\mu_{\text{Dd}}$ (model time steps).

$S_d$: Packet size: The traffic model can use such packet size distribution that suits best for the traffic case under study. Pareto distribution with cut-off is used. Packet Size is defined with the following formula:

$$\text{PacketSize} = \min(X, m)$$

where $X$ is normal Pareto distribution random variable ($\alpha=1.1$, $k=81.5$ bytes) and $m$ is maximum allowed packet size, $m=66666$ bytes. The PDF of the PacketSize is a cut-off Pareto distribution and the mean of $\overline{X}_m = 480$ bytes (see Appendix A). As seen in Table 3.1, some of the distribution parameters are dependent on the traffic channel rate.

For a given traffic channel rate ($R_c$), all the bytes can be translated to time and hence the total activity factor can be estimated by

$$\Gamma_{\text{WWW}} = \Gamma_1 \Gamma_2$$  \hspace{1cm} (3.13)

where $\Gamma_{\text{WWW}}$ represents the overall activity factor, $\Gamma_1$ is the session activity (session times as active and reading times as inactive), $\Gamma_2$ represents the packet call activity (packet times as active and inter-arrival times as inactive). Figure 3.12 compares the approximate activity factor per session with that of the simulation (average taken over 100 WWW open sessions). Similar to the FTP case, we can see that the activity factor highly depends on the channel transmission rate and is very low for a WWW session.
3.5 Other Packet Data Services

There are other data packet models used in this thesis which have been proposed in the literature as either assumptions (see Section 3.7.3) or ETSI GPRS ad hoc evaluation guidelines (see Section 3.7.1-3.7.2) [NAND90, GSM03.64, BRAS97].

3.5.1 Railway Control Data Model

A negative exponential distribution has been proposed for modelling of Railways control applications [GSM03.64]. The mean and the maximum message length are proposed to be 170 bytes and 1000 bytes, respectively.

3.5.2 Mobitex Data Model

The Mobitex model is based on statistics collected from a fleet management application using the Mobitex wireless packet data network in Sweden. This model represents a two-way asymmetric uniform distributions for uplink and downlink traffic load parameterised as [GSM03.64]

\[
\text{Uplink: } 30 \pm 15 \text{ bytes} \\
\text{Downlink: } 115 \pm 57 \text{ bytes}
\]  

(3.14)

3.5.2 Poisson Arrival Packet Data Model

Poisson Arrival Packet Data (PAPD) model is widely used in the literature to model a
simple fixed-rate Poisson distribution for equal size data packets [NAND90, IPRMA].
The packet arrival pattern is shown in Figure 3.13

![Fixed Size Packets](image)

**Figure 3.13 Packet Arrival Pattern**

Assuming each packet contains $F_d$ bits of raw information bits, the aggregate generated bit rate, $R_d$, is given as

$$R_d = \Lambda_d F_d \quad (3.14)$$

where $\Lambda_d$ is the aggregate arrival rate. If the aggregate is generated by $M_d$ number of data packet sources with packet inter-arrival rates of $\lambda_d$, using the Poisson superposition property [PAPO91], the aggregate rate is

$$\Lambda_d = M_d \lambda_d \quad (3.15)$$

It should be noted that terminals with finite buffer sizes would impose different statistical modelling.

### 3.6 Summary

In this chapter, the statistical models for various packet-based services have been reviewed. These services include packet voice, Internet-based services (such as E-mail, FTP, WWW), and other packet services reported in the literature (such as Poisson packet arrival data) and the ETSI GPRS evaluation guidelines (such as Railway control, Mobitex).

Some bursty packet sources, such as WWW and F-VAD PV, show two levels of activity. In such cases, an approximate overall activity factor was developed expressed
in Equation (3.13). Comparison of such an approximation with simulation results proves accurate (see Figure 3.10 and Figure 3.12). Knowing the actual active duration \( T_{on}^{(2)} \) and the overall activity factor, we will be able to calculate the overall average silence duration using Equation (3.1). The idea of two-level activity factor can be easily expanded to include multi-level activities.

The statistical models presented here (which are mostly beyond simple exponential distributions) are incorporated into our evaluation model for simulation and optimisation purposes in the following chapters.
CHAPTER 4

Packet Multiple Access Protocols
(Theory)

4.1 Introduction

This chapter concentrates on the theoretical background and general analysis of packet multiple access protocols which will be followed by numerical evaluations in Chapter 5. Packet-based multiple access protocols are classified and briefly explained. Then, our attention is focused on an important part of most multiple access protocols, random access mechanisms, where guidelines for optimum and stable designing in mixed services environments are provided. A normalised offered load is derived for mixed services which reveals new insight into the behaviour of a protocol enabling performance prediction. An analytical method called “ Aggregate Source Modelling” (ASM) is proposed. This analytical model aggregates various mixed services into a representative service model that can be fed into the monomedia to perform the multimedia system analysis. Two statistical upper-bounds (one for contention-based and another for reservation-based in general) have been derived. The proposed upper-bounds are used to define efficiency factors which enables performance comparison of various multiple access protocols. Finally, design considerations for operational environments (indoor, cellular and satellite systems) with propagation delays and channel impairments are discussed.
4.2 Classifications of Multiple Access Protocols

The classification of multiple access protocols can be made in many ways [TOBA80]. Protocols can be classified in terms of service(s) that they support in a given operational environment [TOBA80]. Here, we propose a novel classification of packet-base multiple access protocols based on the four switching concepts (circuit-, packet-, burst-switching, and hybrid-switched) or a mixture of those discussed in Chapter 2. This classification are presented in Figure 4.1.

Circuit-switched multiple access protocols (or fixed assignment techniques) dedicate a fixed portion of the available channel capacity (usually equal to the service peak-rate) to each terminal. They perform very well with constant-bit-rate connections in terms of both service quality and channel efficiency. However, their performance (e.g., access delay and packet dropping rate) decrease dramatically when supporting bursty traffic. The most common form of these techniques are FDMA and TDMA. This type of protocols can provide circuit-switched bearer services. Further bandwidth can be allocated to a terminal through assigning more than one channel unit. This technique is called High Speed Circuit Switching (HSCS).

In our terminology, a packet-switched protocol sends packets separately. This class of protocols implement packet-switching concept in the wireless environment. The access mechanism could be contention-based such as ALOHA [ABRA73] or carrier sense [TOBA80].

Packet-switched protocols demonstrate low channel utilisation and variable transmission delay especially for “highly active”, bursty, and RT traffic. In burst-
switched protocols, on the other hand, consecutive packets are grouped to form a burst or message. These protocols perform the burst-switching task. Then burst packets are sent continuously once access is accomplished. For each burst, an access to the wireless channel is required, hence, burst-switched protocols could be further classified according to the access mechanism they employ. These access mechanisms could be fixed-assignment (TDMA-reservation [TASA83]) contention-based (random reservation [TASA83, GOOD91, TAAG97a]), polling-based, or any mixture of those. Random Reservation Access Protocols (RRAPs) have been of interest to researcher due to its implementation simplicity and practicality.

Each of the three multiple access types explained above is only suitable for supporting one individual service class. Therefore, in order to support mixtures of services, a hybrid combination of such protocol components are required. We call a protocol of hybrid-switched type if it supports more than one switching concept [MITR93, WEIS95, JAFA98]. Figure 4.2 shows the general format of such protocols.

![General Channels' Structure for Hybrid Multiple Access Protocols](image)

**Figure 4.2 General Channels' Structure for Hybrid Multiple Access Protocols**

Adaptive strategies could also accompany hybrid-switched protocols in which the boundaries between logical channels can be adapted to varying traffic demands, in order to achieve near-optimum performance [WEIS95].
4.3 Proposed Radio Packet Multiple Access Protocols

There have been quite a number of proposals for packet-based protocols. In the first sight such protocols might look very different, however, the key ingredients are very similar. Below, we present a number of important packet-based multi-access protocols.

4.3.1 Reservation-ALOHA (R-ALOHA)

Reservation-ALOHA (R-ALOHA) protocol originally proposed by Crowther [CROW73] contains both elements of slotted ALOHA (packet-switching) and reservation (burst-switching). The R-ALOHA protocol is simple in its principle and easy to implement like slotted ALOHA, it is suitable for users with multi-packet messages (bursts) like any burst-switched (reservation) protocol. In the R-ALOHA protocol, the time axis is divided into slots whose duration is equal to the transmission time of a packets. The slots are organised into transmission frames. The rules for transmitting packets into slots are as follows [TASA83]:

1) If slot \( m \) (say) had a successful transmission by user \( X \) (say) in the previous frame, slot \( m \) is off limits to everyone except user \( X \). Slot \( m \) is said to be reserved by user \( X \). Note that user \( X \) has exclusive access to slot \( m \) as long as he continues to transmit a packet into it in every frame.

2) Those slots in the previous frame which were empty or contained a collision are said to be unreserved or available. Unreserved slots are available in current frame for contention by all users in exactly the same manner as slotted ALOHA.

There are two slightly different protocols in R-ALOHA, depending upon whether an end-of-use flag is included in the header of the last packet before a user gives up his reserved slot [LAM80].

4.3.2 ALOHA Reservation (ALOHA-R)

The ALOHA-Reservation (ALOHA-R) protocol is a burst-switched demand assignment technique with distributed control using an explicit reservation scheme. This protocol uses a channel with a fixed frame structure. The channel is divided into time slots whose
duration is equal to the transmission time of a packet. The slots are organised into frames with a fixed number of slots in each frame. Each frame consists of an access (reservation) sub-frame with a traffic (data) sub-frame. Each slot in the access sub-frame is further subdivided into mini-slots. The mini-slots are for access packets to be used on a contention basis with the slotted ALOHA protocol. The slots in the data sub-frame are for reserved traffic packets.

### 4.3.3 Packet Reservation Multiple Access (PRMA)

A modified version of R-ALOHA, called Packet Reservation Multiple Access (PRMA) [GOOD89], was proposed to deliver an air-interface solution for high capacity environments with low propagation delays such as LANs and indoors. In PRMA, the (uplink) channel is slotted and grouped into fixed length frames. Downstream traffic (base-to-terminal) is scheduled by the base station and can be transmitted either by time sharing the same channel with upstream and downstream (TDD) or over a separate channel (FDD).

Using our classification, PRMA is a hybrid-switched packet protocol. For this protocol, packet terminals are classified as either "periodic" or "random". Periodic (or bursty) traffic sources generate fixed-size packets periodically when in the active state. A typical example of such sources is packet voice with VAD or FTP (see Chapter 3). PRMA assigns burst-switched channels to periodic sources. On the other hand, data packets such as keyboard entries to a computer terminal (TELNET), signalling messages (e.g. contention access packets) and system control information, are labelled as "random". PRMA allocates packet-switched channels to random traffic sources.

In this protocol, packet-switched and burst-switched channels are dynamically shared and no minimum and maximum has been set on numbers or portions of these channels. As shown in Figure 4.3, channels are labelled as "reserved" or "available". A reserved channel is a channel temporarily reserved by a bursty source as opposed to an available channel.
PRMA assumes that the result of contention attempts for a time slot will be acknowledged immediately after the corresponding time slot. Such assumptions has limited the applicability of such protocol to environments with low propagation delays.

Extensive modifications, performance evaluations, and analysis have been carried out for PRMA in error-free and erroneous channels [GOOD91, NAND91, NAND94, WONG93, WU94, PRIS95, QIU96, NARA96, KIM96, GUNT96, FRUL96, BRAN96, CHUA93, BIAN97]. However, due to the mathematical complexity, the analysis and evaluation of such protocols have been limited only to single service or at best two service types (usually voice and low bit rate data) [NAND91, NAND94, WONG93, WU94, QIU96, STER95].

4.3.4 Integrated PRMA (IPRMA)

The Integrated PRMA (IPRMA) [WONG92] protocol is a variation of PRMA which supports packet voice and packet data services. The packet voice sub-system is identical to that of PRMA. In this protocol a minimum number of channels are allocated to higher priority class (i.e. packet voice here), even though, no maximum for the number of packet voice reservations in a frame has been set. Unlike PRMA, IPRMA provides data users with multiple burst-switched channels. Multiple channels are assigned temporarily to data users while simultaneously satisfying the minimum number of channels required for voice terminals.
4.3.5 Frame Reservation Multiple Access (FRMA)

The Frame Reservation Multiple Access (FRMA) protocol is another modified version of PRMA [NARA96]. In this protocol, at the beginning of every frame, the base classifies every time slot as either: (1) reserved by a voice terminal, in which case the identity of the voice terminal that holds a reservation for this time slot is also broadcast, (2) available for packet voice contention, in which case all voice terminals that are in a talkspurt but do not have a reserved slot yet can contend in these slots, or (3) a data slot (packet-switched channel) for use of the data traffic. A contending terminal voice is allowed to content for every available voice contention time slot in a frame. In case a voice terminal succeeds in contention on more than one occasion in a frame, the base reserves one of the time slots for this terminal in an arbitrary manner. Thus, at the end of a frame, each voice terminal can hold a reservation only for at most one time slot.

4.3.6 Burst Reservation Multiple Access (BuRMA)

Based on PRMA, the "Burst Reservation Multiple Access" (BuRMA) protocol [TAAG96b] was proposed to implement RT and NRT mixed packet services for existing circuit-switched systems with minimum modification. Different versions of such protocol was developed for the full-rate and half-rate GSM as well as the DECT (Digital Enhanced Cordless Telephony) transmission system. BuRMA is a hybrid-switched protocol consisting of packet- and burst-switching components. The overheads are designed to a minimum amount by cutting unnecessary headers from packets of the same burst. The packet structure details and performance evaluations of BuRMA are provided in Chapter 6.

4.3.7 PRMA++

In the Advanced TDMA (ATDMA) project of RACE (Research and development in Advanced Communications technologies in Europe), a version of PRMA, known as PRMA++, was developed [DEVI93]. The objective was to develop an Adaptive TDMA air-interface to meet the predicted requirements of UMTS. UMTS is expected to operate in a wide variety of environments from typical urban by environment to open country with mountainous terrain, to indoor/office environments. Therefore, three air-interface types (macro-cells, micro-cells, and pico-cells) based on PRMA++ were designed.
Similar to ALOHA-R and unlike PRMA, PRMA++ has separated traffic (burst-switched) channels from contention (packet-switched) channels in the uplink. Frame structure of the PRMA++ protocol is given in Figure 4.4.

4.3.8 Mitrou's Protocol

This protocol is a hybrid-switched multiple access protocol proposed by Mitrou [MITR90] for a micro-cellular mobile communications system. As no name was given to this protocol, we call it Mitrou's Protocol! The protocol supports three service classes, namely, circuit-switched voice, burst-switched voice, and data. A hybrid multiplexing scheme with no boundaries is employed, which performs statistical multiplexing of connections of three classes at two different levels, the call level (for circuit-switched voice) and the burst-level (for burst-switched voice and data) [MITR90, MITR93].

4.3.9 Time Multiplexed-Base Controlled Multiple Access/Collision Detection (TM-BCMA/CD)

Another RACE ATDMA protocol was proposed under the name of Time Multiplexed-Base Controlled MA/Collision Detection (TM-BCMA/CD) [DUNL93]. It is a purely burst-switched protocol for both packet voice and packet data. Terminals with packets ready for transmission may access the uplink channel at the start of the next mini-slot after an IDLE flag on the downlink. When the base station detects a packet preamble at
the start of a mini-slot it synchronises bit timing and continues receiving the uplink control packet until the end of the cyclic redundancy check (CRC). If the CRC succeeds, indicating that there was no collision (or corruption by interference), the base station broadcasts a BUSY flag with the address of the successful mobile terminal. The mobile, upon receiving the BUSY flag, continues with the information packet transmission. Collisions will occur if two or more mobiles begin transmission at the start of the same mini-slot. The vulnerable interval for collisions is limited to the difference in propagation delay between the individual mobiles and the base station. In the case of a collision, the CRC will fail and the base station broadcasts a STOP flag in the next downlink control channel. Colliding mobiles reschedule the packet transmission after a random interval.

4.3.10 Block Reservation Multiple Access (BRMA)

Similar to PRMA++, Block Reservation Multiple Access (BRMA) [DUNL93] protocol was another proposal for the RACE ATDMA project. BRMA is claimed to be a development of PRMA for cellular systems. This protocol is purely burst-switched, in which, both packet data and packet voice terminals reserve resources in their activity periods. The access mechanism used is random ALOHA. Similar to PRMA, contention attempts (as well as call set-up and handover initialisations) are carried out in the idle time slots. As in FRMA, the reservation status of the time slots are broadcast once at the end of ever transmission frame.

4.4 Random Access Mechanisms

Almost any trunked wireless communication systems requires some form of random access protocols. For example, when a terminal powers on, or intends to set up a call, it needs to inform the base station of its requests (e.g., registration, call set-up, and so forth). Since the base has no contact with the terminal, it could not assign any channel to the terminal.

The random access mechanism was pioneered by the well-known ALOHA protocol [ABRA73]. It was developed in 1970 to allow communications between remote terminals and the central computer at University of Hawaii. In random access
mechanisms, upon generation of a packet a terminal accesses the packet random access channel (PRACH) at random basis regardless of transmission status of other terminals. Thus, if more than one packet is transmitted simultaneously, packets collision occurs and the colliding packets must be retransmitted (see Figure 4.5). In order to resolve and avoid repeated collisions, each retransmission will be attempted after a random time.

![Figure 4.5 Contention Associated with Random Access Protocols](image)

Roberts [ROBE70] developed a protocol with minimal co-ordination between the terminals called Slotted ALOHA. Under this protocol, time is divided into fixed length time slots, and each terminal will delay the transmission of its packet until the next slot boundary. In this case, the throughput will be increased by a factor of two. The priorities in random access mechanisms can be assigned based on PRACHs’ configurations and back-off mechanisms for different service classes.

### 4.4.1 Performance Analysis

In this section, an analysis of random access protocols for a given normalised offered load of $G_o$ is provided. In the ensuing analysis, we should distinguish packets transmitting in a given slot as being either newly generated or ones that have in the past collided with other packets. Therefore, we require the following two additional definitions [KLEI76]:

---

59
\[ q = Pr[\text{newly generated packet is successfully transmitted}] \quad (4.1) \]

\[ q_t = Pr[\text{previously blocked packet is successfully transmitted}] \quad (4.2) \]

It can be shown that the throughput, \( S_a \), \( q \), and \( q_t \) have the following mathematical relationships:

\[ S_a = G_a \frac{q_t}{q_t + 1 - q} \quad (4.3) \]

\[ q = \left[ e^{-G_a P} + G_a P e^{-G_a} \right]^{1/P} e^{-S_a} \quad (4.4) \]

\[ q_t = \left[ \frac{e^{-G_a P} - e^{-G_a}}{1 - e^{-G_a}} \right] \left[ e^{-G_a P} + G_a P e^{-G_a} \right]^{1/P-1} e^{-S_a} \quad (4.5) \]

\( q_t \) can further be approximated by [ROBE73] as

\[ q_t \equiv (1 - P)e^{-G_a} \quad (4.6) \]

where \( S_a \) is the throughout of the random access channel and \( P \) presents permission probability. Equations (4.3)-(4.5) form a set of non-linear simultaneous equations for \( S_a \), \( q \), and \( q_t \) that must be solved to obtain an explicit expression for \( S_a \) in terms of the system parameters \( G_a \) and \( P \). In general, this cannot be accomplished. Computer mathematical tools in this case should be used. We note that as \( P \) approaches zero, these three equations reduce simply to

\[ \lim_{P \to 0} \frac{S_a}{G_a} = \lim_{P \to 0} q = \lim_{P \to 0} q_t = e^{-G_a} \quad (4.7) \]

This results in the famous throughput equation for slotted ALOHA, i.e.

\[ S_a = G_a e^{-G_a} \quad \text{for} \quad P \to 0 \quad (4.8) \]

The average access delay is given by [KLEI76]

\[ D_c = T_c + D_f + \frac{1-q}{q_t} \left( D_f + T_c + \frac{1-P}{2P} T_c \right) \quad (4.9) \]
where $D_c$ is the average contention delay, $D_f$ is the average round-trip propagation delay, and $T_c$ is period of PRACH occurrence. Equation (4.9) shows that as the load increases the average access delay increases. Therefore, it should be noted that maximum throughput does not necessarily mean the lowest access delay. In general, throughput is a parameter of concern for the operator. For example, a throughput of 30% for an operator would mean that their radio resources are used properly in 30 percent of time. However, the end-user is not interested in the channel utilisation. Rather, the end-user is more interested to obtain its required QoS. For most services, delay is a drawback.

### 4.4.2 Non-Prioritised Schemes

As mentioned earlier, in random access mechanisms, one way of defining priorities are through transmission permission probabilities. In order to illuminate priorities between service types, all terminals are required to use the same permission probability, i.e.

$$P_i = P_j, \quad \forall i, j \quad (4.10)$$

where $P_i$ is the permission probability of service class $i$. Depending on the signalling traffic load the number of random access channels could vary. When the request for random access signalling is not high fewer PRACH channels are needed (see Figure 4.6).

![Random Access Protocol for Low Population, No-Priority Multimedia Systems](image)

It should be noted that the above PRACHs’ structure (and the following figures in this report) does not necessarily translate into TDMA. Signalling channels in
communication systems are usually periodic. Therefore, the figure above is a symbolical representation of periodic signalling channels rather than TDMA channels. This period is denoted by $T_c$. The symbol “G” determines that the type of the corresponding signalling channel is “General” and it can be used by any service type. In Figure 4.6, one channel unit is assigned to all service classes.

When the overall population of terminals is not high, signalling, request, and acknowledgement messages will not cause congestion. Therefore, the above allocation is sufficient as long as the calling population is low. However, as soon as the signalling request traffic builds up, the random access mechanism reaches its breaking point at which the frequency of collisions becomes so high that no contending terminal can succeed to access. In order to resolve this situation, either the (average) permission probability should be reduced or more PRACHs must be used. Figure 4.7 shows an example PRACHs’ configuration of this type. In this case, 5 different service classes equally share a PRACH which frequents twice faster than that of Figure 4.7. It should be noted that the permission probability is required to be optimised for each of the two aforementioned cases separately for best performance.

![Figure 4.7 Random Access Protocol for High Population, No-Priority Multimedia Systems](image)

Since terminals contend with the same parameters, therefore, the delay profiles experienced by different service classes are identical. The numerical evaluations of this scheme are performed and discussed in the next chapter.
4.4.3 Prioritised Schemes

Different QoS requirements in mixed services environment result in prioritising one service over the other. Priorities in random access mechanisms can be achieved through either PRACH channels' assignment or permission probability schemes. We number priority classes and assume that the smaller the index number the higher the priority. Therefore, the service class 1 has the highest priority. Then, priority through permission probability is assigned by

\[ P_i > P_j \quad \text{for} \quad i < j \quad (4.11) \]

where \( P_i \) is the permission probability for the service class \( i \). This implies that the higher priority services transmit in PRACHs with higher probability in order to capture channels in their favour. Figure 4.8 shows such a scenario with one general PRACH.

![Diagram showing the prioritisation of different service classes through permission probability.](image)

**Figure 4.8 Random Access Protocol Low Population, Prioritised Multimedia Systems**

Again as the aggregate load on PRACHs increases, there will be need for more channels.
However, if PRACHs are assigned as “General”, the services will not necessarily enjoy the priority schedule. For example, imagine that the signalling frequency from a service class is quite high even though the priority of that particular service is not so high. In this case, most of the channels are tried by that service class. Considering that the higher priority classes use larger permission probability, the collision rate becomes astronomical and hence the breaking point reaches. In such cases, in order to implement a real prioritised scheme, PRACHs could have “dedicated” form where each service class uses its own PRACH(s) as shown in Figure 4.9. Again as the population and the traffic load of a service class increases, more PRACHs will be allocated to that service (see Figure 4.10).

Figure 4.9 Random Access Protocol for High Population, Prioritised Multimedia Systems

Figure 4.10 Random Access Protocol for Higher Population, Prioritised Multimedia Systems
4.4.4 Adaptive Schemes

Up to here, we have assumed that boundaries between PRACHs are fixed. Such configurations are effective as long as the signalling traffic arrives with a fixed average rate. However, in reality the arrival rate at a communications system varies with respect to the time and the environment. For example, the traffic load of a network increases during the working hour and reduces afterwards. Besides, there are times (in a day) when the request for one service is higher than the other. In these cases, fixed allocation of PRACHs to services is wasteful and may result in excessive access delay and congestion. In order to tackle this problem, an adaptive random access mechanism is required.

Figure 4.11 demonstrates the general form of an adaptive scheme in which services are temporarily allocated with some PRACHs and they contend with a variable permission probability.

The permission probability varies in terms of the load and other parameters to be discussed in Section 4.5. The number of allocated channels to each service depends on the activity factor of sources, arrival rate and required signalling load of that service class. Priorities are given by the number PRACH channels and the permission probability scheme.
4.4.5 Dynamic Retransmission Control Schemes

The performance of the random access protocols is adversely affected by collisions caused by contention. Consequently, any method that reduces the collisions increases the performance. We here examine mechanisms in which contending terminals transmit with variable transmission permission probabilities so as to maximise the probability of success.

Dynamic retransmission control schemes can be implemented either in a form of retransmission permission probability or back-off time delay. In retransmission permission schemes, contending terminals (re)transmit with a probability of \( P \). This probability could be different for different service classes (as was mentioned above) and also be dynamically adopted according to the collision experiences. In back-off schemes, retransmission attempts are scheduled with a random delay \( d_r \) (usually assumed exponentially distributed with average \( \bar{d}_r = 1/\lambda_r \)). A general back-off policy is defined as a function \( \lambda_r(a) \), which determines the effective retransmission rate to be used on the \( a \)th retransmission attempt. It is common to consider monotone decreasing functions of the form \( \lambda_r(a) \), \( a=1, 2, ..., A \) where \( A \) is a specified integer, i.e., after \( A \) attempts a constant value of \( \lambda_r(A) \) is used for all successive trails [RAYC90].

In order to stabilise the random access sub-system, there should be maximum allowed number of re-transmission attempts \( (A_{\text{max}}) \) after which the terminal will be blocked. This method is widely used in wireless communication systems such as GSM that sets \( A_{\text{max}}=7 \) [MOUL92].

4.4.5.1 Optimum Variable Permission Probability

The probability of success, \( S(P) \), at a given time, with \( C \) terminals in the contention mode, can be calculated as:

\[
S(P) = P(1 - P)^{C-1}
\]  

(4.12)

where \( P \) is the transmitting permission probability of a contending terminal. It can be easily shown that in order to maximise \( S(P) \) for a given \( C \), the permission probability should be chosen as [TAAG96a, TAAG96b]
where \( P_{opt} \) is the optimum variable permission probability. We call such a scheme in which the permission probability traces the contending traffic as expressed by (4.3), *Optimum Random Access Mechanism*. It goes without saying that such a scheme is impossible to realise since the base station would need to know the number of contending wireless terminals at any given time. The main point is that choosing permission probability as in (4.3) results in the optimum dynamic procedure on permission probability that one can achieve.

### 4.4.5.2 Two-Level Variable Permission Probability Scheme

Experimentally, it can be shown that with a high probability, a terminal does not attempt to access the channel even though there is no one else at the contention phase [TAAG96]. This statistic leads to designing a variable permission probability algorithm in which any terminal transmits the very first request attempt of a newly generated message with a different probability (say \( P_0 \)) from re-transmission probability (say \( P_r \)) so that \( P_0 > P_r \) [TAAG96a]. It will be shown in the evaluation section that such a simple scheme can obtain very close performance to that of the ideal case for low traffic loads.

### 4.4.5.3 Bayesian Broadcast Control Scheme

In the Bayesian broadcast control scheme, the goal is to estimate the instantaneous number of contending (backlogged) terminals, \( C \). Then, this estimate is used to calculate and to periodically broadcast the transmission probability to different classes of backlogged terminals. The transmission probability is computed as

\[
P_r = \frac{G_0}{b_k}
\]

(4.14)

where \( P_r \) represents the permission probability, \( G=G_0=1 \) is the offered traffic load corresponding to the maximum achievable throughput of S-ALOHA, i.e. \( \eta=1/e \) and \( b_k \) is an estimate of the number of backlogged users at the \( k \)-th step defined by
where $\bar{\lambda}$ is either the estimated instantaneous arrival rate or a fixed value of $1/e$. As discussed earlier, the estimation of $\bar{\lambda}$ has been unsuccessful [TAAG96a, BISD96, DIEZ98]. Therefore, for this variable transmission probability scheme, $1/e$ is usually used. The current value of $P_r$ calculated with the updated backlog estimation is assumed to be broadcast by the base station.

### 4.4.5.4 Exponential Back-off Scheme

In this scenario, newly arrived contending terminals transmit with an initial transmission rate of $\lambda_r(I)$, and with every unsuccessful attempt, the retransmission probability decreases with a back-off rate $p$ up to the $A$-th retransmission attempt. Thus, the retransmission probability for the $a$-th transmission attempt can be written as [RAYC90]:

$$\lambda_r(a) = p^a \lambda_r(I) \quad a = 1, 2, ..., A \text{ with } p \leq 1$$

(4.16)

where $\lambda_r(I)$ is the initial value of retransmission rate and $p$ is the back-off rate. Prioritisation can be achieved by selecting different values for the initial retransmission and the back-off rates for each priority class.

### 4.4.5.5 Linear Back-off Scheme

In this scheme the retransmission rate is reduced linearly as follows [RAYC90]:

$$\lambda_r(a) = \lambda_r(I)(1 - p(a - I)) \quad a \leq A \text{ with } p(A - I) \leq 1$$

(4.17)

### 4.4.5.6 $\mu$-Law Back-off Scheme

In this scheme, the retransmission rate changes as follows [RAYC90]:
4 Packet Multiple Access Protocols (Theory)

\[
\lambda_r(a) = \lambda_r(I) \left( \frac{\ln(I + \mu[I - p(a - 1)])}{\ln(I + \mu)} \right) \quad \text{with} \quad p(A - 1) \leq 1 \quad (4.18)
\]

where \( \mu \) is the curvature control parameter.

4.4.5.7 Step-Function Back-off Scheme

The retransmission rate is defined for this scheme as follows [RAYC90]:

\[
\lambda_r(a) = \begin{cases} 
\lambda_r(I) & \text{for } a = 1, 2, \ldots, A - 1 \\
\lambda_r(A) & \text{for } a = A 
\end{cases} \quad (4.19)
\]

4.4.5.8 Stack Based Scheme

The prioritised stack random access algorithm is a stack based scheme [BUOT96]. In this scheme, each terminal counts up its own number of transmission attempts. Upon arrival of a message, the counter \( c \) is reset to \( I \) and the transmission is performed in the very next available channel. If the attempt is unsuccessful (i.e. collided), the counter is updated as follows, depending on the priority class \( i \):

\[
i = 1 \rightarrow c = \begin{cases} 
1 & \text{with probability } \sigma \\
2 & \text{with probability } 1 - \sigma 
\end{cases} \quad (4.20)
\]

\[
i > 1 \rightarrow c = \begin{cases} 
1 & \text{with probability } \beta, \\
2k & \text{with probability } (1 - \beta) \sigma, \\
2k - 1 & \text{with probability } (1 - \beta)(1 - \varphi). 
\end{cases} \quad (4.21)
\]

As long as \( c > 1 \), whenever a collision takes place, \( c \) is incremented by \( i \), otherwise, \( c \) decrements by \( I \). If \( c = 1 \), a retransmission attempt is made and the counter is handled in the way described for the first transmission.

4.4.5.9 Comparison of Retransmission Schemes

In general, using dynamic retransmission schemes, the probability of a collision is reduced in comparison over the non-adaptive retransmission schemes; maximum
throughputs in range of 0.4 to 0.6 can be achieved as compared to 0.37 for slotted ALOHA [RAYC90, CAPE79, GEOR82, LIU83, RAYC85, TAAG96a].

The evaluations presented in [RAYC90] show that although retransmission back-off does not automatically result in performance stability, appropriate back-off policies can provide quite stable characteristics. Specially, it has been demonstrated that all the back-off functions considered above with suitable chosen parameters, offer substantial performance advantages over a non-adaptive retransmission policy [RAYC90]. It is observed that the well-known exponential back-off strategy, while not very far from optimum, is outperformed by alternative functions such as the $\mu$-law and step-function, which tend to delay the onset of retransmission rate reduction until several collisions are experienced [RAYC90]. It will also be shown in Chapter 5 that a simple two-level permission probability could result in very near-optimum performance.

4.5 General Analysis for Mixed Services

4.5.1 Formulation

In general, packet-based protocols differ from each other by the access mechanism (procedure) and design parameters. Hence, a reservation protocol can be defined in terms of

1) The "Access Mechanism" consisting of signalling channels [DEVI93, MITR90], feedback packets, and retransmission strategies.

2) A set of "Protocol Parameters" (PP) representing the channel parameters, and characteristics of the transmission frame.

The access mechanism could be random (see Section 4.4), polling-based, hybrid random / polling, or collision detection access mechanisms. Here, we consider random access protocols for the access mechanism. This class of packet-based protocols with random access mechanism is called contention-based access protocols [JANG94, TAN96, GOOD91, TAAG95, TAAG97a].
4.5.1.1 Protocol Parameters

The protocol parameters (PP) consist of a set of parameters related to the frame structure, channel capacity and the radio channel characteristics. Such a set, \( S_P \), can be expressed with

\[
S_P = \{ R_c, N, H, B, T_f, T_s, T_c, D_f, CDF(D_f), PhyChannel \}
\]

Definitions these parameters are listed in Table 4.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_c )</td>
<td>Total Channel Bit Rate</td>
</tr>
<tr>
<td>( H )</td>
<td>Overhead Bits (^1)</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of Channels in a Frame</td>
</tr>
<tr>
<td>( B )</td>
<td>Packet Payload (in bits)</td>
</tr>
<tr>
<td>( T_f )</td>
<td>Transmission Frame Duration</td>
</tr>
<tr>
<td>( T_s )</td>
<td>Packet Transmission Time</td>
</tr>
<tr>
<td>( T_c )</td>
<td>Period of PRACHs</td>
</tr>
<tr>
<td>( D_f )</td>
<td>Round-Trip Transmission Delay</td>
</tr>
<tr>
<td>PhyChannel</td>
<td>Channel Model Characterising erroneous Radio Channel</td>
</tr>
</tbody>
</table>

4.5.1.2 Performance Metrics

Given the protocol parameters and contention mechanism, system performance metrics can be obtained, through simulation or mathematical analysis by inputting the traffic characteristics. The traffic characteristics of the bursty service class \( i \), \( S_f^{(i)} \), can be formulated as

\[
S_f^{(i)} = \{ R_c^{(i)}, M^{(i)}, R^{(i)}, \Gamma^{(i)}, CDF(T_{on}^{(i)}), CDF(T_{off}^{(i)}), T_{on}^{(i)}, T_{off}^{(i)}, T_h^{(i)}, D_{max}^{(i)}, A_{max}^{(i)} \}
\]

\(^1\) The overhead bits include the training sequence, head and tail bits, guard time bits, in-band signalling flags, terminal’s virtual circuit (or temporary) ID, and all other non-payload bits in a packet.
where \( r \) is the total number of packet service types. Table 4.2 gives a description of such parameters. And the aggregate traffic characteristics, \( S_T \), is defined as

\[
S_T = S_T^{(1)} \cup S_T^{(2)} \cup ... \cup S_T^{(r)}
\]

where \( r \) is the total number of packet service types.

### Table 4.2 Definitions of Typical Traffic Parameters of Service Class \( i (S_T^{(i)}) \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_i )</td>
<td>Population of Service Class ( i ) Terminals</td>
</tr>
<tr>
<td>( R_s^{(i)} )</td>
<td>Average Rate or Source ( i )th (channel encoding included)</td>
</tr>
<tr>
<td>( k^{(i)} )</td>
<td>No. Required Channel Resources of Service Class ( i )</td>
</tr>
<tr>
<td>( T_{on}^{(i)} )</td>
<td>Mean Message Duration (in time)</td>
</tr>
<tr>
<td>( X^{(i)} )</td>
<td>Mean Message Duration (in bytes)</td>
</tr>
<tr>
<td>( T_{off}^{(i)} )</td>
<td>Mean Silence Duration</td>
</tr>
<tr>
<td>( t_{on}^{(i)} )</td>
<td>Minimum Message Duration (in time)</td>
</tr>
<tr>
<td>( \overline{X}^{(i)} )</td>
<td>Average Message Duration (in bytes)</td>
</tr>
<tr>
<td>( t_{off}^{(i)} )</td>
<td>Minimum Silence Duration</td>
</tr>
<tr>
<td>( T_h^{(i)} )</td>
<td>Hangover Duration of Service Class ( i )</td>
</tr>
<tr>
<td>( \Gamma^{(i)} )</td>
<td>Activity Factor of Service Class ( i )</td>
</tr>
<tr>
<td>( A_{max}^{(i)} )</td>
<td>Maximum Retransmission Attempts</td>
</tr>
<tr>
<td>( D_{max}^{(i)} )</td>
<td>Packet Delay Limit of Service Class ( i )</td>
</tr>
</tbody>
</table>

A typical set of performance metrics is given as

\[
S_M^{(i)} = \{D_e^{(i)}, D_q^{(i)}, D_a^{(i)}, S_a^{(i)}, s_a^{(i)}, S_t^{(i)}, s_t^{(i)}, P_{drop}^{(i)}, C^{(i)}, L_q^{(i)}, CDFs\}
\]

and described in Table 4.3.
Table 4.3 Definitions of Performance Metrics for Service Class $i$

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_c^{(i)}$</td>
<td>Average Contention Delay of Service Class $i$</td>
</tr>
<tr>
<td>$D_q^{(i)}$</td>
<td>Average Queueing Delay of Service Class $i$</td>
</tr>
<tr>
<td>$D_a^{(i)}$</td>
<td>Average Access Delay of Service Class $i$</td>
</tr>
<tr>
<td>$D_t^{(i)}$</td>
<td>Average Transmission Delay of Service Class $i$</td>
</tr>
<tr>
<td>$D_r^{(i)}$</td>
<td>Average Retransmission Delay of Erroneous Packets</td>
</tr>
<tr>
<td>$s_t^{(i)}$</td>
<td>Raw Traffic Throughput of Service Class $i$ (in Kbytes/s)</td>
</tr>
<tr>
<td>$s_a^{(i)}$</td>
<td>Raw Access Throughput of Service Class $i$ (in Kbytes/s)</td>
</tr>
<tr>
<td>$S_t^{(i)}$</td>
<td>Normalised Traffic Throughput of Service Class $i$</td>
</tr>
<tr>
<td>$S_a^{(i)}$</td>
<td>Normalised Access Throughput of Service Class $i$</td>
</tr>
<tr>
<td>$P_{drop}^{(i)}$</td>
<td>Packet Dropping Probability of Service Class $i$</td>
</tr>
<tr>
<td>$C^{(i)}$</td>
<td>Subsystem Capacity of Class $i$</td>
</tr>
<tr>
<td>$L_q^{(i)}$</td>
<td>Queue Length of Service Class $i$</td>
</tr>
</tbody>
</table>

The aggregate metrics are

$$S_M = \{D_c, D_q, D_a, S_t, S_a, P_{drop}, C, L_q, CDFs\}$$

Aggregate (average) for each of the performance metrics is the weighted summation of all the service types or

$$Aggragate(Y) = \sum_{i=1}^{r} \alpha_i Y^{(i)} \tag{4.22}$$

where $Y$ is an arbitrary metric and $\alpha_i$ is the weight parameter of service class $i$ which a function of source activity and traffic arrival pattern [KLEI76, GROS85].

The multiplexing gain or sub-system capacity of service class $i$, $C^{(i)}$ is defined as [TAAG97a]
where $\max(M^{(i)})$ is the maximum number of simultaneous terminals of class $i$ while maintaining the required QoS of other in-system terminals. For example, for voice terminals the QoS of concern is the packet dropping ratio, $P_{\text{drop}}$, which should be below the 1% limit [GOOD89]. Mostly the packet dropping ratio has been considered the limiting factor of real-time packet communications, other measuring factors such as mean access delay or access delay distribution have also been utilised in the literature [DEV193, MITR90].

One of the main evaluating factor of various protocols is the throughput defined as [NAND90]

$$S = \frac{\text{Average No. Occupied Logical Channels}}{\text{Total No. Logical Channels}}$$

(4.24)

As explained in Section 4.3, PRACHs and traffic channels could share the same physical resources (as in PRMA) or separate resources (such as PRMA++ and GPRS MAC protocols), in which case two throughput parameters should be considered. We denote the normalised throughput of the traffic channels with $S_t$ and that of the access channels with $S_a$. Throughputs could also be expressed in data amounts sent successfully in the time unit (Kbytes/s). We denote such parameters with $s_t$ and $s_a$, respectively. The average normalised and raw throughputs of the system is calculated by

$$s_t = \sum_{i=1}^{r} S_t^{(i)}$$

(4.25)

$$S_t = \sum_{i=1}^{r} S_t^{(i)}$$

(4.26)

The raw througught and the normalised throughput are mathematically related as
Another important metric for the end-user is the average burst transmission time, $T_t^{(i)}$, which is the sum of the delay components experienced by the burst or

$$T_t^{(i)} = D_c^{(i)} + D_q^{(i)} + D_i^{(i)} + D_r^{(i)}$$  \hspace{1cm} (4.28)

Given the average number of attempts per burst, $A^{(i)}$, and the average contention delay, $D_c^{(i)}$, are related by

$$D_c^{(i)} = A^{(i)} T_c$$  \hspace{1cm} (4.29)

where $T_c$ is the occurrence period of the PRACH. Recalling Little's formula, the queueing delay is calculated as [KLEI75, GROS85]

$$D_q^{(i)} = \frac{L_q^{(i)}}{\lambda^{(i)} M^{(i)} (1 - \Gamma^{(i)})}$$  \hspace{1cm} (4.30)

The transmission delay of the service class $i$ is calculated by

$$D_t^{(i)} = \frac{T_f \bar{X}^{(i)}}{BR^{(i)}}$$  \hspace{1cm} (4.31)

### 4.5.2 Equivalent Channel Load for Mixed Services

Mixed services scenarios could become quite complicated when dealing with large number of service types. Such complications will impose hardship on both analytical and simulation studies. Here, we define two equivalent offered load concepts; the load imposed on PRACHs (access load) and the load on the traffic channels.

#### 4.5.2.1 Offered Access Load

The offered access load is defined as the average number of access arrivals in a PRACH period. Recalling that terminals access the PRACH only at arrival of a burst and assuming that the arrival rate of each user of the service class $i$ is $\lambda^{(i)}$, the average offered access load is calculated as

$$G_a^{(i)} = E(\lambda^{(i)}) T_c$$  \hspace{1cm} (4.34)
where \( E(\lambda^{(i)}) \) is the aggregate access arrival rate for service class \( i \). The total access arrival is the sum of access arrivals from different service classes or

\[
G_a = \sum_{i=1}^{r} G_a^{(i)}
\]  

(4.35)

If we assumed that silence periods of sources follow negative exponential, then the calculation of \( E(\lambda^{(i)}) \) would be easy. However, realistic service models presented in Chapter 3 hardly inherit memoryless properties of Poisson distributions. Here, we show that it is still possible to approximate the arrival distribution with Poisson if the time spent in the silence \( (T_{off}) \) is considerably larger than a packet transmission time \( (T_s) \). Figure 4.12 demonstrates state transition of an ON/OFF source (repeat from Chapter 3). If \( T_{off} >> T_s \) then obviously the departure probability from the OFF state occurring in a \( T_s \) duration is very small.

![Figure 4.12 State Transitions of a ON/OFF Source](image)

Recalling

\[
e^{-\alpha} = 1 - \alpha \quad \alpha << 1
\]  

(4.36)

then

\[
\sigma = 1 - (1 - \sigma) = 1 - e^{-T_s/T_{off}} \quad T_{off} >> T_s
\]  

(4.37)

where Equation (4.37) proves the memoryless properties of source with \( T_{off} >> T_s \). Having arrived at this point, we can now use the approximation (C.15 in Appendix C) thus,

\[
G_a^{(i)} = M^{(i)} (1 - \Gamma^{(i)}) \lambda^{(i)} T_c
\]  

(4.38)
4.5.2.2 Offered Traffic Load

The offered traffic load is defined as the average arrived traffic load in time unit. Hence, according to this definition the offered traffic load generated by service class \( i \) is calculated as

\[
g_t^{(i)} = E(\lambda^{(i)}) X^{(i)}
\]  
(4.39)

where \( g_t^{(i)} \) is the traffic load and \( X^{(i)} \) is the average burst size (in bytes) of service class \( I \). Again, the raw traffic load is calculated as

\[
g_t = \sum_{i=1}^{r} g_t^{(i)}
\]  
(4.40)

For stability analysis and general view, we are more interested in the normalised traffic load which is the average amount of packet arrival with respect to the channel capacity. This normalisation leads to

\[
G_t^{(i)} = E(\lambda^{(i)}) \frac{X^{(i)}}{B} T_s
\]  
(4.41)

where \( G_t^{(i)} \) is the normalised traffic load generated by the service class \( i \). Again, the aggregation results in

\[
G_t = \sum_{i=1}^{r} G_t^{(i)}
\]  
(4.42)

where \( G_t \) is the total normalised traffic load.

4.5.3 Packet Dropping and its Impacts on Performance

The number of dropped packets from the front-end of a burst experiencing \( D_a (> D_{max}) \) delay is [TAAG95]

\[
n_d = \left\lfloor \frac{D_a - D_{max}}{T_f} \right\rfloor
\]  
(4.43)

77
where \( n_d \) presents the number of dropped packets. Similarly, the number of packets, \( n_p \), in a burst with duration of \( t \) seconds is

\[
n_p = \left\lfloor \frac{t}{T_f} \right\rfloor
\]

(4.44)

4.5.3.1 Terminal-Level Packet Dropping Probability

Using (4.40) and (4.41), the packet dropping probability for a message with length of \( t \) seconds and experienced delay of \( D_a \) seconds is given as

\[
P_{\text{drop}}(D_a, t) = f(D_a, t) \frac{D_a - D_{\text{max}}}{T_f} \Delta D_a \Delta t = f(D_a, t) \frac{D_a - D_{\text{max}}}{t} \Delta D_a \Delta t
\]

(4.45)

where \( f(D_a, t) \) is the joint probability density function of delay \( D_a \) and message duration of \( T_{\text{on}} \) seconds with average message duration of \( T_{\text{on}} \) seconds.

Since all the sources are assumed to be independent, the suffered delay and message duration are statistically independent, then (4.45) can be re-written as [PEEB93]

\[
P_{\text{drop}}(D_a, t) = f_{D_a}(D_a) f_{\text{i}}(t) \frac{D_a - D_{\text{max}}}{t} \Delta D_a \Delta t
\]

(4.46)

where \( f_{D_a}(D) \) is the PDF of delay and \( f_i(t) \) is the PDF of message duration. Finally the average packet dropping probability is obtained as

\[
P_{\text{drop}} = \int_{t=T_f}^{D_{\text{max}}+T_{\text{on}}} \int_{D_a=D_{\text{max}}+T_f}^{D_{\text{max}}+T_{\text{on}}} f_{D_a}(D_a) f_{\text{i}}(t) \frac{D_a - D_{\text{max}}}{t} dD_a dt
\]

(4.47)
4.5.3.2 System-Level Packet Dropping Probability

Even though the average packet dropping aforementioned is very accurate, however, most of the time we are interested in the average packet dropping from system's point of view, which is defined and obtained as follows:

\[ P_{\text{drop}} = \frac{E(n_d)}{E(n_p)} = \frac{\int_{D=D_{\text{max}}+T_{on}}^{D_{\text{max}}+T_{on}} f_{D_a}(D_a) \frac{D_a - D_{\text{max}}}{T_f} dD_a}{\int_{t=0}^{T_f} f_t(t) dt} \]  \hspace{1cm} (4.48)

after summarising

\[ P_{\text{drop}} = \int_{D=D_{\text{max}}+T_f}^{D_{\text{max}}+T_{on}} f_{D_a}(D_a) \frac{D_a - D_{\text{max}}}{T_{on}} dD_a \]  \hspace{1cm} (4.49)

4.5.3.3 Service Time Reduction Due to Packet Dropping

Although packet dropping decreases the received voice quality, it also reduces the message lengths which results in lower congestion. Therefore, to take this traffic reduction into account, it is needed to make some modification. We can realistically assume that the reduction does not change the distribution profile of the message duration, however, it modifies the average duration as

\[ T_{on}^{v+1} = E[(1 - P_{\text{drop}}^v)t^v] \]
\[ = (1 - P_{\text{drop}}^v)E(t^v) \]
\[ = (1 - P_{\text{drop}}^v)T_{on}^v \] \hspace{1cm} (4.50)

and, \( P_{\text{drop}}^0 = 0 \)

Hence, in order to obtain the exact value of the performance metrics cited above (throughput, average delay, packet dropping probability, capacity), we need to calculate them by iteration using (4.50).
4.5.4 Mixed Services Analysis Using Aggregate Source Modelling

Analytical study of multimedia systems is very complicated even for simple cases of two or three service types using the conventional mathematical approach, i.e. multidimensional Markov analysis [PAVL94]. As an alternative to multidimensional Markov analysis, an approximate analytical model called “Equilibrium Point Analysis” (EPA) has been proposed [FUKU83, TASA86]. By utilising the EPA, we can analyse multidimensional Markov chains since it is not necessary for the EPA to calculate state transition probabilities due to the assumption that the system is always at an equilibrium point [FUKU83]. Analytical representation of performance measures of these systems could be very useful in network dimensioning, Call Admission Control (CAC), and optimal design of networks. Most traffic models consider infinite population for users where the nature of the arrival process depends only in a negligible way upon the number of customers already in the system [KLEI76]. However, the real world involves no infinite population, e.g. numbers of cellular mobile subscribers and computer consoles in an organisation are evidently finite. Therefore, analysis of multi-service finite queueing system could have numerous applications for the future multimedia-capable communication networks. Furthermore, due to new trends towards use of variable source coding and support of different service classes, statistical multiplexing techniques could be another area of application of finite queue models [GOOD91, DEVI93, MAGL88, SEN89].

In this section, a novel traffic model has been proposed for finite population of on-off sources with different service-time and idle-time profiles. Unlike the classical approach, we try to avoid the multidimensional Markov chains. The proposed analytical model called “Aggregate Source Model” (ASM) aggregates heterogeneous traffic sources into one homogeneous prototype source model. Then, the problem will be confined to a monomedia system analysis/simulation. This model can be applied to analyse a wide range of finite-population resource-sharing (multiplexing) system provided that a few conditions satisfy. Applications of the ASM range from mixed services analysing of FCFS queues to evaluating reservation based protocols for mixed services. As will be shown, the random access protocols can also be modelled with queueing systems. The application of the ASM is further examined for non-Poisson inputs by inputting an Internet-type traffic (a mixture of WWW and FTP) to an FCFS
downlink protocol. The results show the close agreement between the ASM analysis and the mixed services simulation for FCFS queues as well as a random reservation protocol.

4.5.4.1 Aggregate Source Modelling

We consider a system in which there is a finite number of different service types, \( r \), and where \( M_i \) identical independent sources constitute service type \( i \) population. Inter-arrival times (time that a source spends outside the system) of a service class \( i \), form negative exponential distributions with mean rate \( \lambda_i \). In this study, we relax service (holding) times to follow any arbitrary distribution with known mean departure rates \( \mu_i \)'s. Arrivals emanated from sources of different classes enter a single queue governed by a non-preemptive non-prioritised discipline. Arrivals are then picked from the head of the queue and served by one of the \( N \) available servers (see Figure 4.13). In non-prioritised systems, first-order statistics, i.e. the average queueing delay \( D_q \), queue length \( L_q \) and number of users in the system \( L \) experienced by service classes are identical [KLEI76], or

\[
D_q^{(i)} = D_q^{FCFS} \quad \forall i \tag{4.51}
\]

\[
L_q^{(i)} = L_q^{FCFS} \quad \forall i \tag{4.52}
\]

\[
L^{(i)} = L^{FCFS} \quad \forall i \tag{4.53}
\]

where \( D_q^{(i)} \) represents the average queueing delay of service class \( i \).
The study of non-prioritised queueing systems is important because most priority systems are compared to the non-prioritised queueing system. From the conservation laws, it can be proven that as long as the queueing discipline selects arrivals independent of their service time measures (i.e., external priority), then the distribution of the number in the system (as well as the average waiting time) will be invariant to the order of the service. It should be noted that distribution of waiting time will indeed rely on the order in which service is given to arrivals [KLEI76]. Therefore, the non-prioritised system could be exploited for analysis of other priority queues. Such queueing system with \( r \) service types can be defined by three matrices of \( M, \lambda, \) and \( \mu \) representing the number of sources, arrival rates, and departure rates, respectively.

\[
M = \begin{pmatrix} M^{(1)} \\ M^{(2)} \\ \vdots \\ M^{(r)} \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda^{(1)} \\ \lambda^{(2)} \\ \vdots \\ \lambda^{(r)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \vdots \\ \mu^{(r)} \end{pmatrix}
\]  

(4.54)

The objective is to obtain a general source model, named "Representative Aggregate Source Model" (RASM), which represents the traffic generated by all the \( r \)
service classes. This representative model can be defined by three parameters of $M^{RASM}$, $\mu^{RASM}$, and $\lambda^{RASM}$. Figure 4.14 depicts the modelling approach.

Assuming that inter-arrival times follow Poisson distributions, we can aggregate the arriving traffic from class $i$ into one Poisson arrival process with rate of $\Lambda^{(i)}$ [KLEI75]

$$\Lambda^{(i)} = m^{(i)} \lambda^{(i)}$$  \hspace{1cm} (4.55)

where $m^{(i)}$ is instantaneous number of inactive sources of type $i$. Likewise, the aggregate arrival rate generated by all the service classes, $\Lambda$, is

$$\Lambda = \sum_{i=1}^{r} \Lambda^{(i)}$$  \hspace{1cm} (4.56)

It should be noted that the aggregate service distribution cannot be negative exponential even if service time distributions of all the classes are (it can be shown that the resultant distribution will be of type $r$-stage hyper-exponential, $H_r$ [KLEI75]). The average service time perceived by the servers, $l/\mu^{RASM}$, can be calculated as

$$\frac{l}{\mu^{RASM}} = \sum_{i=1}^{r} \frac{E(\Lambda^{(i)})}{\lambda^{(i)}} \frac{l}{\mu^{(i)}}$$  \hspace{1cm} (4.57)
where \( E(\Lambda^{(i)}) \) represents the average arrival rate of finite population of service type \( i \) and is computed and approximated as

\[
E(\Lambda^{(i)}) = \lambda^{(i)} E(m^{(i)}) = \lambda^{(i)} (M^{(i)} - L^{(i)}) \equiv M^{(i)} (1 - \Gamma^{(i)}) \lambda^{(i)}
\] (4.58)

and the summation

\[
\lambda = E\left( \sum_{i=1}^{r} \Lambda^{(i)} \right)
\] (4.59)

where \( \Gamma^{(i)} [= \lambda^{(i)}/(\lambda^{(i)}+\mu^{(i)})] \) is the activity factor. Observations show that the approximation of (4.58) is accurate for bursty sources with \((\lambda\mu<1)\) (see Appendix C, Section C.3) [TAAG98c]. In order to obtain \( M^{RASM} \) and \( \lambda^{RASM} \), for the Rasm, we can write

\[
\Lambda^{RASM} = m^{RASM} \lambda^{RASM}
\] (4.59)

The instantaneous arrival of the Rasm should be equal to that of the aggregate traffic, \( \Lambda^{RASM} \), that is

\[
\Lambda = \Lambda^{RASM}
\] (4.60)

Equation (4.58) should hold for any combination of \( m^{(i)} \)s. The maximum aggregate traffic load should correspond to the maximum arrival generated by the Rasm or

\[
\sum_{i=1}^{r} M^{(i)} \lambda^{(i)} = M^{RASM} \lambda^{RASM}
\] (4.61)

Similarly, by taking average from both sides of (4.60), thus

\[
\sum_{i=1}^{r} M^{(i)} (1 - \Gamma^{(i)}) \lambda^{(i)} = M^{RASM} (1 - \Gamma^{RASM}) \lambda^{RASM}
\] (4.62)

Solving the simultaneous equations of (4.61) and (4.62)
Equations (4.57), (4.63), and (4.64) are the parameters describing the representative source model.

In the above analytical, the following assumptions were made:

**Conditions (1)** - Negative exponential distributions for inter-arrival times of each source,

**Conditions (2)** - Insensitivity of system performance metrics (measures of effectiveness) to service distributions,

**Conditions (3)** - No priority of classes,

**Conditions (4)** - Equal resource reservations for service classes.

Having all these conditions satisfied, the ASM analysis can be applied to any finite population multimedia system as shown in Figure 4.15.

**Figure 4.15 Analysis of Multimedia Systems Using the ASM**

In this approach, the multimedia traffic streams generated by finite population of different service types are aggregated and represented by a single type service (RASM). Then this representative service type can be fed into the monomedia analysis (or
simulation) of the given queueing system. The end result is the average performance measures of the multimedia queueing system.

4.5.4.2 Mixed Services Analysis of Random Access Protocols

Random access protocols can also be viewed as a queueing system with a random selection queueing discipline. As demonstrated in Figure 4.16, arrivals from terminals are stacked in a "Random Access Queue" (RAQ). Every user in the RAQ attempts to access each of the available servers with a pre-defined permission probability. An awaiting terminal is selected from this queue if it is the only one accessing the server \( n \). In this case, the attempt is termed success, and the successful user occupies the server \( n \). If there is more than one user trying to access one of the servers, the attempt is flagged collision. In this case, the collided users stay in the queue and make more attempts for the rest of the available servers. The priority to different service classes are set by the permission probability \( P \); the larger the permission probability the higher the priority.

![Figure 4.16 A Finite Population Random Access Queueing System](image)

4.5.5 Stability and Saturation of System

Contention (random) based packet protocols could become unstable. This instability is due to the random access mechanism. As the offered access load \( (G_o) \) increases, the
transmitted packets keep colliding with each other and no packet is successfully transmitted. Therefore, a long term condition of low throughput and long delay may be experienced due to overload of the PRACH by retransmission traffic. For the frequently encountered scenario in which the PRACH is shared by a finite number of interactive stations, such instability condition can be avoided by selecting an appropriately low permission probability (or long back-off time), resulting in large average transmission delay. The stability can be improved by allocating more PRACHs and deploying adaptive and dynamic retransmission control policies (see Section 4.4).

Having derived the offered access load, $G_a$, we can easily detect the breaking point after which the random access mechanism starts to fail. As $G_a$ is defined as the average number of access request per PRACH, therefore, it is clear that if we have an aggregate access load of more than one ($G_a > 1$) then the PRACH will definitely not be able to carry the generated request traffic. The maximum number of successful attempts per PRACH is one and the excessive arrivals are bound to collide and pile up before the PRACH. This chain-effect results in the random access breaking point [CARL75, KLEI76].

There are cases where the random access mechanism is stable and functional however, the normalised traffic load, $G_t$, is more than unity. In this case, the request access from sources are successful, however, the generated traffic is more than the traffic channel capacity and the requests need to be queued up at the base. In this cases, if the calling population is infinite, then we face overflowing at the queue (buffer) of the base. However, if the total population is known and finite, the queue size can be set to the population size, in which case we refer to the system as saturated.

4.6 Statistical Upper-Bound of Packet Multiple Access Protocols

4.6.1 Description

Apart from radio channel impairments, the reduction in multiplexing gain of packet reservation-based protocols is mainly due to waste of resources at the access stage. The waste could be because of contention, polling and/or a mixture of those, depending on the access mechanism deployed [GOOD89, DEVI93]. These collisions waste available
channel resources and adversely affect the system performance because no terminal having packets, succeeds in gaining the free time slot.

In order to obtain the upper-bound, it should be assumed that both the base and terminals are aware of one another status at any time with no signalling delays involved. Since the base station has complete knowledge of all the terminals, it is then able to organise awaiting terminals in a desired order. Such ordering can be carried out in one of the following ways

1) The transmission permission probability \( P \) varies ideally with the contending population \( C \) by \( P = P_{\text{opt}} = 1/C \) to maximise the probability of success (see Section 4.4.5.1). It is also assumed that contentions occur in available traffic channels and no additional channel is assigned for access. This scheme results in the upper-bound on the random access mechanism. We refer to this idealised case as “Upper-Bound with Contention” [MERA85, TAAG96a];

2) Access is granted to one of accessing terminals at random. We call this scheme the “Perfect Capture Upper-Bound” [JANG94];

3) The accessing users are scheduled with a FCFS discipline. We refer to this perfect scheduling “Contention-Free Upper-Bound” [GROS85, TAAG95, TAAG97a]. When a channel becomes available, it is then granted by the base station to the terminal waiting on the top of the queue.

The option (1) can only be used for contention-based protocols, whereas, the options (2) and (3) can generally be used for all packet multiple access protocol. Using the options (1) or (2) largely minimises the damage caused in access procedure. However, the successful terminal is not necessarily the one which has waited the longest. Therefore, deploying the third option provides the most optimum upper bound.

4.6.2 Analysis of Contention-Free Upper Bound

In order to give priority to the voice traffic, we deploy a non-preemptive queueing discipline on top of the FCFS. A priority discipline is said to be non-preemptive if there is no interruption and the highest priority customer (e.g. RT services) just moves to the
head of the queue to wait its turn [GROS85]. Such priority discipline (shown in Figure 4.17) is known as the “Head-Of-the Queue” (HOQ) [KLEI76].

Figure 4.17 Head-Of-the-Queue (HOQ) Priority Queue

The general analytical analysis of the upper-bound protocol (i.e. general queuing discipline, general arrival distribution, general service distribution, various service classes) is almost impossible. Thus, we here consider a problem where there is a finite calling population of $M_b$ on-off bursty terminals and $M_p$ packet data terminals and sources of each class are homogenous. The priority is given to bursty traffic which implies that arrived bursty terminals overtake data packets in the queue.

4.6.2.1 Analysis of Burst-Switched Sub-System

Assuming that frame duration is short enough and knowing that a user with reservation, is severed for the whole frame duration once a frame, makes possible to consider traffic channels as a bunch of parallel servers [MITR90, TAAG95]. Since data terminals are served packet by packet and each packet takes only one frame duration of the servers’ time, the data traffic has no impacts on the performance metrics of the burst-switched sub-system by deploying a non-preemptive FCFS order. Consequently, the voice subsystem can be analysed independently and separately from the data sub-system [TAAG95]. The burst-switched sub-system can be modelled with an $M/G/N//M_b$ queuing system consisting of Poisson arrival rate of $\lambda_b (=1/T_{off})$, general departure with rate of $\mu_b (=1/T_{on})$, $N$ parallel servers, infinite storage and a finite population of $M_b$ bursty users [TAAG97a, TAAG95, GROS85, KLEI76]. Therefore, the system metrics be calculated as follows:
1) State Probabilities:

\[ P_{n,b}^{upper} = P_{n}^{FCFS} \quad (4.66) \]

2) Delay Distributions:

\[ F_{D,b}^{upper} (D) = F_{D}^{FCFS} (D) \quad (4.67) \]
\[ f_{D,b}^{upper} (D) = f_{D}^{FCFS} (D) \quad (4.68) \]

3) Number of Users in the System

\[ L_{b}^{upper} = L^{FCFS} \quad (4.69) \]

4) Average Talkspurt Access Delay

\[ D_{b}^{upper} = D_{q}^{FCFS} \quad (4.70) \]

5) Throughput:

\[ S_{b}^{upper} = S^{FCFS} \quad (4.71) \]

6) Packet Dropping Probability: Assuming that the burst (e.g. talkspurts) duration is negative exponentially distributed, i.e.

\[ f_{t} (t) = \frac{e^{-t/T_{on}}}{T_{on}} = \mu_{b} e^{-\mu_{b}t} \quad (4.72) \]

the packet dropping probability can be calculated by substituting (4.68) and (4.72) into (4.47) or (4.49) results in packet dropping probability function. Using the equation (4.49), the packet dropping probability can be simplified as
where,

\[ P_{\text{drop}}(t) = \frac{1}{N(M_p - L)} \sum_{n=0}^{M_p-1} \frac{P_n(M_p - n)}{(n-N)!} \left[ (\mu_b ND_{\text{max}} t^{n-N} - t^{n-N+1}) e^{-t} + (n-N+1) I_{n-N}(t) - \mu_b ND_{\text{max}} (n-N) I_{n-N+1}(t) \right] \]  

(4.74)

and,

\[ I_n(t) = \int_0^t e^{-u} u^n du = -t^n e^{-t} + n I_{n-1}(t) \]  

(4.74)

As pointed out before, as the packet dropping in terminals reduces message lengths, then to take this factor into account and calculate the exact values of performance metrics (state probabilities, average access delay, throughput, packet dropping probability) the iteration method is required. The performance metrics for a desired accuracy can be obtained by using (4.50) with a required number of iteration steps.

7) Burst-Switched Sub-system Stability Condition: As in this upper-bound system there is no contention and the bursty terminal population is assumed finite, the burst-switched sub-system will always be stable (see Section 4.5.6) [KLEI76].

### 4.6.2.2 Analysis of Packet-Switched Sub-System

We assume that the inter-arrival times (silent periods) have negative exponential distributions. The packet data services considered here follow PAPD model (see Section 3.5.2, Chapter 3), i.e. a data terminal is assumed to generate packets with a Poisson arrival. Data terminals do not reserve channels (packet-switched), and hence, the service time for a data source is fixed to one packet transmission duration in a frame \( T_f = 1/\mu_p \).

The average aggregate data packet arrival rate per frame duration, \( G_p \), is

\[ G_p = \frac{\text{Aggregate Data Bit Rate}}{\text{Payload Size}} = \frac{M_p R_p}{B} T_f \]  

(4.75)
1) Throughput: Since a user in the queue is assigned the channel as soon as a traffic channel becomes available, no user stays in the queue when a channel is available. Therefore, the throughput of packet-switched sub-system is calculated as [TAAG97a,TAAG95]

\[
S_{upper} = \frac{\text{No. generated data packets in a frame}}{\text{Frame}} = \frac{G_p}{N}
\]  

(4.76)

2) Average Data Packet Access Delay: From non-preemptive queueing systems, the average access delay of the class-two priority (here data traffic) is obtained using Cobham’s formula as [GROS85]

\[
D_{p upper} = \frac{W_0}{(1-\sigma_1)(1-\sigma_2)}
\]  

(4.77)

where \(D_{p upper}\) is the average access delay for data packets, \(W_0\) represents the average of the random duration of time that a channel becomes available when all channels are full, and the parameters \(\sigma_1\) and \(\sigma_2\) are

\[
\sigma_1 = \frac{E(\lambda_p)}{\mu_p} = \frac{(M_p - I_{upper})\lambda_p}{\mu_p}
\]  

(4.78)

\[
\sigma_2 = \sigma_1 + \frac{\lambda_p}{\mu_p}
\]  

(4.79)

where \(E(\lambda_p)\) is the average burst arrival rate of terminals not-in-the-system. Here, we need to make an estimation, as the mathematics for multi-priority multi-channels analysis becomes quite intractable [GROS85]. Since data packets stay in servers much less than bursts, we can realistically assume that \(W_0\) is mainly caused by the voice sources. Then

\[
W_0 = P(\text{all time slots busy})E(w_0|\text{all time slots busy})
\]  

(4.80)

or,

92
3) **Data Sub-System Stability Condition**: In order to maintain the system in equilibrium, the throughput increase due to data traffic cannot exceed the unused channel capacity, expressed mathematically

\[
W_0 = \left( \sum_{n \in N} p^{\text{upper}}_n \right) \left( \frac{1}{N\lambda_p} \right)
\]  

(4.81)

4.6.3 **Modifications for Multi-Level Activity Sources**

As was presented in Chapter 3, some bursty traffic sources such as WWW and F-VAD PV show more than one level of activity. We call such sources multi-level active sources. For such traffic sources, the resultant distribution of silence periods will no longer follow a negative distribution. Even if each silence component follows Poisson distribution, the aggregate distribution will be that of a hyper-exponential [KLEI75, STER95].

As proven in [BUND80], finite-source queues are independent of the arrival distribution as long as the service times have identical negative exponential distributions. Thus, the above analysis performed for bursty sources with one-level activity can still be applied to the multi-level activity sources case by substituting the average silent and active periods in the equations (see Appendix A, Section A.3).

4.7 **Performance Comparison Methods**

4.7.1 **Cross Comparison**

Variability in system parameters has a number of significant impacts on the overall system performance [GOOD91, TAAG95], hence, in order to make a fair comparison, performance of various protocols needs to be examined under the same configurations that includes, bandwidth, radio channel. This way, the performance metrics of the protocols, with a pre-defined QoS, are obtained and shown in single-scaled graphs.
easing the interpretation of the results. This method requires obtaining performances of different protocols for a wide range of protocol parameters through simulation and analysis techniques.

4.7.2 Efficiency Comparison

Even though the first approach gives a completely fair comparison of the protocols, there are, however, times when the performance metrics of protocols are unavailable for the same system configurations. In such cases, a mathematical tool provides the dynamic upper-bound of packet multiple access protocols for a given set of system configuration. By calculating performance metrics of the upper-bound for a given set of system parameters, we can obtain the "efficiency factor" (defined as the normalised value of a system metric in terms of the upper-bound figure) of each protocol. Thereafter, the performance comparison is accomplished through comparing the efficiency factors of the protocols. The efficiency factor for service class $i$ of the burst-switched sub-system is defined as

$$E^{(i)} = \frac{C^{(i)}}{upper(C^{(i)})} = \frac{Max(M^{(i)})}{upper(M^{(i)})}$$

(4.83)

where $E^{(i)}$ is the efficiency factor of bursty service class $i$ of the burst-switched sub-system, and $upper(C^{(i)})$ represents the ideal (upper-bound) capacity of the system of that service class. Similarly, for the efficiency factor of the packet-switched sub-system, the capacity can be defined as the maximum possible aggregate data rate [TAAG97a].

The efficiency factor can alternatively be defined in terms of other performance metrics (e.g. average access delay, packet dropping rate, throughput), here we will, however, only consider the capacity utilisation factor. Having arrived at the efficiency factors, we are now able to judge different protocols and compare their performances fairly with one another.
4.8 Effect of Physical Layer

4.8.1 Channel Impairments

Channel impairments could occur either by mobile channel variations (multipath, shadowing, propagation losses) or by co-channel interference. Smaller cell- (or spotbeam) layouts would further increase the co-channel interference.

Depending on the protocol robustness to channel conditions, transmission errors may affect the overall protocol efficiency differently. In general, channel impairments might cause higher packet loss, reservation loss (in the burst-switched mode) and longer access delays (in the contention mode) [FRUL93, CHUA93].

By adding a few features, the protocol can become robust to channel impairments. For example to suppress the co-channel interference in the contention mode, contention packets could be transmitted with higher power and heavier coded headers [FRUL93]. In order to reduce the reservation losses, reservations could be retained despite the loss of acknowledgement packet or uplink transmission failures. In this case, reservations would be released only when the terminal flags the End of Message (EOM) or a long sequence of packets is lost (i.e. deep fade) [CUA93].

Using interpolation techniques, packets lost from the middle of bursts might be recovered for some services such as speech [WASE88].

4.8.2 Propagation and Processing Delays

Packet multiple access protocols rely heavily on the downlink (forward-link) acknowledgement messages, informing terminals of their reservation status and contention attempts. The time constraints of RT services does not allow the terminal to wait for long acknowledgement latency. Therefore, long propagation delays could severely affect the efficiency of delay-sensitive services. For example, satellite environments are generally associated with long round-trip delays particularly on the bend-pipe with no On-Board-Processing (OBP) available.

To overcome this problem and to assure prompt access for RT services, the access mechanism should be adapted so to rely less on the delay acknowledgement. For
prompt data transmission over long-delay channels, reference [TAAG98a] proposes an access mechanism in which the terminal continues to contend until it receives a positive acknowledgement from the receiver. Similarly, for RT transmissions, the proposed access mechanism in [LEUN93, TAAG98b] allows the terminal to make simultaneous multi-reservations in the available channels and on arrival of a positive acknowledgement for one the reservation attempts, the other reservations are released by the terminal. Otherwise, the multi-reservation procedure would be repeated [TAAG98b]. These techniques will be examined in Chapter 6 and Chapter 7.

4.8.3 Channel Synchronisation

The long differential path delay between the inner and outermost points in a coverage area (satellite spotbeam or cell) imposes timing uncertainty. In circuit-switched communications such as GSM, the network is in touch with the terminal constantly and can detect and take account of the timing errors of the terminal transmissions. However, due to the bursty transmission/reception nature of packet communication systems, the network might be out of touch with the packet terminal for a relatively long time during the inactive periods [TAAG98a, TAAG98b]. The channel synchronisation then becomes critical in the contention mode as the terminals might have lost their timing advance during the inactive period. The channel synchronisation problem should be tackled differently depending on the delay distribution in the system. For small to medium coverage areas (pico- and micro-cells), contention (request) packets could be tailed by long enough guard periods (as in GSM [MOUL92]). This way contention channel could be mapped on traffic channels. Alternatively, in order to keep the synchronisation up-to-date in the idle state, a terminal can send a timing advance periodically acquiring burst in one traffic channel [TAAG98b]. This packet is recognised by the base and as soon as reception timing advance information will be piggybacked on one of the control channels to the terminal [TAAG98b]. If the timing uncertainty is high (due to large coverage area or fast movements as in satellite constellations), the contention channel should be mapped into a separate physical channel (as in ICO [ICO]) or alternatively a positioning technique (such as GPS) should be deployed [TAAG98a].

Furthermore, due to the unbalanced nature of the packet traffic load (see Chapter 3), transmission of timing advance information multiplexed with the traffic channel is
inappropriate for packet date systems. This due to the fact that a user might be only uploading data to the network and not downloading. Therefore the timing advance information would have to be separately transmitted to each terminal on regular basis. However, multiple timing advance information for different terminals can be multiplexed onto the same channel as such control information usually requires a much lower bit-rate pipe compared to that of the traffic channel.

4.9 Conclusions

In this chapter, we have developed and formulated the necessary theoretical background for the following chapters. The applications and evaluations of theories as such will be performed in Chapter 5.

A novel approach to classifying packet-based multiple access protocols has been developed by which any packet access protocols can be mapped into one of four classes of circuit-switched, packet-switched, burst-switched, and hybrid-switched (see Section 4.2). According to this classification, the well-known packet multiple access protocols have been reviewed and classified. These protocols are similar to each other and they differ in few manners, i.e. the access mechanism and channels' structure. The possible access mechanisms could be either of deterministic, random, polling, or a mixture of these (hybrid) (see Section 4.3). We named the important class of packet access protocols that deploy random access mechanisms, random access protocols.

Concentrating on the random access mechanism, a series of schemes for mixed services scenarios have been presented (see Section 4.4). It has been shown that random access mechanisms could be prioritised either through retransmission schemes or access channels (PRACH) structure. For high rates of access requests, several random access channels should be made available to the calling population (it will be shown in the next chapter that by using the concept of offered access load, \( G_a \), the required number of PRACHs for a given traffic scenario can be derived).

To further improve the stability and efficiency of random access mechanisms, dynamic and adaptive features could be added to retransmission control schemes (see Section 4.4.5). With non-adaptive retransmission control schemes, the maximum
throughput will be $37\% = 1/e$ [KLEI76]. On the other hand, throughputs up to 0.4-0.6 can be achieved if dynamic back-off schemes are deployed [RAYC90, CAPE79, GEOR82, LIU83, RAYC85]. A two-level variable permission probability has been proposed in Section 4.4.5.2 which, as will be shown in the next chapter, can optimise the contention mechanism performance close to that obtained with the ideally variable probability ($P_{opt} = 1/C$).

In Section 4.5, a general analysis of packet protocols for mixed services has been presented. Offered loads for mixed services have been derived. We differentiated between the traffic load and the access load. With the access load concept defined as the load imposed on the access channels (PRACHs), stability of the access mechanism can result, while, with the traffic load the overall system stability and saturation points can be obtained. The applications of these two concepts will be shown in the following chapters. A general formula for calculation of packet dropping ratio of RT services has been derived. Then, an analytical method called "Aggregate Source Model" (ASM) was proposed and parameterised such that various service types are modelled with one representative service model. This single service model then is used in conjunction with the monomedia analysis to obtain the multimedia performance measures. The ASM method developed in this chapter is applicable to any finite-population non-prioritised resource-sharing system. As was shown, random access protocols could also be considered queueing systems by introduction of "Random Access Queue" (RAQ) concept.

Statistical upper-bound of packet multiple access protocols are achieved through idealising the access mechanism. Three different approaches to idealisation of the access mechanism have been presented, namely Upper-Bound With Contention, Perfect Capture Upper-Bound, and Contention-Free Upper Bound. The latter has been modelled and analysed both for the burst-switched and the packet-switched sub-system using a FCFS queueing system [TAAG95, TAAG97a]. Based on the upper-bound concept, the efficiency factor of protocols has been defined as the maximum achievable upper-bound capacity of a given protocol. Then, two methods for comparison of packet access protocols have been proposed, namely cross comparison (comparison of protocols under the same conditions) and efficiency-comparison (i.e., comparisons of efficiency of protocols).
Finally, in Section 4.8, the effects of physical layer (such as channel impairments, propagation delays, and channel synchronisation) on the design and the performance of packet multiple access protocols have been discussed. These effects are taken into account in Chapter 6 and Chapter 7, where specific protocols are proposed and evaluated for cellular and satellite systems.
5.1 Introduction

This Chapter employs the techniques and the theories developed in Chapter 4 to evaluate, compare and analyse various packet-based multiple access protocols. Starting with evaluation of random access mechanisms, the performance of slotted ALOAH systems with a non-dynamic retransmission scheme (Chapter 4-Section 4.4.1) is portrayed, followed by evaluation of the two-level variable permission probability scheme (Chapter 4-Section 4.4.5), prioritised and non-priorities mechanisms. The analysis carried out for the upper-bound is verified in Section 5.4 (see Chapter 4-Section 4.6.2) followed by an extensive study of contention-free and with-contention upper bounds for various protocol parameters. Then, based on the offered loads concepts developed in Chapter 4, performance of a typical packet random access protocol (PRMA) is evaluated for a mixture of packet voice (PV), E-mail, FTP, and WWW services (Chapter 3). Using the two comparison methods (cross-comparison and efficiency-comparison), a comprehensive comparison of well-known packet access protocols are carried out. Accuracy and applications of “Aggregate Source Modelling” (ASM) is presented in Section 5.5, where various queueing and multi-access systems are analysed for mixed services. As was mentioned in the previous chapter, by using this method, we can avoid multidimensional Markov chains associated with analysis of multimedia systems. Results show that the ASM analytical method could still be used in presence of non-Poisson inputs referring to the discussion of Chapter 4-Section 4.5.2.1.
5.2 Protocol Parameters Spreadsheet

A list of protocol parameters (PP) of the well-known packet multiple access protocols are provided in Table 5.1 and Table 5.2. These protocols are mainly evaluated for a mixture of packet voice (Table 5.1) and packet data services (Table 5.1) [GOOD91, WONG92, NARA96, TAAG96b, DEVI93, MITR90, DUNL93]. These parameter sets along side with three additions PPs are used throughout this Chapter for numerical evaluations and comparison. Here, $M_v$ and $M_d$ represent maximum allowed packet voice and data terminals, respectively, before the QoS requirements are violated.

| Table 5.1 Parameter Sets of Proposed Reservation Protocols (Voice Subsystem) |
|---|---|---|---|---|---|---|---|---|---|
| **XYMA** | $R_v$ | $R_v$ | $T_{av}$ | $T_{off}$ | $t_{on}$ | $t_{off}$ | $T_f$ | $N$ | $M_v$ |
| Sample PP₁ | 4000 | 32 | 1000 | 1350 | 0 | 0 | 10 | 20 | 0 | 100 | 224 |
| Sample PP₂ | 1800 | 13 | 1410 | 1740 | 0 | 0 | 5 | 10 | 0 | 72 | var |
| Sample PP₃ | 720 | 16 | 1000 | 1350 | 0 | 0 | 10 | 30 | 0 | 30 | var |
| PRMA | 720 | 32 | 1000 | 1350 | 0 | 0 | 16 | 32 | 0 | 20 | 36 |
| IPRMA | 224 | 8 | 1000 | 1350 | 0 | 0 | 20 | 40 | 0 | 20 | 36 |
| PRMA++ (N18) | 450 | 13 | 1410 | 1740 | 10 | 200 | 5 | 10 | 3 | 18 | 28 |
| PRMA++ (N72) | 1800 | 13 | 1410 | 1740 | 10 | 200 | 5 | 10 | 3 | 72 | 139 |
| BRMA | 500 | 8x2 | 1410 | 1740 | 10 | 200 | 5 | 10 | 3 | 17 | 30 |
| BuRMA | 720 | 32 | 1000 | 1350 | 0 | 0 | 16 | 32 | 0 | 20 | 37 |
| FRMA (N11) | 432 | 32 | 1000 | 1350 | 0 | 0 | 11 | 11 | 11 | 11 | 18 |
| FRMA (N20) | 720 | 32 | 1000 | 1350 | 0 | 0 | 16 | 16 | 16 | 20 | 36 |
| Mitrou’s PRMA | 2000 | 10-12 | 500 | 500 | 100 | 50 | 5-10 | 20 | 0 | 150 | 280 |
| TM-BCMA/CD | 2000 | 16 | 1000 | 1350 | 0 | 0 | 20 | 20 | 0 | 105 | 188 |

| Table 5.2 Parameter Sets of Proposed Reservation Protocols (Data Subsystem) |
|---|---|---|---|---|---|
| **XYMA** | $D_d$ | $M_v$, $M_d$ | $R_d$ |
| Sample PP₁ | — | — | — |
| Sample PP₂ | — | — | — |
| PRMA | 150 | 36, 36 | 1.2 |
| IPRMA | 40 | 24, 24 | 2 |
| PRMA++ (N18) | — | — | — |
| PRMA++ (N72) | — | — | — |
| BRMA | — | — | — |
| BuRMA | — | — | — |
| FRMA (N11) | — | — | — |
| FRMA (N20) | 200 | 20, 20 | 3 |
| Mitrou’s PRMA | — | — | — |
| TM-BCMA/CD | 900 | 30, 130 | 4.64 |
5.3 Performance Evaluations of Random Access Mechanisms

5.3.1 Performance of Slotted-ALOHA

The random access mechanism associated with packet access protocols could use separate PRACHs. Performance of such random access sub-systems behave very similarly to that of the slotted ALOHA system.

Figure 5.1 shows numerical solutions to Equations (4.3)-(4.5) where the throughput $S_n$ is plotted as a function of the channel traffic $G_n$ for various values of $P$.

![Figure 5.1 S-ALOHA Throughput Versus Normalised Access Load for Various Permission Probabilities](image)

We note that the maximum throughput at any given $P$ occurs at $G_n=1$. The throughput improves as $P$ decreases, finally, yielding a maximum value of $S_n=1/e=0.368$ for $G_n=1$, $P\to0$. Thus we have the ultimate capacity of this channel supporting a large number of small users in less than 37% of its theoretical maximum (of 1). We note that the efficiency rapidly approaches this limiting value (of $1/e$) as $P$ decreases. It should be noted that higher throughput does not necessarily mean lower access delays. It is clear that as the load increases the access delay grows regardless of the throughput. Therefore, maximisation of the throughput ($P\to0$) would not result in the same as the optimisation of contention access delays.
The performance of slotted ALOHA can be improved by deploying dynamic retransmission schemes (Chapter 4-Section 4.4.5). Using such schemes, normalised throughput as high as 0.6 can be achieved [RAYC90, CAPE79, GEOR82, LIU83, RAYC85, TAAG96a] as opposed to 0.37 of the non-dynamic retransmission scheme (above).

5.3.2 Evaluation of Two-Level Variable Permission Probability Scheme

In Chapter 4-Section 4.4.5.2, a novel and easy-to-implement dynamic retransmission scheme has been proposed. A terminal using such retransmission scheme in the random access mechanism resumes transmission with an initial permission probability of $P_0$ and upon collision access attempts are retransmitted with probability $P_r$.

So as to study and compare two-level variable permission probabilities with non-dynamic and optimum retransmission schemes (Chapter 4-Section 4.4.5.1), the PRMA protocol with packet voice services is considered (see the PP set of Table 5.1).

Figure 5.2 shows the optimum $P_0$ and $P_r$ variations against the traffic load. As the traffic load builds up, both $P_0$ and $P_r$ should be decreased for the optimum solution. As seen, for very high traffic loads the two-level permission probability scheme merges the of the single-level (non-dynamic) scheme.

![Figure 5.2 Optimisation of Two-Level Variable Permission Probability for PRMA](image)
As shown in [GOOD91], the voice-only sub-system PRMA with the considered
protocol parameters can accommodate around 36 simultaneous users. Referring to the
above graph, \( P_0 \) for this case should be assumed high. Figure 5.3-Figure 5.5 show the
average packet dropping percentage experienced of PRMA with high initial permission
probabilities of \( P_0=1, P_0=0.95, P_0=0.9 \), and \( P_0=0.85 \). The performance of such
scenarios are compared against the optimum case (\( P=P_{opt}=1/C \)) and the non-dynamic
case (\( P_t=P_0 \)). Figure 5.6 displays front-end clipping probability distribution functions
(CDF) against number of dropped packets for the six contention procedures.

As shown, the simple two-level variable permission probability scheme performs
very closely to the optimum retransmission scheme (\( P=1/C \)) case and reduces the packet
dropping percentage so that the exemplary system (with \( P_0=0.95, 0.9, 0.85 \)) can
accommodate up to 38 simultaneous conversations with a total of 20 channels (recalling
that packet dropping rate of 1% or less can be tolerated by packet voice services)
[GOOD89].

![Figure 5.3 Voice Packet Dropping Ratio Vs. Traffic Load
\((P_0=1, P_0=0.95, P_0=P_t, P=1/C)\)
Figure 5.4 Voice Packet Dropping Ratio Vs. Traffic Load
\( (P_0=1, P_0=0.90, P_0=P_r, P=1/C) \)

Figure 5.5 Voice Packet Dropping Ratio Vs. Traffic Load
\( (P_0=1, P_0=0.85, P_0=P_r, P=1/C) \)

Figure 5.6 CDF of Voice Packet Dropping \( (M_v=38) \)
5.3.3 Evaluation of Non-Prioritised Schemes with Mixed Services

In Chapter 4-Section 4.4.2, it has been discussed that if various services deploy the same access mechanism, the priorities will be eliminated.

The mixed traffic considered in here consists of three equally populated services of PV, WWW, and FTP (see Chapter 3). Amongst the three, the PV service imposes real-time constraints. Deliberately ignoring the delay-sensitivity of PV terminals, in this section, the performance of a random access mechanism is obtained where no priorities amongst different services are employed.

The operation mode of PRACH is set as “General”. Therefore, all the services can contend for these PRACHs. The configuration of PRACHs is illustrated in Figure 5.7.

![Figure 5.7 PRACHs' Configurations for Non-Prioritised Case Evaluation](image)

Figure 5.10 and Figure 5.9 represent the contention delay and the throughput, respectively, against the mixed traffic load. The graphs are obtained assuming a periodic cycle of 5 ms ($T_c = 5$ ms) and non-dynamic permission probability is optimised for every simulation step. The following graphs are drawn against the total number of users of the three services of PV, WWW, and FTP. It should be noted that if there is no priority in the system, different services experience the same access and contention delays.

Figure 5.8 shows the normalised access load on the PRACH with respect to population of the mixed services. It should be noted that the access load is always less than the traffic load, therefore, quite a number of sources are required to generate a significant access load (see Chapter 4-Section 4.5.2.1).
Figure 5.8 Normalised Access Load Versus Total No. Terminals

The throughput graph (Figure 5.9) shows a constant increase in the channel utilisation until the point of $G_a=1$ after which the random access performance starts declining (breaking point). As seen, the throughput reaches very close to the maximum of non-dynamic retransmission schemes (i.e., $S_a=0.36$) at load of around $G_a=1$ (see Figure 5.1).

Figure 5.9 Aggregate Normalised Throughput Vs. Total No. Terminals (Non-Prioritised)

As seen in Figure 5.10, the contention delay increases as the traffic builds up. This delay is due to collisions associated with random access schemes. The contention delay
in this case is mostly created by the PV traffic rather than those of FTP and WWW. The latter services are less active and hence the access load generated by them is lower (see Chapter 4-Equation (4.38)). Besides, in order to maximise the throughput, in the simulation performed here, permission probability $P$ has been assumed very small for higher loads (e.g., at load $G_a=1$, $P$ was optimised to $0.001$). In such cases, the contention delay, as shown in Figure 5.10, would increase even more dramatically with the offered access load (see Chapter 4-Equation (4.9)).

\[ \text{Figure 5.10 Aggregate Contention Delay Vs. Total No. Terminals (Non-Prioritised)} \]

In Figure 5.11 and Figure 5.12, impacts of PRACHs' periodic cycle on the contention delay and the throughput is presented.

\[ \text{Figure 5.11 Effect of PRACH Periodic Cycle on Contention Delay (Non-Prioritised)} \]
As seen, the system performance is degraded for high periodic cycles. As the channel periodic cycle grows, the signalling load waiting to be served becomes higher. This, in turn, increases the channels load.

![Figure 5.12 Effect of PRACH Periodic Cycle on Throughput](image)

### 5.3.4 Evaluation of Prioritised Schemes with Mixed Services

In this section, a prioritised access mechanism for mixed services of PV, FTP, and WWW is considered. In the considered case, the priority is given to the PV traffic (because of delay-sensitivity) by assigning two PRACHs to PV terminals. Figure 5.13 shows the considered PRACHs’ configurations. It should be noted that the exemplary configuration given below does not necessarily represents the optimum arrangement.

![Figure 5.13 PRACHs' Configurations for Prioritised Case Evaluation](image)

Figure 5.14 and Figure 5.15 portrait contention delays and the throughput, respectively, versus the total number of prioritised terminals. The results show the performance of the PRACH mechanisms for the three different service types (PV, FTP, and WWW) and the aggregate performance which lies in between the graphs.
We can be seen, the PV traffic has the promptest delivery amongst the three service types (as desired). The low access delay of packet voice terminals is due to allocation of two PRACHs. For larger PV population, we will need more than two PRACHs (see). The FTP and WWW traffic, however, were assigned only one PRACH each. Hence, the sub-system performances for these services are worse (lower throughput and higher contention delay).

![Figure 5.14 Contention Delays Vs. Total No. Terminals (Prioritised)](image)

**Figure 5.14** Contention Delays Vs. Total No. Terminals (Prioritised)

![Figure 5.15 Throughput Vs. Total No. Terminals (Prioritised)](image)

**Figure 5.15** Throughput Vs. Total No. Terminals (Prioritised)
5.4 Performance Evaluation of Statistical Upper-Bound

5.4.1 Analysis Verification of Contention-Free Upper-Bound

In Chapter 4-Section 4.6.2, an analysis of contention-free upper bound has been carried out. Here, we compare the simulation results with such analysis using protocol parameter set of PP₁. Figure 5.16 and Figure 5.17 show, respectively, the packet dropping ratio ($P_{drop}^{upper}$) and the capacity ($C_{v}^{upper}$) of the burst-switched sub-system of the contention-free upper-bound, while Figure 5.18 demonstrates the joint throughput ($S^{upper}$) of burst-switched and the packet-switched sub-systems. The close agreement between the simulation and the analysis proves the accuracy of the analysis carried out. The difference between the simulation and analysis results in Figure 5.16 for high packet dropping is due to the approximation we made in Equation (4.45). This approximation is adequate, since for good quality RT transmission, we are only interested in the low-end of the packet dropping graph.

![Figure 5.16 Voice Packet Dropping Vs. Traffic Load (Using PP₁ of Table 5.1)](image)
5.4.2 Study of Upper-Bound Performance

In order to observe the general behaviour of packet multiple access protocols, here we study the protocol performance by varying different protocol parameters [TAAG95]. Two upper-bound protocols are considered here, namely, Contention-Free Upper-Bound and With-Contention Upper-Bound (see Chapter 4-Section 4.6).

5.4.2.1 Effect of Traffic Load

The effects of traffic load on access delay, packet dropping ratio, and throughput are depicted in Figure 5.19-Figure 5.21, respectively. As show the contention-free upper-
bound outperforms that with contention. This is quite expectable since collisions in the access phase degrades the performance of the contention-with upper-bound.

Figure 5.19 Access Delay of Upper-Bounds Vs. Traffic Load (Using PP₁ of Table 5.1)

As seen, for the contention-free upper-bound, the average access delay (as well as packet dropping) graph becomes liner after the so-called "saturation point", \( M \), defined by [KLEI76]
\[ M^* = \frac{N}{\Gamma} \] 

(5.1)

\(M^*\) is exactly the maximum number of terminals that could be scheduled to cause no mutual interference and each user beyond this point would cause all queued-up users to be delayed by a time equal to his entire service time \((T_{on})\). However, as was discussed in Chapter 4-Section 4.5.5, with-contention upper-bound case (or for infinite value of \(M\)) the system cannot become "unstable" [KLEI76, GROS85].

It should also be noted that packet dropping graph is closely related to the experienced delay. Therefore, the trend of \(P_{drop}\) graphs are always similar to those of the access delay.

5.4.2.2 Number of Traffic Channels

Figure 5.17 displays the impact of \(N\) on the burst-switched subsystem capacity, \(C^{upper}\). It reveals that the upper-bound protocol gains its highest efficiency in high bandwidth. However, as seen, the capacity does not increase significantly after exceeding a certain number of time slots. For higher channel bit-rate after this point we do not benefit substantially from the multiplexing gain, hence, in order to boost the overall capacity of the cellular system we should switch to multi-carrier frequency planning. We call this phenomena the "multiplexing gain saturation".

![Graph showing normalised throughputs of upper-bounds vs. traffic load](image-url)
As observed for upper bound with contention, the capacity starts degrading after it reaches its maximum. This is due to the excessive contentions experienced with this upper bound. It should be noted that this effect (instability) can be observed in any packet random access protocol and the case shown in the figure below demonstrates the upper-bound of such protocols.

![Trunking Efficiency in Upper-Bounds](image)

**Figure 5.22 Trunking Efficiency in Upper-Bounds**

### 5.4.2.3 Activity Factor

According to empirical results presented in [BRAD68], changing the voice activity detector threshold has no significant impact on the mean silence duration and it only changes the mean talkspurt duration. Therefore, we achieve different activity factors ($\Gamma$) through changing the mean talkspurt duration ($T_{on}$) while fixing the mean silence duration ($T_{off}$).
Figure 5.23 Effect of Activity Factor on Capacity (Using PP, of Table 5.1)

Figure 5.23 illustrates the upper-bound burst-switched subsystem capacities, \( C_{\text{upper}} \), as a function of voice activity factor. As expected, the bigger the voice activity factor, the less the capacity such that as \( (\Gamma) \) tends to one, the capacity reaches to one, i.e. the circuit-switched capacity (one user per traffic channel).

As seen from Figure 5.22 and Figure 5.23, the difference between upper-bound capacities of the two upper-bound protocols are negligible. This result will be re-confirmed when efficiencies of packet access protocols are compared against each other in Section 5.5.2.2.

5.5 Evaluation of Packet Multiple Access Protocols

5.5.1 Performance of PRMA for Mixed Packet Services

In this section, the PRMA protocol is examined for a range of services, including PV, FTP, WWW, and E-mail. The typical values of this protocol is used here (Table 5.1). Figure 5.24 shows the normalised offered load traffic load generated by each of these services. As shown PV generates the largest portion of the traffic load, followed by FTP, E-mail, and WWW. It is been assumed that a terminal reserves no more than one channel in a frame.
Figure 5.24 Normalised Traffic Load Generated by Individual Service Types

Figure 5.25 demonstrates the performance of a case in which a packet service is multiplexed by PV traffic. The graph shows maximum number of packet terminals for a given number of PV terminals. One of the limiting factor is the packet dropping ratio (considered less than 1% here) and the other is the stability of the system. As the number of sources increase, the access load increases and because there is no separation between access channels and traffic channels in PRMA, the protocol becomes unstable very quickly. As expected, we can accommodate larger number of WWW terminals than E-mail and FTP.

Figure 5.25 Max Possible No. Data Terminals Versus No. PV Terminals
To further examine the performance of PRMA for mixed services, a traffic scenario consisting of PV, WWW, E-mail (equally populated), and FTP (10th of population) services. The population of FTP terminals has been considered less in order to reduce the overshadowing effect of FTP on the other traffic source. The offered traffic load generated by such scenario is depicted in Figure 5.26 followed by its normalised version in Figure 5.27.

![Figure 5.26 Offered Traffic Load Versus Terminal Population](image)

![Figure 5.27 Normalised Traffic Load Versus Terminal Population](image)

Figure 5.26-Figure 5.31 demonstrates the performance metrics of PRMA for this mixed traffic scenario. Generally, it can be seen that the performance of the PRMA
protocol starts declining after $G_t=0.65$. One could have expected this breaking point to be at $G_t=1$. However, due to the lack of separate contention channels (PRACHs), there is no distinction between PRACHs and traffic channels and the actual load imposed on channels are twofold (traffic as well as access loads). Therefore, the protocol reaches its breaking point earlier than $G_t=1$.

**Figure 5.28 Average No. Contention Attempts Versus Normalised Offered Load**

**Figure 5.29 Average Contention Delay Versus Normalised Offered Load**
Figure 5.30 Packet Dropping of PV Terminals Versus Normalised Offered Load

Figure 5.31 Normalised Throughput Versus Normalised Offered Load

5.5.2 Comparative Study

Since our primary aim here is to evaluate the multiple access efficiency, we do not take other design parameters (ranging from the cellular lay-out to transmission techniques) into account. Furthermore, protocols are assumed to operate in an environment free of channels errors, co-channel interference, and capture effect [TAAG97a, FRUL94]. We also do not consider the interleaving depth, since it does not affect the multiplexing gain or the multiple access efficiency.
5.5.2.1 Cross-Comparison

Here we consider three protocols to cross-compare, PRMA, PRMA++, and BuRMA which is a burst-switched multi-access protocol with a two-level variable permission probability random access mechanism (see Chapter 6-Section 6.3) [TAAG96b].

To make a thoroughly fair comparison, performance metrics of protocols should be examined under identical system design parameters. Thus, we used Sample PP2 of Table 5.1 to simulate the three protocols. Contention parameters of the protocols were optimised for each value of $M_v$, separately (permission probabilities and number of reservation slots in case of PRMA++). In order to optimise PRMA++, reservation slots should be evenly distributed over the whole length of the frame [ROBE94].

![Figure 5.32 Voice Packet Dropping Ratio of RRA Protocols Vs. Traffic Load (Using PP2 of Table 5.1)](image)

Figure 5.32 displays the packet dropping percentage with respect to the number of voice-only terminals for the three protocols as well as the upper-bound.
In Figure 5.33, throughputs of these protocols and the upper-bound are shown. To have a fair comparison, for PRMA++, the throughput is measured considering all time slots including contention slots.

Figure 5.34 shows the average access delay of the protocols. The access delay for PRMA++ consists of the contention delay and the queueing delay.
The average number of contention attempts is portrayed in Figure 5.35. The contention attempts are less for PRMA++ due to the dedication of slots for contention-only purposes. The number of contention slots has been optimised for PRMA++ nearing it to the optimum contention attempts (i.e., the upper-bound performance). Looking back at Figure 5.34, it can be concluded that the access delay in PRMA++ is mainly built up in the queue rather than in the contention phase.

As seen, PRMA and BuRMA are in close harmony with the two upper bounds (with-contention and contention-free upper bounds). In PRMA++, some number of time slots are dedicated to reservation (contention) purposes and actual traffic is sent over the traffic channels only. This usage of reservation slots is unnecessary since: 1) if traffic channels are available, there is no need for additional time slots for contention, and, 2) if traffic channels are all occupied, contention would be inappropriate.

5.5.2.2 Efficiency-Comparison

As mentioned earlier, by calculating efficiency factors of a given PP, we can make a comparison between multiplexing gains of protocols. Table 5.1 and Table 5.2 extract crucial parameters of well-known packet-based protocols. To demonstrate a fair comparison, the same QoS factor as the nominal performance of a protocol should be considered in upper-bound calculations. In this study, we consider the with-contention and contention-free upper-bound protocols. Figure 5.36 and Figure 5.37 portray the
voice-only efficiency of ten protocols in an ascending order, using contention-free and with-contention upper-bounds, respectively. For most of the multi-access protocols, packet dropping probability is taken as the limiting factor. However, Mitrou's PRMA has considered the complementary probability distribution function \((CPDF = 1 - CDF)\) of delay or packet loss as the performance metric. To obtain the contention-free upper-bound for this case, Equation (4.67) was utilised.

**Figure 5.36** Voice Sub-System Efficiency of Packet Multiple Access Protocols Using *Contention-Free Upper-Bound* (See Table 5.1)

**Figure 5.37** Voice-Only Efficiency of Packet Multiple Access Protocols Using *Upper-Bound with Contention* (See Table 5.1)
Obtaining similar bar-charts for the data sub-system is more difficult, as most protocols are only studied for voice sources. Figure 5.38 and Figure 5.39 represent the efficiency comparison of data sub-system in mixed voice/data environment the two upper bound protocols.

![Figure 5.38 Data Subsystem Efficiency for Mixed Services Using Contention-Free Upper-Bound (See Table 5.2)](image1)

![Figure 5.39 Data Subsystem Efficiency for Mixed Services Using Upper-Bound with Contention (See Table 5.2)](image2)

It is generally agreed that protocol parameters have great impact on the global performance of a protocol [GOOD91, TAAG95, NAND90]. Consequently, the same protocols with different configurations might be ranked in a dissimilar order.
5.6 **Mixed Services Analysis Using the ASM**

5.6.1 **Analysis of Mixed Services FCFS Queues**

In this section, we consider a FCFS (non-prioritised) queueing discipline, provides a set of arbitrary traffic parameters generated by \( r=15 \) different service classes. It is assumed that inactive periods of sources are negative exponentially distributed. It has been shown that, for finite-source queues, steady-state probabilities (and hence performance measures) are independent of the service distribution profile as long as arrivals follow negative exponential distribution. Likewise, if the service time distribution is negative exponential, the system performance is invariant to the distribution of arrival (thinking) times [GROS85, BUND80]. Sources are allocated one of the available servers. A total number of \( N=23 \) parallel servers is considered. In this section, for evaluation purposes we consider two different multi-traffic multi-server queueing system [TAAG98c].

5.6.1.1 **M/G/N//FCFS with Five Service Types**

In the first example, we consider a system with five different traffic types \((r=5)\) as listed in Table 5.3. These numbers are chosen arbitrarily in order to evaluate the generality of the analytical model. The number of servers is assumed to be 10 \((N=10)\).

| Table 5.3 Traffic Parameters of Example 1 |
|-----|-----|-----|-----|
| \(i\) | \(M_i\) | \(\lambda_i^{-1}\) | \(\mu_i^{-1}\) |
| 1   | 10  | 1000 | 100 |
| 2   | 25  | 100  | 10  |
| 3   | 80  | 892  | 56  |
| 4   | 20  | 2000 | 250 |
| 5   | 14  | 1570 | 200 |

Using the equations (4.57), (4.63), and (4.64), the RASM parameters can be obtained as \( M^{RASM} = 139, \mu^{RASM}=34.7^{-1}, \lambda^{RASM}=377.2^{-1} \).

Using the analysis of Appendix B with the RASM parameters, the measures of effectiveness of the analytical analysis are compared against the corresponding simulation measures. This comparison is demonstrated in Table 5.4.
The close agreement between the simulation and the analysis proves the accuracy of the analytical modelling technique.

### Table 5.4 Measures of Effectiveness Comparison (Example 1)

<table>
<thead>
<tr>
<th></th>
<th>$S$</th>
<th>$L$</th>
<th>$L_a$</th>
<th>$D_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.973</td>
<td>26.2</td>
<td>18.5</td>
<td>72.7</td>
</tr>
<tr>
<td>ASM Analysis</td>
<td>0.996</td>
<td>30.6</td>
<td>20.6</td>
<td>71.9</td>
</tr>
</tbody>
</table>

### 5.6.1.2 M/G//N/FCFS with Fifteen Service Types

In this example, we increase the traffic types to 15 ($r=15$). Table 5.5 provides the traffic parameters used for this example assessment. A total number of $N=23$ is considered for number of parallel servers. The $RASM$ parameters will be $M_{RASM} = 348$, $\mu_{RASM} = 52.7^{-1}$, $\lambda_{RASM} = 662.3^{-1}$.

### Table 5.5 Traffic Parameters of Example 2

<table>
<thead>
<tr>
<th>$i$</th>
<th>$M_i$</th>
<th>$\lambda_i^{-1}$</th>
<th>$\mu_i^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>987</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>3000</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>3000</td>
<td>198</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>5000</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>2764</td>
<td>200</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>3000</td>
<td>500</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>800</td>
<td>80</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>970</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>80</td>
<td>892</td>
<td>56</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>1987</td>
<td>98</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>1000</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>10000</td>
<td>1000</td>
</tr>
<tr>
<td>13</td>
<td>20</td>
<td>2000</td>
<td>250</td>
</tr>
<tr>
<td>14</td>
<td>60</td>
<td>1000</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>25</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5.6 compares the theoretical analysis with the simulation result. Again, as seen, the analysis closely matches with the simulation.

### Table 5.6 Measures of Effectiveness Comparison (Example 2)

<table>
<thead>
<tr>
<th></th>
<th>$S$</th>
<th>$L$</th>
<th>$L_a$</th>
<th>$D_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.978</td>
<td>52.7</td>
<td>31.9</td>
<td>81.1</td>
</tr>
<tr>
<td>ASM Analysis</td>
<td>0.998</td>
<td>59.1</td>
<td>36.1</td>
<td>82.9</td>
</tr>
</tbody>
</table>
5.6.1.3 Applicability of ASM to Mixed Services Analysis with Non-Poisson Inputs

In order to find the applicability of the ASM to the non-Poisson inter-arrival distributions, FTP and WWW models are then fed into the ASM analysis (which was originally designed for Poisson inter-arrivals). Here, we consider a FCFS queueing discipline deployed on the downlink traffic stream consisting of FTP and WWW data blocks. Each data burst (i.e. file for FTP and IP datagram for WWW) is assigned only one of the \( N=20 \) channels. The average active and inactive times (and hence the activity factor) of the FTP and WWW sessions are calculated for a channel rate of \( R_c=32 \, \text{Kb/s} \) (see Figure 3.10 and Figure 3.12). The radio channel is assumed to be error-free, and hence, no channel encoding used (the erroneous channels would have inverse impacts on performance). As explained earlier, the traffic source models represented above follow a mixture of log-normal, Pareto, and Geometrical distributions for inter-arrival times or service times.

![Graph a: Throughput](image)

![Graph b: Average Access Delay](image)

Figure 5.40 Performance of FCFS Queue with Internet Traffic Inputs
a) Throughput and b) Average Access Delay
Figure 5.40 shows the performance metrics (throughput and average access delay) of the above mentioned FCFS queueing system for a mix traffic generated by equally populated FTP and WWW sources. Because of the long reading time associated with WWW sources, for an accurate average, the simulation had to be run for a lengthy time period. The RASM parameters have been calculated for different numbers of sources and then fed into the analysis of $M/G/N//FCFS$ queueing system (see Appendix B) [GROSS85, KLEI76]. The above figures show very close agreement between the proposed ASM method and the simulation results even though in the proposed model Poisson inputs were assumed. As expected with the finite source FCFS queues, the access (queueing) delay becomes linear at the upper region (saturation) [KLEI76].

### 5.6.2 Analysis of Random Access Protocols Using the ASM

Here, we consider a PRMA protocol in which arrivals from different service classes contend with equal permission probability (hence non-prioritised) for available time slots [TAAG98]. We assume negative exponential distribution for inactive periods (Poisson inter-arrival) of each source. In order to examine service distribution independent property of such a protocol, we have considered a finite population of with $T_{on} = 1000 \text{ ms}, T_{off} = 1350 \text{ ms}$ with a frame duration of $T_f = 10 \text{ ms}, P=0.3, N=30$, and $D_{max}=30 \text{ ms}$ [GOOD91].
Figure 5.41 Service Distribution Independence of PRMA
a) Throughput and b) Average Access Delay

Figure 5.41 shows the simulation results for such a system with four different statistical distributions for service times (hybrid distribution represents a mix of exponential and uniform distribution). The performance metrics show very similar values. Hence, we can conclude that our protocol is service distribution independent.

Having satisfied all the four required conditions, we can now use the ASM method to analyse the mixed service case of the reservation-based protocol. Below, we have considered a set of three different service classes \( r = 3 \) as (the time is considered in ms)

\[
M = \begin{pmatrix} M^{(1)} \\ M^{(2)} \\ M^{(3)} \end{pmatrix} \quad \quad \quad T_{\text{off}} = \begin{pmatrix} 1000 \\ 1300 \\ 2000 \end{pmatrix} \quad \quad \quad T_{\text{on}} = \begin{pmatrix} 100 \\ 50 \\ 300 \end{pmatrix}
\]

Assuming a total number of \( N = 10 \) time slots, \( T_f = 10 \text{ ms} \) and \( P = 0.2 \), the simulation results for the mixed and the equivalent RASM monomedia cases are demonstrated in Figure 5.42. Again, the close agreement proves the accuracy of the ASM analytical model.
Random access protocols suffer from collisions experienced during the contention-based access procedure. These collisions could result in system instability. The instability point will be reached when the number of access requests to PRACHs exceeds the *breaking point* after which no user will succeed in gaining access [CARL75]. Simulation results for various traffic configurations further show that the random access system using *ASM* has the same breaking point as that of the multimedia case. Therefore, the monomedia with the *RASM* service can be used for stability study of multimedia random access systems eliminating typical mathematical complications associated with stability studies [TASA83, NAND94].
5.7 Conclusions

In this chapter, packet multiple access protocols have been evaluated, compared, and discussed using the theories and analytical models developed in Chapter 4.

A two-level variable permission probability scheme (Chapter 4-Section 4.4.5) has been evaluated and optimised for PRMA. The results show considerable improvements (close to the upper-bound) over the non-dynamic retransmission schemes.

In the process of designing the random access mechanism, we could see that the normalised offered access load provides a concept based on which an optimum and stable PRACH configuration for mixed services can be obtained. This arrangement of PRACHs should be carried out considering QoS requirements and the traffic scenario. Random access mechanisms, in general, could enter an unstable situation if the calling population exceeds a limit known as breaking point. Hence, the higher the population, the larger the number of PRACHs are needed to provide a prompt and stable access. Besides, in order to prioritise real-time services over the non-real-time, separate and sufficient number of PRACHs should be assigned to real-time service. As shown in the previous section, less active and burstier services (such as WWW and FTP) could share common “General” PRACHs and there is no need for assigning separate PRACHs to each of them. It has also been shown that the periodic cycle of PRACHs could inversely effect the system performance. For large cycles, terminals will generate more signalling requests and hence the probability of collisions increases. On the other hand, although the probability of collision will be reduced (and so will the contention delay), the required signalling channel rate will be higher.

The performance of PRMA for mixed services scenario (consisting of RT PV, FTP, WWW, and E-mail) has been evaluated. Results show that protocols with no distinction between traffic and random access channels (such as PRMA) have less stability than those with separate channels (see Chapter 6 and Chapter 7 for more results on protocols with PRACHs). However, the PRMA protocols (and the like) provide better bandwidth efficiency than protocols with PRACHs.

The analysis (carried out in Chapter 4-Section 4.6.2) of the contention-free upper-bound (both for burst-switched and packet-switched sub-system) has been verified. The
analysis does not only lead to average figures, it also provides more detailed performance metrics such as access delay distribution and state probabilities. The conventional belief is that the upper-bound of burst-switched sub-system of multiple access protocols is the inverse of the activity factor \( (\Gamma^{-1}) \) [GOOD89, DEVI93, MITR90, DUNL93, GOOD91, ROBE94, MITR93]. As observed in this chapter, with low number of traffic channels, the multiplexing gain cannot reach \( \Gamma^{-1} \). In addition, the multiplexing gain (capacity) of the system depends on the QoS requirements. For example, if we could tolerate longer dropping delay limits \( (D_{\text{max}}) \) or higher packet dropping ratio \( (P_{\text{drop}}) \), the system capacity can change dramatically.

Then, two different collation methods (cross-comparison and efficiency-comparison) have been exploited for a comprehensive performance comparison evaluations of ten well-known packet protocols. It was shown that PRMA, FRMA, and BurMA protocols outperform the others for the given system configurations. Nevertheless, due to statistical multiplexing, packet-based protocols demonstrate high capacity and have better performances than those of the circuit-switched. A thoroughly fair comparison of the multiple access protocols is a difficult task. Here, although we have taken into account numerous design parameters affecting the global performance of multiple access protocols, there are more parameters to be considered, such as effect of realistic channels, near/far (partial and perfect capture) effects, propagation and processing delay, co-channel interference, and so forth.

Also, we have verified the "Aggregate Source Modelling" (ASM) analytical method for four resource sharing scenarios: (1) A FCFS queueing system with 5 arbitrary service classes and Poisson inputs, (2) A FCFS queueing system with 15 service types and Poisson inputs, (3) A random reservation access protocol with 3 service types and Poisson inputs, and (4) a FCFS downlink scheduling protocol for the Internet services of FTP and WWW (non-Poisson inputs). The comparisons between the simulation results the corresponding analytical model demonstrate very close agreements. The accuracy of the ASM method is due to the fact that the non-Poisson distribution could well be approximated if the silence durations are considerably larger than a packet transmission time (as proved in Chapter 4-Section 4.5.2.1).
6.1 Introduction

In the previous chapters, the theory and methods for designing and optimisation of multimedia packet-based multiple access protocols have been established. In this Chapter, as an application of the developed techniques, provision of packet-based services for the future cellular communication systems is addressed. The two proposed multiple access protocols in this section can support mixed RT and NRT services for cellular packet communications. The first MAC sub-layer protocol applies to GPRS-type environments, named "Round-Robin Multiple Access" (RRMA) [TAAG97b]. As the name implies, in RRMA, the TDMA transmission frame acts as a Round-Robin queuing system such that its size could vary according to the received traffic load. Using this protocol, no traffic channel is left unused even if only one data terminal is in service. The evaluation results for a mixed traffic combination (Internet-type and other packet data services) demonstrate a high efficiency and low transmission delay of such a protocol with presence of channel errors. We then proceed to propose a multi-access protocol named "Burst Reservation Multiple Access" (BuRMA) protocol capable of simultaneously accommodating RT and NRT services. Using Equilibrium Point Analysis (EPA), the performance of BuRMA is analysed. Then, we apply this multiple access protocol to the uplink direction of the DECT, half- and full-rate GSM systems while simultaneously avoiding changes in the fundamental structure of those systems (unlike the approach presented in [GOOD90]). Then, the system performance of the adapted systems, in terms of packet dropping percentage and channel throughput, is evaluated and displayed taking the feedback delay into account.
6.2 Round-Robin Multiple Access Protocols

6.2.1 Description

To retain compatibility, the burst structures of GSM (such as normal and access bursts) are re-used (see Figure 2.12). In the "Round-Robin Multiple Access" (RRMA) protocol, a GUT with packets to transmit initiates transmission by sending a random access request on the PRACH indicating its TLLI and the number of required time slots or the message length. A two-level variable permission probability scheme is deployed to offset the retransmission and resolve the collision (see Chapter 4-Section 4.4.5.2). On successful contention in the PRACH, if there are enough resources available for the terminal, it will be admitted to the serving queue. However, if the serving queue is full, two scenarios could be considered: (1) the terminal request is rejected and its burst is dropped (see Figure 6.2), or (2) the data terminal enters the access queue at the base station in which case the request reception is acknowledged by the BSS and the data terminal stops contending and waits to be admitted to the serving queue (see Figure 6.3). It will be shown in the next section that, through deploying an access queue at the base station, unnecessary contentions are avoided and the overall performance is improved.

As other reservation protocols that use conventional TDMA structures (see Figure 2.15), if the requested number of slots exceeds the remaining unreserved channels, the request is rejected and hence the remaining unused channels are wasted. In order to exploit all the traffic channel resources a Round-Robin-like system can be employed (see Figure 6.1) [KLEI76].

![Figure 6.1 Virtual Serving Queue Organisation (in this case, N=10 and n=5)]
In such configuration, the variable frame size structure is adopted to cope with the additional required slots to ensure maximum resource utilisation and eliminate any wastage of unused slots. This queue is organised and controlled by the base station. Newly arrived data users join the serving queue (providing that it has not reached its limits, \( N_{\text{max}} \)) and work their way up to the top of the queue in a FCFS fashion or based on a requested/granted priority. Finally, each terminal receives a quantum of service or more. When that quantum expires and if they need more service, they then return to the tail of that same queue and repeat the cycle. The end of a message is flagged by the user in the very last packet. Effectively, if a data customer requires more than one time slot in a frame, it doubles its service quantum (provided that the channel resource is available). Having entered the round-robin queue, data terminals are informed and updated by the control channels on their transmission patterns, i.e. the starting point and the frequency of transmission in the frame. An ARQ can be deployed in the layer three to decrease the frame error rate caused by the physical layer to the upper layers. These acknowledgements can be concatenated to the serving queue rearrangement updates on the downlink control channels.

The instantaneous Round-Robin frame duration, \( T_f \), is

\[
T_f = T_s N
\]  

(6.1)

where \( T_s \) is the time slot duration (quantum time [KLE176]) and \( N \) represents the instantaneous serving queue size that for \( m \) number of users in service, it is obtained by

\[
n = \sum_{i=1}^{m} R_i
\]  

(6.2)

where \( R_i \) is the number of reservations made by user \( i \).
Figure 6.2 Data Transfer Procedure in RRMA with Access Queue
Figure 6.3 Data Transfer in RRMA with no Access Queue
6.2.2 Performance Evaluations

In order to examine the performance of RRMA protocol, an extensive model simulating the $U_m$ interface (the communications link between the BSS and GUTs) has been developed. It is assumed that data terminals do not generate new messages as long as they have a message (burst) for transmission in buffer. Simulations have been performed using an ARQ scheme over the non-interleaved Gilbert-Elliot channel model (Appendix B) [CHAK96, BISC95]. In order to reduce the signalling load, the erroneous packets are retransmitted at the end of the burst. We assume that a PRACH frequents once every 8 time slots, and the serving queue is updated at every 4 TDMA frame (32 time slots = 18.46). Table 6.1 presents the parameters used for the performance evaluation of this protocol. Results are drawn for the two cases of RRMA with and without access queue.

**Table 6.1 System Design Spreadsheet**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Values</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slot Duration</td>
<td>$T_s$</td>
<td>0.577</td>
<td>ms</td>
</tr>
<tr>
<td>Burst Payload</td>
<td>$B$</td>
<td>75.75</td>
<td>bits</td>
</tr>
<tr>
<td>Max Frame Size</td>
<td>$N_{max}$</td>
<td>10</td>
<td>----</td>
</tr>
<tr>
<td>Max Frame Duration</td>
<td>$T_{f_{max}}$</td>
<td>5.77</td>
<td>ms</td>
</tr>
<tr>
<td>Feedback Delay</td>
<td>$D_f$</td>
<td>18.46</td>
<td>ms</td>
</tr>
<tr>
<td>Max No. Retransmission Attempts</td>
<td>$A_{max}$</td>
<td>7</td>
<td>----</td>
</tr>
</tbody>
</table>

6.2.2.1 Effect of Offered Traffic Load

For evaluation purposes a mixture of packet services is considered. This mixture includes equal portions of E-mail, Mobitex, Railway, WWW, and a $10^{th}$ of FTP packet sources (see Chapter 3). For example, a population of 41 consists of 10 E-mail, 10 Mobitex, 10 Railway, 10 WWW, 1 FTP sources. This assumption has been made because of the heavy load imposed by the FTP traffic, which would otherwise undermine the effect of other services. Figure 6.4 show the traffic load ($g_t$) generated by the given population as such. A maximum total population of 840 terminals (corresponding to a raw load of over 70 Kbytes/s) has been envisaged. In Figure 6.5, the normalised offered traffic load is shown that for the given traffic scenario it could reach up to almost 4.5 packets per time slot duration.
Figure 6.4 Offered Load Versus No. Terminals

Figure 6.5 Normalised Offered Load Versus No. Terminals

Figure 6.6-Figure 6.9 portray the effect of the normalised traffic load on the performance metrics. As shown, the system reaches its saturation point after $G_t > 1$ (see Chapter 4-Section 4.5.5).

Figure 6.6 demonstrates the effect of the access queue on the blocking probability ($P_B$). As the normalised traffic load, $G_t$, increases the probability of a traffic channel being available will be reduced. When there is no access queue in operation, the successful contention attempts will be blocked if there is no traffic channel available. This is the reason for higher blocking rate in case of RRMA with no access queue. As was expected, for normalised traffic load higher than unity (the maximum traffic channel capacity), the blocking rate increases dramatically.
The average number of contention attempts with respect to the normalised offered load is drawn in Figure 6.7. As seen, the presence of the access queue is negligible for normalised traffic loads smaller than unity ($G_t < 1$), after which the two graphs start diverting. RRMA with access queue accepts more requests even when all the traffic channels are occupied and as a result, the frame size could reach its maximum of $N_{max} = 10$.

According to Equation (6.1), the frame duration in RRMA is proportionally related to the frame size. Therefore, it is expected that the average transmission delay for the RRMA with access queue become higher as shown in Figure 6.8.
The average access delay for the two cases is depicted in Figure 6.9. The overall access delay for with-access-queue case is far lower than that with the access queue. This is not surprising since the terminals in RRMA with no access queue will be blocked after contention $A_{\text{max}}=7$ times (translating into a maximum access delay of $A_{\text{max}}*T_c=7*4.615=32.3$ ms).

In order to reveal the delay components of RRMA with access queue, the average queueing delay versus the traffic load is shown in Figure 6.10. As seen, the queueing delay reaches a maximum at high loads. This demonstrates the saturation point associated with the finite population queues (see Chapter 4-Section 4.5.5) [KLEI76, GROS85]. Closely related to the queueing delay with the Little’s formula of Equation (4.31) [KLEI75, GROS85], the queue length is presented Figure 6.11.
Figure 6.10 Average Queueing Delay Versus Normalised Offered Traffic Load

Figure 6.11 Average Queue Length Versus Normalised Offered Traffic Load

Last but not least, the effect of traffic load on the channel utilisation is presented in Figure 6.12 and Figure 6.13.

Figure 6.12 Throughput Versus Normalised Offered Traffic Load
In the normalised version of the S-G graph, we can see that the throughput approaches unity. **RRMA** is a burst-switched protocol with extra channels for random access (PRACH). The throughput of the PRACH is limited to the performance of the used random access mechanism. However, provided that an optimised access mechanism is deployed, the normalised throughput of the traffic channels could reach unity. Therefore, the high traffic channel throughput demonstrates the optimality of the used random access mechanism (i.e. a two-level permission probability in this case).

![Graph](image)

**Figure 6.13 Normalised Throughput Versus Normalised Offered Traffic Load**

### 6.2.2.2 Effect of Frame Size

The effect of the traffic load on the average frame size was studied in Figure 6.7. In this section, we concentrate on the effect of the maximum frame size on the **RRMA** protocol performance which will eventually lead to optimisation of $N_{\text{max}}$ for a given traffic scenario and the required traffic load. It is assumed that the normalised traffic load is fixed to $G_t = 7.2$. This load is generated by 100 E-mail together with 100 WWW sources.

As shown in Figure 6.14 and Figure 6.15, the maximum frame size has no impact on the performance of the random access mechanism. This was expected since the traffic load and other protocol parameters (Chapter 4-Section 4.5.1.1) are static.
Figure 6.14 No. Contention Attempts Versus Maximum Frame Size

Figure 6.15 Average Contention Delay Versus Maximum Frame Size

Figure 6.16 shows the average frame size ($N_{ave}$) with respect to the maximum possible size ($N_{max}$). The RRMA without access queue shows a large difference between $N_{ave}$ and $N_{max}$. This is due to high blocking probability of the bursts.
Closely related to the average frame size, the transmission delay ($D_t$) also grows proportionally (see Chapter 4-Equation (4.31)).

Figure 6.16  Average Frame Size Versus Maximum Frame Size

Figure 6.17  Average Transmission Time Versus Maximum Frame Size

Figure 6.18 depicts the average access delay for the two cases. As shown the maximum frame size has a large impact on the access delay for the RRMA with access queue. This decrease in access delay for larger $N_{max}$ is due to the reduction in the queueing delay.
As shown in Figure 6.19, the blocking probability is reduced dramatically with increase in the maximum number of traffic channels. However, it reaches a stable situation after $N_{\text{max}} > 9$. This point shows that the average number of active terminals, therefore, the increase in $N_{\text{max}}$ has little impact on the performance.

Finally, the raw and normalised throughputs of the traffic channels are shown in Figure 6.20 and Figure 6.21, respectively. Again, maximum frame size has great influence on the throughput of RRMA without access queue and little impact on that with access queue. This is due to the effect of blocking on the bursts (see Figure 6.19).
6.3 Burst Reservation Multiple Access Protocol\(^1\)

Here, a packet multiple access protocol, called "Burst Reservation Multiple Access" (BuRMA), is proposed for provision of RT packet services to existing circuit-switched systems [TAAG96b]. Three adapted versions of BuRMA are considered for GSM half/full rates and DECT systems while simultaneously preserving the fundamental structure of these systems (unlike the approach presented in [GOOD90]).

---

\(^1\) This protocol was originally abbreviated to BRMA [TAAG96b], however, due to a name conflict with that of [DUNL93], it was renamed BuRMA [TAAG97a].
6.3.1 Description and Modelling

We consider a two-way wireless packet network, in which the spatially dispersed mobile terminals transmit and receive packets via the base station. In the uplink, packets are transmitted using BuRMA [TAAG96b]. Since the base station is aware of all the arrived packets from the fixed network, it broadcasts the downlink stream through a contention-free protocol. As this downlink is not subject to collisions, its channel throughput is higher than the uplink (see Chapter 4-Section 4.6) [NAND94, TAAG95, WONG93]. This freed capacity can be exploited to broadcast BuRMA feedback messages from the base station. Therefore, we will only focus on the uplink multiple access. A two-level permission probability scheme is deploy to resolve collisions (see Chapter 4-Section 4.4.5.2). Figure 6.22 depicts a \((N+3)\)-state Markov chain model for BuRMA.

![BuRMA Terminal Model with Two-Level Variable Permission Prob.](image)

Each terminal can be in one of the \((N+3)\) states (here \(N\) is the total number of time slots in a TDMA frame). All transitions occur at the end of a time slot [NAND94, NAND91]. Upon arrival of a burst, the terminal moves from the silent state, \(SIL\) into the contending state, \(CON_0\) and starts contending for a reservation with a permission probability of \(P_0\). If it failed to get a reservation in the first contention attempt, it moves to the \(re-contention\) state, \(CON_r\) where it keeps contending with a different permission probability of \(P_r\). Observation shows (see Chapter 5-Section 5.) that choosing \(P_0 > P_r\)
lowers the undesirable contention periods \cite{NAND91, MITR90, TAAG96a}. The state of the system at any given time can be defined by: \( S \) (number of terminals in silence), \( C_0 \) (number of terminals in \( CON_0 \)), \( C_r \) (number of terminals in \( CON_r \)), and reservation status of time slots \((R_i: 0 \leq i < N)\). When the system is at its equilibrium, the average change in any state is expected to be zero \cite{FUKU83}. We denote the state of the system at equilibrium by the corresponding small letters \( \{ s, c_0, c_r, r_i: 0 \leq i < n \} \). As the system is frame-asynchronous \cite{NAND94}:

\[
\frac{d}{dt} s + c_0 + c_r - \sigma = 0 \tag{6.3}
\]

The equilibrium equations at states \( SIL, CON, \) and \( RES_{N-1} \) can be written respectively as:

\[
\gamma_f + c_0 + c_r - s\sigma = 0 \tag{6.4}
\]

\[
c_0 [1 - P_o u_o (1-r)](1-\gamma) - c_r P_r u_r (1-r)(1-\gamma) - c_r \gamma = 0 \tag{6.5}
\]

\[
c_0 P_o u_o (1-r)(1-\gamma) + c_r P_r u_r (1-r)(1-\gamma) + r (1-\gamma_f) - r = 0 \tag{6.6}
\]

Assuming that the total number of terminals in the system is constant for a period of time:

\[
s + c_0 + c_r + rN = M \tag{6.7}
\]

where the undefined symbols are:

\[
\sigma = 1 - e^{-\frac{T_{off}}{T_{off}}} \tag{6.8}
\]

\[
\gamma = 1 - e^{-\frac{T_{on}}{T_{on}}} \tag{6.9}
\]

\[
\gamma_f = 1 - (1-\gamma)^N \tag{6.10}
\]

and \( u_o \) and \( u_r \) are the probabilities that no other backlogged contending terminal has permission to transmit their packets, provided that the successful terminal is in the state \( CON_0 \) or \( CON_r \), respectively.

\[
u_o = (1-P_o)^{c_0-1}(1-P_r)^{c_r} \tag{6.11}
\]
As shown in [NAND91], the equilibrium point analysis for this modified protocol can be followed through solving the equation set of (6.4)-(6.12).

### 6.3.2 BuRMA for GSM

The GSM radio interface is a combination of FDMA and TDMA with Frequency Hopping (FH) [MOUL92]. Each GSM carrier is divided into eight time slots recurring in time every 0.577 ms (or to be more accurate, every $\frac{15}{26}$ ms). Therefore, the TDMA frame duration is $0.577 \times 8 = 4.615$ ms. A time slot carries a burst of traffic or signalling bits (see Figure 2.13).

In this adaptation, modifications should be kept to a minimum. A terminal accomplishes full call set-up and handover as it might do in GSM. All the control channels, i.e. broadcast, common, and dedicated ones, are left intact. In this protocol, a contending terminal uses one of the 57-bit blocks in a normal burst to send the header to the base station. An example of the header structure is given in Figure 6.23.

Apart from source and destination virtual IDs which are assigned to the terminal during the call set-up, a 10-bit NDP field is also added. The NDP indicates the number of packets that the terminal has dropped during the contention period. This number will help the interpolation and reconstruction of lost part of a burst at the other end of communication network.

Once a mobile station succeeds in gaining a reservation, there is no need to retransmit the header, as the base station already knows which terminal in its coverage has reserved the channel. The terminal releases its reservation by flagging the end of the
burst. For packet voice transmission, we can exploit the silence speech frame used for interleaving with the first speech frame of a talkspurt in order to accommodate the header as well as avoiding speech packet dropping. Speech frame interleaving in GSM provides a look-ahead advantage, i.e., the speech terminal requires a delay of 20 ms before it can produce speech bursts. Using such a delay, the terminal can start contending on the arrival of talkspurt even though no speech burst is yet produced.

6.3.3 BuRMA for DECT

In DECT, as displayed in Figure 6.24, the uplink and downlink are duplexed in time on the same carrier (TDD).

![DECT Channel Organisation and Packet Structure](image)

Figure 6.24 DECT Channel Organisation and Packet Structure

Each TDMA frame in DECT consists of 12 uplink and 12 down-links time slots ($N=12$) TDDed (!) in a frame duration of $T_f = 10$ ms. Likewise, we adapt BuRMA for DECT as described above. All the channels apart from the uplink channel are unchanged. BuRMA is applied, in the uplink. Again, the header packets are sent only during the contention. A header structure similar to Figure 6.23 might be also used for the DECT adaptation. In addition, for speech interpolation purposes, an NDP will be included in the header. The feedback packet of a time slot is broadcast from the base station in the corresponding downlink which is delay by half a TDMA frame duration.
6 Cellular Packet Communication

Table 6.2 System Design Spreadsheet

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Full-Rate GSM</th>
<th>Half-Rate GSM</th>
<th>DECT</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel Rate</td>
<td>( R_c )</td>
<td>270.83</td>
<td>270.83</td>
<td>1152</td>
<td>Kb/s</td>
</tr>
<tr>
<td>Source Rate</td>
<td>( R_s )</td>
<td>13</td>
<td>6.5</td>
<td>32</td>
<td>Kb/s</td>
</tr>
<tr>
<td>Frame Duration</td>
<td>( T_f )</td>
<td>4.615</td>
<td>9.23</td>
<td>10</td>
<td>ms</td>
</tr>
<tr>
<td>No. Slots</td>
<td>( N )</td>
<td>8</td>
<td>16</td>
<td>12</td>
<td>----</td>
</tr>
<tr>
<td>Delay Limit</td>
<td>( D_{max} )</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>ms</td>
</tr>
<tr>
<td>Talkspurt Duration</td>
<td>( T_{on} )</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>ms</td>
</tr>
<tr>
<td>Silence Duration</td>
<td>( T_{off} )</td>
<td>1350</td>
<td>1350</td>
<td>1350</td>
<td>ms</td>
</tr>
<tr>
<td>Hangover Duration</td>
<td>( T_h )</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>ms</td>
</tr>
<tr>
<td>Feedback Delay</td>
<td>( D_f )</td>
<td>1.731</td>
<td>1.731</td>
<td>5</td>
<td>ms</td>
</tr>
</tbody>
</table>

6.3.4 Performance Evaluation

Table 6.2 provides a list of system parameters used for numerical analysis. Since voice terminals require prompt delivery of packets, they discard packets delayed over a certain limit, say \( D_{max} \). This limit is determined by the end to end delay budget. In GSM and DECT, the organisation of the uplink direction is derived from the downlink by applying a constant delay, which is known as half-duplexing [MOUL92, WALL90]. As discussed in Chapter 4-Section 4.8.2), this delay could cause additional congestion as contending terminals could not be aware of the result of their contention attempts (or reservation status). It is important to point out that the performance of the proposed air interface is presented on single carrier basis as the overall system capacity is simply the capacity of a single carrier multiplied by the number of carriers (provided that carriers are fully orthogonal). Throughout this section a non-interleaved Gilbert-Elliot statistical channel model has been used (see Appendix B). The two-level permission probabilities have been optimised at every stage for the best performance.

6.3.4.1 Full-Rate GSM Case

Figure 6.26 represents packet dropping percentage of BuRMA for the full-rate GSM against the number of mobile stations. The hangover duration is chosen equal to the speech frame of the speech codec. The feedback is delayed by 3 time slot periods, that is \( 3 \times 0.577 = 1.731 \) ms (see Table 6.2).
As shown in Figure 6.25, the system can accommodate up to 13 terminals while satisfying $P_{\text{drop}} < 0.01$ (a capacity of $13/8 = 1.625$). In Figure 6.26, the channel throughput of the multiple access is displayed with zero frame hangover. In this case (full-rate GSM), the system can utilise up to 63% of the channel resources at the operational point (number of voice terminals equal to 13).

6.3.4.2 Half-Rate GSM Case

In half-rate GSM, the speech channel bit rate is reduced from 22.8 Kb/s to 11.4 Kb/s. The result will be that only every other time slot is necessary for one user and the other one becomes available to another terminal. It can be assumed effectively that the
number of time slots is doubled (see Table 6.2). Again the downlink is delayed by 3 time slot periods. The performance of BuRMA for the half-rate GSM is portrayed in Figure 6.27 (hangover frame duration is 20 ms again). As seen, the system can serve 29-30 simultaneous conversations with a packet dropping limit of 1% (corresponding to a capacity gain of 1.81-1.88, depending on the hangover duration).

![Figure 6.27 BuRMA Performance for Half-Rate GSM](image)

Figure 6.27 BuRMA Performance for Half-Rate GSM

Figure 6.28 demonstrates the channel throughput of BuRMA with the half-rate GSM parameters (with no frame hangover). It is shown that the system exploit up to 75% of the channel resources while simultaneously keeping the average packet dropping to a minimum of 1% (up to 29 voice users).

![Figure 6.28 Channel Throughput of BuRMA for Half-Rate GSM](image)
6.3.4.3 DECT Case

The packet dropping percentage of BuRMA with the DECT transmission system appears in Figure 6.29.

![Figure 6.29 BuRMA Performance for DECT](image)

DECT employs the ADPCM speech, and therefore, we consider the hangover frame duration equal to its TDMA frame duration. The downlink delay is half the TDMA frame duration, 5 ms (see Table 6.2). As displayed in Figure 6.29, the system accommodate 21 simultaneous conversations with $P_{\text{drop}} < 0.01$ (21/12=1.75 users per channel). Figure 6.30 presents the channel utilisation of the protocol with the DECT GSM parameters (with no frame hangover). As seen, voice packets successfully occupy up to 70% of the radio channel resources in the system operational area (number of simultaneous conversations less than 22).

![Figure 6.30 Channel Throughput of BuRMA for DECT](image)
6.4 Conclusions

In this Chapter, we have examined provision of packet-oriented services for cellular systems. This provision could be complementary to the existing circuit-switched systems.

A novel MAC layer protocol, named "Round-Robin Multiple Access" (RRMA), for GPRS has been proposed. The performance of the MAC layer has been examined in presence of channel impairments (using a Gilbert-Elliot error model) for various mixed services scenarios including E-mail, Mobitex, Railway, FTP, and WWW (see Chapter 3). A wide range of performance metrics have been evaluated with respect to offered load and frame sizes. It has been shown that deploying an access queue at the base station could improve the overall protocol performance in terms of blocking probability, traffic channel throughput, contention delays, and protocol stability. It has been shown that RRMA with access queue is insensitive to the maximum frame size. The RRMA protocol shows very high throughput and hence capacity. These improvements, however, are achieved at the cost of adding more complexity to the frame structure management (to be handled by the base station). With the help of normalised traffic load developed in Chapter 4-Section 4.5.2, a set of normalised graphs have been presented in this Chapter which could be used to design an optimum protocol parameter for the given traffic scenario and the required QoS.

Also, a packet random access protocol, called "Burst Reservation Multiple Access" (BuRMA) protocol, has been proposed for RT cellular packet communications [TAAG96a]. An analysis of this protocol has been carried out using the EPA analytical method. The BuRMA protocol has then been adapted for the full-, half-rate GSM, and DECT systems for burst-switching operation. As shown, the capacity of GSM and DECT systems can considerably be improved by deploying BuRMA in uplink carriers in presence of a delayed feedback channel. Performance evaluations for packet voice service result in capacity improvements of 62.5%, 81-88%, and 75% over the circuit-switched DECT, the full-, and half-rate GSM, respectively. BuRMA for the half-rate GSM achieves the highest capacity gain amongst the other two. This was expected as statistically multiplexing protocols perform better in higher bandwidths (recalling that half-rate GSM provides the largest number of time slots) [TAAG95, GOOD91]. It can
be observed that hangover duration does not effect the system performance significantly. Although, on one hand, hangover reduces the talkspurt rate resulting in lower packet dropping, on the other hand, it increases the speech talkspurt duration (i.e. speech activity) resulting in higher packet loss. Since channel throughput is defined as the probability that the channel is successfully occupied, the throughput of a circuit-switched protocol for packet voice is equal to the voice activity factor (in our study, 43%). As seen, BuRMA increases the channel utilisation up to 75% while satisfying a maximum packet dropping of 1%. Although BuRMA was originally devised to exploit a portion of the existing circuit-switched resources, this protocol can be extended towards a smooth transition from circuit- to purely packet-switched radio networks.
Satellite Packet Communications

7.1 Introduction

Satellite systems represent the extreme case in terms of differential and round-trip delays. The use of packet-based protocols such as R-ALOHA, in satellite environments have been limited to NRT data services due to the long round-trip delays associated with the satellite environment [TASA84]. Long round-trip latency of satellite systems could then exceed the delay constraints of RT services such as voice, and video telephony (depending on the satellite altitude and on-board processing availability). In this chapter, concepts for designing multiple access protocols in presence of long acknowledgement delays are proposed. Similar to the terrestrial counterpart, the “Satellite-GPRS” (S-GPRS) is proposed to complement low rate circuit-switched services of constellation systems such as Iridium and ICO. The access mechanism used in this protocol allows the terminal to contend non-stop until a positive acknowledgement is received. Further work presents a novel hybrid multiple access protocol called “Burst Reservation CDMA” (BR-CDMA) which makes RT and NRT bursty packet transmissions possible. Inspired by the hybrid TD-CDMA protocol of the FRAMES project, we highlight the rationale behind using protocols as such. In the proposed protocol, bursty packet terminals are not required to contend for access. Instead, a terminal selects the most suitable time slot and the spreading sequence based on the delayed feedback packet is broadcast once every frame duration. These protocols are analysed, optimised and evaluated for a wide range of packet traffic mixtures. These protocols are applied to ICO and Iridium constellations.
### 7.2 Non-Geostationary Satellite Constellations

1998 was a banner year for the mobile satellite industry. With Orbcomm, Iridium and Globalstar leading the way in terms of the number of satellites launched, an array of services and products are coming onto the market. The first generation Satellite Personal Communication Systems (S-PCS), such as Iridium [IRID] and ICO [ICO], will mainly offer circuit-switched voice and low-rate data communications. Figure 7.1 presents the constellation information of some of the main first generation MSSs.

![Figure 7.1 First Generation Satellite Constellations](image)

As described in Chapter 2, there are presently few packet satellite networks in operation for experimental purposes, such as SATNET, TACNET and WESTAR. In the short to medium term, the S-PCS architecture is likely to remain based on circuit-switched technology. Appendix D provides a brief review of forthcoming S-PCNs.

### 7.3 Satellite GPRS

This Section proposes a novel scheme for data packet applications for S-PCN, an integral part of future global communication systems. In order to keep compatibility
with the terrestrial packet sub-system counterpart proposed for GSM/DCS, i.e., GPRS, we follow the same approach for the Satellite-GPRS (S-GPRS). The concept of GSM-GPRS can be applied to the satellite communications through the use of a packet data sub-network, in order to provide efficient packet data links.

There are certain characteristics associated with satellite communication systems which impact the performance of any such scheme. Amongst these the most significant are the long round-trip-delays associated with satellite communications. Such long round-trip delays requires specific access protocols for initial access request, otherwise, the access delay become intolerable.

The proposed scenario can be applied to any TDMA-based mobile satellite constellation. However, in order to examine the MAC protocol, we have chosen a representative example of MEO satellite system, ICO, and a representative of LEO, Iridium, to evaluate the proposed packet sub-system performance. The ICO constellation system is designed based on the GSM architecture, and therefore an inevitable evolution of ICO will be to join packet data transmission to the system as with GPRS for GSM. It will be shown that if the protocol performs acceptably for ICO, it will work efficiently for lower altitude constellations such as Iridium. A multi-slot reservation-based multiple access protocol is specifically tailored for lengthy round-trip delays and long delay-spread. Then, this MAC layer is examined for the Iridium-like and ICO constellation systems. The latter represents the worst case amongst the proposed non-GEO constellations in terms of propagation delay.

Similar to GPRS, in S-GPRS a terminal registers with the S-GPRS sub-network by accomplishing a fast call set-up (GPRS registration). It then stays in the STANDBY state until packet data communication is required. In the STANDBY mode, the terminal periodically scans the corresponding CCS broadcast channels in order to select the appropriate channels within the given spotbeam and to read the paging channel for the incoming communication.

When communication is required, the S-GUT logs onto the network and moves into the STANDBY mode. During this procedure an identity (TLLI), a shortened
version of the \textit{S-GUT} IMSI, will be assigned to the user terminal. The TLLI is then used
the by either the \textit{S-GUT} or the LES at communication initiation phase.

In case of the user initiated data transfer, the \textit{S-GUT} moves into the
\textsc{Contension} mode where it competes with other users within the same mode for
reservation of PTCHs through a random access mechanism. Once the LES receives the
request for reservation from a \textit{S-GUT} successfully, a paging message, identifying the
address and the amount of the resources, is then forwarded to the corresponding user
terminal. However, if resources are not available, the user will enter a FCFS queue at
the LES. A paging message is then transmitted, preventing the user from further
contention.

\subsection*{7.3.1 Proposed MAC Layer for S-GPRS}

\textit{GPRS} relies on a reservation-based multiple access protocol. However, as mentioned
earlier, reservation access protocols suffer from access delay due to the random access
(ALOHA-like) procedure associated with schemes as such [TAAG97a, TAAG97b,
HÄMÄ95b, CHAK96, HONK94]. In reservation-based protocols, a contending terminal
should hence be acknowledged of its contention attempts. Under circumstances where
the delay in receiving the acknowledgement packet is high, modified protocols are
needed in order to prevent excessive access delay. Satellite environments are generally
associated with long round-trip delays particularly on the bent-pipe or transparent
satellite systems where no On-Board-Processing (OBP) is envisaged. Figure 7.2-Figure
7.3 show average single-hop propagation delays for the ICO and Iridium constellations.
The minimum acknowledgement packet delay for a given elevation angle (without
OBP) is simply four times the propagation delay shown below.
Therefore, one can conclude that implementing GPRS for constellations such as Iridium [IRID] (LEO with OBP) will be less problematic than that of ICO [ICO] (MEO without OBP). The GPRS MAC layer can be divided into two main sections, access mechanism and reservation procedure. For the access mechanism, as is known, the Slotted-ALOHA protocol performs more efficiently than Pure-ALOHA. However, use of S-ALOHA for satellite is limited due to the long path delay difference between the inner and outermost points in a satellite spotbeam (see Figure 7.4 and Figure 7.5).
This long differential delay imposes a long Guard Period (GP) which could exceed the duration of a couple of time slots. For example, as shown in Figure 7.4 the differential path delay for spotbeam no. 100 of ICO (see Figure 7.4, Figure 7.6 and Figure 7.7) could reach up to 6.5 ms, and hence a required GP of 13 ms. However, the necessary GP for inner spotbeam is much lower.
In order not to design the random access protocol for the worst case which would result in low PRACH performance, the LES can divide the PRACH slots into different sizes for inner and outer spotbeam. The LES can broadcast the PRACH slot size of a spotbeam through the corresponding PBCCH. This algorithm can even be used for circuit-switched services (voice) through exploiting the spare bits in the BCCH. Despite waste of some satellite channel resources though GP, long differential delays and use of
S-ALOHA might can give a rise to a phenomenon called time capture effect; assuming that PRACH packets are short in length, it is possible that two packets transmitted in the same channel from two geographically separated terminals could be captured by the receiver [HE97].

Due to the long acknowledgement delay, the S-GUT cannot wait for the acknowledgement of every contention attempts. Therefore, in the proposed protocol here, we assume that the S-GUT continues to contend in the PRACH until it receives a paging from the LES in the PAGCH. This paging message contains reservation information if required resource is available or queueing information when the resources are all occupied. In the latter case, the S-GUT will stop contending and would be paged as soon as available resources are ready.

During the contention period, the S-GUT deploys a back-off algorithm to deal with collisions. We use a two-degree variable permission probability, where the contending terminals transmit with higher probability at their first contention attempt [TAAG96b]. The mobile-initiated data transfer procedure is depicted in Figure 7.8.
7.3.2 Performance Evaluations

For evaluation purposes, we have considered a mixture of equally populated packet services including Mobitex, Railway, E-mail, and WWW (see Chapter 3). The FTP
service has been excluded as it would generate heavy traffic load and undermine the rest of the services. Furthermore, we make the following assumptions:

- It can realistically be assumed that the S-GUT is static due to high speed of the satellite in comparison with that of the S-GUT,
- Power capture at the receiver is not considered,
- The use of half rate convolutional coding is envisaged,
- S-GUTs are uniformly distributed within a spotbeam,
- In order to reduce the access to the PRACH, there is an access queue at the base, erroneous in CONTENTION is interpreted as collision for the terminal and in TRANSMISSION mode results in retransmission of a data packet,
- The propagation channel impairments are modelled by a non-interleaved Gilbert-Elliot statistical channel model (see Appendix B).

Table 7.1 summarises the other parameters used in this study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>LEO66</th>
<th>ICO10</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel Rate</td>
<td>$R_c$</td>
<td>50</td>
<td>36</td>
<td>Kb/s</td>
</tr>
<tr>
<td>Coding Rate</td>
<td>----</td>
<td>1/2</td>
<td>1/2</td>
<td>----</td>
</tr>
<tr>
<td>Slot Duration</td>
<td>$T_s$</td>
<td>6.67</td>
<td>6.67</td>
<td>ms</td>
</tr>
<tr>
<td>Frame Duration</td>
<td>$T_f$</td>
<td>40</td>
<td>40</td>
<td>ms</td>
</tr>
<tr>
<td>Packet Payload</td>
<td>$B$</td>
<td>166.75</td>
<td>120</td>
<td>bits</td>
</tr>
<tr>
<td>Max. Retrans. Attempts</td>
<td>$A_{\text{max}}$</td>
<td>15</td>
<td>15</td>
<td>----</td>
</tr>
<tr>
<td>Permission Probabilities</td>
<td>$P_0, P_r$</td>
<td>optimised</td>
<td>optimised</td>
<td>----</td>
</tr>
<tr>
<td>No. Slots</td>
<td>$N$</td>
<td>8</td>
<td>6</td>
<td>----</td>
</tr>
<tr>
<td>Min Elevation Angle</td>
<td>$E_{\text{min}}$</td>
<td>8</td>
<td>10</td>
<td>degrees</td>
</tr>
<tr>
<td>Min Feedback delay</td>
<td>$D_{f,\text{min}}$</td>
<td>10</td>
<td>145</td>
<td>ms</td>
</tr>
<tr>
<td>Max Feedback Delay</td>
<td>$D_{f,\text{max}}$</td>
<td>33</td>
<td>172</td>
<td>ms</td>
</tr>
<tr>
<td>Altitude</td>
<td>$A_c$</td>
<td>780</td>
<td>10355</td>
<td>Km</td>
</tr>
<tr>
<td>No. Spotbeams</td>
<td>NOSB</td>
<td>48</td>
<td>163</td>
<td>----</td>
</tr>
<tr>
<td>No. Terminals</td>
<td>$M$</td>
<td>variable</td>
<td>variable</td>
<td>----</td>
</tr>
</tbody>
</table>

The offered traffic load generated by these services are drawn in Figure 7.9 with respect to total population size. The normalised version of the offered load is depicted in Figure 7.10 (see Chapter 4-Section 4.5.2).
Figure 7.9 Offered Traffic Load Versus Total No. Terminals

Figure 7.10 Normalised Offered Load Versus Total No. Terminals

Figure 7.11—Figure 7.15 demonstrate the performance of ICO S-GPRS with respect to the offered load. The performance metrics are shown for the minimum (worst-case) and the maximum (best case) elevation angles of ICO (see Table 7.1). As is expected, the system in its minimum visibility (maximum delay, i.e. $E=10$) has the worst performance.

Figure 7.11 shows the average number of contention attempts before succeeding to access to the access queue. Because of propagation delays, in the minimum elevation angle, S-GUTs are required to access the PRACH more often. Due to long feedback delay, terminals are required to access considerable times even in low traffic load.
Figure 7.11 No. Contention Attempts Versus Normalised Offered Traffic Load (ICO)

Figure 7.12 shows the average access queue length versus the normalised traffic load. As the propagation delay increases the collision rate increases and fewer terminals can succeed the access mechanism. Therefore, the case with $E=90$ has a larger access queue size. It can be seen that after $G_t=1$ the access queue size starts rising sharply. The indicates the saturation of the system (see Chapter 4-Section 4.5.5).

Figure 7.12 Average Queue Length Versus Normalised Offered Traffic Load (ICO)

Similarly to the access queue size, the access delay (see Figure 7.13) starts rising after the point of $G_t=1$. Recalling that the access delay consists of contention, feedback, and queueing delay components (see Chapter 4-Section 4.5.1.2).
Figure 7.13 Average Access Delay Versus Normalised Offered Traffic Load (ICO)

Figure 7.14 and Figure 7.15 show the raw and the normalised throughput of the systems. As can be seen, the throughput for the two elevation angles are fairly close. The system's saturation can be observed in the normalised graph where the throughput of the traffic channels approaches unity for normalised offered traffic loads of $G_t > 1$.

Figure 7.14 Throughput Versus Offered Traffic Load (ICO)
Similarly, Figure 7.16-Figure 7.18 show the performance of S-GPRS for the Iridium constellation with the same mixed traffic scenario. In the case of Iridium, the propagation delay is much lower and number of traffic channels are assumed to be larger than those of the ICO (see Table 7.1). Therefore, we expect better performance measures, i.e., lower access delay, fewer contention attempts, smaller queue length, and higher traffic channel throughput.

Figure 7.16 No. Contention Attempts Versus Normalised Offered Traffic Load (Iridium)
Figure 7.17 Ave. Queue Length Versus Normalised Offered Traffic Load (Iridium)

Figure 7.18 Ave. Access Delay Versus Normalised Offered Traffic Load (Iridium)

Figure 7.19 Throughput Versus Offered Traffic Load (Iridium)
7.4 Burst Reservation CDMA for Mixed Services (BR-CDMA)

A novel reservation-based CDMA multi-access protocol is proposed and its applicability for real-time voice and non-real time data communications in the satellite environment is investigated. It is shown that despite the long round-trip delays associated with satellite communication systems, the proposed protocol maintains the inherent advantages of the reservation-based protocols. This is mainly achieved through the use of a simultaneous multi-reservation scheme, whereby a user terminal would not need to delay transmission of real-time services through not having to solely rely on reception of the acknowledgement message. As the proposed protocol is packet-based, data terminals with a wide range of capabilities can be easily integrated into the system. The proposed protocol is compared with the hybrid TDMA/CDMA proposal of FRAMES, a candidate UMTS access scheme. Results show increased flexibility and capacity over the latter under similar conditions.

7.4.1 Hybrid TD-CDMA

Hybrid TDMA/CDMA (TD-CDMA) schemes such as FRAMES FMA1 [ANDE97], selected as one of the terrestrial UMTS multiple access technologies for the unpaired band by the ETSI SMG4 forum, are best examples of such protocols.
Under such a scheme, the user will transmit a wide-band CDMA signal in the TDMA slots. Hence more than one user can transmit in the same slot. Furthermore, depending on the service requirements (bit-rate, RT or NRT) more than one CDMA spreading code can be allocated to any single user in a given time slot, providing flexible multi-rate capabilities.

The signals from different users are burst synchronous, that is timing advances, synchronisation channels and procedures have been incorporated in order to meet all the TDMA synchronisation needs. Nevertheless, the uplink CDMA burst from different user within a given time slot are not chip synchronous, unlike the conventional IS-95 standard, any single user would be able to set up several services such as fax and voice simultaneously, or access higher data rates through the use of multiple spreading sequences in a given time slot. In fact a single user transmitting on all the codes within a given time slot, achieves synchronous transmission which effectively reduces the Multiple Access Interference (MAI) level, thus improving the overall system performance and capacity.

The use of such a scheme is particularly beneficial for efficient resource management of multimedia services as the uplink and downlink traffic loads associated with multimedia services are usually asymmetric. Any system utilising such an access scheme can operate in Time Division Duplex (TDD) mode, eliminating the need for separate uplink and downlink frequencies. The terminal complexity is likely to be reduced as unlike pure CDMA, there is no need for full-duplexed terminals regardless of the transmission mode, TDD or FDD. In addition to the above, many inherent characteristics of the CDMA such as, frequency reuse factor of 1, robustness in multipath environments, soft capacity and soft handover can also be fully exploited in TD-CDMA scheme. Under this scheme, three different types of burst, i.e. the traffic burst, synchronisation burst and the access burst, are considered [NIKU97]. Figure 7.21 shows the structure of one such burst, i.e., the traffic burst.

<table>
<thead>
<tr>
<th>Information Symbols</th>
<th>Training Sequence</th>
<th>Information Symbols</th>
<th>Guard Period</th>
</tr>
</thead>
</table>

Figure 7.21 A Typical TD-CDMA Traffic Burst Structure
Each burst consists of two data fields separated by the training sequence field. Guard bits are tagged to the end of each burst. The length of the training sequence depends on the operational environments, and that of the guard period, on the size of the cell and the delay estimation uncertainty. Due to the strong presence of multi-path and echoes in terrestrial environments, the size of the training sequence field is much larger than is needed for the satellite case.

A link quality control procedure based on a combination of a $E_b/N_0$ estimation and other link quality indicators such as the packet error rate could also be used to maintain the desired QoS [TAAG94].

### 7.4.2 Burst Reservation CDMA Protocol

The TDMA/CDMA hybrid protocol introduced in Chapter 5 is used here as the foundations and a reference point to which relevant performance measures can be compared with.

In this section, an alternative hybrid multiple access protocol based on BuRMA [TAAG96b] and CDMA is proposed. We call this protocol Burst Reservation CDMA (BR-CDMA). In BR-CDMA, a frame consists of equal-sized time slots. In each time slot, bursts of different sources (or of the same source) are multiplexed with different spreading codes. Here a time slot and a spreading code together form a physical channel (PCH).

Here, we consider two types of sources, namely bursty and random (see Chapter 3). An RT bursty source drops, packets delayed over a certain threshold, $D_{\text{max}}$, while it is attempting for access [GOOD91].

Packet multiple access protocols heavily rely on the downlink (forward-link) acknowledgement messages informing terminals of their reservation status (see Chapter 4-Section 4.8.2) [TAAG97a, GOOD91, DEVI93]. Due to long propagation delays over satellite communication channels, these acknowledgement-based protocols are not in fact efficient for delay-sensitive services [TAAG98a, HE97]. To overcome this problem and to provide QoS assurance, when a RT bursty source goes to the ON-STATE, it starts accessing the channel through a simultaneous multi-reservation procedure rather
than contention. In this procedure, the terminal reads the frame status packet (FSP), sent once at the end of each frame. The FSP indicates the spreading codes in use tailed by the user’s ID, therefore, enabling the terminal to recognise available PCHs (when accessing) or to check its reservation status (if reserving). As real-time services cannot tolerate long delays due to acknowledgements, upon arrival of a talkspurt, the terminal temporarily reserves $R_t \geq 1$ number of available PCHs from the pool of free PCHs at random. After a period of time (temporary reservation time), the terminal expects the acknowledgement of its temporary reservation in the FSP. Upon reception of positive acknowledgement, the terminal releases all the temporary reservations but one to keep as permanent. The reservation is released by the terminal in the very last packet of its burst. If it does not receive the acknowledgement, it assumes that its temporary reservation has been lost or collided. Therefore, the terminal is required to repeat the procedure, and in the meantime, similar to all random reservation protocols, it would have to drop the over-delayed packets of RT services. The state transition diagram of a RT-bursty source for such a procedure is shown in Figure 7.22.

![State Transition Diagram of an RT-Bursty Source](image)

**Figure 7.22 State Transition Diagram of an RT-Bursty Source**

As a random source generates occasional packets, it does not require reservation. Therefore, a random source can transmit its packets in the contention phase as modelled in Figure 7.23. Using the FSP, a PCH is selected randomly from the available PCHs. The user should receive an acknowledgement for each transmitted packet within a given time window through the FSP. If an acknowledgement is not received, the packet is assumed lost and re-transmission could take place.
For higher data rate sources, however, it is more appropriate to adopt a reservation-based scheme [GOOD91]. As this type of services are not delay sensitive, the reservation procedure can be separated from the transmission process. Under such a scheme, the user would request the full required bandwidth (number of PCHs) from the network. The maximum bandwidth available (up to the requested value, preferably within the same time slot) will then be allocated to the user before transmission is commenced, similar to the GPRS protocol [TAAG98a, BRAS97]. This model is presented in Figure 7.24.

As shown in [TAAG97a, GOOD91, TAAG97], reservation-based protocols perform best in high bandwidths, where the statistical multiplexing gain is maximised (see Chapter 4-Section 4.4). In the TDMA access scheme, however, the total number of physical channels per frame is limited to the total number of time slots. The total number of possible time slots on a given frequency and bandwidth is quite limited due to intersymbol interference and the required guard period. On the other hand, in the hybrid BR-CDMA and TD-CDMA, the number of PCHs per frequency band is greater than that of the TDMA case. This boosts the multiplexing gain as the number of PCHs per frequency in each slot is greater than that of the TDMA only case.
7.4.3 Performance Evaluation

7.4.3.1 Simulation Model
For evaluation purposes, we consider two types of services for the BR-CDMA protocol: PV and Railway data model (see Chapter 3). QPSK modulation is used and the protocol is applied only to the uplink or the return-link (the most critical link). Convolutional (1/2 rate, constraint length=7) and BCH (1/2 rate) block coding schemes were deployed for voice and data services, respectively. The use of maximal length sequences were assumed. Using similar approaches to that of [BRAN96, VALA97], the achievable BER and packet success rates in Additive White Gaussian Noise (AWGN) channel were calculated. A typical required uplink Eb/No of 5 dB was assumed as a representative case. As for the purpose of comparison, perfect power control has been assumed in order to keep the computation complexity to a minimum. In asynchronous CDMA, the capture effect can be employed in order to improve the contention performance by allowing more than one terminal to use the same code. In here, however, this performance gain was not considered. There are fixed number of codes available in all time slots. Furthermore, non-regenerative payloads were considered, increasing the round-trip delay. For comparison purposes we maintained the FMA1 frame structure, however, the parameters would need further optimisation for the satellite environment (since the FRAMES protocols are designed for terrestrial systems). Moreover, user terminals are assumed to be within a small spotbeam, hence, the difference in their propagation delays is insignificant. The rest of parameters used for the simulation model are presented in Table 7.2.

7.4.3.2 Numerical Results
Figure 7.25-Figure 7.26 demonstrate the throughput and packet dropping ratio of the voice-only system for different round-trip propagation delays. As seen, satisfying the 1% packet dropping limit, the BR-CDMA can accommodate up to 91 and 101 voice terminals for round-trip delays of 33 ms and 10 ms, respectively (i.e. multiplexing gains of 1.42-1.58 over FMA-1). In BR-CDMA, packet dropping is caused by either contentions or collisions between terminals during their temporary reservation. In BR-CDMA, contention occurs when two or more terminals try to resume their temporary
reservations at the same time on the same PCH. Channel utilisation (throughput) is defined as

\[
S = \frac{\text{Ave. no. of successful transmissions in a frame}}{\text{Total no. of possible PCHs in a frame}}
\]

Since voice terminals make temporary reservations and start transmission instantaneously, as observed, the performance measures are relatively insensitive to the propagation delay.

It has also been noted that by increasing the maximum tolerable voice delay \(D_{\text{max}}\) from 30 ms to 60 ms, multiplexing gains of up to 1.625 and 1.45 for propagation delays of 33 ms and 10 ms can be achieved, respectively.

---

**Table 7.2 System Design Spreadsheet**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Values</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>(W)</td>
<td>1.6</td>
<td>MHz</td>
</tr>
<tr>
<td>Channel Rate</td>
<td>(R_c)</td>
<td>2.167</td>
<td>Mchip/s</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>(R_s)</td>
<td>8</td>
<td>Kb/s</td>
</tr>
<tr>
<td>Packet Payload</td>
<td>(B)</td>
<td>166.75</td>
<td>bits</td>
</tr>
<tr>
<td>Coding Rate</td>
<td>(T_f)</td>
<td>4.615</td>
<td>ms</td>
</tr>
<tr>
<td>Frame Duration</td>
<td>(S)</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Number of Slots</td>
<td>(N)</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Spreading Factor</td>
<td>(N_c)</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>No. Temp. Reservation</td>
<td>(R_t)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Delay Limit</td>
<td>(D_{\text{max}})</td>
<td>30</td>
<td>ms</td>
</tr>
<tr>
<td>Talkspurt Duration</td>
<td>(T_{on})</td>
<td>1410</td>
<td>ms</td>
</tr>
<tr>
<td>Silence Duration</td>
<td>(T_{off})</td>
<td>1740</td>
<td>ms</td>
</tr>
<tr>
<td>Min Elevation Angle</td>
<td>(E_{\text{min}})</td>
<td>8</td>
<td>degrees</td>
</tr>
<tr>
<td>Min Feedback Delay</td>
<td>(D_{f,\text{min}})</td>
<td>10</td>
<td>ms</td>
</tr>
<tr>
<td>Max Feedback Delay</td>
<td>(D_{f,\text{max}})</td>
<td>33</td>
<td>ms</td>
</tr>
<tr>
<td>Altitude</td>
<td>(A_c)</td>
<td>780</td>
<td>Km</td>
</tr>
<tr>
<td>No. Voice Terminals</td>
<td>(M_v)</td>
<td>variable</td>
<td></td>
</tr>
<tr>
<td>No. Data Terminals</td>
<td>(M_d)</td>
<td>variable</td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.25 Voice Sub-System Throughput Vs. Number of Terminals

![Voice Sub-System Throughput Vs. Number of Terminals](image1)

Figure 7.26 Packet Dropping Ratio versus Number of Terminals

(PV+Railway)

![Packet Dropping Ratio versus Number of Terminals](image2)

Figure 7.27-Figure 7.28 show the performance of the system for a mixed voice/data services (number of voice terminals is fixed to 90, i.e. $M_v = 90$). Permission probability for data terminals has been optimised. As a data terminal sends a channel request packet (through ALOHA-like mechanism) and waits to receive the acknowledgement, voice terminals are not affected much by the data traffic. Transmission time (Figure 7.27) comprises of two components: the access delay and the serving time (i.e. the duration needed to transmit the multi-packet message after the access-grant). Unlike the voice-only system, the propagation delay has direct impact on
transmission times of data terminals, since users are bound to wait for the acknowledgement of each contention attempts they make.

![Graph](image1)  
**Figure 7.27 Average Trans. Time Vs. Number of Railway Terminals (Mv=90)**

For the same reason, as shown in Figure 7.28, the channel throughput for the higher propagation delay is lower.

![Graph](image2)  
**Figure 7.28 Total Throughput Vs. Number of Railway Terminals (Mv=90)**

Simulation results show a fairly constant BER in the order of $10^{-6}$. This is due to the fact that the maximum number of users is limited by the MAC layer only and levels of multiple access interference stay well below the acceptable threshold at all times.
7.5 Conclusions

In this Chapter, after a review of the emerging and existing S-PCN constellation, two packet access protocols have been devised for data and RT packet communications over long-delayed satellite channels. The random access mechanisms proposed here rely less on the feedback acknowledgement.

For the satellite data transmission, the application of GPRS to S-PCN was investigated. Using the proposed access mechanism, terminals keep contending until a positive acknowledgement has been received. It is concluded that S-GPRS will perform more efficiently for lower altitude constellations, i.e. LEOs (such as Iridium), as opposed to MEOs (such as ICO). Evaluations have been carried out for a mixture of packet services. The results are generally comparable to that of the terrestrial GPRS.

The normalised traffic load once again has helped to predict the performance metrics. In general, low elevation angles result in longer access delay and slightly lower channel throughput. However, the N-GEO satellite constellations provide satellite diversity, which in turn increases the average elevation angle resulting in lower access delays. From the above, it can also be concluded that lower altitude satellites will achieve lower access delay time and higher capacity. Furthermore, the use of full-duplex terminals would enhance the performance of GPRS over satellite at the cost of increased terminal complexity. The long round-trip delays of such systems was found to be the main limiting factor.

Also, a dynamic multiple access protocol, called "Burst Reservation-Code Division Multiple Access" (BR-CDMA), has been proposed for RT packet services. This robust-to-propagation-delay protocol can accommodate three main service types, posing as a burst-switched protocol for high bit rate or RT services or as a packet-switched protocol for low bit rate applications. The protocol was then evaluated and optimised for a Iridium-type constellation. The proposed BR-CDMA protocol, fully exploits the CDMA characteristics to its benefit while maintaining the flexibility features of reservation protocols such as statistical multiplexing. Although the proposed protocol is capable of supporting multi-reservation high bit rate services (GPRS-like services), we limited the evaluation to bursty voice and low rate packet data in this chapter. Future
work could evaluate the proposed protocol for more complicated services. Performance evaluations show improved capacity, that is, multiplexing gains of up to $1.58$ and $1.625$ (for $D_{\text{max}}$ of 30 and 60 ms respectively) over FMA-1, while satisfying the quality of service requirements at all time. It has been observed that terminals are not affected by the internal interference as the number of possible transmissions in a slot is limited. Further improvements can be made by modifying the design parameters to suit satellite environments, for example, the frame duration ($T_j$) can be further optimised. Although this protocol was mainly proposed to operate in satellite environments (with long round-trip delays), with some minor modifications, it can be expected to perform better in cellular and indoor systems. Furthermore, through the use of advanced code allocation algorithms, efficient resource management could enhance the overall performance and efficiency of the system.
Conclusions and Further Research

8.1 Summary and Conclusions

Packet communications is an emerging technology which will marry Information Technology (IT), computer communications, and telecommunication services. With phenomenal advances in this field, various real-time and non-real-time services could co-exist in a unified network. Naturally, it is expected that packetised services offered in wired networks, such as the Internet, will seamlessly be provided to the wireless. This requires optimum and wise management of the scarce radio resources. In the evolution from the circuit-switching technology towards the packet-switching, initially parts of radio resources will be dedicated to packet-switching technologies (e.g. GPRS and S-GPRS), which it could eventually lead to full dedication of resources to packet services.

The main objective of this thesis was to provide a generic study on multiple access design, evaluation, and optimisation for wireless packet communications. Fundamentals for designing multiple access protocols for mixed services have been established. The design of an optimum medium access control (MAC) layer in the wireless is very different from that of the wired. The radio medium is unreliable and unpredictable. The derived points in this thesis could be designed for different environments and scenarios. The MAC should have dynamic and adaptive features to support unspecified future services.

Equivalent offered loads have been proposed for access and traffic loads for mixed services. The access load is the load imposed on the random access channels and the traffic load is the load on the traffic channels. Mathematical derivations of such loads show good accuracy in comparison with that of the simulation (see Appendix E).
Conclusions and Further Investigations

In process of mathematical derivation of offered loads, it was proved that non-Poisson arrivals could still be approximated by a Poisson process provided that the average silence duration is much larger than a packet transmission time. This estimation has been shown to be accurate. Generally, the normalised offered load gives a new insight to mixed traffic study of packet access protocols. The normalised and raw offered loads are exploited throughout this thesis to predict and interpret system performance in presence of various traffic scenarios. Also these offered load concepts could evidently be used for stability and saturation study.

Despite numerous proposals for packet multiple access protocols, no analytical tool had been available for performance comparison of protocols as such. Therefore, an efficiency factor was defined which normalises the system capacity with that of the upper-bound. This upper-bound is mathematically derived and analysed. Thereafter, a comprehensive comparison of 15 such protocols was carried out.

It is known in resource sharing systems that the larger the number of servers, the higher the maximum multiplexing gain would be. The study of the upper-bound system, however, reveals that multiplexing gain reaches a saturation point after which increase in the number of servers has little impacts on the multiplexing gain. These results could be used in designing the future multi-carrier packet systems.

A novel analytical method called "Aggregate Source Modelling" (ASM) was proposed for analysis of mixed services resource sharing systems such as queues and multi-access protocols. In this method, various traffic classes can be aggregated into one representative service type, whereby the mixed service analysis/simulation could be carried out by a monomedia analysis/simulation. Results presented in this thesis proved the accuracy of the method for queueing systems as well as multi-access protocols.

Based on the theoretical discussion carried out, a series of packet multiple access protocols have been proposed for cellular and satellite systems. These protocols intend to provide real-time and non-real-time packet communications. Evaluations for these systems have been carried out for various traffic scenarios (Internet-type traffic and more) an system parameters. The systems of concern were GSM and GPRS (for cellular) and Iridium and ICO (for satellite).
8.2 Further Research Topics

In this thesis, we considered service models regardless of their mobility features. However, the load perceived by a cell (or a spotbeam) depends on the cell size and the mobility model of active sources in the cell.

The call might have spent some time in the mother cell before it is handed into the reference cell. The terminal stays in the reference cell for a random duration (residence or dwell time [ZONO97]) which depends on the cell topology, terminal speed and direction. This is the duration that the terminal imposes traffic load on the reference cell. This traffic load remains in the reference cell until the call is completed or handed out to the adjacent cell(s). It can be shown that, the PDF of the cell dwell time (and related statistics, such as expected holding time and resident time) is different from that of call holding time [ZONO97]. One research topic could be to obtain truncated distributions for each service distribution presented in Chapter 3.

In Chapter 4 and 5, we have studied the pure protocol efficiency of multi-access protocols. However, in reality, the overall protocol efficiency largely depends on the operational environment (e.g. indoor, cellular, or satellite) which determines interfering conditions propagation delays, the system target (e.g. type of services, capacity, and required performance), and the technological status (such as interference suppression techniques). Further work in this area could concentrate on expanding the protocol efficiency to include physical layer impacts such as channel impairments and co-channel interference.

In Chapter 5, it was mentioned that the aggregate source modelling (ASM) could be used for stability study of random access protocols. However, further work is required for stability analysis using the ASM.

The ASM analytical method presented in Chapter 4 considered single server reservation for each service class. However, in multimedia systems, it is expected that services will employ multiple channel reservation (multi-server per service type). Multi-reservation (or multi-rate) changes the mean and the distribution that a user occupies a server. Let \( R \) be the reservation factor, then the time spent in servers is
where \( x \) is the time spent in the server with single reservation and \( y \) is that with \( R \) parallel reservation. Therefore, another area for future research could address multi-reservation ASM analysis by co-operating the above distributions above different service classes.

Also further work could look into the expansion of the ASM for priority-based queues. This could possibly be accomplished through the “Conservation Laws”.


References

780, Apr 28-May 1, 1996.


References


[ELLI] www.ellipso.com


[ETSI97] ETSI/TR 101 112 V3.1.0, “Universal mobile telecommunications system (UMTS); Selection procedures for the choice of radio transmission technology of the UMTS (UMTS 30.03 version 3.1.0),” November 1997.


[ICO] www.ICO.com/nonshock/system.htm


[KARL98] P. Karlsson and A. Arvidsson, “On the characteristics of WWW traffic and the relevance to ATM,” Sixth IFIP Workshop on Performance Modelling and Evaluation of ATM Networks (IFIP ATM’ 98), Ilkley, West Yorkshire, UK, pp. 13/1-
References


References


[MOUL92] M. Mouly, M. B. Pautet, GSM-The System for Mobile Communications,


[ORBC] www.orbc.com


1994.


References


[SPOT] www.spotbeam.com/contents.htm


References


[TAAG98f] Pouya Taaghol, R. Tafazolli, "Multimedia traffic modelling for finite population queueing systems," *First Conf. on Mathematics in Communications*, University of Loughborough, UK, Dec 98.


[TELE] www.teledesic.com

References


Statistical Distributions

In this appendix, the distributions used in the report are briefly explained. Following denotations are used: $f_x(x)$ represents the Probability Density Function (PDF) [PEEB93], $F_x(x)$ is the Cumulative Distribution Function (CDF) [PEEB93], $\bar{X}$ and $\sigma^2$ are the mean and the variance of the distribution, respectively.

A.1 Uniform Distribution

The simplest, and probably the most used, statistical distribution is that of the uniform distribution. The highlights of this distribution are given below.

\[
\begin{align*}
    f_x(x) &= \frac{1}{x_{max} - x_{min}}, & x_{min} \leq x \leq x_{max} \\
    F_x(x) &= \frac{x - x_{min}}{x_{max} - x_{min}}, & x \geq x_{min} \\
    \bar{X} &= \frac{x_{min} + x_{max}}{2} \\
    \sigma^2 &= \frac{(x_{max} - x_{min})^2}{12}
\end{align*}
\]  

(A.1)

The uniform distribution can be used to generate random numbers of other distributions. In general, we can compute a random quantity $X$ with the continuous, strictly increasing distribution $F(X)$ by setting

\[X = F^{-1}(U)\]

where $U$ is uniformly distributed random number ranging from 0 to 1; this works because the probability that $X \leq x$ is the probability that $F^{-1}(U) \leq x$, i.e., the probability
that \( U \leq F(x) \), and this is \( F(x) \) [KNUT81]

### A.2 Negative Exponential Distribution

Negative exponential distributions are commonly used in traffic modelling and analysis. Researchers are interested in the analytical simplicity associated with this distribution (or equivalently Poisson models) because of its memoryless property.

The exponential distribution has a single non-negative parameter \( \lambda \) and the principal characteristics are given as

\[
\begin{align*}
    f_x(x) &= \lambda e^{-\lambda x}, \quad x \geq 0 \\
    F_x(x) &= 1 - e^{-\lambda x}, \quad x \geq 0 \\
    \bar{X} &= \frac{1}{\lambda} \\
    \sigma^2_x &= \frac{1}{\lambda^2}
\end{align*}
\]

(A.2)

### A.3 Hyper-Exponential

The \( R \)-Stage hyper-exponential distribution denoted by \( H_R \) has the following characteristics

\[
\begin{align*}
    f_x(x) &= \sum_{i=1}^{R} \alpha_i \lambda_i e^{-\lambda_i x}, \quad x \geq 0 \\
    F_x(x) &= \sum_{i=1}^{R} \alpha_i \left(1 - e^{-\lambda_i x}\right), \quad x \geq 0 \\
    \bar{X} &= \sum_{i=1}^{R} \frac{\alpha_i}{\lambda_i} \\
    \sigma^2_x &= \sum_{i=1}^{R} \frac{2\alpha_i}{\lambda_i^2} - \mu^2
\end{align*}
\]

(A.3)

And,
A.4 Extreme Distribution

As the name implies, the extreme distribution dies very fast as the random variable increases. The principle characteristics of such a distribution is as follows

\[
\begin{align*}
\sum_{i=1}^{R} \alpha_i &= 1 \\

f_x(x) &= \frac{1}{\beta} e^{-\left(\frac{x-\alpha}{\beta}\right)} e^{-\left(\frac{x-\alpha}{\beta}\right)}, \quad x \geq 0 \\
F_x(x) &= e^{-\left(\frac{x-\alpha}{\beta}\right)}, \quad x \geq 0
\end{align*}
\]

(A.4)

A.5 Normal (Gaussian) Distribution

The normal distribution is used extensively in statistics because of its tractability and application for describing many statistical processes. It is parameterised by two values \( \mu \) and \( \sigma, \infty \leq \bar{X} \leq \infty \) and \( \sigma_x > 0 \). The principal characteristics are given as

\[
\begin{align*}
\bar{X} &= \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{x-\bar{X}}{2\sigma_x^2}\right)} \\
\sigma_x^2 &= \sigma_x^2
\end{align*}
\]

(A.5)

A.6 Pareto Distribution

The Pareto distribution (also referred to as the power-law distribution, the doubly-exponential, and the hyperbolic distribution) has been used to model distributions of incoming exceeding a minimum value, and size of asteroids, islands, cities and extinction events [KAUF93, PAXS95].
The Pareto distribution is defined by

\[
\begin{align*}
    f_x(x) &= \frac{\alpha k^\alpha}{x^{\alpha+1}}, & x \geq k \\
    F_x(x) &= 1 - \left(\frac{k}{x}\right)^\alpha, & x \geq k \\
    \bar{X} &= \frac{k\alpha}{\alpha - 1}, & \alpha > 1 \\
    \sigma_x^2 &= \frac{k^2\alpha}{(\alpha - 2)(\alpha - 1)^2}, & \alpha > 2
\end{align*}
\]  

(A.6)

where \( k \) is the location parameter and \( \alpha \) is the shape parameter. If \( \alpha \leq 2 \), then the distribution has infinite variance, and if \( \alpha \leq 1 \), then it has infinite mean.

The Pareto distribution is noteworthy for having a very heavy upper tail, an important property when considering self-similarity in network traffic [LELA94]. It has also been found that a Pareto distribution with \( 1.05 < \alpha < 1.25 \) is a good model for the amount of CPU time consumed by an arbitrary process [LELA86]. In communications, heavy-tailed distributions have been used to model telephone call holding times [DUFF94] and frame sizes for VBR video [GARR94].

**A.7 Pareto Distribution with Cut-Off**

By limiting a Pareto random variable to \( m \), the resultant distribution will be

\[
f_m(x) = \begin{cases} 
    \frac{\alpha k^\alpha}{x^{\alpha+1}}, & k \leq x < m \\
    \phi, & x = m 
\end{cases}
\]

(A.7)

where \( \phi \) is the probability that \( x > m \). Thus,

\[
\phi = \int_m^\infty f_m(x)dx = \left(\frac{k}{m}\right)^\alpha, \quad \alpha > 1
\]

Then, the mean of \( f_m(x) \) can be calculated as
A.8 Log-Normal Distribution

If a variable $X$ is normally distributed with mean $\mu$ and variance $\sigma^2$ then the log-normal variable $Y$ is found from the transformation

$$Y = e^X$$

so that

$$\ln Y = X$$

which is normally distributed. The product of a large number of positive random variables tends to have a lognormal distribution so this distribution is often used to model errors that are a product of a large number of factors.

The details of the distribution is as follow

$$\begin{align*}
  f_X(x) &= \frac{1}{\sigma x \sqrt{2\pi}} e^{-(\ln x - \mu)^2/2\sigma^2}, \quad x > 0 \\
  \bar{X} &= e^{\mu + \frac{\sigma^2}{2}} \\
  \sigma_x^2 &= e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)
\end{align*}$$

(A.8)

A.9 Log-Extreme Distribution

If the random variable $Y = \log X$ has an "extreme" distribution, then $X$ has a log-extreme distribution. These distributions are often with base-2 logarithm and written as $\log_2$-extreme.

A.10 Geometric Distribution

The geometric distribution represents the number of trails that occur in a sequence of
Bernoulli trails until the first success is encountered. In the sequence of trails the probability of success is kept fixed and equal to $p$, $0 < p < 1$, and the trails are all independent. The characteristics of this distribution are

$$
\begin{align*}
    f_x(x) &= (1-p)^{x-1}p, & x = 1, \ldots, \infty \\
    F_x(x) &= \frac{1-(1-p)^x}{p} \\
    \overline{X} &= \frac{1}{p} \\
    \sigma_x^2 &= \frac{1-p}{p^2}
\end{align*}
$$

(A.9)

**A.11 Cauchy Distribution**

The Cauch distribution with location parameter $k$ (also the median) and scale parameter $\alpha > 0$, is characterised by

$$
\begin{align*}
    f_x(x) &= \left\{\pi\alpha\left[1+\left(\frac{x-k}{\alpha}\right)^2\right]\right\}^{-1} \\
    F_x(x) &= \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-k}{\alpha}\right) \\
    \overline{X} &= \text{Do not exist} \\
    \sigma_x^2 &= \text{Do not exist}
\end{align*}
$$

(A.10)

**A.12 Cauchy Distribution with Cut-Off**

By limiting a Cauchy random variable to $0 < x < m$, the mean of the resultant distribution will be

$$
\overline{X}_m = \int_0^m xf_x(x)dx = \frac{1}{\pi}\left[\alpha \ln\left(\frac{\cos \beta_1}{\cos \beta_2}\right) + k(\beta_2 - \beta_1)\right]
$$

(A.11)

where $\beta_1$ and $\beta_2$ are

$$
\beta_1 = \tan^{-1}\left(-\frac{k}{\alpha}\right)
$$

$$
\beta_2 = \tan^{-1}\left(\frac{m-k}{\alpha}\right)
$$
Statistical Distributions

In this appendix, the distributions used in the report are briefly explained. Following denotations are used: $f_{x}(x)$ represents the Probability Density Function (PDF) [PEEB93], $F_{x}(x)$ is the Cumulative Distribution Function (CDF) [PEEB93], $\bar{X}$ and $\sigma_{x}^{2}$ are the mean and the variance of the distribution, respectively.

### A.1 Uniform Distribution

The simplest, and probably the most used, statistical distribution is that of the uniform distribution. The highlights of this distribution are given below.

\[
\begin{align*}
    f_{x}(x) &= \begin{cases} 
    \frac{1}{x_{\text{max}} - x_{\text{min}}} & , \quad x_{\text{min}} \leq x \leq x_{\text{max}} \\
    0 & , \quad \text{else} 
    \end{cases} \\
    F_{x}(x) &= \begin{cases} 
    \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} & , \quad x \geq x_{\text{min}} \\
    0 & , \quad x < x_{\text{min}} 
    \end{cases} \\
    \bar{X} &= \frac{x_{\text{min}} + x_{\text{max}}}{2} \\
    \sigma_{x}^{2} &= \frac{(x_{\text{max}} - x_{\text{min}})^2}{12}
\end{align*}
\] (A.1)

The uniform distribution can be used to generate random numbers of other distributions. In general, we can compute a random quantity $X$ with the continuous, strictly increasing distribution $F(X)$ by setting

\[X = F^{-1}(U)\]

where $U$ is uniformly distributed random number ranging from 0 to 1; this works because the probability that $X \leq x$ is the probability that $F^{-1}(U) \leq x$, i.e., the probability
that $U \leq F(x)$, and this is $F(x)$ [KNUT81]

A.2 Negative Exponential Distribution

Negative exponential distributions are commonly used in traffic modelling and analysis. Researchers are interested in the analytical simplicity associated with this distribution (or equivalently Poisson models) because of its memoryless property.

The exponential distribution has a single non-negative parameter $\lambda$ and the principal characteristics are given as

\[
\begin{align*}
    f_x(x) &= \lambda e^{-\lambda x}, \quad x \geq 0 \\
    F_x(x) &= 1 - e^{-\lambda x}, \quad x \geq 0 \\
    \bar{X} &= \frac{1}{\lambda} \\
    \sigma_x^2 &= \frac{1}{\lambda^2}
\end{align*}
\] (A.2)

A.3 Hyper-Exponential

The $R$-Stage hyper-exponential distribution denoted by $H_R$ has the following characteristics

\[
\begin{align*}
    f_x(x) &= \sum_{i=1}^{R} \alpha_i \lambda_i e^{-\lambda_i x}, \quad x \geq 0 \\
    F_x(x) &= \sum_{i=1}^{R} \alpha_i \left(1 - e^{-\lambda_i x}\right), \quad x \geq 0 \\
    \bar{X} &= \sum_{i=1}^{R} \frac{\alpha_i}{\lambda_i} \\
    \sigma_x^2 &= \sum_{i=1}^{R} \frac{2\alpha_i}{\lambda_i^2} - \mu^2
\end{align*}
\] (A.3)

And,
A.4 Extreme Distribution

As the name implies, the extreme distribution dies very fast as the random variable increases. The principle characteristics of such a distribution is as follows

\[
\begin{align*}
\alpha_i & = 1 \\

A.5 Normal (Gaussian) Distribution

The normal distribution is used extensively in statistics because of its tractability and application for describing many statistical processes. It is parameterised by two values \(\mu\) and \(\sigma, \infty \leq x \leq \infty\) and \(\sigma_x > 0\). The principal characteristics are given as

\[
\begin{align*}
\left\{ 
\begin{array}{l}
\frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\bar{X})^2 / 2\sigma^2} \\
\bar{X} = \bar{X} \\
\sigma_x^2 = \sigma_x^2 
\end{array}
\right.
\end{align*}
\]

A.6 Pareto Distribution

The Pareto distribution (also referred to as the power-law distribution, the doubly-exponential, and the hyperbolic distribution) has been used to model distributions of incoming exceeding a minimum value, and size of asteroids, islands, cities and extinction events [KAUF93, PAXS95].
Appendix A: Statistical Distributions

The Pareto distribution is defined by

\[
\begin{align*}
    f_x(x) &= \frac{\alpha k^\alpha}{x^{\alpha+1}}, & x \geq k \\
    F_x(x) &= 1 - \left(\frac{k}{x}\right)^\alpha, & x \geq k \\
    \bar{X} &= \frac{k\alpha}{\alpha - 1}, & \alpha > 1 \\
    \sigma_x^2 &= \frac{k^2\alpha}{(\alpha - 2)(\alpha - 1)^2}, & \alpha > 2
\end{align*}
\]

(A.6)

where \( k \) is the location parameter and \( \alpha \) is the shape parameter. If \( \alpha \leq 2 \), then the distribution has infinite variance, and if \( \alpha \leq 1 \), then it has infinite mean.

The Pareto distribution is noteworthy for having a very heavy upper tail, an important property when considering self-similarity in network traffic [LELA94]. It has also been found that a Pareto distribution with \( 1.05 < \alpha < 1.25 \) is a good model for the amount of CPU time consumed by an arbitrary process [LELA86]. In communications, heavy-tailed distributions have been used to model telephone call holding times [DUFF94] and frame sizes for VBR video [GARR94].

A.7 Pareto Distribution with Cut-Off

By limiting a Pareto random variable to \( m \), the resultant distribution will be

\[
f_m(x) = \begin{cases} 
    \frac{\alpha k^\alpha}{x^{\alpha+1}}, & k \leq x < m \\
    \varphi, & x = m
\end{cases}
\]

(A.7)

where \( \varphi \) is the probability that \( x > m \). Thus,

\[
\varphi = \int_m^\infty f_m(x) \, dx = \left(\frac{k}{m}\right)^\alpha, \quad \alpha > 1
\]

Then, the mean of \( f_m(x) \) can be calculated as
Appendix A: Statistical Distributions

\[ \bar{X}_m = \int_{-\infty}^{\infty} x f_m(x) \, dx = \frac{\alpha k - m}{\alpha - 1} \]  

\[ f_x(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0 \]  

\[ \bar{X} = e^{\mu + \sigma^2/2} \]  

\[ \sigma_x^2 = e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1) \]  

\section*{A.8 Log-Normal Distribution}

If a variable \( X \) is normally distributed with mean \( \mu \) and variance \( \sigma^2 \), then the log-normal variable \( Y \) is found from the transformation

\[ Y = e^Y \]

so that

\[ \ln Y = X \]

which is normally distributed. The product of a large number of positive random variables tends to have a lognormal distribution so this distribution is often used to model errors that are a product of a large number of factors.

The details of the distribution are as follow

\[ f_x(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0 \]  

\[ \bar{X} = e^{\mu + \sigma^2/2} \]  

\[ \sigma_x^2 = e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1) \]  

\section*{A.9 Log-Extreme Distribution}

If the random variable \( Y = \log X \) has an "extreme" distribution, then \( X \) has a log-extreme distribution. These distributions are often with base-2 logarithm and written as \( \log_2\)-extreme.

\section*{A.10 Geometric Distribution}

The geometric distribution represents the number of trials that occur in a sequence of
Bernoulli trails until the first success is encountered. In the sequence of trails the probability of success is kept fixed and equal to \( p, 0 < p < 1 \), and the trails are all independent. The characteristics of this distribution are

\[
\begin{align*}
    f_x(x) &= (1-p)^{x-1} p, & x = 1, \ldots, \infty \\
    F_x(x) &= \frac{1-(1-p)^x}{p} \\
    \overline{X} &= \frac{1}{p} \\
    \sigma_x^2 &= \frac{1-p}{p^2}
\end{align*}
\] (A.9)

### A.11 Cauchy Distribution

The Cauch distribution with location parameter \( k \) (also the median) and scale parameter \( \alpha > 0 \), is characterised by

\[
\begin{align*}
    f_x(x) &= \frac{\pi\alpha}{\left[1 + \left(\frac{x-k}{\alpha}\right)^2\right]^{\frac{3}{2}}} \\
    F_x(x) &= \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x-k}{\alpha}\right) \\
    \overline{X} &= \text{Do not exist} \\
    \sigma_x^2 &= \text{Do not exist}
\end{align*}
\] (A.10)

### A.12 Cauchy Distribution with Cut-Off

By limiting a Cauchy random variable to \( 0 < x < m \), the mean of the resultant distribution will be

\[
\overline{X}_m = \int_0^m x f_x(x) \, dx = \frac{1}{\pi} \left[ \alpha \ln\left(\frac{\cos\beta_1}{\cos\beta_2}\right) + k(\beta_2 - \beta_1) \right]
\] (A.11)

where \( \beta_1 \) and \( \beta_2 \) are

\[
\begin{align*}
    \beta_1 &= \tan^{-1}\left(\frac{k}{\alpha}\right) \\
    \beta_2 &= \tan^{-1}\left(\frac{m-k}{\alpha}\right)
\end{align*}
\]
APPENDIX

Non-Interleaved Gilbert-Elliot Statistical Channel Model

The propagation channel plays an important role in the overall design and evaluation of any mobile communication system. It directly impacts many low layer system aspects from the choice of access scheme, channel coding and modulation, all the way to the higher layer signalling and protocol requirements. It is, therefore, important to base the optimisation of various system parameters on accurate propagation channel models, representing the intended communication environment as realistically as possible.

In this thesis, we have used a well-known statistical channel model proposed by Gilbert-Elliot (see Figure B.1) [BISC95, CHAK96]. In this model, the channel is error free when in the GOOD state. The error are produced only in the BAD state with a probability of $\sigma$. The time spent in each state is exponentially distributed. The error model parameters are: Packet Error Rate ($\bar{P_e}$) and the mean error length $D_e$.

![Two-State Statistical Channel Model](image)

Figure B.1 Two-State Statistical Channel Model
These parameters are related to each other as follows

\[ P_{bg} = \frac{1}{D_b} \]  
(B.1)

\[ P_B = \frac{P_e}{\sigma} \]  
(B.2)

\[ P_G = 1 - P_B \]  
(B.3)

\[ P_G P_{gb} = P_B P_{bg} \]  
(B.4)

\[ D_G = \frac{D_B}{P_B} - D_B \]  
(B.5)

\[ P_{gb} = \frac{1}{D_G} \]  
(B.6)

Table B.1 provides a list of parameters typically used for evaluation purposes [CHAK96].

<table>
<thead>
<tr>
<th>Variable</th>
<th>Notation</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet Error Rate</td>
<td>( \overline{P_e} )</td>
<td>0.01</td>
</tr>
<tr>
<td>Error Rate in Bad State</td>
<td>( \sigma )</td>
<td>0.4</td>
</tr>
<tr>
<td>Mean Error Length</td>
<td>( D_B )</td>
<td>10</td>
</tr>
<tr>
<td>BAD to GOOD state Transition Probability</td>
<td>( P_{bg} )</td>
<td>0.1</td>
</tr>
<tr>
<td>GOOD to BAD State Transition Probability</td>
<td>( P_{gb} )</td>
<td>0.00256</td>
</tr>
<tr>
<td>GOOD State Probability</td>
<td>( P_G )</td>
<td>0.975</td>
</tr>
<tr>
<td>BAD State Probability</td>
<td>( P_B )</td>
<td>0.025</td>
</tr>
</tbody>
</table>
C.1 Introduction

The mathematical definition of finite population may differ from that of the real world. In the real world, everything could be considered finite; total numbers of cars, people, computer hosts, mobile cellular subscribers, and so forth. On the other hand, from analytical viewpoints, when the population exceeds a limit, it is said to be infinite. This limit is determined by the problem and the governing laws.

Likewise, we can define finite and infinite populations for queueing systems. A population is termed infinite if the nature of the arrival process depends only in a negligible way upon the number of customers already in the queueing system [KLEI76]. Otherwise, the population is tagged finite. As will be explained in the following Section, the population type impacts analysis of queueing systems and imposes special properties.

Here, we consider a queueing system serving a finite population of $M$ users. Each user could be in two modes, “thinking” and “service” (see Figure C.1).

![Figure C.1 Service Arrival Pattern of One Source](image)

The thinking or silence mode is when the user does not need to be served, and the service mode represents the period of time that the user requires to be served. The mean
periods of thinking and service modes are denoted by $\lambda$ and $\mu$, respectively. When the user requires service, it enters a First-Come-First-Served (FCFS) queue.

### C.2 Birth-Death Model

A birth-death process is a special case of a Markov process in which transitions are only permitted to neighbouring states [KLEI75]. For example, if each state represents the number of users, the system state of $n$ can only be transited to states $n\pm 1$. This restriction allows to solve much further in many cases. The continuous-time birth-death process is of more interest to us than the discrete-time, hence, we do not consider the discrete-time case in our study.

The birth-death model for the finite population $M/G/M/N$ queueing system is shown in Figure C.2. The arrival rate from each state depends on the number of users already in the system. Assuming that the arrival rate of each user follows a Poisson distribution, the aggregate arrival rate is again a Poisson process. Likewise, the departure rate varies in terms of the number users already being served. As the number of active users exceeds the total number of servers, the departure rate will become constant $(\lambda N)$. 

![Figure C.2 Birth-Death Model for Finite M/G/M/N](image)

### C.3 Analysis

Let $M$ be the finite calling population size. The time that each calling unit spends outside the system is random with a mean of $1/\lambda$. The service times follow a negative exponential distribution with departure rate of $\mu$. The awaiting customers are served by one of the $N$ available servers from the top of the queue, i.e. FCFS queueing discipline. Using the birth-death theory for an $M/G/N/M$ queuing system, we derive the following formulae [GROS85, KLEI75]
Appendix C: Finite Population Monomedia M/G/N//FCFS Queue

\[ P_{n}^{FCFS} = \begin{cases} \binom{M}{n} \rho^n p_0 & 0 \leq n < N \\ \binom{M}{n} \frac{n!}{N^{n-N} N!} \rho^n p_0 & N \leq n \leq M \end{cases} \]  

(C.1)

where \( P_{n}^{FCFS} \) is the equilibrium probabilities and defined as the probability that number of users in the system at any given time is \( n \), and \( \rho = \lambda / \mu \). Recalling

\[ \sum_{n=0}^{M} P_{n}^{FCFS} = 1 \]  

(C.2)

\( P_0 \) is calculated as

\[ P_{0}^{FCFS} = \left[ \sum_{n=0}^{N-1} \binom{M}{n} \rho^n + \sum_{n=N}^{M} \binom{M}{n} \frac{n!}{N^{n-N} N!} \rho^n \right]^{-1} \]  

(C.3)

The average number of users in the system, \( L_{FCFS} \), is easily calculated from

\[ L_{FCFS} = \sum_{n=0}^{M} n P_{n}^{FCFS} \]  

(C.4)

or,

\[ L_{FCFS} = P_{0}^{FCFS} \left[ \sum_{n=0}^{N-1} n \binom{M}{n} \rho^n + \frac{1}{N} \sum_{n=N}^{M} \binom{M}{n} \frac{n!}{N^{n-N} N!} \rho^n \right] \]  

(C.5)

When there are more users in the system than the number of servers, users are bound to wait. So the average queue length, \( L_{q}^{FCFS} \), is

\[ L_{q}^{FCFS} = \sum_{n=N}^{M} (n - N) P_{n}^{FCFS} \]  

(C.6)

or,
Appendix C: Finite Population Monomedia M/G/N//FCFS Queue

\[ L_q^{FCFS} = L^{FCFS} - N + P_0^{FCFS} \sum_{n=0}^{N-1} \binom{M}{n} p^n \]  

(C.7)

and from Little's result, the mean time that a customer spends in the system, \( W^{FCFS} \), can be calculated as

\[ W^{FCFS} = \frac{L^{FCFS}}{E(\lambda)} \]  

(C.8)

where \( E(\lambda) \) represents the average arrival rate generated by the sources and is computed as

\[ E(\lambda) = \lambda(M - L) \]  

(C.9)

where \( \Gamma \) is the source activity factor defined as

\[ \Gamma = \frac{\lambda}{\lambda + \mu} = \frac{\rho}{\rho + I} \]  

(C.10)

Furthermore, the average waiting time experienced by a source is obtained by

\[ D_q^{FCFS} = W^{FCFS} - \frac{1}{\mu} \]  

(C.11)

The CDF of the access delay is known as

\[ F_D^{FCFS}(D) = \Pr(d \leq D) = 1 - \sum_{n=N}^{M-1} \frac{P_n^{FCFS}(M-n)}{(M-L)} \sum_{i=0}^{n-N} \frac{(\mu ND)^i}{i!} e^{-\mu ND} \]  

(C.12)

differentiating (C.12) and then simplifying, the PDF of access delay is as follows

\[ f_D^{FCFS}(D) = e^{-\mu ND} \frac{\mu N}{M-L} \sum_{n=N}^{M-1} \frac{P_n^{FCFS}(M-n)(\mu ND)^{n-N}}{(n-N)!} \]  

(C.13)

Another measure of effectiveness, throughput (or resource utilisation) is defined as the probability that a server is full [TAAG95]
Appendix C: Finite Population Monomedia M/G/N//FCFS Queue

\[ S = \text{probability that a server channel is occupied} = \frac{\text{average number of occupied servers}}{\text{total number of servers}} \]

where \( S \) is throughput. Therefore, for the FCFS case the throughput is calculated by

\[
S_{\text{FCFS}}^{\text{FCFS}} = \frac{1}{N} \left( \sum_{n=0}^{N-1} nP_n^{\text{FCFS}} + N \sum_{n=N}^{M} P_n^{\text{FCFS}} \right) \quad (C.14)
\]

It has been shown that, for finite-source queues, steady-state probabilities (and hence performance measures) are independent of the service distribution profile as long as arrivals follow negative exponential distribution. Likewise, if the service time distribution is negative exponential, the system performance is invariant to the distribution of arrival (thinking) times [GROS85, BUND80].

As far as the stability of the system is concerned, we perceive immediately that finite-source queues cannot be "saturated" or become unstable (for \( \lambda/\mu < \infty \)) [KLEI76].

C.4 Activity Approximation

There are cases where the number of users in the system, \( L_{\text{FCFS}} \), is not available, however, an approximation of it could ease the analysis (see Chapter 4, Section 4.5.5).

The claim is that the activity of sources can be approximated by normalising the number users in the system by the total population size as expressed by

\[
\Gamma = \frac{L_{\text{FCFS}}}{M} \quad (C.15)
\]

The accuracy of the above estimation is verified for various population, activity factors, and number of servers in Figure C.3. As can be seen from Figure C.3.a, regardless of the calling population the estimation (C.15) is very accurate for activities of less than 0.5 (i.e. \( \lambda/\mu < 1 \)). Figure C.3.a further proves the accuracy of the estimation (C.15) for low activity sources. As discussed in Chapter 3, most bursty packet services show very low activity, and therefore, the above approximation can be widely used in analysis of bursty packet services.
Figure C.3 Comparison of Estimation and Analysis: a) Effect of Activity Factor and Population, b) Effect of Server Number and Population
APPENDIX

An Overview of Main Satellite Personal Communication Systems

D.1 ORBCOMM

The Orbcomm system uses 137-138 MHz and 400 MHz frequencies for subscriber downlink (transmissions down to mobile or fixed data communication devices) and 148-150 MHz frequencies for subscriber uplink (up to the satellites). The space segment is an array of up to 36 small communication satellite in circular LEO orbits approximately 825 km above the Earth. Twenty-eight of the satellites are planned to be put in service by the third quarter of 1998 with the remaining to be launched during 1999. The system is capable of sending and receiving two-way alphanumeric packets, similar to two-way paging or e-mail. Orbcomm is claimed to become the world’s first commercial wireless data and messaging communications services to industries. Current applications include intermittent monitoring of industrial assets such as pipelines, storage tanks and construction equipment, and tracking of mobile assets such as trailers, railway cars and shipping containers. These services will only be available in the United States at first, but will eventually be offered world-wide [ORBC, BERM98].

D.2 IRIDIUM

Scheduled to begin service towards the end of 1998, the Iridium system combines the convenience of terrestrial wireless systems with the global reach of a 66-satellite circular LEO constellation orbiting 780 km above the Earth. Iridium will provide voice, data, fax and paging services to those individuals and businesses who need global communications capability and are willing to pay for the convenience of a handheld
wireless phone or belt-worn pager. Using personalised number and Iridium phone, customers will be able to place and receive calls (or page) virtually anywhere in the world. Twelve gateways will provide global service initially [IRID, BERM98].

**D.3 GLOBASTAR**

With its planned 48-satellite LEO constellation, Globalstar will provide low-cost, high-quality telephony and other digital telecommunication services such as data transmission, paging, fax, and position location to areas currently underserved, or not served, by existing wireline and cellular telecommunication systems. The constellation altitude is $1414$ km. Qualcomm is manufacturing a tri-mode handset that will switch automatically from terrestrial analogue or digital cellular to the Globalstar’s CDMA-based satellite network [GLOB, BERM98].

**D.4 ICO**

Digital voice, data, fax and a variety of messaging services are planned for ICO’s MEO satellite mobile communications service. ICO believes its potential customers fall into four categories: existing cellular users who want service in areas covered by incompatible cellular systems; users in areas with no terrestrial coverage; speciality sectors, such as aeronautical, maritime and long-distance land transport operators; and people living in rural and remote areas lacking adequate telecommunications infrastructure. The basic terminal will be handheld and similar in size, weight and design to current pocket-sized cellular units. Phones will be dual mode, capable of working with satellite, cellular and PCS systems based on GSM as well as US, Japanese and other cellular standards such as CDMA, D-AMPS and PDC. They are scheduled for service roll-out in the year 2000. Leading mobile communications manufacturers NEC, Samsung and Mitsubishi are developing the terminals [ICO, BERM98].

**D.5 ELLIPSO**

The Ellipso system offers fully nation-wide service to each country it serves, thereby providing service to users located anywhere within national boundaries, no matter how isolated or remote. To accomplish this, Ellipso uses a constellation of elliptical MEO satellite. The constellation consists of 17 satellite in innovative inclined and equatorial orbits. In addition to low cost telephony, Ellipso also offers other digital services, such
Appendix D: An Overview of Main Satellite Personal Communication Systems

as data transfer, facsimile, paging, voice mail, messaging, and geo-positoning [ELLI, BERM98].

D.6 ECCO

Constellation Communications Inc. (CCI) is developing a big LEO global satellite system. But First, it will launch ECCO Equatorial, a 12-satellite constellation providing communications service to the world’s equatorial regions-targeting populations with the greatest demand for basic telephony service. One satellite ring covering the equatorial regions of Asia, Africa, Central and South America can reach 25 to 30 percent of the world’s population. CCI will offer both fixed site and mobile telephony. Dual mode mobile handsets will be a little bigger than cellular phones. Once ECCO Equatorial deploys in 2001, CCI will launch an additional 42 satellite in seven planes, completing its global coverage. This strategy allow CCI to begin service quickly, providing a revenue stream while its global system is being completed [BERM98, ECCO].

D.7 Other Constellations

The first generation of S-PCN such as Iridium, Globalstar, ICO and Ellipso are mainly concerned with the provision of telephony and low bit rate data services. However, it is expected that by the turn of the century mobile multimedia services such as videotelephony, teleworking and video-conferencing would be much in demand. Higher service rates with significantly better quality of service would hence have to be supported. On that note, a number of research/development efforts into the second generation S-PCN such as Teledesic [TELE], Celestri [CELE], Skybridge, EuroSkyWay, WEST, EAST, and HORIZONS [MAST97, SPOT] have been initiated. The use of advanced regenerative satellite payloads with fast on-board switching capabilities is envisaged. Any such system would have to provide a wide range of services both autonomously and integrated with existing terrestrial networks fulfilling the mobility, inter-activity, and large bandwidth requirements of such services.
Simulation Methodology

Throughout this study, a comprehensive simulation model has been developed in the C programming language. The method of simulation employed in the models is known as discrete-event with discrete-time [WATK93]. Figure E.1 illustrates the general idea and the modular steps behind the deployed simulation methodology.

Figure E.1 Simulation Methodology
E.1 Simulation Core (Scheduler)

The Simulation Core (Scheduler) has been embedded in main() of the developed C program. In a normal C application, a single main() procedure can call user routines\(^1\) and library routines\(^2\) which, in turn, can call other user routines and library routines as necessary. In contrast, the user routines in a model are essentially services called upon by the scheduler function. The scheduler function is capable of controlling the activation and termination of user routines as it alone has knowledge of the sequence of future events and the routines which must be called. User modules can, in turn, call upon other library routines and user routines and user routines as dictated by the logic of the model.

In this study, the following user modules have been developed:

E.1.1 Update Clock

This user module is responsible for updating the simulation clock and all time-related parameters and variables (such as contention delays and queueing delays). The simulation clock is increased by the simulation step \((T_s)\) at every update. The simulation process is terminated when the simulation clock reaches the simulation time \((SimTime)\) set in the initialisation module. The \(SimTime\) is set such that rarest event can be captured at least 100 times in simulation process in order to provide an accurate averaging.

E.1.2 Update Arrival

This user module handles the arrival of packets and packet bursts. The arrived messages will be put at the end of access queue.

E.1.3 Update Access Queue

This user module organises the queue according to the service priority of the queue. As was shown in Chapter 4 Section 4.5.4.2, a random multiple access protocol can also be modelled with a queueing system with a random access mechanism.

---

\(^1\) A user routine is a routine developed by the modeller.

\(^2\) A library routine is a routine supplied by the C library.
E.1.4 Update Access

This user module updates the access of the users in the access queue. In the contention-free case, the access will be simply given to the user on the top of the queue, whereas, in the random access case, the success will be determined though a pre-defined random access mechanism.

E.1.5 Update Servers

This user module updates the status of the servers (traffic channels or time slots). Traffic channels are dynamically monitored; the unused channels will be tagged available and occupied channels will labelled full.

E.2 Random Number Generators

Various random number generators have been developed which are used throughout the simulation for traffic modelling. These random number generators correspond to all necessary statistical distributions for this study. Random number generators of various distributions are derived from uniform as shown in Appendix A Section A.1.

E.3 Statistical Models of Services

Random number generators are used together to form the statistical models of services. As shown in Chapter 3, a packet traffic source can be modelled by defining its activity level and the distributions of its active and inactive periods.

E.4 Access Mechanism

Following the concepts presented in Chapter 4 Section 4.5.4.2, the access mechanism of the multi-access system under study has been modelled. These access mechanism could be contention-based, contention-free, prioritised random, or a prioritised contention-free. Through generalisation of the access mechanism, it has been made possible to produce simulation models for various multiple-access systems by just replacing the access mechanism.

E.5 Initialisation

This module in the program initiates the simulation with the given input parameters. The initialisation procedure includes setting the simulation time (SimTime), the warm-
up time, the simulation step ($T_s$), arrays, queues, terminals, re-setting the simulation clock, and the performance-gathering parameters (such as total access delay and number of arrivals).

**E.7 Post-Processing**

This module is responsible for post-processing of the statistics gathered from the simulation process. These calculations comprise of calculating PDFs, CDFs, averages, summations, radios and percentages.

**E.8 Output**

Finally, this module outputs the post-processed data on the screen or into various data files.