Optimum Hybrid Error Correction Scheme under Strict Delay Constraints

Dissertation

Zur Erlangung des Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
der Naturwissenschaftlich-Technische Fakultät I
- Mathematik und Informatik -
der Universität des Saarlandes

von

Guoping Tan

Saarbrücken, Deutschland 2009

Dekan: Uni.-Prof. Dr. Joachim Weickert

Tag des Kolloquiums: 24. März 2009

Die Mitglieder des Prüfungsausschusses:

Der Vorsitzenden: Herrn Prof. Dr. Joachim Weickert
Die Berichterstatter: Herrn Prof. Dr.-Ing, Thorsten Herfet
Herrn Prof. Dr. Raimund Seidel

Der akademischen

Mitarbeiter: Herrn Dr. Michael Breuss

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, daß ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen order indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Diese Arbeit wurde bisher weder im In- noch im Ausland in gleicher order ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgeleget.

Saarbrücken, den 25. März 2009

Acknowledgments

A number of people have helped with the accomplishment of this thesis. First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Thorsten Herfet, for his encouragement, support, and invaluable advices during the whole period of my Ph.D study.

I am also grateful to Professor Raimund Seidel for taking the role of my second supervisor and reviewing my thesis.

I would like to give my special thankfulness to Professor Shixin Cheng, who is from Southeast University, Nanjing, China, for his kind support and encouragement. His suggestions to my research work have been very valuable.

Many thanks to Manuel Gorius for a fruitful cooperation during the research work. And I wish to acknowledge my appreciation to Diane Chlupka and Zakaria Keshta for their kind assistance during my Ph.D. study in the telecommunications lab. I also appreciate all of the other colleagues and friends for their friendship and companionship. I really had a wonderful and fruitful time in the whole period of my Ph.D study. I cannot mention them all, but I would like to name Aleksej Spenst, Eric Haschke, Jochen Miroll, Sylvia Kümmel, Zhao Li, Dr. Bingyang Wu, Dr. Qifan Chen, Dr. Zhigang Zhou, Dr. Jing Xu, Dr. Yang Lan, Dr. Wei Li, and Dr. Fang Wang.

Last but not least, I would like to express my special love and thanks to my wife Lianxiang Liu and my daughter Wenrui Tan for their endless love and support. I also owe a lot to my parents for their great loves. This thesis would not have been possible without their support.

Kurzzusammenfassung

In Paket-basierten drahtlosen Netzwerken benötigen Medien-basierte Dienste oft Multicast-fähigen Transport, der quasi-fehlerfreie Übertragung unter strikten Zeitgrenzen garantiert. Außerdem beeinflussen sowohl Multicast als auch Zeitbegrenzungen stark die Architektur von Auslöschungs-Fehlerschutz (Erasure Error Recovery, EER). Daher stellen wir eine allgemeine Architektur der EER vor und untersuchen ihre Optimierung in dieser Arbeit. Die Architektur integriert alle wichtigen EER-Techniken: Automatic Repeat Request, Forward Error Correction und Hybrid ARQ. Jede dieser EER-Techniken kann als Spezialfall der Hybrid Error Correction (HEC) angesehen werden. Da das Gilbert-Elliot (GE) Auslöschungs-Fehler-Modell für einen weiten Bereich von Paket-basierten drahtlosen Netzwerken als gültig erwiesen wurde, präsentieren wir in dieser Arbeit die allgemeine Architektur und deren Optimierung basierend auf dem GE Kanalmodell. Zweck der Optimierung ist es, eine gewisse Ziel-Paketfehlerrate unter strikten Zeitgrenzen effizient zu erreichen. Durch die Optimierung für ein gegebenes Echtzeit-Mutlicast-Szenario kann die insgesamt benötigte Redundanz-Information minimiert werden. Dies erfolgt durch automatische Auswahl des optimalen HEC Schemas unter all den Schemata, die in die Architektur integriert sind. Das optimale HEC-Schema kann die Shannon Grenze so nahe wie möglich, dynamisch, entsprechend dem derzeitigen Kanalzustand, erreichen.

Short Abstract

In packet-based wireless networks, media-based services often require a multicast-enabled transport that guarantees quasi error free transmission under strict delay constraints. Furthermore, both multicast and delay constraints deeply influence the architecture of erasure error recovery (EER). Therefore, we propose a general architecture of EER and study its optimization in this thesis. The architecture integrates overall existing important EER techniques: Automatic Repeat Request, Forward Error Correction and Hybrid ARQ techniques. Each of these EER techniques can be viewed as a special case of Hybrid Error Correction (HEC) schemes. Since the Gilbert-Elliott (GE) erasure error model has been proven to be valid for a wide range of packet based wireless networks, in this thesis, we present the general architecture and its optimization based on the GE channel model. The optimization target is to satisfy a certain target packet loss level under strict delay constraints efficiently. Through the optimization for a given real-time multicast scenario, the total needed redundancy information can be minimized by choosing the best HEC scheme automatically among the entire schemes included in the architecture. As a result, the performance of the optimum HEC scheme can approach the Shannon limit as closely as possible dynamically according to current channel state information.

Zusammenfassung

Die wachsende Nachfrage nach Echtzeit-Multimedia Diensten über das derzeitige Internet führte zur Konzeption des Future Media Internet (FMI)[1]. Im FMI wird der Großteil der übertragenen Bits aus Medien-basierten Diensten stammen, die üblicherweise strikte Übertragungszeiten gewährleisten müssen, aber eine bestimmte Fehlerrate tolerieren können. Außerdem nimmt man an, dass die Internet Protocol Television (IPTV) die Killer-Applikation des Next-Generation (NG) Internet sein wird. Mit der größer werdenden Nachfrage nach mobilen Multimedia-Diensten brauchen die Echtzeit Multimedia-Dienste wie IPTV im NG Internet oft Unterstützung zur Multicast-Kommunikation in drahtlosen, IP-basierten Netzwerken. Bis jetzt ist es allerdings immer noch eine Herausforderung, die unterschiedlichen Bedürfnisse nach Dienstgüte (Quality of Service, QoS) der verschiedenen mobilen Real-time Multimedia Multicast (RMM) Anwendungen zu unterstützen. Dieses Problem wird in der vorliegenden Arbeit behandelt und es wird versucht, eine perfekte Lösung zu finden, die verschiedenen QoS-Anforderungen verschiedener RMM-Dienste zu erfüllen.

Drahtlose Kanäle sind jedoch fehlerbehaftet, begrenzt in der Bandbreite und zeit-invariant durch Fading-Effekte, Interferenzen, die Mobilität des Nutzers etc. Gewöhnlich muss man Fehlerschutz-Techniken anwenden, um akzeptable Qualität für Multicast-Dienste über drahtlose Netzwerke zu erhalten. Im Allgemeinen verwendet man bitweise Kanalcodierung in der Physical Layer, um die Transport-Zuverlässigkeit zu erhöhen. Allerdings kann die bitweise Kanalcodierung keine Bündelfehlern (Burst Error) beheben, die länger als die entsprechenden Codeworte sind. Dieses Problem bedingt die Beschädigung von Paketen auf Layer 2 (OSI Model[2]) oder auf höheren Ebenen, so dass auf diesen Ebenen Paketverlust stattfindet. Diese Paketverluste können durch einen alternativen Ansatz korrigiert werden,

1 http://www.futuremediainternet.eu/

2 http://de.wikipedia.org/wiki/OSI-Modell

wobei Erasure Error Recovery (EER) Mechanismen auf Paket-Ebene betrachtet werden. Bei Verwendung von Kanalcodierung auf Paket-Ebene, statt auf Bit-Ebene, werden Pakete als Code-Symbole dieser EER Schemata angesehen. Diese Arbeit wird Fehler-Codierungs-Mechanismen auf Layer 2 oder auf höheren Schichten betrachten, um Übertragungs-Zuverlässigkeit in drahtlosen Netzwerken zu garantieren. Des Weiteren wird gezeigt, dass das so genannte Gilbert-Elliot (GE) Kanalmodell [Mus89] bestehend aus einer Markov Kette mit zwei Zuständen geeignet ist, diese Frame-Verluste in Fading-Kanälen bzw. die Paketverluste in Wireless LANs, wie den IEEE 802.11a[3], IEEE 802.11b[4] etc, zu modellieren. In dieser Arbeit wird daher das GE Kanalmodell eingesetzt, um die Leistungsfähigkeit verschiedener EER-Schemata auszuwerten.

Andererseits verlangen diese Multimedia Multicast Dienste oft, im Gegensatz zu Nicht-Echtzeit-Diensten, strikt Zeit-begrenzte und quasi-fehlerfreie (Quasi Error Free, QEF) Übertragung. In der Tat können strikte Zeitgrenzen im Entwurf des Echtzeit-Transports Multimedia-basierter Dienste die Forschungsarbeit in vielen Bereichen stark beeinflussen: Network Coding, Multicast Routing, Fehlerschutz-Mechanismen etc. Diese Arbeit konzentriert sich vor allem auf folgendes Thema: Wie beeinflussen Multicast und strikte Zeitgrenzen den Entwurf von zuverlässigen Echtzeit-Multicast-Protokollen durch die Anwendung verschiedener EER-Techniken? Die vorliegende Arbeit ist eine einleitende Forschungsarbeit über die Betrachtung strikter Zeitbeschränkungen als fundamentale Grenze. Eine elementare Untersuchung des Einflusses strikter Zeitbedingungen auf andere aktuelle Forschungsbereiche wie Network Coding, Multicast Routing usw. wird erforderlich.

Traditionell gibt es zwei grundlegende EER-Schemata: Automatic Repeat reQuest (ARQ) und Forward Error Correction (FEC). Des Weiteren bezeichnet man Schemata, die sowohl ARQ als auch FEC integrieren, als Hybrid ARQ. In dieser Arbeit wird jede dieser EER Techniken als ein Spezialfall der Hybrid Error Correction (HEC) betrachtet. Wenn diese Schemata auf ein gegebenes RMM-Szenario angewendet werden, verbleiben immer noch

[3] http://de.wikipedia.org/wiki/IEEE_802.11a

[4] http://de.wikipedia.org/wiki/IEEE_802.11b

zwei kritische Fragen: Welche sind die optimalen Parameter für diese verschiedenen EER Schemata und was ist das beste Schema, um die quasi-fehlerfreie Übertragung unter strikten Zeitbeschränkungen effizient zu garantieren? Unter Berücksichtigung dieser beiden Fragen wird eine allgemeine Architektur der EER entwickelt, die alle der oben genannten EER Schemata kombiniert. Unter Verwendung der allgemeinen Architektur liefern wir in dieser Arbeit ein allgemeines mathematisches Framework, um deren Leistungsfähigkeit zu analysieren und die Parameter zu optimieren. Durch Optimierung der allgemeinen Architektur kann die insgesamt benötigte Redundanz-Information (RI) minimiert werden, indem automatisch das optimale HEC-Schema unter den in die allgemeine Architektur integrierten Schemata gewählt wird. Unsere Forschungen zeigen, dass das optimale HEC-Schema die unterschiedlichen QoS-Anforderungen unter strikten Zeitbedingungen perfekt garantieren kann und dabei alle existierenden EER-Schemata übertrifft. Tatsächlich kann das optimale HEC-Schema die sog. Shannon-Grenze so nahe wie möglich, dynamisch, entsprechend des derzeitigen Kanalzustandes, erreichen. Als Ergebnis bieten unsere Beiträge in dieser Arbeit eine nützliche Anleitung zur Entwicklung zuverlässiger RMM-Protokolle in zukünftigen IP-basierten Netzwerken.

Zum Schluss geben wir einen kleinen Überblick über die Publikationen, die aus dieser Arbeit entstanden sind. Teile der Arbeit wurden auf den folgenden beiden Symposien vorgetragen: Erstes und zweites *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (*ISBMSB*), die durch die IEEE Broadcasting Society finanziert wurden. Teile der Arbeit wurden auf folgenden Konferenzen präsentiert: *Die dritte und vierte IEEE International Conference on Wireless Communications, Networking and Mobile Computing* (i.e. WiCOM2007 and WiCOM2008), beide größere Konferenzen, finanziert von IEEE Communications Society; und der *European Wireless Conference* 2008 (i.e. EW2008). Teile der Arbeit wurden in den folgenden beiden Journalen veröffentlicht: *IEEE Transactions on Broadcasting*, March, 2007; and the *International Journal of Communications, Network and System Sciences,* June, 2008.

Abstract

The rapidly increasing demand on real-time multimedia services over current Internet has led to the creation of the Future Media Internet (FMI) concept[1]. In the FMI, the majority of bits transported will be media-based services, which usually require strict delivery time but can tolerate a certain error rate. In addition, it was even claimed that Internet Protocol Television (IPTV) is the killer application for the next-generation (NG) Internet. With the increase of the demands of mobile multimedia services, the real-time multimedia services such as IPTV in NG Internet often require the support of multimedia multicast communications in the IP-based wireless networks. Up to now, however, how to support the diverse Quality of Services (QoS) requirements for different mobile Real-time Multimedia Multicast (RMM) applications is still a challenge. We thus address this issue in this thesis and try to find a perfect solution for guaranteeing the diverse QoS requirements of different RMM services.

Wireless channels, however, are error-prone, bandwidth-limited and time-varying channels due to fading effects, interferences and user mobility etc. Usually, we have to employ some error recovery schemes to provide the acceptable quality for multicast services over wireless networks. In general, we employ bit-wise channel coding in physical layer for increasing the transport reliability. However, the bit-wise channel coding can not deal with burst errors longer than code words. This problem will cause the corruption of packets in Layer 2 (OSI Model[2]) or higher layers so that packet loss happens in those layers. This packet loss issue can be overcome by an alternative approach, where packet level erasure error recovery (EER) schemes are considered. Using packet level channel coding instead of bits, packets are seen as code symbols in those EER schemes. This thesis will address the error coding schemes over erasure error channel, which actually are the solutions in Layer 2 or higher layers for guaranteeing the transport reliability in wireless networks. Moreover, the so-

[1] http://www.futuremediainternet.eu/

[2] http://de.wikipedia.org/wiki/OSI-Modell

called Gilbert-Elliott (GE) channel model [Mus89] with the two-state Markov chain was proved to be adequate for modeling the frame losses in slow fading channels and the burst packet losses in Wireless LANs such as IEEE 802.11a[3] and IEEE 802.11b[4] and so on. In this thesis, therefore, we will adopt the GE channel model as the erasure error channel model to evaluate the performance of all kinds of EER schemes.

On the other hand, unlike non-real-time services, those mobile multimedia multicast services often require a strict delay-bounded transport that guarantees quasi error free (QEF) transmission. We thus need to consider the influences of strict delay constraints when designing suitable EER schemes for reliable QEF transmissions. In fact, strict delay constraints in the design of real-time multimedia based services transport can deeply influence the research works in a wide variety of fields: Network coding, multicast routing, error correction schemes etc. In this thesis, we will address the following big issue: How do the multicast and the strict delay constraints influence the design of reliable real-time multicast protocols by employing all kinds of EER techniques? This thesis work is an initial research work of considering the strict delay constraints as the fundamental limit, which is an elementary investigation on the influence of the strict delay constraints to other hot research areas such as network coding, multicast routing and so on.

Traditionally, there are two basic EER schemes: Automatic Repeat reQuest (ARQ) schemes and Forward Error Correction (FEC) schemes. The schemes integrating ARQ schemes and FEC schemes are referred to Hybrid ARQ schemes. Each of these EER techniques will be viewed as a special case of Hybrid Error Correction (HEC) schemes in this thesis. When employing these schemes for a given RMM scenario, there are still two critical questions needed to be answered: For guaranteeing the QEF transmission under the strict delay constraints efficiently, what are the optimum parameters for those different EER schemes and which scheme is the best scheme? By addressing these two questions, we thus develop a general architecture of EER combing all of the EER schemes mentioned above.

[3] http://de.wikipedia.org/wiki/IEEE_802.11a

[4] http://de.wikipedia.org/wiki/IEEE_802.11b

Using this general architecture, we contribute a general mathematical framework to analyze its performance and optimize its parameters in this thesis. By optimizing the general architecture, the total needed Redundant Information (RI) can be minimized by choosing the optimum HEC scheme automatically among the overall schemes integrated in the general architecture. Our studies in this thesis show that the optimum HEC scheme can guarantee the diverse QoS requirements under strict delay constraints perfectly and outperforms all of the existing EER schemes. In fact, the performance of the optimum HEC scheme can approach the Shannon limit as closely as possible dynamically according to current channel state information. As a result, our contributions in this thesis provide a useful guide for designing reliable RMM protocols in future IP-based networks.

Finally, we give a brief overview on the publications resulted from this work. Parts of this thesis work have been presented at the following two symposiums: *The* 1st *and the* 2nd *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (*ISBMSB*), which are sponsored by the IEEE Broadcasting Society. Parts of this thesis work have been presented at the following conferences: *The* 3rd *and the* 4th *IEEE International Conference on Wireless Communications, Networking and Mobile Computing* (i.e. WiCOM2007 and WiCOM2008), which are major conferences sponsored by IEEE Communications Society; and the *European Wireless Conference* 2008 (i.e. EW2008). Parts of this thesis work have been published in the following two journals: *IEEE Transactions on Broadcasting*, March, 2007; and the *International Journal of Communications, Network and System Sciences,* June, 2008.

Contents

List of Figures

List of Tables

Abbreviations

3GPP	Third-generation Partnership Project
3GPP2	Third-generation Partnership Project 2
ACK	Acknowledgment
ACKs	Acknowledgments
AFEC	Adaptive Forward Error Correction
AHEC	Adaptive Hybrid Error Correction
AL	Application Layer
ALC	Asynchronous Layered Coding
AP	Access Point
ARQ	Automatic Repeat reQuest
BCMCS	BroadCast MultiCast Services
CC	Correlation Coefficient
CDMA	Code Division Multiple Access
CRC	Cyclic Redundancy Check
CSI	Channel State Information
DTV	Digital Television
DVB	Digital Video Broadcast
DVB-C	Digital Video Broadcast - Cable
DVB-H	Digital Video Broadcast - Handheld

DVB-IPI	Digital Video Broadcast - Internet Protocol Infrastructure
DVB-S	Digital Video Broadcast - Satellite
DVB-T	Digital Video Broadcast - Terrestrial
EER	Erasure Error Recovery
FEC	Forward Error Correction
FLUTE	File Delivery over Unidirectional Transport
FMI	Future Media Internet
FSMC	Finite State Markov Channel
GBN	Go-Back-N
GE	Gilbert-Elliott
GSM	Global System for Mobile communications
HARQ	Hybrid Automatic Repeat reQuest
HDTV	High Definition Television
HEC	Hybrid Error Correction
HEC-PR	Hybrid Error Correction with Packet Repetition technique
HTTP	HyperText Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPI	Internet Protocol Infrastructure
IPTV	Internet Protocol Television
LAN	Local Area Network
LCT	Layered Control Transport
LDPC	Low Density Parity Check

LT	Luby Transform
MAC	Medium Access Control
MBMS	Multimedia Broadcast/Multicast Services
MF	Mathematical Framework
MLE	Maximum Likelihood Estimation
MPEG	Moving Picture Experts Group
MRT	Multicast Reliable Transport
MTA	Markov-based Trace Analysis
NACK	Negative-Acknowledgment
NACKs	Negative-Acknowledgments
NC	Network Coding
NECC	Network Error Correction Coding
NG	Next-Generation
NORM	NACK–Oriented Reliable Multicast
OSI	Open Systems Interconnection
P2P	Peer-to-Peer
PC	Personal Computer
PL-FEC	Packet Level Forward Error Correction
PLM	Packet Loss Model
PLR	Packet Loss Ratio
PR	Packet Repetition
PS	Packet-Switched
RI	Redundancy Information
RMM	Real-time Multimedia Multicast
RMTP	Real-time Multicast Transport Protocol

RMTPs	Real-time Multicast Transport Protocols
RS	Reed-Solomon
RTP	Real-time Transport Protocol
RTP/AVPF	Extended RTP Profile for RTCP-based Feedback
RTSP	Real Time Streaming Protocol
RTCP	Real-time Transport Control Protocol
RTT	Round Trip Time
SDTV	Standard Definition Television
SR	Selective Repeat
SW	Stop-and-Wait
TCP	Transmission Control Protocol
TPM	Transition Probability Matrix
UDP	User Datagram Protocol
UEP	Unequal Error Protection
UMTS	Universal Mobile Telecommunications System
VB	Virtual Block
WLAN	Wireless Local Area Network
WHN	Wireless Home Network

Notations

$E(\mathrm{X})$	The expected value of a random variable X
$\Pr(y)$	Calculating the probability of the expression y
$\ln(x)$	The natural logarithm of x
$\frac{\partial \Lambda}{\partial x}$	The partial derivation of Λ with respect to x
$\lfloor x \rfloor$	The largest integer which is not greater than x
$\lceil x \rceil$	The smallest integer which is greater than or equal to x
$\hat{x}$	The estimated value for x
$\binom{m}{h}$	The number of ways h objects can be chosen from among m objects without repetition
$P_{B\|B}$	The one transition probability of from "B" state to "B" state in GE channel model
$P_{G\|G}$	The one transition probability of from "G" state to "G" state in GE channel model
$P_{G\|B}$	The one transition probability of from "B" state to "G" state in GE channel model
$P_{B\|G}$	The one transition probability of from "G" state to "B" state in GE channel model
$P_t[m]$	The *m-step* transition probability matrix in GE channel model
$P_{B\|B}[m]$	The *m-step* transition probability of starting in the "B" state and ending in the "B" state in GE channel model
$P_{G\|G}[m]$	The *m-step* transition probability of starting in the "G" state and ending in the "G" state in GE channel model

$P_{G\|B}[m]$	The *m-step* transition probability of starting in the "B" state and ending in the "G" state in GE channel model
$P_{B\|G}[m]$	The *m-step* transition probability of starting in the "G" state and ending in the "B" state in GE channel model
PLR_{GE}	The average packet loss ratio in GE channel model
X	A random variable representing the error-burst-length in GE channel model
Y	A random variable representing the error-free-length in GE channel model
P_{X}^{j}	The probability of the random variable X being j
P_{Y}^{j}	The probability of the random variable Y being i
$\hat{P}_{G\|G}$	The maximum likelihood estimation of $P_{G\|G}$
$\hat{P}_{B\|B}$	The maximum likelihood estimation of $P_{B\|B}$
ρ	The correlation coefficient of GE channel model
Γ	A random variable representing the loss probability of a packet transmitted in GE channel
Γ'	A random variable representing the loss probability of each retransmission packet in the first retransmission stage when retransmission-based schemes used
$P(b,d,$ CSI)	The probability of b packets lost in a sequence of d packets in the GE channel with the CSI of ($P_{G\|G}$, $P_{B\|B}$)
$P_G(b,d)$	The probability of b errors in d transmissions with the channel ending in state "G" with the GE channel model
$P_B(b,d)$	The probability of b errors in d transmissions with the channel ending in state "B" with the GE channel model
$\mathrm{PLR}_{\mathrm{target}}$	Target packet loss ratio requirement
$\mathrm{D}_{\mathrm{target}}$	Target delay requirement
P_e	Original link packet loss ratio
R_d	Constant source multimedia multicast data rate
(n, k)	The ideal systematic erasure code adopted by the AFEC scheme, which is with a codeword of n symbols along with k source symbols

k_{lim}	The maximum allowable length of the parameter k with the AFEC scheme due to the constrict delay constraints
n_{opt}	The optimum parameter n with the AFEC scheme for a certain multicast scenario
I_k	A random variable representing the number of lost data packets in a group of k data packets after decoding using the FEC decoder with ideal (n,k) code, which is used for analyzing the performance of the AFEC scheme
$P_{G\mid G}(j)$	The transition probability of from state "G" to state "G" for the *j-th* receiver in one multicast scenario with GE model
$P_{B\mid B}(j)$	The transition probability of from state "B" to state "B" for the *j-th* receiver in one multicast scenario with GE model
$CSI(j)$	Channel State Information of certain parameters $(P_{G\mid G}(j), P_{B\mid B}(j))$ for the *j-th* receiver in one multicast scenario with GE model
$\rho(j)$	The CC of the GE channel model for the *j-th* receiver
$\vec{C}_{CSI}$	A vector denotes the CSI of overall receivers in a multicast scenario, which is defined as $\vec{C}_{CSI} = \{CSI(j) \mid 1 \le j \le N_{recv}\}$
$P_G(j)$	The steady state probability of being in state "G" for the *j-th* receiver in one multicast scenario with $CSI(j)$
$P_B(j)$	The steady state probability of being in state "B" for the *j-th* receiver in one multicast scenario with $CSI(j)$
$X(j)$	A random variable representing the error-burst-length during the first transmission for the *j-th* receiver in one multicast scenario with $CSI(j)$
$Y(j)$	A random variable representing the error-free-length during the first transmission for the *j-th* receiver in one multicast scenario with $CSI(j)$
$RTT(j)$	Average round trip time for the *j-th* receiver in one multicast scenario, one way delay is $RTT(j)/2$. In this thesis, it does not include any time caused by protocols (e.g. t_{rw} and t_{sw} below)
N_{recv}	The number of receivers in the multicast scenario. This parameter is also viewed as the group size in a multicast scenario in the thesis
l_p	The length of one packet in unit of bytes
B_w	The channel bandwidth in unit of Mbps
t_s	The time of average interval between two continuous original data packets at the sender

t_d	The time duration of one packet occupied in the transmission at the sender. Note that it is no more than t_s
t_{rw}	The average waiting time at each receiver. It is the time between the time when in original stream the latest packet loss occurs and the time when the corresponding NACK is sent. Note it has included the time of loss detection at the receiver. Note that the average waiting time at each receiver is identical due to the same process for all of the receivers
t_{sw}	The average waiting time at the transmitter, which is the time between receiving a NACK message and the time when the corresponding packets required by the NACK message are retransmitted
t_{rp}	The maximum possible time duration occupied by all of the retransmission packets in each retransmission round, which is viewed as same for all of the receivers in a multicast scenario
$t_{lp}(j)$	The time duration from the time the latest packet loss occurs at the j-th receiver to the time it possibly receives this required packet, which is $RTT(j)+ t_{sw}+ t_{rw}$
$N_{rr,max}(j)$	The maximum possible number of retransmission rounds for the j-th receiver with the HEC-PR scheme, where $1 \leq j \leq N_{recv}$
$N_{rt}^q(j)$	The number of copies on each retransmission packet for the j-th receiver at the sender during the q-th retransmission round with the HEC-PR scheme, where $1 \leq q \leq N_{rr,\max}(j)$
$\vec{N}_{rt}(j)$	A vector denotes the parameters $N_{rt}^q(j)$ in each retransmission round with the HEC-PR scheme, which is defined as $\vec{N}_{rt}(j) = \left\{N_{rt}^q(j) \middle\vert 1 \leq q \leq N_{rr,\max}(j)\right\}$
$N_{rt,\max}(j)$	The maximum possible number of retransmissions for each missing data packet for the j-th receiver at the sender with the HEC-PR scheme, which is given by: $N_{rt,\max}(j) = \sum_{q=1}^{N_{rr,\max}(j)} N_{rt}^q(j)$
$\mathrm{I}_{\mathrm{X}(j)}(w)$	A random variable represents the number of lost packets among the X(j) lost packets for the j-th receiver with w (where $0 \leq w \leq N_{rr,max}(j)$) retransmission rounds with the HEC-PR scheme, which is used for analyzing the performance of the HEC-PR scheme
Φ^m	A finite space with the number of dimensions varying from 1 to m
k	The number of source data packets in one encoding block with the general architecture of EER, where $k \geq 1$

N_p	The number of redundant packets in one encoding block in the first transmission with the general architecture of EER
$N_{rr,max}$	The maximum possible number of retransmission rounds with the general architecture of EER
N_{cc}^q	A constant coefficient in the q-th (where $1 \le q \le N_{rr,max}$) retransmission round with the general architecture of EER, which is the number of multiples for the number of required redundant packets in the q-th retransmission round
$\vec{N}_{cc}$	A vector denotes the parameters N_{cc}^q in each retransmission round with the general architecture of EER, which is defined as $\vec{N}_{cc} = \left\{ N_{cc}^q \middle\| 1 \le q \le N_{rr,\max} \right\}$
$\mathrm{I}_k(j,w)$	A random variable represents the number of lost data packets in one encoding block of k source data packets after the *j-th* receiver experiences w retransmission rounds with the general architecture of EER, where $1 \le w \le N_{rr,max}$
$\mathrm{N}_{req}(j,w)$	A random variable represents the number of redundant packets required by the *j-th* receiver in the w-th retransmission round with the general architecture of EER, which is for recovering all of the k data packets in one block, and where $0 \le \mathrm{N}_{req}(j,w) \le k$ and $1 \le w \le N_{rr,max}$
$\mathrm{N}_{rea,\max}^w$	A random variable represents the maximum number of redundant packets required by all of the receivers in the w-th retransmission round with the general architecture of EER
Z	An aggregate represents all of sequence number of the receivers in a multicast scenario, i.e. $\mathrm{Z} = \{ i \| 1 \le i \le N_{recv} \}$
$\mathrm{Z}(h,m)$	An aggregate with h elements represents the result on the m-th way of choosing h elements from the overall elements of Z without repetition, where $0 \le h \le N_{recv}$ and $1 \le m \le \binom{N_{recv}}{h}$. It is denoted by $\mathrm{Z}(h,m) = \{ e(i,m) \| 1 \le i \le h \}$
$\overline{\mathrm{Z}}(h,m)$	An aggregate with $N_{recv}-h$ elements represents the complement of $\mathrm{Z}(h,m)$ in the aggregate Z, which is denoted by $\overline{\mathrm{Z}}(h,m) = \{ \overline{e}(i,m) \| 1 \le i \le N_{recv} - h \}$
Z^j	An aggregate with N_{recv}-1 elements defined as $\mathrm{Z}^j = \mathrm{Z} - \{ j \}$

$Z^j(h,m)$ An aggregate with h elements representing the result on the m-th way of choosing h elements from the overall elements in Z^j without repetition, where $0 \leq h \leq N_{recv}$ and $1 \leq m \leq \begin{pmatrix} N_{recv}-1 \\ h \end{pmatrix}$. It is denoted by $Z^j(h,m)=\{e^j(i,m) \mid 1 \leq i \leq h\}$

$\overline{Z}^j(h,m)$ An aggregate with N_{recv}-h-1 elements represents the complement of $Z^j(h,m)$ in the aggregate Z^j, which is denoted by $\overline{Z}^j(h,m)=\{\overline{e}^j(i,m) \mid 1 \leq i \leq N_{recv}-h-1\}$

Chapter 1

Introduction

In the near future Internet, the majority of bits transported will be media-based services, which usually require strict delivery time but can tolerate a certain error rate. The rapidly increasing demand on real-time multimedia services over current Internet has led to the creation of the Future Media Internet (FMI) concept[1]. Further, it was claimed that Internet Protocol (IP) [Rfc0a] Television (IPTV) is the killer application for the next-generation (NG) Internet [Xia07]. Additionally, with the increase of the demands of group-oriented real-time multimedia services such as video conferencing, distance educations, mobile entertainment services and interactive games explosively, the real-time multimedia services such as IPTV in NG Internet often require the support of multimedia multicast communications in the NG IP-based wireless networks. Actually, it has led to the mobile multimedia multicast communications have become a critically important component of the NG IP-based wireless networks. For example, the current cellular networks such as UMTS, CDMA2000, WiMAX[2] or extensions of those have been extended with IP multicast transport, e.g. with 3GPP Multimedia Broadcast/Multicast Services (MBMS) [3gp05a] [3gp05b] or 3GPP2 BroadCast MultiCast Services (BCMCS) [3gp06], which provide the possibility of distribute real-time multimedia services for mobile users via IP multicast data

[1] http://www.futuremediainternet.eu/

[2] http://en.wikipedia.org/wiki/WiMAX

over point-to-multipoint radio bearers. Moreover, the IP multimedia subsystem (IMS) was recently specified by the 3GPP as an overlay framework over 3GPP Packet-Switched (PS) domain [3gp05c], where IMS is designed to support real-time voice over IP and other IP-based multimedia services. In fact, IMS is not limited to cellular networks but is expected to be used to support IP multimedia services in the integrated all-IP networks [Agr08]. Consequently, mobile multicast in the NG wireless networks has received more and more research attention [Jin06] [Zha07].

Since wireless channels are error-prone, bandwidth-limited and time-varying due to fading effects, interferences and user mobility etc. [Pro00], we usually must employ some error recovery schemes to provide a real-time service of acceptable quality over wireless networks. In addition, most of the successful wireless networks (e.g. 3GPP and 3GPP2) adopt the IP as a network layer simplifying the integration of wireless networks into internet networks. This results in user applications seeing the wireless channel as an IP packet based channel with erasure errors, i.e. IP packet losses. On the other hand, users usually do not have direct access to the physical or even the Medium Access Control (MAC) layer in real systems. Therefore, it will be very convenient by employing application layer (AL) erasure error recovery (EER) schemes to guarantee the diverse QoS requirements for different users. Based on the background introduced above, in this thesis, we will model the wireless channels as erasure error channels for studying how to guarantee the QoS requirements for real-time multicast services efficiently with all kinds of EER schemes.

Unlike the traditional reliable transmission of data over Internet, the real-time multimedia services can tolerate a certain packet loss ratio (PLR) but must have strict delay constraints. As a matter of fact, strict delay constraints in the design of real-time multimedia based services transport can deeply influence the research works in a wide variety of fields: Network Coding (NC) [Ahl00], multicast routing, error correction scheme etc. For example, the architectural design challenges in implementing network coding for peer-to-peer live multimedia streaming have been addressed in [Wan07a] [Wan07b]. The Quality of Services (QoS) requirements of multimedia services, however, can deeply vary from application to application. In the NG IP-based wireless networks, how to support those diverse QoS requirements under strict delay constraints for different mobile multimedia multicast applications is still a challenge.

In this thesis, therefore, we will address the following big issue: How do the multicast and the strict delay constraints influence the design of reliable real-time multicast protocols by employing all kinds of EER techniques? By answering this question, we try to find a perfect solution to satisfy the diverse QoS requirements under strict delay constraints in the NG IP-based wireless networks.

1.1. State of the Art

Traditionally, there are two basic categories of EER techniques: Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC) erasure coding. The integrated FEC / ARQ schemes are referred to as Hybrid Error Correction (HEC) schemes in this thesis. The HEC schemes can be divided in to two categories: One is referred to Type I Hybrid ARQ (HARQ) scheme, which sends a certain amount of redundant packets in the first transmission; the other is referred to Type II HARQ scheme, which does not send any redundant packets in the first transmission but send parity packets when a retransmission is required. All of those EER techniques mentioned above will be discussed in detail in Chapter 2. In recent years, the researchers have contributed many reliable protocols using different EER techniques mentioned above.

With pure ARQ based techniques, Weldon's point-to-point scheme was the first selective repeat (SR) ARQ scheme that allowed multiple copies of frames to be sent [Wel82]. Afterwards, the studies have shown that the ARQ based error control schemes for point-to-point communication or single receiver in multicast scenario can outperform all the existing point-to-point schemes [Cha92]. In recent years, many researchers have considered extending the pure ARQ based schemes to reliable multicast protocols [Flo97] [Ott04] [Pej96] [Zor01]. The erasure error control in multicast protocols geared towards real-time multimedia applications have been proposed and analyzed in [Ott04] [Pej96]. It is found that the ARQ based schemes are appropriate for real-time applications, and actually can be quite effective [Pej96].

On the other hand, many researchers also had studied how to employ the FEC alone based EER schemes for reliable transmissions [Ada04] [Bye98] [Fuj04] [Riz98] [Yee95], in which many of them were designed for reliable multicast protocols (e.g. [Ada04] [Fuj04]). Especially, to adapt the variable channel conditions, more and more attentions have been paid on the adaptive solutions for EER in reliable real-time transmission protocols. For instance, by adapting the code rate of Reed-Solomon (RS) code [Mor02] to the variable channel condition, recently, some authors proposed Adaptive FEC (AFEC) for providing point-to-point real-time services [Ela98]; some other authors also proposed AFEC schemes for providing reliable real-time multicast services [Ana07] [Par98].

However, many studies had shown that HEC schemes can be much more efficient for recovering missing packets than the schemes with either FEC or ARQ alone [Den95] [Dja99] [Lin82] [Yan93]. For example, the Type I HARQ scheme has been employed for providing reliable real-time multicast services in [Rub99], and proposed for providing multicast services with completely reliability of zero loss in [Lee05]. Furthermore, the traditional Type II HARQ scheme has been proved to be very bandwidth-efficient for reliable non-real-time multicast to a large number of receivers [Ada04] [Maj02] [Non98]. To provide flexible and efficient error control for guaranteeing a certain PLR requirement of mobile multicast services, based on Type II HARQ technique, an adaptive HEC based mobile multicast protocol with low complexity graph-code was recently proposed in [Du05] [Zha06a].

Although up to now many EER schemes (e.g. those mentioned above) are proposed for designing reliable real-time multicast protocols, however, all of them did not consider taking the strict delay constraints as a fundamental limit. Therefore, using ARQ and Packet Repetition (PR) techniques, we developed a HEC-PR scheme for satisfying a certain PLR requirement under strict delay constraints and optimized its performance in [Tan07a]. Afterwards, it is found that the better performance can be achieved by combining the HEC-PR scheme with traditional Type I HARQ scheme [Tan07c] [Tan08a]. Nevertheless, there is still a critical question needed to be answered: under strict delay constraints, which scheme is optimal in a real-time multimedia multicast scenario for guaranteeing a certain PLR requirement? By addressing this question, we thus develop a general architecture of erasure error recovery combing all of the HEC schemes mentioned above in [Tan08b] [Tan08c]. Through optimizing the general architecture under strict delay constraints, the total needed Redundancy Information (RI) can be minimized by choosing the best scheme automatically among the entire schemes included in the architecture. As a result, in this thesis, we will take the strict delay constraints as a fundamental limit for designing the optimum HEC scheme for reliable real-time multicast protocols.

1.2. Thesis Outline

We now go through the remaining Chapters of the thesis one by one.

Chapter 2 firstly presents those existing traditional EER mechanisms: Pure ARQ, FEC alone and Hybrid ARQ (HARQ) schemes. Then, this Chapter proposes a general architecture of EER integrated all of the most important EER schemes. Afterwards, we will also introduce the parameters of the general architecture and interpret how the general architecture works in this Chapter.

Chapter 3 introduces two channel models: independent identical distribution (*i.i.d*) channel model and Gilbert-Elliott (GE) [Mus89] (originated in [Gil60] and [Ell63]) channel model, which are often used for evaluating the performances of different EER schemes over erasure error channels by many researchers. Then, depending on the evaluation results on the parameters of the channel model with a practical test bed, this Chapter explains why the GE channel model is more accurate than the i.i.d channel model in real systems. Therefore, we will use the GE channel model to evaluate the performances of all kind of EER mechanisms in this thesis. For the convenience of analysis and simulation works, this Chapter finally presents a method to produce the parameters of the GE channel model.

Chapter 4 presents how to calculate the end-to-end delay budget for different EER mechanisms in this thesis. These results will be used for evaluating the end-to-end delay for all kinds of EER schemes in this thesis. Upon these delay budgets, we then can design the optimum EER scheme considering the strict delay constraints.

Chapter 5 explains how the GE channel model applies to different EER schemes. It is very important to understand how we evaluate the performances of those EER schemes over GE channels.

Chapter 6 proposes an Adaptive FEC (AFEC) scheme to satisfy a certain PLR requirement of the real-time services under strict delay constraint with the minimum needed RI. Depending on the application of channel model for FEC schemes presented in Chapter 5, first, this Chapter presents how to compute the PLR performance of the packet level FEC schemes with ideal erasure codes. Using the knowledge of the end-to-end delay budget for FEC schemes introduced in Chapter 4, then, it proposes an optimization method for optimizing the parameters of the AFEC schemes. This Chapter finally presents the analysis results on the AFEC schemes by applying them in some typical scenarios.

Chapter 7 proposes a pure ARQ based HEC scheme with Packet Repetition (HEC-PR) technique, which satisfies a certain PLR requirement of the real-time services under strict delay constraint with the minimum needed RI. The HEC-PR scheme is mainly motivated by those multicast applications with small group size. Similar to Chapter 6, depending on the application of channel model for ARQ schemes presented in Chapter 5, this Chapter presents how to compute the PLR and RI performances of HEC-PR schemes. Using the knowledge of the end-to-end delay budget for ARQ schemes introduced in Chapter 4, then, it also proposes an optimization method for optimizing the parameters of HEC-PR schemes. This Chapter then presents the analysis results on HEC-PR schemes with some typical scenarios, and finally compares it with AFEC schemes proposed in Chapter 6.

Chapter 8 investigates a general mathematical framework to compute the PLR and RI performance of the general architecture proposed in Chapter 2. Similar to Chapter 6 and Chapter 7, depending on the application of channel model for the general architecture of EER presented in Chapter 5, this Chapter first studies how to calculate the PLR performance of the general architecture by combining some important probabilities. Then, it investigates how to calculate the total need RI of the general architecture. Finally, this Chapter obtains two explicit formulas to calculate the PLR and RI performances of the general architecture of EER, respectively.

Chapter 9 presents how to optimize the parameters of the general architecture based on the general mathematical framework achieved in Chapter 8. Using the knowledge of the end-to-end delay budget for the general architecture of EER introduced in Chapter 4, this Chapter also proposes an efficient greedy algorithm to searching for the optimum parameters of the general architecture.

Chapter 10 presents the analysis results of optimizing the general architecture under strict delay constraints depending on some typical multicast scenarios. Upon the analysis results, this Chapter presents how the multicast and the strict delay constraints influence the optimum architecture of EER. The effects of all kinds of systems parameters to the optimum architecture are also discussed in this Chapter.

Chapter 11 presents the simulation results for some typical scenarios with ns-2[3]. These simulation results will be compared with the analysis results to see if they can match with each other.

Chapter 12 concludes this thesis and discusses the possible future works by extending this work.

1.3. Thesis Contributions

As mentioned above, the strict delay constraints will influence deeply a wide variety of research areas such as NC, multicast routing, EER schemes etc. We thus take the strict delay constraints as a fundamental limit for designing reliable Real-time Multimedia Multicast (RMM) protocols in this thesis. Furthermore, since the RMM protocols usually must employ an EER scheme to overcome the packet loss problem in packet-based networks, this thesis will focus on the influence of the strict delay constraints on all kinds of EER schemes such as FEC, ARQ and HEC schemes and so on. Actually, both FEC and ARQ are two basic EER schemes, and any HEC scheme can be viewed as a combination of these two basic schemes. Those different existing EER schemes will be discussed in Chapter 2. Finally, as mentioned in Chapter 1.1, there is a very interesting question on this topic: Which scheme is the best scheme for any given multicast scenario? To answer this question, we propose a general architecture integrated nearly all of the existing EER schemes in Chapter 2.3.

Following the idea above, therefore, the main contributions of the thesis will address the following issues:

[3] The Network Simulator ns-2. http://www.isi.edu /nsnam/ns.

1. What is the accurate packet loss model in practical systems?

First of all, since the Packet Loss Model (PLM) plays a very important role in evaluating the performance of all kinds of EER schemes, we need to understand the actual PLM in practical systems. Therefore, we study the accurate PLM in practical systems by measurements over a real test bed in this thesis. It is found that the GE model with two-sate Markov chain is a good approximation for the PLM in our practical systems. The detail discussions on the PLM will be presented in Chapter 3. Base on this finding, we thus will analyze the performance of all kinds of EER schemes based on GE channel model in the whole thesis.

2. How do the strict delay constraints influence the optimization of pure FEC schemes?

In order to understand the influence of the strict delay constraints on the optimization on the pure FEC schemes, obviously, we need to know how to compute the end-to-end delay budget for this kind of schemes. Additionally, since the GE model is adopted for analyzing all kinds of EER schemes, we also need to understand how the GE model applied in this kind of schemes. Therefore, before analyzing the influence of the strict delay constraints on the pure FEC schemes, we will introduce how to calculate the end-to-end delay budget for them in Chapter 4.1, and present how to apply the GE model for them in Chapter 5.1. Depending on the knowledge of delay budget and channel model, afterwards, we will address this question in Chapter 6. Through answering this question, an Adaptive FEC (AFEC) scheme is proposed in this chapter, which can adapt to the variable channel conditions. The main advantage of the AFEC scheme is that no any feedback channel is needed when it applied in practical systems.

3. How do the strict delay constraints influence the optimization of pure ARQ schemes?

Similar to answer the question 2, we will first introduce how to compute the end-to-end delay budget for pure ARQ schemes in Chapter 4.2, and then present how to apply the GE model for them in Chapter 5.2. Afterwards, we study how to optimize the schemes with pure ARQ techniques under the strict delay constraints in Chapter 7. To distinguish with traditional ARQ schemes, the optimization of pure ARQ schemes proposed in this chapter is denoted by HEC-PR scheme in the whole thesis. The main attractive advantage of the HEC-PR scheme is the tiny complexity of its practical implementations, because the practical system with this scheme does not need any complicated encoding or decoding algorithm for recovering lost packets.

4. How do the strict delay constraints influence the optimization of the general architecture of EER?

Finally, we focus on the optimization problem of the general architecture of EER under strict delay constraints for any given multicast scenario. Similar to the question 2 and question 3, we will first introduce how to model the end-to-end delay budget for the general architecture of EER in Chapter 4.3, and then present how to apply the GE model in the analysis for the general architecture in Chapter 5.3. Due to the complexity of this question, afterwards, we answer this question through the following two chapters: the performance analysis on the general architecture of EER will be presented in Chapter 8 and its optimization problem will be addressed in Chapter 9. Through answering this question, we can obtain the optimum parameters of the general architecture for any given multicast scenario. According the current channel state information of a multicast scenario, since the optimization method can choose the best HEC scheme (keep in mind that both pure FEC schemes and pure ARQ schemes can be viewed as special cases of HEC schemes) automatically among all of the HEC schemes integrated in the general architecture, it is denoted by Adaptive HEC (AHEC) scheme in the whole thesis.

As a result, the contributions in this thesis provide a useful guide to design reliable RMM protocols. The analysis framework under the strict delay constraints contributed by this thesis also can be extended for further research in other research areas such as NC, multicast routing etc.

Chapter 2
Erasure Error Recovery in Bidirectional Channels

In this Chapter, we first discuss the packet loss issue in packet-based networks. To overcome the packet loss problem, then, we introduce the existing traditional erasure error recovery (EER) techniques in bidirectional channels. Finally, we propose a general architecture integrating all of the important existing EER techniques.

2.1. Packet Loss Issue

As mentioned in Chapter 1, packet loss issue is a major problem on decreasing the quality of Real-time Multimedia Multicast (RMM) over packet based networks. Since IPTV services have been claimed to be killer applications in the FMI, we take the IPTV services as a typical example to explain the packet loss issue for RMM services over packet based networks. A simplified protocol stack for IPTV service offerings is shown in Figure 2.1.

<table>
<tr><td colspan="5">IPTV Service Offerings</td></tr>
<tr><td rowspan="2">RTSP</td><td colspan="2">File Based Content</td><td colspan="2">Real Time Content</td></tr>
<tr><td>HTTP</td><td>FLUTE/LCT/ALC</td><td></td><td>RTP/RTCP</td></tr>
<tr><td colspan="2">TCP</td><td colspan="3">UDP</td></tr>
<tr><td colspan="2">IP Unicast</td><td colspan="3">IP Multicast</td></tr>
<tr><td colspan="5">Link Layer</td></tr>
<tr><td colspan="5">Physical Layer (Cable, Fiber, Wireless, etc.)</td></tr>
</table>

Figure 2.1: The simplified protocol stack for IPTV services delivery

As shown in this figure, the Internet Engineering Task Force (IETF) has defined components and protocols that enable delivering file-based content and real-time content to single or multiple users in parallel over IP-based networks. The Real Time Streaming Protocol (RTSP) [Sch98] can be used for choosing the actual delivery channels for transporting the audio/video (A/V) streams. For example, the RTSP provides a means to choose Transmission Control Protocol (TCP) [Rfc0b] for unicast non-real-time services or User Datagram Protocol (UDP) [Rfc0c] for real time multicast services.

For non-real-time services, the A/V streams are multiplexed and encapsulated in a file format. Then users can download them and perform local playback. For the unicast-transport of such file-based content in download delivery services via the HyperText Transfer Protocol (HTTP) [Fie97], the TCP can be used to guarantee the reliability of download. For the case of file-based content distribution over IP multicast, the IETF has specified a File Delivery over Unidirectional Transport (FLUTE) [Pai04] protocol for unidirectional file delivery. The FLUTE is an IETF protocol based on the Layered Control Transport (LCT) [Lub02a] protocol and the Asynchronous Layered Coding (ALC) [Lub02b] protocol.

For real-time services, the A/V streams are usually multiplexed into an MPEG-2 [Itu0a] transport stream in most deployed IPTV multicasting services (e.g. DVB-IPI proposed in [Ets0f]). Then, the MPEG-2 packets are encapsulated either directly in UDP or in Real-time

Transport Protocol (RTP) [Sch96] /UDP. The Real-time Transport Control Protocol [Sch96] is used for sending information to receivers on transmission statistics and feedback to the sender on the quality of the RTP packets distribution. As a result, for providing RMM services over IP based networks, UDP or RTP will be adopted.

For the support of high-quality RMM services over IP based networks, the physical layer usually provides reliability options such as various FEC schemes within an IP based network. However, at the physical layer, these error coding schemes generally have limits in terms of complexity and memory availability to overcome impulse noise and burst errors [Lub08]. Moreover, they do not provide end-to-end reliability across the network. Therefore, it is usually needed to employ error coding techniques at the link layer or upper layers to provide high-quality RMM services. Since the data transmitted at the link layer or upper layers are in the form of packets, the solutions at these layers are viewed as packet level solutions in this thesis. Note that a corrupted packet is generally detected and discarded at the link layer or the IP layer with Cyclic Redundancy Check (CRC), which results in a packet loss in this layer or upper layers [Tan02]. By checking the sequence number of those packets received, a receiver can exactly know which packets are lost. The loss or corruption of data packets thus can be viewed as erasure errors due to the receivers having the knowledge of the location of errors. It indicates that all of the traditional packet level EER techniques can be used to overcome the packet loss problem. By making use of those packet level EER techniques, therefore, this thesis focuses on studying how to guarantee the QoS requirements of RMM services over packet based networks in the most efficient way.

Actually, in practical implementations for some EER techniques with encoding blocks (e.g. FEC techniques), a certain amount of overhead messages must be added for each packet to indicate the location of the packet in an encoding block. The receiver then can recover those missing packets in the encoding block using the location information of lost packets. Note that in OSI model based networks the protocols are layered. As mentioned above, the error detection is carried out in lower layers (e.g. IP layer), and then the corrupted packets are discarded so that erasure error occurs at the upper layers (e.g. UDP layer). That is, the error location information will be given directly at the upper layers resulting in seeing an erasure error channel. Therefore, we only focus on the RI on top of erasure error channels in this thesis, and do not consider the effect of the overhead messages needed for error detections to the RI performance of EER techniques. In case that the overhead messages can not be neglected in practical systems, the framework of calculating the RI performance proposed in this thesis can be extended by adding this part for studying its influence in the future.

In addition, it is quite costly and also difficult to provide high-quality RMM services by employing packet level EER techniques at the link layer or the IP layer [Lub08]. Furthermore, neither UDP nor RTP presently provides mechanisms to guarantee the high-quality of RMM services. Recently, many researchers thus propose to use Application Layer (AL) EER techniques to reduce or eliminate packet loss to achieve the quality requirement of RMM services over IP based networks. For example, a pure ARQ based scheme at the RTP level was proposed in [Ott04]; Michael Luby et al. proposed to use the AL-FEC scheme in IPTV services [Lub08]; the HEC scheme was adopted by the NACK–Oriented Reliable Multicast (NORM) Protocol [Ada04]. Nevertheless, all of the EER techniques discussed in this thesis can be viewed as the solutions of from the link layer to the application layer. That is, they are suitable for practical implementations at the link layer or any of the upper layers.

2.2. Traditional EER Techniques

From Chapter 1, we have known that, in order to overcome the packet loss problem in bidirectional channels, there are two basic mechanisms: Automatic repeat request (ARQ) and forward error correction (FEC). In the following, based on these two basic mechanisms, we will introduce three kinds of traditional EER techniques used for the support of high-quality RMM services, respectively. To compare them with each other, we will also discuss the total needed Redundancy Information (RI) when applying those schemes in some multicast scenarios. For the convenience of analysis intuitively, lets N_{recv} denote the number of receivers in a multicast scenario, and it is assumed that all of the receivers are independent and have the same original link PLR of P_e in this section.

2.2.1 Automatic Repeat Request

Automatic Repeat reQuest (*ARQ*) is an error control method for data transmission which uses feedbacks and timeouts to achieve reliable data transmission. A feedback is a message sent by the receiver to the sender to indicate that it has lost or correctly received a data packet. A timeout is a reasonable point in time at the sender or the receiver. Both the sender and the receiver can use a timeout for detecting packet losses. As a major example, TCP is a typical ARQ based protocol using timeout at the sender, which is couple with RTT estimation for setting the reasonable values for the retransmission timer [Tan02]. In a multicast scenario, after the sender knowing which data packets lost, the sender will multicast the lost data packets to all of the receivers for recovering the missing data packets. Typically, an ARQ based scheme is composed of three parts: packet loss detection mechanism, feedback mechanism and retransmission mechanism. We now discuss different mechanisms for the three parts in the following, respectively.

1. Mechanisms of packet loss detection:

According to the entity that performs loss detection, sender-based and received-based mechanisms can be distinguished.

- For sender-based mechanisms, receivers return positive ACKs to indicate which data packets having been received. The sender is responsible for loss detection, and thus must maintain and process status information associated with every single receiver. Such a scheme obviously has problems in case of large number of ACKs, and these ACKs may cause congestion and losses in the sender's neighborhood [Car97].
- For receiver-based mechanisms, each receiver is responsible for loss detection, and therefore maintains its own state information. When a packet loss is detected, the

receiver sends a NACK to the sender by unicast or multicast to indicate which data packet needs to be retransmitted. The study shows that receiver-based mechanisms are far more scalable than sender receiver-based mechanisms for large groups and provide higher throughput [Tow97].

Actually, when the channel is in good state, the number of NACKs caused by receiver-based mechanisms is much less than that of ACKs caused by sender-based mechanisms. The reason is that the number of NACKs will be less than that of ACKs in case of the original link PLR being less than 50%. Since in practical wireless networks the original link PLR is far less than 50% (e.g. the worst PLR is only about 10% in WLAN with IEEE 802.11 [Fuj04]), receiver-based mechanisms generally produce much less feedbacks than sender-based mechanisms. It indicates that the processing load at the sender can also be reduced when applying receiver-based mechanisms. Due to this reason, most Real-time Multicast Transport Protocols (RMTPs) for high speed networks proposed to use receiver-based mechanisms (i.e. based on NACKs) rather than sender-based mechanisms (i.e. based on ACKs) [Arm92] [Bra93] [Ott04] [Pej96] [Pin94]. Because of the efficiency of this approach, in this thesis, it is always assumed that receiver-based mechanisms are adopted for packet loss detection in those retransmission based schemes.

2. Mechanisms of Feedback:

As mentioned above, the mechanisms of feedback are associated with the mechanisms of packet loss detection. That is, if sender-based loss detection mechanisms are adopted, receivers send positive ACKs to the sender; if receiver-based loss detection mechanisms are adopted, however, receivers send NACKs to the sender. Since it is assumed that receiver-based loss detection mechanisms are adopted in the whole thesis, receivers will always send a NACK if a packet loss detected.

In addition, receivers have two options for sending a NACK: by unicast only to the sender, or by multicast to the sender and all other receivers. On the one hand, some authors have suggested that receivers send NACKs to the whole group by multicasting for allowing the receivers to implement feedback suppression schemes [Pin94] [San90]. On the other hand, these alternatives are compared for RMTPs in [Pej96]. Multicasting NACKs is proved to be useful when receivers are locally concentrated with small propagation delay; when receivers are far apart with large propagation delay, however, multicasting NACKs is proved to have a negative impact on delay properties, due to the large propagation delay of NACKs to other receivers and to the time receivers must wait to achieve the desired suppression of NACKs [Pej96]. Therefore, these two schemes provide a way of tradeoff between the feedback suppression and the delay performance.

However, in this thesis, we only consider the total needed RI caused by the pure erasure error recovery while not consider the needed RI caused by those control messages (e.g. NACKs). Under this consideration, the RI performance is only associated with the total number of retransmission data packets at the sender in the pure ARQ based schemes. As a result, the mechanisms of sending NACKs have nothing to do with the performance analysis in this thesis. In other words, our analysis results are suitable for the retransmission based schemes with unicasting NACKs and/or multicasting NACKs. Nevertheless, we would like to suggest adopting multicasting NACKs to overcome the NACK-implosion problem at the sender in practical systems.

Finally, in order to see how good the optimum performance of the EER schemes with feedback channels can achieve, we assume that the feedback channel for NACKs is error-free in the whole thesis. In other words, the performance obtained in this thesis can be viewed as an "upper anchor" of the performance of the EER schemes with feedback channels. It is because the NACKs loss obviously decreases the performance of the EER schemes. The performance of the EER schemes with non-error-free feedback channels still needs further study. Nevertheless, we would like to point out that the core results in the thesis still can be used for many practical systems. The reason is as follows: On the one hand, since many practical systems give higher priority to control messages such as NACKs to guarantee the reliability of transmissions, the quasi error free (QEF) transmission of NACKs thus can be achieved so that the influence of the NACKs loss can be neglected; On the other hand, we also can overcome the NACks loss problem by adding a margin to the performance analysis for the EER schemes.

3. Mechanisms of Retransmission:

There are three conventional retransmission mechanisms: Stop-and-Wait (SW), Go-Back-N (GBN) and Selective Retransmission (SR) [Tan02].

- SW is the simplest kind of ARQ method, in which the sender sends a packet and then waits for ACKs from all of receivers before sending next packet. If the ACK of any receiver among all of the receivers does not reach the sender before a certain time, known as the timeout, the sender sends the same packet again. This mechanism is inefficient compared to other two mechanisms, because the time between packets, if the ACK and the data packet are received successfully, is at least the largest RTT among all of the receivers. Due to its low efficiency and long latency [Tan02], SW is not considered presently in all of the proposed RMTPs.

- In GBN based ARQ schemes, the sender continues to send a number of packets specified by a window size even without receiving an ACK from the receiver. The receiver keeps track of the sequence number (SN) of the next packet it expects to receive, and sends the SN with every ACK it sends. The receiver will simply discard any packet that does not have the exact SN it expects. Once the sender has sent all of the packets in its window, it will check if the SN of the last ACK it received equals to the SN of the final packet sent in this window. If yes, it indicates all of the packets in this widow have been received by the receiver correctly. If not, the sender will go back to the SN of the last ACK it received from the receiver process. Afterwards, the sender will retransmit all unacknowledged packets in order, starting with the SN of the last ACK it received.

Actually, GBN based ARQ schemes have been widely used for guaranteeing the reliability of transmission of non-real-time services. For instance, the basic protocol used by TCP entities is the sliding widow protocol [Tan02], which is actually a version of GBN based ARQ schemes. It has been proved that TCP can work very well for guaranteeing the reliability of transmission for non-real time services. However, TCP does not provide any mechanism to guarantee the end-to-end delay of transmission so that it is not suitable for real-time services with the strict delay constraints. Also, this approach can waste a lot of bandwidth if the PLR is high [Tan02]. The reason is clear: many of the packets received correctly by the receiver will also be retransmitted, because in this approach, if any packet is lost, this packet and all of the following packets in the same widow will be retransmitted. To avoid this problem, SR based ARQ schemes can be adopted.

- When SR based ARQ schemes used, at the receiver, a corrupted packet received is discarded and required retransmission at the sender, but the packets received correctly will be buffered and do not need any retransmission for them. By this approach, only those corrupted packets will be retransmitted at the sender, which obviously results in less RI than the GBN based schemes. As a result, the SR based schemes can be more efficient than the GBN based schemes [Tan02]. Actually, in most RMTPs, it is suggested to use a SR scheme rather than GBN due to its efficiency [Arm92] [Bra93] [Ott04] [Pej96] [Pin94]. Based on the analysis above, in our thesis, we will also adopt SR as the retransmission mechanism if needed.

As a result, when discussing an EER scheme with retransmission techniques in this thesis, rather than focus on a particular transport protocol, we shall consider a generic retransmission based scheme with the following features:

- An SR, NACK-only retransmission scheme is used;
- The feedback channel for NACKs is assumed to be error-free;
- All of the receivers are independent.

In wireless networks, the receivers can be viewed as independent due to different locations with different wireless channels. In some other networks (e.g. wired networks or the networks with multi-hops), however, packet losses of the receivers are usually correlated because of shared links [Lac00] [Mos00]. Note that there are two basic mechanisms to response NACKs received at the sender: The simplest way is that the sender retransmits all of the packets required by those NACKs from different receivers without repetition detection. In this case, the needed RI for a multicast scenario with independent receivers is same to for the multicast scenario with correlated receivers. The other way is more advanced: The sender only retransmits the different packets required by those NACKs from different receivers with repetition detection. In this case, the scheme can benefit from the correlation among all of the receivers in a multicast scenario. In other words, more the receivers are correlated, more the needed RI can be reduced. Especially, when all the receivers sharing an identical link, SR based ARQ schemes for this case can be viewed as used for point to point communication so that the best performance can be achieved. In this thesis, we will analyze the performance of the ARQ scheme with the simplest way in Chapter 7, and analyze its performance with the more advanced way in Chapter 10.

Obviously, the performance of the ARQ scheme mainly depends on its parameters: The number of retransmission rounds and the number of transmissions at each transmission stage. Note that the number of retransmission rounds is actually limited by the strict delay constraints when applying it for RMM services. The end-to-end delay budget for ARQ schemes will be discussed in Chapter 4.2. In addition, when multiple copies are allowed to be transmitted in each transmission stage, it can be viewed as a kind of HEC scheme with channel coding of Packet Repetition (PR) techniques. As mentioned in Chapter 1, to distinguish it with those traditional ARQ techniques, it is denoted by HEC-PR schemes in the whole thesis.

Finally, let's consider the needed RI when applying this general retransmission based scheme in a multicast scenario with N_{recv} receivers. To compare the performance of the ARQ scheme with other schemes intuitively, we assume that the simplest way is adopted at the sender for the retransmissions in this Chapter. Then, when all of the receivers are independent and have the same original link PLR of P_e, the RI caused by the first retransmission will be simply $N_{recv} \cdot P_e$ if those retransmission packets are retransmitted only once. Intuitively, we now can see that the total needed RI of the pure ARQ based schemes increase linearly with the number of receivers in the current multicast scenario. We will compare it with other schemes in the following sections.

2.2.2 Forward Error Correction

Forward Error Correction (*FEC*) is a method commonly used to overcome packet losses in RMM services over packet-based networks. FEC techniques enable a receiver to correct errors and/or losses without further interaction with the sender (i.e. no feedback channel needed). Usually, FEC codes can correct both bit-distortion errors and erasures. In coding theory, a **bit-distortion error** is defined as a corrupted symbol in an unknown position, while an **erasure** is a corrupted symbol in a known position [Mor02]. As mentioned in Chapter 2.1, however, we only address the packet loss issue in packet-based networks in this thesis. Packet losses can be viewed as erasure errors, because the location information of those lost packets can be found out at the receiver by checking the overhead messages of the packets received. It means that the receiver knows exactly the error positions. For this case, one packet can be viewed as one symbol in one FEC block. We thus call it packet-level FEC (PL-FEC) schemes in the whole thesis. In the following, we discuss those erasure codes widely used for RMTPs.

An erasure code has two most prominent input parameters: the number of source symbols, k, and the sum of source and parity symbols, n. In the whole thesis, an erasure code is denoted by (n, k) code. Since the symbols in erasure codes used in this thesis are equivalent to packets, for the conveniences of description, we will use the term “packets” instead of the term “symbols” in the following Chapters. An ideal (n, k) code is able to reconstruct the k source packets by using any subset of at least k out of the n packets. Note that the value of k can vary from one source block to the next in the PL-FEC solutions [Lub08]. According to the coding theory, a **systematic code** is one in which the input k source packets are embedded in the encoded output block of n packets; similarly, a **non-systematic code** is one in which the encoded output block does not contain the input k source packets [Mor02]. Actually, most of the erasure codes are linear codes, which can be used in systematic forms or non-systematic forms. But for overcoming the packet loss issue with PL-FEC schemes, however, a systematic code is more efficient than a non-systematic code. It can be explained by the failure decoding process for one encoding block of n packets: using a systematic code, part of the k source packets might be received correctly due to all the k source packets being embedded in the encoded output block of n packets; while using a non-systematic code, none of the k source packets can be received correctly due to there being no source packets in the encoded output block of n packets, which results in all of the k source packets lost. Therefore, we will assume that systematic erasure codes are adopted for the PL-FEC schemes in the whole thesis.

Reed-Solomon (RS) codes are a widely used type of erasure codes, which operate on non-binary symbols [Mor02]. An RS code with parameters of (n, k) can recover up to n-k lost packets. It thus performs an ideal erasure code. More importantly, most RS codes are available as systematic codes so that they can work efficiently. The encoding cost of RS codes can be expected to be a linear function of n-k. However, the decoding algorithm of RS codes has a cost of $O(kl^2)$ (where $l \leq \min(k, n\text{-}k)$ denotes the number of erasures) [Riz97] Due to the decoding complexity of RS codes, therefore, the most commonly used RS codes operate on symbols of bytes and restrict the code parameters as $k \leq n \leq 255$ [Riz97] [Roc04]. Nevertheless, we would like to point out that RS codes are still very suitable for many cases in reliable RMTPs. The reason is clear: When applying erasure codes for real-time services, the value of k usually has to be limited in a small range for many cases due to the strict delay constraints. It indicates that the code parameters with $k \leq n \leq 255$ can be satisfied in many practical systems so that the efficient RS codes can be adopted for them.

The limitations of RS codes have inspired the development of new codes offering large block lengths k [Mac05]. It is a general class of rate-less erasure codes and referred as fountain codes, which is developed on the base of Low Density Parity Check (LDPC) codes [Mor02]. LDPC codes were first introduced by Gallager in 1960 [Gal60] [Gal62]. However, LDPC has been almost completely forgotten for the next 30 years and has only been rediscovered in 1995 by MacKay and Neal [Mac99]. Afterwards, LDPC has been largely improved by Luby, Shokorollahi and et al. Their works lead to the fruitful inventions of such codes as Tornado [Bye98], Luby Transform (LT) [Lub02c] and Raptor [Sho06]. LT codes are the first efficient and practical fountain codes. However, LT codes can not give a construction with constant encoding and linear decoding cost without sacrificing the error probability [Lub08]. Raptor codes, an extension of LT codes, are a class of fountain codes with fixed encoding and linear decoding cost [Sho06]. Raptor codes also have very low reception overhead and are available as systematic codes. These properties make Raptor codes very attractive for the purposes of IPTV data communication. As a result, Raptor codes have been standardized by DVB for IPTV applications [Ets0g] and by 3GPP for MBMS services [3gp05a].

Despite the excellent performance of Raptor codes for large block size and manageable encoding/decoding complexity, they are not idea erasure codes. Although ideal erasure codes with very large block sizes are theoretically possible, they are presently impractical in terms of memory usage and computational complexity. The non-ideal on Raptor codes can be realized by the so called reception overhead ε (where $\varepsilon > 0$): Instead of only exactly k packets needed to recover the k source packets by ideal erasure codes, Raptor codes require on average $(1+\varepsilon)k$ packets to recover the k source packets [Sho06]. It means that Raptor codes are less bandwidth-efficient than RS codes. Furthermore, considering the computational complexity of the Raptor codes, each output symbol is generated using $O(\log(1/\varepsilon))$ arithmetic operations, and the original symbols are recovered from the collected ones with $O(k \cdot \log(1/\varepsilon))$ arithmetic operations [Sho06].

In this thesis, without specifying practical erasure codes used in PL-FEC schemes, it is always assumed that **ideal systematic erasure codes** used in PL-FEC schemes. Obviously, an ideal erasure code is taken as the “upper anchor” of the performance of erasure codes. Recently, the IETF Multicast Reliable Transport (MRT) working group[1] performed evaluation

[1] http://www.ietf.org/html.charters/rmt-charter.html

of FEC codes. The evaluation results show that LDPC based codes are more suitable for large block transfers over unidirectional channels, while RS codes are more appropriate for small block sizes and real time services [Naf08]. Considering the practical implementations, therefore, RS codes can be considered in case of small block sizes (i.e. $k \leq n \leq 255$) and fountain codes (e.g. Raptor codes) can be adopted in case of large block sizes.

Now we focus on the total needed RI of a PL-FEC scheme with an ideal (n, k) erasure code in a multicast scenario with N_{recv} receivers. When an ideal (n, k) code used, the total RI can be calculated by simply $(n-k)/k$. In contrast to ARQ schemes, since parity packets can be used for recovering different lost packets at different receivers, the total RI of PL-FEC schemes has nothing to do with the number of receivers. The reason is that the ideal (n, k) code only needs to be designed according to the PLR requirement of the receiver with the worst link PLR, and then it can guarantee the same PLR requirement for all other receivers.

In the following, we take an example to compare PL-FEC schemes with ARQ schemes. Without loss of generality, we consider applying different schemes for a multicast scenario with N_{recv}=3 and P_e=0.01. It is assumed that both a PL-FEC scheme with the ideal (k+1, k) erasure code and an ARQ scheme with only one retransmission round can satisfy the PLR requirement of those receivers. Therefore, the total needed RI of the PL-FEC scheme is $1/k$, while the total needed RI of the ARQ scheme is $N_{recv} \cdot P_e = 0.03$. Note that the parameter k is variable due to the variable multicast data rate, packet size, RTT, strict delay constraints and so on. If k is set to 20, then the RI of the PL-FEC scheme will be 0.05, which is more than the needed RI by the ARQ scheme; however, if k can be set to 50, then the RI of the PL-FEC scheme will be 0.02, which is less than the needed RI by the ARQ scheme. From this example, it is found that ARQ schemes outperform PL-FEC schemes in some cases; while in some other cases PL-FEC schemes outperform ARQ schemes. Motivated by these findings, in this thesis, we will study how to find the best scheme under a certain multicast scenario among ARQ schemes, FEC schemes and Hybrid ARQ schemes (see Chapter 2.2.3) by looking up the optimum parameters for them.

2.2.3 Hybrid ARQ

Hybrid ARQ (*HARQ*) are a type of HEC schemes integrated FEC / ARQ techniques. The first idea for a system that allowed for both error correction and error detection with retransmission requests was introduced by Wozencraft and Horstein [Woz61]. Their system, now known as a HARQ scheme, provides significantly improved performance over pure ARQ schemes. Afterwards, many researchers studied applying all kinds of bit-level HARQ techniques which generally implemented in physical layer for repairing both bit-distortion errors and erasure errors, some known as Type I HARQ schemes in which the sender proactively sending a certain amount of redundant bits [Cha85] [Dru83] [Mor89] [Sin77] [Yam80], some others known as Type II HARQ schemes in which the sender only resending redundant bits in case of receivers required [Lin82] [Pur91] [Wic94] [Yan93]. The overviews of the development of such bit-level HARQ schemes can be found in [Cos98] [Lot07].

Nowadays, the ideas of HARQ techniques have been absorbed by many reliable multicast protocols to overcome the packet loss problem, and also have been proved to be very efficient [Ada04] [Den95] [Du05] [Non97] [Rub98] [Zha06a] [Zha07]. These HARQ schemes only address erasure errors for reducing or eliminating lost packets. They are thus packet-level (PL) solutions. As mentioned above, we address only erasure errors in this thesis. Therefore, only packet-level HARQ (PL-HARQ) schemes are considered in the whole thesis. Traditionally, PL-HARQ schemes can also be divided into two categories: Type I PL-HARQ schemes and Type II PL-HARQ schemes. To simplify the description in the thesis without confusions, we will still use the term "HARQ schemes" while not the term "PL-HARQ schemes" in the following Chapters. We now explain them applied in multicast scenarios in more detail:

1. For Type I HARQ schemes, the sender sends a certain amount of parity packets in the first transmission. If the final PLR after reconstruction at a receiver is still too high, ARQ will be used to retransmit. In type I HARQ schemes proposed in [Den95] [Tan07c], it is suggested that only data packets are retransmitted if these repairs for lost packets are insufficient. However, studies have shown that retransmitting parity

packets are more efficient than retransmitting data packets for real-time multicast services [Rub98] [Tan08b]. This is because all of the receivers can benefit from those retransmission parity packets required by any receiver among them. Therefore, in our thesis, all of the retransmission packets are assumed to be parity packets when Type I HARQ schemes adopted.

2. For Type II HARQ schemes, the sender does not send any parity packets in the first transmission, but send parity packets when a retransmission is required by a receiver. As shown in [Non97], this approach is very bandwidth-efficient for completely reliable multicast to a large number of receivers. By employing LDPC codes in Type II HARQ schemes, recently, Du and Zhang et al. found that this approach is also efficient for real-time multicast services [Du05] [Zha06a] [Zha07]. We thus also consider employing this approach for designing suitable RMTPs in this thesis.

Apparently, the performances of these two HARQ schemes depend on their parameters used. Note that the strict delay constraints will limit both the encoding block size and the number of retransmission rounds. The detail information on the delay budgets for different schemes can be found in Chapter 4. When Type I HARQ schemes used for RMTPs, the performances depend on the following parameters: the number of retransmission rounds; the number of source packets in one encoding block (denoted by k); the number of parity packets sent in the first transmission (denoted by N_p) and the number of parity packet resent in different retransmission stage. For Type I HARQ schemes, the total needed RI thus is composed of two parts: one of them is the needed RI in the first transmission, which can be computed by N_p/k; the other part is the needed RI caused by the retransmission parity packets during all of the retransmission rounds. When Type II HARQ schemes used for RMTPs, the performances also depend on the parameters as same as Type I HARQ schemes by setting $N_p \equiv 0$. For type II HARQ schemes, therefore, the total needed RI only depends on total number of retransmission parity packets during all of the retransmission rounds.

From the analysis above for the total needed RI of these two schemes, we can find that it is really hard to say which scheme perform better for a given multicast scenario. Nevertheless, we can compare them with each other in an intuitive way. It is assume that only one retransmission round is allowed for a real-time multicast scenario with N_{recv} receivers, and both of the HARQ schemes can satisfy the target PLR requirement. In case that P_e is far less than $1/k$ so that $N_{recv} \cdot P_e$ is also less than $1/k$, the total needed RI of Type I HARQ schemes obviously is more than $1/k$ due to at lease one parity packet transmitted in the first transmission. When Type II HARQ schemes adopted for this case, however, the total needed RI is at most $N_{recv} \cdot P_e$ due to all of the receivers sharing those retransmission parity packets for recovering different lost packets. Therefore, Type II HARQ schemes can perform better than Type I HARQ schemes. On the other hand, in case that P_e is equal to $1/k$, the total needed RI of Type I schemes with N_p=1 can be only slightly more than $1/k$, because most of lost packets have been recovered in the first transmission stage. It means that the number of parity packets required during the retransmission stage is very small. But for Type II schemes, the total needed RI can be far more than $1/k$ when there are large number of receivers in this multicast scenario, because each receiver requires retransmitting parity packets independently so that many of them can not share the parity packets due to different lost time. Usually, the parameter k is small due to the strict delay constraints. It indicates that $1/k$ is generally relative large. Intuitively, therefore, Type I HARQ schemes outperform Type II HARQ schemes in case of large average link PLR (e.g. no less than $1/k$); while in case of small average link PLR (e.g. far less than $1/k$), Type II HARQ schemes will outperform Type I HARQ schemes. This intuitive result has been proved to be correct in our studies [Tan08b] [Tan08c].

2.2.4 Discussions

Generally speaking, for reliable multicast transport under strict delay constraints, HARQ schemes are more bandwidth efficient than pure ARQ or FEC alone schemes. However, for some special cases, FEC alone schemes could be the best scheme, e.g. no retransmission round is allowed. For some other special cases, pure ARQ could be the best scheme, e.g. there is only one receiver in a multicast scenario or the entire receivers having high correlation. That is, given a certain multicast scenario under strict delay constraints, it is possible for any of those schemes could be chosen as the most efficient scheme. Our main task in this thesis is to find the perfect solution for any given multicast scenario under strict delay constraints. In the following, we will take some examples to discuss the performance of the different EER schemes introduced above.

First, for comparing HARQ schemes with pure ARQ based schemes, Figure 2.2 gives an example where different packets are lost for different receivers in the first transmission in a multicast scenario with three receivers. As shown in this figure, two different mechanisms are used for recovering the lost packets in different receivers. One is NACK-only, SR based ARQ scheme (i.e. Scheme "A" as shown in Figure 2.2), in which the sender multicasts those different data packets required by all the receivers. Note that if all the receivers lose the same data packet in the SR based ARQ scheme, the sender send only this data packet required while not other data packets. The other is a Type II HARQ scheme (i.e. Scheme "B" as shown in Figure 2.2), in which the sender retransmits only one parity packet allowing all the receivers to recover their different lost packets. From this figure, we can see that the total needed RI is 100% when the ARQ scheme used, while the RI is only about 33.3% when the Type II HARQ scheme used. Therefore, in case that the receives are independent in a multicast scenario so that their lost data packets are often different, Type II HARQ schemes perform much more efficiently than ARQ schemes, because parity packets retransmitted can be used for recovering different lost packets at different receivers.

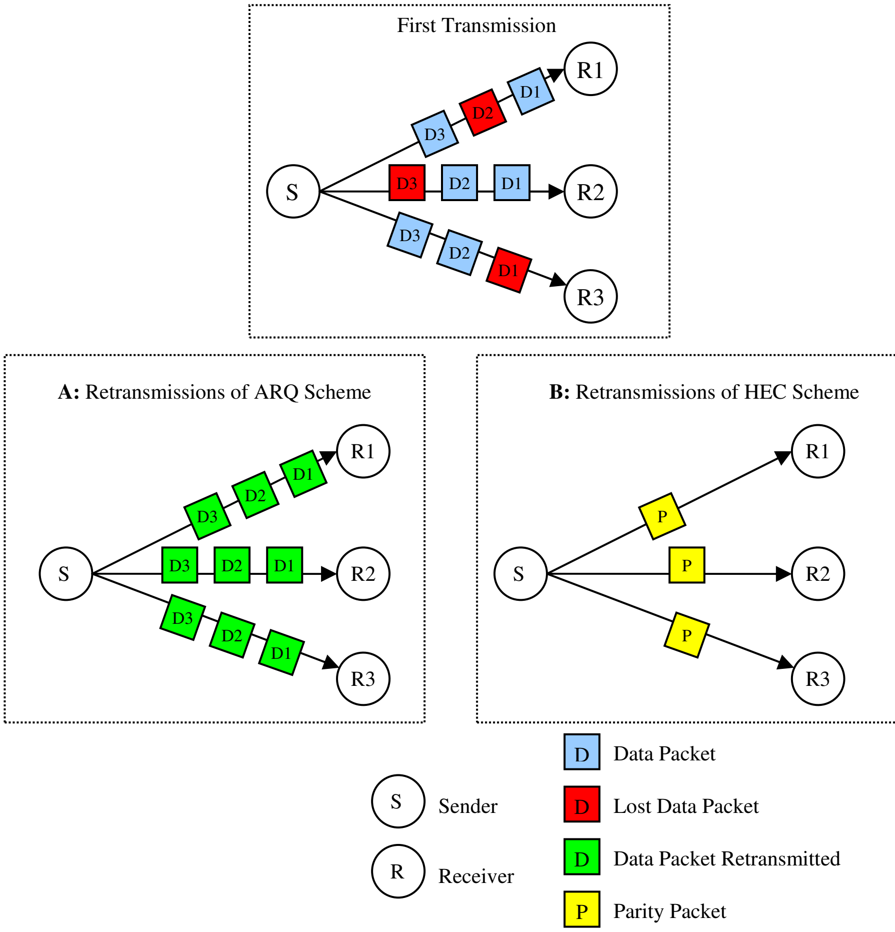


Figure 2.2: Retransmissions of ARQ Scheme and HARQ Scheme

However, under strict delay constraints, the performance of the ARQ scheme and the HARQ scheme is associated with the RTT and the length of packet interval. Note that the end-to-end delay of the ARQ scheme only depends on the RTT, and the end-to-end delay of the HEC scheme depends on both the RTT and the length of packet interval due to the FEC blocks. Considering a multicast scenario with very small RTT but very large packet interval that more than the strict delay constraints, the ARQ scheme obviously is the best and the exclusive choice. However, in case that a multicast scenario has very large RTT (e.g. only one retransmission round is allowed) but very small packet interval, the FEC scheme or the HARQ scheme should be the better choice. As a result, under the strict delay constraints, the performance of all of kinds of EER scheme is associated the end-to-end delay budget for them, which will be discussed in Chapter 4.

Then, we compare HARQ schemes with FEC schemes. Considering a case that no retransmission round is allowed for a multicast scenario due to the strict delay constraints and large RTT, a pure FEC scheme obviously is the best and exclusive choice. Theoretically, FEC schemes can be always the best choice if the block size is large enough to achieve the Shannon limit [Mor02]. However, when applied FEC techniques in real-time services, the FEC block size generally has to be limited in a reasonable small range due to the delay limits. It indicates that FEC schemes with erasure codes of less efficient code rate usually has to be adopted for these cases so that HARQ schemes with retransmissions can be more efficient than pure FEC schemes.

We now take an example to compare HARQ schemes with FEC schemes. Note, for HARQ schemes with retransmissions under the same delay constraints, the FEC block size has to be set shorter than that used in pure FEC schemes. Without loss of generality, it is assumed that there is a multicast scenario with N_{recv}=10 and P_e=0.001, in which the parameter k can be set to 50 when pure FEC schemes used; while the parameter k can only be set to 20 due to large RTT when HARQ schemes with one retransmission round used. For this case, we assume that a pure FEC scheme with the ideal (51, 50) erasure code can satisfy the target PLR requirement. Then, the total needed RI of the pure FEC scheme is 0.02 (i.e. 1/50). Obviously, when Type I HARQ schemes applied for this case, the total needed RI will be at least 0.05 (i.e. 1/20). When Type II HARQ schemes adopted for this case, however, the total needed RI will be less than $N_{recv} \cdot P_e = 0.01$ if only the minimum number of parity packets needed to be retransmitted in the unique retransmission round. For this case, therefore, Type II HARQ schemes will be the best choice. In addition, as analyzed previously, Type I HARQ schemes can outperform Type II HARQ scheme in case of multicast scenarios with large link PLR. For this case with P_e=0.05, without loss of generality, we assume that Type I HARQ schemes outperform Type II HARQ schemes. For this case, we assume that a pure FEC scheme with the ideal (54, 50) erasure code can satisfy the target PLR requirement. Then, the total needed RI of the pure FEC scheme is 0.08 (i.e. 4/50). When a Type I HARQ scheme applied for this case, it is assumed that the parameter N_p can be set to one to guarantee the target PLR requirement. The total needed RI of the Type I HARQ scheme will be only slightly more than 0.05 (i.e. 1/20), because the number of the retransmission parity packets are small after the recovery process in the first transmission stage. Therefore, Type I HARQ schemes will be the best choice for this case with large link PLR. According to this example, we can conclude that for some scenarios Type I HARQ schemes will be the best choice while for some other scenarios Type II schemes being the best choice.

To sum up, from the discussions for different schemes above, we know that the performance of different schemes depends on their parameters used for a given multicast scenario. For pure ARQ schemes, there is only one parameter: The number of retransmission rounds. For so called HEC-PR schemes based on pure ARQ techniques, besides the number of retransmission rounds, there are additional parameters: the number of copies transmitted for each data packet at each transmission stage. Pure FEC schemes with ideal erasure codes have two parameters: The number of source packets in one encoding block and the sum of source and parity packets in the encoding block. For HARQ schemes integrated ARQ and FEC techniques, they have the most parameters: The number of source packets in one encoding block, the number of retransmission rounds and the number of parity packets sent at each transmission stage. Note that the strict delay constraints for RMM services will limit the range of the values for two parameters: The number of retransmission rounds and the number of source packets in one encoding block. Therefore, the strict delay constraints have a significant effect on the performance of those schemes which depending on one of the two parameters or both. In order to guarantee a predefined target PLR requirement under the strict target delay constraints with those EER techniques introduced above, for a given real-time multicast scenario, there are two big questions need to answer: What are the optimum parameters for each of them and which one is the best scheme? To answer these two questions, in this thesis, we propose a general EER architecture integrated overall EER techniques introduced above. Through analyzing the performances of the general architecture and optimizing its parameters, we finally can answer these two questions.

2.3. General Architecture

As mentioned before, there is a very interesting question not being answered before: Which EER scheme is the best scheme under a certain real-time multicast scenario? Motivated by answering this question, we propose a general EER architecture in this Chapter, which has integrated nearly all of the EER techniques introduced above. The framework of the general EER architecture is shown in Figure 2.3.

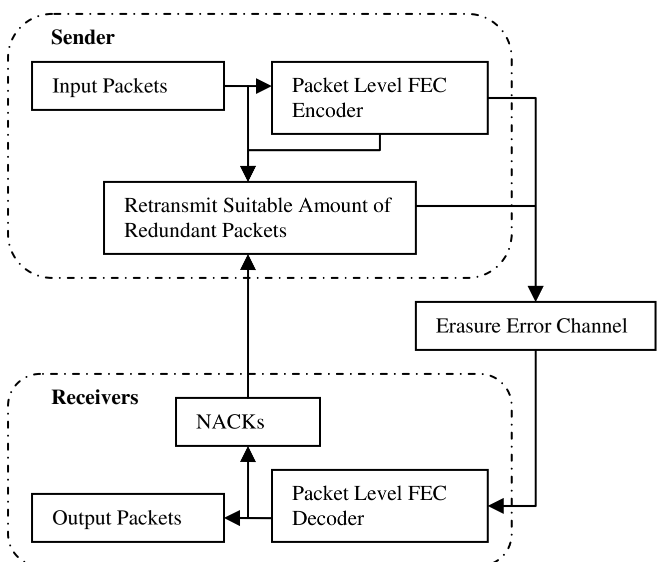


Figure 2.3: The General Architecture of Erasure Error Recovery

As shown in this figure, the sender first transmits encoding blocks to all receivers using the packet level FEC encoder (keep in mind that the scheme with each encoding block of single packet is identical to not applying FEC). In this thesis, it is always assumed that perfect erasure codes (e. g. RS codes) are used and the number of source data packets is k in one encoding block. That is, upon reception of any k packets of one encoding block, one receiver can recover all the data packets. Otherwise, the receiver will send NACKs to the sender to repair the missing data packets. To explain the architecture in more detail, we now define the essential parameters of the architecture in the following table:

Table 2.1: Parameters of the general Architecture

Symbol	Definition
k	the number of source data packets in one encoding block, where $k \geq 1$
N_p	the number of redundant packets in one encoding block in the first transmission
$N_{rr,max}$	the maximum possible number of retransmission rounds
N_{cc}^{q}	a constant coefficient in the q-th (where $1 \leq q \leq N_{rr,max}$) retransmission round, which is the number of multiples for the number of required redundant packets in the q-th retransmission round

Remarks: The general architecture has integrated nearly all of the existing EER techniques, which can be understood as follows: Considering the parameters of the general architecture, in case of $k>1$, the architecture acts as three different schemes: First, if no any retransmission round needed (i.e. $N_{rr,max}=0$), the architecture acts as a pure FEC scheme; secondly, in case of $N_p>0$, this architecture behaves as the traditional Type I HARQ scheme; Thirdly, in case of $N_p=0$, this architecture actually acts as the traditional Type II HARQ scheme. Especially, when k is set to one, no any parity packet will be produced so that this architecture acts as pure ARQ based schemes. In this case, the redundant packets during all of the retransmissions are always the copies of source data packets. As a result, this architecture indeed integrates all of the EER techniques introduced before.

Based on an example shown in Figure 2.4, in the following, we explain the operations of the architecture in more detail:

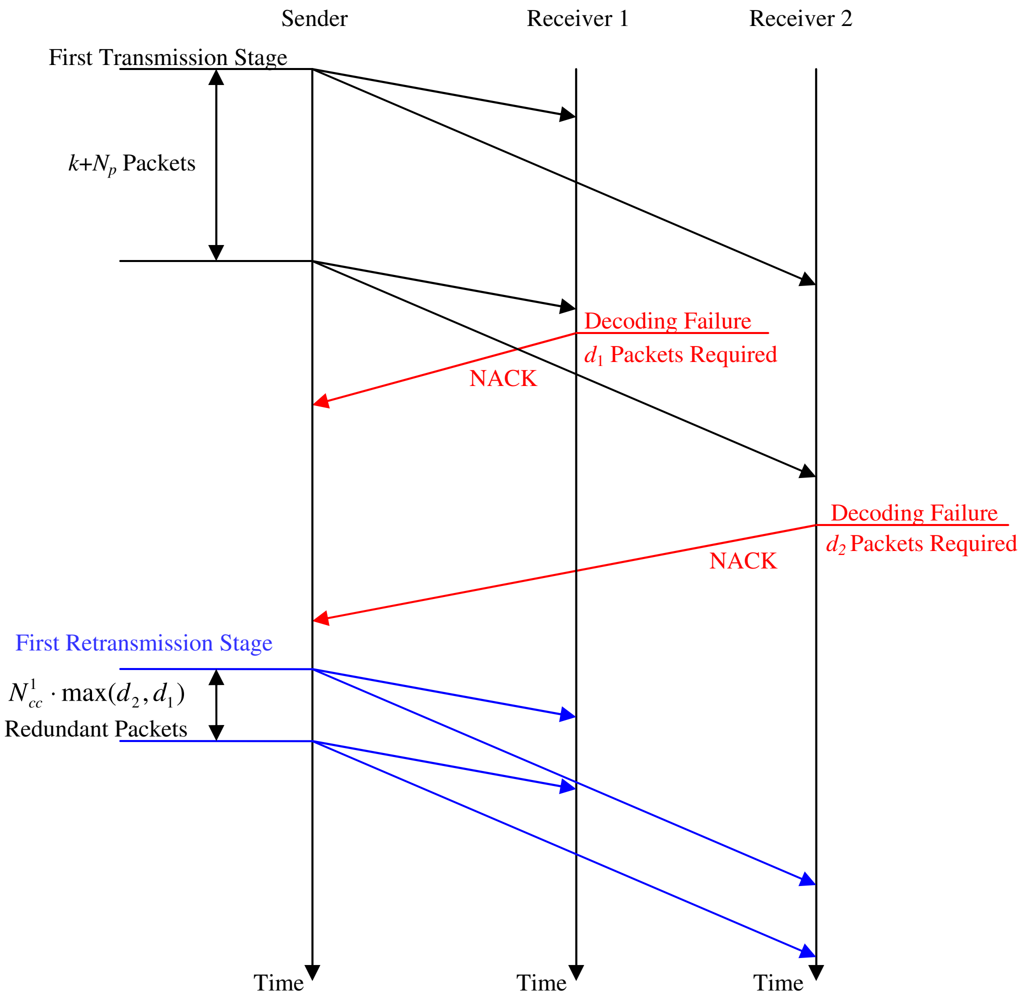


Figure 2.4: An Example of The First Retransmission Round on the General Architecture

1. As shown in this figure, in the first transmission, the sender transmits the packets in the form of FEC encoding block of k+N_p packets (i.e. each encoding block includes k data packets and N_p redundant packets) to all the receivers immediately.

2. At the receiver, if no less than k packets for one encoding block are received, it can recover all of the k data packets and forward them to the application immediately. Otherwise, the receiver firstly calculates the amount of the essential redundant packets needed for this block. In case of the NACKs transmitted by multicast, it then can check the NACKs from other receivers for this retransmission round. If the receiver finds that

the amount of redundant packets required by any NACK message is no less than itself, it will do nothing. Otherwise, the receiver feedbacks a NACK message to the sender. By this way, those NACKs can be suppressed efficiently for overcoming the NACK-implosion problem. From Figure 2.4, we can see that these two receivers required d_1 and d_2 redundant packets separately for the first retransmission round.

3. Upon getting those NACKs for one encoding block during each retransmission round, the sender will multicast a certain number of redundant packets to all of the receivers immediately with one copy (or multiple copies in case of k=1) of these retransmission packets. Obviously, the number of redundant packets sent at the sender in each retransmission round plays a very important role in the performance of the architecture. We thus take an example to show how to calculate the number of redundant packets in each retransmission round at the sender. Without loss of generality, in the q-th retransmission round, it is assumed that the sender receives a NACK for one block with maximum h redundant packets (e.g. h=max(d_1,d_2) in the example as shown in Figure 2.4). Note that in case of k>1, the redundant packets will be parity packets, which should not be repeated in any retransmissions. In contrast, in case of k=1, the redundant packets will be always multiple copies of source data packets, which produce no any parity packet in any retransmissions. Based on the parameters of the architecture, therefore, the sender will retransmit $N_{cc}^{q} \cdot h$ different parity packets in case of k>1 or N_{cc}^{q} copies of the h source data packets required in case of k=1 to all of the receivers.

According to the description above, we now know that the performance of the general architecture mainly depends on its parameters as shown in Table 2.1: such as the number of retransmission rounds; the number of redundant packets with the first transmission and retransmissions; the number of copies of redundant packets within each retransmission round and so on. Our remaining tasks are to theoretically analyze the performance of the architecture and then optimize its performance for satisfying a certain PLR requirement under strict delay constraints with minimum total needed RI.

To investigate the performance of the architecture under strict delay constraints, we need to introduce some essential preliminary knowledge: The channel model used in the analysis and simulations will be presented in Chapter 3; the end-to-end delay budgets on different schemes will be discussed in Chapter 4; and Chapter 5 will present how to apply the features of the channel model in our analysis work for different EER schemes.

Chapter 3

Channel Model

Recalling the introduction in Chapter 1, we know that wireless channels are error-prone channels, and we have to employ some error recovery schemes to provide acceptable quality for multicast services over wireless networks. Usually, we employ bit-wise channel coding in physical layer for increasing the transmission reliability. However, the bit-wise channel coding can not deal with burst errors longer than code words. This problem will cause the corruption of packets in Layer 2 (OSI Model) or higher layers so that packet loss happens in those layers. As introduced in Chapter 2.1, the packet loss issue can be viewed as erasure error in Layer 2 or higher layers. This problem can be overcome by an alternative approach, where packet level erasure error recovery (EER) schemes introduced in Chapter 2 are considered. Using packet level channel coding instead of bits, packets are seen as code symbols in those EER schemes. This thesis will address the error coding schemes over erasure error channel, which actually are the solutions in Layer-2 or higher layers for guaranteeing transmission reliability in wireless networks. Since the packet loss is a kind of typical erasure error, we will use these two terms "packet loss" and "erasure error" in the whole thesis without difference.

To demonstrate Shannon limit [Sha48] for channel coding in erasure error channel, an erasure error channel with erasure error rate of P_e is shown in Figure 3.1.

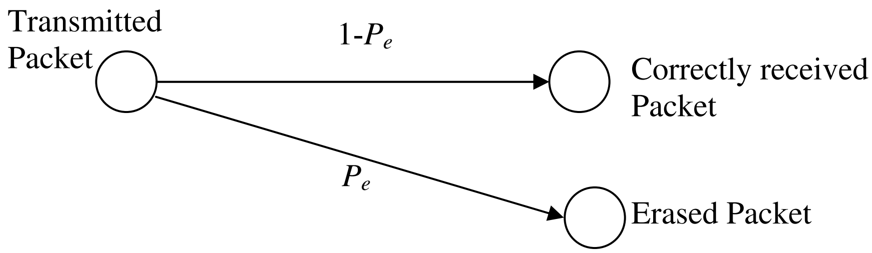


Figure 3.1: Erasure Error Channel

As show in Figure 3.1, the probability of the transmitted packet erasure is P_e, and the probability of the packet received correctly is 1- P_e. Therefore, the PLR caused by this erasure error channel will be P_e. Now let n denote the length of the code word used by a channel code, and lets k denote the number of information symbols in the code word. Then, the code rate of the channel code can be expressed as k/n. According to Shannon's coding theorem, the code rate of channel coding used has to be less than 1- P_e for reaching reliable transmission with the original link PLR equal to P_e [Lin83]. Therefore, the needed RI (denoted by the occupation rate of original useful information, i.e. $(n$-$k)/k$) of the Shannon limit will limited by:

$$\frac{k}{n} \leq 1 - P_e \Leftrightarrow \frac{n}{k} \geq \frac{1}{1-P_e} \Leftrightarrow \frac{n-k}{k} \geq \frac{P_e}{1-P_e} \tag{3.1}$$

Upon (3.1), the Shannon limit on the minimum needed RI can be calculated by:

$$RI_{\lim} = \frac{n-k}{k} = \frac{P_e}{1-P_e} \tag{3.2}$$

Note that the RI_{lim} is only the minimum needed RI for erasure error coding, which does not include the needed RI for those messages of transmitting the needed parameters of the erasure error coding etc., and also does not include the needed overhead messages for indicating the erasure error locations. By making use of the general architecture of EER proposed in Chapter 2.3, this thesis will focus on how to minimize its RI performance and close to the Shannon limit dynamically by adapting the current channel state.

According to the memory features of the channel, the erasure error channel model can be divided into two categories: memory-less erasure error channel model and memory erasure error channel model. In the following, we will introduce two classical channel models which widely used in the analysis works by most of the researchers: one is a memory-less channel model with "independent and identically distributed" (i.i.d) feature; the other is a memory channel model with two-state Markov chain.

3.1. i.i.d Channel Model

The i.i.d channel model is actually a kind of memory-less channel, which infers that the erasures among different packets are uncorrelated. Generally speaking, in this channel, the loss probability of each packet is independent and identical, where we do not need to consider the types of the packet transmitted (e.g. the packet is in original stream or in retransmission stream; the packet is data packet or parity packet etc.). This feature will simplify the analysis work very much for those retransmission based HARQ schemes.

Now let's take an example for explaining the i.i.d channel model in more detail. Let P_e be the link PLR of an erasure error channel with the i.i.d channel model. It is assumed there are total m packets sent over this channel. Since all of the m packets have the same loss probability of P_e, the probability of h packets lost among the m packets can be calculated by a simple Bernoulli formula:

$$\Pr(h \text{ packets lost} \mid m \text{ packets sent}) = \binom{m}{h} (P_e)^h (1 - P_e)^{m-h} \tag{3.3}$$

Using the Bernoulli formulas, it will be very convenient to evaluate the performances of all kinds of EER mechanisms over the i.i.d channel model. Based on this channel model, we could have a look on the optimum performance by integrating different EER mechanisms used in RMTPs [Tan07a] [Tan08a].

However, the erasures among those packets sent are usually correlated in wireless networks due to collisions, inferences etc. Many recent studies, furthermore, have shown that the simplified Gilbert-Elliott (GE) model is a very good approximation for the packet loss model in a wireless channel [Kha03] [Tan07b]. Therefore, we should evaluate those EER schemes in wireless channels using the much more accurate packet loss profile based on the GE model. Therefore, to evaluate the performance of the EER schemes over wireless networks more accurately, many researchers model the erasure error channel model as GE model with memory feature. In the following, we will introduce this kind of memory erasure error channel model and evaluate its accuracy by experiments in practical test bed.

3.2. GE Channel Model

In recent years, many researchers have proved that Markov Models with memory features can be used for modeling wireless channels very well. For example, the Finite State Markov Channel (FSMC) models were proposed to model the bit-level errors for Raleigh fading channels [Tan00]. The two-sate Markov chain was proved to be effective to model frame losses in slow fading channels [Tur02] [Tur99] [Zor95] [Zor98]. Furthermore, a Markov-based Trace Analysis (MTA) model algorithm for modeling the frame losses in GSM networks was proposed in [Kon01] [Kon03]. Ji et al. compared the performance of the MTA, a full state *k*-order model, a Hidden Markov Model (HMM) and an extended ON (error-free) / OFF (error-filled) model in capturing the frame losses of GSM-based networks [Ji04]. Similarly, with the rapidly growing ubiquity of WLANs based on IEEE 802.11, Khayam et al. proposed a hierarchical Markov Model for the byte-level errors in IEEE 802.11b link layer [Kar03] [Kha03]. Also, the FSMC model was proposed for modeling the frame losses in the link layer of IEEE 802.11b and IEEE 802.11a [Ara03]. Although the two-state Markov chain was claimed to be inadequate for modeling the frame losses in GSM-based networks [Ji04] [Kon01] [Kon03], and also inadequate for modeling the bit-level or byte-level errors in WLANs with IEEE 802.11 [Ara03] [Kar03] [Kha03], the simplified Gilbert-Elliot (GE) Channel Model with the two-state Markov chain was proved to be adequate for modeling the busty packet losses in IEEE 802.11a [Tan07b] and in IEEE 802.11b [Kar03] [Kha03]. In this thesis, we will model the busty packet losses in wireless networks as the simplified GE model to evaluate the performance of all kinds of EER techniques. Although the GE channel model might be not suitable for many other types of wireless networks (e.g. GSM, CDMA, WiMAX etc.), it is expected to have the similar results as analyzed in this thesis.

The GE channel model is a two-state Markov chain with one "Good" (G) state and one "Bad" (B) state as shown in Figure 3.2.

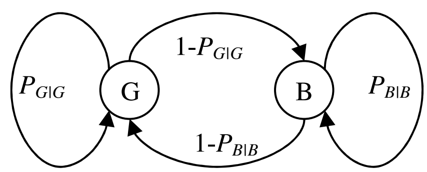


Figure 3.2: Hidden Markov Model of the GE Channel Model

Where each state corresponds to a specific channel quality, in other words, in the "G" state the PLR is very low (denoted by PLR(G)) while in the "B" state the PLR is very high (denoted by (PLR(B))). In this thesis, it is assumed that PLR(G)=0 and PLR(B)=1 because it corresponds to the most reasonable choices for real scenarios. As mentioned above, this model is actually a simplified GE model, because the original GE channel model allowed PLR(G) to be nonzero [Mus89]. Once again, we would like to point out that the core results obtained in this thesis are expected to be similar to those with the original GE channel model.

In this thesis, we use the convention of assuming that channel state transitions occur at the beginning time of each packet interval. As shown in Figure 3.2, at each packet interval, the channel changes to a new state with the transition probabilities $1-P_{G|G}$ for going from state "G" to state "B" and $1-P_{B|B}$ for changing from state "B" to state "G". Now we define the conditional transition probability of from state "*a*" (i.e. "G" or "B") to state "*b*" (i.e. "G" or "B") in *m*-steps as $P_{b|a}[m]$. Apparently, as shown in Figure 3.2, its one-step transition probabilities can be compactly specified in the form of a transition probability matrix (denoted by $P_t[1]$):

$$P_t[1] = \begin{bmatrix} P_{G|G}[1] & P_{B|G}[1] \\ P_{G|B}[1] & P_{B|B}[1] \end{bmatrix} = \begin{bmatrix} P_{G|G} & 1-P_{G|G} \\ 1-P_{B|B} & P_{B|B} \end{bmatrix} \tag{3.4}$$

From $P_t[1]$, the *m-step* Transition Probability Matrix (TPM, denoted by $P_t[m]$) then is given by [Tri82]:

$$\begin{aligned} P_t[m] &= \begin{bmatrix} P_{G|G}[m] & P_{B|G}[m] \\ P_{G|B}[m] & P_{B|B}[m] \end{bmatrix} = \left(P_t[1]\right)^m \\ &= \begin{bmatrix} \dfrac{1-P_{B|B}+(1-P_{G|G})(P_{G|G}+P_{B|B}-1)^m}{2-P_{G|G}-P_{B|B}} & \dfrac{(1-P_{G|G})\left(1-(P_{G|G}+P_{B|B}-1)^m\right)}{2-P_{G|G}-P_{B|B}} \\ \dfrac{(1-P_{B|B})\left(1-(P_{G|G}+P_{B|B}-1)^m\right)}{2-P_{G|G}-P_{B|B}} & \dfrac{1-P_{G|G}+(1-P_{B|B})(P_{G|G}+P_{B|B}-1)^m}{2-P_{G|G}-P_{B|B}} \end{bmatrix} \end{aligned} \tag{3.5}$$

The complete proof on this formula can be found in [pp. 313-315, [Tri82]]. Now let P_G and P_B be the steady state probabilities of being in states "G" and "B", respectively. By calculating the limiting state probability of $P_t[m]$ (i.e. $\lim_{m\to\infty} P_t[m]$), these two steady state probabilities can be obtained [Tri82]:

$$P_G = \lim_{m\to\infty} P_{G|G}[m] = \lim_{m\to\infty} \frac{1-P_{B|B}+(1-P_{G|G})(P_{G|G}+P_{B|B}-1)^m}{2-P_{G|G}-P_{B|B}} \overset{|P_{G|G}+P_{B|B}-1|<1}{=} \frac{1-P_{B|B}}{2-P_{G|G}-P_{B|B}} \tag{3.6}$$

$$P_B = \lim_{m\to\infty} P_{B|G}[m] = \lim_{m\to\infty} \frac{(1-P_{G|G})\left(1-(P_{G|G}+P_{B|B}-1)^m\right)}{2-P_{G|G}-P_{B|B}} \overset{|P_{G|G}+P_{B|B}-1|<1}{=} \frac{1-P_{G|G}}{2-P_{G|G}-P_{B|B}} \tag{3.7}$$

Also, the proofs of (3.6) and (3.7) can be found in [pp. 321-322, [Tri82]]. Using (3.6) and (3.7), the average packet loss rate caused by the GE channel can be expressed in the following form:

$$PLR_{GE} = \mathrm{PLR(G)}P_G + \mathrm{PLR(B)}P_B \tag{3.8}$$

Since PLR(G)=0 and PLR(B)=1, by substituting (3.6) and (3.7) into (3.8), the PLR of the simplified GE channel is given by:

$$PLR_{GE} = \frac{1 - P_{G|G}}{2 - P_{G|G} - P_{B|B}} = P_B \tag{3.9}$$

Furthermore, let the random variables X and Y be the error-burst-length and the error-free-length, respectively. The Probability Distribution Functions (PDF) of X and Y are geometric and given by [Tri82]:

$$P_{\mathrm{X}}^{j} = \Pr(\mathrm{X} = j) = (P_{B|B})^{j-1}(1 - P_{B|B}), \qquad j \in \{1,2,...\} \tag{3.10}$$

$$P_{\mathrm{Y}}^{i} = \Pr(\mathrm{Y} = i) = (P_{G|G})^{i-1}(1 - P_{G|G}), \qquad i \in \{1,2,...\} \tag{3.11}$$

Note that (3.10) and (3.11) also can be obtained intuitively from Figure 3.2: It actually is (*j*-1) or (*i*-1) state transitions to its own state plus the *j*-th or *i*-th state transition to the other state. All of the *j* or *i* steps is calculated independently so that we can get (3.10) and (3.11) immediately. Finally, using (3.10) and (3.11), we then can get the expected values *E*(X) and *E*(Y) :

$$E(\mathrm{X}) = \sum_{j=1}^{\infty} j \cdot P_X^j = \sum_{j=1}^{\infty} j \cdot ((P_{B|B})^{j-1}(1 - P_{B|B})) = \frac{1}{1 - P_{B|B}} \tag{3.12}$$

$$E(\mathrm{Y}) = \sum_{i=1}^{\infty} i \cdot P_Y^i = \sum_{i=1}^{\infty} i \cdot ((P_{G|G})^{i-1}(1 - P_{G|G})) = \frac{1}{1 - P_{G|G}} \tag{3.13}$$

3.3. How to Produce Parameters

Because the average error burst length has a noticeable impact on the performance of error correcting codes over the GE channel [Yee95], we need to study the impact of *E*(X) on different error correction schemes. For the convenience of analysis and simulations, here we introduce a more meaningful quantity in the GE channel: The correlation coefficient (CC, defined as ρ in this paper) of two consecutive erasure packets. Now lets random variable Γ denote the loss probability of a packet transmitted in GE channel, which is defined as:

$$\Gamma = \begin{cases} 0 & \text{the packet transmitted in "G" state} \\ 1 & \text{the packet transmitted in "B" state} \end{cases} \tag{3.14}$$

Then, $E(\Gamma)$ and $E(\Gamma^2)$ can be computed as follows, respectively:

$$\begin{aligned} E(\Gamma) &= 0 \cdot P_G + 1 \cdot P_B = P_B = PLR_{GE} \\ E(\Gamma^2) &= 0^2 \cdot P_G + 1^2 \cdot P_B = PLR_{GE} \end{aligned} \tag{3.15}$$

Furthermore, let random variable Γ_1 denote the loss probability of a packet transmitted in GE channel, and let random variable Γ_2 denote the loss probability of the packet transmitted following the packet with Γ_1 in GE channel. Upon Γ_1 and Γ_2, the parameter ρ then can be defined as:

$$\rho = \frac{E((\Gamma_1 - PLR_{GE})(\Gamma_2 - PLR_{GE}))}{\sigma^2} \tag{3.16}$$

Where σ^2 is the variance of Γ , which is given by:

$$\sigma^2 = E(\Gamma - PLR_{GE})^2 \tag{3.17}$$

Using (3.15), then, Eq. (3.17) can be rewritten as:

$$\sigma^2 = E(\Gamma - PLR_{GE})^2 = E(\Gamma^2) - (PLR_{GE})^2 = P_B - (PLR_{GE})^2 \tag{3.18}$$

Moreover, $E(\Gamma_1 \cdot \Gamma_2)$ can be computed intuitively as follows:

$$\begin{aligned} E(\Gamma_1 \cdot \Gamma_2)) &= (P_{\mathrm{G}} \cdot P_{B|\mathrm{G}}) \cdot 0 \cdot 1 + (P_{\mathrm{G}} \cdot P_{\mathrm{G|G}}) \cdot 0 \cdot 0 + (P_{\mathrm{B}} \cdot P_{\mathrm{G|B}}) \cdot 1 \cdot 0 + (P_B \cdot P_{B|B}) \cdot 1 \cdot 1 \\ &= P_B \cdot P_{B|B} \end{aligned} \tag{3.19}$$

Using (3.18) and (3.19), we now can calculate (3.16) as:

$$\rho = \frac{E((\Gamma_1 - PLR_{GE})(\Gamma_2 - PLR_{GE}))}{\sigma^2} = \frac{E(\Gamma_1 \cdot \Gamma_2) - (PLR_{GE})^2}{\sigma^2} = \frac{P_B \cdot P_{B|B} - (PLR_{GE})^2}{P_B - (PLR_{GE})^2} \tag{3.20}$$

Note that PLR_{GE}=P_B here, substituting (3.9) into (3.20), we then have:

$$\begin{aligned}
\rho &= \frac{P_B \cdot P_{B|B} - (PLR_{GE})^2}{P_B - (PLR_{GE})^2} = \frac{P_B \cdot P_{B|B} - (P_B)^2}{P_B - (P_B)^2} = \frac{P_{B|B} - P_B}{1 - P_B} \\
&= \frac{P_{B|B} - \dfrac{1 - P_{G|G}}{2 - P_{G|G} - P_{B|B}}}{1 - \dfrac{1 - P_{G|G}}{2 - P_{G|G} - P_{B|B}}} = \frac{(2 - P_{G|G} - P_{B|B})P_{B|B} - 1 + P_{G|G}}{1 - P_{B|B}} \\
&= \frac{(1 - P_{G|G} - P_{B|B})P_{B|B} - 1 + P_{B|B} + P_{G|G}}{1 - P_{B|B}} = \frac{(1 - P_{G|G} - P_{B|B})(P_{B|B} - 1)}{1 - P_{B|B}} \\
&= P_{G|G} + P_{B|B} - 1
\end{aligned} \tag{3.21}$$

Note both $P_{G|G}$ and $P_{B|B}$ are conditional probabilities which values only being in the range of between 0 and 1. The value of ρ thus can take all possible values between -1 and 1. We now can express the group of parameters ($P_{G|G}$, $P_{B|B}$) in terms of the more meaningful quantities PLR_{GE} and ρ by solving (3.9) and (3.21):

$$\begin{aligned}
P_{B|B} &= PLR_{GE} + \rho(1 - PLR_{GE}) \\
P_{G|G} &= (1 - PLR_{GE}) + \rho PLR_{GE}
\end{aligned} \tag{3.22}$$

Upon (3.22), given a certain PLR_{GE} and ρ for a scenario, we then can obtain the parameters of the GE channel model. To understand the relationship between the PLR_{GE} and ρ, we will discuss them for some typical cases in the following section.

3.3.1 Discussions

Above all, in order to show the meaning of the parameter ρ, we also express the *m-step* TPM in terms of PLR_{GE} and ρ, which can be obtained by substituting (3.22) into (3.5), i.e.:

$$
\begin{aligned}
P_t[m] &= \begin{bmatrix} P_{G|G}[m] & P_{B|G}[m] \\ P_{G|B}[m] & P_{B|B}[m] \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{(1-\rho)(1-PLR_{GE})+PLR_{GE}(1-\rho)\rho^m}{1-\rho} & \dfrac{PLR_{GE}(1-\rho)\left(1-\rho^m\right)}{1-\rho} \\ \dfrac{(1-\rho)(1-PLR_{GE})\left(1-\rho^m\right)}{1-\rho} & \dfrac{PLR_{GE}(1-\rho)+(1-\rho)(1-PLR_{GE})\rho^m}{1-\rho} \end{bmatrix} \\
&= \begin{bmatrix} (1-PLR_{GE})+PLR_{GE}\rho^m & PLR_{GE}(1-\rho^m) \\ (1-PLR_{GE})(1-\rho^m) & PLR_{GE}+(1-PLR_{GE})\rho^m \end{bmatrix}
\end{aligned}
\tag{3.23}
$$

Based on (3.23), we then can discuss the meaning of ρ in GE model. First, we discuss three special cases (i.e. ρ=0 and ρ=±1) in the following:

1. In case of ρ=0 (i.e. $P_{G|G}+P_{B|B}=1$), the $P_t[m]$ can be written as follows:

$$
P_t[m] = \begin{bmatrix} P_{G|G}[m] & P_{B|G}[m] \\ P_{G|B}[m] & P_{B|B}[m] \end{bmatrix} = \begin{bmatrix} 1-PLR_{GE} & PLR_{GE} \\ 1-PLR_{GE} & PLR_{GE} \end{bmatrix} \tag{3.24}
$$

   From (3.24), we can see that all of the four transition probabilities only depend on the original link PLR of PLR_{GE}. It infers that the channel acts as the *i.i.d* channel model for this case.

2. In case of ρ=1, the $P_t[m]$ can be written as follows:

$$
P_t[m] = \begin{bmatrix} P_{G|G}[m] & P_{B|G}[m] \\ P_{G|B}[m] & P_{B|B}[m] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.25}
$$

   As shown in (3.25), the channel state will always stay in "G" state or "B" state with the probability of 1, which only depends on the initial channel state. That is, there are no state transitions for this case.

3. In case of ρ=-1, the calculation of $P_t[m]$ can be analyzed as follows:

First, from (3.21), we can know that this case only occurs with $P_{G|G}$=$P_{B|B}$=0. Then, by (3.7), we can know that the stable probability of the channel state being in "B" state is 50% for this case (i.e. P_B=50%). That is, the link PLR of the channel for this case is 50% (i.e. PLR_{GE}=1/2). Upon the analysis above, the $P_t[m]$ in (3.23) then can be expressed as the following form:

$$P_t[m] = \begin{bmatrix} P_{G|G}[m] & P_{B|G}[m] \\ P_{G|B}[m] & P_{B|B}[m] \end{bmatrix} = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & m \text{ is even} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & m \text{ is odd} \end{cases} \tag{3.26}$$

From (3.26), we can see that the channel state transition occurs at each step for this case. For example, if the initial channel state is "B", the channel state must be "G" after one step transition, and the channel state must be "B" after two- steps transition and so on. That means that "B" state and "G" state always appear alternately at any two continues transmissions, which results in the link PLR of 50%.

Secondly, we also take an example to illustrate the meaning of ρ in GE model for common cases as in [Her08]. Figure 3.3 shows the *m*-steps transition probabilities for a state trellis staring in the "B" state and ending in "B" state (i.e. $P_{B|B}[m]$ in $P_t[m]$ calculated by (3.23)) with PLR_{GE}= 50% and various ρ between -1 and 1.

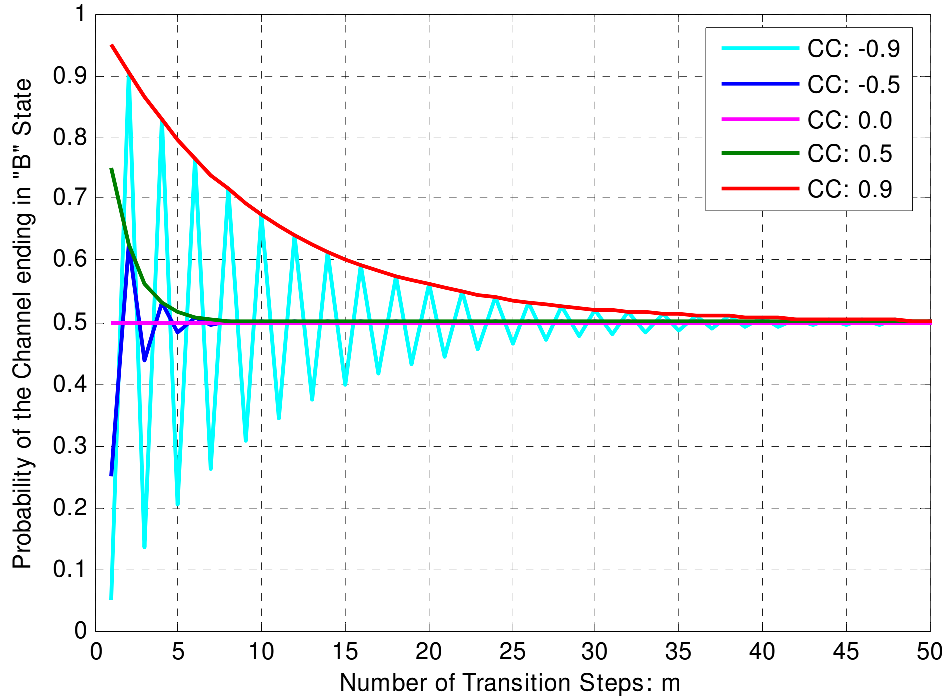


Figure 3.3: An example on the m-steps transition probabilities with PLR_{GE}=50%

As shown in Figure 3.3, when the time is long enough (i.e. the value of m is large enough), the probability of the channel ending in "B" state (i.e. $P_{B|B}[m]$) will always be the steady state probability of the channel state being in "B" state (i.e. P_B=0.5) for all of these cases. This result obviously corresponds to the theory on the steady state probabilities introduced in Chapter 3.2, i.e. the limiting state probability of $P_t[m]$ is always the steady state probability. Furthermore, from this figure, we can see that the probabilities of $P_{B|B}[m]$ always vary concussively with the parameter ρ of negative values. This is because the tendency of the channel state transition for the next step is always against the current channel state when the parameter ρ is of negative values. When the parameter ρ is of positive values, however, the tendency of the channel state transition for the next step always continues the current channel state. As shown in this figure, therefore, the channels with positive CC change states much more smoothly than those with negative CC. It indicates that the channels with positive CC are more stable than those with negative CC. Actually, the values of CC are usually positive in real systems (e.g. see Table 3.2). We thus investigate the performance of all kinds of EER schemes based on positive CC in this thesis. However, we must point out that it is not difficult to extend the framework proposed in this thesis for the scenarios with negative CC.

Finally, we list parts of the parameters calculated by (3.22) in the following table to show the features of the GE channel in more detail.

Table 3.1: Produced Parameters for GE channel

PLR_{GE}		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$\rho=0.00$	$P_{G\|G}$	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.92	0.91	0.90
	$P_{B\|B}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$\rho=0.05$	$P_{G\|G}$	0.9905	0.9810	0.9715	0.9620	0.9525	0.9430	0.9335	0.9240	0.9145	0.905
	$P_{B\|B}$	0.0595	0.0690	0.0785	0.0880	0.0975	0.1070	0.1165	0.1260	0.1355	0.145
$\rho=0.10$	$P_{G\|G}$	0.9910	0.9820	0.9730	0.9640	0.9550	0.9460	0.9370	0.9280	0.9190	0.910
	$P_{B\|B}$	0.1090	0.1180	0.1270	0.1360	0.1450	0.1540	0.1630	0.1720	0.1810	0.190

From this table, we can see that the $P_{B|B}$ increases with a growing ρ under a certain link PLR. Also, Eq. (3.12) indicates that the average burst length increases with a growing $P_{B|B}$. As a result, the impact of the average burst length on the erasure error recovery schemes can be studied by varying the parameter ρ.

3.4. Parameters Evaluation

To evaluate whether the GE channel model matches the features of erasure error in real wireless networks, we implemented a test bed with IEEE 802.11a in our labs [Gor07]. Through the test bed, we can distribute digital TV via access point (AP) to the computers with wireless card. Roughly speaking, the data stream will be encapsulated in continuous RTP packets, and then distributed to the receivers via UDP/IP/802.11a protocol stack. Now we model the length of consecutive erroneous or correct RTP packets transmitted over UDP/IP/802.11a protocol stack as geometric distribution. From (3.10) and (3.11), we can perform the GE model to a set of experimental data. Since we need to estimate the parameters $P_{G|G}$ and $P_{B|B}$, we determine the maximum likelihood estimation (MLE) of $P_{G|G}$ and $P_{B|B}$, i.e. $\hat{P}_{G|G}$ and $\hat{P}_{B|B}$. Without loss of generality, we take the random variable bust error length X as an example. According to the theory of MLE[1], the likelihood function on parameter $\hat{P}_{B|B}$ can be expressed as:

$$L = \prod_{i=1}^{N} P^{n_i}_{\mathrm{X};\hat{P}_{B|B}} \tag{3.27}$$

Where n_i (i=1,2,…,N) is a set of observations of error burst length and N is the total number of observations. Substituting (3.10) into (3.27), we get:

$$\begin{aligned} L = \prod_{i=1}^{N} P^{n_i}_{\mathrm{X};\hat{P}_{B|B}} &= (1-\hat{P}_{B|B})^{N} \prod_{i=1}^{N} (\hat{P}_{B|B})^{n_i - 1} \\ &= (1-\hat{P}_{B|B})^{N} (\hat{P}_{B|B})^{\sum_{i=1}^{N} n_i - N} \end{aligned} \tag{3.28}$$

Now taking the natural logarithm of both sides, we have:

$$\Lambda = \ln(L) = N \ln(1-\hat{P}_{B|B}) + (\sum_{i=1}^{N} n_i - N) \ln(\hat{P}_{B|B}) \tag{3.29}$$

Then we can calculate the partial derivation of Λ with respect to $\hat{P}_{B|B}$, we obtain,

$$\frac{\partial \Lambda}{\partial \hat{P}_{B|B}} = -\frac{N}{1-\hat{P}_{B|B}} + \frac{\sum_{i=1}^{N} n_i - N}{\hat{P}_{B|B}} \tag{3.30}$$

By setting (3.30) equal zero, we have,

$$\frac{\partial \Lambda}{\partial \hat{P}_{B|B}} = -\frac{N}{1-\hat{P}_{B|B}} + \frac{\sum_{i=1}^{N} n_i - N}{\hat{P}_{B|B}} = 0 \tag{3.31}$$

At last, after rearranging (3.31), we obtain,

[1] http://www.weibull.com/AccelTestWeb/mle_maximum_likelihood_parameter_estimation.htm

$$\frac{1}{N}\sum_{i=1}^{N} n_i = \frac{1}{1-\hat{P}_{B|B}} \tag{3.32}$$

From (3.32), we know that the MLE of $P_{B|B}$ is determined using the expectation expression (3.12). Similarly, the MLE of $P_{G|G}$ is also determined using the expectation expression (3.13). The GE model is described by two geometric PDFs, one for lost packet burst and another for the number of consecutive arrived packets, which are denoted by P_{X}^{j} and P_{Y}^{j}, respectively. Then we can find an estimation of $P_{G|G}$ or $P_{B|B}$ from observed runs and bursts.

3.5. Evaluation Results

In this section, using the method of MLE introduced above, we estimated the distribution parameters $P_{G|G}$ and $P_{B|B}$ of GE model according to the observations, and then compare them with experimental results. First of all, we describe the basic scenarios used in our experiments based on our laboratory environments. We considered sending real time multimedia signals by RTP/UDP/IP protocol stack over IEEE 802.11a. Note that any erroneous packet is discarded at UDP/IP level. The RTP packet loss was monitored by checking the sequence number of the received RTP packets. In our measurement, the RTP payload size is set to 1316 bytes with 7 MPEG packets (a packet length of 188 bytes). Then, the IP packet size will be 40+1316=1356 bytes, where the additional 40 bytes include 20 bytes for IP header, 8 bytes for UDP header and 12 bytes for RTP header [Ets0f]. Note that the number of bytes transmitted for one IP packet in the physical layer actually is more than 1356 bytes due to the additional MAC header, CRC and channel coding used etc. It indicates that the bits data rate transmitted in the physical layer will be higher than the bits data rate in the RTP level or IP layer. We choose from Standard Definition Television (SDTV) (about 4Mbps) to High Definition Television (HDTV)[2] (about 12Mbps) as the application services over IEEE 802.11a. The source multimedia data are encapsulated into RTP stream with constant data rate. Then, each RTP packet is transmitted in the unit of packet interval of t_s.

[2] http://www.timefordvd.com/tutorial/DigitalTVTutorial.shtml

To evaluate the parameters of GE model for this case, we always use the convention that state transitions occur at the beginning of each RTP packet transmitted, which is shown in Figure 3.4.

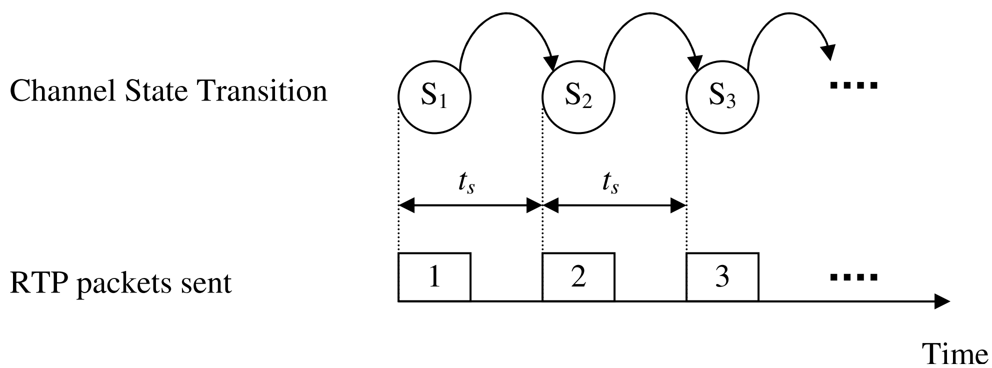


Figure 3.4: Channel state transition for the RTP packets sent in the evaluations of the parameters of GE channel model

From Figure 3.4, we can see that the channel sate is S_1 when the first RTP packet sent. Then, the channel state will change to S_2 when the second RTP packet sent and so on. For each RTP packet sent, the channel state (e.g. S_1) is either "G" or "B". Then, the probability of the channel state being "G" or "B" for the next RTP packet sent (e.g. S_2) will depend on the parameters $P_{G|G}$ and $P_{B|B}$ of GE channel model. For example, the calculation of the Pr(S_2="G") can be expressed as follows:

$$\Pr(S_2=\text{“G”}) = \Pr(S_1=\text{“G”}) \cdot P_{G|G} + \Pr(S_1=\text{“B”}) \cdot (1-P_{B|B}) \qquad (3.33)$$

By applying the GE channel model for our experiments in this way, we then evaluate its parameters using the MLE method introduced in Chapter 3.3.

We build a typical scenario based on the infrastructure mode for the measurements in our laboratory, which is shown in Figure 3.5. In this scenario, the two stations access real-time multimedia data from DVB servers via the Access Point (AP).

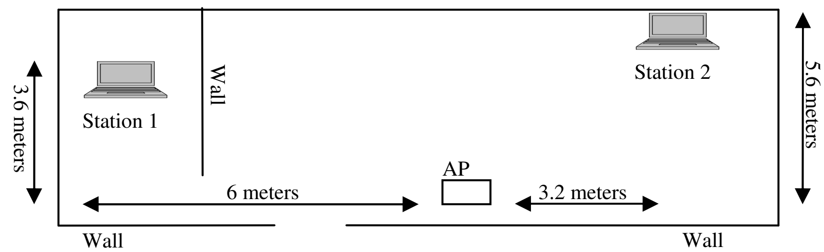


Figure 3.5: Experimental Scenario

As shown in Figure 3.5, in this experimental scenario, there is one AP that can distribute data to two stations with wireless card. In our experiments, we distribute one constant data rate stream of 7Mbps, 9Mbps and 13Mbps (RTP payload data rate) to two stations, respectively. The bandwidth in physical layer is set to 18Mbps. The total number of RTP packets for each experiment is about 20 million. By these experimental data we can estimate the parameters for the GE model. Table 3.2 shows these estimated parameters and the average error-burst-length (which is denoted by E(X) here and can be obtained by (3.12) upon the estimated $P_{B|B}$) for the two stations with three different experiments.

Table 3.2: Estimated Parameters for the GE model

Data Rate	Station 1				Station 2			
	$P_{G\|G}$	$P_{B\|B}$	ρ	E(X)	$P_{G\|G}$	$P_{B\|B}$	ρ	E(X)
7Mbps	0.9852	0.0286	0.0138	1.029	0.9957	0.0441	0.0398	1.046
9Mbps	0.9857	0.0315	0.0172	1.033	0.9945	0.0460	0.0405	1.048
13Mbps	0.9840	0.0341	0.0181	1.035	0.9935	0.0474	0.0409	1.050

First, from this table, we can see that the two stations have different parameters of GE model in these three scenarios. Therefore, the two stations can be viewed as independent receivers. In addition, as shown in Table 3.2, the CC of GE models (i.e. the parameter ρ) for these two stations increase with the increase of the multicast data rate. For example, the parameter ρ increases about 31.2% for **Station 1** and 0.11% for **Station 2** when the data rate increased from 7Mbps to 13Mbps. This phenomenon can be explained by the relationship between the packet interval t_s and the packet duration t_d. Note these two parameters are associated with the multicast data rate R_d and the channel bandwidth B_w. Let l_p be the length of the data packet in unite of bytes, then the value of t_d can be calculated according to the current channel bandwidth B_w and the packet length l_p, i.e.:

$$t_d = \frac{8l_p}{B_w} \tag{3.34}$$

And the value of t_s depends on the multicast data rate R_d and the packet length l_p, i.e.:

$$t_s = \frac{8l_p}{R_d} \tag{3.35}$$

From (3.34) and (3.35), we know that the relationship between the packet interval and the packet duration can be expressed as: $t_s = (B_w / R_d)t_d$ .To demonstrate the effect of the multicast data rate to the parameter ρ, Figure 3.6 shows the channel state transitions for different multicast data rate in a GE channel with a certain channel bandwidth of B_w and a constant packet length of l_p.

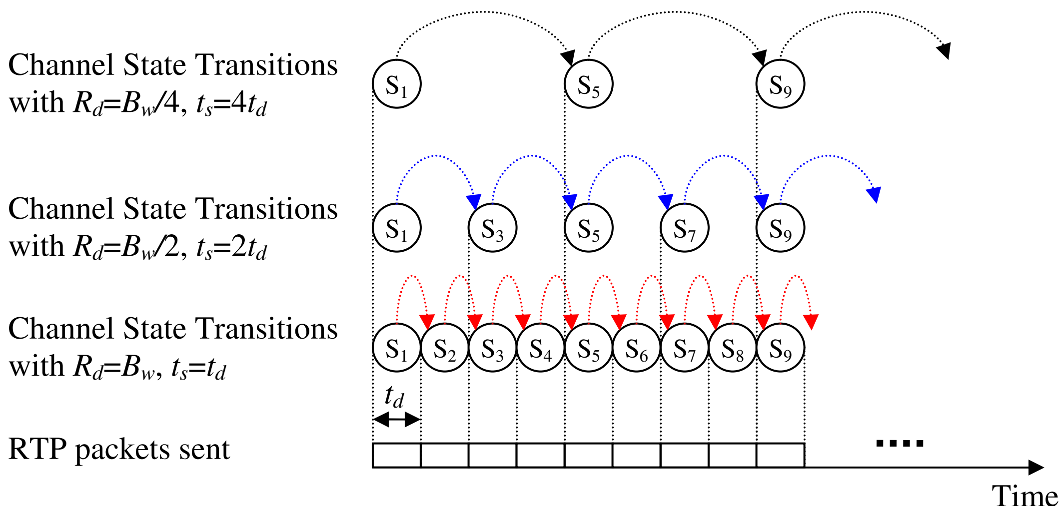
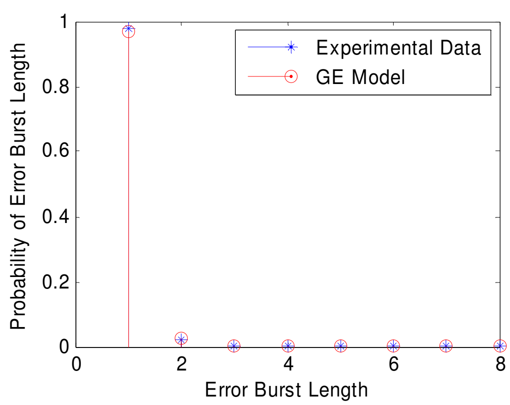


Figure 3.6: Channel state transitions with different source data rate and fixed bandwidth

As shown in Figure 3.6, when the channel is of full load (i.e. R_d=B_w), the channel state transitions occur at the beginning of each unit time of t_d. With the decreasing of the multicast data rate, the channel state transitions can be obtained by sub-sampling those with R_d=B_w. For example, in case of R_d=B_w/4, the channel state transitions can be obtained by sub-sampling those with R_d=B_w in the unit length of 4t_d. That is, one channel state transition with R_d=B_w/4 is equivalent to four channel state transitions with R_d=B_w. Let ρ' be the CC of the GE channel in case of R_d=B_w. Therefore, according to the analysis above and the *m-step* TPM denoted by (3.23), we can know that the CC of the GE channel in case of R_d=B_w/4 or R_d=B_w/2 is actually $(\rho')^4$ or $(\rho')^2$. It indicates that the CC of the GE channel with high date rate is more than that with low date rate in case of $0<\rho'<1$. In case of $\rho'=0$ or $\rho'=1$, however, all of them will be identical so that it has nothing to do with the variable multicast data rate. Since the CC of the GE channel satisfies $0<\rho'<1$ in this practical system, it will increase with the increase of the multicast data rate.

However, from Table 3.2, we can know that the average error-burst-length increases only about 0.6% for **Station 1** and 0.4% for **Station 2** when the data rate increases from 7Mbps to 13Mbps. It indicates that the increased multicast data rate does not increase the average error-burst-length significantly. Therefore, the GE channel model can be viewed as relative stable for each receiver in a multicast scenario with variable multicast data rate.

Finally, in the following, we also compare the distribution of error-free-length and error-burst-length of GE model with the experimental data for these two stations. Without loss of generality, we will take the **Station 1** as the example. The results are shown in Figure 3.7, Figure 3.8 and Figure 3.9. As shown in these three figures, the distribution of the error burst length and error free length of the experimental determined counterpart corresponds to that of GE model very well for all of the three cases. That is, the simplified GE model is very representative for the packet loss model in the multimedia multicast over IEEE 802.11a scenario. Therefore, the simplified GE model can be used to assist in the design of communication systems, evaluate the performance of all kinds of AL-HEC schemes in WLANs.



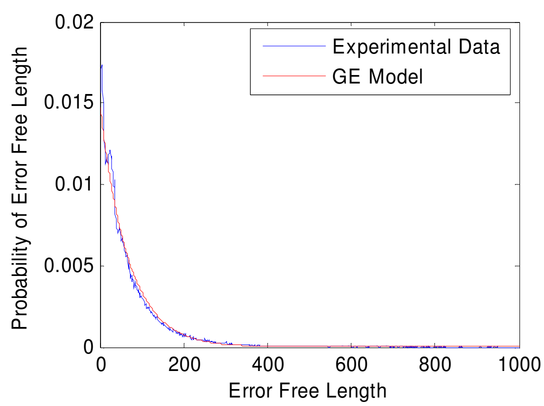


Figure 3.7: Probability of Error Burst Length and Error Free Length at Station 1 with about 7Mbps

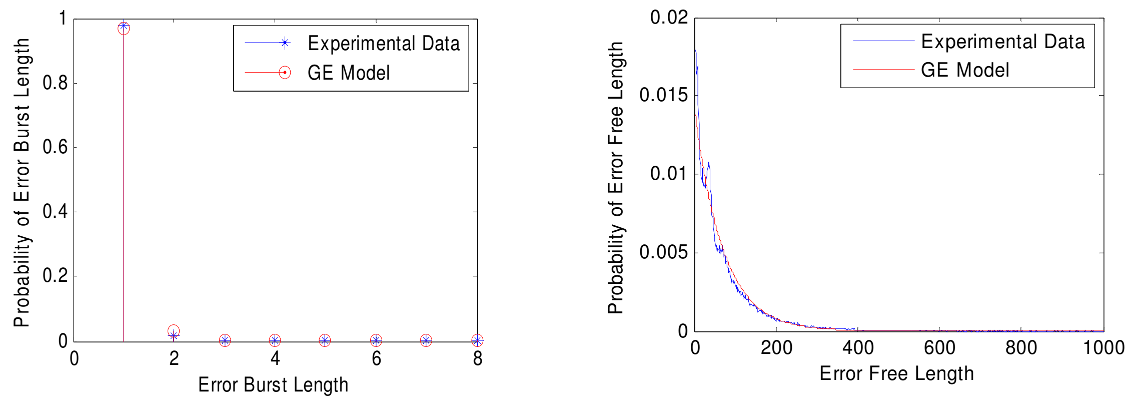


Figure 3.8: Probability of Error Burst Length and Error Free Length at Station 1 with about 9Mbps

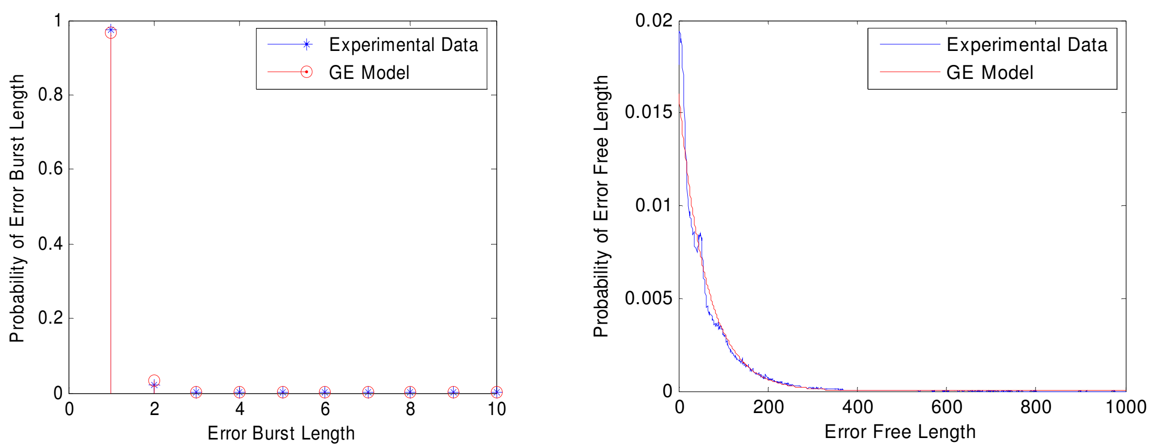


Figure 3.9: Probability of Error Burst Length and Error Free Length at Station 1 with about 13Mbps

Chapter 4
End-to-End Delay Budget

Since we consider the performance of those EER schemes under strict delay constraints in this thesis, we need to do the budgets of the end-to-end delay for different schemes. In the following, we first present how to calculate the end-to-end delay budget for FEC schemes, ARQ schemes, respectively. Then, we introduce how to model the end-to-end delay for the proposed general architecture by combining the delay budgets for ARQ and FEC schemes.

4.1. FEC Delay

Since FEC alone based EER schemes will transmit those packets in the form of blocks without any retransmission, the delay caused by the FEC mainly comes from the length of the block (i.e. the number of packets in one encoding block). Without loss of generality, here we assume that the (n, k) code is used for the FEC scheme, where n is the length of the code word and k is number of source data packets, and the number of parity packets is n-k. Since it is always assumed that PL-FEC is used for EER schemes in this thesis, one packet thus can be viewed as one symbol in PL-FEC schemes. Note, when the source data packets are transmitted in the packet interval of t_s in case of without any PL-FEC scheme, the time duration of transmitting one group of k source data packets will occupy total $(k\text{-}1)\cdot t_s+t_d$, where t_d is the packet duration. According to the information theory, in order to guarantee the reliability of transports, the system must satisfy $B_w \geq R_d$. Otherwise, it is impossible to guarantee the reliability of transports due to the packet losses caused by the congestion in the channel. Therefore, in this thesis, it is always assumed that the channel bandwidth is enough to satisfy the target PLR requirement. This means that we always have $B_w \geq R_d$. From (3.34) and (3.35), we can find that the packet duration t_d is always no more than the packet interval t_s in case of $B_w \geq R_d$. Especially, in case of $B_w >> R_d$, we will have $t_d << t_s$.

Now applying the (n, k) code to one group of k source data packets, we assume that the group of n encoded packets will also can be transmitted in the time duration of $(k\text{-}1)\cdot t_s + t_d$ in the packet interval of $((k\text{-}1)/(n\text{-}1))\cdot t_s$. This assumption is based on the two reasons: First, to save the time of recovering missing data packets at receivers for real-time distributions, those parity packets should be transmitted as soon as possible. Secondly, it can simplify the analysis and the design works for the optimum scheme. This is because under this assumption the number of parity packets has nothing to do with the delay budget; we then can only focus on the contribution of the parameter k for the delay budget. This assumption could be realistic when the bandwidth B_w is large enough and the encoding time is short enough, since the $n\text{-}k$ parity packets could be produced immediately as soon as all of the k source data packets obtained from the application. Note that if the source data rate is R_d for this case, the data rate of the encoded stream will be $((n\text{-}1)/(k\text{-}1))\cdot R_d$. In practical systems, although the packet interval for transmitting the encoded packets might be more or less than $((k\text{-}1)/(n\text{-}1))\cdot t_s$, the analysis results are still suitable for those cases in which the error between the practical packet interval and the value of $((k\text{-}1)/(n\text{-}1))\cdot t_s$ is small. In case that this error can not be neglected, however, it needs further study to find the perfect solution based on the framework proposed in this thesis.

To demonstrate the end-to-end delay in the FEC scheme with (n, k) code in more detail, Figure 4.1 shows the time consumed by transmitting a full encoding block from the sender to one receiver.

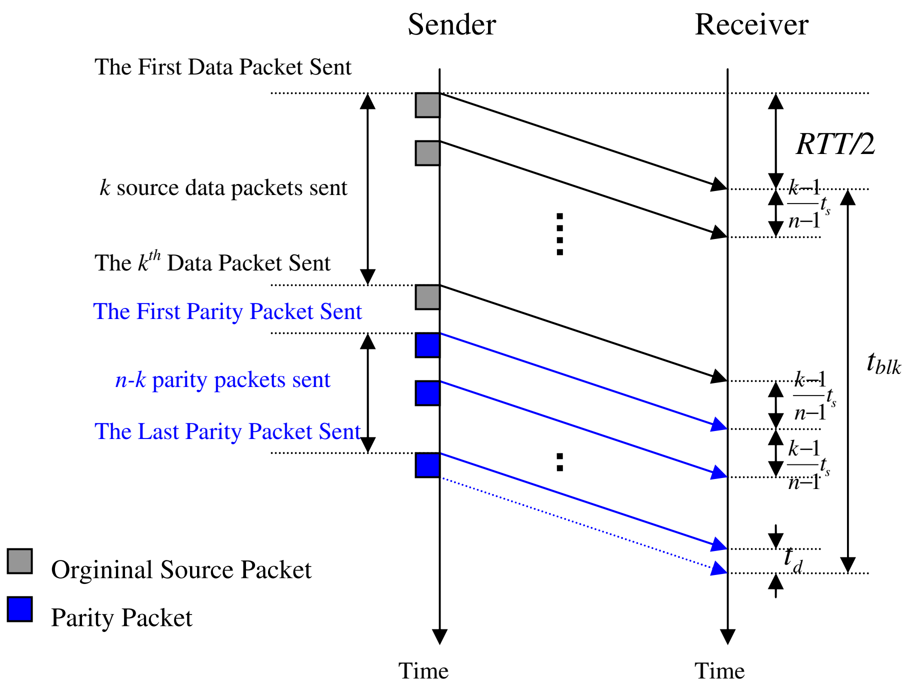


Figure 4.1: End-to-end Delay Budget for FEC Schemes with (n, k) code

As shown in Figure 4.1, the one-way delay of from the sender to the receiver is half of the Round Trip Time (RTT) (i.e. $RTT/2$) in the network. In the whole thesis, for the convenience of analysis, we assume that the one-way delay is always half of the RTT. Since systematic erasure codes are always assumed to be adopted in this thesis, the k source data packets belong to one FEC encoding block are first transmitted to the receiver with the packet interval of $((k-1)/(n-1))\cdot t_s$. Following the k source data packets, the remaining n-k parity packets for this encoding block are also transmitted with the packet interval of $((k-1)/(n-1))\cdot t_s$. From Figure 4.1, we can find that the maximum time of the whole block sent occupies exactly:

$$t_{blk}=(n-1)\cdot ((k-1)/(n-1))\cdot t_s+t_d=(k-1)\cdot t_s+t_d \tag{4.1}$$

Apparently, if no any error for those data packets happens, the end-to-end delay in the FEC scheme will be fixed to *RTT/2*. However, if some data packets lost at the receiver, they can not be recovered until enough number of parity packets received.

Now we analyze the maximum possible delay by recovering those missing data packets. Under the assumption above, considering the recovery delay for those missing data packets at the receiver, the worst case will be the following case: The first data packet is lost, and it can not be recovered until the last parity packet received at the end of the last packet duration. From Figure 4.1, it is easy to calculate the end-to-end delay caused by recovering the first data packet in the worst case, which is exactly $RTT/2 + t_{blk}$. Because we consider the EER schemes under strict delay constraint in this thesis, we must use the maximum possible end-to-end delay budget (i.e. $RTT/2 + (k-1) \cdot t_s + t_d$) for designing the optimum FEC scheme or HARQ schemes combined the FEC scheme. As a result, when the FEC scheme is used for real-time distributions, the maximum possible end-to-end delay budget (i.e. $RTT/2 + t_{blk}$) should be guaranteed to be no more than the target delay requirement.

4.2. ARQ Delay

In contrast to the FEC based schemes, the end-to-end delay in the ARQ based schemes mainly comes from the retransmissions. Therefore, we should consider the delay caused by those retransmission packets. In the following, Figure 4.2 shows the diagram of the end-to-end delay budget for the first retransmission round in an ARQ based scheme.

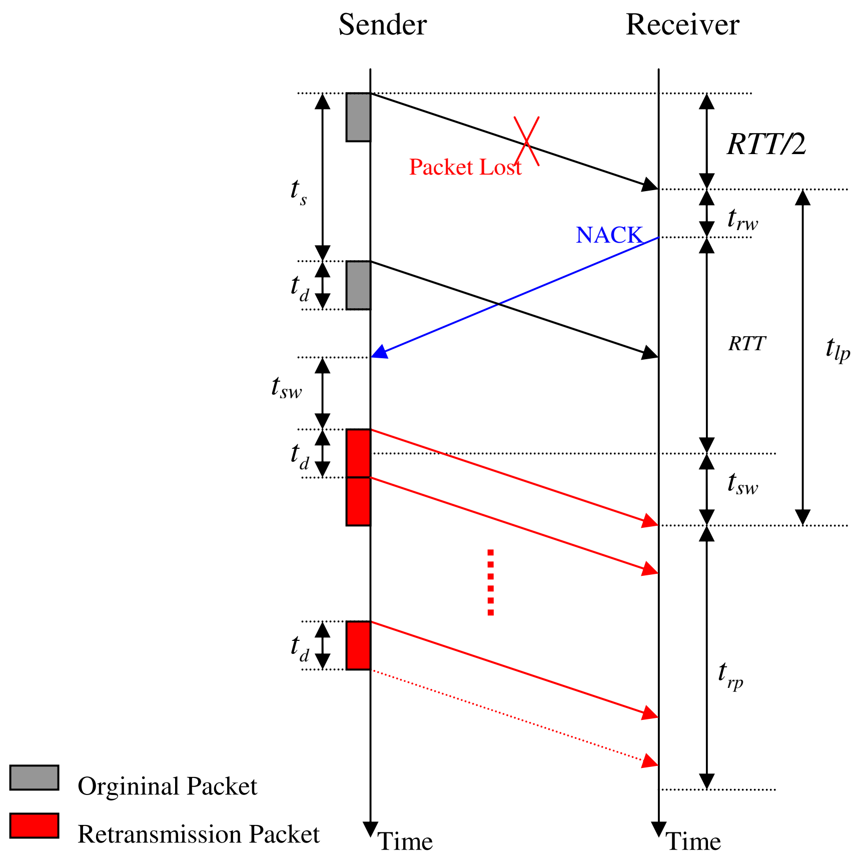


Figure 4.2: End-to-end Delay Budget for ARQ Schemes

Similar to Figure 4.1, in Figure 4.2, the one-way delay of from the sender to the receiver is also *RTT*/2. Additionally, let t_{rw} be the waiting time at the receiver, which is the time between the time the latest packet loss occurs and the time when the corresponding NACK is sent; let t_{sw} be the waiting time at the transmitter, which is the time between receiving a NACK message and the time when the corresponding packets required by the NACK message start to be retransmitted. As shown in Figure 4.2, upon the NACK message received, the sender will retransmit those source data packets required as soon as possible to save the recovery time at the receiver. Obviously, the time duration occupied by those retransmission packets depends on the number of packets retransmitted and the packet duration t_d.

For the convenience of analysis, then, let t_{rp} denote the maximum possible time duration occupied by all of the retransmission packets in each retransmission round. As a matter of fact, the value of t_{rp} can be set to a fixed value with a certain margin in the end-to-end delay budget. That is, the value of t_{rp} is always set large enough to guarantee that all of the retransmission packets can be transmitted during the time of t_{rp}. Now let's take an example to explain the evaluation of the parameter t_{rp}. First, note that the value of t_{rp} mainly depends on the average error-burst-length (denotes by E(X) here) and the packet duration t_d. As mentioned in Chapter 4.1, in case of B_w>>R_d, the packet duration t_d will be far less than the packet interval t_s. When B_w is so large enough that t_s is much larger than $E(X)\cdot t_d$, then we can simply set t_{rp} to t_s in the delay budget for this case. Following this idea, the value of t_{rp} can be set to a fixed value of multiple t_s according to the relationship between B_w and R_d. Especially, according to (3.34), we can know that $t_d \to 0$ in case of $B_w \to \infty$. Therefore, we can always set t_{rp} to t_s in case of $B_w = \infty$, which is actually the ideal condition in the end-to-end delay budget for the retransmission based schemes. For the convenience of comparing the performance of all kinds of EER schemes with each other fairly, in this thesis, we will analyze their performances under some typical multicast scenarios in the ideal condition.

Finally, from Figure 4.2, we can find that the maximum possible end-to-end delay for those retransmission packets in the first retransmission round is $RTT/2 + (t_{rw} + RTT + t_{sw}) + t_{rp}$. In the following, we begin to analyze the maximum end-to-end delay caused by retransmissions in ARQ based schemes. We define the symbol "t_{lp}" as the time duration from the time the latest packet loss occurs at the receiver to the earliest time it possibly receives the required packets. From Figure 4.2, we can see that t_{lp} can be easily computed by t_{rw}+RTT+t_{sw}. For an ARQ based scheme, let $N_{rr,max}$ be the maximum allowable number of retransmission rounds. Afterwards, based on the analysis above, it is easy to know that the maximum possible end-to-end delay for the first retransmission round is $\frac{RTT}{2} + t_{lp} + t_{rp}$. Accordingly, for the ARQ scheme with total $N_{rr,max}$ retransmission rounds, the delay budget can be calculated by:

$$D_{ARQ} = \frac{RTT}{2} + N_{rr,\max} \cdot (t_{lp} + t_{rp}) \tag{4.2}$$

Upon (4.2), we then can evaluate the maximum possible end-to-end delay in ARQ based schemes. Finally, we can use it for designing the optimum ARQ scheme under strict delay constraints by limiting D_{ARQ} no more than the target delay requirement.

4.3. Model of the End-to-End Delay for the General Architecture

For the general architecture proposed in Chapter 2.2, the end-to-end delay mainly includes two parts: One part comes from the length of the FEC encoding block in the first transmission; the other part comes from those retransmissions for data packets or parity packets. Based on the parameters of the architecture defined in Table 2.1 and the analysis in 4.1, it is easy to know that the time consumed by the block of (k+N_p) packets in the first transmission will be t_{blk} of (k-1)·t_s+t_d. Combining t_{blk} and the delay caused by $N_{rr,max}$ retransmissions rounds, which actually can be viewed as ARQ delay and can be computed by (4.2), we then get the total end-to-end delay for the general architecture immediately:

$$\begin{aligned} D &= \frac{RTT}{2} + t_{blk} + N_{rr,\max} \cdot (t_{lp} + t_{rp}) \\ &= \frac{RTT}{2} + (k-1) \cdot t_s + t_d + N_{rr,\max} \cdot (t_{lp} + t_{rp}) \end{aligned} \tag{4.3}$$

When applying the general architecture for the EER under strict delay constraints, in the following Chapters, we will use (4.3) as the end-to-end delay budget for the general architecture.

Chapter 5
Application of GE Model

Since we use GE model as the erasure error channel model to evaluate the performance of all kinds of EER schemes, in this Chapter, we will introduce how to apply the GE model in different schemes in this thesis. According to the introduction of GE model in Chapter 3, we know that it is actually a kind of memory channel with two state (i.e. “B” state and “G” state) Markov chain. That is, the erasures of the packets are time correlated so that we have to consider the sequence of those packets transmitted over GE channel. In the following, first, we will present how to apply GE model in FEC schemes and ARQ schemes, respectively. Because the general architecture proposed in Chapter 2.2 is actually composed of these two basic schemes, then, we introduce how to apply GE model in the general architecture based on these two basic cases.

5.1. Applied in FEC Schemes

As the FEC scheme described in Chapter 4.1, in the FEC scheme the sender will transmit those packets in the form of blocks. For the convenience of analysis as in [Yee95], a stream of all of those encoding blocks sent can be viewed as only one Virtual Block (VB) sent over the GE channel model, where the initial channel state before transmitting this VB is in stable state. That is, the probability of the initial channel state being in “B” state (or “G” state) is the stable probability of P_B (or P_G), which can be calculated by (3.6) and (3.7) using the parameters of the GE channel. By modeling the stream of those blocks as one VB, we then can analyze the channel state transition for each packet sent in the VB. In this thesis, as applied the GE model in evaluations in Chapter 3.4, we always use the convention that state transitions occur at the beginning of a packet is transmitted. Figure 5.1 shows the diagram of channel state transitions for the VB transmitted over GE channel model.

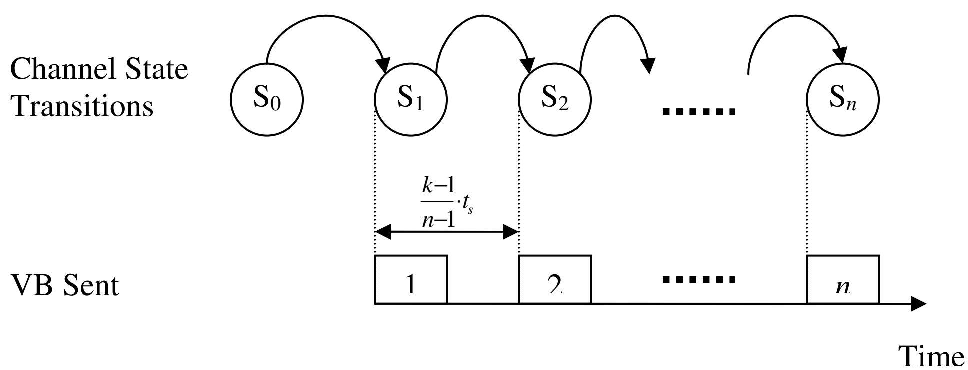


Figure 5.1: Channel state transitions for the VB transmitted with (n, k) code

As shown in this figure, the initial state of the GE channel is S_0, and the state transition occurs at the beginning of each packet is transmitted. Based on the assumption above, we can obtain the probability of S_0 being in different states, i.e. Pr(S_0="B")=P_B and Pr(S_0="G")=P_G. Based on these two initial probabilities, we then can derive the probability of the state being in "B" state or "G" state for each packet sent in the VB.

In the following, we take the calculations of Pr(S_1="B") and Pr(S_1="G") as examples for explanations. Now we assume that the parameters of the GE model is $P_{G|G}$ and $P_{B|B}$. Additionally, let $P_{B|G}$ be the one transition probability of from "G" state to "B" state, and let $P_{G|B}$ be the one transition probability of from "B" state to "G" state. Upon the features of GE model as shown in Figure 3.2, we obviously have $P_{B|G}$=1- $P_{G|G}$ and $P_{G|B}$=1- $P_{B|B}$. Note the state of S_1 being in "B" state only occurs under the following two conditions: One is that the S_0 is in "B" state and then transits to "B" state when the first packet sent; the other is that the

S_0 is in "G" state and then transits to "B" state when the first packet sent. Combining these two cases, we then can obtain Pr(S_1="B") immediately:

$$\begin{aligned}\Pr(S_1=\text{"B"}) &= \Pr(S_0=\text{"B"})\, P_{B|B}+\Pr(S_0=\text{"G"})\, P_{B|G} \\ &=P_B P_{B|B}+ P_G(1-P_{G|G})\end{aligned} \tag{5.1}$$

Similarly, we also can compute the Pr(S_1="G"):

$$\begin{aligned}\Pr(S_1=\text{"G"}) &= \Pr(S_0=\text{"B"})\, P_{G|B}+\Pr(S_0=\text{"G"})\, P_{G|G} \\ &=P_B(1-P_{B|B})+ P_G P_{G|G}\end{aligned} \tag{5.2}$$

Following this thought, we then can obtain Pr(S_i="B") and Pr(S_i="G") (where $1 \le i \le n$) for each packet sent in the VB. It results in the convenience for analyzing the performance of the FEC scheme with (n, k) code, which will be described in detail in Chapter 6.

Finally, we would like to point out a noticeable phenomenon when applying the GE model in FEC schemes: because the data rate of the encoded stream is usually higher than the source data stream, the parameters of the GE model for the encoded stream may be different from those for the original source data stream. According to the evaluation results in Chapter 2.4, higher the data rate is, bigger the CC of the GE model is. However, due to the small range of the CC of the GE model (e.g. far less than 0.1) in real systems, the small changes of the parameters on the GE model usually have no effect on the accuracy of the analysis results for FEC schemes.

5.2. Applied in ARQ Schemes

For the ARQ based schemes over GE channel model, we know that the PDF of error-burst-length can be computed by (3.10). When a burst error happens, according to the ARQ scheme, the receiver will send a NACK message for requiring the retransmissions of all of the missing data packets included in the burst error. When evaluating the performance of the retransmission based schemes over GE channel, we need to know the loss probability of those retransmission packets in each retransmission stage. To demonstrate how we evaluate the loss probability for those retransmission packets, Figure 5.2 shows the diagram of channel state transition for two retransmission packets in the first retransmission round with the error-burst-length of two.

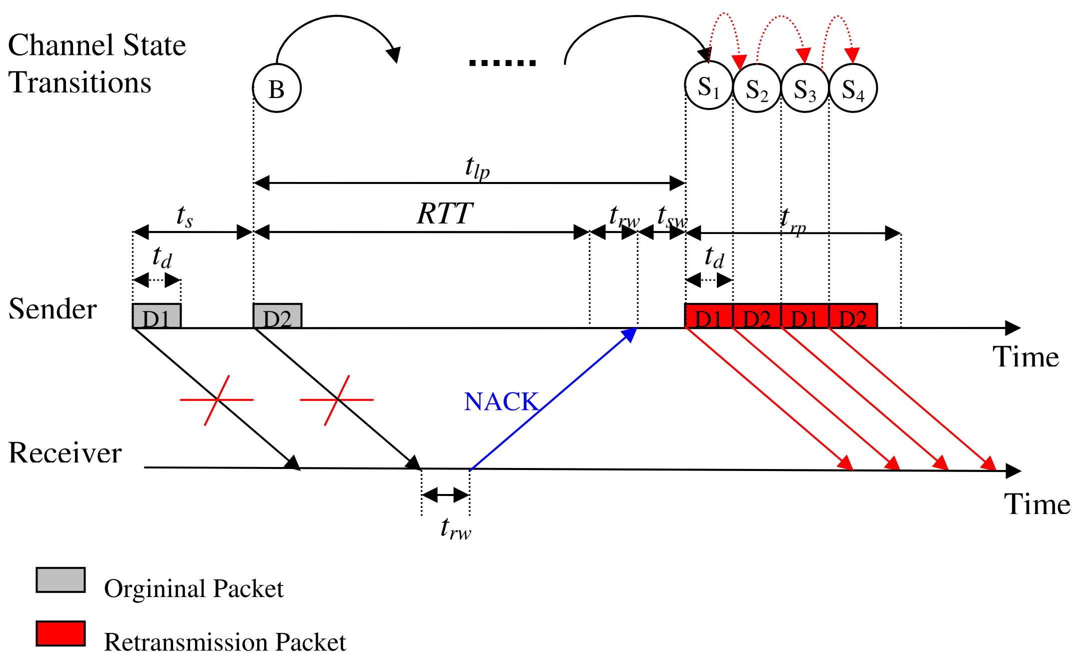


Figure 5.2: Channel state transitions for retransmission packets in the first retransmission round in ARQ Schemes

As shown in Figure 5.2, there are two continuous data packets (i.e. "D1" and "D2" shown in this figure) lost resulting in an error-burst-length of two occurs. According to the mechanism of ARQ, then, the receiver will send a NACK message to the sender for requiring the recovery of these two missing data packets (i.e. "D1" and "D2"). From Figure 5.2, we can see that the sender will send two copies of the missing data packets required continuously as soon as possible. Note, this mechanism is completely corresponding to the ARQ scheme described in Chapter 4.2.

Now focus on the channel state transition in the first retransmission round, it is clear that the channel is in "B" state when the packet "D2" is lost. From Figure 5.2, after the time duration of t_{lp}, we can see that the first retransmission of the missing data packets occurs. During the time of from the packet "D2" lost to the start time of retransmitting the first packet (i.e. t_{lp} as shown in this figure), we use the convention that state transitions occur at the beginning of a time slot of unit length of t_s. Additionally, let S_1 be the channel state at the beginning of retransmitting the first packet, which can be found in Figure 5.2. During the time of t_{lp}, therefore, the maximum possible number of channel state transitions will be:

$$T_{lp}=\left\lceil \frac{t_{lp}}{t_s} \right\rceil \tag{5.3}$$

Note that in this case the initial state is "B" as shown in Figure 5.2, using (3.5), the probability of S_1 being "B" or "G" can be obtained immediately by calculating the T_{lp}-step transition probability, i.e.: $\Pr(S_1=\text{"B"})=P_{B|B}[T_{lp}]$ and $\Pr(S_1=\text{"G"})=P_{G|B}[T_{lp}]$.

Additionally, according to the evaluation results for practical systems in Chapter 3.5, we know that the parameters of GE channel model are variable with the increase of the multicast data rate in the same channel. From Figure 5.2, we can see that the data rate of the retransmission stream in the time duration of t_{rp} is much higher than the data rate of the original transmission stream. As analyzed in Chapter 3.5, therefore, the CC of the GE model for the retransmission stream will be larger than that for the original data stream. It indicates that the parameters of GE model for the retransmission stream is different from those for the original transmission stream. Furthermore, each retransmission packet in the retransmission stream will have different loss probability in case that the CC of the GE model is not of zero or one. Due to these two characters on retransmission streams, it is very complicated and a big challenge for achieving the accurate mathematical framework on the performance of the retransmission based EER schemes. Instead of deriving the accurate formulas for the performance of the retransmission based EER schemes, therefore, we focus on the upper-band of the performance for retransmission based EER schemes in this thesis. In other words, we will study how well the EER schemes can work in the worst situation. The worst situation can be realized by the following assumption: In each retransmission round, all of the retransmission packets have the same loss probability to that of the retransmission packet with the largest loss probability. As a matter of fact, we have the following assertion:

Assertion: *Among all of the retransmission packets at each retransmission round, the first retransmission packet has the largest loss probability.*

The proof of this assertion is attached in the appendix as shown in Chapter 5.4. Upon this assertion, in order to derive the upper-band of the performance for retransmission based EER schemes in this thesis, we will assume that all of the retransmission packets at each retransmission round have the same loss probability to the first retransmission packet in this retransmission round. Actually, as analyzed in Chapter 5.4, the error between the assumption and the accurate loss probabilities for those retransmission packets is very small so that it can be neglectable in many practical systems. It means that the upper-band of the performance contributed in this thesis is very close to the accurate performance in many practical systems. Furthermore, we will adopt the upper-band of the performance in the assumed worst situation to design the optimum parameters for those retransmission based EER schemes. Therefore, the optimum parameters of those EER schemes can always satisfy the target QoS requirements very well.

Now we analyze the loss probability of the first retransmission packet at each retransmission round. First, we focus on the loss probability of the first retransmission packet in the first retransmission round as shown in Figure 5.2. For the convenience of analysis, let random variable Γ' be the loss probability of the first retransmission packet (i.e. the first "D1" in Figure 5.2) in the first retransmission round, which is defined as:

$$\Gamma' = \begin{cases} 0 & S_1 = "G" \\ 1 & S_1 = "B" \end{cases} \tag{5.4}$$

Based on the analysis above, then, the expected value of Γ' can be obtained by:

$$E(\Gamma') = 0 \cdot \Pr(S_1 = "G") + 1 \cdot \Pr(S_1 = "B") = P_{B|B}[T_{lp}] \tag{5.5}$$

According to the assertion introduced above, all of the retransmission packets in the first retransmission round are assumed to have the same loss probability to the first retransmission packet, which can be computed by (5.5) (i.e. $P_{B|B}[T_{lp}]$). Note, since each source data packet is retransmitted twice in this example (i.e. two copies of D1 and D2 as show in Figure 5.2), the loss probability of each source data packet in the first retransmission round is thus $\left(P_{B|B}[T_{lp}]\right)^2$. As a result, the loss probability of each source data packet in each retransmission round is associated with the number of copies sent in the retransmission round (i.e. N_{cc}^q defined in Table 2.1 for the general architecture with k=1). That is, the loss probability of each source data packet in the first retransmission round is $\left(P_{B|B}[T_{lp}]\right)^{N_{cc}^1}$. Similarly, the loss probability of each source data packet in the second retransmission round is $\left(P_{B|B}[2T_{lp}]\right)^{N_{cc}^2}$ due to the longer delay budget (i.e. $2T_{lp}$) as described in Chapter 4.2. Following this idea, finally, we can extend the result to a general case in the ARQ scheme: The loss probability of each source data packet during the q-th retransmission round will be $\left(P_{B|B}[q \cdot T_{lp}]\right)^{N_{cc}^q}$. In this thesis, we will use this general formula to calculate the loss probabilities of those source data packets at each retransmission round.

5.3. Applied in the General Architecture

The general architecture proposed in Chapter 2.3 is actually composed of the FEC scheme and the ARQ scheme introduced above. Recalling the parameters of the general architecture, in case of k>1, it behaves as the FEC scheme in the first transmission so that the results of Chapter 5.1 is adaptable for the first transmission stage. In this case, if some retransmission rounds needed, the retransmission packets will be parity packets, which can be modeled as retransmission packets as for the GE model applied in the ARQ scheme in Chapter 5.2. The only difference between the parity packets retransmission and the data packets retransmission is: To keep the efficiency of retransmissions, each independent parity packet will be retransmitted only once, while each missing data packet is allowed to be retransmitted multiple times in the ARQ based schemes. As a result, the loss probability of those retransmission parity packets in the q-th retransmission round will always be $P_{B|B}[q \cdot T_{lp}]$. In case of k=1, especially, the general architecture behaves as the pure ARQ schemes so that all of the results in Chapter 5.2 is suitable for this case.

Strictly speaking, due to the memory feature of GE channel and the variable parameters of the channel model, each packet at all of the transmission stages has different loss probability. As analyzed in Chapter 5.2, therefore, it is still a big challenge to achieve the accurate performance for the general architecture. However, through making a reasonable assumption for the worst situation as described in Chapter 5.2, we can derive a tight upper-band of the performance for the retransmission based schemes integrated in the general architecture. In the following chapters, we will analyze and optimize the performance for three kinds of schemes: pure FEC scheme, pure ARQ based HEC-PR scheme and the general architecture of EER proposed in Chapter 2.3.

5.4. Appendix: Proof of the Assertion

In this appendix, we give a complete proof for the assertion introduced in Chapter 5.2. Without loss of generality, we focus on the first retransmission round as similar as shown in Figure 5.2. Note that channel state transitions in the GE channel also occur at the beginning of a time slot of unit length of t_d. Here we would like to point out that the parameters of the GE model may be different from those in which channel state transitions occur at the beginning of a time slot of unit length of t_s for original data stream. In this proof, we model the GE channel as channel state transitions occurring at the beginning of each unit time of t_d. For the convenience of description, in this proof, let ρ' be the CC of this GE channel model (where $0 \le \rho' \le 1$), let PLR'_{GE} be the average PLR of the GE channel (where $0 \le PLR'_{GE} \le 1$), and let random variable Γ'_i (where i is a positive integer) be the loss probability of the i-th retransmission packet in the first retransmission round, which is defined as:

$$\Gamma'_i = \begin{cases} 0 & \mathrm{S}_i = \text{"G"} \\ 1 & \mathrm{S}_i = \text{"B"} \end{cases} \tag{5.6}$$

Then, during the time of t_{lp}, the maximum possible number of channel state transitions is:

$$T'_{lp} = \left\lceil \frac{t_{lp}}{t_d} \right\rceil \tag{5.7}$$

As shown in Figure 5.2, note that the initial state is “B” in the retransmission stage. Using (3.5) and (5.7), the probability of S_i being “B” or “G” can be obtained immediately by calculating the transition probability, i.e.:

$$\begin{aligned} &\Pr(\mathrm{S}_i=\text{“B”})=P_{B|B}[T'_{lp}+i-1] \\ &\Pr(\mathrm{S}_i=\text{“G”})=P_{G|B}[T'_{lp}+i-1] \end{aligned} \tag{5.8}$$

Upon (5.8), then, the expected value of Γ'_i can be obtained immediately by:

$$E(\Gamma'_i) = 0 \cdot \Pr(\mathrm{S}_i = \text{"G"}) + 1 \cdot \Pr(\mathrm{S}_i = \text{"B"}) = P_{B|B}[T'_{lp} + i - 1] \tag{5.9}$$

Based on (3.23) and (5.9), therefore, we have:

$$E(\Gamma'_i) = P_{B|B}[T'_{lp} + i - 1] = PLR'_{GE} + (1 - PLR'_{GE})(\rho')^{T'_{lp}+i-1} \tag{5.10}$$

Depending on (5.10), the value of $E(\Gamma'_i) - E(\Gamma'_1)$ with i>1 can be expressed as:

$$\begin{aligned} E(\Gamma_i') - E(\Gamma_1') &= PLR_{GE} + (1 - PLR_{GE}')(\rho')^{T_{lp}'+i-1} - (PLR_{GE}' + (1 - PLR_{GE}')(\rho')^{T_{lp}'}) \\ &= (1 - PLR_{GE}')(\rho')^{T_{lp}'+i-1} - (1 - PLR_{GE}')(\rho')^{T_{lp}'} \\ &= (1 - PLR_{GE}')(\rho')^{T_{lp}'}((\rho')^{i-1} - 1) \\ &\leq 0 \end{aligned} \tag{5.11}$$

According to (5.11), the values of $E(\Gamma_i')$ thus always satisfy $E(\Gamma_i') \leq E(\Gamma_1')$ (where *i*>1) due to $0 \leq \rho' \leq 1$ and $0 \leq PLR_{GE}' \leq 1$. It means that the loss probability of the first retransmission packet in the first retransmission stream is always the largest. This result obviously holds true for each retransmission stream at each retransmission stage. The proof of the assertion is completed. In fact, as shown in (5.11), when the value of ρ' is far less than one (e.g. $\rho' \leq 0.1$) and the value of T_{lp}' is far more than one, the error between $E(\Gamma_i')$ and $E(\Gamma_1')$ will be very small so that it is neglectable. For many practical systems, this result is true due to the small CC of the GE channel and the small packet duration t_d.

Finally, we would like to point out two special cases in which the values of $E(\Gamma_i')$ in each retransmission stream are identical. First, in case of ρ'=0, upon (5.10) we have:

$$E(\Gamma_i') = E(\Gamma_j') = PLR_{GE}' \qquad \forall i \neq j \tag{5.12}$$

In this case, the channel behaves as an *i.i.d* channel so that all of the packets have the same loss probability to the link PLR. Secondly, in case of ρ'=1, upon (5.10) we have:

$$E(\Gamma_i') = E(\Gamma_j') = 1 \qquad \forall i \neq j \tag{5.13}$$

Eq. (5.13) indicates that the loss probability of the entire retransmission packets is always 100% in this case. As discussed in 3.3.1, it is because that the initial channel state is "B" and there are no channel state transitions in the case of ρ'=1.

Chapter 6

Adaptive FEC Scheme

In this chapter, we will introduce an Adaptive Forward Error Correction (AFEC) scheme using packet level FEC (PL-FEC) techniques alone. As introduced in Chapter 2.2.2, in this thesis, it is assumed that perfect systematic erasure codes (e.g. RS codes) are used for PL-FEC schemes. As mentioned before, more recent codes like LDPC [Gal60] [Gal62] or Fountain / Raptor-codes [Sho06] do scale well for long blocks and offer advantages concerning computational efficiency. In the thesis, however, a perfect erasure code is taken as the “upper anchor”. Additionally, the block sizes for many practical application scenarios (e.g. DVB services over WLAN) can well be solved by RS codes with acceptable computational complexity. For the convenience of description, the perfect FEC code is denoted by (n, k) code in this chapter, where k denotes the number of data symbols per codeword and n denotes the codeword length. In the thesis, the basic idea of the AFEC scheme is that it will guarantee the target PLR requirement with the minimum total needed redundancy information (RI) under strict delay constraints. Since the total needed RI of the AFEC scheme only depends on the code rate (i.e. k/n) used, this scheme will optimize its code rate according to the current channel state. In the following, first, we will introduce the principle of PL-FEC techniques. Then, we will present how to calculate its performance. Finally, we will present how to design the optimum parameters for the AFEC scheme.

6.1. Performance of Packet Level FEC

The principle of PL-FEC can be explained by a structure of one coding block of n packets. Figure 6.1 shows the structure of the coding block transmitted within packets protected by the ideal systematic (n, k) code.

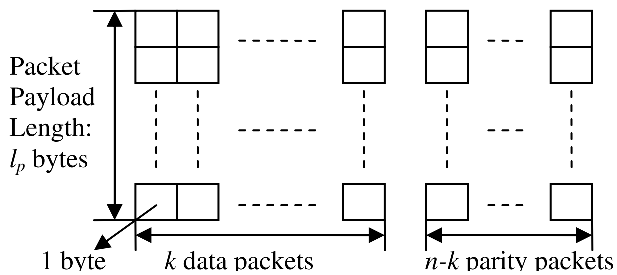


Figure 6.1: Applying ideal systematic (n, k) code at the packet level, which forms an FEC coding block in n packets

As shown in Figure 6.1, the source data packet stream is divided into blocks each consisting of k consecutive data packets with a length of l_p bytes. The (n, k) code is applied to each row containing k data packets in order to produce a group of (n-k) parity packets. The coding block is transmitted by packets in the form of columns. Without loss of generality, here we assume exactly one packet per column. According to the theory of perfect erasure codes introduced in Chapter 2.2.2, the receiver only needs to correctly receive any k of these n columns to be able to recover all the k source data packets. Therefore, the PLR performance of PL-FEC schemes is exactly the same as the performance of the ideal systematic (n, k) code.

We now focus on analyzing the performance of PL-FEC schemes with ideal (n, k) code. First, we need to calculate the probability of b errors in a sequence of d symbols in erasure error channel. When applied the GE model in FEC schemes, as introduced in Chapter 5.1, we use the convention that state transitions occur at the beginning of a time slot of unit length and then a packet is transmitted. Now let $P(b,d,$ CSI) denote the probability of b packets lost in a sequence of d packets in the GE channel with the CSI of ($P_{G|G}$, $P_{B|B}$). Let $P_G(b,d)$ be the probability of b errors in d transmissions with the channel ending in state "G". Similarly, let $P_B(b,d)$ be the probability of b errors in d transmissions with the channel ending in state "B". Follow the thought presented by (5.1) and (5.2) in Chapter 5.1, then, $P(b,d)$ can be calculated using a recursion approach given by [Yee95]:

$$P(b,d,\mathrm{CSI}) = P_G(b,d) + P_B(b,d) \tag{6.1}$$

For d=1,2,3… and b=0,1,2,…,d,

$$\begin{aligned} P_G(b,d) &= P_G(b,d-1)P_{G|G}(1-P_G) + P_B(b,d-1)(1-P_{B|B})(1-P_G) \\ &\quad + P_G(b-1,d-1)P_{G|G}P_G + P_B(b-1,d-1)(1-P_{B|B})P_G \end{aligned} \tag{6.2}$$

$$\begin{aligned} P_B(b,d) &= P_B(b,d-1)P_{B|B}(1-P_B) + P_G(b,d-1)(1-P_{G|G})(1-P_B) \\ &\quad + P_B(b-1,d-1)P_{B|B}P_B + P_G(b-1,d-1)(1-P_{G|G})P_B \end{aligned} \tag{6.3}$$

The initial conditions are:

$$\begin{aligned} P_G(b,0) = P_B(b,0) = 0, b \neq 0 \\ P_G(0,0) = P_G \\ P_B(0,0) = P_B \end{aligned} \tag{6.4}$$

Note that with these initial conditions, all numerical values calculated by this algorithm will be steady state results. Additionally, we would like to point out that, for the simplified GE model with P_G=0 and P_B=1, the right hand sides of (6.2) and (6.3) can be reduced to two terms, i.e.:

$$P_G(b,d) = P_G(b,d-1)P_{G|G} + P_B(b,d-1)(1-P_{B|B}) \tag{6.5}$$

$$P_B(b,d) = P_B(b-1,d-1)P_{B|B}P_B + P_G(b-1,d-1)(1-P_{G|G})P_B \tag{6.6}$$

Now we begin to derive the PLR performance of the ideal (n, k) code in the GE channel of parameters ($P_{G|G}, P_{B|B}$). To simplify the description below, we define an additional random variable as follows:

- I_k: a random variable representing the number of lost data packets in a group of k data packets after decoding using the FEC decoder with ideal (n,k) code.

Upon the definition of I_k, the PLR performance of the (n,k) code actually can be computed by $E(\mathrm{I}_k)/k$. That means we only need to calculate the expected value of I_k. To obtain $E(\mathrm{I}_k)$, we firstly have to find out the PDF of I_k. For the convenience of description, here we assume that the value of I_k is i and there are b packets lost in a group of n packets. If b is not more than n-k, the number of packets received in a group of n packets will be at least k so that all of the k data packets can be recovered. Obviously, the value of I_k is zero in this case. On the other hand, if the value of I_k is more than zero, there are exactly i (where $1 \le i \le k$) data packets lost in the group of n packets after decoding with the (n,k) code. It indicates that at least $\max(n\text{-}k+1,i)$ and at most $(n\text{-}k+i)$ packets are lost in this group. That is, the value of b is in the range of $[\max(n\text{-}k+1,i), n\text{-}k+i]$. Let $P_d(i,b)$ denote the probability of i data packets lost under the condition of b packets lost in a group of n packets. In other words: Among all of the b packets lost in the group, there are i data packets lost among all of the k data packets and b-i parity packets lost among all of the n-k parity packets. To simplify the calculation of the conditional probability $P_d(i,b)$, it is assumed that all packets in the group of n packets have the same loss probability. Under this assumption, $P_d(i,b)$ can be computed by:

$$P_d(i,b) = \frac{\binom{k}{i}\binom{n-k}{b-i}}{\binom{n}{b}} \tag{6.7}$$

Note that the calculation of $P_d(i,b)$ by (6.7) can be viewed as accurate when the CC of the GE model is small enough (e.g. $\rho<0.1$). However, for larger CC, the more accurate formula of calculating this probability needs further investigation, and then should be compared with this simplified calculation. Based on the analysis above, using (6.1) and (6.7), we then obtain the PDF of I_k as follows:

$$\mathrm{Prob}(\mathrm{I}_k = i) = \sum_{b=\max(n-k+1,i)}^{n-k+i} P(b,n,\mathrm{CSI})P_d(i,b), i = 1,2,...,k. \tag{6.8}$$

Although this formula is not accurate enough for those scenarios with large CC due to the inaccuracy of computing $P_d(i,b)$, the effect of the inaccuracy to the analysis results in this thesis will be very small due to the small values of the conditional probability obtained by calculating $P(b,n,\mathrm{CSI})\cdot P_d(i,b)$. Following (6.8), the expected value of I_k thus can be calculated by:

$$E(\mathrm{I}_k)=\sum_{i=1}^{k}\sum_{b=\max(n-k+1,i)}^{n-k+i} i\times P(b,n,\mathrm{CSI})P_d(i,b) \tag{6.9}$$

Finally, the PLR performance of the (*n, k*) code is given by:

$$PLR_{FEC}(n,k,\mathrm{CSI})=\frac{E(\mathrm{I}_k)}{k}=\frac{1}{k}\sum_{i=1}^{k}\sum_{b=\max(n-k+1,i)}^{n-k+i} i\times P(b,n,\mathrm{CSI})P_d(i,b) \tag{6.10}$$

From the analysis above it follows that (6.10) is also the PLR performance of the PL-FEC scheme with an ideal (*n, k*) code over the GE channel with parameters ($P_{G|G}, P_{B|B}$). Here we take an example to show the PLR performance of the PL-FEC scheme: First, we assume that the length of k is 30 and the link PLR is 0.05 in the GE channel with ρ=0.0 or 0.08; then the parameters ($P_{G|G}, P_{B|B}$) can be obtained by solving (3.22); finally, the PLR performance of the PL-FEC scheme with ideal (n,30) codes can be calculated by (6.10) and part of results are listed in Table 6.1.

Table 6.1: PLR Performance of the PL-FEC scheme with ideal (n,30) codes given an original link PLR of 0.05

n		32	34	36	38	40
PLR Performance	ρ=0.0	2.32×10^{-2}	4.04×10^{-3}	3.63×10^{-4}	2.04×10^{-5}	8.12×10^{-7}
	ρ=0.08	2.50×10^{-2}	5.96×10^{-3}	8.95×10^{-4}	9.91×10^{-5}	8.80×10^{-6}

From this table, we can see that the PLR performance of the PL-FEC scheme will decrease with the increase of the ρ. This is because the average burst length increases with the growing ρ resulting in more packets lost for one encoding block. It indicates that the parameter ρ of the GE channel model has a significant impact on the performance of PL-FEC schemes, and its effect should be considered when designing FEC schemes for reliable communications over erasure error channels with GE model.

6.2. AFEC Scheme

Based on the PLR performance of PL-FEC schemes presented in the Chapter 6.1, we now can optimize the parameters for AFEC schemes. Note that in AFEC schemes all of the receivers can share identical parameters (i.e. an identical ideal (*n*, *k*) code) without any feedback. That is because, if an AFEC scheme can guarantee the QoS requirements for the receiver with the worst link state in a multicast scenario, it can also guarantee the same QoS requirements for any of other receivers in the multicast scenario. Without loss of generality, it is assumed that the first receiver is with the worst situation in a multicasting scenario. In other words, the first receiver has the worst CSI (denoted by CSI(1)) with the largest link PLR and the largest RTT (denoted by *RTT*(1)). Our remaining task is to design optimum parameters of the AFEC scheme for satisfying the target PLR (i.e. $\mathrm{PLR}_{\mathrm{target}}$) requirement under strict delay constraints (i.e. $\mathrm{D}_{\mathrm{target}}$) with the minimum total needed RI.

First, it is known that the strict delay requirement will limit the number of data packets in one encoding block. As analyzed in Chapter 4.1, the end-to-end delay budget for FEC schemes mainly depends on the encoding block size. According to the analysis result on the delay budget presented in Chapter 4.1, for those data packets received by the first receiver, the maximum possible end-to-end delay is mainly composed of the two parts: the one-way delay in the first transmission (which is $RTT(1)/2$); the decoding delay caused by the length of the encoding block (which is $(k\text{-}1)\cdot t_s+t_d$). Therefore, to satisfy the target delay requirement for those data packets transmitted for the first receiver, the maximum possible end-to-end delay must satisfy:

$$\frac{RTT(1)}{2}+(k-1)t_s+t_d \le \mathrm{D}_{\mathrm{target}} \tag{6.11}$$

Note here we do not consider the effect of the delay caused by the encoding and decoding algorithm with a certain erasure code. Usually, this kind of delay can be neglected if an efficient algorithm adopted in powerful PC (e.g. [Riz97]). In case that it can not be neglected in real systems, the problem also can be solved by simply adding a margin for evaluating the end-to-end delay. From (6.11) we now can derive the maximum allowable value for the parameter k, i.e.:

$$k_{\lim} = \left\lfloor \frac{D_{\text{target}} - \frac{RTT(1)}{2} - t_d}{t_s} \right\rfloor + 1 \tag{6.12}$$

Where $\lfloor x \rfloor$ denotes the largest integer not greater than x. Note that in (6.12) the system parameters t_s, t_d and D_{target} usually can be viewed as fixed in practical systems for the given QoS requirements. Therefore, the parameter k only depends on the parameter $RTT(1)$. In other words, the AFEC scheme only needs to adapt the length of k to the worst link delay in the current multicast scenario. Finally, for practical applications, we want to point out that the length of the k should be set as long as possible so that more efficient code rates for ideal erasure codes can be adopted.

Now regarding the total needed RI when applying the AFEC scheme with an ideal (n, k_{lim}) code, it can be calculated as follows:

$$RI_{AFEC} = \frac{n - k_{\lim}}{k_{\lim}} \tag{6.13}$$

Note that here we do not consider the needed overhead messages for transmitting the FEC parameters to the receivers. Usually, the payload size in one coding packet is much bigger than the overhead size so that its effect can be neglected.

From the analysis above, we know that the length of the parameter k for the ideal FEC codes in one block is actually limited by the strict delay constraints (i.e. k_{lim}). Based on k_{lim}, our remaining task is to find the most efficient FEC code to satisfy the target PLR requirement. The final optimization problem can thus be expressed as:

$$RI_{AFEC,opt} = \arg\min(RI_{AFEC}) \tag{6.14}$$

Subject to:

$$PLR_{FEC}(n, k_{\lim}, \mathrm{CSI}(1)) \leq \mathrm{PLR}_{\mathrm{target}}$$

Relying on a certain value of k_{lim}, apparently, we then can find the minimum n (denoted by n_{opt}) to satisfy the PLR requirement for the first receiver with CSI(1) by solving (6.14). As a result, the minimum total needed RI of the AFEC scheme will be:

$$RI_{AFEC,opt} = \frac{n_{opt} - k_{\lim}}{k_{\lim}} \tag{6.15}$$

Not that the number of the receivers has nothing to do with the total needed RI of the AFEC scheme.

6.3. Performance Analysis

From the formulas of calculating the performance of the AFEC scheme, we can find that its performance mainly depends on the parameters (n, k) of the ideal FEC code used. Especially, the parameter k will be limited by such system parameters as the strict delay constraint, the multicast data rate and the packet size and so on used in the communication. As long as the length of the parameter k is fixed, we then can obtain the minimum n to guarantee the target PLR requirement for current scenario by solving (6.14). Actually, the length of k can be decided previously based on the explicit scenario in real systems. Therefore, we only need to find the minimum n for the FEC code used in the AFEC scheme. Accordingly, the performance of the AFEC scheme will mainly depend on the length of k used. So, in the following, we will firstly analyze the effect of the parameter k to the AFEC scheme. On the other hand, from the features of the GE model introduced in Chapter 2, we know that the Correlation Coefficient (CC) of GE model has a significant effect on the error-burst length. Apparently, the error-burst length will directly influence the number of lost packets in one encoding block with FEC codes so that the performance of the AFEC scheme is associated with the CC of GE model. Consequently, we will finally present how the CC of GE model influences the performance of the AFEC scheme.

6.3.1. The effect of the parameter *k*

As mentioned above, since the length of k has a great effect on the performances of the AFEC scheme, in this section, we study the effect of k to this scheme. Note that the different lengths of k refer to different delay constraints or different constant source data rates if the packet size is constant. Also, the performance of the AFEC scheme has nothing to do with the number of receivers in a multicast scenario. We only need to design the parameters for the receiver with the worst link PLR in a multicast scenario. Then, the AFEC scheme can satisfy the PLR requirement for all of the receivers in the multicast scenario. Without loss of generality, it is assumed that the target PLR requirement is 10^{-6} and the parameter k_{lim} varies from 20 to 200 under certain delay constraints. Moreover, we do not consider the effect of the CC of GE model in this section, i.e. the parameter ρ is set to zero here. By solving (6.14), we now search for the optimum parameters for the AFEC scheme with any given original link PLR and length of k, part of the results are listed in the following table.

Table 6.2: Perfect (n, k_{lim}) codes for the AFEC scheme to meet the PLR (10^{-6}) requirement under certain delay constraints

Original Link PLR		0.01	0.03	0.05	0.07	0.09	0.10
n_{opt}	k_{lim}=40	45	49	52	54	57	58
	k_{lim}=120	128	135	141	146	151	154

As shown in Table 6.2, the value of n_{opt} increases with the increase of the original link PLR under a certain value of k_{lim}. It is not difficult to understand: given a certain PLR requirement, larger the original link PLR is, more RI the AFEC scheme needs. To see how the length of the k_{lim} influences the RI performance of the AFEC scheme, we also obtain those perfect (n, k_{lim}) codes for the AFEC scheme under a certain original link PLR. Then, we can compute the total needed RI of this scheme by (6.15), where part of the results is shown in Figure 6.2.

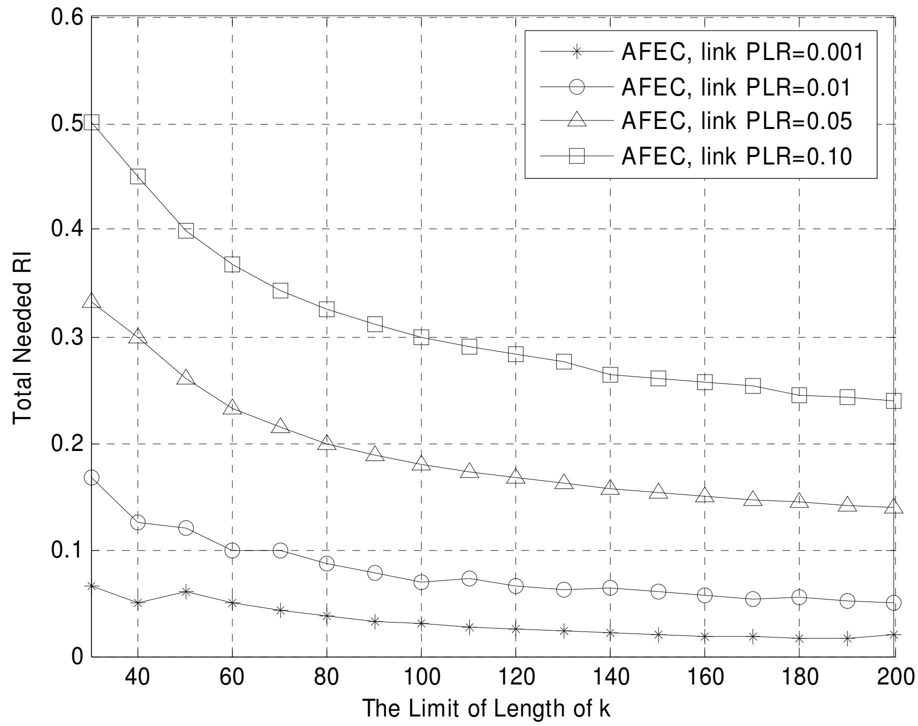


Figure 6.2. The total needed RI of the AFEC scheme with ρ=0

From Figure 6.2, it is found that the parameter k indeed has a significant effect on the performance of the AFEC scheme. The total needed RI of the AFEC scheme will decrease with the increase of the length of the parameter k. The result can be explained by the feature of the AFEC scheme: Note that the performance of the AFEC scheme depends on only the code rate of the (n, k) code, which has to adopt short length for the parameter k due to the strict limit of delay. To reach the target PLR requirement, therefore, only small code rates can be used for the (n, k) code in the AFEC scheme which results in a large amount of RI. That is, the total needed RI for the AFEC scheme with smaller k is much more than that with larger k. As a result, when applying the AFEC scheme for practical systems, larger the length of the parameter k is allowed, better the performance can be achieved.

In addition, there are some strange inflection points in Figure 6.2, which are caused by the large scale of the code rate of the (n, k) code. Now we take an example of the AFEC scheme with a link PLR of 0.001 for explanation: the minimum n is 42 for the code with k=40 to reach the target PLR requirement (i.e. the needed RI is 5%); however, the minimum n is 53 for the code with k=50 to reach the same target PLR requirement (i.e. the needed RI is 6%), because the RI is only 4% if the n is decreased to 52 with k=50 for the code, which is not enough to reach the same target PLR requirement. As a result, the needed RI of the AFEC scheme with k=50 is more than that with k=40. It infers that we should choose the length of k carefully to avoid this problem when applying it in practical systems, which can be implemented by adjusting the packet size or multicast data rate.

6.3.2. The effect of CC of GE Channels

Up to now, it is always assumed that ρ =0 during the analysis above. That is, the effect of ρ is not taken into account for the AFEC so far. For studying the effect of ρ, we now calculate the total needed RI with different ρ using the AFEC scheme, where the parameters of the GE channel are produced by the method introduced in Chapter 3.5. Without loss of generality, it is assumed that the target PLR requirement is 10^{-6} and the parameter k is fixed to 40 under certain delay constraints. Moreover, the CC of GE model (i.e. ρ) varies from 0.0 to 0.5. Under these general assumptions, By solving (6.14), we now search for the optimum parameters for the AFEC scheme with certain original link PLR and different ρ, part of the results are listed in Table 6.3.

Table 6.3: The optimum parameter n for the AFEC scheme with (n, 40) codes to meet the target PLR (10^{-6}) requirement

CC of GE Model		0.0	0.1	0.2	0.3	0.4	0.5
n_{opt}	Original Link PLR=0.01	45	47	49	52	55	59
	Original Link PLR=0.05	52	54	57	60	64	69
	Original Link PLR=0.10	58	61	64	68	72	78

From Table 6.3, we can find that the length of n_{opt} increases not only with the increase of the original link PLR but also with the increase of the CC of GE channel. It indicates that the total needed RI of the AFEC scheme will also increase not only with the increase of the original link PLR but also with the increase of the CC of GE channel. To illustrate of the effect of the CC of GE channel in more detail, we also obtain those perfect (n, 40) codes for the AFEC scheme with a certain original link PLR. Similar to Chapter 3.3.1, then, we also can compute the total needed RI of this scheme by (6.15). Figure 6.3 shows the results on the total needed RI of the AFEC scheme under different CC of GE channel.

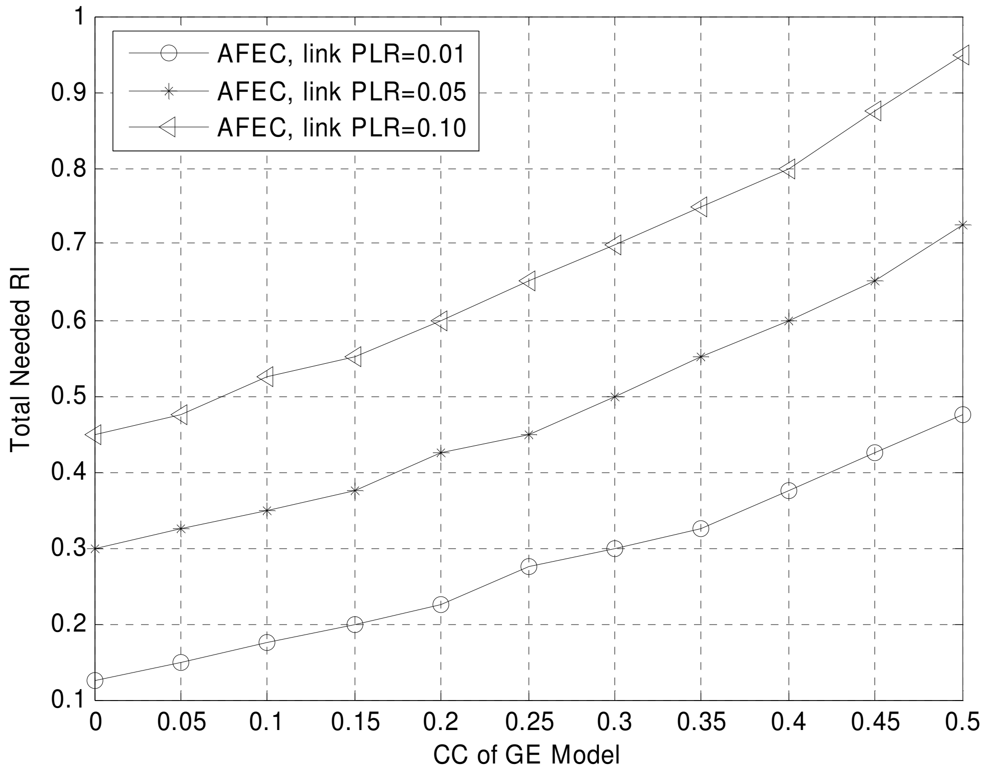


Figure 6.3. The total needed RI of the AFEC scheme under different CC with GE model

As shown in Figure 6.3, we can see that the CC of GE model (i.e. the parameter ρ) plays a very important role on the performance of the AFEC scheme. The total needed RI of the AFEC scheme increase almost linearly with the increase of the parameter ρ. This phenomenon can be explained by the features of the GE channel. First, from Table 3.1 we can see that the value of $P_{B|B}$ increases with the increase of ρ. From (3.12), moreover, we know that the average value of the error-burst length in the GE channel increases with the increase of $P_{B|B}$. Consequently, the average number of missing data packets in one encoding block will increase with the increase of ρ. Therefore, in case of the parameter ρ increased, the AFEC scheme has to employ more parity packets to achieve a certain target PLR requirement. Because the length of the parameter k is usually short, even though the number of parity packets increase only a little for one encoding block, the total needed RI of the AFEC scheme still increases very much due to the inefficient code rate having to be adopted. To eliminate the passive effect of ρ in the AFEC scheme, the interleaving method proposed by [Yee95] can be employed, which however increases the complexity of the implementation and the latency. As a result, the effect of ρ must be taken into account and eliminated for the AFEC scheme.

6.4. Summary

For guaranteeing a certain target PLR requirement under strict delay constraints, an Adaptive FEC (AFEC) scheme with ideal erasure codes is proposed in this Chapter. The main ideal of the AFEC scheme is: The code rate of erasure codes automatically adapts to the worst channel state among all of the receivers in a multicast scenario. The great advantage of the AFEC scheme is: No feedback channels needed for this scheme. Theoretically, the AFEC scheme can be always the best choice for reliable transmissions if the length of the codeword is allowed to be without limits. As discussed in Chapter 2.2.2, however, the length of the codeword has to be constrained in practical systems due to the computational complexity. Especially, for RMM services, the number of source data packets (i.e. the length of the parameter k) in an encoding block has to be limited by the multimedia data rate and the strict delay constraints. Therefore, the performance of the AFEC scheme has to be constrained by the restrained length of the parameter k. Our studies in this Chapter show that this parameter k indeed has a significant effect on the performance of the AFEC scheme: Larger the length of the parameter k is allowed, better performance the AFEC scheme can achieve. Furthermore, since the CC of the GE channel model (i.e. the parameter

ρ) influences the distribution of the error-burst length deeply, this parameter also has a significant impact on the performance of the AFEC scheme. The analysis results show: To achieve a certain PLR requirement under strict delay constraints, bigger the value of the parameter ρ is, more the RI of the AFEC scheme needs. It indicates that the effect of the parameter ρ must be taken into account and eliminated for the AFEC scheme. As a result, both the parameter k and the parameter ρ affect the performance of the AFEC scheme significantly for RMM services in real systems. To overcome its shortages, we will consider combining FEC schemes with retransmission based ARQ schemes in the following Chapters.

Chapter 7
HEC-PR Scheme

Motivated by developing a Hybrid Error Correction (HEC) scheme for multimedia multicast services over Wireless Home Networks (WHNs), we will present a HEC scheme based on Packet Repetition (PR) technique in this Chapter. Strictly speaking, the HEC-PR scheme belongs to a kind of retransmission based scheme with pure ARQ techniques. The major feature of the HEC-PR scheme presented in this thesis is that multiple copies of each packet are allowed in each transmission stage. In the following, we will firstly give an overview on the background of developing the HEC-PR scheme. Then, we will present how to calculate its performance and optimize its parameters. Finally, by applying this scheme in a typical scenario, we will analyze its performance and compare it with the AFEC scheme presented in Chapter 6.

7.1. Background

With the increase in the consumers' demand for high quality digital TV (DTV), Digital Video Broadcasting (DVB) [1] project has developed solutions for DTV for almost all kinds of traditional transmission media [Rei06]: DVB-S [Ets0a] and DVB-S2 [Ets0b] for satellite broadcasting, DVB-C for use in cable networks [Ets0c], DVB-T for terrestrial transmission systems [Ets0d] and DVB-H for broadcasting to battery-powered devices [Ets0e]. Because

[1] Digital Video Broadcasting (http://www.dvb.org)

broadband access networks based on IP have become available in many parts of the world, it has been enabled the implement of DTV multicasting to home via broadband IP networks. The first DVB specification [Ets0f] describes the transport of MPEG-2 [Itu0a] based DVB services over IP based networks. With the imminent arrival of MPEG-4 Part 10 or ITU H.264 [Itu0b], DVB services over IP multicast has been also extended for this kind of transport streams. Furthermore, with the advent of home servers and multiple terminals used in one apartment or house, home networking is becoming more and more important. In the recent years, some researchers have studied the system for DTV multicasting to home over wired broadband IP access network [Luo05]. However, with the rapid development of broadband wireless networks, more and more attention has turned to distributing real time multimedia services over wireless home networks (WHNs). Recently, wireless LANs (WLANs) based on IEEE 802.11 [Iee0a] have come into widespread use for accessing the Internet and are expected to be used for DTV multicast services over home and nomadic networks. In the following, we will take the DVB services over IP based WHN using IEEE 802.11 as an example for explaining the importance of EER schemes.

From the specification [Ets0f] we know that the MPEG-2 transport stream is encapsulated into Real-time Transport Protocol (RTP) according to RFC 1889 [Rfc0a] in conjunction with RFC 2250 [Rfc0b], and then the RTP packets are transmitted by UDP/IP [Rfc0c] [Rfc0d] protocols. RTP does, however, not guarantee any quality-of-service (QoS) for real-time applications. Moreover, in WLANs with IEEE 802.11, collisions frequently occur when multiple wireless stations transmit packets since each station acts independently. When collisions occur, unicast packets are retransmitted on the MAC level, but multicast packets are lost because the MAC protocol does not use a retransmission control for multicast packets. Therefore, it is inevitable that some receivers will experience packet losses when the RTP/UDP/IP protocols are employed for multicasting in IEEE 802.11. As a matter of fact, it was discovered in [Fuj04] that the multicast PLR in IEEE802.11 based wireless LANs can reach 10% with the multicast date rate of about 500kbps. Hence, it is essential to employ error control techniques to guarantee the target PLR requirement on RTP level (e.g.

10^{-6}, refer to [Ets0f]) under strict end-to-end delay constraints (e.g. 100ms, refer to [Ets0f]) for DVB services. Because both the Round Trip Time (RTT) and multicast group size are usually very small for DVB services over WHN (e.g. the typical RTT is around 20 ms and the group size is less than 7[2]), the retransmission-based repair techniques can be used in such systems for overcoming the packet loss problem.

In order to perform a packet repair for the case with "small group" and "sufficiently small RTT", a modified RTP profile to repair missing packets by low delay feedback was proposed in [Ott04] (RTP/AVPF). Alternatively, the so called Negative-Acknowledgment (NACK) - Oriented Reliable Multicast (NORM) protocol [Ada04] is designed to provide reliable transport of data over IP multicast networks. NORM is a protocol centered on the use of selective NACKs to request repairs of missing packets and also provides for the use of PL-FEC techniques for efficient multicast repair and optional proactive transmission robustness. We have studied the performances of both RTP/AVPF and NORM by simulation using ns-2 [Tan06]. The studies showed that the RTP/AVPF cannot meet the high PLR requirement of 10^{-6} for DVB systems within the latency of 100ms even when the PLR of the wireless link is only about 3%. NORM can meet the PLR requirement, but the latency is too large, and the maximum latency is even more than 2 seconds, which is intolerable for DTV services. On the other hand, for guaranteeing the high PLR requirement with FEC alone by the NORM agent, the needed redundancy information (RI) will be large.

Alternatively, retransmission-based schemes can be considered for applications with "small group" and "sufficiently small RTT". For example, different retransmission-based schemes for error control in multicast protocols geared toward real-time multimedia applications are analyzed in [Pej96]. It is found that retransmission schemes are appropriate for such applications, and actually can be quite effective. In fact, as introduced before, the study had shown that the retransmission based error control schemes for point-to-point communication or single receiver in multicast scenario can outperform all the existing point-

[2] 5.1 multichannel audio plus wireless display

to-point schemes [Cha92]. Based on the studies above, we will propose a retransmission based HEC scheme with Packet Repetition techniques, which is always denoted by HEC-PR scheme in the thesis. Our target is to design the optimum parameters for the HEC-PR scheme, which will guarantee the predefined target PLR requirement under strict delay constraints with the minimum total needed RI.

Actually, the retransmission based ARQ scheme has been studied by many researchers. As in most high speed protocols, it is suggested to use a Selective Repeat (SR) scheme rather than go-back-n (GBN) [Arm92] [Bra93]. For point-to-point communication, some researchers proposed SR schemes that allowed multiple copies of a packet to be sent for retransmissions [Cha92] [Wel82] for reducing the delay, although the single copy is the optimum choice (for both transmissions and retransmissions) in ideal ARQ schemes [Wan89]. For point-to-multipoint communication, the analysis in [Lim95] aims to optimize data throughput only. In [Wan89] the authors aim to minimize the time between the first transmission of a packet and when the packet is received correctly by all receivers. In most of the schemes mentioned above, the transmitter waits for some time-out period to collect ACKs and/or NACKs from all receivers. This would be a logical step for improving data throughput in reliable point-to-multipoint communications, but it could be a serious drawback for real-time applications, since all receivers will be forced to accept the end-to-end delay of the worst receiver. Therefore, for multicast real-time multimedia, it is better to retransmit a packet immediately upon getting a NACK for the cases with small group size, which leads to much lower probabilities of dropping a packet, and slightly lower average packet delay [Pej96]. All of these schemes, however, do not consider the case of meeting a certain PLR requirement under strict delay constraints for RMM services. In this chapter, we will address this issue by presenting the HEC-PR scheme for RMM services over WHN with small group size and small RTT.

7.2. Introduction of the HEC-PR Scheme

Recalling the ARQ schemes introduced in Chapter 2.2.1, rather than focus on a particular transport protocol, an SR, NACK-only based retransmission based mechanism will be used for the HEC-PR scheme proposed in this Chapter. To illustrate the HEC-PR scheme in more detail, Figure 7.1 shows the basic operations at receivers with this scheme and Figure 7.2 also shows the basic operations at the sender with the scheme.

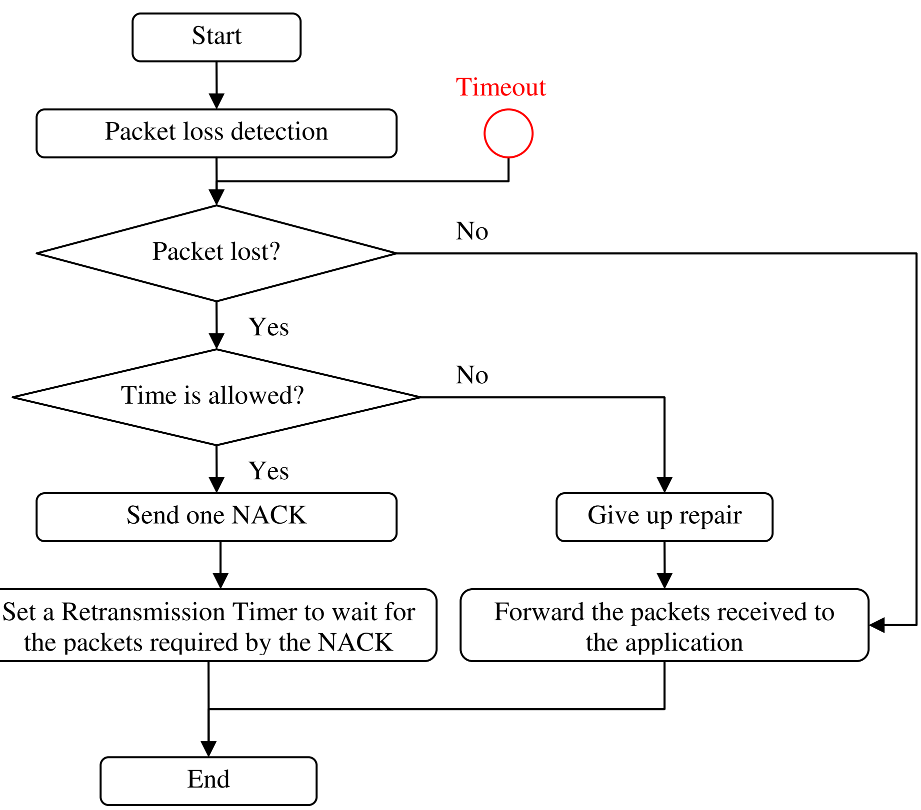


Figure 7.1: Operations at Receivers with the HEC-PR Scheme

As shown Figure 7.1, receivers will carry out the packet loss detection with the HEC-PR scheme. When packet losses detected, the receiver will first check if the time is allowed for further retransmissions. If not, it will give up repairing those lost packets and forward the packets received to the application. If yes, the receiver will send a NACK message as soon as possible to the sender for requiring retransmitting those lost packets, and at the same time it sets a retransmission timer for waiting for the lost data packets required by this NACK message. If all of the required missing packets are received, the timer will be deleted. If the timer is timed out, the receiver will check if all of the required lost data packets regarding this NACK message are received. If not, the receiver will check if the time is still allowed for more retransmission rounds; if the time is allowed, the receiver will repeat the repair procedure again.

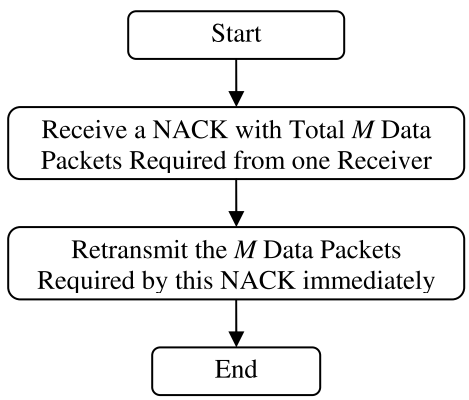


Figure 7.2: Operations at the Sender with the HEC-PR Scheme

From Figure 7.2, we can find that the operations of the sender indeed can be simplified very much when the receiver-based loss detection mechanism adopted. The sender, when receiving a NACK message with total *M* packets required for retransmissions from a receiver, as discussed in Chapter 2.2.1, there are two ways to respond this NACK message at the sender. The simplest way is that the sender will resend all of the *M* data packets immediately. The other way is more advanced: The sender will carry out repeat detection by checking whether the *M* data packets were retransmitted before in this retransmission round; then, the sender will only resend those non-repeat data packets in this retransmission round. In this Chapter, we only consider the simplest way for the HEC-PR scheme. The more advanced way will be considered in Chapter 8 and Chapter 9.

The SR, NACK-only based HEC-PR scheme actually has a number of advantages in the context of real-time multicast communications. It causes less traffic and processing at the sender [Pin94], since NACKs occur much less frequently than ACKs. Additionally, the timeout operation can be greatly simplified at the sender: All packets will be discarded after a fixed period which is bounded by the requirement of the end-to-end delay (i.e. D_{target}). According to the mechanisms of the HEC-PR scheme introduced above, we know that performance of the HEC-PR scheme mainly depends on the number of retransmission rounds and the number of copies for the retransmission packets during each retransmission round. However, what is the optimum performance of the retransmission based HEC-PR scheme and how it can achieve the optimum performance are still questions. To answer these questions, we will present how to evaluate the performances of the HEC-PR scheme and optimize its parameters in the subsequent sections. Now, the essential system parameters are defined and summed up in Table 7.1.

Table 7.1: System Parameters

Symbol	Definition
N_{recv}	The number of receivers in a multicast scenario
CSI(*j*)	Channel State Information of certain parameters ($P_{G\|G}(j)$, $P_{B\|B}(j)$) for the *j-th* receiver with GE model
$\rho(j)$	The CC of the GE channel model for the *j-th* receiver
X(*j*)	A random variable representing the error-burst-length during the first transmission for the *j-th* receiver with *CSI*(*j*)
Y(*j*)	A random variable representing the error-free-length during the first transmission for the *j-th* receiver with *CSI*(*j*)
$P_B(j)$	The steady state probability of being in state "B" for the *j-th* receiver with *CSI*(*j*)
RTT(*j*)	The round trip time for the *j-th* receiver, one way delay is *RTT*(*j*) /2. This definition is same to that in Chapter 4
$t_{lp}(j)$	The time duration from the time the latest packet loss occurs at the *j*-th receiver to the time it possibly receives this required packet, which is $RTT(j)+ t_{sw}+ t_{rw}$. This definition is same to that in Chapter 4

7.3. PLR Performance of the HEC-PR Scheme

Now let's focus on the PLR performance of the HEC-PR scheme introduced above. From the introduction above, we know that in this scheme one or multiple copies of the required data packets will be retransmitted in each retransmission stage at the sender. To calculate its performance, first of all, we need to define the essential parameters for the HEC-PR scheme, which are shown in Table 7.2.

Table 7.2: Parameters on the HEC-PR Scheme

Symbol	Definition
$N_{rr,max}(j)$	the maximum possible number of retransmission rounds for the *j-th* receiver with the HEC-PR scheme, where $1 \leq j \leq N_{recv}$.
$N_{rt}^{w}(j)$	the number of copies on each retransmission packet for the *j-th* receiver at the sender during the *w-th* retransmission round with the HEC-PR scheme, where $1 \leq w \leq N_{rr,\max}(j)$.
$N_{rt,\max}(j)$	the maximum possible number of retransmissions for each missing data packet for the *j-th* receiver at the sender with the HEC-PR scheme, which is given by: $N_{rt,\max}(j) = \sum_{w=1}^{N_{rr,\max}(j)} N_{rt}^{w}(j)$.

Due to the independence of the receivers, for the HEC-PR scheme with the simplest retransmission way at the sender, we only need to analyze the performances for only one receiver (without loss of generality, it is assumed to be the *j*-th receiver). From now on, for the convenience of description, we define an additional random variable as follows:

- $\mathrm{I}_{\mathrm{X}(j)}(w)$: A random variable representing the number of lost packets among the X(*j*) lost packets for the *j-th* receiver with *w* (where $0 \leq w \leq N_{rr,max}(j)$) retransmission rounds.

Since the expected values of the original error-burst length and the error-free length can be computed by (3.12) and (3.13) (i.e. $E(\mathrm{X}(j))=1/(1-P_{BB}(j))$ and $E(\mathrm{Y}(j))=1/(1-P_{GG}(j))$), the PLR performance of the HEC-PR scheme for the *j-th* receiver can be calculated as: $E(\mathrm{I}_{\mathrm{X}(j)}(w))/(E(\mathrm{X}(j))+E(\mathrm{Y}(j)))$. It actually is the final PLR at the *j-th* receiver after all of the retransmission packets experienced *w* retransmission rounds. The parameters

($P_{G|G}(j)$, $P_{B|B}(j)$) of *CSI*(j) can also be estimated by some accurate methods (e.g. the Maximum Likelihood Estimation method proposed in Chapter 3.3). It is assumed that both the sender and the receiver have perfect knowledge of the *CSI*(j) in this thesis. How to obtain the perfect information of *CSI*(j) at both the sender and the receivers needs further studies in the future. Note that $E(\mathrm{I}_{\mathrm{X}(j)}(0))$ (i.e. w=0) is actually equal to $E(\mathrm{X}(j))$ so that the PLR performance of the HEC-PR scheme is equal to the original link PLR of PLR_{GE} with *CSI*(j) for the *j-th* receiver. In the following, it is assumed that the w is more than zero for the HEC-PR scheme.

Now the remaining task is to calculate $E(\mathrm{I}_{\mathrm{X}(j)}(w))$ for this receiver, which can be derived from the probability of the GE channel still being in state "B" after X(j) packets experienced w retransmission rounds. As analyzed in Chapter 5.2, this transition probability depends on the duration time of $t_{lp}(j)$ for retransmission packets for the j-th receiver due to the memory feature of the GE channel. Similar to (5.3), the value of $t_{lp}(j)$ can be expressed in discrete form by the number of packet intervals:

$$T_{lp}(j) = \left\lceil \frac{t_{lp}(j)}{t_s} \right\rceil \tag{7.1}$$

Upon (7.1), therefore, the loss probability of each retransmission data packet during the q-th retransmission round is $\left(P_{B|B}[q \cdot T_{lp}(j)]\right)^{N_{rt}^{q}(j)}$. According to (3.9) and (3.23), the value of $P_{B|B}[q \cdot T_{lp}(j)]$ can be calculated by:

$$P_{B|B}[q \cdot T_{lp}(j)] = P_B(j) + (1 - P_B(j)) \cdot \rho(j)^{q \cdot T_{lp}(j)} \tag{7.2}$$

Remark: Considering practical implementations, because the $P_B(j)$ is always less than one and $\rho(j)$ usually being less than 0.1 in real systems, the second part of the $P_{B|B}[q \cdot T_{lp}(j)]$ in (7.2) is negligible when $q \cdot T_{lp}(j)$ is much larger than one. In RMM systems with high data rate, as a fact, t_{lp} is usually multiple times of t_s resulting in T_{lp} being much larger than one. Therefore, the value of $P_{B|B}[q \cdot T_{lp}(j)]$ can be approximated as $P_B(j)$ with $q \geq 1$ in most real systems so that the calculation of (7.2) can be simplified very much.

Base on the analysis above, using (3.10) and (7.2), the PDF of $\mathrm{I}_{\mathrm{X(j)}}(w)$ then can be expressed in this form:

$$\Pr(\mathrm{I}_{\mathrm{X}(j)}(w) = g) = P_{X(j)}^{g} \cdot \prod_{q=1}^{w} \left(P_{B|B}[q \cdot T_{lp}(j)]\right)^{N_{rt}^{w}(j)} \tag{7.3}$$

$$\text{Where: } g = 1,2,\ldots,\infty, 1 \leq w \leq N_{rr,\max}(j)$$

Follows (7.3), the expected value of $\mathrm{I}_{\mathrm{X(j)}}(w)$ can be expressed as:

$$E(\mathrm{I}_{\mathrm{X}(j)}(w)) = \sum_{g=1}^{\infty} g \cdot \Pr(\mathrm{I}_{\mathrm{X}(j)} = g) = \sum_{g=1}^{\infty} g \cdot P_{X(j)}^{g} \cdot \prod_{q=1}^{w} \left(P_{B|B}[q \cdot T_{lp}(j)]\right)^{N_{rt}^{w}(j)} \tag{7.4}$$

Upon the above analysis, after experiencing w (where $1 \leq w \leq N_{rr,\max}(j)$) retransmission rounds with the HEC-PR scheme for the *j-th* receiver, the PLR performance of the HEC-PR scheme is:

$$PLR_{\mathrm{HEC}-PR}(j,w) = \frac{E(\mathrm{I}_{\mathrm{X}(j)}(w))}{E(\mathrm{X}(j)) + E(\mathrm{Y}(j))} \tag{7.5}$$

Note, as mentioned above, the value of $P_{B|B}[q \cdot T_{lp}(j)]$ can be approximated as $P_B(j)$ in case of $q \geq 1$ and the CC of GE channel being small enough. In this chapter, we adopt this approximation as the value of $P_{B|B}[q \cdot T_{lp}(j)]$ to evaluate the performance of the HEC-PR scheme quickly. More accurate performance of the HEC-PR scheme will be discussed in Chapter 8 and Chapter 9. Using this approximation, Eq. (7.4) then can be simplified as this form:

$$E(\mathrm{I}_{\mathrm{X}(j)}(w)) = \sum_{g=1}^{\infty} g \cdot P_{X(j)}^{g} \cdot \prod_{q=1}^{w} \left(P_{B|B}[q \cdot T_{lp}(j)]\right)^{N_{rt}^{w}(j)} \tag{7.6}$$

$$= \sum_{g=1}^{\infty} g \cdot P_{X(j)}^{g} \cdot (P_B(j))^{\sum_{q=1}^{w} N_{rt}^{q}(j)}$$

In fact, Eq.(7.6) indicates that the loss probability of each group of the retransmission packets during each retransmission round can be approximated by the steady state probability $P_B(j)$ when $T_{lp}(j)$ is large enough. As a result, after experiencing $N_{rr,max}(j)$ retransmission rounds, following (7.5) and (7.6), the PLR performance of the HEC-PR scheme for the *j-th* receiver turns out to be:

$$PLR_{HEC-PR}(j, N_{rr,\max}(j)) = \frac{E(\mathrm{I}_{\mathrm{X}(j)}(N_{rr,\max}(j)))}{E(\mathrm{X}(j)) + E(\mathrm{Y}(j))} \tag{7.7}$$

$$= \frac{\sum_{g=1}^{\infty} g \cdot P_{X(j)}^{g} \cdot (P_B(j))^{\sum_{q=1}^{N_{rr,\max}(j)} N_{rt}^{q}(j)}}{E(\mathrm{X}(j)) + E(\mathrm{Y}(j))}$$

$$= \frac{\sum_{g=1}^{\infty} g \cdot P_{X(j)}^{g} \cdot (P_B(j))^{N_{rt,\max}(j)}}{E(\mathrm{X}(j)) + E(\mathrm{Y}(j))}$$

By looking into (7.7), we can find that the simplified mathematical expression is actually a function with two parameters: *CSI*(*j*) and $N_{rt,max}(j)$, which is denoted by $f_{PLR,HEC-PR}(N_{rt,\max}(j), CSI(j))$ in the Chapter.

7.4. RI Performance of the HEC-PR Scheme

To evaluate the performance of the HEC-PR scheme and compare it with the AFEC scheme introduced in Chapter 6, now let's consider the total needed RI for the *j-th* receiver with the HEC-PR scheme. Note that for the *j-th* receiver the needed RI with the HEC-PR scheme in the *q-th* retransmission round can be derived from the parameter $N_{rt}^{q}(j)$ and the final PLR with (*q*-1)-*th* retransmission rounds. That is, for the *j*-th receiver, the needed RI in the q-th retransmission round can be calculated by $N_{rt}^{q}(j)PLR_{H-PR}(j,q-1)$, where the value of $PLR_{H-PR}(j,q-1)$ can be obtained by (7.5). In this chapter, it is assumed that no any redundancy packets transmitted in the first transmission. Thus, there is no any RI in the first transmission. General cases of allowing adding RI in any transmission stage will be discussed in Chapter 8. Finally, the needed RI for the j-th receiver with the HEC-PR scheme proposed in this chapter is:

$$RI_{HEC-PR}(j,N_{rr,\max}(j)) = \sum_{q=1}^{N_{rr,\max}(j)} N_{rt}^{q}(j)PLR_{\mathrm{HEC}-PR}(j,q-1) \tag{7.8}$$

In order to simplify the description, we define a vector as $\vec{N}_{rt}(j) = \left\{N_{rt}^{q}(j) \middle| 1 \le q \le N_{rr,\max}(j)\right\}$ upon the parameters of the HEC-PR scheme. Now Eq.(7.8) can be expressed as a function with the parameters: $CSI(j)$ and $\vec{N}_{rt}(j)$, which is denoted by $f_{RI,HEC-PR}(\vec{N}_{rt}(j),CSI(j))$ in this chapter. Due to the independence of the receivers and the simplest retransmission way at the sender, all of the lost packets for each receiver will be retransmitted without any suppression. The total needed RI for all of the N_{recv} receivers with the HEC-PR scheme thus is:

$$RI_{total,HEC-PR} = \sum_{j=1}^{N_{recv}} f_{RI,HEC-PR}(\vec{N}_{rt}(j),CSI(j)) \tag{7.9}$$

Form (7.9), we can see that the total needed RI of the HEC-PR scheme will increase significantly with the increase of the number of receivers. The main reason is that all of the receivers are independent and any receiver can not benefit from the retransmission data packet required by other receivers.

7.5. Optimization of the HEC-PR Scheme

Relying on the mathematical framework proposed for the HEC-PR scheme above, we now introduce a method to design the suitable parameters of the scheme for the *j-th* receiver. Given a certain multicast scenario, from Chapter 7.2, we know that for the *j*-th receiver with $T_{lp}(j)$ and $CSI(j)$, the performance of the HEC-PR scheme depends only on the parameter $\vec{N}_{rt}(j)$. Now the task is to find out the most suitable $\vec{N}_{rt}(j)$ for the HEC-PR scheme to guarantee the target PLR requirement under a strict delay constraint with the minimum needed RI.

First of all, to achieve the target PLR requirement for the *j-th* receiver, the PLR performance of this scheme must satisfy:

$$f_{PLR,HEC-PR}(N_{rt,\max}(j), CSI(j)) \le \mathrm{PLR}_{\mathrm{target}} \tag{7.10}$$

Note that we have assumed that the $CSI(j)$ are known perfectly by measurements and estimations in this thesis. Depending on (7.10), then, we can search for the minimum value for the parameter $N_{rt,\max}(j)$ (which is defined as $\hat{N}_{rt,\max}(j)$ here) using numerical algorithm.

On the other hand, according to the delay budget for ARQ schemes presented in Chapter 4.2, we know that the maximum possible number of retransmission rounds for the HEC-PR scheme is limited by the strict delay constraints. For those retransmission packets in the *j-th* receiver, as analyzed in Chapter 4.2, the maximum possible end-to-end delay includes two parts: the link delay in the first transmission (which is $RTT(j)/2$), and the delay for the $N_{rr,max}(j)$ retransmission rounds (which is $N_{rr,max}(j)(t_{lp}(j)+t_{rp})$). Note, as mentioned in Chapter 4.2, the value of t_{rp} can be set a fixed value for all of the receivers in a certain multicast scenario. To satisfy the delay requirement for the *j-th* receiver, therefore, the maximum possible number of retransmission rounds $N_{rr,max}(j)$ must be limited by:

$$N_{rr,\max}(j)(t_{lp}(j)+t_{rp})+\frac{RTT(j)}{2}\le \mathrm{D}_{\text{target}} \tag{7.11}$$

Relying on (7.11), the parameter $N_{rr,max}(j)$ of the HEC-PR scheme can be decided by:

$$\hat{N}_{rr,\max}(j)=\left\lfloor \frac{\mathrm{D}_{\text{target}}-\dfrac{RTT(j)}{2}}{t_{lp}(j)+t_{rp}} \right\rfloor \tag{7.12}$$

In order to make sure that $N_{rr,max}(j)$ is not more than $N_{rt,\max}(j)$ (for the reasonable consideration in practical systems), we define $\hat{\hat{N}}_{rr,\max}(j)=\min(\hat{N}_{rr,\max}(j),\hat{N}_{rt,\max}(j))$. We then obtain a reasonable range for the parameter $N_{rr,max}(j)$ to satisfy the delay requirement, which is: $1\le N_{rr,\max}(j)\le \hat{\hat{N}}_{rr,\max}(j)$. Note that for each possible value of $N_{rr,max}(j)$, we should choose the parameter $\vec{N}_{rt}(j)$ with $\sum_{q=1}^{N_{rr,\max}(j)} N_{rt}^{q}(j)=\hat{N}_{rt,\max}(j)$ for this scheme to guarantee the target PLR requirement. Obviously, the vector $\vec{N}_{rt}(j)$ is limited in a finite space $\Phi^{\hat{\hat{N}}_{rr,\max}(j)}$ (i.e. $\vec{N}_{rt}(j)\in\Phi^{\hat{\hat{N}}_{rr,\max}(j)}$), which is given by:

$$\Phi^{\hat{\hat{N}}_{rr,\max}(j)}=\left\{\left[a_1,a_2,\ldots,a_{N_{rr,\max}(j)}\right]\;\middle|\;\begin{array}{l}1\le N_{rr,\max}(j)\le \hat{\hat{N}}_{rr,\max}(j)\\ \sum_{k=1}^{N_{rr,\max}(j)} a_k=\hat{N}_{rt,\max}(j)\\ a_i\ge 1, N_{rr,\max}(j)\ge i\ge 1\end{array}\right\} \tag{7.13}$$

Note that the size of the search space $\Phi^{\hat{\hat{N}}_{rr,\max}(j)}$ is determined by the parameters $\hat{N}_{rt,\max}(j)$ and $\hat{\hat{N}}_{rr,\max}(j)$. Given certain $\hat{N}_{rt,\max}(j)$ and $\hat{\hat{N}}_{rr,\max}(j)$, the space size is $\begin{pmatrix}\hat{N}_{rt,\max}(j)-1\\ \hat{\hat{N}}_{rr,\max}(j)-1\end{pmatrix}$.

As a result, since our optimum target is to satisfy the target QoS requirements with the minimum needed RI, the optimization problem for the *j-th* receiver is:

$$RI_{opt,HEC-PR}(j) = \underset{\vec{N}_{rt}(j)\in\Phi^{\hat{\hat{N}}_{rr,\max}(j)}}{\arg\min}\ (f_{RI,HEC-PR}(\vec{N}_{rt}(j), CSI(j))) \tag{7.14}$$

Subject to:

$$f_{PLR,HEC-PR}(N_{rt,\max}(j), CSI(j)) \le \mathrm{PLR}_{\mathrm{target}}$$

Note that the space $\Phi^{\hat{\hat{N}}_{rr,\max}(j)}$ is usually very small for practical RMM services over WHNs. For example, to achieve a target PLR of 10^{-6} under the delay constraints of 100ms for DVB services, both $\hat{N}_{rt,\max}(j)$ and $\hat{\hat{N}}_{rr,\max}(j)$ are not more than 5 with small link PLR so that the full space size is actually small. Therefore, we can obtain the optimal $\vec{N}_{rt}(j)$ (denoted by $\vec{N}_{rt,\ opt}(j)$ in this chapter) by full searching algorithm. On the other hand, considering the practical implementation, we can determine the optimum parameters $\vec{N}_{rt,\ opt}(j)$ off-line based on different values of $t_{lp}(j)$ and $CSI(j)$, and then make tables. The real systems then can adopt the optimum parameters by simply looking up those tables upon the current $t_{lp}(j)$ and $CSI(j)$.

Finally, since it is assumed that the sender retransmits all of the data packets required by all of the receivers without repetition detection in this chapter, we can design the suitable parameters of the HEC-PR scheme for each receiver independently by solving (7.14). As a result, using (7.9), the minimum total needed RI for all of the N_{recv} receivers can be computed by:

$$\begin{aligned} RI_{total,HEC-PR} &= \sum_{j=1}^{N_{recv}} RI_{opt,HEC-PR}(j) \\ &= \sum_{j=1}^{N_{recv}} f_{RI,HEC-PR}(\vec{N}_{rt,opt}(j), CSI(j)) \end{aligned} \tag{7.15}$$

7.6. Performance Analysis

In this section, we firstly analyze the performance of the HEC-PR scheme over the GE channel model; and then compare it with the AFEC scheme. For the convenience of comparison and analysis, we make two assumptions as follows: The entire receivers experience an independent GE channel with the same level of original link PLR and the same ρ ; the channel bandwidth is always so large enough that all of the retransmission packets in each retransmission round can be retransmitted during the time duration of one packet interval (i.e. $t_{rp}=t_s$), which actually can be viewed as the ideal condition as analyzed in Chapter 4.2. Therefore, from (3.34) and (3.35), it is known that we always have $t_s>t_d$ for this case. Here we consider DVB services over WHNs with a group size of less than 7, RTT of less than 15ms and a wireless link PLR of up to 10% when the video multicast data rate is more than 500Kbps [Fuj04]. However, it should be clear that the mathematical framework of the HEC-PR scheme is suitable for any wireless multicast scenario under strict delay constraints. Now we apply the HEC-PR scheme and the AFEC scheme in a typical scenario with the common system parameters, which are summarized in Table 7.3.

Table 7.3: System Parameters for Analysis

PLR Requirement: PLR_{target}	10^{-6}
Delay Constraints: D_{target}	100ms
Packet Loss Model:	GE Model
Data Rate: R_d	4Mbps
RTT:	15ms
$t_{rw}+t_{sw}$:	0ms
Packet Size:	1250bytes
Original Link PLR: P_e	10^{-3}~10^{-1}

Note, in the analysis, the packet interval t_s can be computed according to the multimedia data rate and the packet size (i.e. t_s=2.5ms with the multicast data rate of 4Mbps in this case); and the parameters of the GE model can be produced upon the parameters ρ and P_e by the method introduced in Chapter 3.3.

7.6.1. Performance Analysis with Single Receiver

Now we begin to analyze the performance of the HEC-PR scheme with only one receiver (i.e. N_{recv}=1). Then, all of the parameters described in the Section above can be simplified as those without the sequence number for a special receiver. Therefore, those parameters $\hat{N}_{rt,\max}(j), \hat{\hat{N}}_{rr,\max}(j), \vec{N}_{rt}(j)$ etc. for the j-th receiver will be denoted by $\hat{N}_{rt,\max}, \hat{\hat{N}}_{rr,\max}, \vec{N}_{rt}$ etc. in this section. To simplify the analysis, the effect of the correlation coefficient (CC) of the GE model is not considered in this section, i.e. the CC of the GE model is set to zero. Its effect on the HEC-PR scheme will be analyzed later.

First of all, let's focus on the influence of the system parameters RTT and t_s to the maximum allowable number of retransmission rounds $\hat{N}_{rr,\max}$ on the HEC-PR scheme. Without loss of generality, it is assumed that the link PLR is less than 10% in this case. By solving (7.10), then, we can derive that $\hat{N}_{rt,\max}$ is always not more than 5 with the target PLR requirement of 10^{-6} under the strict delay constraints of 100ms. In other words, when applying the proposed HEC-PR scheme for this case with $\hat{N}_{rt,\max} = 5$, the target PLR requirement can be met under the target delay constraints. Therefore, the $\hat{N}_{rt,\max}$ can be fixed to 5 for this case. Using this fixed value of $\hat{N}_{rt,\max}$, as a result, we can obtain the maximum allowable number of retransmission rounds $\hat{N}_{rr,\max}$ for this receiver with different RTT and t_s by (7.12). Figure 7.3 shows the calculated results of $\hat{N}_{rr,\max}$ for different RTT and t_s with $P_e = 10^{-1}$ and $\hat{N}_{rt,\max} = 5$.

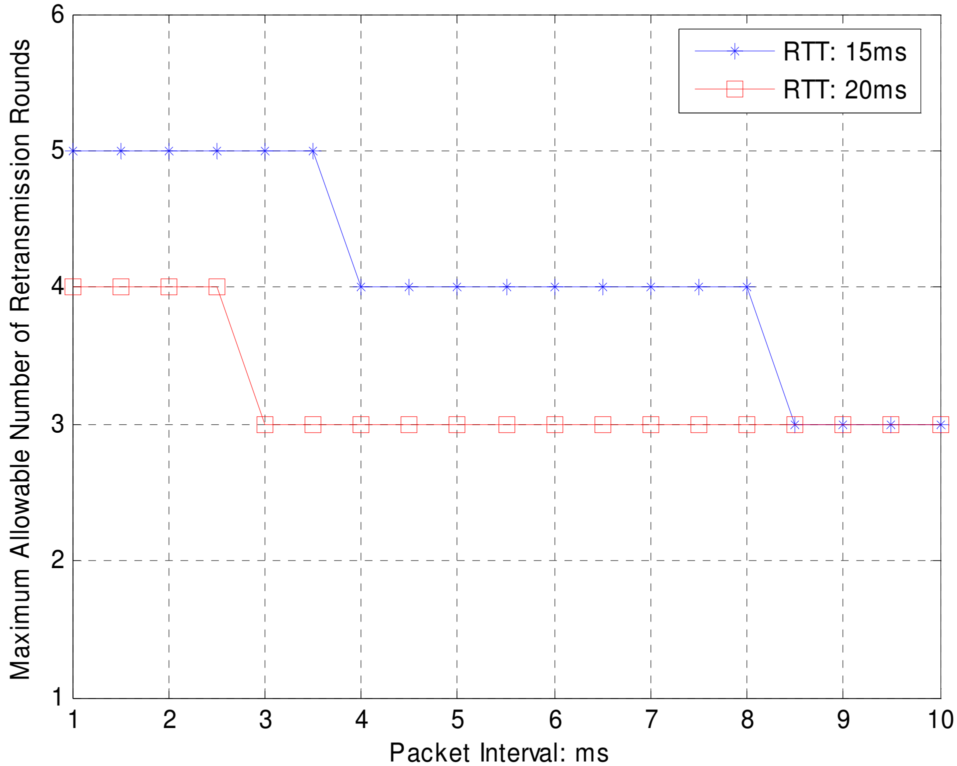


Figure 7.3: The maximum allowable number of retransmission rounds $\hat{N}_{rr,\max}$ for single receiver with $P_e = 10^{-1}$ and $\hat{N}_{rt,\max} = 5$

As shown in Figure 7.3, both t_s and RTT have a noticeable effect on the parameter $\hat{N}_{rr,\max}$. Especially, it shows that this parameter is very sensitive to RTT. For example, when the packet interval t_s varies between 1.0ms and 8.0ms, the value of $\hat{N}_{rr,\max}$ with RTT=15ms is always more than that with RTT=20ms, which indicates that within a range of data rate the HEC-PR scheme can have larger value of $\hat{N}_{rr,\max}$ for the receivers with smaller RTT. On the other hand, Figure 7.3 also shows that this parameter has the same value for some range of t_s, which indicates that within a range of data rate the value of $\hat{N}_{rr,\max}$ is identical if the RTT is not variable. That is, the parameter $\hat{N}_{rr,\max}$ is not sensitive to the multicast data rate. Since the parameter RTT is usually stable in practical systems, we can obtain a stable parameter

$\hat{N}_{rr,\max}$ for the receiver in the HEC-PR scheme. Therefore, even if the source multicast data rate is variable; the parameter of the HEC-PR scheme is also fixed so that it can work robustly.

Applying the HEC-PR scheme in the typical scenario with the parameters as shown in Table 7.3, in the following, we analyze the performances of the HEC-PR scheme using the mathematical framework proposed above. Above all, according to different original link PLR, we find out the minimum $N_{rt,max}$(i.e. $\hat{N}_{rt,\max}$) to satisfy the target PLR requirement by solving (7.10). Furthermore, using (7.12), the maximum allowable value of $N_{rr,max}$ (i.e. $\hat{N}_{rr,\max}$) can also be obtained so that we can get the parameter $\hat{\hat{N}}_{rr,\max}$ by computing min($\hat{N}_{rt,\max}$, $\hat{N}_{rr,\max}$). Actually, from Figure 7.3, we can see that the parameter $\hat{N}_{rr,\max}$ will be 5 for this case with t_s=2.5ms and RTT=15ms. Thus, based on the system parameters in Table 7.3, both $\hat{N}_{rt,\max}$ and $\hat{\hat{N}}_{rr,\max}$ are computed with different link PLR and summarized in the flowing Table 7.4.

Table 7.4: $\hat{N}_{rt,\max}$ and $\hat{\hat{N}}_{rr,\max}$ for the HEC-PR scheme for this typical scenario

Original Link PLR: P_e	0.001	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$\hat{N}_{rt,\max}$	1	2	3	3	4	4	4	5	5	5	5
$\hat{\hat{N}}_{rr,\max}$	1	2	3	3	4	4	4	5	5	5	5

Form Table 7.4, we can see that the maximum value of the parameter $N_{rr,max}$ for the HEC-PR scheme can be set to five for a typical scenario with P_e=0.10 and RTT=15ms. In other words, the parameter $N_{rr,max}$ of the HEC-PR scheme can be vary from zero to five. Actually, for each possible value of $N_{rr,max}$, we can search for the most suitable parameters $\vec{N}_{rt}$ for the HEC-PR scheme, which is called *local optimum solution* of the scheme in the following. For understanding the correlation between the *local optimum solution* and the *global optimum solution*, we analyze the *local optimum solution* by searching for $\vec{N}_{rt}$ according to each possible value of $N_{rr,max}$ (where $1 \le N_{rr,\max} \le \hat{\hat{N}}_{rr,\max}$) with $\hat{N}_{rt,\max} = 5$ and $\hat{\hat{N}}_{rr,\max} = 5$. By solving (7.14), we obtain the *local optimum solutions* for the HEC-PR scheme, which are shown in Table 7.5.

Table 7.5: The *local optimum solution* of the HEC-PR scheme for single receiver with P_e=0.10

$N_{rr,max}$	$\bar{N}_{rt}$			
	N_{rt}^1	N_{rt}^2	N_{rt}^3	N_{rt}^4
1	5	-	-	-
2	1	4	-	-
3	1	1	3	-
4	1	1	1	2

Relying on this table, using (7.8), we can obtain the needed RI for the local optimum results with different $N_{rr,max}$. Figure 7.4 plots the calculated results for the HEC-PR scheme with $N_{rr,max}$=1~4 . In order to show the limit of the performance of the HEC-PR scheme, Figure 7.4 also shows the Shannon limit obtained by (3.2) for erasure error channels.

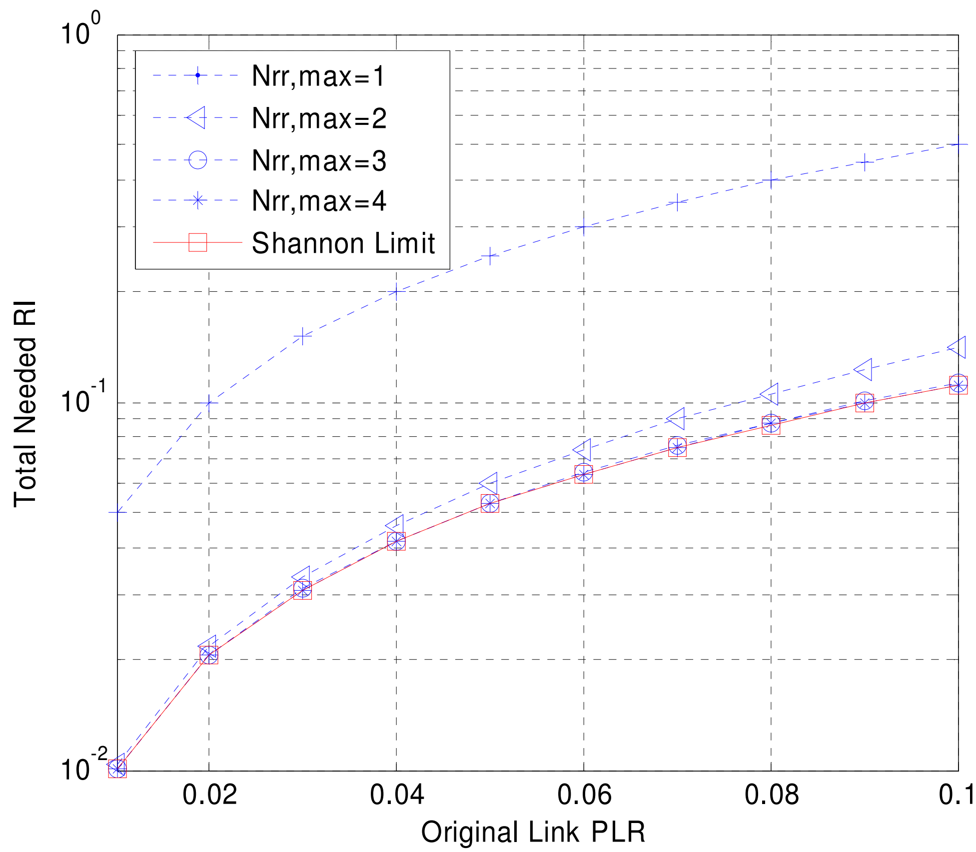


Figure 7.4: The total needed RI by the *Local Optimum Solutions* for a single receiver with P_e=0.10

As shown in Figure 7.4, the parameter $N_{rr,max}$ has a significant effect on the total needed RI for the HEC-PR scheme. For example, this figure shows that the total needed RI of the HEC scheme with $N_{rr,max}$=1 is much more than that with $N_{rr,max}$>1. The reason is clear: As shown in Table 7.5, when only one retransmission round is allowed, multi-copies of those required packets have to be retransmitted in the first retransmission stage to achieve the target PLR requirement. However, when multiple retransmission rounds are allowed, the multi-copies of those required packets only need to be retransmitted in the last retransmission round. As a matter of fact, the total number of required packets in the first retransmission round will be much more than that required in the following retransmission rounds due to the previous recovery process at the receiver. Therefore, more the number of retransmission rounds are allowed, less the total needed RI will be needed.

Furthermore, Figure 7.4 shows that the needed RI of the scheme with $N_{rr,max}$=2 is only a little more than that with $N_{rr,max}$>2. That is, to make sure that the HEC-PR scheme can perform well, the parameter $N_{rr,max}$ should be set at least to two. Additionally, from this figure we can see that the performance of the HEC-PR scheme with $N_{rr,max}$=3 is almost the same to that with $N_{rr,max}$=4 (Actually, which is nearly the same to the global optimum solution for this case). Therefore, to close to the best performance for the HEC-PR scheme, the parameter $N_{rr,max}$ should be set at least to three. In other words, the parameter $N_{rr,max}$ should be no less than three for this case.

Finally, from Figure 7.4, we can see that the performance of the HEC-PR scheme with $N_{rr,max}$=3 can even reach the Shannon limit. This means that under the strict delay constraint the HEC-PR scheme can guarantee the target PLR requirement using only three retransmission rounds with very small needed RI. Since it is very hard to achieve the Shannon limit by using existing FEC techniques only, the proposed HEC-PR scheme with small number of retransmission rounds is much more attractive than that with FEC alone for this case with a single receiver.

As a summary, considering the practical implementation, we must point out that the most reasonable value for $N_{rr,max}$ should be three for this case; although the HEC-PR scheme with $N_{rr,max}$=5 is the much more close to the global optimum solution (Note that it is also close to the Shannon limit). The main reason is that the performance of the HEC-PR scheme with $N_{rr,max}$=3 is almost same to that with $N_{rr,max}$=4, and it causes less latency and also easier to implement due to the less number of retransmission rounds.

7.6.2. Performance Comparison

In this section, based on the typical scenario as shown in Table 7.3, we analyze the performance of the HEC-PR scheme using the mathematical framework proposed in previous sections. In the following, we will choose the value of $N_{rr,max}$ in designing the HEC-PR scheme based on this rule: Achieve almost the optimum performance with the smallest number of retransmission rounds. According the analysis above, we know that the parameter $N_{rr,max}$ on the HEC-PR scheme should be set to three with RTT=15ms for achieving almost the optimum performance. Actually, it can be viewed as the most reasonable choice for this case by setting $N_{rr,max}$=3 with the nearly optimum performance. Similar to the analysis above, we firstly design the parameters of the HEC-PR scheme for the scenario with only one receiver. Follow this idea, we find out the minimum $\hat{N}_{rt,\max}$ to satisfy the PLR requirement by solving (7.10), and obtain the most reasonable value for the parameter $\hat{N}_{rr,\max}$ by (7.12). In fact, based on Table 7.4, both $\hat{N}_{rt,\max}$ and $\hat{\hat{N}}_{rr,\max}$ can be obtained directly, which are summarized in Table 7.6.

Table 7.6: Reasonable $\hat{N}_{rt,\max}$ and $\hat{\hat{N}}_{rr,\max}$ for the HEC-PR scheme for this typical case with single receiver

Original Link PLR: P_e	0.001	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$\hat{N}_{rt,\max}$	1	2	3	3	4	4	4	5	5	5	5
$\hat{\hat{N}}_{rr,\max}$	1	2	3	3	3	3	3	3	3	3	3

Using the Table 7.6, upon (7.14), we then can search for the local optimum parameters $\vec{N}_{rt}$ for the HEC-PR scheme with different original link PLR. At a result, the search results regarding $\vec{N}_{rt}$ with different range of original link PLR are shown in Table 7.7.

Table 7.7: The search results of $\vec{N}_{rt}$ for the HEC-PR scheme with different link PLR for the scenario with single receiver

Original Link PLR P_e	$\vec{N}_{rt}$		
	N_{rt}^1	N_{rt}^2	N_{rt}^3
$P_e \leq 10^{-3}$	1	-	-
$10^{-3}<P_e\leq 10^{-2}$	1	1	-
$10^{-2}<P_e\leq 3\times 10^{-2}$	1	1	1
$3\times 10^{-2}<P_e\leq 6\times 10^{-2}$	1	1	2
$6\times 10^{-2}<P_e\leq 10^{-1}$	1	1	3

Note, since the same level of original link PLR and independence among all of the receivers are assumed, the total needed RI is simply the multiple of the needed RI for single receiver derived upon Table 7.7. Depending on those parameters in Table 7.7, we now can calculate the minimum total needed RI of overall receivers by (7.15), which is shown in Figure 7.5.

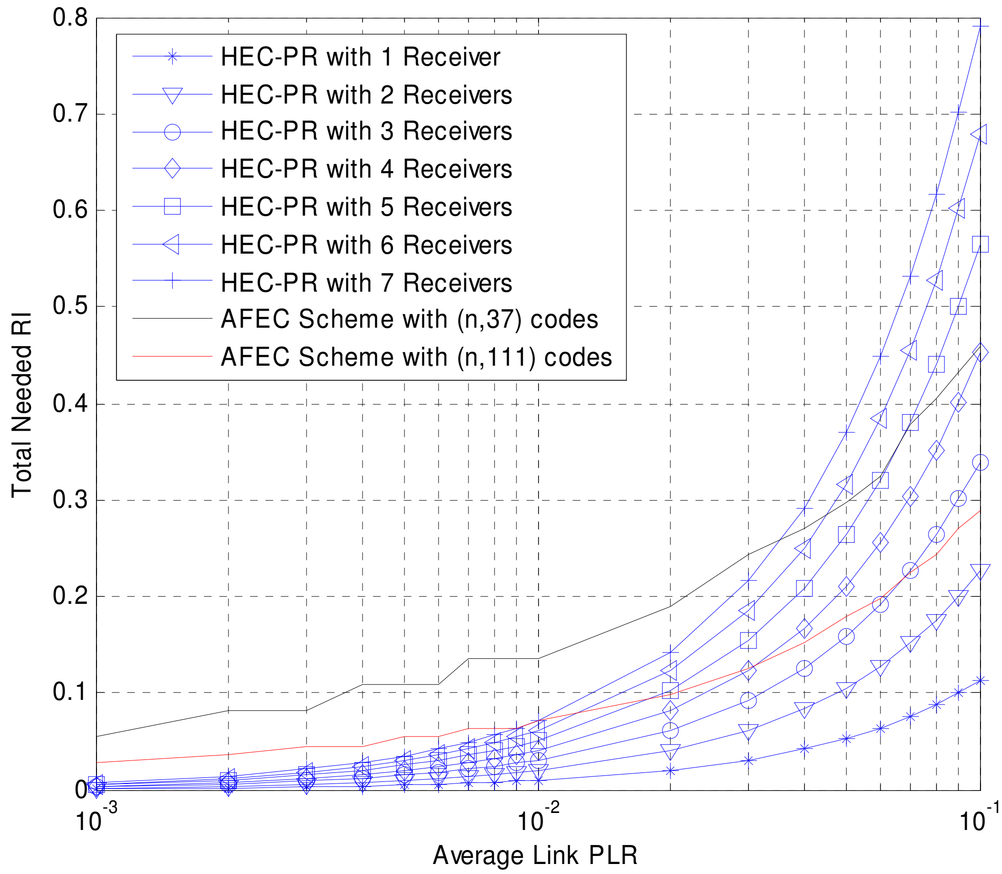


Figure 7.5. The Total Needed RI for the HEC-PR Scheme and the AFEC Scheme

On the performance of the HEC-PR scheme shown in Figure 7.5, although the parameters of the HEC-PR scheme are designed for the scenarios with multicast data rate of 4Mbps, we must point out that it also can be viewed as the performance of the HEC-PR scheme with the multicast data rate higher than 4Mbps. The reason is clear: Higher data rate is, small packet interval is, which results in larger $N_{rr,max}$; however, the HEC-PR scheme has achieved almost the optimum performance with only $N_{rr,max}$=3. Therefore, for the scenarios with higher data rate, the performance of the HEC-PR scheme will be almost the same to that as shown in Figure 7.5 with $N_{rr,max}$=3.

To compare the HEC-PR scheme with the AFEC scheme introduced in Chapter 3, in the following, we also design the parameters of the AFEC scheme for the scenario with SDTV and HDTV services, respectively. According to the AFEC scheme, first, we need to calculate the k_{lim} for the ideal FEC code used under the delay constraints (100ms) by (6.12). Note that under the ideal condition we always have $t_s > t_d$, and the transmission of each source data packet occupies exactly one packet interval of t_s. Therefore, to compare the performance of the AFEC scheme with the HEC-PR scheme fairly and guarantee the strict delay requirement, we use t_s instead of t_d in (6.12) for obtaining the parameter k_{lim}. According to (6.12), it is easy to know that the parameter k_{lim} is 37 for the SDTV service with R_d=4Mbps and 111 for the HDTV service with R_d=12Mbps. Then, we can find the perfect n to meet the target PLR requirement (10^{-6}) upon the original link PLR level by solving (6.14). These results are shown in Table 7.8 as follows.

Table 7.8: Perfect (n_{opt}, k_{lim}) codes of the AFEC scheme to meet the target PLR (10^{-6}) requirement under strict delay constraints of 100ms

Original Link PLR: P_e		0.001	0.01	0.03	0.05	0.07	0.09	0.10
n_{opt}	k_{lim}=37	39	42	46	48	51	53	54
	k_{lim}=111	114	119	125	131	136	141	143

Based on Table 7.8, the total needed RI for the AFEC can be computed by (6.15) for these two cases, which are also plotted in Figure 7.5. Note that the number of the receivers in multicast scenario has no any effect on the total needed RI for the AFEC scheme.

As shown in Figure 7.5, when the multicast data rate is 4Mbps (or 12Mbps), the performance of the HEC-PR scheme with $N_{recv} \leq 4$ (or $N_{recv} \leq 2$) correspondingly is always better than that of AFEC scheme. Also, in case of the average link PLR being less than 0.01, the HEC-PR scheme always outperforms the AFEC scheme. Note that the link PLR level for each receiver is usually different in practical systems. Therefore, in many real scenarios, the

HEC-PR can also perform better than the AFEC scheme. Let's take an example for explaining this point in the following. It is assumed that there is only one receiver with link PLR of 10^{-1} while the average link PLR is only about 0.04 for the case with N_{recv}=7 and R_d=4Mbps. From Figure 7.5, we can see that the total needed RI of the AFEC scheme will be about 45%, since the AFEC scheme must adopt the worst case for designing the optimum FEC code. When applying the proposed HEC-PR scheme for this scenario, however, the minimum total needed RI will be about 25%, which is much less than that of AFEC scheme. Obviously, the average link PLR is usually less than the worst one, which indicates that the HEC-PR scheme can actually perform better than the AFEC scheme in many real scenarios. There are mainly two reasons for explaining this: One is that the AFEC scheme must adopt the worst case for designing parameters; the other is that the length of k_{lim} the FEC code limits its performance, since the short length of k_{lim} has to be adopted for satisfying the strict delay constraint.

However, if both the average link PLR and the number of receivers are high enough, the AFEC scheme will perform better than the HEC-PR scheme. For instance, in the case with N_{recv}=7 and R_d=12Mbps as shown in Figure 7.5, when the link PLR for each receiver is about 10^{-1}, the minimum total needed RI of the HEC scheme will be about 80% while that of AFEC scheme is less than 30%, which indicates that the AFEC scheme for these cases with high average link PLR and large group size can be considered.

7.6.3. The Effect of CC of GE Model

As for the AFEC scheme, for studying the effect of ρ to the HEC-PR scheme, we now calculate the total needed RI for one receiver with different ρ, where the parameters of the GE channel are also produced by the method introduced in Chapter 3.5. The performance of the HEC-PR scheme is shown in Figure 7.6.

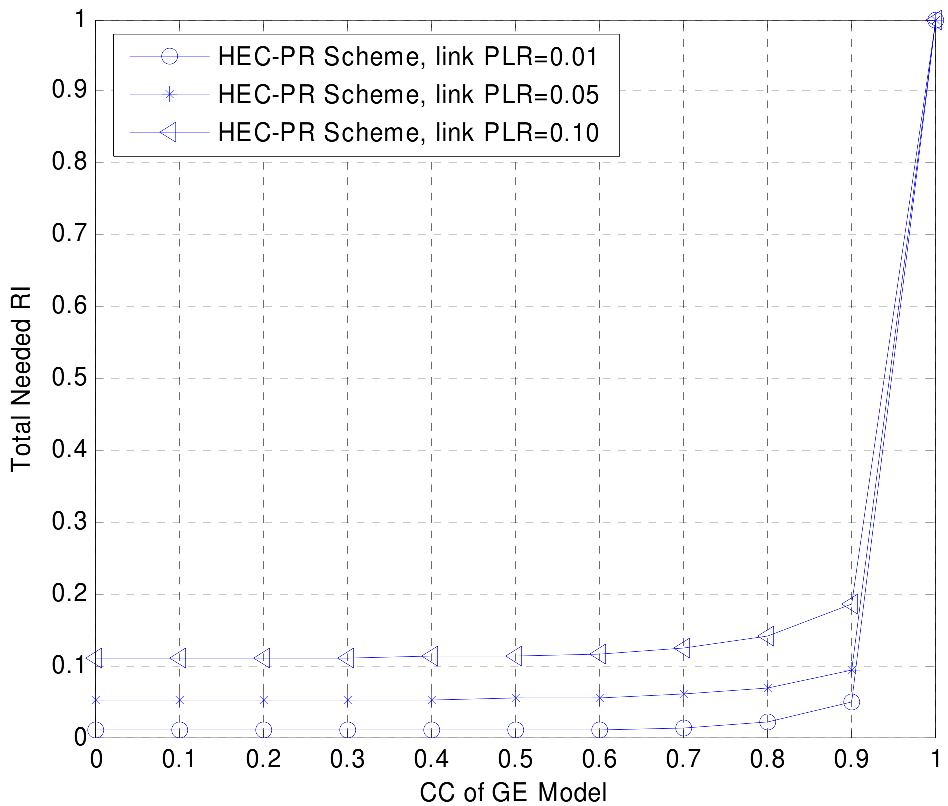


Figure 7.6. The Total Needed RI for the HEC-PR Scheme with different ρ

As shown in Figure 7.6, it is found that the parameter ρ has almost no effect on the HEC-PR scheme in case of ρ<0.7, which means that the HEC-PR scheme with even high CC in GE channel can perform always as well as that with ρ=0. The reason is clear: On the one hand, unlike the AFEC scheme as introduced in Chapter 3, the HEC-PR scheme does not employ encoding or decoding technique in the form of block FEC coding, which results in that the variable average error-burst length in GE channel does not influence the HEC-PR scheme so much as in the AFEC scheme. On the other hand, according to our measurements results in real systems as presented in Chapter 3.5, the ρ is usually very small in real systems (e.g. ρ <0.1), the probability of the error-burst length of large value (e.g. more than 10) is quite small so that in most cases the HEC-PR scheme actually can work perfectly under strict delay constraints. Therefore, when the CC of the GE channel is small, the average error-burst length has no effect on the performance of the HEC-PR scheme. Especially, in case of $\rho=1$, the state of the GE channel will only depend on the initial state of the channel so that the HEC-PR scheme fails in this case. As a result, in real systems with small values on the CC of GE channel, the effect of CC to the HEC-PR scheme can be neglected.

7.7. Summary

In this chapter, a Hybrid Error Correction scheme with Packet Repetition techniques (HEC-PR) is proposed to guarantee the target PLR requirement under strict delay constraints with minimum needed RI. The analysis results show that the HEC-PR scheme can work very well in the multicast scenarios with small group size and small average link PLR due to its efficiency and simplicity. Therefore, the HEC-PR scheme is a good candidate technique for the cases with small group size and small link PLR (e.g. traditional DVB services over Wireless Home Networks). However, the needed RI of the HEC-PR scheme will increase linearly with the increase of the group size so that it is not suitable for those multicast scenarios with large group size (e.g. the wide-area wireless networks such as WiMAX, 3GPP etc.). Actually, for multimedia multicast services, many studies have shown that the traditional Type I Hybrid ARQ and Type II Hybrid ARQ scheme outperform those schemes based on pure ARQ technique or FEC alone. Based on those findings, by proposing a general architecture integrating all of the important existing EER techniques, we will study the influence of the multicast and strict delay constraints to the general architecture of EER in the following chapters. Also, we will present how to analyze and optimize the performance of the general architecture of EER.

Chapter 8
The General Mathematical Framework

In this chapter, based on the application of GE channel model for the general architecture of EER introduced in Chapter 5.3, we will present how to analyze the PLR and RI performance of the general architecture of EER, which is proposed in Chapter 2.3. Theoretically, we also can design the parameters of the general architecture (see Table 2.1) for each receiver independently as similar as for HEC-PR schemes proposed in Chapter 7. However, it is very hard to implement in practical systems if different FEC codes are used for different receivers in the first transmission. To simplify the practical implementation, therefore, it is assumed that the general architecture will adopt the same parameters for each receiver in this thesis. However, by dividing a multicast scenario into multiple parallel independent unicast scenarios, it is possible to extend the analysis results in this thesis for designing different parameters for different receivers. For the convenience of description in the following analysis, two additional random variables are defined:

- $\mathrm{I}_k(j,w)$: a random variable representing the number of source data packets lost in one encoding block of k source data packets after the *j-th* receiver experiences w retransmission rounds, where $1 \le w \le N_{rr,max}$;
- $\mathrm{N}_{req}(j,w)$: a random variable representing the number of redundant packets required by the *j-th* receiver in the w-th retransmission round, which is for recovering all of the k data packets in one block, and where $0 \le \mathrm{N}_{req}(j,w) \le k$ and $1 \le w \le N_{rr,max}$.

Obviously, in the w-th retransmission round, in order to recover all of the missing data packets for each receiver that received fewer than k packets for one block, at least $\mathrm{N}_{req,\max}^{w}$ redundant packets need to be retransmitted at the sender:

$$\mathrm{N}_{rea,\max}^{w} = \max(\mathrm{N}_{req}(1,w), \mathrm{N}_{req}(2,w), \ldots, \mathrm{N}_{req}(N_{recv}, w)) \tag{8.1}$$

Note that $\mathrm{N}_{req,\max}^{w}$ is also a random variable. Since it is assumed that the feedback channel is error-free, the random variable $\mathrm{N}_{req,\max}^{w}$ always reflects the true maximum number of redundant packets required for one block by the worst receiver.

8.1. PLR Performance

This section proposes a mathematical framework (MF) for analyzing the PLR performance of the general architecture of EER. The basic idea of the MF is to compute the average number of lost data packets in one block of k source data packets for the general architecture. To achieve this MF, we firstly need to analyze the probability of decoding failure for one block with the parameters of the architecture defined in Table 2.1. Depending on the decoding failure probability, we finally can obtain the average number of lost data packets within one block of k source data packets in each transmission stage. Following this idea, we will present how to compute the PLR performance of the general architecture in the following sections.

8.1.1. Probability of Decoding Failure

In this section, we present how to calculate the probability of decoding failure for one receiver (without loss of generality, it is assumed to be the j-th receiver) in each retransmission stage. Before computing this probability for the j-th receiver, we need to know the loss probability of each redundant packet in each retransmission round. We thus define two probabilities as follow: Let $P_B(w, j)$ be the loss probability of each redundant packet in the w-th retransmission round for the j-th receiver; and let $P_{recv}(r,s,w,j)$ be the probability of r packets received in case of s different redundant packets retransmitted in the w-th retransmission round for the j-th receiver. As analyzed in Chapter 5.3, the loss probability of each parity packet during the w-th retransmission round is $P_{B|B}[w\cdot T_{lp}(j)]$, which can be calculated by (7.2). Note, even in case of k=1, all of the copies of the retransmission data packets can also be viewed as parity packets, because each of them can be used for recovering the only data packet in one block. Therefore, similar to (7.2), the value of $P_B(w,j)$ can be computed by:

$$P_B(w,j) = P_{B|B}[w\cdot T_{lp}(j)] = P_B(j) + (1-P_B(j))\cdot \rho(j)^{w\cdot T_{lp}(j)} \tag{8.2}$$

Upon (8.2), therefore, $P_{recv}(r,s,w,j)$ is given by:

$$P_{recv}(r,s,w,j) = \binom{s}{r}(1-P_B(w,j))^r (P_B(w,j))^{s-r} \tag{8.3}$$

Now recalling the EER mechanism of the general architecture described in Chapter 2.3, the sender will multicast $k+N_p$ packets to all of the receivers in the first transmission. As a matter of fact, the decoding failure for the j-th receiver only occurs possibly in the condition of at least N_p+1 packets lost in the first transmission. In the following, for the convenience of description, let the symbol b denote the number of lost packets in the first transmission for the j-th receiver, where $k+N_p \geq b > N_p$. Additionally, for the first retransmission round, let the symbol s denote the maximum number of redundant packets required by all of the receivers, i.e. $\mathrm{N}^1_{req,\max} = s$, obviously where we have $k \geq s \geq b-N_p$. According to the general architecture described in Chapter 2.3, then, the sender will multicast $N^1_{cc}\cdot s$ additional redundant packets to all of the receivers in the first retransmission round. For the j-th receiver in the first retransmission round, the decoding failure will occur in case that the number of received redundant packets (which is denoted by r^1) is less than $b-N_p$, which probability is $P_{recv}(r^1, s, 1, j)$. In the condition of b packets lost in the first transmission for the j-th receiver and s redundant packet required for the first retransmission round, using (8.3), the total probability of decoding failure in the first retransmission stage for the j-th receiver (which is denoted by $P_f(s, b, 1, j)$ in this thesis) can be obtained directly by:

$$P_f(s, b, 1, j) = \sum_{r^1=0}^{b-N_p-1} P_{recv}(r^1, s, 1, j) \tag{8.4}$$

Afterwards, let's focus on the probability of decoding failure in the w-th retransmission round for the j-th receiver with w>1. In the following, let r^w be the number of redundant packets received in the w-th retransmission round by the j-th receiver; let $r^w_{\max}$ be the maximum possible number of redundant packets received in the w-th retransmission round by the j-th receiver if decoding failure occurs. Note, it is still a challenge to compute the accurate PDF of $\mathrm{N}^w_{req,\max}$ with w>1 due to the combinations of NACKs from all of the receivers. Instead of deriving the accurate MF for the general architecture, therefore, we derive the upper-band of its performance by assuming that the $\mathrm{N}^w_{req,\max}$ only depends on the value of $\mathrm{N}_{req}(j, w)$ in case of w>1. Under this assumption, we have the following theorem on calculating the upper-band of probability of decoding failure for the j-th receiver:

***Theorem* 1**: *Suppose that there are b (where k+ $N_p \geq b > N_p$) packets lost in one encoding block of k+N_p packets in the first transmission for the j-th receiver and* $\mathrm{N}^1_{req,\max} = s$ *(where k≥s≥b-N_p); let $P_{f,upper}$(b,s,w,j) be the upper-band of the probability of decoding failure for the j-th receiver after it experiences w retransmission rounds. For w=2,3,…, the upper-band of the probability can be recursively calculated by:*

$$P_{f,upper}(b,s,w,j) \le P_{f,upper}(b,s,w-1,j)\left(\sum_{r^w=0}^{r^w_{\max}} P'_{recv}(r^w, N^w_{cc} \cdot \mathrm{N}_{req}(j,w), w, j)\right) \tag{8.5}$$

Where:

$$\mathrm{N}_{req}(j,w) = \begin{cases} 0, b - N_p \le \sum_{q=1}^{w-1} r^q \\ b - N_p - \sum_{q=1}^{w-1} r^q, otherwise \end{cases}$$

$$r^w_{\max} = \max\{0, \mathrm{N}_{req}(j,w) - 1\}$$

$$P'_{recv}(r^w, N^w_{cc} \cdot \mathrm{N}_{req}(j,w), w, j) = \begin{cases} 0, \mathrm{N}_{req}(j,w) = 0 \\ P_{recv}(r^w, N^w_{cc} \cdot \mathrm{N}_{req}(j,w), w, j), \text{otherwise} \end{cases}$$

Initialization:

$$P_{f,upper}(b,s,1,j) = P_f(s,b,1,j) = \sum_{r^1=0}^{r^1_{\max}} P_{recv}(r^1, s, 1, j) \tag{8.6}$$

Where:

$$r^1_{\max} = b - N_p - 1$$

Proof: For the first retransmission round, as shown in (8.6), the result is straightforward and hence the proof is omitted. For the remaining retransmission rounds (i.e. w>1), as a matter of fact, the value of $\mathrm{N}_{req}(j,w)$ depends on the value of $b - N_p - \sum_{q=1}^{w-1} r^q$. Note, in case of $\mathrm{N}_{req}(j,w) = 0$, that it indicates that the number of redundant packets received during the previous w-1 retransmission rounds by the j-th receiver is equal to or more than the number of redundant packets required by this receiver, which means the j-th receiver has received redundant packets enough for recovering all of the missing data packets so that the probability of the decoding failure is zero. Otherwise, according to the general architecture, the sender will retransmit $N^w_{cc} \cdot \mathrm{N}^w_{req,\max}$ redundant packets in the w-th retransmission round due to the combination of all of the NACKs from overall receivers. Note that the probability $P_{recv}(r^w, N^w_{cc} \cdot \mathrm{N}_{req}(j,w), w, j)$ is always no more than the probability $P_{recv}(r^w, N_{cc}^w \cdot \mathrm{N}_{req,\max}(w), w, j)$ due to $\mathrm{N}_{req,\max}(w) \geq \mathrm{N}_{req}(j,w)$. As a result, by combining the probability $P_{recv}(r^w, N_{cc}^w \cdot \mathrm{N}_{req}(j,w), w, j)$ with the probability of decoding failure during all of the previous *w*-1 retransmission rounds (i.e. $P_{f,upper}(b,s,w-1,j)$) for the *j*-th receiver, we can obtain the upper-band of the probability $P_{f,upper}(b,s,w,j)$ immediately. The proof of the theorem is completed.

Remarks: The $P_{f,upper}(b,s,w,j)$ is actually a very tight upper-band of the probability of decoding failure for the *j*-th receiver with *w* retransmission rounds, because the calculation for this probability in the first retransmission round is accurate and the probabilities of the decoding failure in the remaining retransmission rounds is very small. Therefore, this upper-band can be used for evaluating a roughly accurate PLR performance for the *j*-th receiver with the general architecture. More importantly, this upper-band provides a good margin for keeping the final PLR below a certain PLR level so that a fixed target PLR requirement can be guaranteed very well.

8.1.2. Upper-band of PLR Performance

In the following, let's focus on how to evaluate the PLR performance for the *j*-th receiver with the general architecture of EER. To evaluate its PLR performance, we need to calculate the expected value of the number of missing data packets in one encoding block of *k* source data packets after the receiver experiences *w* (where $0 \leq w \leq N_{rr,max}$) retransmission rounds. In other words, the PLR performance can be calculated by $E(\mathrm{I}_k(j,w))/k$, which is the final PLR at the *j-th* receiver after it experiences *w* retransmission rounds.

Same to the definition in Chapter 6.1, first, let *P*(*m,d,*CSI) be the probability of *m* packets lost in a sequence of *d* packets in GE channel with CSI, which can be calculated by (6.1). Afterwards, we introduce two useful probabilities based on *P*(*m,d,*CSI): One is the PDF of $\mathrm{N}_{req,\max}^1$ (i.e. $\Pr(\mathrm{N}_{req,\max}^1 = i)$), which is denoted by $P_{N_{req,\max}}^i$; the other is the probability of

$\mathrm{N}^1_{req,\max}$ of i in the condition of $\mathrm{N}_{req}(j,1)$ of c (i.e. $\Pr(\mathrm{N}^1_{req,\max}=i \mid \mathrm{N}_{req}(j,1)=c)$), which is denoted by $P_{req}(i,c,j)$. The detail derivations on these two important probabilities are attached in Appendix A and Appendix B, respectively.

Then, let $P_d(i,b)$ denote the probability of i data packets lost in case of b packets lost in one encoding block of $k+N_p$ packets in the first transmission. This probability is same to the definition in Chapter 6.1 and can be calculated by (6.7).

Relying on the three important probabilities introduced above: $P_d(i,b)$, $P_{req}(i,c,j)$ and $P_{f,upper}(b,s,w,j)$, we now have the following lemma on evaluating the upper-band of the PDF of $\mathrm{I}_k(j,w)$:

***Lemma* 1**: *Based on* ***Theorem* 1**, *using $P_d(i,b)$ and $P_{req}(i,c,j)$, the PDF of $\mathrm{I}_k(j,w)$ will satisfy:*

$$\Pr(\mathrm{I}_k(j,w)=i) \le \sum_{b=\max(N_p+1,i)}^{N_p+i} \sum_{s=b-N_p}^{k} P_d(i,b) P_{req}(s,b-N_p,j) P_{f,upper}(b,s,w,j) \tag{8.7}$$

Where i=1,2,…,k

Proof: First, the value of $\mathrm{I}_k(j,w)$ being i (where $1 \le i \le k$) means that the number of data packets lost is i in one block of $k+N_p$ packets for the j-th receiver in the first transmission, and the receiver also does not receive enough redundant packets for recovering these i data packets within w retransmission rounds. Secondly, as a matter of fact, it indicates that there are b (where $\max(N_p+1,\ i) \le b \le N_p+i$) packets lost in the block of $k+N_p$ packets for the j-th receiver in the first transmission, in which the probability of there being i data packets lost can be computed by $P_d(i,b)$. According to the general architecture, then, the receiver will require b-N_p redundant packets for retransmission at the sender for recovering the missing i data packets. However, at the same time, the sender will possibly send $N^1_{cc} \cdot N^1_{rt} \cdot s$ (where b-$N_p \le s \le k$) redundant packets due to the combination of all of the NACKs from overall receivers, which probability is exactly $P_{req}(s, b-N_p, j)$. Furthermore, based on the ***Theorem* 1**, in case of b packets lost in the first transmission and maximum s redundant packets required for the first retransmission, the upper-band of the probability of the decoding failure for the j-th receiver within w retransmission rounds can be calculated by $P_{f,upper}(b, s, w, j)$. As a result, combining all of those probabilities as analyzed above, we can obtain the upper-band of the probability of $\mathrm{I}_k(j, w)$ being i as shown in (8.7) immediately. The proof of the lemma is completed.

Depending on ***Lemma* 1**, finally, we have the following theorem on evaluating the upper-band of final PLR for the j-th receiver with $N_{rr,max}$ retransmission rounds:

Theorem* 2: After the j-th receiver experiences $N_{rr,max}$ retransmission rounds, the final PLR at this receiver will satisfy:*

$$PLR(j, N_{rr,\max}) = \frac{E(\mathrm{I}_k(j, N_{rr,\max}))}{k} = \frac{\sum_{i=1}^{k} i \cdot \Pr(\mathrm{I}_k(j, N_{rr,\max}) = i)}{k}$$
$$\le \frac{1}{k} \sum_{i=1}^{k} \sum_{b=\max(N_p+1,i)}^{N_p+i} \sum_{s=b-N_p}^{k} i \cdot P_d(i,b) P_{req}(s, b-N_p, j) P_{f,upper}(N_{rr,\max}, b, s, j) \tag{8.8}$$

This theorem can be achieved from Lemma 1 directly; hence its proof is omitted.

For simplifying the description, we define additional three vectors as following: $\vec{N}_{cc} = \{N_{cc}^q | 1 \le q \le N_{rr,\max}\}$ and $\vec{C}_{CSI} = \{CSI(j) | 1 \le j \le N_{recv}\}$. By looking into (8.8), we can find that the upper band of the PLR performance of the architecture for the j-th receiver is actually a function of: k, N_p, $\vec{N}_{cc}$, N_{recv} and $\vec{C}_{CSI}$, which is denoted by $f_{PLR,UP}^{j}(k, N_p, \vec{N}_{cc}, N_{recv}, \vec{C}_{CSI})$ in this thesis.

8.2. RI Performance

This chapter develops a MF for calculating the total needed RI with the general architecture of EER, which includes two parts: One is the common part for all of the receivers in the first transmission, which is N_p/k; the other is the part in the retransmissions, which is caused by the retransmissions of redundant packets for all of the receivers. Note, in case of $\mathrm{N}_{req,\max}^{w} = i$, that the sender will transmit total $i \cdot N_{cc}^{w}$ redundant packets in the w-th retransmission round. Now let $P_{\mathrm{N}_{req,\max}}^{i}(w)$ be the PDF of $\mathrm{N}_{req,\max}^{w}$ (i.e. $P_{\mathrm{N}_{req,\max}}^{i}(w) = \Pr(\mathrm{N}_{req,\max}^{w} = i)$), where the accurate PDF of $\mathrm{N}_{req,\max}^{1}$ can be found in Appendix A and the derivation of the PDF of $\mathrm{N}_{req,\max}^{w}$ with w>1 can be found in Appendix C. Upon the PDF of $\mathrm{N}_{req,\max}^{w}$, then, we have the following theorem regarding the MF of computing the total needed RI for the general architecture:

***Theorem* 3**: *Suppose that the general architecture with the parameters of k, N_p, $\vec{N}_{cc} = \{N_{cc}^{q} | 1 \le q \le N_{rr,\max}\}$ for a multicast scenario with N_{recv} receivers of $\vec{C}_{CSI} = \{CSI(j) | 1 \le j \le N_{recv}\}$, after all of the receivers experience $N_{rr,max}$ retransmission rounds, the total needed* RI *will be*:

$$
\begin{aligned}
RI &= \frac{N_p}{k} + \frac{1}{k}\sum_{w=1}^{N_{rr,\max}} N_{cc}^{w} \cdot E(\mathrm{N}_{req,\max}^{w}) \\
&= \frac{N_p}{k} + \frac{1}{k}\sum_{w=1}^{N_{rr,\max}} N_{cc}^{w} \cdot \left(\sum_{i=1}^{k} i \cdot P_{\mathrm{N}_{req,\max}}^{i}(w)\right) \\
&= \frac{N_p}{k} + \frac{1}{k}\sum_{w=1}^{N_{rr,\max}} \sum_{i=1}^{k} i \cdot N_{cc}^{w} \cdot P_{\mathrm{N}_{req,\max}}^{i}(w)
\end{aligned}
\tag{8.9}
$$

Proof: As mentioned above, we can divide the total needed RI into two parts: one is the common part for all of the receivers in the first transmission, which is N_p/k; the other is the part in the retransmissions. For the second part in the transmission rounds, we know that in the w-th retransmission round the probability of the maximum number of redundant packets required of i (where $1 \le i \le k$) for one encoding block is $P^i_{\mathrm{N}_{req,\max}}(w)$. Additionally, according to the architecture, the sender will retransmit total $N^w_{cc} \cdot i$ redundant packets for recovering those missing data packets for each receiver in the w-th retransmission round. Therefore, in the w-th retransmission round, the expected value of the total number of redundant packets retransmitted for one encoding block is exactly $N^w_{cc} \cdot E(\mathrm{N}^w_{req,\max})$ (where $E(\mathrm{N}^w_{req,\max}) = \sum_{i=1}^{k} i \cdot P^i_{\mathrm{N}_{req,\max}}(w)$), which results in RI of $N^w_{cc} \cdot E(\mathrm{N}^w_{req,\max}) / k$ needed. Combining the needed RI in the first transmission and all of the $N_{rr,max}$ retransmission rounds, we can obtain (8.9) immediately. This proves the theorem 3.

By looking into (8.9), we can find that the total needed RI of the general architecture actually is also a function of: k, N_p, $\vec{N}_{cc}$, N_{recv} and $\vec{C}_{CSI}$, which is denoted by: $f_{RI}(k, N_p, \vec{N}_{cc}, N_{recv}, \vec{C}_{CSI})$ in this thesis.

8.3. Summary

This chapter develops a general Mathematical Framework (MF) to analyze the performance of the general architecture of EER proposed in Chapter 2.3. Given the parameters of the general architecture for a multicast scenario, the final PLR at the j-th receiver can be evaluated by $f^j_{PLR,UP}(k, N_p, \vec{N}_{cc}, N_{recv}, \vec{C}_{CSI})$ proposed in Chapter 8.1, in which the first three parameters are the parameters of the general architecture and the last two parameters are the parameters of the current multicast scenario. Actually, it is a very tight upper-band of the PLR performance of the general architecture. The more accurate PLR performance on the general architecture needs further study. Nevertheless, this upper-band is accurate enough to achieve the optimization results on optimizing the general architecture in this thesis due to its tight property. Secondly, this chapter also investigates a formula to analyze the total needed RI of the general architecture, i.e. $f_{RI}(k, N_p, \vec{N}_{cc}, N_{recv}, \vec{C}_{CSI})$ proposed in Chapter 8.2, which has the same parameters to its PLR performance. As a result, base on this proposed general MF of analyzing the PLR and RI performance for the general architecture, we then can search out its optimum parameters for any given multicast scenario. In the following chapter, we will present how to optimize the parameters of the general architecture for a certain multicast scenario using the general MF, and then propose an efficient algorithm to achieve the optimal parameters.

Chapter 9
Optimization Problem

Based on the general mathematic framework proposed in Chapter 8, in this Chapter, we will present how to optimize the parameters of the general architecture of EER proposed in Chapter 2.3. Note that all of the receivers share identical parameters for this general architecture. Therefore, if the architecture can guarantee the QoS requirements for the worst receiver, it can also guarantee the same QoS requirements for every receiver in the multicasting scenario. Without loss of generality, it is assumed the first receiver is the one with the worst situation in a multicast scenario. In other words, the first receiver has the largest link PLR and the largest RTT. Our remaining task is to design optimum parameters for the general architecture, which will satisfy a certain target PLR requirement for the first receiver under strict delay constraints with the minimum total needed RI.

9.1. Optimization Problem

First of all, from the end-to-end delay budget analyzed in Chapter 4.3, we can know that the delay requirement will limit the number of data packets in one coding block and the number of retransmission rounds. In the following, upon the end-to-end delay budget for the general architecture, we will derive the boundary for the parameters $N_{rr,max}$ and k based on the strict delay constraints. For those retransmission packets in the first receiver, according to (4.3), the maximum possible end-to-end delay must satisfy:

$$\frac{RTT(1)}{2}+(k-1)\cdot t_s+t_d+N_{rr,\max}\cdot(t_{lp}(1)+t_{rp})\leq \mathrm{D}_{\text{target}} \tag{9.1}$$

Because the value of k is at least 1 for the general architecture, the maximum allowable number of retransmission rounds is constrained by:

$$\hat{N}_{rr,\max}=\left\lfloor \frac{\mathrm{D}_{\text{target}}-\frac{RTT(1)}{2}-t_d}{t_{lp}(1)+t_{rp}} \right\rfloor \tag{9.2}$$

Therefore, for the general architecture, the parameter $N_{rr,max}$ will be limited in the range of between zero and $\hat{N}_{rr,\max}$. Then, let $k(w)$ denote the function of calculating the length of k of the general architecture with w retransmission. By solving (9.1), this function can be expressed as:

$$k(w)=\left\lfloor \frac{\mathrm{D}_{\text{target}}-w\cdot\left(t_{lp}(1)+t_{rp}\right)-\frac{RTT(1)}{2}-t_d}{t_s} \right\rfloor+1 \tag{9.3}$$

Where $0\leq w\leq \hat{N}_{rr,\max}$

Note that in (9.3) the parameter k will only rely on the parameter w if $t_{lp}(1)$, t_{rp}, t_s, t_d, $RTT(1)$ and $\mathrm{D}_{\text{target}}$ are fixed. Note, given a certain QoS requirements for a multicast scenario, the system parameters mentioned above actually are fixed and measurable. Therefore, the length of k will only depend on the parameter w for a certain multicast scenario under predefined QoS requirements.

Finally, depending on the general mathematical framework of calculating the PLR and RI performance proposed in Chapter 8, using (9.2) and (9.3) , our optimization problem can be summarized as the following form:

$$RI_{AHEC,opt} = \arg\min f_{RI}\left(k\left(N_{rr,\max}\right), N_p, \vec{N}_{cc}, N_{recv}, \vec{C}_{CSI}\right) \tag{9.4}$$

Subject to:

$$0 \le N_{rr,\max} \le \hat{N}_{rr,\max}$$

$$f^{1}_{PLR,UP}\left(k\left(N_{rr,\max}\right), N_p, \vec{N}_{cc}, N_{recv}, \vec{C}_{CSI}\right) \le \mathrm{PLR}_{\mathrm{target}}$$

Remarks: If k is set to one, the general architecture acts as similar as the HEC-PR scheme presented in Chapter 7; but here the sender is more intelligent than that in Chapter 7, because it can suppress those repeat retransmission requirements by different receivers. If k is set to more than one and N_p is set to more than zero, the general architecture acts as the traditional Type I HARQ scheme. If k is set to more than one and N_p is set to zero, the architecture acts as the traditional Type II HARQ scheme. Note that in the traditional Type I and Type II HARQ schemes, multiple copies of a parity packet only can be counted as one redundant packet at the receivers. For the consideration of efficiency, we only adopt new parity packets instead of multiple copies of redundant packets at each retransmission round. As a result, the architecture can choose the best scheme automatically among the HEC-PR scheme, traditional Type I and Type II HARQ scheme by solving (9.4).

Obviously, using a full search algorithm, we can obtain the optimal parameters for the architecture: k, N_p and $\vec{N}_{cc}$. In many cases, however, the size of the full search space will be very large so that it is very hard to achieve the optimum parameters by the full search algorithm. In the following, therefore, we will present an efficient greedy algorithm on searching for the optimum parameters of the general architecture.

9.2. Greedy Algorithm

In this chapter, we propose an efficient greedy algorithm to solve the optimization problem introduced above. First, we need to calculate $\hat{N}_{rr,\max}$ by (9.2). Notice that only FEC scheme can be adopted for the architecture with $\hat{N}_{rr,\max} < 1$. In the following, it is assumed that

$\hat{N}_{rr,\max} \geq 1$ for the architecture to present the greedy algorithm. Apparently, the parameter $N_{rr,max}$ will be limited in the range of between zero and $\hat{N}_{rr,\max}$ (i.e. $0 \leq N_{rr,max} \leq \hat{N}_{rr,\max}$). Since the parameter k is limited by the parameter $N_{rr,max}$, therefore, the maximum possible length for the parameter k for each possible value of $N_{rr,max}$ is given by $k(N_{rr,\max})$ according to (9.3). Moreover, to limit the size of the search space, we can set an upper band for the parameter N_p for each possible value of $N_{rr,max}$ as $N_{p,max}(N_{rr,max}) = k(N_{rr,max}) \cdot RI_{lim}$, where RI_{lim} should be set to far more than the worst PLR among all of the receivers to guarantee the optimum parameters included in the search space.

Using $\hat{N}_{rr,\max}$ and $N_{p,max}(N_{rr,max})$, we can find out the optimum $\vec{N}_{cc}$ for each possible combination of $N_{rr,max}$ and N_p. Afterwards, we can obtain the final global optimum parameters. This task can be carried out by an efficient greedy algorithm: For each combination of $N_{rr,max}$ and N_p, the algorithm will always increase the minimum RI at each stage until the target PLR requirement being satisfied by increasing the total number for the $\vec{N}_{cc}$. For the convenience of description, we define the operation of adding one for $\vec{N}_{cc}$ in the w-th retransmission round as:

$$f_{+1}(\vec{N}_{cc}, w) = \{N_{cc}^{1}, N_{cc}^{2}, \ldots, N_{cc}^{w-1}, N_{cc}^{w}+1, N_{cc}^{w+1}, \ldots, N_{cc}^{N_{rr,\max}}\} \tag{9.5}$$

As a result, for any case, we only need to increase the total needed RI by increasing the total number of $\vec{N}_{cc}$ until the target PLR requirement satisfied.

Intuitively, the increased RI at each stage should be the minimal by the operation of adding one for $\vec{N}_{cc}$ in the last retransmission round (i.e. $f_{+1}(\vec{N}_{cc}, N_{rr,\max})$). This result can be explained by the following fact: For any receiver, the number of redundant packets required in the last retransmission round is always no more than that in any previous retransmission round. Therefore, we can obtain the locally optimum results by the operation of $f_{+1}(\vec{N}_{cc}, N_{rr,\max})$ at each stage for each combination of $N_{rr,max}$ and N_p, which results in a greedy algorithm with the hope of finding out the global optimum parameters. The details on the greedy algorithm are as follows:

Table 9.1: Greedy Algorithm

```
set N_{rr,max}=0; N_p=0;
PLR = f^1_{PLR,UP}(k(0), N_p, 0, N_recv, C_CSI);

while PLR>PLR_target do
   N_p= N_p +1;
   PLR = f^1_{PLR,UP}(k(0), N_p, 0, N_recv, C_CSI);
endwhile

k_opt=k(0); N_{p,opt}=N_p; N_{cc,opt} = 0;
RI_opt = f_RI(k(0), N_{p,opt}, 0, N_recv, C_CSI);

for N_{rr,max}=1 to N̂_{rr,max}
   k_lim= {1, k(N_{rr,max})};
   for N_p=0 to N_{p,max}(N_{rr,max})
      for i=1 to 2
         N_cc = {N^w_cc = 1 | 1 ≤ w ≤ N_{rr,max}};
         PLR = f^1_{PLR,UP}(k_lim(i), N_p, N_cc, N_recv, C_CSI);

         while PLR>PLR_target do
            N_cc = f_{+1}(N_cc, N_{rr,max});
            PLR = f^1_{PLR,UP}(k_lim(i), N_p, N_cc, N_recv, C_CSI);
         endwhile

         RI= f_RI(k_lim(i), N_p, N_cc, N_recv, C_CSI);

         if (RI<RI_opt)
            RI_opt=RI; k_opt=k_lim(i); N_{p,opt}=N_p; N_{cc,opt} = N_cc;
         endif

      endfor
   endfor
endfor
```

Where vectors are marked: $\vec{N}_{cc}$, $\vec{N}_{cc,opt}$, $\vec{C}_{CSI}$.

Based on Table 9.1, now we have the following assertion:

***Assertion**: In the algorithm as shown in* Table 9.1*, the increased RI at each stage is always minimal by the operation of* $f_{+1}(\vec{N}_{cc}, N_{rr,\max})$ *for each combination of* $N_{rr,max}$ *and* N_p*. Therefore, this algorithm is a greedy algorithm, which always makes the locally optimum choice at each stage with the hope of finding out the global optimum parameters.*

The complete proof on this assertion can be found in Appendix D. Note, to compare with the HEC-PR scheme proposed in Chapter 7 fairly, we should constrainedly set the parameter k to one for each possible value of $N_{rr,max}$ to see if it is the optimum scheme. As show Table 9.1, the value of the element of more than one in $\vec{N}_{cc}$ only exists possibly in the last retransmission round. Depending on the greedy algorithm above, finally, we can obtain the optimum parameters (i.e. k_{opt}, $N_{p,opt}$ and $\vec{N}_{cc,opt}$) for the general architecture.

9.3. Advantages of the Greedy Algorithm

In the following, let's take an example to demonstrate the advantages of the proposed greedy algorithm. Considering a practical scenario with following parameters: D_{target}=100ms, RTT=4.0ms, t_s=1.25ms (e.g. multimedia data rate of 8Mbps and packet size of 1250bytes) and $t_{rw}+ t_{sw}$ =4ms. Similar to the analysis in Chapter 7.6, we also assume that $t_{rp}=t_s$ for the transmission time of retransmission packets and use t_s instead of t_d in the end-to-end delay budget. Based on these parameters, then, we can obtain the parameter $\hat{N}_{rr,\max}$ by (9.2), which is 10. According the optimization problem as shown in (9.4), we need to search for the optimum $\vec{N}_{cc}$ for each possible value of $N_{rr,amx}$ in the range of from 0 to 10. Because we have no any knowledge on the range of the optimum elements for the $\vec{N}_{cc}$, we have to predefine a big enough upper band (which is denoted by M here) for each element of the $\vec{N}_{cc}$ to guarantee the optimum parameters included in the search space. Now we analyze the size of search space by *full search* and the *greedy algorithm* for this case, respectively:

1. *Full Search Algorithm*: Note, under certain $N_{rr,amx}$ and M, the number of possible values on $\vec{N}_{cc}$ is $M^{N_{rr,\max}}$. Hence, the size of the full search space for this case will be $\sum_{i=0}^{\hat{N}_{rr,\max}} M^i$. In case of M=10, the size of the full search space is exactly $\sum_{i=0}^{10} 10^i$. Furthermore, because the calculations of the RI and PLR performance are also complicated, it will be very hard to obtain the optimum $\vec{N}_{cc}$ for this case by full search.
2. *Greedy Algorithm*: Under certain $N_{rr,max}$, since only the last element of $\vec{N}_{cc}$ is variable, we thus need no more than M steps to obtain the optimal $\vec{N}_{cc}$. In case of M =10 and $N_{rr,max}$=10, the size of the search space by the greedy algorithm is only 10 while the size of the full search space is 10^{10}. Therefore, the greedy algorithm is much faster than the full search algorithm.

Finally, we must point out that the optimum parameters achieved by the greedy algorithm are indeed the global optimum results in most cases. The reason is: For each receiver, since the total number of redundant packets required in the last retransmission round is actually much less than that in any previous retransmission stage, the total needed RI with the $\vec{N}_{cc} = \{1,1,...,1,2,1\}$ is usually far more than that with the $\vec{N}_{cc} = \{1,1,...,1,1,M\}$ of small M. Therefore, through the proposed greedy algorithm, the final optimum parameters with the form of $\vec{N}_{cc} = \{1,1,...,1,1,M\}$ are usually the real global optimum parameters.

9.4. Summary

Making use of the Mathematical Framework (MF) presented in Chapter 8, we discussed how to optimize the parameters of the general architecture of EER in this chapter. The optimization target is to minimize the total needed RI with guaranteeing a certain target PLR requirement under strict delay constraints. To achieve this target, we build up an optimization problem using the MF of calculating the PLR and RI performance of the general architecture. By solving this optimization problem, we then can obtain the optimum parameters the general architecture. Generally speaking, we always can solve the optimization problem by full search algorithms. However, for many multicast scenarios, the search space will be very large due to the large number of free parameters so that it is very difficult to obtain the optimum results by full search algorithms. In order to overcome this shortage, we also proposed an efficient greedy algorithm to solve the optimization problem in this chapter. In the following chapter, we will analyze the performance of the optimum architecture of EER using the proposed greedy algorithm.

Chapter 10

Analysis Results

In this chapter, first, we will analyze the optimization results for the architecture using the proposed greedy algorithm in Chapter 9.2; and then compare it with the AFEC scheme presented in Chapter 6 and the HEC-PR scheme presented in Chapter 7. Afterwards, we will study the effects of the length of k, the group size and the CC of GE channel on the optimum architecture separately. Similar to the analysis in Chapter 7.6, for the convenience of analysis and comparison, we make some general assumptions as follows: The entire receivers experience independent GE channel with the same level of original link PLR and the same CC in the GE channel; and we also assume that the channel bandwidth is always enough so that we can set $t_{rp}=t_s$ for the transmission time of retransmission packets. Moreover, we still use t_s instead of t_d in the end-to-end delay budget for the general architecture to keep a certain margin. In addition, it is assumed that the average link PLR is less than 10^{-1} and the waiting time at both the receivers and the sender is neglected. However, for any real-time multicast scenario with any average link PLR and non-zero waiting time, the similar procedure introduced in this chapter is also suitable for the performance analysis and comparison. Now we apply the general architecture in the typical scenarios with the system parameters as shown in Table 10.1.

Table 10.1: System Parameters

PLR Requirement: PLR_{target}	10^{-6}
Latency Constraint: D_{target}	100ms
Multimedia Data Rate: R_d	0.2~4Mbps
Packet Loss Model	GE Model
RTT	5~50ms
$t_{rw}+t_{sw}$:	0ms
Encoding Packet Length: L	1250bytes
Original Link PLR: P_e	10^{-3}~10^{-1}

10.1. Optimization Results

Based on the systems parameters as shown in Table 10.1, above all, we can design the optimum parameters for the general architecture. In this case, since the original link PLR is no more than 10%, the RI_{lim} can be set to 40% here so that the $N_{p,\max}$ can be computed for each possible $N_{rr,max}$. Using the greedy algorithm presented in Chapter 9.2, we obtain the optimum parameters of the architecture with different link PLR, media data rate, group size and so on. Parts of the optimum parameters of the general architecture for the multicast scenario with R_d=4Mbps are shown in Table 10.2 and 10.3.

Table 10.2: Optimum Parameters of the Architecture with ρ=0, R_d=4Mbps and N_{recv}=4

Average Link PLR P_e	Optimum Parameters of the Architecture				
	k	N_p	$\vec{N}_{cc}$		
			N_{cc}^1	N_{cc}^2	N_{cc}^3
0.001	23	0	1	1	-
0.01	23	0	1	2	-
0.03	23	1	1	2	-
0.05	23	1	1	3	-
0.07	16	0	1	1	3
0.09	16	1	1	1	3
0.10	16	1	1	1	4

Table 10.3: Optimum Parameters of the Architecture with ρ=0, R_d=4Mbps and P_e= 0.03

N_{recv}	Optimum Parameters of the Architecture				
	k	N_p	$\vec{N}_{cc}$		
			N_{cc}^1	N_{cc}^2	N_{cc}^3
1	1	0	1	1	1
2	16	0	1	1	2
3	23	0	1	3	-
4	23	1	1	2	-
5	23	1	1	2	-
6	23	1	1	2	-
7	23	1	1	2	-

From these two tables, we can see that the proposed greedy algorithm can automatically choose the best scheme for the general architecture integrating different schemes according to the current scenario. Typically, as shown in Table 10.3, the HEC-PR scheme is chosen as the optimum scheme when the scenario is with N_{recv}=1; the Type I HARQ scheme is chosen as the optimum scheme when the scenario is with N_{recv} of more than 3; the Type II HARQ scheme is chosen as the optimum scheme when the scenario is with N_{recv}=2, 3. Apparently, those systems parameters (such as group size, RTT, average link PLR etc.) have a significant effect on the optimization results for the architecture. In the following, we will discuss the influences of some typical system parameters on the optimum architecture of EER.

10.1.1. Influence of Multicasting

To analyze the influence of the group size on the general architecture of EER under strict delay constraints, by solving (9.4) using the greedy algorithm proposed in Chapter 9.2, we obtain the optimum parameters for the architecture with the group size of varying from 1 to 20. Meanwhile, we also get the optimization results with ρ=0.0 and ρ=0.1 to study the effect of the CC in the GE channel, which are shown in Figure 10.1 and Figure 10.2, respectively. In these two figures, each rectangle means one scheme chosen as the optimal scheme by optimizing the general architecture for the scenario determined by the left-down point in the rectangle: Where the HEC-PR scheme is mapped on “Red” areas, the Type I HARQ scheme is mapped on “Green” areas and the Type II HARQ scheme is mapped on “Blue” areas. Additionally, the optimum maximum number of retransmission rounds is also inserted in the rectangle.

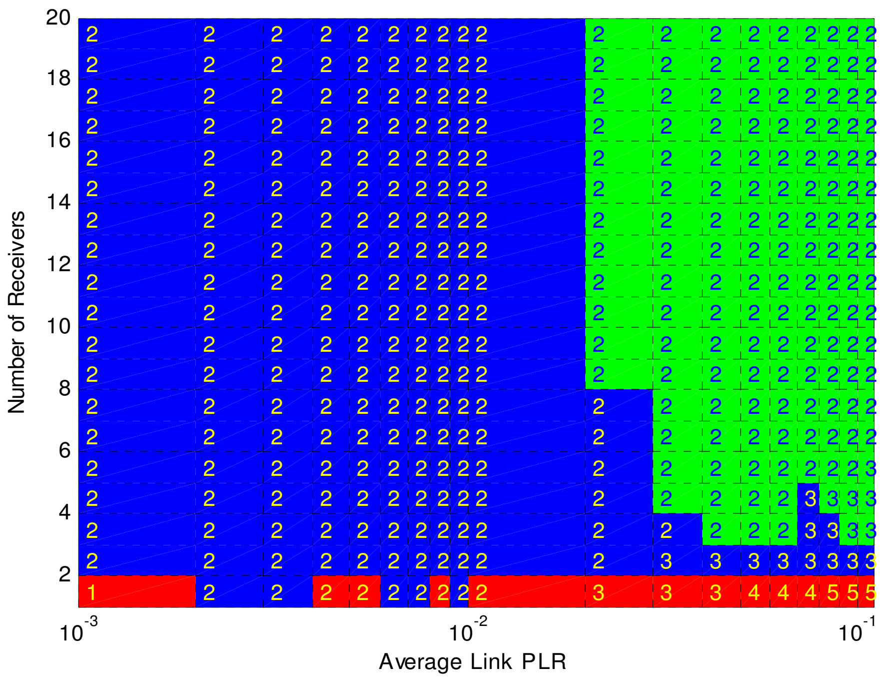


Figure 10.1 Optimization Results on the Architecture with R_d=4Mbps and ρ=0.0

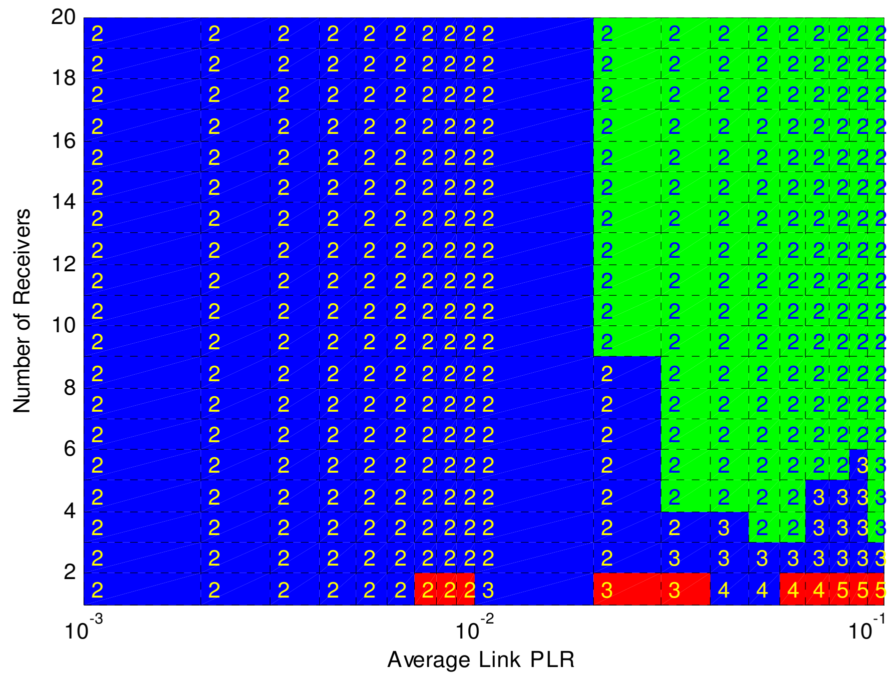


Figure 10.2 Optimization Results on the Architecture with R_d=4Mbps and ρ=0.1

From Figure 10.1 and Figure 10.2, we can see that the global trend is very clear: When the group size is very small (e. g. only one receiver), the HEC-PR scheme will be the best scheme in many cases; with the increase of the group size, the Type II HARQ scheme will become the best one if the average link PLR is relative small (e. g. no more than 0.02); otherwise, the Type I HARQ scheme will be the best scheme if both the group size and the average link PLR are large enough.

This trend can be explained by the features of the three HEC schemes. First, the HEC-PR scheme has been proved to be very efficient for those scenarios with small group size and small link PLR in Chapter 7. However, because the receivers can not share common parity packets for recovering different missing data packets with the HEC-PR scheme, its total needed RI will increase with the group size almost linearly so that it fails when used in such as multicast scenarios with large group size and large link PLR. Type I or Type II HARQ schemes, however, since different receivers always can share common parity packets for recovering different missing data packets, are much more efficient than the HEC-PR scheme when the group size is large enough. Secondly, comparing the Type I HARQ scheme with the Type II HARQ scheme, the Type I scheme sends at least one redundant packet in the first transmission; the Type II scheme, however, only retransmits minimum needed redundant packets according to the NACKs of receivers. Therefore, in the area with lower PLR, the margin of PLR performance of the Type I HARQ scheme is much looser than that of the Type II HARQ scheme if the parameter k is small (where the large code rate has to be adopted in the first transmission), which results in the fact that under the same delay constraints the Type II scheme usually can achieve the target PLR requirement with less needed RI than the Type I scheme in this area. On the other hand, in the area with higher PLR, the number of retransmissions can be reduced efficiently by sending a certain amount of parity packets in the first transmission with the Type I scheme, which results in the fact that under the same delay constraints the Type I scheme usually can guarantee the target PLR requirement with less needed RI than the Type II scheme in this area. These results show that the intuitive analysis results on different EER schemes in Chapter 2.2.4 are correct.

Some “strange” areas, however, appear in these two figures: There are some holes and gaps in those maps. For example, the HEC-PR scheme intuitively should cover the whole bottom line in Figure 10.1 and Figure 10.2; however, it is interrupted by the Type II HARQ scheme. This phenomenon can be understood from the following two aspects: On the one hand, the HEC-PR has nearly the same performance to the Type II HARQ scheme in the areas with gaps. That is, both of them can work nearly perfectly in those areas. Therefore, we can use the HEC-PR scheme for those gaps instead of the Type II HARQ scheme without decreasing the performance too much. On the other hand, the Type II HARQ scheme with the smallest needed RI actually results in slightly higher PLR than the HEC-PR scheme in the areas with gaps, but both of them can satisfy the target PLR requirement under the same strict delay constraints. In other words, although the HEC-PR scheme has better PLR performance than the Type II HARQ scheme in those special areas, the scheme with the smallest needed RI is chosen as the optimal scheme. Because the Type II HARQ scheme can use the allowed time more efficiently than the HEC-PR scheme in those areas (which can be realized by the needed maximum number of retransmissions), it needs slightly less RI than the HEC-PR scheme for guaranteeing the target PLR requirement within the same delay boundary. As a result, the optimization algorithm always chooses the scheme with the smallest needed RI under an upper limit of delay and PLR. However, the granularity of the architecture parameters (k,N_p etc.) leads to a discrete finite-set on the RI performance of those schemes, which results in those holes and gaps. Actually, the same holds true for the boundaries between the Type I HARQ scheme and the Type II HARQ scheme, which will be demonstrated in more detail in the next section. Finally, we must point out the optimization algorithm actually provides a good way for the tradeoff between the total needed RI and the PLR performance on the general architecture of EER.

Now focusing on the optimum $N_{rr,max}$ for the architecture, we can see that the total trend is still very clear: Along with the increase of the average link PLR with the group size of no more than six, the value of $N_{rr,max}$ increases with the increase of the average link PLR; along with the increase of the group size under certain average link PLR, however, it will decrease with the increase of the group size. It is not difficult to understand: On the one hand, in the areas with higher average link PLR, the optimum scheme have to adopt more number of retransmission rounds to achieve the same target PLR requirement than it does in those areas with lower average link PLR. On the other hand, with the increase of the group size, each receiver will have more chance to benefit from the parity packets required by other receivers at each retransmission stage. It indicates that the target PLR can be achieved faster than those with small group size. Therefore, the optimum parameter $N_{rr,max}$ can be decreased with the increase of the group size.

Additionally, we would like to point out a noticeable phenomena on those optimum $N_{rr,max}$ in the boundary areas with holes and gaps: Along with the increase of the average link PLR, the optimum $N_{rr,max}$ either keeps invariable or increases by one while the optimum scheme is changed. For example, for the scenarios with N_{recv}=3 as shown in Fig.3.1, the Type II HARQ scheme with $N_{rr,max}$=2 is the best scheme in case of P_e=0.03 while the Type I HARQ scheme with $N_{rr,max}$=2 becomes the best scheme in case of P_e=0.04~0.06, which indicates that the Type I scheme can make use of two retransmission rounds more efficiently than any other schemes in those areas. However, when the average link PLR is increased from 0.06 to 0.07 in this case, the Type II HARQ scheme becomes the optimum scheme again by increasing the parameter $N_{rr,max}$ from two to three, which indicates that the Type II scheme can make use of three retransmission rounds in the most efficient way for this case. As a result, the propose algorithm can not only choose the optimum scheme with the minimum total needed RI for achieving a certain PLR requirement, but also make the needed maximum number of retransmission rounds as small as possible.

When comparing Figure 10.1 with Figure 10.2, finally, we can find out that the maps in these two figures are very similar, which means the total trend on the choice of architectures is identical. That is, the optimization results are not sensitive to the small CC of the GE channel (e. g. $\rho \leq 0.1$). Therefore, as analyzed above, we can adopt closed maps for the choice of the architecture of EER with decreasing the performance only a little so that the architecture can be implemented simply and work robustly.

10.1.2. Influence of RTT and Media Data Rate

For practical systems, the RTT and media data rate are also very important parameters. In this chapter, we thus discuss the influence of the RTT and media data rate on the architecture under strict delay constraints. Similar to the discussions in Chapter 10.1.1, we obtain the optimum results of the architecture with the scenarios of N_{recv}=5 and ρ=0.0, where the RTT varies from 5ms to 50ms and the media data rate varies from 0.2Mbps to 2.0Mbps. Meanwhile, we also search out the optimization results with P_e=0.01 and P_e=0.1 to study the effect of the average link PLR, which are shown in Figure 10.3 and Figure 10.4, respectively.

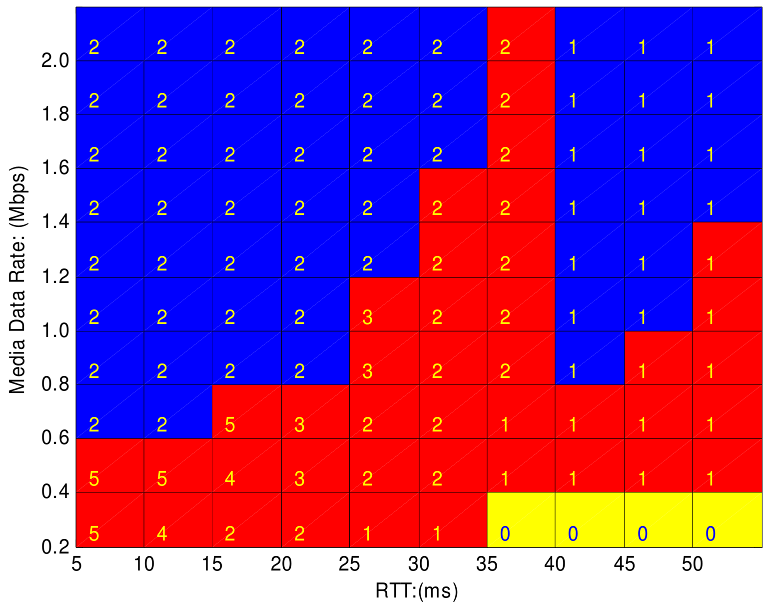


Figure 10.3 Optimization Results on the Architecture with N_{recv}=5, P_e=10^{-2} and ρ=0.0

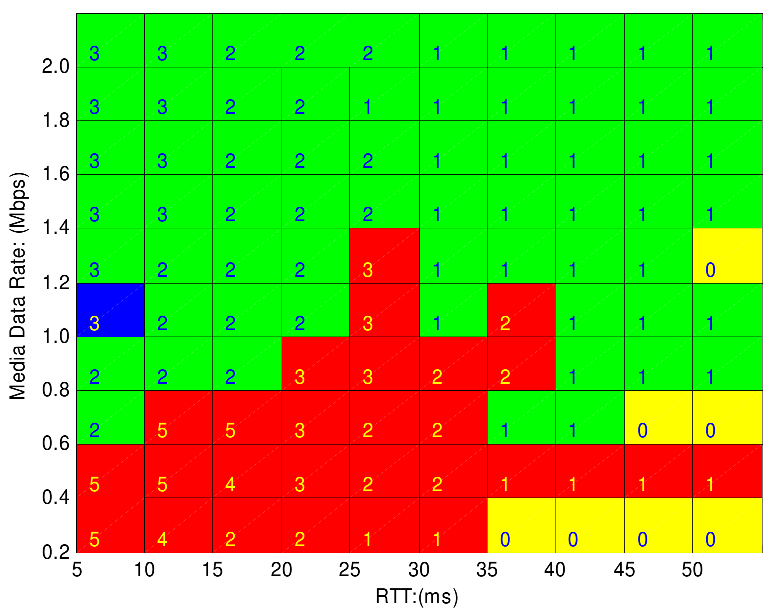


Figure 10.4 Optimization Results on the Architecture with N_{recv}=5, P_e=10^{-1} and ρ=0.0

In these two figures, the meanings of the colors and numbers are same to those in Figure 10.1 and Figure 10.2, besides that the AFEC scheme is mapped on "Yellow" areas.

First, as shown in Figure 10.3 and Figure 10.4, we find that the optimum $N_{rr,max}$ decreases with the increase of the RTT. The reason is clear: Under a certain delay constraints, the maximum allowable number of retransmission rounds has to be reduced with the increase of the RTT. Especially, as shown in these two figures, the AFEC scheme has to be adopted when the media data rate is low and the RTT is large, since no any retransmission round is allowed for those cases. Secondly, from Figure 10.3, we can find that the more efficient HEC-PR scheme and Type-II HARQ scheme can be adopted when the media data rate is high enough. That is because the receivers with higher data rate can response faster than those with lower data rate, which results in one or more retransmission rounds can be carried out even though the RTT is relative large. Comparing Figure 10.4 with Figure 10.3, we can see that the Type I HARQ scheme becomes the best scheme instead of the Type II HARQ scheme for most cases when the average link PLR increase from 0.01 to 0.1. Obviously, this result is similar to that shown in Figure 10.1 and Figure 10.2. Furthermore, for each combination of the RTT and the media data rate, the greedy algorithm can not only find out the optimum scheme with the minimum total needed RI but also make use of each retransmission stage in the most efficient way.

As a result, by adjusting the number of retransmission rounds and changing schemes integrated in the general architecture of EER, the proposed greedy algorithm always can provide a good tradeoff between the PLR performance and the RI performance by adapting to all kinds of system parameters (e.g. RTT, media data rate, group size, average link PLR etc.).

10.1.3. Type I HARQ vs. Type II HARQ

In order to realize those "strange" holes and gaps in the boundaries between the Type I HARQ scheme and the Type II HARQ scheme in more detail, we compare them with each other in this chapter. Through optimizing the general architecture with different limitations as explained in Chapter 10.1.1, we can obtain the minimum total needed RI for them separately. Part of the results on these two schemes with ρ=0 are shown in Figure 10.5.

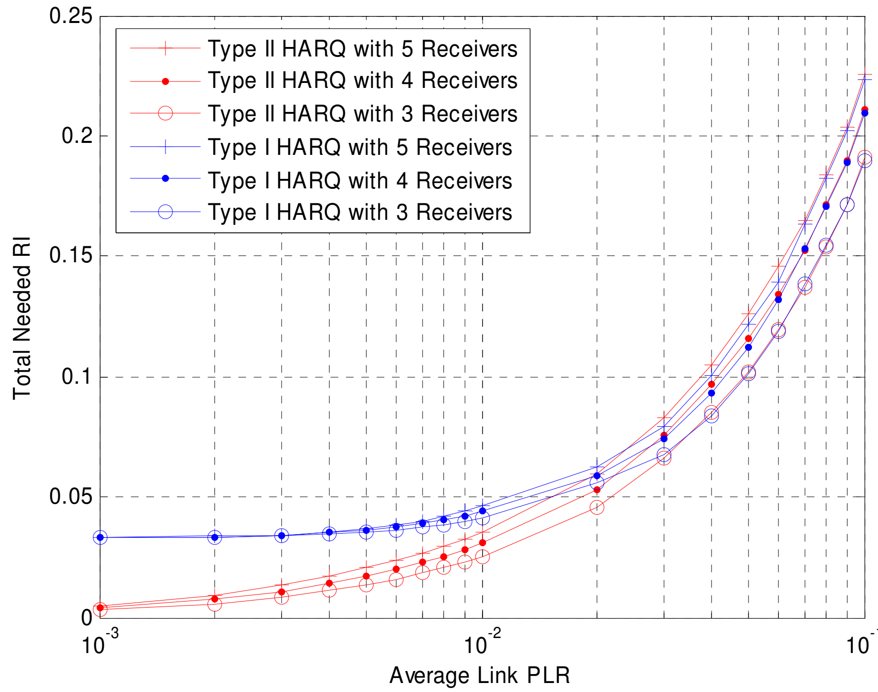


Figure 10.5 Type I HARQ vs. Type II HARQ with R_d=4Mbps and ρ=0.0

From this figure, first, we can see that the Type II scheme outperforms the Type I scheme in the area with lower PLR (e. g. less than 0.01); and the result will be the opposite in the area with higher PLR (e. g. for the scenarios with N_{recv}=5 and the average link PLR in the range of between 0.03 and 0.06). These results prove that our analysis above on the choice between them in different areas is correct. Secondly, from Figure 10.1, we can see that the gaps between these two schemes appear in the area of PLR varying from 0.07 to 0.09. By observing Figure 10.5, an interesting effect can be observed: Both of them perform nearly perfectly in the same area of PLR varying from 0.07 to 0.10. This result indicates that our explanations on the holes and gaps in the boundaries between these two schemes are credible. Actually, from this figure, we can see that the Type I HARQ scheme always performs perfectly or very nearly perfectly for the scenarios with higher average link PLR. Therefore, we can use the Type I HARQ scheme instead of the Type II HARQ scheme for the areas with gaps to form closed areas with nearly the perfect performance. The same holds true for the gaps between the HEC-PR scheme and the Type II HARQ scheme.

10.1.4. Practical Considerations

Considering the practical implementation for the optimum architecture as shown in Figure 10.1 with small group size, we can find that the optimum architecture will change very frequently when the scenarios vary in those boundary areas with holes or gaps. Actually, we can use suitable schemes instead of those holes and gaps to form smoother maps. Based on this idea, we obtain closed maps using the greedy algorithm for the same scenarios as in Figure 10.1 with small group size (i.e. $N_{recv}\leq 7$). In this thesis, the architecture derived from the smoothed maps is identified as sub-optimum architecture; the optimum parameters of the smoothed architecture are identified as sub-optimum parameters. The smoothed maps are shown in Figure 10.6, where the meanings of colors and numbers are same to Figure 10.1.

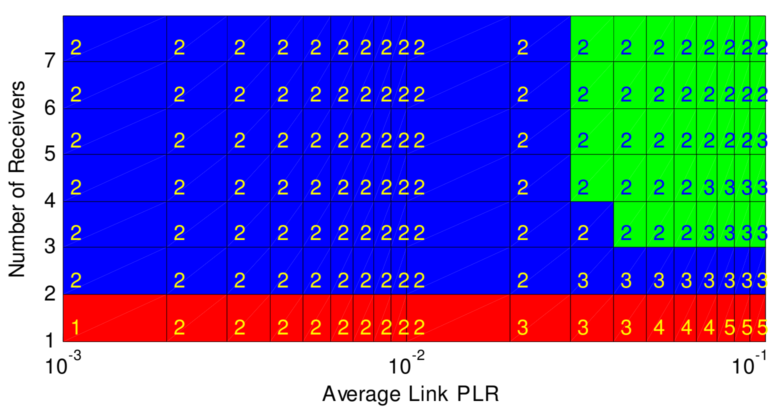


Figure 10.6 Smoothed Maps on the General Architecture

Now comparing Figure 10.6 with Figure 10.1 with $N_{recv}\leq 7$, we can find an interesting phenomenon: The optimum $N_{rr,max}$ for both of them are completely identical. It indicates that the sub-optimum architecture does not need to increase $N_{rr,max}$ for achieving the same target

PLR requirement. Therefore, the sub-optimum scheme can also make the optimum $N_{rr,max}$ as small as possible. On the other hand, we can see that the maps in Figure 10.6 are much smoother than Figure 10.1 with small group size. To compare the sub-optimum architecture with the optimum architecture in more detail, we list those optimum parameters in Table 10.4 and sub-optimum optimum parameters in Table 10.5 with the same scenarios.

Table 10.4: Optimum Parameters of the General Architecture with N_{recv}= 3

PLR	k	N_p	$\bar{N}_{cc}$			RI
			N_{cc}^1	N_{cc}^2	N_{cc}^3	
0.07	16	0	1	1	3	0.1368
0.08	16	0	1	1	3	0.1541
0.09	16	1	1	1	3	0.1714
0.10	16	1	1	1	4	0.1900

Table 10.5: Sub-optimum Parameters of the General Architecture with N_{recv}= 3

PLR	k	N_p	$\bar{N}_{cc}$			RI
			N_{cc}^1	N_{cc}^2	N_{cc}^3	
0.07	16	1	1	1	3	0.1382
0.08	16	1	1	1	3	0.1546
0.09	16	1	1	1	3	0.1714
0.10	16	1	1	1	4	0.1900

First, from Table 10.5, it is found that the same sub-optimum parameters can be adopted in a wide range of average link PLR (e.g. from 0.07 to 0.09). From Table 10.4, however, we can see that the optimum parameters have to be changed when the average link PLR varies between 0.08 and 0.09. It proves that the sub-optimum architecture is more stable than the optimum architecture. More importantly, comparing the needed RI in Table 10.5 with Table 10.4, we can find that the needed RI with the sub-optimum parameters is only a little more (e.g. about 1% for the case with PLR of 0.07) than that with the optimum parameters. It indicates that the sub-optimum architecture has nearly the same performance as the optimum architecture. As a result, to form smoothed maps, we can adopt the sub-optimum architecture instead of the optimum architecture with nearly the best performance. Based on the smoothed maps, actually, we can build up tables such as Table 10.5 in prior and then subsample those tables. Afterwards, the full tables can be rebuilt at any time by interpolating the subsampled tables due to the smoothness of those maps. This indicates a look-up technique with subsampled tables will be very attractive. Thus, considering practical implementations, there are two intuitive ways to implement the sub-optimum architecture:

1. Using the proposed greedy algorithm, first, we can obtain those sub-optimum parameters off-line according to all kinds of scenarios. Then, those parameters can be made as tables and saved in the memory at both the sender and the receivers. Finally, in practical systems, the parameters of the architecture for current scenario can be determined by looking up those tables immediately.
2. The smoothness of the maps indicates that there is very strong correlation on the sub-optimum parameters of the architecture among all kinds of scenarios. Therefore, based on the subsampled tables on those sub-optimum parameters of the architecture, we can derive the parameters for any current multicast scenario by some simple mathematical formulas. In this way, only a small part of the tables (i.e. the subsampled tables) is needed to be saved in the memory. When the full tables are very large and the memory of end-devices is very limited, consequently, this method will be much more attractive than the way of saving and looking up the full tables.

10.2. Performance Comparisons

In this chapter, we compare the optimum architecture of EER with the AFEC scheme presented in Chapter 6 and the HEC-PR scheme proposed in Chapter 7. For the convenience of description, the scheme derived from optimizing the general architecture of EER is denoted by Adaptive HEC (AHEC) scheme in this thesis. Depending on the system parameters shown in Table 10.1, we then can design the optimum parameters for different EER schemes separately for a typical scenario with RTT=15ms, R_d=4Mbps and ρ=0.0. Finally, Figure 10.7 shows the total needed RI by optimizing those three different EER schemes: The AHEC scheme, the HEC-PR scheme and the AFEC scheme.

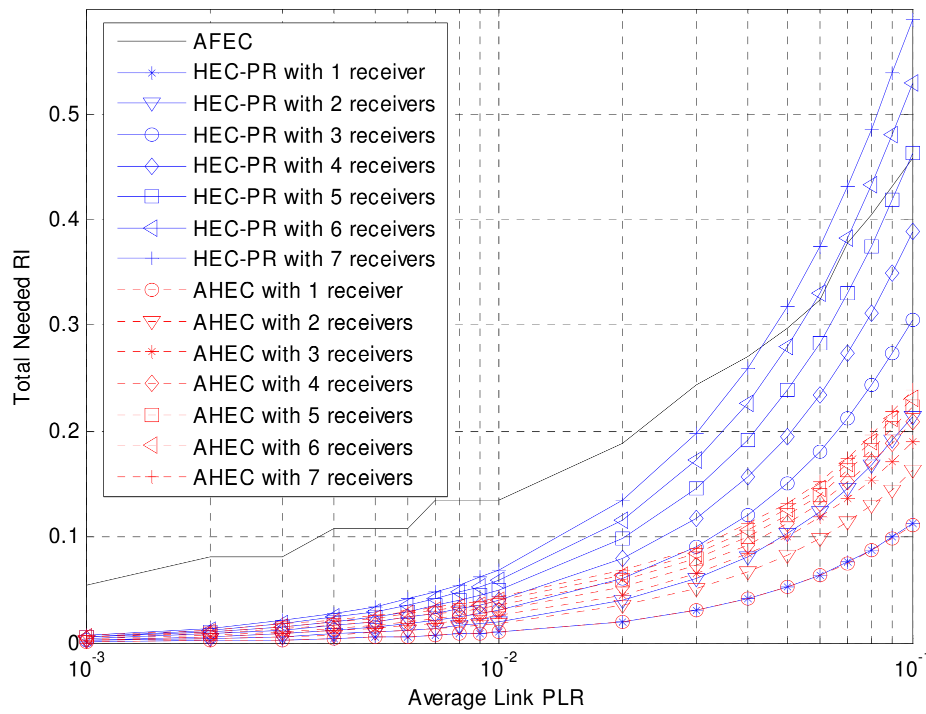


Figure 10.7 Performance Comparison under the Scenario with RTT=15ms, R_d=4Mbps and ρ=0.0

From this figure, we can see that the total needed RI of the HEC-PR scheme increase with the increase of the number of receivers significantly but not strict linearly. The reason is clear: Although all of the receivers are independent, they also can recover some missing packets by retransmitting a small part of identical packets with the HEC-PR scheme. Note that the probability of sharing identical packets among different receivers will increase with the increase of the number of receivers. It leads to the total needed RI of the HEC-PR scheme increase with the number of receivers significantly but not strict linearly. In other words, the speed of the increase of the total needed RI will slow down with the increase of the number of receivers.

Secondly, as shown in Figure 10.7, we can see that the AHEC scheme always outperforms the AFEC scheme. The reason is clear: The very short length of k with the AFEC scheme limits its performance, because the AFEC scheme can not adopt very efficient code rates to satisfy the QoS requirements due to its short codeword length. For the AHEC scheme, however, the retransmission technique can compensate this shortage so that it can outperform the AFEC scheme.

Moreover, although the total needed RI of the AHEC scheme also increases slightly with the increase of the group size, it outperforms the HEC-PR scheme for those scenarios with the group size of more than one. The main reason is that the AHEC scheme has overcome the shortage of the HEC-PR scheme by combining the Type I and Type II HARQ schemes. In other words, using the Type I or Type II HARQ scheme, each receiver has more chance to benefit from sharing the common parity packets required by other receivers with the increase of group size, which results in the total needed RI increasing slightly with the increase of group size. Therefore, the AHEC scheme can outperform the HEC-PR scheme with large group size.

Finally, we would like to point out that the performance of the HEC-PR scheme is very close to the AHEC scheme under the scenarios with small group size (e.g. $N_{recv} \leq 7$) and small average link PLR (e.g. $P_e < 10^{-2}$). Because the implementation of the HEC-PR scheme is simple due to without encoding and decoding algorithms, the HEC-PR scheme should be considered for such as those real-time multicast scenarios with small group size and small average link PLR.

10.3. The Effect of Important Parameters

From the analysis results above, we can see that all kinds of system parameters, such as the group size, average link PLR, RTT, multimedia data rate etc., have a noticeable impact on the performance of the optimum architecture of EER. In order to see how the group size in a multicast scenario influences the performance of the general architecture in more detail, first, we will analyze its effect on the performance of the optimum architecture in this chapter. Then, from (9.3), we can see that the system parameters, such as RTT, multimedia data rate and the packet size, affect directly the value of the parameter k in the general architecture of EER under strict delay constraints. To study its effect, we will analyze the performance of the optimum architecture with the parameter k of different values in following sections. Similar to the analysis for the AFEC scheme in Chapter 6.3.2 and for the HEC-PR scheme in Chapter 7.6.3, finally, we will also study the effect of the CC of GE channel model to the performance of the optimum architecture of EER.

10.3.1. The Effect of the Group Size

This section focuses on the effect of the group size to the performance of the AHEC scheme. In order to design the optimum parameters of the general architecture with different group size, for the convenience of comparison, we adopt the same system parameters as shown in Table 7.3 with ρ=0.0. Similar to the analysis in Chapter 10.1, we then can design the optimum parameters of the general architecture for this scenario with different group size. Based on those optimum parameters similar to Table 10.2 and Table 10.3, finally, we can compute the total needed RI of the AHEC scheme by (8.9), which is shown in Figure 10.8.

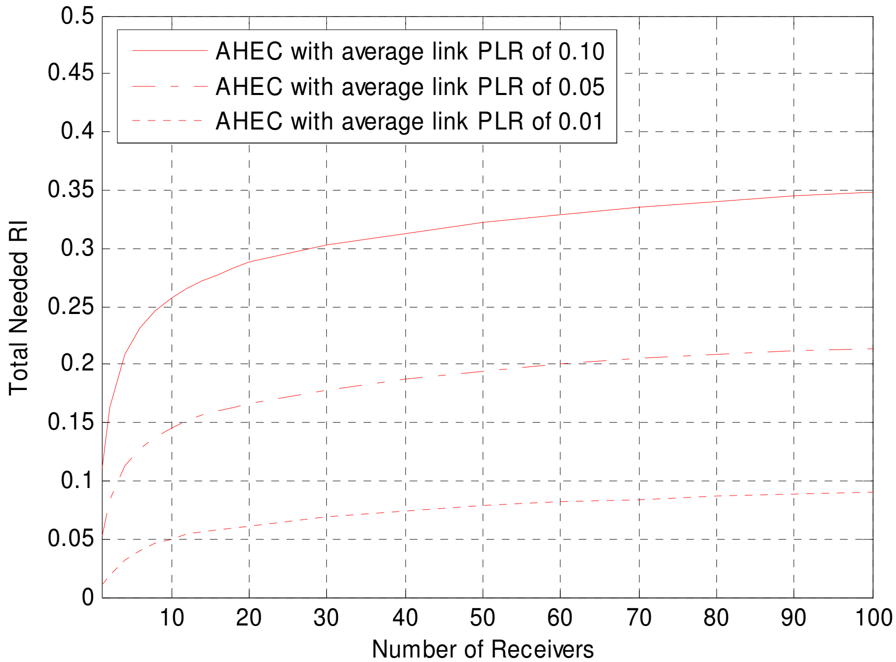


Figure 10.8 The effect of the group size under the scenarios with RTT=15ms, R_d=4Mbps and ρ=0.0

As shown in this figure, the total needed RI of the AHEC scheme increase significantly with the increase of group size of from 1 to 10. But when the number of receivers is more than 10, it will increase very slightly with the increase of group size, which indicates that the stable performance of the AHEC scheme can be achieved for the multicast scenarios with large group size. It is not difficult to understand: With the increase of the number of receivers, more and more receivers can share same redundant packets for recovering different missing data packets, which results in the total needed redundant packets will increase slightly with the increase of the number of receivers. In addition, Figure 10.8 shows that the total needed RI of the AHEC scheme with N_{recv}=100 and average link PLR of 0.10 is about 0.35 while that of the AFEC scheme is about 0.45 as shown in Figure 7.5, which indicates that the proposed AHEC scheme can outperform the AFEC scheme for the scenarios with even very large group size. As a result, the AHEC scheme is also a good candidate technique for guaranteeing the reliable real-time multicast services with large group size.

10.3.2. The Effect of the Parameter *k*

Since the length of *k* has a significant effect on the performance of the Type I or Type II HARQ scheme integrated in the general architecture, in the following, we study the effect of *k* for the optimum architecture. To demonstrate the effect of the parameter *k*, we firstly design the optimum parameters for the general architecture with the same parameters as shown in Table 7.3 under different average link PLR. For the convenience of analysis, however, we will only study the performance of the general architecture for the multicast scenario with N_{recv}=5, $N_{rr,max}$=2 and ρ=0.0. Finally, upon those optimum parameters with different length of *k*, the total needed RI of the optimum architecture can be obtained and then shown in Figure 10.9.

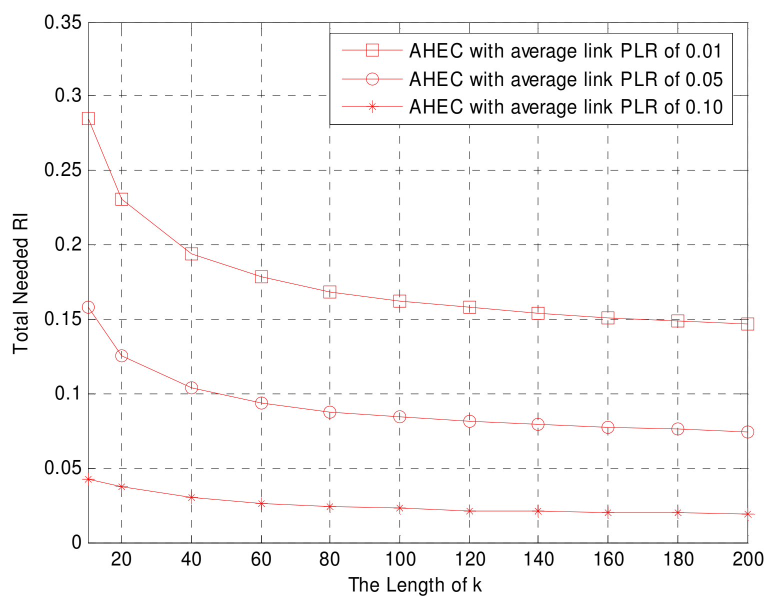


Figure 10.9 The effect of the parameter *k* under the scenario with RTT=15ms, R_d=4Mbps, N_{recv}=5, $N_{rr,max}$=2, and ρ=0.0

Note that although here we only show the results for the optimum architecture with the length of k being less than 200, it should be clear that the general architecture is suitable for any length of k upon requirements; moreover, the higher k is employed by the ideal FEC code, the more efficient code rate can be adopted. Additionally, the variable value of k means different multimedia data rate under certain delay constraints if the packet size is constant. Obviously, this feature of the AHEC scheme can simplify its implementation in real-time multicast scenarios with variable source data rate.

From this figure, we can see that the total needed RI decreases significantly when the parameter k increases from 10 to 60. When k is more than 60, however, this parameter has only a little effect on the performance of the AHEC scheme. Note that different k values mean different delay constraints or different source data rates if the packet size is fixed. Therefore, under certain delay constraints with fixed packet size, the higher the multicast source data rate is, the better performance can be achieved in the AHEC scheme. Moreover, since the stable good performance can be obtained if the parameter k is more than 60, a suitable fixed short length of k ($\geq$60) can be always adopted when the data rate is high enough to provide good delay performance. On the other hand, the parameter k is only associated with the delay constraints if both the source data rate and the packet size are fixed: to guarantee a certain PLR requirement, shorter delay constraints the system has, shorter length of the parameter k in the AHEC scheme has to be adopted so that more RI needed. Therefore, the AHEC scheme also provides a good way for the tradeoff between the total needed RI and delay constraints by choosing different k.

10.3.3. The Effect of CC of GE Channel

Finally, to study the effect of the CC of GE channel on the AHEC scheme derived from optimizing the general architecture of EER, we study its performance with a multicast scenario of the same parameters as shown in Table 7.3. For the convenience of analysis, it is assumed that the scenario is with N_{recv}=5 and ρ=0.0~1.0. Afterwards, we can obtain the optimum parameters of the AHEC scheme with different CC of GE channel. Based on those optimum parameters, we then can obtain the total needed RI of the AHEC scheme with the variable CC of GE channel, which is shown in Figure 10.10.

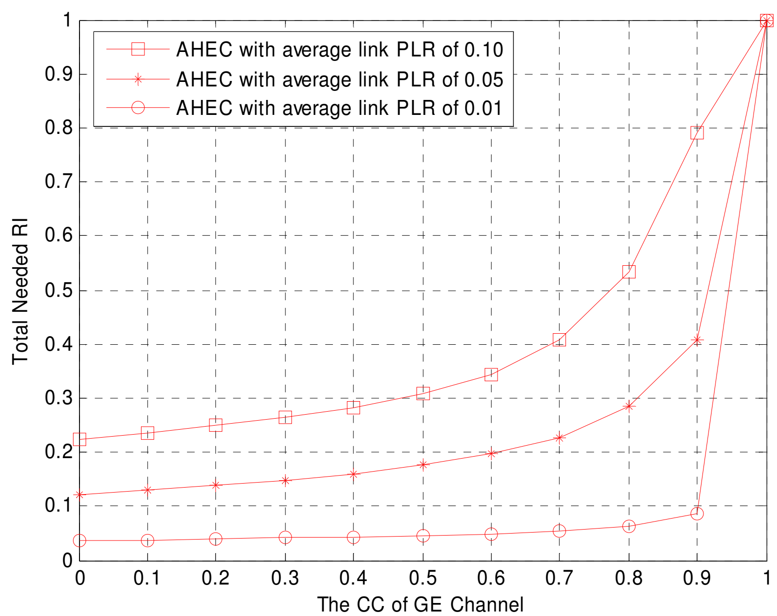


Figure 10.10 The effect of the CC under the scenarios with RTT=15ms, R_d=4Mbps and N_{recv}=5

From Figure 10.10, it is found that the CC of GE channel has a noticeable effect on the performance of the AHEC scheme. The total needed RI of the AHEC scheme increases significantly with the increase of the CC in the GE channel. This phenomenon can be explained by the features of the GE channel. First, from Table 3.1 we can see that the value of $P_{B|B}$ increases with the increase of ρ. From (3.12), moreover, we know that the average value of the error-burst length in the GE channel increases with the increase of $P_{B|B}$. As a result, the average number of missing data packets in one block increases with the increase of ρ. The AHEC scheme thus has to employ more parity packets with the increase of ρ to achieve a certain target PLR requirement. Especially, in case of ρ=1, the state of the GE channel will only depend on the initial state of the channel so that any erasure error recovery scheme fails in this case.

However, in case of $\rho \leq 0.1$, the total needed RI of the AHEC scheme is nearly the same to that with $\rho=0.0$. Fortunately, the CC of GE channel is indeed small in practical systems (it is usually less than 0.1 according to the practical evaluation results as presented in Table 3.2). Therefore, the AHEC scheme works nearly perfectly for most cases in real systems. As a result, the effect of ρ of GE channel on the AHEC scheme usually is negligible in practical systems.

10.4. Summary

In this chapter, based on the GE channel model, we analyzed the performance of the general architecture of EER with the greedy algorithm of optimizing its parameters. The optimization target is to satisfy a certain PLR requirement under strict delay constraints with minimum total needed RI for RMM services. Through the analysis, we have found:

1. The traditional Type II HARQ scheme can usually perform perfectly when the group size is large while the average link PLR is small;
2. When both the group size and the average link PLR are large enough, the traditional Type I HARQ scheme can outperform all of the other schemes;
3. The total needed RI of the AHEC scheme will increase significantly with the increase of the group size if the group size is small. However, when the group size is large enough, the total needed RI will increase very slightly with the increase of the group size. That means the AHEC scheme can be very suitable to the scenarios with large group size;
4. The studies show that the CC of GE channel with small value has small impact on the design and the performance of the AHEC scheme. In most practical systems, the CC is usually very small (e.g. $\rho \leq 0.1$) so that its effect can be neglected;

5. The length of k has a significant effect on the performance of the AHEC scheme. Langer length of k is adopted, better performance is achieved. It indicates that the AHEC scheme is very suitable in the real-time multicast systems with high data rate. Also, it provides a good way for the tradeoff between the total needed RI and the strict delay constraints.

As a result, both the multicast and the strict delay boundary indeed influence deeply the results on optimizing the general architecture of EER. Theoretically, the AHEC scheme derived from optimizing the general architecture of EER can always work very well, because it can choose the best scheme automatically among overall schemes integrated in the general architecture. Apparently, as mentioned before, these findings provide a useful guide to design reliable real-time multicast protocols.

Chapter 11

Simulation Results

We carried out simulations by ns-2 for validating the mathematical framework proposed in this thesis. The basic protocol stack used for the simulations is RTP/UDP/IP. The packet loss model is set to the simplified GE model, which has been presented in Chapter 3.2. Moreover, the parameters of the GE model are produced as the method presented in Chapter 3.3. In the following sections, we present the simulation results and compare them with analysis results.

11.1. Simulation for the AFEC Scheme

From Chapter 6, we know that the performance of the AFEC scheme only depends on the ideal FEC code. It has nothing to do with the group size. The t_{sw} and t_{rw} have also no effect on the performance of the AFEC scheme due to the absence of any retransmission round. Therefore, for the AFEC scheme, we only need to validate the analysis results for the performance of an ideal (n, k) code. Without loss of generality, it is assumed that k is 30 and the length of n is variable between 30 and 40. Figure 11.1 shows the simulation results and the corresponding analysis results (refer to Table 6.1) with the original link PLR of 0.05 and different CC (ρ=0.0 and ρ=0.08).

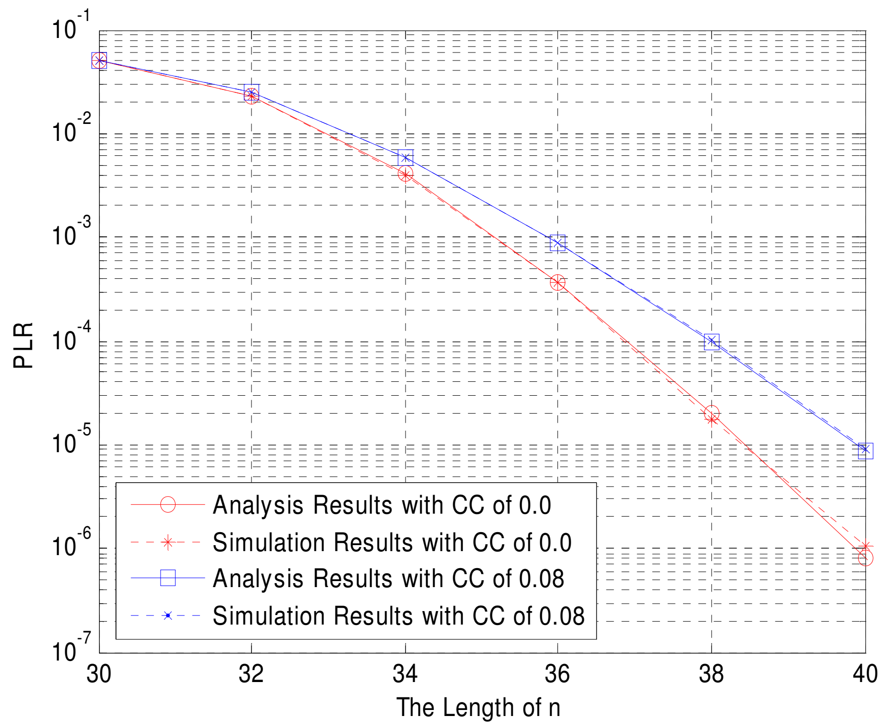


Figure 11.1. The PLR performances of the (n,30) codes with link PLR of 0.05

As shown in this figure, both the analysis results and the simulation results indicate that the PLR performance of the FEC code decreases significantly with the increase of the CC in the GE channel. It proves that the CC of the GE channel has a significant effect on the performance of the AFEC scheme. Furthermore, Figure 11.1 also shows that the analysis results match the simulation results very well, which indicates that the proposed formula in this paper is a good approximation for real systems. Actually, we only need to evaluate the performance of FEC codes roughly for designing optimum parameters for the AFEC scheme due to the large scale of the performances of FEC codes. Therefore, the mathematical formula is accurate enough for designing suitable (n, k) codes for the AFEC scheme.

11.2. Simulation for HEC Schemes

In this section, we run simulations with the HEC-PR scheme presented in Chapter 7 and the AHEC scheme coming from optimizing the general architecture of EER, respectively. Here the target PLR requirement is set to 10^{-4} and the delay constraint is set to 100ms. It is assumed that there are five receivers in a multicast scenario. For the convenience of comparing the simulation results with the analysis results fairly, we also set the bandwidth of the channel is much larger than the original multicast data rate. Then, the RTT for each receiver is assumed to be no more than 10ms as in WLANs. Additionally, note that in practical systems each receiver usually does not feedback NACK immediately. To avoid the implosion of NACKs, it often delays a random time to transmit the NACK by setting a random feedback timer. To make sure the end-to-end delay being in the range of 100ms, thus, we must consider the effect of the t_{sw} and t_{rw} in our simulations for these two HEC schemes. Here we set total of them to 10.0ms depending on the feedback timer used in our simulations. In the simulations, the retransmission packets with a delay of more than 100ms will be discarded. The system parameters are summarized in Table 11.1.

Table 11.1: System Parameters for Simulations

Target PLR Requirement	10^{-4}				
End-to-end Delay Constraints	100ms				
Number of senders	1				
Number of receivers	5				
Bandwidth	40Mbps				
Multimedia Data Rate	4Mbps				
Packet size (bytes)	1250				
$t_{rw}+t_{sw}$:	10.0ms				
Link Delay (ms)	4.0	3.5	5.0	2.5	4.5
Channel Model	GE Model				
Group I					
CC of GE channel	0.0				
Original Link PLR	0.05	0.05	0.05	0.05	0.05
Group II					
CC of GE channel	0.0				
Original Link PLR	0.05	0.08	0.02	0.01	0.10
Group III					
CC of GE channel	0.05				
Original Link PLR	0.05	0.05	0.05	0.05	0.05
Group IV					
CC of GE channel	0.05				
Original Link PLR	0.05	0.08	0.02	0.01	0.10

As shown in Table 11.1, to show the effect of the CC of the GE channel, we design the same link PLR features for all of the receivers in Group I (or Group II) with the Group III (or Group IV), but set a different CC for the GE channel. On the other hand, in order to simulate real multicast scenarios, we set different link PLR for different receivers in Group II and Group IV.

At the beginning, we design the parameters for the HEC-PR scheme using the mathematical framework proposed in Chapter 7. Also, we design the optimum parameters for the AHEC scheme using the general mathematical framework proposed in Chapter 9. For simplifying the simulations, we design the same parameters for all of the receivers with each scheme according to the worst case of every group. The parameters of the four groups are summarized in Table 11.2.

Table 11.2: Parameters for different groups with the HEC-PR scheme and the AHEC scheme

Group No.	HEC-PR			AHEC			
	N_{rt}^1	N_{rt}^2	N_{rt}^3	k	N_p	$\bar{N}_{cc}$	
						N_{cc}^1	N_{cc}^2
I	1	1	1	20	1	1	1
II	1	1	1	20	2	1	1
III	1	1	1	20	1	1	1
IV	1	1	1	20	2	1	1

Then, we run simulations for the four different groups with these two HEC scheme. The total number of sending RTP packets is 100,000,000 for every simulation. Furthermore, we run 10 times randomly for each group with each HEC scheme and then compute the average value of those simulation results. Finally, those average values are viewed as the final simulation results. The PLR performances for the four groups are shown in Table 11.3-7.6, respectively. In the following Tables, 'SR' denotes 'Simulation Results' and 'AR' denotes 'Analysis Results'.

Table 11.3: PLR performance for Group I with two HEC schemes

Receiver No.	Link PLR	PLR after correction with different schemes			
		HEC-PR		AHEC	
		SR	AR	SR	AR
1	5%	6.25×10^{-6}	6.25×10^{-6}	5.12×10^{-6}	5.15×10^{-6}
2	5%	6.27×10^{-6}		5.12×10^{-6}	
3	5%	6.19×10^{-6}		5.09×10^{-6}	
4	5%	6.24×10^{-6}		5.05×10^{-6}	
5	5%	6.08×10^{-6}		5.05×10^{-6}	

Table 11.4: PLR performance for Group II with two HEC schemes

Receiver No.	Link PLR	PLR after correction with different schemes			
		HEC-PR		AHEC	
		SR	AR	SR	AR
1	5%	6.11×10^{-6}	6.25×10^{-6}	8.70×10^{-8}	9.94×10^{-8}
2	8%	4.04×10^{-5}	4.10×10^{-5}	2.31×10^{-6}	2.33×10^{-6}
3	2%	1.32×10^{-7}	1.60×10^{-7}	0.0	1.84×10^{-10}
4	1%	1.10×10^{-8}	1.00×10^{-8}	0.0	1.49×10^{-12}
5	10%	0.99×10^{-4}	1.00×10^{-4}	0.99×10^{-7}	1.01×10^{-7}

Table 11.5: PLR performance for Group III with two HEC schemes

Receiver No.	Link PLR	PLR after correction with different schemes			
		HEC-PR		AHEC	
		SR	AR	SR	AR
1	5%	6.14×10^{-6}	6.25×10^{-6}	5.28×10^{-6}	5.56×10^{-6}
2	5%	6.20×10^{-6}		5.52×10^{-6}	
3	5%	6.08×10^{-6}		5.42×10^{-6}	
4	5%	6.25×10^{-6}		5.52×10^{-6}	
5	5%	6.24×10^{-6}		5.45×10^{-6}	

Table 11.6: PLR performance for Group IV with two HEC schemes

Receiver No.	Link PLR	PLR after correction with different schemes			
		HEC-PR		AHEC	
		SR	AR	SR	AR
1	5%	6.22×10^{-6}	6.25×10^{-6}	1.18×10^{-7}	1.19×10^{-7}
2	8%	4.08×10^{-5}	4.10×10^{-5}	2.41×10^{-6}	2.54×10^{-6}
3	2%	1.90×10^{-7}	1.60×10^{-7}	0.0	3.08×10^{-10}
4	1%	1.20×10^{-8}	1.00×10^{-8}	0.0	3.89×10^{-12}
5	10%	0.99×10^{-4}	1.00×10^{-4}	1.06×10^{-5}	1.06×10^{-5}

From Table 11.3-7.6, we can see that the simulation results match the analysis results very well, which proves that the MF proposed in this thesis is accurate enough for designing the optimum parameters for all kinds of EER schemes. On the other hand, as shown in these tables, both of the two HEC schemes can satisfy the target PLR requirement very well for all of the receivers in the four simulation groups. Especially, as shown in Table 11.3-7.6, the final PLR in the simulations with the AHEC scheme is always no more than the PLR obtained by analysis. Also, the final PLR obtained by the simulations is very close to the analysis results. It proves that the analysis results are indeed a very tight upper-band of the PLR performance for the general architecture of EER. Therefore, when adopting the upper-band performance of the general architecture to design its optimum parameters, the AHEC scheme can guarantee the final PLR performance very well.

In the following, simulation results for the maximum delay of useful retransmission packets with the HEC-PR scheme and the AHEC scheme are also shown in Table 11.7.

Table 11.7: Simulation Results for Maximum Delay of useful Retransmission Packets

Receiver No.		Maximum Delay of retransmission packets: ms			
		Group I	Group II	Group III	Group IV
1	HEC-PR	88.0	88.6	90.4	89.9
	AHEC	94.7	93.3	95.0	93.3
2	HEC-PR	84.7	90.7	90.0	91.5
	AHEC	93.7	92.9	95.6	92.9
3	HEC-PR	93.4	87.9	92.5	88.1
	AHEC	96.5	94.4	96.3	94.4
4	HEC-PR	86.5	80.7	90.2	84.4
	AHEC	96.8	93.1	96.0	93.1
5	HEC-PR	84.6	92.9	88.9	92.8
	AHEC	99.6	92.2	93.7	92.1

As shown in Table 11.7, the maximum delay of retransmission packets with the two HEC schemes has been controlled in 100ms for all of the four groups, which indicates that both of them can satisfy the target PLR requirement under the strict delay constraints. Furthermore, from this table, we can see that the maximum delay with the HEC-PR scheme is often a little less than that with the AHEC scheme. It indicates that the HEC-PR scheme responds faster than the AHEC scheme. The reason is clear: Because the Type I HARQ scheme is employed by the AHEC scheme for these multicast scenarios, each receiver has to spend more time to decide if one NACK is needed for retransmissions due to the FEC block decoding delay.

At last, by simulations we also calculated the total needed RI for the four groups with the two HEC schemes, which is summarized in Table 11.8.

Table 11.8: Total Needed RI for the four groups with different HEC schemes

HEC Scheme	Total Needed RI							
	Group I		Group II		Group III		Group IV	
	SR	AR	SR	AR	SR	AR	SR	AR
HEC-PR	26.32%	26.31%	28.13%	28.10%	26.32%	26.31%	28.14%	28.10%
AHEC	12.19%	12.18%	16.29%	15.71%	12.60%	12.59%	16.67%	16.08%

Obviously, Table 11.8 shows that the simulation results correspond to the analysis results very well. It also validates our analysis results for the performances of the two HEC schemes on the total needed RI. Also, as shown in this table, the total needed RI of the AHEC scheme is much less than that of the HEC-PR scheme. It is because the more efficient Type I HARQ scheme is used by the AHEC scheme for these multicast scenarios. Finally, when comparing the total needed RI shown in Group I (or Group II) with that in Group III (or Group IV), we can found that the total needed RI of the HEC-PR scheme with ρ=0 is nearly same to that with ρ=0.05, which validates that the CC of the GE channel has no apparent impact on its performance. For the AHEC scheme, we can see that the CC of the GE channel has a little effect on its performance, which also validates the analysis results on the impact of the CC of the GE channel to the AHEC scheme.

11.3. Summary

Using ns-2, we carried out simulations to validate the analysis results for three EER schemes proposed in this thesis: The AFEC scheme, the HEC-PR scheme and the AHEC scheme. It is found that the simulation results match the analysis results very well, which indicates that the MF contributed in this thesis is correct and useful. Also, it is proved that this framework is very suitable for designing the optimum parameters for different EER scheme in practical multicast scenarios. Especially, for the general architecture of EER, the MF of calculating the upper-band performance can be used for designing its optimum parameters for a multicast scenario, which can guarantee the target PLR requirement of the real-time services under the strict delay constraints very well.

Chapter 12
Conclusion

In order to guarantee the QoS requirements of Real-time Multimedia Multicast (RMM) services over packet-based networks, we must employ erasure error recovery (EER) schemes in real-time multicast protocols. Since the real-time services usually have strict end-to-end delay requirements, this thesis has studied the optimization problem of all kinds of EER techniques under strict delay constraints. First of all, this thesis has presented how to optimize the EER scheme with pure FEC techniques, which results in the adaptive FEC (AFEC) scheme as described in Chapter 6. Then, the thesis has presented how to optimize the EER scheme with pure ARQ techniques, which results in the HEC-PR scheme as described in Chapter 7. Finally, by proposing a general architecture integrating all of the important existing EER schemes, the thesis has presented how to analyze its performance in Chapter 8 and presented how to optimize its parameters in Chapter 9. In the following sections, we will summarize the contributions of this thesis and point out some interesting future works along with this work.

12.1. Contributions

The thesis contributions are concluded below.

1. By building a practical test bed for real-time multimedia services over Wireless LAN (WLAN) with IEEE 802.11 [Iee0a], we evaluated the characters of the erasure errors in the channels. It is found that the erasure error channels match the GE channel model very well, which infers that we can use GE channel model to accurately evaluate the performance of all kinds of EER mechanisms. These results were presented in Chapter 3.
2. Using FEC technique alone, an Adaptive FEC (AFEC) scheme is proposed for guaranteeing the target PLR requirement of real-time multicast services under strict delay constraints. By adapting the code rate of the FEC code used to the worst original link PLR in current multicast scenario, the total needed RI of the AFEC scheme can be minimized. The main advantage of the AFEC scheme is that no any feedback channel is needed in this case. The main disadvantage of the AFEC scheme is that the short code word usually has to be adopted due to the strict delay constraint, which leads to the less efficient erasure codes with large code rate having to be adopted for this scheme. This weakness obviously will decrease the transmission efficiency. The detail discussions on the AFEC scheme were presented in Chapter 6.
3. For a kind of typical practical multicast scenarios with multimedia services over wireless home networks (WHNs), they have the following features: “small group size”, “small round trip time” and “small link PLR” etc. For example, for DVB services over WHNs, the group size is usually less than 7, the round trip time (RTT) is usually less than 20ms, and the link PLR is usually less than 10%. To guarantee the target PLR requirement for this kind of applications under strict delay constraints, we proposed a pure ARQ based HEC scheme with Packet Repetition (RP) technique in Chapter 7. This scheme is denoted by HEC-PR scheme in this thesis. Our studies reveal that the HEC-PR scheme can work very well for those RMM applications with “small group size”, “small round trip time” and “small link PLR”. However, the

needed RI of the HEC-PR scheme will increase linearly with the increase of the group size so that it is not suitable for those multicast scenarios with large group size. These results on the HEC-PR scheme were discussed in Chapter 7.

4. To overcome the defects of the AFEC scheme and the HEC-PR scheme mentioned above, a general architecture integrating all of the important existing EER mechanisms is proposed in Chapter 2.3. Then, based on GE channel model, we contribute a general mathematical framework to analyze the performances of the proposed general architecture. This mathematical framework was presented in Chapter 8.
5. Based on the contributed mathematical framework, we then can optimize the parameters of the general architecture under strict delay constraints. To reduce the time of searching for the optimum results, we also contribute an efficient greedy algorithm for achieving the optimum parameters of the general architecture. The optimization problem on the general architecture under strict delay constraints were discussed in Chapter 9.
6. We contribute the analyze results on the optimum architecture of ERR under different typical multicast scenarios. By optimizing the parameters of the general architecture according to the system parameters of the current multicast scenario, the optimum HEC scheme with the minimum total needed RI can be found out for the scenario. In fact, the performance of the proposed optimum HEC scheme can close to the Shannon limit as closely as possible dynamically. The analysis results were discussed in Chapter 10. Finally, we also contributed the simulation results with ns-2 to validate the analysis results in Chapter 11.

12.2. Future Works

The general framework contributed in this thesis can be utilized and extended for further research in wide areas. In the following, we would like to point out some avenues for future research:

- How to implement the optimum architecture of EER in practical systems?

Although the performance of the AHEC scheme by optimizing the general architecture of EER is very attractive, how to implement it in practical systems still needs further study. For example, given a multicast scenario, how to obtain the optimum parameters in time at both the sender and the receivers is still a challenge. Especially, how to obtain the accurate Channel State Information (CSI) of current multicast scenarios is also a problem. Apparently, the accuracy of the CSI has directly impact on the performance of the AHEC scheme. A potential solution for this problem is that the AHEC scheme can cooperate with some efficient channel prediction techniques.

- Extend the framework to the packet-based networks with multi-hops

The framework contributed in this thesis can be viewed as a perfect solution for hop-to-hop connections. Therefore, it will be a very interesting research direction by extending this general framework to the packet-based networks with multi-hops, such as Peer-to-Peer (P2P) networks and mobile Ad hoc networks etc.

- Extend the framework to other types of networks

As mentioned in Chapter 3.2, the simplified GE channel model used in this thesis is not good enough for many other types of wireless networks (e.g. GSM, CDMA, WiMAX etc.). Using the accurate channel models for those networks instead of the simplified GE model, we then can study the optimum performance of the general architecture of EER over those different types of networks. Furthermore, as mentioned before, the derivation of the accurate mathematical framework on the general architecture is still a bigger challenge and needs further study. Nevertheless, we would like to point out that all of them are expected to have the similar results as analyzed in this thesis.

- Joint source-channel coding under strict delay constraints

The joint source-channel coding technique is also an interesting research area for providing reliable transmissions of real-time services. The studies have shown that it can improve the performance of the error control scheme with source coding or channel coding alone and is worthy of further research [Liu06]. The AHEC scheme proposed in this thesis can be viewed as a solution with pure channel coding techniques. Therefore, the performance of the AHEC scheme can be improved by combing with source coding techniques. For example, when the multimedia data is transported in type of a MPEG stream, it can be divided into different classes according to the importance of video data in a coded stream. Accordingly, Unequal Error Protection (UEP) based error control scheme can be employed for the transmission of MPEG streams [Cai00] [Maj02] [Sch01] [Ziv02]. Obviously, the UEP based error control scheme is based on concrete source coding techniques. In this case, we then can integrate the UEP based error control scheme in the general architecture to improve the performance of the AHEC scheme.

- Cross-layer design and optimization under strict delay constraints

In most existing packet-based networks, different error protection strategies exist at different layers (e.g. physical layer, MAC layer, transport layer, and application layer etc.) of the protocol stack. Therefore, a cross-layer design is desirable to provide an optimal overall performance for the transmission of real-time services. For example, in recent years, some researchers have focused on the cross-layer solutions for the reliable transmission of video streaming over wireless networks [Baj07] [Bou07] [Dja07] [Sch03] [Sha05]. Up to now, however, all of them do not take the strict delay constraints as a fundamental limit in the cross-layer solutions. It thus needs further research for the issues of the cross-layer design and optimization under strict delay constraints. As described in Chapter 3, our solution in the thesis can be viewed as a kind of solution in Layer-2 or upper layers. Therefore, it will be an attractive research direction by extending the general framework for one layer (e.g. Layer-2, or any of upper layers) to cross-layer design and optimization under strict delay constraints.

- Develop the network error correction coding theory under strict delay constraints

Network Coding (NC) [Ahl00] is a new research area that has attracted significant interest from engineers and computer scientists. In recent years, many studies have shown that, in emerging areas such as Ad-hoc networks, P2P networks etc., the multicast services can benefit from NC not only in throughput improvements and a high degree of robustness [Wan07a] [Wan07b], but also in establishing minimum-cost multicast connections[Lun06] [Wu07]. When NC is employed for the multicast services in packet erasure networks, the existing solutions for reliable multicast have to be re-thought. For example, the general architecture of EER proposed in this thesis actually can be viewed as a kind of end-to-end solution: Packets are encoded only at the sender and decoded at the end receivers. If intermediated nodes existed in the multicast scenario, they are only allowed to replicate and forward packets. When applying NC for this case, however, the intermediate nodes are allowed to do linear combinations of previously received packets and then send them out. This feature has attracted more and more attention by researchers. Many of them have begin to develop a novel error correction theory to improve the reliability of multicast with NC, which results in the emergence of the concept of Network Error Correction Coding (NECC) proposed by Cai and Yeung [Cai02] [Cai06] [Yeu06]. Recently, the theory of NECC has been largely developed by Koetter, Silva, Zhang etc. [Koe07] [Sil07] [Zha06b]. A good overview of the recent progresses on the NECC theory can be found in [Zha08]. Up to now, however, the existing NECC theory does not consider the strict delay constraints as a fundamental limit. Therefore, our work in the thesis can be utilized as an initial work for developing the NECC theory under the strict delay constraints, which is of great interest for improving the performance of real-time multicast services over packet-based networks with NC.

12.3. Publications

The final section gives a brief summary on the publications that resulted from the thesis work:

- Journals

Parts of this thesis work have been published in the following two journals: *IEEE Transactions on Broadcasting*, March, 2007 [Tan07a]; and the *International Journal of Communications, Network and System Sciences,* June, 2008 [Tan08a].

- Conferences

Parts of this thesis work have been presented at the following conferences: *the* 3rd *and the* 4th *IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, China (i.e. WiCOM2007 and WiCOM2008) [Tan07c] [Tan08c]; and the *European Wireless Conference* 2008, Prague, Czech Republic (i.e. EW2008) [Tan08b].

- Symposiums

Parts of this thesis work have been presented at the following two symposiums: *the* 1st *and the* 2nd *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (*ISBMSB*), USA [Tan06] [Tan07b].

Appendix A

PDF of $\mathrm{N}^1_{req,\max}$

To derive the PDF of $\mathrm{N}^1_{req,\max}$, first, we define two basic probabilities: One is the probability of $\mathrm{N}_{req}(j,1)$ being i, which is denoted by $P^{=i}_{req}(j)$; the other is the probability of $\mathrm{N}_{req}(j,1)$ being less than i, which is denoted by $P^{<i}_{req}(j)$. Using (6.1), these two probabilities can be calculated as follows, respectively:

$$P^{=i}_{req}(j) = \Pr(\mathrm{N}_{req}(j,1) = i) = P(N_p + i, N_{blk}, CSI(j)) \tag{A.1}$$

$$P^{<i}_{req}(j) = \Pr(\mathrm{N}_{req}(j,1) < i) = \sum_{g=0}^{N_p+i-1} P(g, N_{blk}, CSI(j))$$

Note that different receivers have different channel state information. Therefore, for the convenience of description in the analysis, we need to define some special aggregates. First, we define all of sequence number of the receivers as an aggregate $\mathrm{Z} = \{i | 1 \le i \le N_{recv}\}$. Then, let $\mathrm{Z}(h,m)$ denote an aggregate with h elements representing the result on the m-th way of choosing h elements from the overall N_{recv} elements of Z without repetition. Obviously, $\mathrm{Z}(h,m)$ is a subset of Z, where $0 \le h \le N_{recv}$ and $1 \le m \le \binom{N_{recv}}{h}$. Let $e(a,m)$ (where $1 \le a \le h$) denotes the a-th element of $\mathrm{Z}(h,m)$, then we have $\mathrm{Z}(h,m) = \{e(a,m) | 1 \le a \le h\}$.

On the other hand, let $\overline{\mathrm{Z}}(h,m)$ denote the complement of $\mathrm{Z}(h,m)$ in the aggregate Z. Obviously, $\overline{\mathrm{Z}}(h,m)$ is an aggregate with N_{recv}-h elements, which is also a subset of Z. Let $\bar{e}(b,m)$ (where $1 \le b \le N_{recv}-h$) denote the b-th element of $\overline{\mathrm{Z}}(h,m)$, then we have: $\overline{\mathrm{Z}}(h,m) = \{\bar{e}(b,m) | 1 \le b \le N_{recv} - h\}$.

Now set $\mathrm{N}^1_{req,\max} = i$ (where $1 \le i \le k$), which means that in the first transmission there are at least h receivers (where $1 \le h \le N_{recv}$) lost N_p+i packets and the other N_{recv}-h receivers lost less than N_p+i packets. Let $P_{N_{req,\max}}(i,h)$ be the probability of h receivers lost N_p+i packets and the other N_{recv}-h receivers lost less than N_p+i packets. Based on the definitions above, $P_{N_{req,\max}}(i,h)$ can be calculated by:

$$P_{N_{req,\max}}(i,h) = \sum_{g=1}^{\binom{N_{recv}}{h}} \prod_{a=1}^{h} P_{req}^{=i}(e(a,g)) \prod_{b=1}^{N_{recv}-h} P_{req}^{<i}(\bar{e}(b,g)) \tag{A.2}$$

Then, Let $P^i_{N_{req,\max}}$ denote the probability of $\Pr(\mathrm{N}^1_{req,\max} = i)$ (where $1 \le i \le k$). Upon (A.2), we can get the PDF of $\mathrm{N}^1_{req,\max}$ directly:

$$P^i_{N_{req,\max}} = \Pr(\mathrm{N}^1_{req,\max} = i) = \sum_{h=1}^{N_{recv}} P_{N_{req,\max}}(i,h) \tag{A.3}$$

Appendix B

Derivation of $P_{req}(i,c,j)$

In this appendix, let's consider the probability $P_{req}(i,c,j)$ for the j-th receiver. Now we define an aggregate $\mathrm{Z}^j = \mathrm{Z} - \{j\}$ with N_{recv}-1 elements. Similar to the definitions of $\mathrm{Z}(h,m)$ and $\overline{\mathrm{Z}}(h,m)$ in the Appendix A, let $\mathrm{Z}^j(h,m) = \{e^j(i,m) \mid 1 \le i \le h\}$ denote the aggregate with h elements representing the result on the m-th way of choosing h elements from the overall N_{recv} -1 elements in Z^j without repetition, and let $\overline{\mathrm{Z}}^j(h,m) = \{\overline{e}^j(i,m) \mid 1 \le i \le N_{recv} - h - 1\}$ denote the complement of $\mathrm{Z}^j(h,m)$ in Z^j. Then in Z^j, let $P^i_{N_{req,\max}}(h,j)$ be the probability of h receivers lost N_p+i packets and the other $N_{recv} - h - 1$ receivers lost less than N_p+i packets, which can be obtained by:

$$P^i_{N_{req,\max}}(h,j) = \sum_{g=1}^{\binom{N_{recv}-1}{h}} \prod_{a=1}^{h} P^{=i}_{req}(e^j(a,g)) \prod_{b=1}^{N_{recv}-h-1} P^{<i}_{req}(\overline{e}^j(b,g)) \tag{B.1}$$

Note that the calculation of $P_{req}(i,c,j)$ should be divided into two parts due to two different cases in scenarios:

1. The one part is the probability of $\Pr(\mathrm{N}^1_{req,\max}=i,\mathrm{N}_{req}(j,1)=c)$ with i=c, in which case the number of missing packets are no more than N_p+c in one block for any receiver among all of the N_{recv}-1 receivers in Z^j. Using (B.1), $P_{req}(i,c,j)$ can be expressed as:

$$\begin{aligned}P_{req}(c,c,j)&=\Pr(\mathrm{N}^1_{req,\max}=c,\mathrm{N}_{req}(j,1)=c)\\&=P(N_p+c,N_{blk},CSI(j))\sum_{h=0}^{N_{recv}-1}P^c_{N_{req,\max}}(h,j)\end{aligned} \tag{B.2}$$

2. The other part is the probability of $\Pr(\mathrm{N}^1_{req,\max}=i,\mathrm{N}_{req}(j,1)=c)$ with i> c, in which case at least one receiver among all of the N_{recv}-1 receivers in Z^j loses N_p+ i packets in one block and all of the other receivers lose less than N_p+ i packets in the block. Similarly, using (B.1), $P_{req}(i,c,j)$ can be expressed as:

$$\begin{aligned}P_{req}(i,c,j)&=\Pr(\mathrm{N}^1_{req,\max}=i,\mathrm{N}_{req}(j,1)=c)\\&=P(N_p+c,N_{blk},CSI(j))\sum_{h=1}^{N_{recv}-1}P^i_{N_{req,\max}}(h,j)\end{aligned} \tag{B.3}$$

To integrate (B.2) and (B.3) for the expression of $P_{req}(i,c,j)$, we define a function $f_{cmr}(x1,x2)$ with two parameters (where $x1$≥$x2$) as follows:

$$f_{cmr}(x1,x2)=\begin{cases}0, x1=x2\\1, x1>x2\end{cases} \tag{B.4}$$

Based on (B.2), (B.3) and (B.4), finally, the calculation of $P_{req}(i,c,j)$ can be expressed as the following form:

$$P_{req}(i,c,j)=P(N_p+c,N_{blk},CSI(j))\sum_{h=f_{cmr}(i,c)}^{N_{recv}-1}P^{\max(i,c)}_{N_{req,\max}}(h,j) \tag{B.5}$$

Appendix C

PDF of $\mathrm{N}_{req,\max}^{w}$ with *w*>1

From Appendix A, we know that the calculation of the accurate PDF of $\mathrm{N}_{req,\max}^{1}$ is very complex due to the assemble operations. Apparently, the calculation of the accurate PDF of $\mathrm{N}_{req,\max}^{w}$ with w>1 will be much more complicated than that for the accurate PDF of $\mathrm{N}_{req,\max}^{1}$. As a matter of fact, in the retransmission round with w>1, the needed RI is much less than that in the previous transmission stages. Therefore, we can make an assumption to simplify the calculation of the PDF of $\mathrm{N}_{req,\max}^{w}$ with w>1. That is, for each receiver in the multicast scenario with N_{recv} receivers, using (8.2), it is assumed that all of the retransmission packets in the w-th retransmission round have the same average loss probability as follows:

$$\hat{P}_B(w) = \frac{\sum_{j=1}^{N_{recv}} P_B(w,j)}{N_{recv}} \tag{C.1}$$

Note that this assumption does not influence the accuracy of the final result very much, because the needed RI with w>1 occupies only a very small part in the whole total needed RI. Then, for each receiver, let $P_{los}(i,l,w)$ be the probability of i retransmission packets lost with total l retransmission packets retransmitted at the sender in the w-th retransmission round:

$$P_{los}(i,l,w)=\binom{l}{i}\left(\hat{P}_B(w)\right)^i\left(1-\hat{P}_B(w)\right)^{l-i} \tag{C.2}$$

Depending on (C.2), we define two basic probabilities for each receiver in the case of total l redundant packets retransmitted at the sender for one encoding block in the w-th retransmission round: One is the probability of the number of redundant packets lost being i in all of the l redundant packets, which is denoted by $P^{=i}_{los}(l,w)$ and given by:

$$P^{=i}_{los}(l,w)=P_{los}(i,l,w) \tag{C.3}$$

The other is the probability of the number of redundant packets being less than i in all of the l redundant packets, which is denoted by $P^{<i}_{los}(l,w)$ and given by:

$$P^{<i}_{los}(l,w)=\sum_{m=0}^{i-1}P_{los}(m,l,w) \tag{C.4}$$

Without loss of generality, suppose that there are l redundant packets transmitted at the sender for one encoding block in the w-th retransmission round. After those l redundant packets experienced the w-th retransmission round, let $P_{los,max}(i,l,h,g,w)$ be the probability of exactly h receivers with i redundant packets lost while other receivers with less than i redundant packets lost among g receivers. Relying on (C.3) and (C.4), we thus have:

$$P_{los,\max}(i,l,h,g,w)=\binom{g}{h}\left(P^{=i}_{los}(l,w)\right)^h\left(P^{<i}_{los}(l,w)\right)^{g-h} \tag{C.5}$$

Furthermore, we make another assumption: The value of $\mathrm{N}^{w}_{req,\max}$ is only determined by those receivers with $\mathrm{N}_{req}(j,w-1)=\mathrm{N}^{w-1}_{req,\max}$. Because the receivers with $\mathrm{N}_{req}(j,w-1)=\mathrm{N}^{w-1}_{req,\max}$ obviously have more chance to decide the value of $\mathrm{N}^{w}_{req,\max}$ than those receivers required less number of redundant packets, this assumption is thus reasonable in real systems. Under this assumption, afterwards, we have the following lemma:

***Lemma* 2**: *Let $P_{req,max}(i,h,w)$ be the probability of there being exactly h receivers with $\mathrm{N}_{req}(j,w)=\mathrm{N}^{w}_{req,\max}=i$ for the w-th retransmission round among all of the N_{recv} receivers. Under the assumption above, using* (C.5)*, for w*=2,3,…*, this probability can be recursively calculated by:*

$$P_{req,\max}(i,h,w)=\sum_{m=i}^{k}\sum_{g=h}^{N_{recv}}P_{req,\max}(m,g,w-1)\cdot P_{los,\max}\left((N_{cc}^{w-1}-1)\cdot m+i,N_{cc}^{w-1}\cdot m,h,g,w-1\right) \tag{C.6}$$

Where i=1,2,…,*k*; *h*=1,2,…,N_{recv}

Initialization:

$$P_{req,\max}(i,h,1)=P_{N_{req,\max}}(i,h) \tag{C.7}$$

Proof: For the first retransmission round (i.e. w=1), first, there will be exactly h receivers requiring i redundant packets if there are exactly h receivers with $\mathrm{N}_{req}(j,1)=\mathrm{N}^{1}_{req,\max}=i$, which probability is identical to $P_{N_{req,\max}}(i,h)$ as defined by (A.2). For w≥2, then, suppose that for the w-th retransmission round there are exactly h (where 1≤h≤N_{recv}) receivers with $\mathrm{N}_{req}(j,w)=\mathrm{N}^{w}_{req,\max}=i$ (where 1≤i≤k) while other receivers required less than i redundant packets. That indicates, for the (w-1)-th retransmission round, that there are g receivers (where h≤g≤N_{recv}) with $\mathrm{N}_{req}(j,w-1)=\mathrm{N}^{w-1}_{req,\max}=m$ (where i≤m≤k), which probability obviously is $P_{req,\max}(m,g,w-1)$. Then, in case of $\mathrm{N}^{w-1}_{req,\max}=m$, the sender will transmit $N_{cc}^{w-1}\cdot m$ additional redundant packets for this encoding block in this retransmission round. Furthermore, note that the value of $\mathrm{N}^{w}_{req,\max}$ only depends on those g receivers with $\mathrm{N}_{req}(j,w-1)=\mathrm{N}^{w-1}_{req,\max}$ under the assumption. Therefore, after those $N_{cc}^{w-1}\cdot m$ additional redundant packets experienced the (w-1)-th retransmission round, in case of there being exactly h receivers with $(N_{cc}^{w-1}-1)\cdot m+i$ redundant packets lost among those g receivers (i.e. those h receivers have received exactly m-i redundant packets in the (w-1)-th retransmission round), there will be exactly h receivers with $\mathrm{N}_{req}(j,w)=\mathrm{N}_{req,\max}^{w}=i$ in the w-th retransmission round. This probability can exactly expressed as $P_{los,\max}\left((N_{cc}^{w-1}-1)\cdot m+i, N_{cc}^{w-1}\cdot m, h, g, w-1\right)$, which can be calculated by (C.5). Combining this probability with $P_{req,\max}(m,g,w-1)$ for all kinds of conditions, therefore, we can obtain the $P_{req,\max}(i,h,w)$ as shown in (C.6) immediately, which proves the lemma.

Using ***Lemma* 2**, finally, we have the following lemma for calculating the PDF of $\mathrm{N}_{req,\max}^{w}$ with w>1:

***Lemma* 3**: *Let* $P_{\mathrm{N}_{req,\max}}^{i}(w)$ *denote the probability of* $\Pr(\mathrm{N}_{req,\max}^{w}=i)$. *For the first retransmission round* (i.e. w=1)*, we have:*

$$P_{\mathrm{N}_{req,\max}}^{i}(1)=\Pr(\mathrm{N}_{req,\max}^{1}=i)=P_{N_{req,\max}}^{i} \tag{C.8}$$

For w=2,3,…, *using* ***lemma* 2** *and* (C.5)*, we have:*

$$\begin{aligned} &P_{\mathrm{N}_{req,\max}}^{i}(w)=\Pr(\mathrm{N}_{req,\max}^{w}=i) \\ &=\sum_{m=i}^{k}\sum_{g=1}^{N_{recv}} P_{req,\max}(m,g,w-1)\left(\sum_{h=1}^{g} P_{los,\max}\left((N_{cc}^{w-1}-1)\cdot m+i, N_{cc}^{w-1}\cdot m, h, g, w-1\right)\right) \end{aligned} \tag{C.9}$$

Where i=1,2,…,k

Proof: For the first retransmission round, the result is straightforward and hence its proof is omitted. For $w\geq 2$, suppose that $\mathrm{N}_{req,\max}^{w}=i$ (where $1\leq i\leq k$), it means that in the (w-1)-th retransmission round there are g (where $1\leq g\leq N_{recv}$) receivers with $\mathrm{N}_{req}(j,w-1)=\mathrm{N}_{req,\max}^{w-1}=m$ (where $i\leq m\leq k$), which probability can be obtained immediately according to ***Lemma 2*** and denoted by $P_{req,max}(m,g,w-1)$. Then, in case of $\mathrm{N}_{req,\max}^{w-1}=m$, the sender will transmit additional $N_{cc}^{w-1}\cdot m$ redundant packets for this encoding block in this retransmission round. Furthermore, note that the value of $\mathrm{N}_{req,\max}^{w}$ only depends on those g receivers with $\mathrm{N}_{req}(j,w-1)=\mathrm{N}_{req,\max}^{w-1}$. Therefore, in the (w-1)-th retransmission round, when there are at least one receiver with $(N_{cc}^{w-1}-1)\cdot m+i$ redundant packets lost (i.e. at least one receiver received only *m*-*i* redundant packets in the (*w*-1)-th retransmission round) among those *g* receivers, the maximum number of redundant packets required for the *w*-th retransmission round will be *i* (i.e. $\mathrm{N}_{req,\max}^{w}=i$), which probability can be obtained by calculating $\sum_{h=1}^{g} P_{los,\max}\left((N_{cc}^{w-1}-1)\cdot m+i, N_{cc}^{w-1}\cdot m, h, g, w-1\right)$ using (C.5). Combining this probability with the $P_{req,max}(m,g,w\text{-}1)$ derived from ***Lemma* 2**, we get the $P_{\mathrm{N}_{req,\max}}^{i}(w)$ as shown in (C.9) immediately, which completes the proof.

Appendix D

Proof of the Assertion on the Greedy Algorithm

In this Appendix, we will give the complete proof for the assertion on the greedy algorithm presented by Table 9.1. Note, to prove this assertion, we only need to prove that the increased RI caused by the operation of $f_{+1}(\vec{N}_{cc}, N_{rr,\max})$ is always minimal. First, according to the ***lemma* 3** in Appendix C, we know that the PDF of $\mathrm{N}^{w}_{req,\max}$ can be expressed as:

$$\Pr(\mathrm{N}^{w}_{req,\max} = i) = P^{i}_{\mathrm{N}_{req,\max}}(w) \tag{D.1}$$

Where i=1,2,…,*k*

Then, let $f_{+1}(\Delta RI, w)$ be the increased RI by the operation of $f_{+1}(\vec{N}_{cc}, w)$. Based on ***Theorem* 3** presented in Chapter 8.2, we can compute the $f_{+1}(\Delta RI, w)$ as follows:

$$\begin{aligned} f_{+1}(\Delta RI, w) &= \frac{N_p}{k} + \frac{1}{k}\left(\sum_{q=1}^{w-1}\sum_{i=1}^{k} i \cdot N^{q}_{cc} \cdot P^{i}_{\mathrm{N}_{req,\max}}(q) + \sum_{i=1}^{k} i \cdot (N^{w}_{cc}+1) \cdot P^{i}_{\mathrm{N}_{req,\max}}(w) + \right. \\ &\left. \sum_{q=w+1}^{N_{rr,\max}}\sum_{i=1}^{k} i \cdot N^{q}_{cc} \cdot P^{i}_{\mathrm{N}_{req,\max}}(q)\right) - \left(\frac{N_p}{k} + \frac{1}{k}\sum_{q=1}^{N_{rr,\max}}\sum_{i=1}^{k} i \cdot N^{q}_{cc} \cdot P^{i}_{\mathrm{N}_{req,\max}}(q)\right) \\ &= \frac{1}{k}\sum_{i=1}^{k} i \cdot P^{i}_{\mathrm{N}_{req,\max}}(w) = \frac{1}{k} E(\mathrm{N}^{w}_{req,\max}) \end{aligned} \tag{D.2}$$

Now let $P_{\mathrm{N}_{req,\max}}(i,s,w|w-1)$ be the probability of $\Pr(\mathrm{N}^{w}_{req,\max}=i|\mathrm{N}^{w-1}_{req,\max}=s)$. According to this definition, for any positive integer s of no more than k (i.e. $0<s\leq k$), the PDF of $P_{\mathrm{N}_{req,\max}}(i,s,w|w-1)$ will satisfy:

$$P_{\mathrm{N}_{req,\max}}(i,s,w|w-1)=\Pr(\mathrm{N}^{w}_{req,\max}=i|\mathrm{N}^{w-1}_{req,\max}=s)=\begin{cases}0, & i>s\\ p''(i), & i\leq s\end{cases} \tag{D.3}$$

Where:

$$0<p''(i)<1$$
$$\sum_{i=0}^{s}p''(i)=1$$

Furthermore, upon the definition of $P_{\mathrm{N}_{req,\max}}(i,s,w|w-1)$ and Bayes' law, for any positive value $i\leqslant k$, we have:

$$P^{i}_{\mathrm{N}_{req,\max}}(w)=\sum_{s=i}^{k}P^{i}_{\mathrm{N}_{req,\max}}(w-1)P_{\mathrm{N}_{req,\max}}(i,s,w|w-1) \tag{D.4}$$

Consequently, using (D.4), the expected value of $E(\mathrm{N}^{w}_{req,\max})$ can be expressed as:

$$E(\mathrm{N}^{w}_{req,\max})=\sum_{i=1}^{k}i\cdot P^{i}_{\mathrm{N}_{req,\max}}(w)=\sum_{i=1}^{k}\sum_{s=i}^{k}i\cdot P^{i}_{\mathrm{N}_{req,\max}}(w-1)P_{\mathrm{N}_{req,\max}}(i,s,w|w-1) \tag{D.5}$$

Through deploying (D.5) and combining the like terms of $P^{i}_{\mathrm{N}_{req,\max}}(w-1)$ (where $1\leq i\leq k$), we then have:

$$\begin{aligned}
E(\mathrm{N}^{w}_{req,\max})&=P^{1}_{\mathrm{N}_{req,\max}}(w-1)P_{\mathrm{N}_{req,\max}}(1,1,w|w-1)+P^{2}_{\mathrm{N}_{req,\max}}(w-1)\sum_{i=1}^{2}i\cdot P_{\mathrm{N}_{req,\max}}(i,2,w|w-1)+\\
&\quad ...+P^{k}_{\mathrm{N}_{req,\max}}(w-1)\sum_{i=1}^{k}i\cdot P_{\mathrm{N}_{req,\max}}(i,k,w|w-1)\\
&<P^{1}_{\mathrm{N}_{req,\max}}(w-1)P_{\mathrm{N}_{req,\max}}(1,1,w|w-1)+P^{2}_{\mathrm{N}_{req,\max}}(w-1)\sum_{i=1}^{2}2\cdot P_{\mathrm{N}_{req,\max}}(i,2,w|w-1)+\\
&\quad ...+P^{k}_{\mathrm{N}_{req,\max}}(w-1)\sum_{i=1}^{k}k\cdot P_{\mathrm{N}_{req,\max}}(i,k,w|w-1)\\
&=P^{1}_{\mathrm{N}_{req,\max}}(w-1)P_{\mathrm{N}_{req,\max}}(1,1,w|w-1)+2\cdot P^{2}_{\mathrm{N}_{req,\max}}(w-1)\sum_{i=1}^{2}P_{\mathrm{N}_{req,\max}}(i,2,w|w-1)+\\
&\quad ...+k\cdot P^{k}_{\mathrm{N}_{req,\max}}(w-1)\sum_{i=1}^{k}P_{\mathrm{N}_{req,\max}}(i,k,w|w-1)
\end{aligned} \tag{D.6}$$

From the features of $P_{\mathrm{N}_{req,\max}}(i,s,w|w-1)$ as shown in (D.3), we know that for any positive integer $s \geq i \geq 1$ it always satisfies $0 < \sum_{i=1}^{s} P_{\mathrm{N}_{req,\max}}(i,s,w|w-1) < 1$. Upon (D.6), we thus have:

$$E(\mathrm{N}_{req,\max}^{w}) < P_{\mathrm{N}_{req,\max}}^{1}(w-1)\cdot 1 + 2\cdot P_{\mathrm{N}_{req,\max}}^{2}(w-1)\cdot 1 + ... + k\cdot P_{\mathrm{N}_{req,\max}}^{k}(w-1)\cdot 1 \tag{D.7}$$
$$= \sum_{i=1}^{k} i\cdot P_{\mathrm{N}_{req,\max}}^{i}(w-1) = E(\mathrm{N}_{req,\max}^{w-1})$$

Therefore, for any positive integer $N_{rr,max}$, the expected value of $\mathrm{N}_{req,\max}^{w}$ is a strict monotonic decreasing function of the variable w, i.e. it satisfy:

$$E(\mathrm{N}_{req,\max}^{N_{rr,\max}}) < E(\mathrm{N}_{req,\max}^{N_{rr,\max}-1}) < ... < E(\mathrm{N}_{req,\max}^{2}) < E(\mathrm{N}_{req,\max}^{1}) \tag{D.8}$$

Relying on (D.2) and (D.8), given a certain value of k, we thus always have:

$$f_{+1}(\Delta RI, N_{rr,\max}) < f_{+1}(\Delta RI, N_{rr,\max}-1) < ... < f_{+1}(\Delta RI, 2) < f_{+1}(\Delta RI, 1) \tag{D.9}$$

Therefore, by the operation of adding one for $\vec{N}_{cc}$ in the last retransmission round, the increased RI is always minimal. That is, the proposed algorithm in Table 9.1 can always make the locally optimum choice by the operation of $f_{+1}(\vec{N}_{cc}, N_{rr,\max})$ at each stage with the hope of finding the global optimum result. It thus is indeed a greedy algorithm and the proof for the assertion is completed.

Bibliography

[3gp05a] 3GPP TS 26.346 V6.1.0, Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service; *Protocols and Codecs*, June 2005.

[3gp05b] 3GPP TS 23.246 V6.9.0, Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service; *Architecture and functional description*, Dec. 2005.

[3gp05c] 3GPP TS 23.228 V7.0.0, Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2 (Release 7), Valbonne, France 2005.

[3gp06] 3GPP2 Std. C.S0054-A v1.0, *CDMA2000 high rate broadcast-multicast packet data air interface specification*, Feb. 2006.

[Ada04] B. Adamson, C. Bormann, M. Handley and J. Macker, "Negative-Acknowledgment (NACK) – Oriented Reliable Multicast (NORM) Protocol," RFC 3940, Nov. 2004.

[Agr08] P. Agrawal, J. H. Yeh, and etc., "IP Multimedia Subsystems in 3GPP and 3GPP2: Overview and Scalability Issues," *IEEE Commun. Mag.*, pp. 138–45, Jan. 2008.

[Ahl00] R.Ahlswede, N.Cai, S.R.Li, and R.W. Yeung, "Network Information Flow," *IEEE Trans. on Information Theory*, vol.46, no.4, pp.1204-1216, July 2000.

[Ana07] B. Anas and S. Takuro, "Adaptive FEC Reliable Multicast MAC Protocol for WLAN," *IEEE 66th Vehicular Technology Conference*, VTC-2007 Fall, pp. 244 – 248, Sept. 30 2007-Oct. 3 2007.

[Ara03] J. Arauz and P. Krishnamurthy, "Markov Modeling of 802.11 Channels," *IEEE VTC 2003-Fall,* vol.2, pp.771-775, Oct. 2003.

[Arm92] S.Armstrong, A.Feirer, and K. Marzullo, "Multicast transport protocol," RFC 1301, Feb. 1992.

[Baj07] Ivan V. Bajić, "Efficient Cross-Layer Error Control for Wireless Video Multicast," *IEEE Trans. Broadcasting*, vol.53, No.1, pp.276-285, Mar. 2007.

[Bou07] A. Bouabdallah, M. Kieffer and etc., "Evaluation of Cross-Layer Reliability Mechanism for Satellite Digital Multimedia Broadcast," *IEEE Trans. Broadcasting*, vol.53, No.1, pp.391-404, Mar. 2007.

[Bra93] R.Braudes and S.Zabele, "Requirements for Multicast protocols," RFC 1458, TASC, Reading, MA, May 1993.

[Bye98] J. Byers, M.Luby, M.mitzenmacher and A.Rege, "A digital fountain approach to reliable distribution of bulk data," in *proceedings of ACM SIGCOMM'98*, 1998.

[Cai00] J. Cai, Q.Zhang, W.Zhu, and C.W.Chen, "An FEC-based error control scheme for wireless MPEG-4 video transmission," in *Proc. IEEE Wireless Communications and Networking Conference*, vol.3, pp.1243-1247, Chicago, USA, Sep.2000.

[Cai02] N. Cai and R. W. Yeung, "Network coding and error correction," in *Proc. 2002 IEEE Inform. Theory Workshop*, pp. 119–122, 20-25 Oct. 2002.

[Cai06] N. Cai and R. W. Yeung, "Network error correction, part II: Lower bounds," *Commun. Inform. Syst.*, vol. 6, no. 1, pp.37–54, 2006.

[Car97] G.Carle and E.W. Biersack, "Survey of Error Recovery Techniques for IP-Based Audio-Visual Multicast Applications," *IEEE Network*, vol.11, no.6, pp.24-36, Nov./Dec. 1997.

[Cha85] D. Chase, "Code combining—A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Commun.*, vol. COM-33, pp. 385–393, May 1985.

[Cha92] S.R.Chandran and S.Lin, "Selective-repeat-ARQ schemes for broadcast links," *IEEE Trans. Commun.,* vol.40, no.1, pp.12-19, Jan.1992.

[Cos98] D.J. Costello, and et al., "Applications of Error-Control Coding," *IEEE Trans. Inform. Theory*, vol. 44, pp.2531–2560, Oct. 1998.

[Den95] H.R. Deng and M.L.Lin, "A Type I Hybrid ARQ system with Adaptive Code Rates," *IEEE Transactions on Communications*, COM-43 (2/3/4):733-737, February/March/April 1995.

[Dja07] I. Djama and T. Ahmed, "A Cross-Layer Interworking of DVB-T and WLAN for Mobile IPTV Service Delivery," *IEEE Trans. Broadcasting*, vol.53, No.1, pp.382-390, Mar. 2007.

[Dja99] H. Djandji, "An Efficient Hybrid ARQ Protocol for Point-to-Multipoint Communication and Its Throughput Performance," *IEEE Trans. Vehicular Technology*, vol.48, No.5, pp.1688-1698, Sep. 1999.

[Dru83] A. Drukarev and D. J. Costello Jr., "Hybrid ARQ error control using sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-29, pp.521–535, July 1983.

[Du05] Qinghe Du and Xi Zhang, "Adaptive Low-Complexity Erasure Correcting Code Based Protocols for QoS Driven Mobile Multicast Services," *IEEE Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 22-24 Aug. 2005.

[Ela98] M. Elaoud and P. Ramanathan, "Adaptive use of error-correcting codes for real-time communication in wireless networks," *IEEE INFOCOM '98*. Vol. 2, pp.548 – 555, 29 March-2 April 1998.

[Ell63] E.O. Elliott, "Estimates of error rate for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol.42, pp.1977-1997, Sept. 1963.

[Ets0a] ETSI EN 300 421, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services.

[Ets0b] ETSI EN 302 307, Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broad-band satellite applications.

[Ets0c] ETSI EN 300 429, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for cable systems.

[Ets0d] ETSI EN 300 744, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television.

[Ets0e] ETSI EN 302 304 V1.1.1 (2004-11), Digital Video Broadcasting (DVB); Transmission System for Hand-held Terminals (DVB-H).

[Ets0f] ETSI TS 102 034 V0.14 (Draft, 2003-05), Digital Video Broadcasting (DVB); Transport of DVB Services over IP-based Networks, Part 1: MPEG-2 Transport Streams.

[Ets0g] ETSI TS 102 034 V1.3.1, Transport of MPEG 2 Transport Stream (TS) based DVB Services over IP Based Networks, DVB Blue Book A086rev5, Oct. 2007.

[Fie97] R. Fielding, UC. Irvine, J. Gettys, J. Mogul, H. Frystyk and T. Berners-Lee, "The Hypertext Transfer Protocol - HTTP/1.1 (RFC: 2068)," tech. rep., Network Working Group, January 1997.

[Flo97] S. Floyd, V. Jacobson, L.Ching-Gung, S.McCanne and L.Zhang, "A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing," *IEEE/ACM Tran. Networking*, vol.5, no.6, pp.784-803, Dec. 1997.

[Fuj04] H. Fujisawa, K. Aoki, M. Yamamoto, and Y. Fujita, **"**Estimation of multicast packet loss characteristic due to collision and loss recovery using FEC on distributed infrastructure wireless LANs"**,** in *Proc. IEEE WCNC* Vol.1, Page(s): 399 – 404, 21-25 March 2004.

[Gal60] R.G.Gallager, "Low Density Parity Check Codes," In PhD thesis, MIT, 1960.

[Gal62] R.G.Gallager, "Low Density Parity Check Codes," *IEEE Transactions on Information Theory*, 8(1), pp.21-28, Jan. 1962.

[Gil60] E.N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol.39, pp.1253-1265, Sept. 1960.

[Gor07] Manuel Gorius, "Application Layer Hybrid Error Coding – (AL-HEC) for In-Home Multicast digital TV Distribution," Master's thesis, Saarland University, Germany, 2007.

[Her08] Thorsten Herfet, "Future Media Internet: Video & Audio transport – A New Paradigm," *Lecture for WS2008/09*, Telecommunications Lab, Saarland university, Germany, 2008.

[Iee0a] IEEE 802.11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications (1999).

[Itu0a] ITU-T Rec.H.262 and ISO/IEC 13818-2 (MPEG2), "Generic Coding of Moving Pictures and Associated Audio Information – Part 2: video," Nov. 1994.

[Itu0b] ITU-T Rec.H.264 and ISO/IEC 14496-10 (MPEG4-AVC), "Advanced Video Coding for Generic Audiovisual Services," July 2005.

[Ji04] P. Ji, B. Liu and et al., "Modeling Frame-level Errors in GSM Wireless Channels," *An International Journal on Performance Evaluation*, Available at www.ComputerSceienceWeb.com, vol.55, pp.165-181, 2004.

[Jin06] N. Jindal and Z.Q.Luo, "Capacity Limits of Multiple Antenna Multicast," *IEEE Int'I. Symp. Infor. Theory,* Seatlle, WA, pp.1841-1845, July 2006.

[Kar03] S. Karande, Syed A. Khayam and et al., "Analysis and Modeling of Errors at the 802.11b Link Layer," *IEEE ICME*, July 2003.

[Kha03] Syed A. Khayam and H. Radha, "Markov-based Modeling of Wireless Local Area Networks" , *Proceedings of the 6th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, San Diego, CA, USA, 2003.

[Koe07] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, pp. 791–795, 24-29 June 2007.

[Kon01] A. Konrad, B.Y. Zhao, A.D. Joseph and R. Ludwig, "A Markov-based Channel Model Algorithm for Wireless Networks," *ACM MSWiM*, July 2001.

[Kon03] A. Konrad, B.Y. Zhao and A.D. Joseph, "A Markov-based Channel Model Algorithm for Wireless Networks," Wireless Networks 9, Kluwer Academic Publishers, pp. 189-199, 2003.

[Lac00] M.S. Lacher, J. Nonnenmacher, and E.W. Biersack, "Performance comparison of centralized versus distributed error recovery for reliable multicast," *IEEE Trans. on Networking,* Vol. 8, No.2, pp. 224-239, April 2000.

[Lee05] I-C. Lee, C.-S. Chang, and C.-M.Lien, "On the throughput of multicasting with incremental forward error correction," *IEEE Trans. on Information Theory,* pp. 900-918, Vol. 51, Issue.3, March 2005.

[Lim95] H.M. de Lima and O.C.M.B. Duarte, "Performance of repeated retransmission GB(N) schemes for point-to-multipoint noisy channels," *Proceedings of the 38th Midwest Symposium on* Circuits and Systems, Vol.1, pp.1-4, Aug. 1995.

[Lin82] S.Lin and P.S: Yu, "A hybrid ARQ scheme with parity retransmission for error control of satellite channels," *IEEE Trans. Commun.*, vol.COM-30, pp.1701-1719, July. 1982.

[Lin83] S. Lin and D. J. Costello, "Error Control Coding: Fundamentals and Applications", Englewood Cliffs, NJ: Prentice-Hall, 1983.

[Liu06] Z. Liu, H. Liu and Y. Wang, "An adaptive joint source and channel coding scheme for H.264 video multicasting over wireless LAN", *Journal of Zhejiang University Science A*, vol.5, pp.837-846, July 2006.

[Lot07] C. Lott, O. Milenkovic and E. Soljanin, "Hybrid ARQ: Theory, State of the ART and Future Directions," *IEEE Information Theory Workshop on IT for Wireless Networks*, Solstrand, July 2007.

[Lub02a] M. Luby, et al., “Layered Coding Transport (LCT) Building Block”**,** *IETF*, RFC3451, 2002.

[Lub02b] M. Luby, et al., “Asynchronous Layered Coding (ALC) Protocol Instantiation”**,** *IETF*, RFC3450, 2002.

[Lub02c] M. Luby, “LT codes,” *In 43rd IEEE Symposium on Foundations in Computer Science*, Nov. 2002.

[Lub08] M. Luby, T. Stockhammer et al., “Application Layer FEC in IPTV Services,” *IEEE Commun. Mag.,* pp.94-101, May 2008.

[Lun06] D.S.Lun, N. Ratnakar and et al., “Minimum-Cost Multicast Over Coded Packet Networks,” *IEEE Trans. Information Theory*, vol.52, no.6, pp.2608-2623, June 2006.

[Luo05] C. Luo, J. Sun and H. Xiong, “Design and Implementation of an End-to-End System for DTV Multicasting to Home Over Broadband IP Access Network,” *IEEE Trans. on Broadcasting*, Vol. 51, Issue 3, Page(s):348 – 353, Sep. 2005.

[Mac05] D.J.C. MacKay, “Fountain Codes,” *Proc. IEEE Commun*., vol.152, no.6, pp.1062-68, Dec. 2005.

[Mac99] D.J.C. MacKay, “Good Error-Correcting Codes Based on Very Sparse Matrices,” *IEEE Tran. Information Theory*, vol.45, no.2, pp.399-431, Mar.1999.

[Maj02] A. Majumdar, D.G.Sachs, I.V.Kozintsev, K.Ramchandran and M.M.Yeung, “Multicast and Unicast Real-Time Video Streaming over Wireless LANs,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol.12, No.6, pp.524-534, June 2002.

[Mor02] R. H. Morelos-Zaragoza, “The Art of Error Correcting Coding,” John Wiley, 2002.

[Mor89] S. Morgera and V. Oduol, “Soft decision decoding applied to the generalized type-II hybrid-ARQ scheme,” *IEEE Trans. Commun.*, vol.37, pp. 393–396, Apr. 1989.

[Mos00] M. Mosko, J.J. Garcia-Luna-Aceves, "An analysis of packet loss correlation in FEC-enhanced multicast trees," 2000 *International Conference on Network Protocols*, pp.151-161, 2000.

[Mus89] M. Mushkin, and I. Bar-David, "Capacity and Coding for the Gilbert-Elliott Channels," *IEEE Trans. on Information Theory*, vol.35, No.6, pp.1277-1290, Nov. 1989.

[Naf08] A. Nafaa, T. Taleb, and L. Murphy, "Forward Error Correction Strategies for Media Streaming over Wireless Networks," *IEEE Comm. Mag.,* Jan. 2008.

[Non97] J. Nonnenmacher, E.W. Biersack and et al., "Parity-Based Loss Recovery for Reliable Multicast Transmission," *ACM SIGCOMM'97*, pp.289-300,Cannes, France, Sept. 1997.

[Non98] J. Nonnenmacher, E.W. Biersack and et al., "Parity-Based Loss Recovery for Reliable Multicast Transmission," *IEEE/ACM Trans. Networking*, Vol. 6, Aug. 1998.

[Ott04] J. Ott, S. Wenger, N. Sato, C. Burmeister, and J. Rey, "Extended RTP profile for RTCP-based feedback," *IETF RFC*, draft-ietf-avt-rtcp-feedback-11.txt, August 2004.

[Pai04] T. Paila, M. Luby, R. Lehtonen, and V. Roca, "FLUTE-File Delivery Over Unidirectional Transport"**,** *IETF*, RFC3926, 2004.

[Par98] K. Park, and W. Wang, "AFEC: An Adaptive Forward Error Correction Protocol for End-to-End Transport of Real-Time Traffic," 7th *International Conference on Computer Communications and Networks*, pp.196-205, 12-15 Oct. 1998.

[Pej96] S. Pejhan, M. Schwartz, and D. Anastassiou, "Error control using retransmission schemes in multicast transport protocols for real-time media**",** *IEEE/ACM transaction on networking*, Volume 4, Issue 3, Page(s):413 – 427, June 1996.

[Pin94] S. Pingali, D. Towsley, and J. Kurose, "A comparison of sender-initiated and receiver-initiated reliable multicast protocols**",** in *Proc ACM SIGMETRICS*, Nashville, TN, May 1994.

[Pro00] J.G. Proakis, Digital Communications. McGraw-Hill, 4th ed., New York, 2000.

[Pur91] M. B. Pursley and S. D. Sandberg, “Incremental redundancy transmission for meteor-burst communications,” *IEEE Trans. Commun.*, vol. 39, pp. 689–702, May 1991.

[Rei06] U.H.Reimers, “DVB-The Family of International Standards for Digital Video Broadcasting,” *Proc. IEEE*, vol.94, no.1, pp.173-182, Jan. 2006.

[Rfc0a] RFC 791, Internet Protocol, IETF, Sep. 1981.

[Rfc0b] RFC 793, Transmission Control Protocol, IETF, 1981.

[Rfc0c] RFC 768, User Datagram Protocol, IETF, Aug. 1980.

[Rfc0d] RFC 2250, RTP Payload Format for MPEG1/MPEG2 Video, IETF, Jan. 1998.

[Riz97] L. Rizzo, “Effective Erasure Codes for Reliable Computer Communication Protocols”, *ACM Computer Communication Review*, Vol.27, n.2, pp.24-36, Apr. 1997.

[Riz98] L. Rizzo, and L.Vicisano, “RMDP: An FEC-based Reliable Multicast Protocol for wireless Environments,” *Mobile Computing and Communications Review*, vol.2, Number 2, April 1998.

[Roc04] V. Roca and C.Neumann, “Design, Evaluation and Comparison of Four Large Block FEC Codecs, LDPC, LDGM, LDGM staircase and LDGM Triangle, plus a Reed-Solomon Small Block FEC codec,” INRIA res.rep. RR-5525, June 2004.

[Rub98] D. Rubenstein, J.Kurose, and D. Towsley, “Real-time Reliable Multicast Using Proactive Forward Error Correction,” Technical Report 98-19, University of Massachusetts, Mar. 1998.

[Rub99] D. Rubenstein, S. Kasera, D. Towsley and J.Kurose, “Improving reliable multicast using active parity encoding services (APES),” *IEEE INFOCOM’99*, New York, vol.3, pp.1248-1255, 21-25 March 1999.

[San90] R.M. Sanders, and A.C. Weaver, "Xpress Transport Protocol (XTP) - A Tutorial," *ACM SIGCOMM Computer Communication Review*, vol.20, Issue 5, pp. 67-80, USA, 1990.

[Sch01] M. van der Schaar and H. Radha, "Unequal packet loss resilience for fine-granular-scalability video," *IEEE Trans. Multimedia*, vol.3, pp.381-394, Dec. 2001.

[Sch03] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive Cross-Layer Protection Strategies for Robust Scalable Video Transmission Over 802.11 WLANs," *IEEE Journal on Selected Areas in Communications*, vol.21, pp.1752-1763, Dec. 2003.

[Sch96] H. Schulzrinne, S. Casner, and et al., "RTP – A Transport Protocol for Real – time Applications," RFC 1889, Jan. 1996.

[Sch98] H. Schulzrinne, A. Rao and R. Lanphier, "The Real Time Streaming Protocol (RFC 2326)," tech. rep., Internet Engineering Task Force (IETF), http://-src.doc.ic.ac.uk/packages/rfc/rfc2326.txt, April 1998.

[Sha05] Y. Shan, "Cross-Layer Techniques for Adaptive Video Streaming over Wireless Networks," *EURASIP Journal on Applied Signal Processing,* vol. 2, pp.220-228, 2005.

[Sha48] Claude E. Shannon: "A Mathematical Theory of Communication", The Bell Systems Technical Journal, Vol. 27, July, October, 1948.

[Sho06] A. Shokrollahi, "Raptor Codes," *IEEE Trans. Information Theory*., vol. 52, no.6, pp.2551-2567, June 2006.

[Sil07] D. Silva and F. R. Kschischang, "Rank-metric codes for priority encoding transmission with network coding," in *Proc.10th Canadian Workshop Inform. Theory*, Edmonton, Alberta, Canada, , pp. 81–84, 6-8 June 2007.

[Sin77] P. Sindhu, "Retransmission error control with memory," *IEEE Trans. Commun*., vol. COM-25, pp. 473–479, May 1977.

[Tan00] C.C. Tan, N.C. Beaulieu, "On First-Order Markov Modeling for the Rayleigh Fading Channel", *IEEE Tran. Com*., Vol.48, No.12, pp.2032-2040, Dec. 2000.

[Tan02] Andrew S. Tanenboum, “Computer Networks”, Third Edition, Prentice Hall International Inc., 2002.

[Tan06] Guoping Tan and Th. Herfet, “RTP-Level Hybrid Error Correction for DVB Systems in Wireless Home Networks”, *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, April, 2006.

[Tan07a] Guoping Tan and Th. Herfet, “Optimization of an RTP Level Hybrid Error Correction Scheme for DVB Services Over Wireless Home Networks Under Strict Delay Constraints,” *IEEE Trans. Broadcasting*, Vol. 53, Issue 1, pp.297 – 307, Mar. 2007.

[Tan07b] Guoping Tan, Th. Herfet and M. Gorius, "Evaluation of the Performance of a Hybrid Error Correction Scheme for DVB Services over IEEE 802.11a", *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (ISBMSB)*, Orlando, USA, March 28th-29th, 2007.

[Tan07c] Guoping Tan and Th. Herfet, "Application Layer Hybrid Error Correction with Reed-Solomon Code for DVB Services over Wireless LANs," *the 3rd IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China, Sep. 2007.

[Tan08a] Guoping Tan and Th. Herfet, “A Novel Adaptive Hybrid Error Correction Scheme for Wireless DVB Services,” *International Journal of Communications, Network and System Sciences*, July 2008.

[Tan08b] Guoping Tan and Th. Herfet, “On the Architecture of Erasure Error Recovery under Strict Delay Constraints,” *European Wireless 2008 (EW2008)*, Prague, June 22nd-25th, 2008.

[Tan08c] Guoping Tan and Th. Herfet, “An Efficient Greedy Algorithm on Optimizing the Architecture of Erasure Error Recovery under Strict Delay Constraints,” *Accepted by the 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2008)*, 2008.

[Tow97] D. Towsley, J. Kurose, and S. Pingali, “ A Comparison of Sender-Initiated and Received-Initiated Reliable Multicast Protocols”, *IEEE JSAC*, vol.15,No.3, pp.398-406, 1997.

[Tri82] S. K. Trivedi, "Probability and Statistics with Reliability, Queuing and Computer Applications", *Prentice-Hall*, Englewood Cliffs, NJ, 1982.

[Tur02] W. Turin, M. Zorzi, "Performance Analysis of Delay-Constrained Communications Over Slow Rayleigh Fading Channels," *IEEE Tran. Wireless Commun*. Vol.1, No.4, pp.801-807, Oct. 2002.

[Tur99] W. Turin, M. Zorzi, "Performance Analysis of Delay-Constrained Communications Over Diverse Burst-Error Channels," *IEEE VTC'99*, pp.1305-1309, 1999.

[Wan07a] M.Wang and B.Li, "Network Coding in Live Peer-to-Peer Streaing," *IEEE Tran. Multimedia*, vol.9, no.8, pp.1554-1567, Dec. 2007.

[Wan07b] M.Wang and B.Li, , "R2: Random Push with Random Network Coding in Live Peer-to-Peer Streaing," *IEEE Journal on Selected Areas in Communications*, vol.25, no.9, pp.1655-1666, Dec. 2007.

[Wan89] J.L. Wang and J.A.Silverster, "Delay minimization of the adaptive go-back-N ARQ protocols," *in Proc. IEEE INFOCOM*, pp. 584-593, 1989.

[Wel82] E.J.Weldom, Jr., "An improved selective repeat ARQ strategy," *IEEE Trans. Commun.*, vol.COM-30, pp.480-486, Mar. 1982.

[Wic94] S. B. Wicker and M. Bartz, "Type-II hybrid-ARQ protocols using punctured MDS codes," *IEEE Trans. Commun*., vol. 42, pp. 1431–1440, Apr. 1994.

[Woz61] J. M. Wozencraft and M. Horstein, "Coding for two-way channels," Tech. Rep. 383, Res. Lab. Electron., MIT, Cambridge, MA, Jan. 1961.

[Wu07] Y. Wu, "Network coding for wireless networks," Microsoft Research Technical Report, MSR-TR-2007-90, July 2007.

[Xia07] Y. Xiao, X. Du et al., "Internet Protocol Television (IPTV): The Killer Application for the Next-Generation Internet," *IEEE Commun. Mag.,* pp.126-134, Nov. 2007.

[Yam80] H. Yamamoto and K. Itoh, "Viterbi decoding algorithm for convolutional codes with repeat request," *IEEE Trans. Inform. Theory*, vol. IT-26, pp.540–547, Sept. 1980.

[Yan93] Q. Yang and V.K. Bhargava, "Delay and Coding Gain Analysis of a Truncated Type-II Hybrid ARQ Protocol," *IEEE Trans. Vehicular Technology*, vol.42, No.1, pp.22-32, Sep. 1993.

[Yee95] J.R. Yee and E.J. Weldon, Jr., "Evaluation of the performance of error-correcting codes on a Gilbert channel"**,** *IEEE Tran. On Communications*, Volume 43, Issue 8, Aug. 1995.

[Yeu06] R. W. Yeung and N. Cai, "Network error correction, part I: Basic concepts and upper bounds," *Commun. Inform. Syst.*, vol. 6, no. 1, pp. 19–36, 2006.

[Zha06a] X. Zhang and Q. Du, "Adaptive Low-Complexity Erasure-Correcting Code-Based Protocols for QoS-Driven Mobile Multicast Services Over Wireless Networks," *IEEE Trans. on Vehicular Technology,* Vol. 55, Issue 5, pp. 1633-1647, Sept. 2006.

[Zha06b] Z. Zhang, "Network error correction coding in packetized networks," in *Proc. 2006 IEEE Inform. Theory Workshop*, Chengdu, China, pp. 433–437, 22-26 Oct. 2006.

[Zha07] X. Zhang and Q. Du, "Cross-Layer Modeling for QoS Driven Multimedia Multicast/Broadcast over Fading Channels in Mobile Wireless Networks," *IEEE Commun. Mag.,* pp.62-70, Aug. 2007.

[Zha08] Z. Zhang, "Some Recent Progresses in Network Error Correction Coding Theory," in *the 4th Workshop on Network Coding, Theory and Applications* (NetCod 2008), Hong Kong, China, 3-4 Jan. 2008.

[Ziv02] A. Ziviani, B.E. Wolfinger, and et. al., "Joint Adoption of QoS Schemes for MPEG Streams," *Kluwer Academic Publishers*, Printed in the Netherlands, Oct. 2002.

[Zor01] M. Zorzi, "Some results on error control for burst-error channels under delay constraints", *IEEE Trans. On. Vehicular Technology*, Vol. 50, Issue 1, Page(s):12 – 24, Jan. 2001.

[Zor95] M. Zorzi, R. R. Rao, and L. B. Milstein, "On the accuracy of a first-order Markov Model for data transmission on fading channels," *Proc. IEEE ICUPC*'95, pp. 211-215, November 1995.

[Zor98] M. Zorzi, R. R. Rao, and L. B. Milstein, "Error Statistics in Data Transmission over Fading Channels," *IEEE Tran. Commun*. Vol.46, No.11, pp.1468-1476, Nov. 1998.