School of Electrical Engineering and Computing

A System for Improving the Quality of Real-Time Services on the Internet

Stephan Bettermann

This thesis is presented for the Degree of

Doctor of Philosophy

of

Curtin University

December 2013

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: …… Date: ……

Abstract

Real-time Internet services are becoming more popular every day, and *Voice over Internet Protocol* (VOIP) is arguably the most popular of these, despite the quality and reliability problems that are so characteristic of VOIP. This thesis proposes to apply a routing technique called *Fully Redundant Dispersity Routing* to VOIP and shows how this mitigates these problems to deliver a premium service that is more equal to traditional telephony than VOIP is currently.

Fully Redundant Dispersity Routing uses the path diversity readily available in the Internet to route complete copies of the data to be communicated over multiple paths. This allows the effect of a failure on a path to be reduced, and possibly even masked completely, by the other paths. Significantly, rather than expecting changes of the Internet that will improve real-time service quality, this approach simply changes the manner in which real-time services use the Internet, leaving the Internet itself to stay the way it is.

First, real VOIP traffic in a commercial call centre is measured (1) to establish a baseline of current quality characteristics against which the effects of Fully Redundant Dispersity Routing may be measured, and (2) as a source of realistic path characteristics. Simulations of various Fully Redundant Dispersity Routing systems that adopt the measured VOIP traffic characteristics then (1) show how this routing technique mitigates quality and reliability problems, and (2) quantify the quality deliverable with the VOIP traffic characteristics measured. For example, quantifying quality as a *Mean Opinion Score* (MOS) estimated from the measurements with the International Telecommunication Union's E-model, slightly more than 1 in every 23 of the VOIP telephone calls measured in the call centre is likely to be perceived to be of a quality with which humans would be less than very satisfied. Simulations carried out for this thesis show that using just two paths adopting the same measurements, Fully Redundant Dispersity Routing may increase quality to reduce that proportion to slightly less than 1 in every 10 000 VOIP telephone calls.

Next, a mathematical model called the *Qualitative Characteristics Estimation Model* (QCE-model) is presented for estimating the quality for Fully Redundant Dispersity Routing systems using the E-model. The QCE-model describes the packet loss and packet loss

burstiness characteristics of each path with a 4-state Markov model. The packet loss and packet loss burstiness characteristics of the Fully Redundant Dispersity Routing system as a whole may then be computed by combining the Markov models of the paths used by the system. From these computed characteristics, the QCE-model may then, along with a delay estimate and knowledge of the media encoding scheme, compute a quality estimate for the system using the E-model. The QCE-model is applied to a mathematical model of path characteristics constructed from the measured VOIP traffic characteristics called the *Packet Loss and Packet Loss Burstiness Model* (PLB-model). Besides demonstrating the relationships between the quality-determining factors for Fully Redundant Dispersity Routing, the application also quantifies the deliverable quality that may be expected for these, even for conditions that have not been observed.

The accuracy of the QCE-model is demonstrated by comparing its results to simulation results, both based on the measured VOIP traffic characteristics, and analytically. While these do show a discrepancy, that discrepancy is so small as to be all but indiscernible to humans. Unsurprisingly, the discrepancy is zero in the absence of threats to quality. However, even when synthesising VOIP telephone calls by adopting measured VOIP characteristics but condensing the quality threats significantly to highlight that discrepancy, 50% of the modelled MOS estimates are within 1.50E−02 of the simulated MOS estimates, and 98% within 5.90E−02. The limitations of the QCE-model are analysed, and the distribution of discrepancies quantified.

Finally, a description of how Fully Redundant Dispersity Routing may be deployed to improve quality in a real setting is offered. Matters such as service provision and management are addressed. This description shows that a system that uses Fully Redundant Dispersity Routing to improve the deliverable quality of VOIP telephony is feasible.

Acknowledgements

I wish to thank, first and foremost, my supervisor Associate Professor Yue Rong for sharing his knowledge, for his level-headed guidance and advice, and for his steadfast support. In particular, I wish to thank him for giving me his time so liberally whenever I asked for it. I am immensely grateful for all of his enthusiastic help. Likewise, I wish to thank my supervisor Professor John Siliquini for his continued guidance, his knowledge, experience and his inspiring insights. I owe deep gratitude to him for generously continuing to give me his time even after leaving the university. His enduring strivings for simplicity have helped greatly in improving the quality of this work.

I am also indebted to Professor Kevin Fynn for his counsel and support. Furthermore, I wish to thank Curtin University for providing me with a supportive research environment and scholarships that helped fund the research.

It is with considerable pleasure that I would like to thank Dr Thomas O'Neill for the support and help that he has given me over the years. Without him convincing me that I could do this, I might never have started this research.

I am genuinely indebted and thankful to all the friends and family that have stood by me throughout this research. I am grateful for your friendship and support. In particular, I owe immense gratitude to my parents Karl-Heinz and Petra Bettermann for their patience, support, encouragement, faith and counsel.

Finally, I owe my deepest gratitude to my two sons Simon and Jaiden Bettermann for showing me what is truly important in life. I wish to thank you both for sharing my joy over every written page, for helping me draw the figures (especially for getting the circles and arrows in figure 13 right), for checking that the figures are in the right place, for being excited about me getting close to finishing, and for helping me carry the thesis to the chancellory for submission. It is with the deepest love that I dedicate this thesis to you.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Glossary

Term	Description
ALG	An *Application Level Gateway* (ALG) injects functionality of one application into another application. For example, by equipping a router with a SIP ALG, the router may be able to support VOIP applications better.
Burst Length	The number of lost and received packets in a loss burst.
Burst Ratio	An indication of the burstiness of loss bursts in relation to independent (that is, random) loss.
Delay	The time that it takes for data to travel from one point to another point. Sometimes also referred to as *latency*.
Delay Variation	The variation in delay of one packet in relation to another. Sometimes also referred to as *jitter*.
E-model	The E-model is a computational model for estimating telephonic speech transmission quality as a MOS objectively from observable telephone system characteristics such as delay, packet loss, packet loss burstiness, and the method used to encode the media containing the speech being transmitted.
FEC	Forward Error Correction (FEC) techniques guard against potential transmission errors by adding redundant information to the data being transmitted. Using this redundant information, receivers may be able to recover data lost during transmission, at least in part.

Term	Description
ITU	The *International Telecommunication Union* (ITU) is a United Nations agency, responsible for information and communication technologies.
ITU-T	The *ITU Telecommunication Standardization Sector* (ITU-T) is a division of the ITU that is responsible for coordinating telecommunications standards.
Jitter	See *delay variation*.
Latency	See *delay*.
Loss Burst	A region in a packet stream of elevated packet loss.
MOS	The *Mean Opinion Score* (MOS) is a measure of the perceived quality of a telephone conversation by a human.
NAT	*Network Address Translation* (NAT) is a protocol for mapping network addresses between address spaces. For example, NAT may be used to map between a private network address and a public network address.
pcap	Packet capturing library for capturing network traffic.
PESQ	The *Perceptual Evaluation of Speech Quality* (PESQ) process is a subjective process for measuring telephonic speech quality.
PLB-model	The *Packet Loss and Packet Loss Burstiness Model* (PLB-model) is a mathematical model constructed in this thesis from VoIP traffic characteristics measured in this thesis in a commercial call centre.
POLQA	The *Perceptual Objective Listening Quality Assessment* (POLQA) process is a subjective process for measuring telephonic speech quality.

Term	Description
QCE-model	The *Qualitative Characteristics Estimation Model* (QCE-model) is a mathematical model developed in this thesis for estimating the characteristics that determine the quality delivered by a fully redundant dispersity routing system.
RTCP	The *RTP Control Protocol* (RTCP) provides information about an RTP stream.
RTP	The *Real-time Transport Protocol* (RTP) is a standard format for communicating media over the Internet.
SDP	The *Session Description Protocol* (SDP) is a protocol that describes information about communication sessions. It may be used by SIP to describe information about the sessions being initiated or managed.
SIP	The *Session Initiation Protocol* (SIP) is a protocol for starting, managing and ending sessions, such as Internet telephony calls.
SSH	The *Secure Shell* (SSH) is a network protocol for secured communications. Significantly it includes tunnelling functionality.
TCP	The *Transmission Control Protocol* (TCP) is a protocol for sending data over the Internet reliably. Specifically, TCP recovers from data being damaged, lost, duplicated, or delivered out of order. It is one of the core protocols in the Internet.
Tunnelling Protocol	The encapsulation of one protocol in another, to allow the encapsulated protocol to be carried as a payload in the encapsulating protocol.

Term	Description
UDP	The *User Datagram Protocol* (UDP) is a protocol for sending messages called *datagrams* without delivery guarantees. Datagrams sent using UDP may get lost, multiple copies may arrive at its destination, and datagrams may arrive out of order. It is one of the core protocols in the Internet.
VOIP	The *Voice over Internet Protocol* (VOIP) term describes approaches that use the Internet Protocol for achieving telephony.
Wall-Clock Resolution	The smallest possible increment in time that may be quantified by the wall-clock.

List of Symbols

Symbol	Description
$\mathcal{A}$	Set of cumulative delay variations for (corresponding) packets in $\mathcal{M}$.
a_i	Cumulative delay variation of packet i, where $1 < i \leq M$; that is, the delay variation of packet i in relation to the first packet.
b	Probability of packet loss. Packet loss occurs in loss (that is, bad) states.
b_{best}	The lowest possible packet loss that a dispersity routing system may deliver.
b_{worst}	The highest possible packet loss that a dispersity routing system may deliver.
BurstR	The E-model burst ratio, an indication of the burstiness of loss bursts in relation to independent (that is, random) loss.
c	The count of lost packets in a call profile. Lost packets are identified using the RTP packet sequence numbers of the packets received and recorded in the call profile.
c_i	The ith element in $\mathcal{C}$ and which describes the number of packets lost on path r_i.
$c_{i,j}$	The count of transitions from state i to j in a call profile.
$\mathcal{C}$	The set of lost packet counts on each path such that c_i, the ith element in $\mathcal{C}$, describes the number of packets lost on path r_i.
$\mathcal{D}$	Set of delays experienced by packets traversing the corresponding paths in $\mathcal{P}$.

Symbol	Description
d_i	Delay applied in addition to l to packet i in a dispersity routing simulation.
e_i	The ith loss burst in $\mathcal{E}$.
$\mathrm{f}(k; \mathcal{C}, n)$	The frequency of k packet losses occurring in a scenario.
g	Probability of receiving a packet successfully. Successful packet receipt occurs in good states.
$\mathcal{G}$	Discrete state space $\mathcal{G} = \{1,2\}$, that is $\mathcal{G} \subset \mathcal{Z}$, represents the receive states Gap Receive and Burst Receive respectively.
$\mathrm{G}_{\mathrm{min}}$	Minimum number of consecutively received packets in a gap.
$j_{i,k}$	Delay variation, also known as jitter, for packet i in relation to packet k, where $1 < i \leq M$, $1 \leq k < M$, and $k < i$.
l	Delay experienced by packets traversing a path due to the latency of that path.
$\mathcal{L}$	Set of indices of the columns in $\mathbf{W}$ that represent state transitions to a loss state on all N paths.
l_i	Delay experienced by packets traversing path $i \in \mathcal{P}$.
$l(i)$	Given $\mathbf{C} = \mathbf{A} \bigotimes \mathbf{B}$, this function computes for index i in $\mathbf{C}$, the corresponding index in $\mathbf{A}$ (the left operand).
L	The length of a call profile, which is the sum of received and lost packets in that call profile.

Symbol	Description
$\mathcal{M}$	A set of packets comprising a message, such as the set of received packets in a call profile.
M	Cardinality of set $\mathcal{M}$; that is, $\lvert\mathcal{M}\rvert$, the number of packets in message $\mathcal{M}$.
m_i	The ith packet in $\mathcal{M}$; that is, the ith sub-message of $\mathcal{M}$.
N	Number of paths used by a dispersity routing system; that is, $N = \lvert\mathcal{P}\rvert$.
$\mathcal{N}$	Set of indices of the rows in $\mathbf{W}$ that do *not* represent a state transition from a loss state on all N paths.
$\mathbf{P}$	State transition matrix $\mathbf{P}$ expressing the state transition probabilities, such that 1 $p_{i,j}$, the element in row i and column j, is the probability of a transition from state $i \in \mathcal{Z}$ to state $j \in \mathcal{Z}$ occurring, and 2 $\sum_{i\in\mathcal{Z}} \sum_{j\in\mathcal{Z}} p_{i,j} = 1$.
$\mathcal{P}$	A set of paths.
p	Probability of transitioning to the bad state from the good state in the 2-state Markov chain used by the E-model.
P(burst)	Probability of a dispersity routing system traversing from a receive state to a loss state.
$p_{i,j}$	Probability of a transition occurring from state i to state j.
$\mathbf{p}_k^{(i)}$	The kth column vector of the ith matrix $\mathbf{P}_i$.

Symbol	Description
$p_{i,j}^{(k)}$	The (i, j)th element of the kth matrix $\mathbf{P}_k$.
$P(\text{loss})$	Probability of packet loss by a dispersity routing system.
$\mathrm{p}(k; \mathcal{C}, n)$	The probability density function of the discrete distribution describing the discrepancy in packet loss of a dispersity routing system.
Q	The delay adopted for the de-dispersion buffer.
$\mathcal{Q}_i$	Arrival time in the de-dispersion buffer of non-lost packet i. This buffer may act as a scheduling queue to compensate for the differences in path delays increasing the probability of delay variation in the packets delivered by the dispersity routing system.
$\mathcal{R}$	Set of receive times for the (corresponding) packets in $\mathcal{M}$.
r_i	The ith path, where $r_i \in \mathcal{P}$.
$r(i)$	Given $\mathbf{C} = \mathbf{A} \bigotimes \mathbf{B}$, this function computes for index i in $\mathbf{C}$, the corresponding index in $\mathbf{B}$ (the right operand).
$\mathcal{S}$	Set of send times for the (corresponding) packets in $\mathcal{M}$.
t_i	The ith point in time.
T	A sequence of truth values (that is, values that may be in one of two states: either true or false), where $\top$ (that is, true) represents a lost packet and $\perp$ (that is, false) a received packet.
$\mathbf{W}$	State transition matrix for a dispersity routing system.
$w_{i,j}^{(x)}$	As a simplification, $w_{i,j}$ in $\mathbf{W}$ may be computed as $w_{i,j}^{(N)}$.

Symbol	Description
$\mathcal{X}$	Discrete state space $\mathcal{X} = \{3,4\}$, that is $\mathcal{X} \subset \mathcal{Z}$, represents the two loss states Burst Loss and Gap Loss respectively,
X	Cardinality of set $\mathcal{X}$; that is, $\lvert\mathcal{X}\rvert$.
$\mathcal{Z}$	Discrete state space $\mathcal{Z} = \{1,2,3,4\}$ represents the states Gap Receive, Burst Receive, Burst Loss and Gap Loss respectively.
Z	Cardinality of set $\mathcal{Z}$; that is, $\lvert\mathcal{Z}\rvert$.
α	A gap comprising of consecutively received packets in a call profile being synthesised. The initial gap, α_0, has between 1 and $G_{\min}$ received packets, subsequent gaps, α_x (where $x > 0$), have $G_{\min}$ received packets.
β	A loss burst in a call profile being synthesised.
δ_x^y	The frequency $\delta_x^y \in \Delta$ of MOS discrepancy $x \in [-4,4]$ occurring, assuming $y \in \{\text{bursty,random}\}$ packet loss.
Δ	The set of frequencies with which MOS discrepancies occur in a particular dispersity routing system.
$\mathcal{E}$	The ordered set of loss bursts extracted from all measured call profiles, where $e_i \in \mathcal{E}$ is the ith loss burst in $\mathcal{E}$.
$\lambda(z_1, z_2, \ldots, z_N)$	Function that computes the index in $\mathbf{W}$ of the Kronecker product of the column vectors $\left\{\mathbf{p}_{z_i}^{(i)} : i = \{1,2,\ldots,N\}\right\}$.
π	A part of a packet, such as one of the 8 bits of an octet of a packet.

Symbol	Description
υ	An encoding of a part, π, of a packet.
$\psi(i)$	Function that quantifies the impact of loss burst e_i.
$\omega(k; \mathcal{C}, n)$	The frequency of k packet losses occurring in a dispersity routing system communicating a stream of n packets using paths losing the packet counts in $\mathcal{C}$.

Chapter 1 • Introduction

Real-time services on the Internet, such as *Voice over Internet Protocol* (VOIP), are rising in popularity. However, the Internet was never designed, nor built, for real-time services. Rather, what the fathers of the Internet sought primarily was resilience to the kinds of attacks feared during the cold war. Consequently the Internet is a best-effort network. Meeting the demands of real-time services was, at the time, simply not a goal.

While that best-effort nature is not a problem in itself (indeed, generally that pragmatic nature is accepted as being characteristic of the Internet), we are accustomed to and expect quality and reliable real-time services, such as those provided by traditional telephony service providers. Unfortunately real-time services are more sensitive to the distortions to which a best-effort network may subject them than non-real-time services are, because real-time services have the additional constraint of *timeliness*. That is, events must occur within a certain period of time because when they do not, quality suffers. Mechanisms, such as *forward error correction* (FEC), that have evolved for dealing with the faults that a best-effort network may bring to bear on non-real-time services, may not be appropriate for real-time services that have that additional constraint of timeliness.

Therefore, a gap exists between (1) the levels of service that real-time services demand from the Internet and (2) the levels of service that the Internet provides consistently. This gap is significant enough for quality and reliability problems to be characteristic of real-time services such as VOIP on the Internet. One way to address these problems without changing the Internet itself (which would be a substantial, if not futile, task), is to change the way that real-time services use the Internet. This thesis proposes to do just that by utilising the *path diversity* readily available in the Internet.

In particular, this thesis concentrates on using path diversity to improve the quality of VOIP on the Internet. There are three reasons for choosing to concentrate on VOIP. First, mature tools are available for measuring VOIP quality, both subjectively and objectively, thus facilitating accurate quantification of quality improvements. Second, real VOIP traffic is readily available for measuring, both (1) offering a baseline against which any improvements

made possible by using path diversity may be measured and (2) making simulations based on that data realistic. Third, currently VOIP is arguably the most popular real-time service.

Simulations presented in this thesis show that harnessing path diversity using a form of routing known as *fully redundant dispersity routing* [1]–[6] can address the quality and reliability problems that are characteristic of real-time services. These simulations draw on actual VOIP traffic data measured for this thesis in a commercial call centre, both for establishing a baseline of current quality performance and as a source of realistic path characteristics. A mathematical model called the *Qualitative Characteristics Estimation Model* (QCE-model) developed in this thesis can predict the quality performance that a fully redundant dispersity routing system is likely to deliver. Simulations and analysis show the accuracy of that model.

Just like the Internet, fully-redundant dispersity routing is a best-effort approach. However, it makes real-time traffic less susceptible to the kinds of distortions possible in a best-effort network, increasing the likelihood that a service delivers an accustomed and expected level of quality. This is accomplished using resources readily available, providing users with increased control over the quality they are likely to experience.

The remainder of this chapter outlines the background of this thesis, states the research goals, enumerates papers published in relation to this work, discusses the contributions made, and concludes with an outline of the structure of the rest of this thesis. In addition to introducing fully redundant dispersity routing as an approach to improving the quality of real-time services on the Internet, the background places this work in context by introducing alternative approaches of exploiting path diversity, and describes how quality may be quantified.

1.1 BACKGROUND AND CONTEXT

Currently VOIP is *not* on par with traditional telephony. That quality and reliability problems are so characteristic of real-time services such as VOIP, that VOIP over the public Internet is all but synonymous with poor telephony, is evidence of this inequality. Improving the network used for real-time services to serve the needs of real-time services better is one approach to improving quality. However, this approach is an option only when the network is

within reach. When the network or even just the used segments are not within reach, such as when they belong to others that are located in different geographical localities and jurisdictions, other approaches for improving quality may be called for.

Ordinarily, data on the Internet travels along a single path from source to destination, even though that path may change at any time and for any number of reasons. This arrangement is entirely rational when seeking to use the Internet efficiently. However, when that pursuit of efficiency prevents satisfaction of other objectives, such as delivering a reliable and quality real-time service, that objective may need to be placed into context. After all, a service that fails to deliver a satisfactory level of service is less desirable than one that succeeds; a telephony service that cannot be used to communicate — *the fundamental purpose for which it exists* — is of little use.

1.1.1 DISPERSITY ROUTING

By actively replicating the data over multiple paths, however, the effect of a failure on one path may be reduced or even masked completely by the other paths [7][8]. In essence, fault-tolerance may be obtained by introducing redundancy through data replication over multiple concurrent paths. Delivering data to its intended destination using multiple paths at the same time is not novel to this thesis. Communicating data from source to destination by sending it towards multiple nodes in the general direction of the destination was first considered as a routing technique called *selective flooding* [9][10]. Despite being dismissed as inefficient, it was also recognised at the time that variations of this technique may be useful. As shown in this thesis, the usage of multiple paths is indeed useful for improving the quality of real-time services.

The first to describe a data communications system that employs, for the benefits that it brings to data communication, multiple concurrent paths to communicate data is generally accepted to be Nicholas Maxemchuk [1]–[6][11]–[14]. Called *dispersity routing*, Maxemchuk identifies three forms:

1. *non-redundant*,
2. *fully redundant*, and
3. *partially redundant*.

Non-redundant dispersity routing seeks to deliver the combined throughput of multiple paths. It does this by dividing the data to be communicated among the paths such that each path is given a subset of the data, and the set of paths collectively communicates a single instance of the data. Effectively inverse multiplexing, this approach makes no attempt to tolerate communication faults such as loss. Its sole goal is increased performance by combining the resources of multiple paths and spreading the data over these paths.

In contrast, fully redundant dispersity routing seeks to tolerate communication faults by replicating the data to be communicated among multiple diverse paths. Each path is given a complete instance of the data to be communicated. Given N paths, the data is sent N times, one instance for each of the N paths. Note that each path is given one instance only. Assuming that paths have independent failure behaviours, actively replicating the data along multiple paths gives these paths the opportunity to reduce, or even mask completely, the effect of a failure on other paths. Because the data is replicated across the paths rather than along them, as is the case with FEC, this approach delivers the timely data redundancy that is so crucial for resilient real-time communications.

Partially redundant dispersity routing combines non-redundant dispersity routing with fully redundant dispersity routing by encoding subsets of the data into blocks using techniques such as erasure codes and then sending these blocks along the set, or subset, of paths. The goal of this approach is to balance the performance gains possible with non-redundant dispersity routing against the quality gains possible using fully redundant dispersity routing.

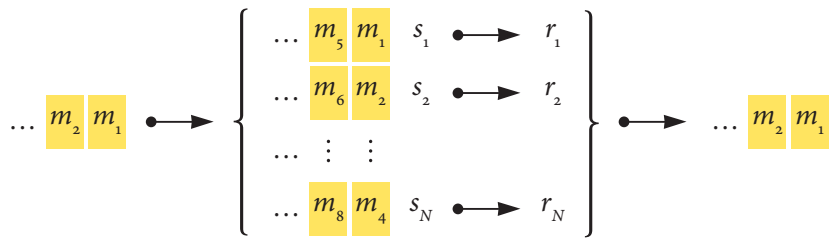

Figure 1: *A non-redundant dispersity routing system of N paths communicating n packets. Each path is given a subset of the data, such that the set of paths collectively communicate a single instance of the data. The dispersity routing system in this example uses N = 4 paths.*

It may be convenient in a packet-switched network for the division into subsets of the data to be communicated to be packet-based. For example, given that a stream $\mathcal{M}$ of M packets $m_1, m_2, \ldots, m_M$ is to be communicated as depicted in figure 1 using a non-redundant dispersity routing system of N paths, partition the packets into N subsets, that is $s_1, s_2, \ldots, s_N$. The first subset, s_1, comprises every Nth packet beginning with the first packet, the second subset, s_2, comprises every Nth packet beginning with the second packet, and so on. Formally, subset $s_i \in \{s_1, s_2, \ldots, s_N\}$ comprises packets $\{m_x \in \mathcal{M} : \forall y \in \mathbb{Z}, y \geq 0, x = yN + i, x \leq M\}$ [3]–[4][6]. Each subset is then given to one of the N paths $r_1, r_2, \ldots, r_N$, for example by giving subset s_i to path r_i. Non-redundant dispersity routing systems are not used in this thesis.

In contrast, in a fully redundant dispersity routing system the data to be communicated does not need to be divided. Rather, the data is given in its entirety to each of the N paths [1][6] which then communicate it, as shown in figure 2. Division is not necessary to achieve fully redundant dispersity routing. However, as this may result in N instances of the data arriving at the recipient, up to $N - 1$ duplicates must be identified and discarded at the recipient. If the data itself does not facilitate the identification of duplicates at the recipient, the data may need to be encapsulated with metadata that does facilitate this identification of duplicates. In a packet-switched network for example, each packet may be encapsulated with an identifier, so that the recipient may deliver the first instance of each encapsulated packet only, discarding the other, duplicate, instances. That is, subset $s_i \in \{s_1, s_2, \ldots, s_N\}$ comprises $\{\langle x, m_x \rangle : m_x \in \mathcal{M}\ \forall y \in \mathbb{Z}, y \geq 0, x = yN + i, x \leq M\}$. At the recipient, only the first instance

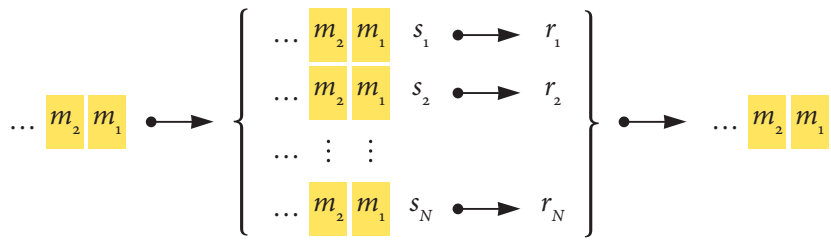

Figure 2: *A fully-redundant dispersity routing system of N paths communicating n packets. Each path is given a complete instance of the data to be communicated, such that the data is sent N times, one instance for each of the N paths.*

of $\langle x, m_x \rangle$ that arrives is accepted, and its contained m_x is then delivered from the system. The other $N - 1$ instances of $\langle x, m_x \rangle$ that may arrive are discarded.

Partially redundant dispersity routing systems, which are also not used in this thesis, may be arranged in various ways. However, the arrangement may be constrained by the number of paths available and the encoding employed, such as a modulo two sum [6], a $(5, 4)$ Hamming code [3]–[4], or a $(7, 4)$ Hamming code [1]–[2][6]. Furthermore, when used for real-time communications, the need for timely delivery of data may further constrain the arrangement. The example depicted in figure 3 of a partially redundant dispersity routing system is based on an example in [6]. It encodes 4 bits (depicted as π) of the data (arriving as packets $m_1, m_2, \ldots, m_M$) into 3 encoded bits (depicted as υ) and gives the block of 4 bits and 3 encoded bits to 7 paths, each path communicating 1 bit. The recipient can decode the block once (1) the 4 bits or (2) any 5 of the 7 bits in the block are received, and then deliver the 4 bits sent.

The concepts of dispersity routing have been used in many ways [11]. For example, non-redundant dispersity routing has been used and adapted in *aggressive transmission* [15], *parallel communications* [16], *inverse multiplexing* [17]–[18], *striping* [19]–[20], *channel striping* [18], *network striping* [21], *channel diversity* [22], *multipath transmission* [23], and in [14][24]–[26]. Similarly, fully redundant dispersity routing, the form used in this thesis, has been employed as *mesh routing* [27] as *simple replication* [28], and in [15][29]–[33], and partially

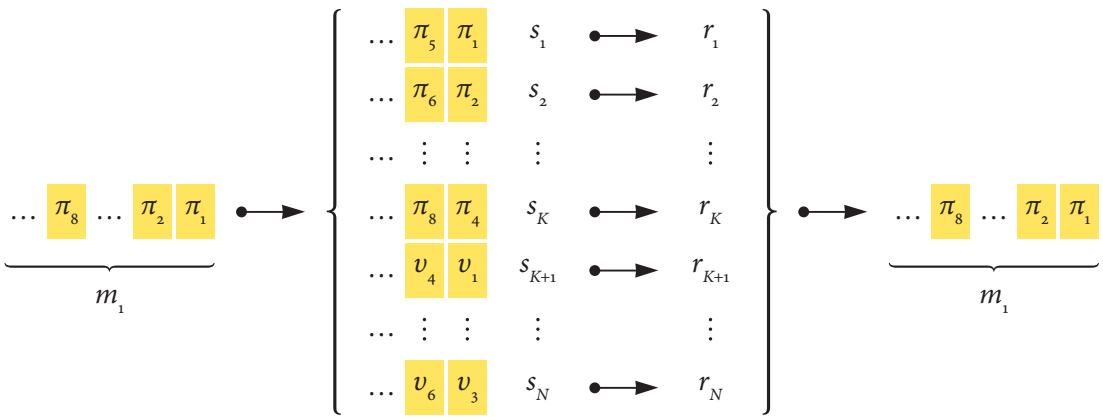

Figure 3: *A partially-redundant dispersity routing system encodes subsets of the data into blocks using techniques such as erasure codes. In this example, $K = 4$ bits (shown as π) are encoded into $N - K = 3$ bits (shown as υ), and the resulting $N = 7$ bits are then sent on the N paths.*

redundant dispersity routing has been used and adapted as *multi-path transmission* [34]–[35] and in [13][28][31][36]–[42]. The work in this thesis differs from these efforts in that it shows how fully-redundant dispersity routing can improve the deliverable quality of real-time communications, focusing in particular on VOIP. Tools are provided for estimating the quality that may be expected from fully-redundant dispersity routing systems, the accuracy of these estimates are quantified, and deployment considerations are presented.

In summary, whereas non-redundant dispersity routing uses additional paths to parallelise data communication for performance gains, fully and partially redundant dispersity routing use additional paths to introduce data redundancy for quality gains. Partially redundant dispersity routing compromises between these two goals. While it may not have been the original intention [12] in [1], dispersity routing may be adapted at the network layer, directing the manner in which packets, rather than bits, are communicated to their destinations. This is similar to the approach taken by [43], who evaluates dispersity routing adapted to operate at the application level. At its most minimalist, each packet may be viewed as a message in its own right to be dispersity routed. In this thesis, dispersity routing is adopted at the network layer, and directs the manner in which packets are communicated. For the sake of brevity, dispersity routing refers to fully redundant dispersity routing for the remainder of this thesis.

1.1.2 ALTERNATIVE APPROACHES TO DISPERSITY ROUTING

Forward error correction and path switching are alternative approaches to dispersity routing for dealing with communication errors, and which just like dispersity routing do not require changes to the Internet itself. This section describes these approaches.

1.1.2.1 FORWARD ERROR CORRECTION

Forward error correction techniques seek to provide data with a degree of resilience against data loss by adding redundant information to it [44]–[46]. Using that redundant information along with the data that is not lost, it may be possible for forward error correction to recover, at least in part, lost data [47].

However, when used to help protect data against loss in a data communications system, clearly the redundant information added for that purpose increases the amount of total data that must be communicated. If communication occurs along a single path, this approach causes increased demand on that path. Indeed, at times that additional demand may be the very cause of the data loss that forward error correction is attempting to recover [45]–[46][48]–[50].

In a real-time setting where data delivery must be timely, the redundant information needed to recover any lost data must be available timely enough to deliver the recovered data in time. When faced with *bursts* of loss, forward error correction techniques may not be able to recover lost data in time, especially when the duration of a loss burst exceeds the time constraint [35][49][51].

1.1.2.2 Path Switching

At any point in time, path switching delivers data from source to destination along a single path only. In addition, path switching at all times maintains a pool of backup paths in preparation for that path developing quality-degrading characteristics. When that occurs, path switching reacts by switching to one of the backup paths in its pool [15][29][52]–[57].

However, switching may also occur pre-emptively before quality degradation occurs, thus avoiding outages that begin with a degradation occurring, continue through the detection of the degradation and end only once the switch completes. For instance, to avoid the outage that would happen when reacting only once degradation has occurred, switching to a backup path may take place when degradation is predicted to occur on the current path. Similarly, switching to a backup path may also occur when its quality-affecting characteristics are superior in the long term to that of the current path.

The ability to predict accurately which paths perform better over long time scales may be sufficiently beneficial to warrant actively probing paths. Even though probing does tax paths, accurate predictions help to make path switching more accurate in its decision (1) whether to switch or not (2) when and (3) to what backup path. Accurate predictions enable path switching to avoid switching from a path that is about to recover from a degradation, to a

backup path that is about to experience the same, or even worse, degradation than the current path.

1.1.3 INCREASING QUALITY BY INCREASING INTERNET USE

Sending multiple copies of the data to be communicated over multiple paths clearly makes more *overall* use of the Internet than sending it just once along a single path. However, that increased usage has purpose: it facilitates the improvement of quality, and that gain in quality may warrant the increase in overall usage.

Considering that the Internet *exists* to be used and that users typically *pay* in one way or another to use it, the decision as to whether that usage (and therefore cost) is warranted or not lies with the user. This is true for any service that uses the Internet, such as other increasingly popular services like Internet television and on-demand video streaming. Compared to the costs of traditional telephony, however, the increase in cost is likely to be negligible. In the end, it is for each individual user to decide if the additional cost incurred by fully redundant dispersity routing is warranted by the gain in quality for a premium real-time service.

Any usage in addition to the minimum that fully redundant dispersity routing imposes on the Internet is parallelised over a set of paths. Compare this to FEC which adds additional data to the only path used, and path switching which may tax paths (1) to determine which paths to select for inclusion in its set of backup paths, and (2) to help predict which path switch is most likely to be beneficial over long time periods.

1.1.4 QUANTIFYING TELEPHONIC SPEECH TRANSMISSION QUALITY

An attractive measure of telephonic speech transmission quality, as perceived by humans, is the *Mean Opinion Score* (MOS) [58]–[65]. What makes it so attractive is the existence of both (1) well-defined processes for measuring that quality as a MOS, and (2) clear semantics that relate the perception of that quality by humans to concise scales of MOS values. Note however that other real-time services may be served better by quality measures other than a MOS.

Fundamentally, a MOS is a number. Some processes for determining the MOS that quantifies the quality of a speech transmission *subjectively* are defined in [59] (although [59] also defines processes for purposes other than determining a MOS). Each of these MOS-determining processes in essence defines an opinion scale that maps a set of *opinions* to a numerical *score* that denotes a quality. To determine the quality of a speech transmission, each individual of a set of humans judges the quality according to that scale, resulting in a set of opinion scores. The *arithmetic mean* of that resulting set of opinion scores is the mean opinion score, the MOS, for that speech transmission, according to that opinion scale.

Most of the scales in [59], such as the *conversation opinion scale* depicted in table 1, have scores in the range 1 to 5, where 1 denotes the lowest and 5 the highest quality possible. However, not all scales are in that range; the scale for the Comparison Category Rating method, for example, is in the range -3 to 3. To identify the process (and, thus, the scale) used to quantify a MOS, the symbol used to denote that MOS is typically suffixed (or postfixed where suffixing is not possible) [59]–[60]. For example, the symbol for the MOS determined using the conversation opinion scale depicted in table 1 is $\mathrm{MOS}_{\mathrm{C}}$. (The exception is the MOS resulting from the *listening-quality scale*, which yields a "*mean listening-quality opinion score*, or simply *mean opinion score* … represented by the symbol MOS" [59].)

Besides the subjective processes in [59] for measuring quality, *objective* processes exist as well. Two such processes, *Perceptual Evaluation of Speech Quality* (PESQ) [66] and

Table 1: *The conversation opinion scale defined in* [59] *for judging the quality of a conversation. Each of a group of humans individually scores the quality by forming an opinion of the connection used. The arithmetic mean of the scores is the Mean Opinion Score (MOS).*

Opinion	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Perceptual Objective Listening Quality Assessment (POLQA) [67], determine quality by comparing (1) the speech *after* transmission (and possibly degraded by that transmission) through the telephone system with (2) the speech *before* it was transmitted (and before possibly being degraded by that transmission) through the telephone system. However, these processes require both the transmitted speech and the speech before transmission. In everyday telephone system usages, the listeners do not have the speech before transmission; if they did, there would be no need for the transmission and, therefore, the possibly degraded speech after transmission.

Another process for measuring quality objectively is the E-model [58]. In contrast to the two processes above however, the E-model does not require the speech before transmission or, indeed, the speech after transmission. Rather, the E-model determines quality from observable telephone system characteristics such as delay, packet loss, packet loss burstiness, and the method used to encode the media being transmitted (that is, the *codec*). From these observable characteristics the E-model may, "for transmission planning purposes" [58], provide "a prediction of the expected quality, as perceived by the user" [63].

The E-model is a computational model that yields a *transmission rating factor* (R-factor), a number in the range 0 to 100, where 0 indicates a very bad quality and 100 a very good

Table 2: *Subset of the E-model parameters from* [58] *used to compute MOS estimates in this thesis. Default values are used for all other parameters. The permitted range is shown for each parameter. For illustration purposes, the default values for these parameters are also shown here.*

Parameter	**Symbol**	**Default Value**	**Permitted Range**
Random Packet-loss Probability	Ppl	0	0 – 20
Burst Ratio	BurstR	1	1 – 2
Absolute Delay in echo-free Connections	Ta	0	0 – 500
Equipment Impairment Factor	Ie	0	0 – 40
Packet-loss Robustness Factor	Bpl	1	1 – 40
Number of Quantization Distortion Units	qdu	1	1 – 14

quality. An estimate of conversational quality (represented by the symbol MOS_{CQE}) may be computed from an R-factor [58]. For the remainder of this thesis, unless qualified otherwise, MOS refers to the conversational quality estimate (MOS_{CQE}) computed with the E-model using the default values defined for each E-model parameter except for those enumerated in table 2. When the value for a parameter is outside the defined range permitted for that parameter, the MOS estimate is computed as undefined, unless explicitly permitted in combination with other parameter values. For example, the burst ratio may exceed 2 when packet loss is less than 2% [58]. That the E-model permits only values in the ranges defined for its parameters is a limitation of the E-model.

However, quality determinations by the E-model are estimates on a scale that "should be viewed as a continuum of perceived quality varying from high quality through medium values to a low quality" [63]. While "the boundaries between different ranges" of quality cannot be fixed [63], [58][62]–[63] nevertheless define a "provisional guide for the relation between" [58] quality scores and categories of speech transmission quality and user satisfaction [68]. This provisional guide is shown in table 3. For a complete definition of the E-model, including its limitations, the reader is referred to [58].

Table 3: *Provisional guide defined in* [58][62]–[63] *for interpretation of MOS estimates. Each MOS estimate shown is the minimum for an interpretation in this provisional guide. Quality rated with a MOS below the lowest minimum shown is not recommended.*

Minimum MOS	Speech Transmission Quality	User Satisfaction
4.34	Best	Very satisfied
4.03	High	Satisfied
3.60	Medium	Some users dissatisfied
3.10	Low	Many users dissatisfied
2.58	Poor	Nearly all users dissatisfied

1.2 RESEARCH GOALS

The primary goal of this research is to investigate and quantify the effectiveness of fully redundant dispersity routing as a system for improving the quality of real-time services on the Internet, focusing in particular on VOIP. In support of this primary goal, the secondary goals of this research then are to (1) develop a new mathematical model that characterises a fully redundant dispersity routing system, (2) quantify the improvements possible using both simulations based on real VOIP traffic measurements and the developed model, (3) establish the accuracy of the model, and (4) outline how fully redundant dispersity routing may be used in a real setting.

Both packet loss and packet loss burstiness probabilities will be considered by the new mathematical model, which has not been done before. In addition to the measured VOIP traffic data, the developed model will be applicable to conditions interpolated and extrapolated from the measured VOIP traffic data to illustrate the quality that may be delivered in conditions that have not yet been observed. The accuracy of the developed mathematical model will be quantified for observed conditions using simulations, and as a probability distribution of discrepancies. Finally, a concrete example will be provided that illustrates how two points on the Internet will be connected using dispersity routing.

1.3 PUBLICATIONS RELATED TO THE THESIS

The following two papers related to this thesis were presented and published at peer reviewed conferences. Both contain material that resulted from the work for this thesis, and this thesis builds on these papers.

1. S. Bettermann and Y. Rong, "Effects of Fully Redundant Dispersity Routing on VOIP Quality," in *Proc. 2011 IEEE Int. Workshop Technical Committee on Communications Quality and Reliability* (*CQR*), Naples, USA, 2011.
2. S. Bettermann and Y. Rong, "Estimating the Deliverable Quality of a Fully Redundant Dispersity Routing System," in *Proc. 17th Asia-Pacific Conf. on Communications* (*APCC*), Kota Kinabalu, Malaysia, 2011.

1.4 CONTRIBUTIONS

This thesis makes a number of contributions. First, it proposes the use of fully redundant dispersity routing to improve the quality of the real-time service on the Internet. The effectiveness of that approach as a system for improving the quality of VOIP, currently a very popular and seemingly a commercially viable real-time service on the Internet, is established.

Next, a new mathematical model that characterises the quality determining characteristics of a fully redundant dispersity routing system is developed, called the *Qualitative Characteristics Estimation Model* (QCE-model). The accuracy of the QCE-model in predicting the deliverable quality that may be expected from a fully redundant dispersity routing system with known path characteristics is established. This is achieved using both simulations and by determining the probability density function of discrepancies between the packet loss probability estimate by the QCE-model and the packet loss probabilities possible in a given configuration. The packet loss probability is the major qualitative characteristic estimated by the QCE-model.

Furthermore, the improvements to VOIP quality that are possible with fully redundant dispersity routing are quantified using both simulations based on real VOIP traffic expressly measured for this thesis and the QCE-model developed in this thesis. The QCE-model is applied to both the VOIP traffic characteristics observed in the VOIP measurements as well as to a mathematical model of path characteristics constructed from the measured VOIP traffic characteristics called the *Packet Loss and Packet Loss Burstiness Model* (PLB-model). In addition to quantifying the effectiveness in improving deliverable quality, the simulations also help to illustrate salient features in how this is accomplished.

The measurements are taken in a real setting where professional staff maintains the environment in an optimum condition. Those measurements serve as a baseline of current performance against which improvements may be measured. Moreover, the measurements are drawn upon by the simulations for the characteristics experienced by packets traversing simulated paths in fully redundant dispersity routing systems by adopting the actual characteristics measured.

Lastly, this thesis presents an outline of how fully redundant dispersity routing may be used in a real setting. In particular, it exemplifies how the developed mathematical model applied to characteristics extrapolated from the measured data may be used as a planning tool.

1.5 OUTLINE OF THESIS

The remainder of this thesis is structured as follows. Following the introduction, chapter 2 describes how actual VOIP traffic data is measured in a commercial call centre for the purposes of (1) establishing a baseline of current performance against which changes may be measured, and (2) providing real measured path characteristics to simulations. Next, chapter 2 explains how fully redundant dispersity routing may be used for real-time communications, details how it may be simulated, and shows how it improves the quality of real-time communications, focusing on VOIP. A concrete simulation using this measured data is used to illustrate and highlight salient features and effects, such as lowering packet loss, packet loss burstiness, delay and delay variation (see section 2.4.2). Three sets of dispersity routing simulations using the measured VOIP traffic data are then used to demonstrate the effectiveness of dispersity routing in improving quality.

The next chapter, chapter 3, looks at how much dispersity routing can improve the quality of VOIP. A new mathematical model called the *Qualitative Characteristics Estimation Model* (QCE-model) is developed for estimating the characteristics affecting the quality that may be expected from any given dispersity routing system. Along with knowledge of the method used to encode the VOIP media being communicated, the E-model may then be used to estimate the VOIP quality deliverable by that dispersity routing system. Next, a new mathematical model called the *Packet Loss and Packet Loss Burstiness Model* (PLB-model) is constructed from the VOIP traffic measurements. The QCE-model is then applied to the PLB-model to estimate the VOIP quality that may be expected from dispersity routing systems of 2 – 6 paths, where the paths adopt characteristics interpolated and extrapolated by the PLB-model from the data measured in the commercial call centre. Together with quality estimates for a single path adopting the same characteristics, this application of the QCE-model achieves two objectives. First, it shows what quality dispersity routing can deliver and

under what conditions. Second, it conveys a sense of the relationships that exist between the parameters affecting the deliverable quality that may be expected of a dispersity routing system. Therefore, this application of the QCE-model may be useful as a planning tool.

Chapter 4 evaluates the accuracy of the QCE-model by comparing the quality estimates computed by simulation with quality estimates computed by the QCE-model for three sets of simulations, all for dispersity routing systems of two paths. The first simulation is of a dispersity routing system adopting every possible combination of the measured VOIP traffic data, the second for 25 sets of dispersity systems with elevated quality threats, and the third for every combination of packet loss possible for a stream of 17 packets. Discrepancies between modelled and simulated quality estimates are analysed, and the distribution of discrepancies is quantified as a probability distribution. Limitations of the QCE-model are examined, as is the impact of the limitations of the E-model on the evaluation of accuracy between the simulated and modelled quality estimates.

Chapter 5 describes how dispersity routing may be deployed in a real-world setting to improve deliverable VOIP quality. Matters such as motivations for using dispersity routing and forms of deployment are discussed. Next, an example is presented that illustrates how two points on the Internet may be connected using dispersity routing. This is followed by an examination of security considerations, service provision and configuration, and an overview of management issues.

Finally, chapter 6 concludes the thesis by summarising the research, itemising the main research contributions and discussing further research.

Chapter 2 • Using Fully Redundant Dispersity Routing

This chapter begins by describing fully redundant dispersity routing in the context of VOIP, and shows how that form of dispersity routing can improve the quality of the VOIP real-time service. It does so using example simulations with the goal of highlighting salient effects on quality-effecting characteristics. For example, one of these characteristics is *delay variation*, that is, the variation in *delay* of one packet in relation to another, where delay is the time that it takes for data to travel from one point to another point. Dispersity routing may reduce delay variation by the competition between the paths. For the remainder of this thesis, unless otherwise specified, for the sake of brevity delay refers to the time that it takes for data to travel from one point to another point.

The majority of the material in this chapter, all of which was produced for this thesis, has been peer reviewed and published in [7]–[8]. However, this chapter extends and builds upon the material presented in [7]–[8] in the following major points:

1. Terminology, mathematical symbols and notation has been harmonised with this thesis.
2. This chapter uses a more recent data set than the one used in [7]–[8]. The VOIP quality observed in this data set is higher than in that observed for [7]–[8], which translates into smaller gains delivered by dispersity routing.
3. This newer data set was collected for this thesis by a custom data collector written specifically for this thesis (see section 2.1), which operates at a lower level than the one used in [7]–[8] resulting in more accurate data.
4. The absence of space constraints in this thesis permits more detail than [7]–[8] in:
   a. examining currently deliverable VOIP quality (see section 2.2),
   b. describing the usage of dispersity routing to improve the quality of VOIP (see section 2.3),
   c. illustrating the effects of dispersity routing on VOIP (see section 2.4), and
   d. exploring the effectiveness of dispersity routing in improving the quality of VOIP (see section 2.5).
5. A new simulation is described in this chapter (see section 2.5.3). This simulation shows the quality improvements possible with dispersity routing, by simulating

exhaustively how dispersity routing could have improved quality in the observed environment if it had been used in that environment.

2.1 MEASURING REAL VOIP TRAFFIC

This section describes how real VOIP traffic was measured for this thesis in a commercial call centre. VOIP traffic was measured for the following reasons. First, it establishes a baseline for comparisons between (1) the quality currently seen in a real VOIP environment that does not use dispersity routing, and (2) the quality possible when using dispersity routing. Second, it offers insight into the actual threats to quality that an approach seeking to improve quality may face. Third, it provides real path characteristics for simulations and modelling.

To measure VOIP traffic, custom software was written that captures packets moving through the network of the telephone system handling a subset of the telephone calls in a commercial call centre. Figure 4 depicts a block diagram of that software. Captured packets are analysed in real-time just sufficiently enough to isolate relevant *Session Initiation Protocol*

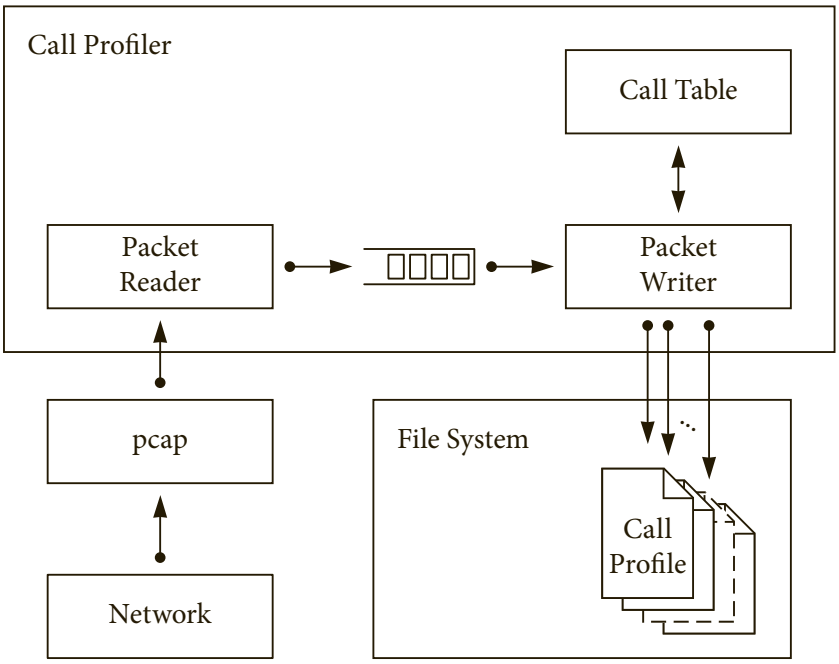

Figure 4: *Block diagram of custom software written to measure VOIP traffic. Packets are read from the network using* pcap*, interesting calls and interesting packets associated with these interesting calls are identified, and for each such call a text file called a* call profile *is created.*

(SIP), *Real-time Transport Protocol* (RTP) and *RTP Control Protocol* (RTCP) packets, identify VOIP telephone calls of interest, attribute the isolated packets to a VOIP telephone call, and to gather pertinent information about each VOIP telephone call (such as a possible delay estimate). VOIP telephone calls of interest are those passing through the public Internet; that is, *incoming* calls from an external source to a local destination, and *outgoing* calls from a local source to an external destination. Similarly, RTP and RTCP packets of interest are those that have passed through the public Internet; that is, packets sent from an external source to a local destination.

For each VOIP telephone call of interest, a file called a *call profile* is written to the file system. An example of a call profile is depicted in the appendix. Each call profile is a text file that contains statistical information about the VOIP characteristics of that VOIP telephone call, as well as contextual information about the VOIP telephone call. In particular, data identifying the participants of the VOIP telephone call and any media data contained in the RTP packets are expressly *not* accessed, stored or altered. Each call profile, besides recording a delay estimate for the VOIP telephone call and contextual information about the VOIP telephone call, primarily contains the data needed to (1) identify lost RTP packets, (2) order RTP packets, and (3) compute the delay variation of any RTP packet in relation to a preceding RTP packet.

To compute a delay estimate for the call, RTCP packets sent by the external participant to the local participant of the call are inspected. From the sender and receiver reports contained in these RTCP packets, *round-trip propagation delays* may be computed [69]. Assuming that the incoming and outgoing paths between the two participants of the call are symmetric, a delay estimate may be computed for the call as a moving average of the halved round-trip propagation delays [69].

2.2 CURRENTLY DELIVERABLE VOIP QUALITY

From the data measurements taken as described above, quality estimates may be computed as MOS estimates using the E-model. This section presents and discusses the quality estimates for the real VOIP traffic measurements taken in a commercial call centre for this thesis in the period beginning 1 August 2011 and ending 31 January 2012.

The telephone system of the call centre is connected to two VOIP service providers through the public Internet across continental Australia. Within the call centre, utilisation of the network hosting the telephone system is negligible. Connection to the public Internet is through a dedicated connection for the telephone system to an Internet service provider. Professional network engineers on staff maintain the telephone system, the network and the connection to the Internet service provider in optimum condition.

Between 1 August 2011 and 31 January 2012, 6265 VOIP telephone calls totalling over 313 hours were measured, encompassing 56 397 249 received RTP packets and RTP packets identified as lost. Every RTP packet found to contain media capable of communicating speech, established by inspecting the payload type (PT) field of the RTP packet header [69] of that RTP packet, contained media encoded as A-law, also known as PCMA. The recorded data comprises 814.5 MB (that is, $814.5 \cdot 10^6$ octets) of maximally compressed call profiles, which, as shown in the appendix, are text files in *Extensible Markup Language* (XML) format.

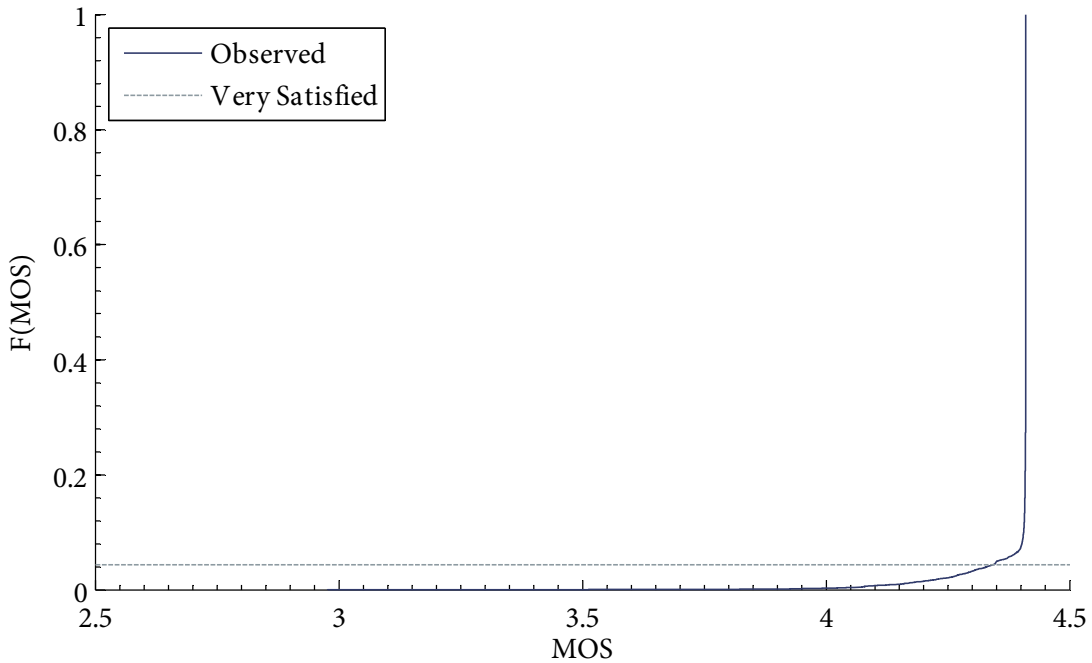

Figure 5: *Cumulative distribution of MOS estimates for measured VOIP telephone calls. A MOS of 4.34 or above may be interpreted as a perceived quality with which users are 'very satisfied'. Approximately 95.64% of VOIP telephone calls measured have an estimated MOS of at least 4.34.*

Figure 5 plots the cumulative distribution of the MOS estimates computed using the E-model for these measured VOIP telephone calls. Also shown in figure 5 is the proportion of VOIP telephone calls with a MOS estimate of at least 4.34. A MOS estimate of 4.34 is the minimum MOS for calls that may be interpreted to be of a quality rating with which users are *very satisfied* [58].

As can be seen, approximately 4.36% of measured VOIP telephone calls — that is, slightly more than 1 in 23 calls — are estimated to have a quality with which users are *less* than very satisfied. Consumers of traditional telephony services, in contrast, are used to the idea that their fixed-line telephone calls are near flawless almost all the time. Therefore, figure 5 supports the view that while in this case most of the time the quality of VOIP telephony is such that most users are very satisfied with it, quality and reliability problems are characteristic of VOIP.

Figure 6 depicts a scatter plot of the probability of packet loss against the MOS estimate for each measured VOIP telephone call. Revealing a trend that follows a MOS curve, the plot

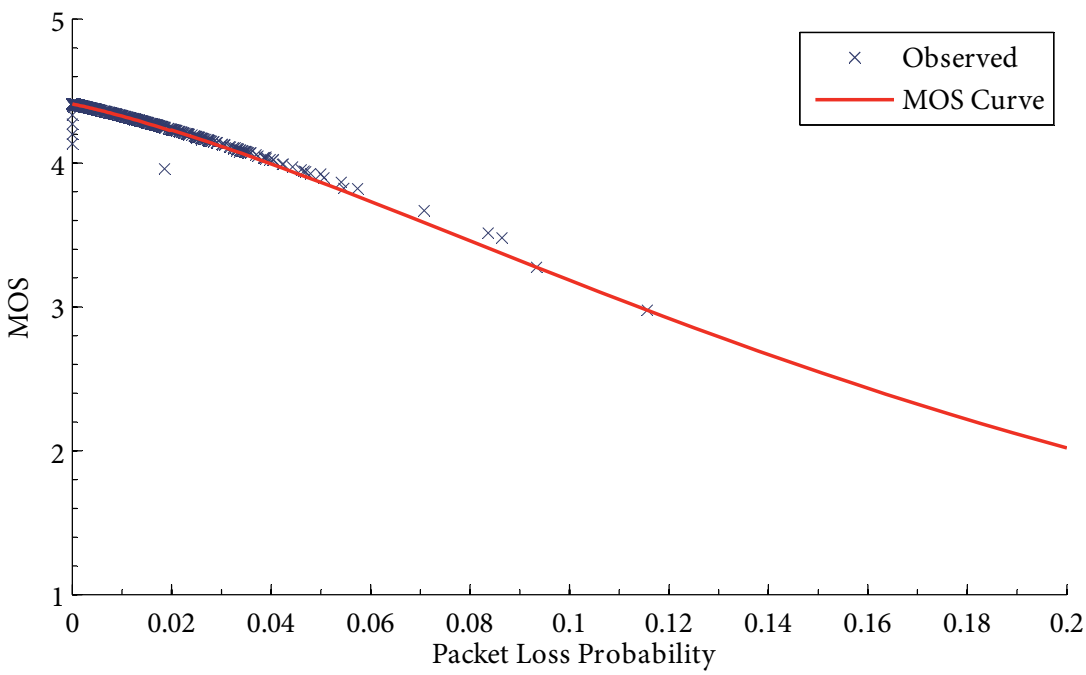

Figure 6: *Scatter plot for measured VOIP telephone calls of observed packet loss probability against estimated MOS. Observable trend follows MOS curve variant on packet loss probability only, using the mean burst ratio and delay for all measured VOIP telephone calls.*

illustrates clearly that the predominant cause of these quality and reliability problems is packet loss. The points below the MOS curve in figure 6 between packet loss probabilities 0 and 0.02 are due to the estimated delay for those VOIP telephone calls exceeding 100 ms, and which thus impact on the MOS estimate. Note that only 0.14% of the measured VOIP telephone calls have a delay estimate exceeding 100 ms, with the 95th percentile of the delays estimated for the measured calls being 57.67 ms, and the 99.7th percentile being 79.14 ms.

As a reference, figure 6 plots a MOS curve for packet loss probabilities in the range 0 to 0.2, with all other parameters of the MOS curve constant. These constant parameters to the MOS curve are as follows. The *Burst Ratio* and *Absolute Delay* parameters are computed as the mean burst ratio and the mean delay estimate respectively for all call profiles [58]. Codec-derived parameters, such as *Equipment Impairment Factor*, *Packet-loss Robustness Factor* and *Number of Quantization Distortion Units*, are derived from the call profiles, which all use the same codec and, therefore, have the same values for these parameters [64]. The remaining parameters adopt the default values as defined by the E-model [58].

2.3 USING DISPERSITY ROUTING FOR VOIP

To improve the deliverable quality of real-time VOIP communications, this thesis proposes to use dispersity routing. This section details how dispersity routing may be used for real-time VOIP communications.

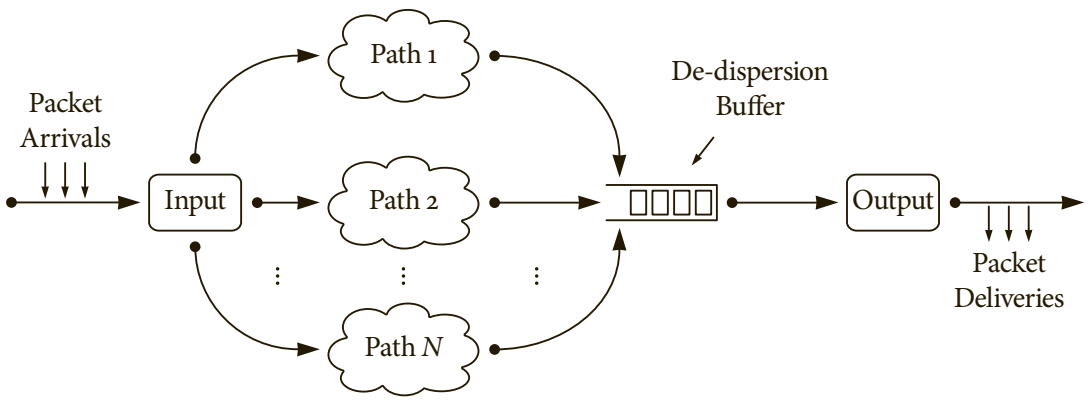

Figure 7: *Dispersity routing system of N paths. Packets enter the dispersity routing system on the left, are encapsulated and traverse all N paths concurrently, pass through a de-dispersion buffer, and leave the dispersity routing system on the right.*

In the interests of simplicity and clarity, this thesis assumes a dispersity routing system for each direction of communication between two fixed points on the Internet engaging in real-time VOIP communications. Figure 7 depicts a dispersity routing system for one direction of communication. Packets from the source enter the dispersity routing system on the left of the figure, are copied and encapsulated (possibly fragmenting the packet) such that for a system of N paths there are exactly N encapsulating instances of each packet. One encapsulating packet is then given to each path for delivery through the dispersity routing system, such that each path is given exactly one instance of every encapsulating packet in the stream.

While traversing a path, a packet may experience any number of events. For example, the packet may be lost, it may be corrupted irreparably and discarded as lost, it is bound to experience delay due to the latency of the path, and it may experience variations in delay, possibly causing it to arrive out of order. Any encapsulating packet that successfully traverses a path enters the de-dispersion buffer. This buffer discards all but the first instance of each encapsulating packet to arrive, and schedules delivery of the encapsulated packets of the rest, appropriately de-fragmented where necessary, from the system on the right of the figure.

Similar to a tunnelling protocol, a dispersity routing system encapsulates any packets it receives for communication through itself. This encapsulation serves a number of purposes. First, it identifies each packet as one that is delivered through a dispersity routing system along one of its paths. Second, it holds the sequence number assigned to the packet, and which is used by the dispersity routing system to deliver only the first of each packet received and to discard the rest. Third, it holds the arrival time of the packet into the dispersity routing system, and which may be used by the de-dispersion buffer to schedule delivery of the packet from the dispersity routing system. Lastly, it holds any data needed to manage fragmentation (when fragmentation is necessary) of the encapsulating packet into the packets communicating that encapsulation. For the sake of simplicity, it is assumed hereafter that fragmentation is not necessary.

Selection of paths for a dispersity routing system is assumed to be a manual process, initially chosen at deployment, and subject to change once operational as part of its ongoing management and maintenance (see section 5.6). The number and choice of paths selected for use by the dispersity routing system connecting any two particular points is assumed to be a

decision based on factors including an understanding of the network between the two points, the characteristics of available paths, and the deliverable quality sought. The planning tools presented in the next two chapters in this thesis may be used to assist in that decision.

For example, the latencies of available paths are likely to differ from one another. Ideally, however, the set of paths chosen for a real-time VOIP communications system that uses dispersity routing comprises paths with comparable delay. Similarly, ideally the paths chosen for inclusion in the set of paths for the dispersity routing system are uncorrelated in their packet loss and delay variation characteristics. Two paths that tend to lose packets at the same time are unlikely to be as beneficial to a dispersity routing system as two paths that do not tend to lose packets at the same time.

As shown by the Howard Street Tunnel Fire in Baltimore [70], a single event may cause multiple paths to fail concurrently for significant time periods. The required degree of resilience to events like this determines the degree of geographical diversity required of the paths chosen. Prudent selection of paths based on an understanding of the physical routes taken by these paths may be necessary when resilience to these kinds of failures is required. This approach of configuring a dispersity routing system for VOIP suffices for VOIP applications such as connecting the two telephone systems of two branch offices of some organisation over the public Internet, because once in place, these points relocate rarely. A branch office, for example, changes location only infrequently. So does the private VOIP handset of an individual, as do the servers of a VOIP service provider to which that handset may connect over the public Internet.

2.4 EFFECTS OF DISPERSITY ROUTING ON VOIP COMMUNICATIONS

To illustrate the effects of dispersity routing on real-time VOIP communications, this section simulates a single exemplary dispersity routing system comprising of three paths. For each path, the simulation adopts a call profile (as measured above) and draws from that call profile the effects that the path has on the packets traversing this path. The effects on packets drawn from call profiles in simulations for this thesis are (1) packet loss, (2) delay, and (3) delay variations. In essence, the characteristics measured with the call profile that is adopted for the path are *replayed* as effects of the path onto the packets traversing that path.

2.4.1 SIMULATING DISPERSITY ROUTING OF VOIP COMMUNICATIONS

Effects are drawn from a call profile in order using the RTP packet sequence number to reconstruct the packet sequence at the sender, and then applied to the packets traversing the path in that order. That is, the effects of packet loss, delay, and delay variations measured for the first packet sent are applied to the first packet traversing the path for which the call profile is adopted. Equally, the effects measured for the second packet sent are applied to the second packet traversing the path, and so on. The duration of the simulation is constrained by the minimum length of the call profiles adopted by the paths, where the length of a call profile is the number of lost and received packets. Effects are drawn only from packets containing media, other packets are ignored.

Adopting the same call profile for multiple paths at the same time causes these paths to have the same effects on the packets traversing them; it causes the paths to be correlated. In a simulation of a dispersity routing system, clearly the collective contribution of such paths is equal to the contribution of just one single path adopting that call profile. When choosing paths for a dispersity routing system, ideally paths with uncorrelated loss and delay behaviours are chosen. Likewise, when selecting call profiles for the paths in a simulation of a dispersity routing system, any call profile is adopted for *one* path in that simulation only. No more than one path in a particular simulation will adopt a given call profile at the same time. Since the paths chosen for a dispersity routing system are ideally uncorrelated in loss and delay behaviours, this constraint reflects that a dispersity routing system requires different paths. There is little point in a dispersity routing system using the same path multiple times.

Packet loss is identified using the RTP packet sequence numbers of the packets received and recorded in the call profile. A packet loss is applied to the corresponding packet traversing the path by discarding that packet from the path. Packets that are not discarded from the path adopt, as the delay experienced when traversing the path due to the latency of the path, the delay estimate recorded in the call profile. Additionally, the packet adopts the delay variation for the packet in the call profile. Both the delay and the delay variation are adopted by including them in the computation of the time that (1) a packet encapsulation

arrives at the de-dispersion buffer and (2) the first copy of each arriving packet is delivered from the system.

Let $\mathcal{M}$ be the set of M received packets in a call profile, and let $\mathcal{S}$ and $\mathcal{R}$ be the sets of send and receive times respectively for the packets in $\mathcal{M}$, such that the send time of packet i is $\mathcal{S}_i$ and the send time of packet j is $\mathcal{S}_j$. Similarly, the receive time of packet i is $\mathcal{R}_i$ and that of packet j is $\mathcal{R}_j$. Assuming the clocks at the sender and receiver increment at the same rate, the delay variation j for packet i in relation to packet k, where $1 < i \leq M$, $1 \leq k < M$, and $k < i$, may be computed as

$$j_{i,k} = (\mathcal{R}_i - \mathcal{R}_k) - (\mathcal{S}_i - \mathcal{S}_k). \tag{1}$$

The *cumulative* delay variation a for received packet i, where $1 < i \leq M$ (that is, the delay variation of packet i in relation to the first packet) is given by

$$a_i = \sum_{x=1}^{i} j_{x,x-1} = (\mathcal{R}_i - \mathcal{R}_1) - (\mathcal{S}_i - \mathcal{S}_1). \tag{2}$$

Since delay variation and cumulative delay variation are computed in relation to a predecessor, and the first packet has no predecessor, delay variation and cumulative delay variation cannot be computed for the first packet. However, as a packet that traverses a path with constant delay l takes by definition at least l to traverse that path, the minimum delay in addition to l is 0 (zero). That delay in addition to l is the delay variation for that packet. Therefore, let the delay experienced by the first packet in addition to l be estimated as $-\min(\mathcal{A})$ where $\mathcal{A} = \left\{a_x : x \in \{x \in \mathcal{M} : x > 1\}\right\}$. To adopt the delay variations observed for the packets in $\mathcal{M}$, the delay in addition to l for received packet i is then given by

$$d_i = \begin{cases} -\min(\mathcal{A}), & i = 1 \\ d_1 + a_i, & i > 1 \end{cases}. \tag{3}$$

Finally, computation of arrival time $\mathcal{Q}$ in the de-dispersion buffer of non-lost packet i traversing a path with delay l is given by

$$\mathcal{Q}_i = \mathcal{S}_i + l + d_i. \tag{4}$$

To compute the delay variation and cumulative delay variation for a packet recorded in a call profile measured, let the receive time be a wall-clock reading taken at the time the packet

is received, and which is recorded in the call profile as the *logged timestamp*. Similarly, let the send time be the RTP packet timestamp formed and included in the packet at the sender when the packet is sent, and which is recorded in the call profile as the *RTP timestamp*.

A wall-clock time duration may be computed from the difference between the two RTP packet timestamps of a pair of RTP packets using the clock rates of the media contained in the packet sequence bounded by the packet pair. Lost packets are assumed to contain media whose loss impacts on the perceived quality of the VOIP telephone call, and which must, therefore, be considered when estimating that quality. For the purposes of computing delay variation and cumulative delay variation, the clock rates of the media that lost packets are assumed to contain, are assumed to be the same as the clock rate of the first sent packet in the packet pair. Indeed, the clock rates of all media contained in any given call profile measured above were found to be always the same.

The delay variation between the two packets shown in the call profile portion in the appendix, for example, may be computed as follows using the clock rate of the media contained in these packets, which is 8000 samples a second:

$$\begin{aligned} j_{i,k} &= (\mathcal{R}_i - \mathcal{R}_k) - (\mathcal{S}_i - \mathcal{S}_k) \\ &= (2.022339 - 2.001407) - \frac{(1\,773\,402\,406 - 1\,773\,402\,246)}{8000} \\ &= 0.020932 - \frac{160}{8000} \\ &= 0.000932 \text{ second,} \end{aligned} \tag{5}$$

where the logged timestamps adopted from the appendix for the receive times are shown in equation (5) as seconds since 2011-12-27T15:11:00+0800, with a resolution of 10^{-6} seconds. The RTP timestamps adopted for the send times are shown exactly as in the appendix.

Although the de-dispersion buffer may be used to delay packets in the same fashion that a de-jitter buffer delays packets to counter delay variations, for the sake of simplicity no delay is adopted by the de-dispersion buffer in this thesis. Rather, the system delivers the first copy of each packet as soon as it arrives in the de-dispersion buffer.

Drawing path behaviour from measured data is similar to the approach taken by [71]. However, unlike their approach, the simulation in this section does not use the

measurements to create stochastic processes from which packet loss, delay and delay variation are then drawn. Packet loss, for example, may be modelled with a Gilbert model [72], a Gilbert-Elliott model [73], or a Markov chain [47]–[48][54][72], and delay variations may be modelled with a shifted gamma distribution [74]–[75]. In this section though, the simulation draws packet loss, delay, and delay variations directly from the call profile adopted for the path, to be as realistic as possible. More importantly, this direct usage enables direct comparison of the simulation results to the measured call profiles adopted in the simulation.

2.4.2 RESULTS OF EXEMPLARY DISPERSITY ROUTING SYSTEM

This section describes a simulation using the method defined above of a single dispersity routing system that comprises three paths. The express purpose of this simulation is to demonstrate the effects of dispersity routing salient to improving the quality of real-time VOIP communications. The call profiles chosen for this simulation were chosen deliberately and principally for this purpose. Consequently, the three call profiles chosen have similar delay estimates, and very high packet loss rates and large burst ratios that individually result in very low MOS estimates. Note that these call profiles were measured earlier than the call profiles described above, in a period beginning 24 November 2009 and ending 16 September 2010.

Table 4: *The effect of dispersity routing on Delay, Quality, Loss, and Mean Burst Length is illustrated by showing their values at the paths used by the dispersity routing system, and at the output delivered by the simulated dispersity routing system.*

Measuring Location	Estimated Delay (ms)	Estimated Quality (MOS)	Loss (Packets)	Mean Burst Length (Packets)
Path 1	57.0	1.50	1419	33.44
Path 2	56.5	1.54	1409	31.38
Path 3	52.0	1.43	1465	34.02
Output	52.0	4.38	35	8.50

Table 4 summarises the effect of dispersity routing by presenting the delay, quality, loss and mean burst length measured for the call profiles adopted by the three paths and for the output computed by the simulation. Prominent in this table is the increase in estimated quality from a mean MOS estimate of 1.49 at the paths to a MOS estimate of 4.38 at the output. Despite using only paths that may be interpreted to deliver an experience worse than one with which *nearly all users would be dissatisfied* (see table 3), the dispersity routing simulation delivers an experience that may be interpreted as *very satisfying*. The deliverable MOS estimate of 4.38 is much closer to the maximum possible MOS estimate of 4.41 for loss-less communication at the output, than the mean MOS estimate of 1.49 at the paths.

The cause for this increase in MOS is a reduction in packet loss and burstiness as quantified by the mean burst length. Mean packet loss on the three paths of 1431 packets (of 7955 packets) is reduced to 35 packets (of 7955 packets) on the output by dispersity routing. Similarly, burstiness is reduced from a mean burst length of the three paths of $\frac{33.44 + 31.38 + 34.02}{3} = 32.95$ packets to 8.5 packets at the output.

Further effects may be observed in figure 8, which depicts cumulative delay variations (labelled in the figure as *jitter* for brevity) for the three paths and the output of the simulated system for the 4000 packets beginning at packet 400. Bursts of packet loss are visible clearly as blocks of missing cumulative delay variation.

Of the three paths used here, path 3 has the lowest delay (see table 4). Thus, packets traversing that path arrive at the de-dispersion buffer before those traversing the other paths, unless path 3 either loses them or delays them enough to arrive later than those traversing the other paths. This is visible clearly for the loss burst on path 3 of packets 1535 – 1693. Since path 2 is also experiencing a loss burst (of packets 1523 – 1557), masking the loss is left to path 1 initially, the path with the highest delay. Path 1 is able to mask the loss until it too begins to lose packets 1554 – 1561. Consequently, all three paths lose packets 1554 – 1557, barely visible as a small loss burst on the output. Path 2 resumes masking packet loss beginning with packet 1558, with path 1 also able to mask packet loss beginning with packet 1562.

Another notable effect is the reduction in cumulative delay variation that happens when packets experiencing *low* cumulative delay variation out-compete packets with *high* cumulative delay variation [7][30]–[31][35][76]–[77]. Despite cumulative delay variation not

affecting the quality estimate in this section (by dropping packets that are delayed excessively through high cumulative delay variation and, thus, affecting the MOS estimate as packet loss), this effect is nevertheless observable in figure 8. Packet 722 experiences a peak in cumulative delay variation on path 3. However, on the output, packet 722 does not. Another path out-competes path 3, and delivers packet 722 earlier than path 3 does.

The effect is illustrated further by the loss burst on path 3 of packets 4063 – 4268. Most of

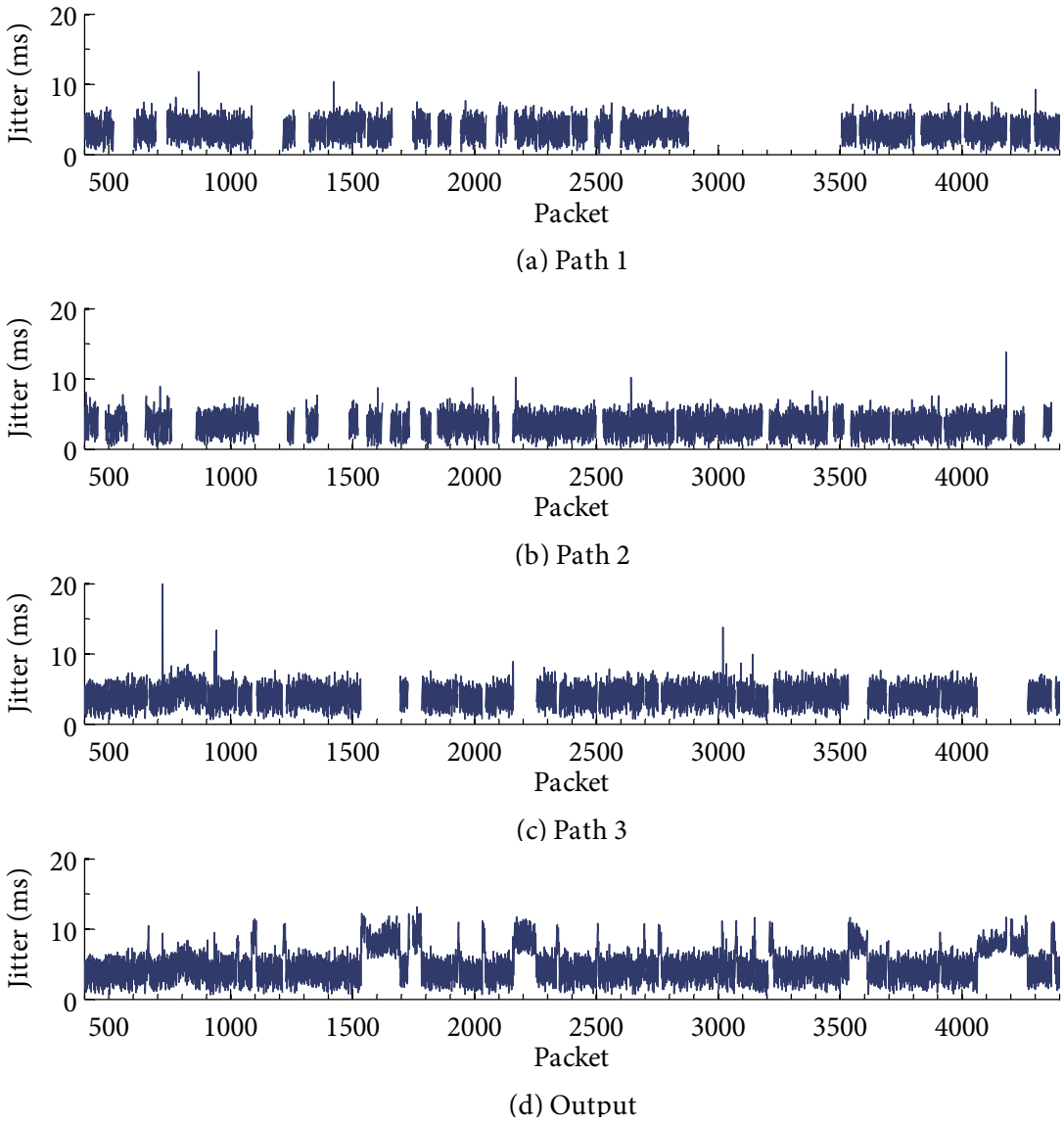

Figure 8: *Subset of simulation illustrating loss and cumulative delay variations (labelled simply as* jitter *in this figure) experienced by packets traversing paths 1 – 3 shown by (a) – (c) respectively, and observed at output (d). Loss is depicted by a gap in cumulative delay variation.*

that loss is masked by paths 1 and 2. However, just like packet 722, when packet 4182 on path 2 experiences high cumulative delay variation, path 1 outcompetes it. Clearly visible also is the smaller range on the output in cumulative delay variation for packets 4063 – 4268 than for packets outside of that period. Because the delays of paths 1 and 2 are very close, competition between these two paths has the opportunity to reduce cumulative delay variation on the output.

A negative effect of dispersity routing is an increased probability of delivering packets out of order. This happens when a path recovering from packet loss delivers a packet *before* a path masking the loss delivers an *earlier* packet that was lost on the recovering path, because the earlier packet is delayed excessively on the masking path. There are two causes for this excessive delay. First, the delay of the masking path may be much higher than that of the recovering path. Second, the packet may experience high cumulative delay variation. Assume that packet m_1 is sent at time t_1 over path r_1 with delay l_1 and cumulative delay variation a_1, and that packet m_2 is sent at time t_2 over path r_2 with delay l_2 and cumulative delay variation a_2, where $t_1 < t_2$. Packet m_1, despite being sent *earlier*, will arrive *later* than m_2 when $t_1 + l_1 + a_1 > t_2 + l_2 + a_2$.

Another negative effect is the possibility of greater delay variation that may occur when paths with higher delay begin and end masking loss for a path with a lower delay. Figure 8 illustrates this clearly by the elevated cumulative delay variation on the output for packets 1535 – 1693 for example. When paths 1 and 2 begin to mask the loss burst on path 3, cumulative delay variation increases. Because paths 1 and 2 have a higher delay, packets simply take longer to traverse these paths.

The probability of these negative effects occurring may be reduced by scheduling delivery from the de-dispersion buffer of the first instance of each packet, instead of delivering them as soon as they arrive. This is similar to the way that a de-jitter buffer schedules packets for delivery to reduce delay variation (also known as jitter). Given a set of path delays $\mathcal{D}$, a de-dispersion buffer of size $\max(\mathcal{D}) - \min(\mathcal{D})$ compensates for the difference in delays causing delay variation and packet re-ordering. However, for illustration purposes, in the simulations in this thesis de-dispersion buffers never delay packets, but, rather, always deliver them as soon as they arrive.

2.5 EFFECTIVENESS OF DISPERSITY ROUTING IN IMPROVING QUALITY

This section presents the results of three simulations that demonstrate how effective dispersity routing is at improving the quality of real-time VOIP communications. They do so by adopting actual VOIP traffic measured in a commercial call centre (as described above) for the behaviour of the paths used in the simulated dispersity routing systems. By adopting these measurements, comparisons of the estimated quality deliverable by these simulated dispersity routing systems may be made against the estimated quality measured in the commercial call centre.

2.5.1 IMPROVEMENTS IN QUALITY FOR OVERALL CONDITIONS

The first simulation comprises dispersity routing systems of 2 – 6 paths [7], and conveys the overall improvements in estimated quality by drawing path effects from the full set of call profiles measured between 1 August 2011 and 31 January 2012. Let i be the number of paths in range 2 – 6. For each i, compose at random a maximum of 10 000 *unique* simulation *scenarios*, where each scenario is a collection of i call profiles (that is, one call profile for each of the i paths; thus i call profiles) selected, at random, from the set of available call profiles. The (unordered) collection of i call profiles selected for a particular scenario contains i different call profiles; that is, no two call profiles selected for a scenario are the same. Furthermore, all scenarios have different (unordered) collections of call profiles; that is, no two scenarios have collections of call profiles that contain the same i call profiles (in any order).

Figure 9 plots for each i the probability that the MOS estimated for the simulation output of each scenario of i paths is at least 4.34 (that is, delivering call qualities that users would perceive as very satisfying; see table 3). For comparison, figure 9 also includes the probability for non-dispersity routing systems (that is, with 1 path) by including the probability that a call profile in the set of available call profiles has an estimated MOS of at least 4.34.

As can be seen, 95.64% of the calls profiled between 1 August 2011 and 31 January 2012 have an estimated MOS of at least 4.34; that is, 4.36% of measured calls are of a quality that users would perceive to be less than very satisfying. Using dispersity routing, the simulations

increase the probability of delivering a call quality with which users would perceive to be very satisfied to 0.9999 with just two paths, and to 1 with three or more paths. With just two paths, dispersity routing delivers a telephony service of a quality that is more on par with traditional telephony than VOIP currently delivers in the commercial call centre. That is, with two paths the proportion of calls with which users would perceive to be very satisfied increases from slightly more than 1 in every 23 to slightly less than 1 in every 10 000 (the probability of 0.9999 above was rounded to four decimal places).

2.5.2 IMPROVEMENTS IN QUALITY FOR EXTREME CONDITIONS

The second simulation is similar to the first simulation, except that, to explore the effectiveness of dispersity routing in the most extreme conditions observed, it does not draw on the full set of available call profiles. Rather, it draws on the subset of available call profiles comprising the call profiles with the worst 100 MOS estimates.

Let i be the number of paths in range 2 – 6. For each i, compose a maximum of 10 000

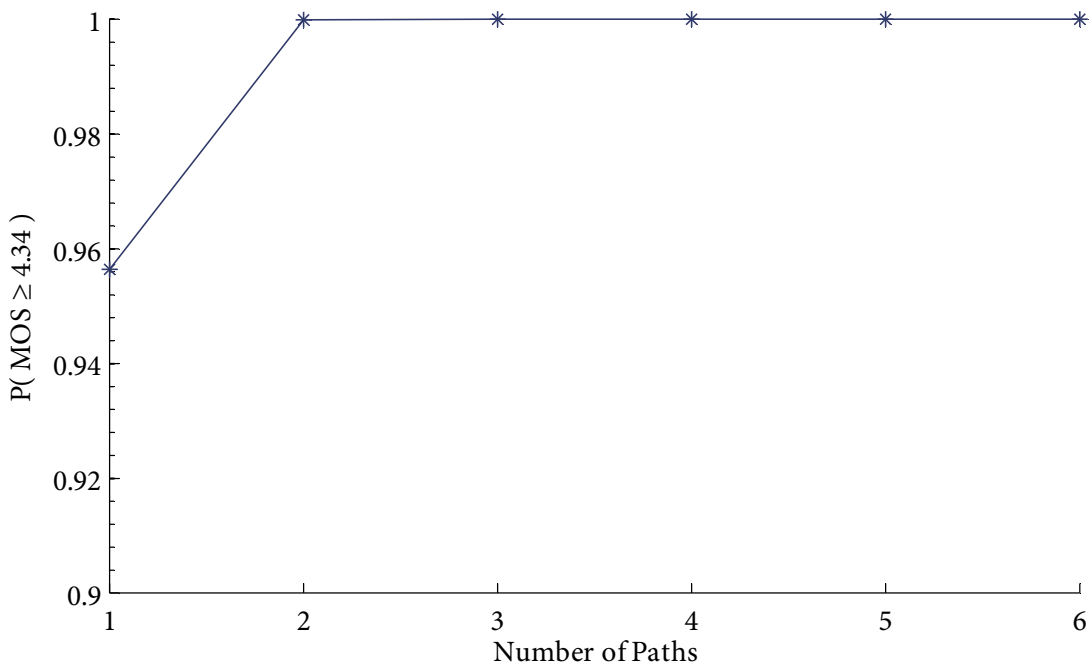

Figure 9: *Probabilities of estimated MOS being at least 4.34 for overall observed conditions with 1 path (no dispersity routing) and 2 – 6 paths (with dispersity routing). Dispersity routing with two paths already yields significant improvements over no dispersity routing.*

unique simulation *scenarios* at random, where each scenario is a collection of i call profiles (that is, one call profile for each of the i paths; thus i call profiles) selected at random from the subset of available call profiles. No two call profiles selected for a given scenario are the same. Furthermore, no two scenarios have collections of call profiles that contain the same call profiles. For two paths, the maximum number of unique scenarios possible is $\binom{100}{2} = 4950$, for three or more paths the sought maximum of 10 000 unique scenarios are possible.

Figure 10 depicts six cumulative distribution functions, and for each identifies its lowest 5th percentile with a vertical dashed line. Viewed from left to right, the first 5th percentile is for the first cumulative distribution function, the second 5th percentile is for the second cumulative distribution function, and so on. The left-most cumulative distribution function shown in figure 10 is for the MOS estimates of the subset of available call profiles with the worst 100 MOS estimates, and represents the quality observed for these call profiles without dispersity routing. From left to right, the remaining five cumulative distribution functions are the output MOS estimates from the simulations of dispersity routing systems employing 2 –

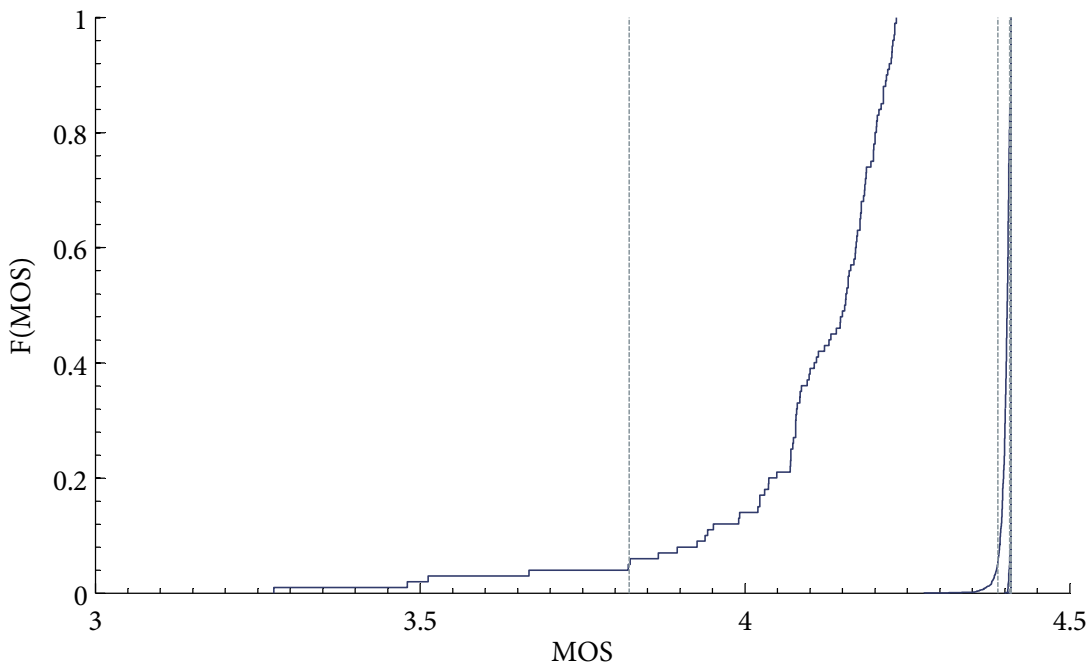

Figure 10: *Cumulative distributions of output MOS estimates in extreme conditions for systems with (from left to right) 1 – 6 paths. Also shown is the lowest 5th percentile of these distributions, also from left to right. The results for the 3 path system obscure the results of 4 – 6 path systems.*

6 paths respectively. The last three cumulative distribution functions and their lowest 5th percentiles are obscured by the cumulative distribution function and its lowest 5th percentile for the dispersity routing system employing 3 paths, because the simulation results become increasingly alike.

The shift discernible in figure 10 in the cumulative distribution function with increasing paths of the output MOS estimate towards the maximum possible MOS of 4.4094 for a lossless output illustrates the impact of dispersity routing on deliverable quality. As the number of paths employed by dispersity routing increases, the deliverable quality tends towards equating that of lossless communication, thereby showing that even in extreme conditions dispersity routing can improve quality. However, that trend is one of diminishing returns [32], as illustrated by figure 11 which plots the 5th percentiles for the cumulative distribution functions as shown in figure 10. Clearly observable are the increasingly smaller gains purchased by dispersity routing with additional paths.

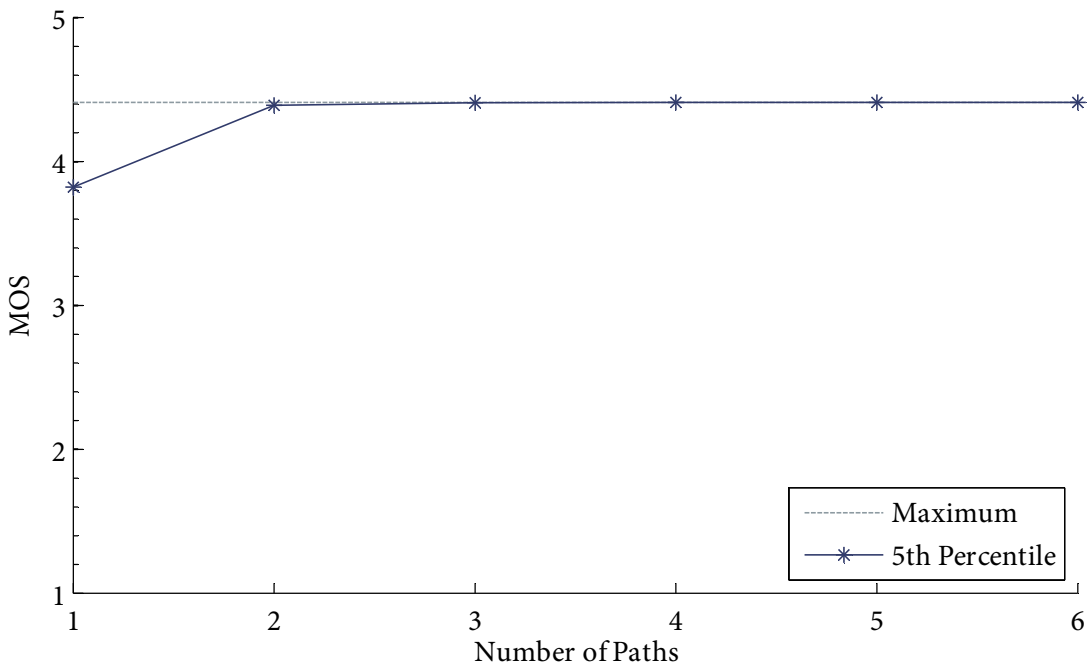

Figure 11: *The 5th percentiles of output MOS estimates in extreme conditions. The largest gain is achieved changing from no dispersity routing to dispersity routing with 2 paths. Additional paths yield increasingly diminishing returns, trending towards the MOS for lossless communication.*

2.5.3 IMPROVEMENTS IN QUALITY USING JUST TWO PATHS

The third simulation is of a dispersity routing system comprising two paths only, but which adopts every possible combination of the full set of call profiles measured in the commercial call centre. While, as shown above, increasing numbers of paths yield increasingly smaller improvements in quality, a dispersity routing system of just two paths is less likely to deliver a quality as high as a system of three or more paths. Therefore, a dispersity routing system of two paths represents the minimum improvements in quality possible with dispersity routing. By adopting every possible combination of the call profiles measured in the commercial call centre, this simulation illustrates the minimum improvements in quality possible with dispersity routing in that commercial call centre.

Figure 12 depicts the cumulative distribution function of the MOS estimates for the output of the simulated dispersity routing systems. Of the $\binom{6265}{2} = 19\,621\,980$ possible scenarios, 6067 have a MOS less than 4.34. Thus, 99.97% of the scenarios in this simulation deliver a MOS of 4.34 or above; that is, a quality with which users may be interpreted to be

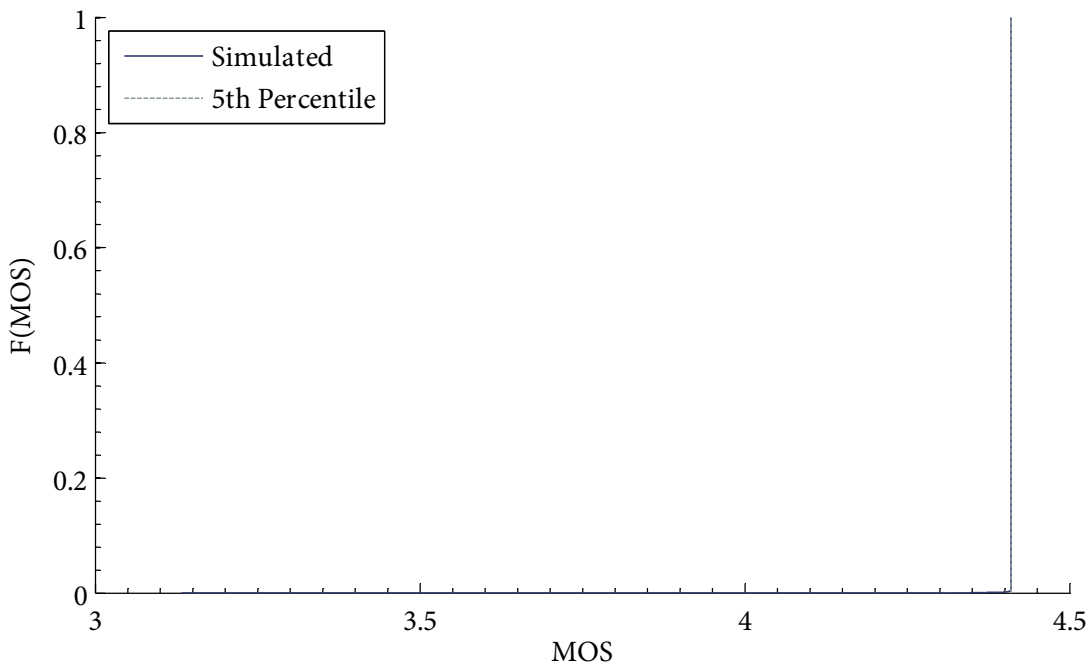

Figure 12: *Cumulative distribution function of output MOS estimates from a dispersity routing system of 2 paths for* every *combination of measured call profiles. With just two paths, in this simulation dispersity routing increases 'very satisfied' calls from 95.64% to 99.97%.*

very satisfied. Indeed, the 5th percentile of that cumulative distribution function is the maximum possible MOS of 4.41 for lossless communication. The quality improvement is illustrated vividly in figure 12 by the vertical dashed line for the 5th percentile almost entirely concealing the cumulative distribution of the simulated MOS estimates. It is further illustrated by contrasting the cumulative distribution of currently observable MOS estimates in the commercial call centre as depicted in figure 9 against the cumulative distribution of simulated MOS estimates in figure 12.

With just two paths, dispersity routing in this simulation improves the quality in the commercial call centre from 95.64% calls perceived to be very satisfying to 99.97%. This is lower than the 99.99% achieved by the first simulation above (see section 2.5.1). However, as that simulation comprises 10 000 scenarios chosen at random for 2 paths, the discrepancy equates to just $10\,000 \cdot 0.02\% = 2$ scenarios.

Chapter 3 • Quantifying the Improvements to Quality

This chapter proposes a mathematical model, called the QCE-model, for estimating the quality that may be expected from a dispersity routing system of known characteristics. That model is then applied to characteristics extrapolated from measurements taken in the commercial call centre (see section 2.1) to show the effectiveness of dispersity routing in improving quality. In addition, the application of the QCE-model may also be used as a planning tool, by illustrating the relationships between numbers of path, packet loss probabilities, and quality that might be expected.

The majority of the material in this chapter, all of which was produced for this thesis, has been peer reviewed and published in [8]. However, this chapter extends and builds upon the material presented in [8] in the following major points:

1. Terminology, mathematical symbols and notation has been harmonised with this thesis.
2. The discussion on delay is more detailed, with the selection of zero as a reasonable value for Q in equation (6) supported by data collected for this thesis (see section 2.1).
3. Optimisation for computing matrix $\mathbf{W}$ in equation (8) is given as equation (16).
4. Computation of the E-model burst ratio is described in greater detail than in [8].
5. The discussion on the relationships between packet loss, packet loss burstiness, the numbers of paths used in a dispersity routing system and the quality most likely delivered by a dispersity routing system is more detailed.
6. This chapter uses a more recent data set than the one used in [8]. The VOIP quality observed in this data set is higher than in that observed for [8], which translates into smaller gains delivered by dispersity routing.
7. This newer data set was collected for this thesis by a custom data collector written specifically for this thesis, which operates at a lower level than the one used in [8] resulting in more accurate data.
8. The application of the QCE-model has been extended to include directly observed characteristics to illustrate the agreement of MOS estimates computed through the PLB-model with those computed for directly observed characteristics.
9. The analysis is more detailed than in [8].

3.1 THE QCE-MODEL FOR ESTIMATING QUALITY

To estimate the deliverable quality that a particular dispersity routing system is most likely to deliver, the packet loss and loss burstiness characteristics are computed for that system. Together with an estimate of the delay and identification of the codec used, the E-model may then compute a MOS estimate from these computed characteristics.

The delay of a dispersity routing system (that is, the time that it takes a packet to traverse that dispersity routing system) may be estimated from (1) the delay experienced by the packets traversing the paths participating in the dispersity routing system and (2) the delay adopted for the de-dispersion buffer. Formally, given the set of delays $\mathcal{D}$ for the paths participating in the dispersity routing system and the delay Q adopted by the dispersity routing system for the de-dispersion buffer, the delay D of a dispersity routing system may be estimated as

$$D = \min(\mathcal{D}) + Q\,. \tag{6}$$

As ideally paths are chosen that are comparable in delay (see section 2.3), and as the E-model does not consider delay to impact on quality until it exceeds 100 milliseconds [58], a value of zero may be assumed for Q in most cases without impact. Of the VOIP telephone calls measured (see section 2.1), 99.7% have delay estimates of 79.14 milliseconds or less, and 95% of 57.67 milliseconds or less. Selection of a non-zero value for Q is discussed in section 5.5.

The packet loss and loss burstiness characteristics of a path may be modelled by a Markov model [72]–[73] such as the 4-state Markov model [78]–[83] depicted in figure 13. This model distinguishes between periods of high packet loss and periods of low packet loss. A high packet loss period, referred to as a *loss burst*, is not necessarily a period of total packet loss. While packets *are* lost during a loss burst, some packets during a loss burst may not be lost. Conversely, a low packet loss period, known as a *gap*, is not necessarily a period of absolutely no packet loss at all; some packets during a gap may *not* be received. The states of the 4-state Markov model as shown in figure 13 describe the four possible combinations of loss burst and gap with packet loss and receipt. That is, the receiving of packets while in a gap is denoted with the *Gap Receive* state and the losing of packets while in a gap with the *Gap Loss* state.

Similarly, the losing of packets while in a loss burst is denoted with the *Burst Loss* state and the receiving of packets while in a loss burst with the *Burst Receive* state.

Gaps and loss bursts are distinguished by fixing the minimum number of consecutively received packets in a gap, where that constant is called G_{min}. To be considered a part of a gap, any lost packet in a gap must be separated by at least G_{min} consecutively received packets from any other lost packet. Any period that is not a gap is a loss burst. Since at least G_{min} consecutively received packets separate each lost packet in a gap from other lost packets, a packet loss in a gap is, necessarily, a single, isolated, packet loss [78]. This thesis fixes G_{min} to the value 16, a value recommended for G_{min} by [80].

Let the discrete state space $\mathcal{Z} = \{1,2,3,4\}$ represent the states Gap Receive, Burst Receive, Burst Loss and Gap Loss respectively. Furthermore, let the state transition matrix **P** express the state transition probabilities, such that $p_{i,j}$, the element in row i and column j, is the probability of a transition from state $i \in \mathcal{Z}$ to state $j \in \mathcal{Z}$ occurring.

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,4} \\ \vdots & \ddots & \vdots \\ p_{4,1} & \cdots & p_{4,4} \end{bmatrix} \tag{7}$$

Unlike the typical right stochastic matrix where each row vector sums to unity (that is, $\sum_{j \in \mathcal{Z}} p_{i,j} = 1$ where $i \in \mathcal{Z}$), here all the elements of state transition matrix **P** sum to unity (that is, $\sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{Z}} p_{i,j} = 1$). The difference is that, instead of $p_{i,j}$ quantifying the probability of, being in state i, going to state j as opposed to the other states, $p_{i,j}$ instead quantifies the probability of transitioning from state i to j as opposed to all other possible state transitions.

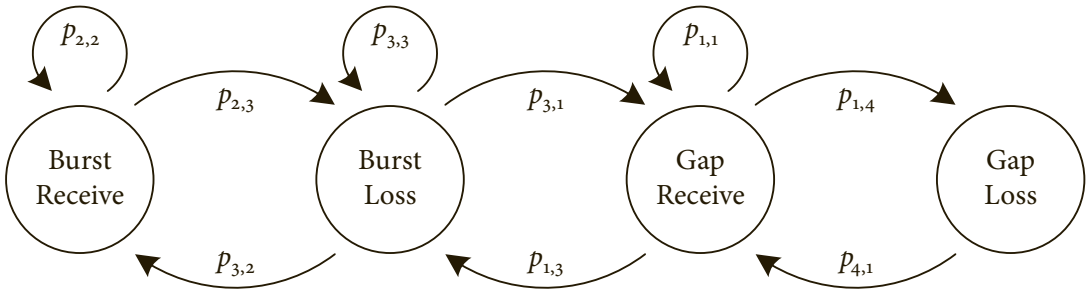

Figure 13: *A 4-state Markov model considers periods of high loss as* loss bursts *and all other periods as* gaps. *Packets are lost in the* Burst Loss *and* Gap Loss, *and received in the* Gap Receive *and* Burst Receive *states.*

This formulation is convenient when computing the packet loss and loss burstiness characteristics of a fully redundant dispersity routing system as shown below, because the probability of every state transition combination experienced by the paths used by the system may then be computed readily.

Clearly, the probability of a particular state transition combination may be computed from the probabilities of the state transitions in that combination, and $p_{i,j}$ quantifies the probability of transitioning from state i to j. However, when computing the probability of a state transition combination experienced by the paths of a dispersity routing system, being independent each path may be in any of the states in $\mathcal{Z}$ transitioning to any state in $\mathcal{Z}$ with the probabilities quantified in the state transition matrix for that path. Therefore, instead of $p_{i,j}$ quantifying, as it does in a typical right stochastic matrix, the probability of transitioning from state i to j when in state i (that is, given that the probability of being in state i is 1), it is convenient for it to quantify instead the probability of transitioning from state i to j when in any of the states in $\mathcal{Z}$. The formulation of state transition matrix $\mathbf{P}$ above does so directly. This makes possible ready computation of the probability of each state transition combination the paths of a fully redundant dispersity routing system may experience from the state transition matrices of these paths.

It is clear that the probability of packet loss, b, is the sum of the probabilities of transitioning to the packet loss states Burst Loss and Gap Loss. Formally, given discrete state space $\mathcal{X} = \{3,4\}$, that is $\mathcal{X} \subset \mathcal{Z}$, that represents the two packet loss states Burst Loss and Gap Loss respectively, $b = \sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{X}} p_{i,j}$. Besides the probability of receiving a packet, g, being $g = 1 - b$, g is also the sum of the probabilities of transitioning to the two packet receipt states Gap Receive and Burst Receive. For completeness, given discrete state space $\mathcal{G} = \{1,2\}$, that is $\mathcal{G} \subset \mathcal{Z}$, that represents the packet receipt states Gap Receive and Burst Receive respectively, $g = \sum_{i \in \mathcal{Z}} \sum_{j \in \mathcal{G}} p_{i,j}$.

The packet loss and loss burstiness characteristics of a dispersity routing system using a set of N, where $N \geq 2$, paths $\mathcal{P} = \{1,2,\ldots,N\}$ each characterized by state transition matrix $\mathbf{P}_{i \in \mathcal{P}}$, may be described by the Kronecker product of these matrices. In combination with the way in which state transition matrix $\mathbf{P}$ is defined — that is, all the elements of $\mathbf{P}$ sum to unity; see equation (7) — the Kronecker product computes the probabilities of all state transition

combinations collectively characterising the packet loss and loss burstiness characteristics of paths $\mathcal{P}$. The sum of the probabilities of those combinations that fully redundant dispersity routing is unable to mask quantifies the packet loss and loss burstiness characteristics of that fully redundant dispersity routing system. For completeness, the packet loss and loss burstiness characteristics of a single-path (that is, non-dispersity routing) system may be characterized by the state transition matrix $\mathbf{P}_{i=1}$ of its only path. Therefore, the packet loss and loss burstiness characteristics of a system using a set of N paths, where $N \geq 1$, may be described by $\mathbf{W}$ as,

$$\mathbf{W} = \begin{cases} \mathbf{P}_{i=1}, & \text{if } N = 1 \\ \bigotimes_{i=1}^{N} \mathbf{P}_i, & \text{if } N > 1 \end{cases}. \tag{8}$$

Let $Z = |\mathcal{Z}| = 4$ be the cardinality (that is, the number of elements) of set $\mathcal{Z}$, and $X = |\mathcal{X}| = 2$ be the cardinality of set $\mathcal{X}$. Clearly, X^N columns of $\mathbf{W}$ (that is, those representing a state transition to a packet loss state on all N paths) contain probabilities of simultaneous packet loss on all N paths. Therefore, the sum of these X^N columns is the probability of simultaneous packet loss on all N paths. Given matrices $\mathbf{P}_i$, where $i = \{1,2,\dots,N\}$, let $\mathbf{p}_k^{(i)}$ be the kth column vector of the ith matrix $\mathbf{P}_i$. Furthermore, let function $\lambda(z_1, z_2, \dots, z_N)$ compute for $N > 1$ the index in $\mathbf{W}$ of the Kronecker product of the column vectors $\left\{\mathbf{p}_{z_i}^{(i)} : i = \{1,2,\dots,N\}\right\}$. For $N = 1$, let $\lambda(z_1, z_2, \dots, z_N)$ equate to the identity function. Therefore, let $\lambda(z_1, z_2, \dots, z_N)$ be defined as

$$\lambda(z_1, z_2, \dots, z_N) = \begin{cases} z_1, & \text{if } N = 1 \\ 1 + \sum_{i=1}^{N} (z_i - 1)\, Z^{(N-i)}, & \text{if } N > 1 \end{cases}. \tag{9}$$

Formally, the set of indices of the X^N columns in $\mathbf{W}$ that represent state transitions to a packet loss state on all N paths (that is, the set of indices in $\mathbf{W}$ of the Kronecker products of column vectors $\left\{\mathbf{p}_{z_i}^{(i)} : z_i \in \mathcal{X}, i = \{1,2,\dots,N\}\right\}$ that represent transitions to a loss state on the N paths) then is

$$\mathcal{L} = \left\{\lambda(z_1, z_2, \dots, z_N) : z_i \in \mathcal{X}, i = \{1,2,\dots,N\}\right\}. \tag{10}$$

Therefore, the probability of packet loss by a system with these N paths — which is the probability of simultaneous packet loss on all N paths — is

$$P(\text{loss}) = \sum_{i=1}^{Z^N} \sum_{j \in \mathcal{L}} w_{i,j} \ . \tag{11}$$

Given that the set of indices of the rows in $\mathbf{W}$ that do not represent a state transition from a packet loss state on all N paths is

$$\mathcal{N} = \left\{ r \in \left\{ 1, 2, \ldots, Z^N \right\} : r \notin \mathcal{L} \right\}, \tag{12}$$

the probability of the system traversing from a packet receipt state to a packet loss state is

$$P(\text{burst}) = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{L}} w_{i,j} \ . \tag{13}$$

Computation of $\mathbf{W}$ becomes expensive for large numbers of paths; given N paths, $\mathbf{W}$ is a $Z^N \times Z^N$ matrix. However, as computation of $P(\text{loss})$ and $P(\text{burst})$ requires only a subset of the elements in $\mathbf{W}$, computation may be simplified by computing only those elements actually needed. Let $l(i)$ and $r(i)$ be functions that, for $\mathbf{C} = \mathbf{A} \bigotimes \mathbf{B}$ (where $\mathbf{A}$ is an M-by-N matrix and $\mathbf{B}$ is a Z-by-Z matrix), compute for index i in $\mathbf{C}$, the corresponding indices in $\mathbf{A}$ and $\mathbf{B}$ respectively. That is, $c_{i,j} = a_{l(i),l(j)} b_{r(i),r(j)}$ for $1 \leq i \leq mZ$ and $1 \leq j \leq nZ$. Given $\lfloor x \rfloor$ evaluates to the floor of x, let these functions be defined as

$$l(i) = \left\lfloor \frac{i-1}{Z} \right\rfloor + 1 \tag{14}$$

and

$$r(i) = i - Z \left\lfloor \frac{i-1}{Z} \right\rfloor . \tag{15}$$

Therefore, given that $p_{i,j}^{(k)}$ is the (i, j)th element of the kth matrix $\mathbf{P}_k$, $w_{i,j}$ in equations (11) and (13) may be computed as $w_{i,j}^{(N)}$, which is defined recursively as

$$w_{i,j}^{(x)} = \begin{cases} p_{i,j}^{(x)}, & \text{if } x = 1 \\ w_{l(i),l(j)}^{(x-1)} p_{r(i),r(j)}^{(x)}, & \text{if } x > 1 \end{cases} . \tag{16}$$

The E-model characterises loss burstiness as a *burst ratio*, BurstR, that may be calculated using a 2-state Markov model [58] and which captures "very short-term dependencies

between lost packets, i.e., consecutive losses" [79]. This is in contrast to the 4-state Markov model used in the QCE-model to compute the packet loss and loss burstiness characteristics of a fully redundant dispersity routing system from the packet loss and loss burstiness characteristics of the paths used by that system. In essence, the Markov models used differ because they are used for different purposes.

In the *loss* state of this 2-state Markov model the probability of packet loss is 1 [79], as opposed to the *receive* state where the probability of packet loss is 0 (zero). This is a specialisation of the Gilbert model [72], obtained by fixing the probability, h, of correct reception in the B (that is, the *Bad*, or lossy) state to be 0. It could also be considered a specialisation of the Gilbert-Elliott model [73] as depicted in figure 14. The Gilbert-Elliott model is an extension of the Gilbert model, adding the probability, k, of correct reception in the G (that is, the *Good*, or non-lossy) state. In the Gilbert model, the probability of correct reception in the G state is fixed at 1; that is, the Gilbert model is a Gilbert-Elliott model with the probability, k, of correct reception in the G state fixed at 1.

Let p be the probability of transitioning to the packet loss state from the packet receipt state, computed as

$$p = \frac{P(\text{burst})}{1 - P(\text{loss})} \,. \tag{17}$$

The burst ratio for the E-model may then be calculated as

$$\text{BurstR} = \frac{P(\text{loss})}{p} \,. \tag{18}$$

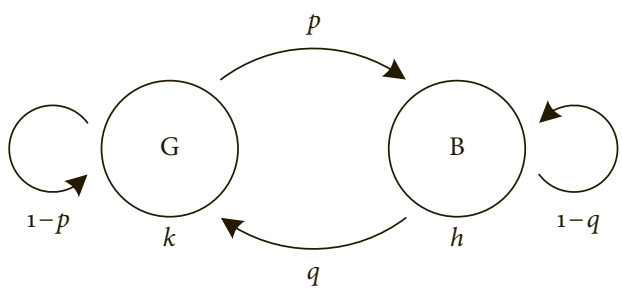

Figure 14: *The Gilbert-Elliott model comprises (1) a 2-state Markov model and (2) the probabilities of* correct *reception when in these states. In the Gilbert model (which the Gilbert-Elliott model extends) loss is possible only in the* B *state (that is, k is fixed to 1)* [72].

Having computed the packet loss probability in equation (11), the burst ratio in equation (18), estimated the delay in equation (6), and knowing the codec used, the E-model can then compute a rating factor from which it can then compute a MOS estimate. The default values recommended by the E-model [58] are adopted for all parameters except for those derived from the above [63]. However, MOS estimates are computed only for parameter values that are within the ranges permitted by [58]. For parameter values outside the permitted ranges, the MOS estimate is computed as undefined, because "the results obtained [with these parameters] will not have been validated" and, "therefore, the use of such values should be avoided" [63].

3.2 ESTIMATING QUALITY WITH THE QCE-MODEL

The QCE-model described in section 3.1 above is applied in this section to estimate the quality that dispersity routing systems of 2 to 6 paths are most likely to deliver. In addition to illustrating the most likely quality improvements deliverable by dispersity routing, this application of the QCE-model also conveys a sense of the relationships between (1) packet loss, (2) packet loss burstiness, (3) the numbers of paths used by a dispersity routing system and (4) the quality most likely delivered by that dispersity routing system.

To make the quality estimates and conveyed sense of relationships realistic, the characteristics of the paths are adopted from the actual VOIP traffic measurements described in section 2.1 above. However, while the QCE-model supports paths with *differing* packet loss and loss burstiness characteristics, the model application in this section assumes the *same* characteristics for each path of a dispersity routing system. By constraining the packet loss and packet loss burstiness characteristics to be the same for the paths in this application of the model, a sense of the sought relationships described above is able to emerge. In chapter 4, paths with differing packet loss and loss burstiness characteristics are included in the applications of the QCE-model described there.

In addition to applying the QCE-model to the measured VOIP traffic characteristics, the QCE-model is also applied to a mathematical model constructed from these measurements called the PLB-model and which is described in section 3.2.1 below. The PLB-model makes possible deliverable quality estimates for packet loss rates that were not observed in the VOIP

traffic measurements. For example, an exact loss rate of 10% was not observed; the closest loss rates observed were 9.34% and 11.56%. Similarly, a loss rate of exactly 20% was not observed either; the closest were 11.56% and 29.59%. Nor was a loss rate in excess of 49.53% observed.

The PLB-model adopts the observed packet loss and packet loss burstiness characteristics, and for any packet loss probability in the range 0 to 1 computes for that packet loss probability the closest fitting characteristics to the observed packet loss and packet loss burstiness characteristics. Furthermore, as well as applying the QCE-model to the PLB-model by assuming the burstiness characteristics computed by the PLB-model, the QCE-model is also applied to the PLB-model assuming random packet loss. This approach illustrates the significance of including packet loss burstiness in quality estimates.

Both applications of the QCE-model — that is, to the observed characteristics and to the characteristics from the PLB-model — are presented in section 3.2.2 below. Finally, in order to place the results for dispersity routing into context with non-dispersity routing, they are presented together with quality estimates for a single-path system that assumes the same packet loss and packet loss burstiness characteristics.

3.2.1 THE PLB-MODEL FOR RELATING LOSS TO STATE TRANSITIONS

The PLB-model relates packet loss probabilities ranging from 0 to 1 to state transition matrices of the 4-state Markov model used in section 3.1 to model the packet loss and loss burstiness characteristics of a path. It does this by fitting a second degree polynomial for each of the 9 state transitions possible (as enumerated in table 5) in the 4-state Markov model used to characterise packet loss and loss burstiness. Each polynomial is a linear least-squares fitting of a set of 6264 points, constrained to the expected values at packet loss probabilities 0 and 1. Each point maps the packet loss probability (as the independent variable) observed for a real VOIP telephone call measured in a commercial call centre to the value of that polynomial's state transition probability (as the dependent variable) observed for that call. The packet loss probability and the state transition probabilities are observed for a VOIP call by identifying lost RTP packets and computing the resulting packet loss probability and 4-state Markov model state transition probabilities from the call profile of that call. Let c be the count of lost packets and L the total number of lost and received packets in the call profile of a call. Clearly,

the packet loss probability b for that call is $b = \frac{c}{L}$. Similarly, let $c_{i,j}$ be the count of transitions from state i to state j in the call profile of a call. The state transition probability $p_{i,j}$ of state transition matrix **P** for that call then is $p_{i,j} = \frac{c_{i,j}}{L}$.

All 9 polynomials are constrained to the expected state transition probabilities at packet loss probabilities 0 and 1. Because at packet loss probability 0 there is no loss, the only state transition possible at packet loss probability 0 is Gap Receive to Gap Receive. Therefore, all polynomials are constrained to state transition probability 0 at packet loss probability 0, *except* for the polynomial for Gap Receive to Gap Receive. That polynomial is constrained instead to state transition probability 1 at packet loss probability 0, because at zero packet loss the only state transition possible is from Gap Receive to Gap Receive.

Similarly, because at packet loss probability 1 the only state transition possible is Burst

Table 5: *Coefficients for polynomials that map the loss rate to the state transition probabilities in the state transition matrix of the 4-state Markov model used to characterise the loss and loss burstiness properties observed for the 6264 measured VOIP telephone calls.*

From	To	Coefficient 1	Coefficient 2	Coefficient 3
Gap Receive	Gap Receive	1.1536E+00	−2.1536E+00	1.0000E+00
Gap Receive	Burst Loss	−1.0066E−01	1.0066E−01	0.0000E+00
Gap Receive	Gap Loss	−1.4381E−01	1.4381E−01	0.0000E+00
Burst Receive	Burst Receive	−7.3623E−01	7.3623E−01	0.0000E+00
Burst Receive	Burst Loss	−1.7286E−01	1.7286E−01	0.0000E+00
Burst Loss	Gap Receive	−1.0066E−01	1.0066E−01	0.0000E+00
Burst Loss	Burst Receive	−1.7286E−01	1.7286E−01	0.0000E+00
Burst Loss	Burst Loss	4.1734E−01	5.8266E−01	0.0000E+00
Gap Loss	Gap Receive	−1.4381E−01	1.4381E−01	0.0000E+00

Loss to Burst Loss, all polynomials are constrained to state transition probability 0 at packet loss probability 1, *except* for the polynomial for Burst Loss to Burst Loss. That polynomial is constrained instead to state transition probability 1 at packet loss probability 1, since at total packet loss the only state transition possible is from Burst Loss to Burst Loss. For completeness, table 5 presents the coefficients for the second degree polynomials determined by the linear least-squares fitting using the packet loss and loss burstiness characteristics measured for the 6264 VOIP telephone calls as described in section 2.1 above. Letting c_1, c_2, and c_3 be coefficients 1, 2 and 3 respectively of a state transition catalogued in table 5, the probability of that state transition at packet loss probability $x \in [0,1]$ may then be computed as $\mathrm{p}(x) = c_1 x^2 + c_2 x + c_3$.

3.2.2 APPLYING THE QCE-MODEL

In this section, the QCE-model is applied to three sets of data. First, the QCE-model is applied to the PLB-model described in section 3.2.1 adopting the measured characteristics (see section 2.1) for packet loss rates ranging from 0 to 1, that is, assuming bursty loss characteristics. Second, the model is applied for the same packet loss rate range to the same PLB-model described but assuming random (that is, non-bursty) loss. Third, the model is applied to packet loss and packet loss burstiness characteristics observed in the real VOIP traffic measurements (see section 2.1).

Let $i \in \{1,2,\ldots,6\}$ be the number of paths, and $b \in \{0,0.01,\ldots,1\}$ be the packet loss probability. That is, in this section the QCE-model is applied to a sample of possible packet loss probabilities in the interval [0,1] at an interval of 1%. For each value of i, compute, using the QCE-model and the E-model,

1. a MOS estimate for the packet loss and packet loss burstiness characteristics computed by the PLB-model for each value of b assuming *bursty* loss,
2. a MOS estimate for the same characteristics computed by the PLB-model for each value of b assuming *random* loss, and
3. a MOS estimate for the packet loss and packet loss burstiness characteristics observed for each measured VOIP telephone call (see section 2.1 above).

To compute a MOS estimate assuming bursty loss, first compute the state transition matrix **P** in equation (7) of the 4-state Markov model that models the packet loss and loss burstiness characteristics of the paths. The PLB-model computes matrix **P** by evaluating the second degree polynomial for each of the 9 state transitions possible at loss rate b using the coefficients computed above and shown in table 5. The deliverable quality may then be estimated from **P** as a MOS estimate using the QCE-model and the E-model as described in section 3.1 above. Since in this section the paths in the dispersity routing system assume the same packet loss and loss burstiness characteristics, matrix **P** is used for each of the i paths.

For convenience, computation of a MOS estimate assuming non-bursty loss is identical to the above, except that the burst ratio, BurstR, is fixed at 1. However, that computation is equivalent to computing a MOS estimate for loss rate b directly, again while fixing the burst ratio to 1.

To compute a MOS estimate for each measured VOIP telephone call, the state transition matrix **P** is computed from the packet loss and packet loss burstiness characteristics observed for the call. From **P** the MOS estimate is then computed using the QCE-model and the E-model as above.

Figure 15 depicts the MOS estimates thus computed for, from left to right, a single path system and dispersity routing systems using 2 to 6 paths respectively. All paths in each of these systems experiences packet loss probabilities in the range 0 to 1 at 1% intervals. However, only MOS estimates for parameter values that are within their permitted ranges [58][63] are shown. Consequently, for the single path system for example, MOS estimates are undefined for packet loss probabilities in excess of 0.2, and are, thus, not shown in figure 15.

The six solid curves in figure 15 plot the MOS estimates assuming bursty packet loss for the packet loss and packet loss burstiness characteristics computed by the PLB-model for packet loss probabilities 0 to 1 at 1% intervals. These curves are, from left to right, for a single path system and dispersity routing systems of 2 to 6 paths respectively. Alongside the solid curves, the corresponding six dashed curves in figure 15 plot the MOS estimates assuming non-bursty packet loss for the same packet loss, packet loss burstiness characteristics, and systems. MOS estimates for the packet loss and packet loss burstiness characteristics observed for the measured VOIP telephone calls are depicted as crosses, also, from left to

right, for a single path system and dispersity routing systems of 2 to 6 paths. As can be seen, the MOS estimates assuming burstiness and the MOS estimates for directly observed packet loss and packet loss burstiness characteristics are in good agreement. This indicates that the PLB-model relates packet loss probabilities to the state transitions of the 4-state Markov model used in section 3.1 to model the packet loss and loss burstiness characteristics of a path relatively accurately.

In addition to the MOS estimates, figure 15 also marks with horizontal dashed lines the minimum MOS values for the five user satisfaction experience interpretations enumerated in table 3. From top to bottom these lines correspond to the minimum MOS values for 'very satisfied' to 'nearly all users dissatisfied'. While the minimum MOS values do not *fix* the boundaries between the user satisfaction experience interpretations [58][62]–[63], they nevertheless do offer as a guide an indication of the estimated experience at that level.

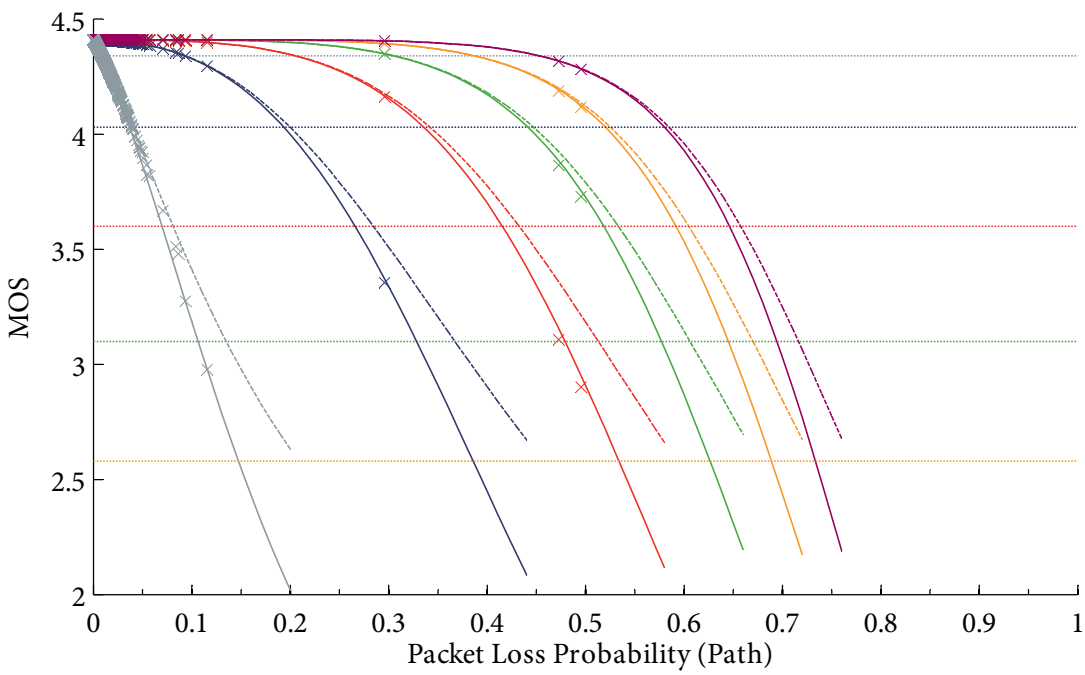

Figure 15: *Deliverable MOS estimates for systems (solid curves from left to right) of 1 – 6 paths. Dashed curves show corresponding estimates assuming non-bursty loss. Crosses show estimates for observed characteristics. Horizontal lines mark minimum user satisfaction MOS.*

3.2.3 ANALYSIS OF RESULTS

By illustrating the most likely quality improvements deliverable by dispersity routing based on the observed packet loss and packet loss burstiness characteristics of measured VOIP traffic, figure 15 demonstrates the capacity of dispersity routing to improve quality in a real environment. Futhermore, a sense of the relationships between packet loss, packet loss burstiness, numbers of paths, and deliverable quality emerges.

As a first observation, additional paths plainly result in increased quality, albeit with diminishing returns as already observed earlier in section 2.5. In figure 15 the gains delivered by adding a third path are less than the gains delivered by adding a second path, the gains delivered by the forth less than the gains by the third, and so on. Figure 16 more clearly illustrates the increases in quality and diminishing returns shown in figure 15 by plotting the MOS increases purchased with each path addition. The largest increase in estimated MOS is of 1.98 shown in figure 15 and figure 16 at packet loss probability 0.2 going from a non-dispersity routing system to a dispersity routing system of 2 paths each with the same packet

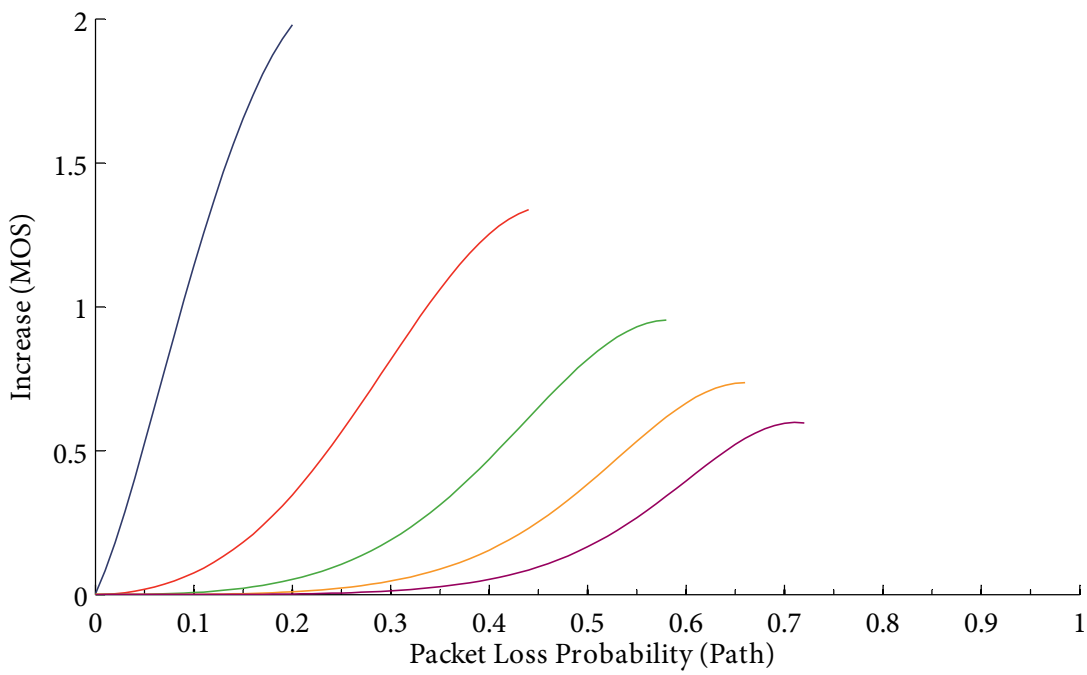

Figure 16: *Improvements in quality due to dispersity routing. The left-most curve plots increase in MOS for packet loss probabilities 0 to 1 going from non-dispersity to dispersity routing with 2 paths. Each curve is for an additional path; the right-most curve is for moving from 5 to 6 paths.*

loss probability of 0.2. At that packet loss rate, the provisional guide from table 3 interprets the quality delivered by the non-dispersity routing system as being 'not recommended'. Dispersity routing with just 2 paths at that packet loss rate increases the quality interpretation by 2 degrees to 'some users dissatisfied'. Indeed, the increase is only 0.03 MOS short of the minimum MOS for a 'satisfied' quality interpretation. Adding another path increases quality interpretation another 2 degrees to the highest possible interpretation of 'very satisfied'.

Figure 17 plots for a non-dispersity and for dispersity routing systems using 2 – 6 paths the highest tolerable packet loss probabilities assuming bursty loss for the five user satisfaction experience interpretations enumerated in table 3. The bottom curve is for the maximum packet loss probability that still delivers an interpretation of 'very satisfied', whereas the top curve plots the maximum packet loss probability that still allows an interpretation of 'nearly all users dissatisfied'. Diminishing returns of additional paths are also visible in this figure as increasing numbers of paths resulting in fewer gains in the maximum

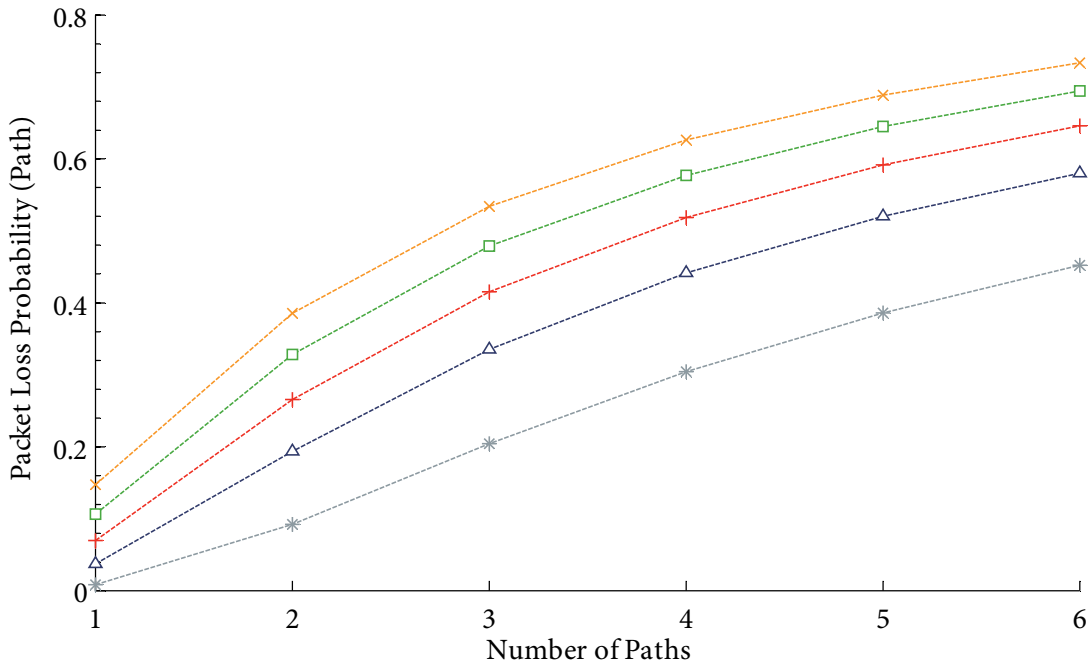

Figure 17: *Maximum tolerable packet loss probabilities on each path for the minimum MOS estimates of the five user satisfaction interpretations of (from bottom to top) 'very satisfied' (the dashed grey, bottom, line) to 'nearly all users dissatisfied' (the dashed orange, top, line), for 1 – 6 paths adopting bursty loss characteristics.*

packet loss probability that still result in that user satisfaction experience interpretation. It is clear from figure 15 and figure 17 that to deliver a user satisfaction experience interpretation of 'very satisfied', dispersity routing employing 2 paths may not use paths with packet loss probabilities exceeding 0.09. However, figure 17 elucidates clearly that increasing the numbers of path allows the same quality goal to be satisfied using paths experiencing higher packet loss probabilities. For example, increasing the numbers of paths employed from 2 to 6 allows dispersity routing to deliver the same quality of 'very satisfied', despite each path experiencing packet loss probabilities of up to 0.45.

Besides conveying the increases in quality that dispersity routing makes possible, figure 15 also reveals the degrees of freedom that dispersity routing awards. Summarily, to satisfy a particular quality goal, (1) additional paths may be added, (2) the packet loss rate may be lowered, (3) the packet loss burstiness may be decreased, or (4) a combination of these measures may be chosen. While not all of these measures may be feasible in a particular situation, figure 15 nevertheless does relate how much each measure may contribute; and knowing how much measures contribute is key to evaluating their cost-effectiveness.

Conversely, figure 15 exposes how vulnerable quality is to the reverse of these measures. Should a path be lost or the packet loss or packet loss burstiness probability increase, quality suffers. Figure 15 quantifies that effect on quality. Knowing how much these impact also helps here in assessing the cost-effectiveness of guarding against these vulnerabilities.

Further visible in figure 15 is the difference in the curves for bursty and non-bursty loss. This difference demonstrates clearly the importance of including burstiness in any MOS estimates. In figure 15 the MOS estimates that assume random loss are up to 0.61 MOS higher than those that assume bursty loss. Another change discernible in figure 15 is a change in the shape of the MOS curves as the number of paths increase. The cause of this change is dispersity routing lowering the packet loss probability and packet loss burstiness through using additional paths.

Figure 18 exemplifies how dispersity routing lowers packet loss burstiness with increasing paths. The leftmost curve plots the probabilities of a packet loss burst starting for a single path system (that is, a non-dispersity routed system), interpolated and extrapolated from the measured VOIP traffic data using the PLB-model for packet loss probabilities from 0 to 1.

From left to right, the remaining curves plot the probabilities of a packet loss burst starting for the same packet loss probabilities for dispersity routing systems of 2 to 6 paths. The packet loss burst start probabilities are computed with the QCE-model for that dispersity routing system from the packet loss burst start probabilities interpolated and extrapolated using the PLB-model for packet loss probabilities from 0 to 1.

Besides again exhibiting the diminishing returns of increasing paths, figure 18 illustrates that dispersity routing lowers packet loss burstiness [16][34]. Decreasing burstiness contributes to increasing quality. This effect is visible in figure 15, where the solid curves for deliverable MOS estimates get closer with each additional path to the corresponding dashed curves for the MOS estimates assuming non-bursty (that is, random) loss. For example, the two curves (that is, the solid curve and the dashed curve) for a dispersity routing system of 6 paths are much closer than the two curves for a dispersity routing system of 2 paths.

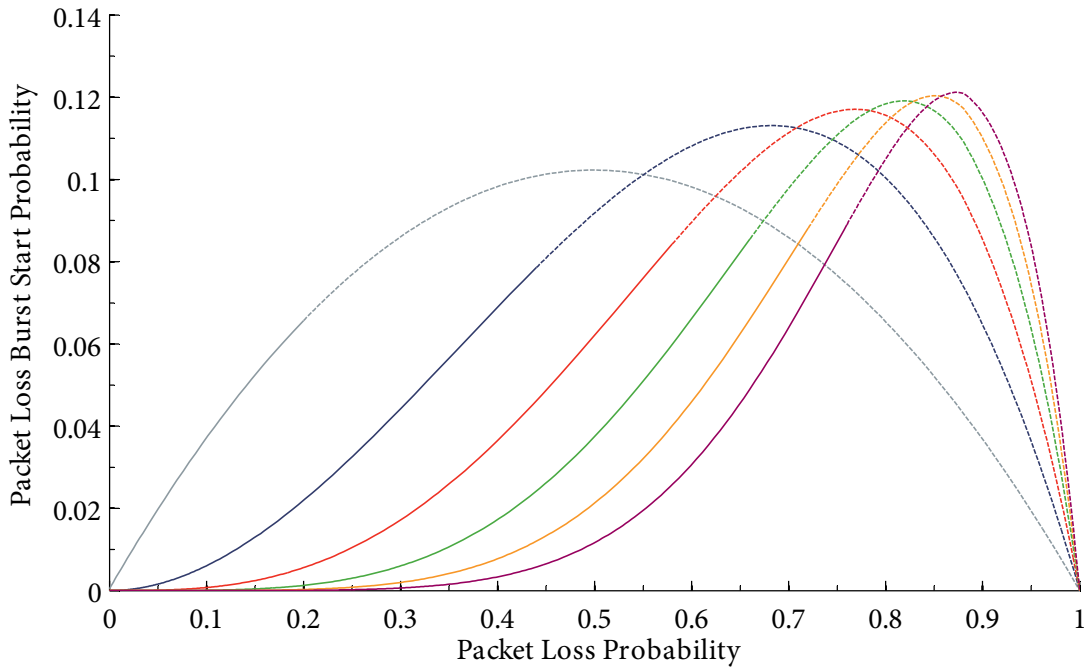

Figure 18: *Packet loss burst start probabilities for, from left to right, systems of 1 – 6 paths, computed using the PLB-model for the single path system, and the QCE-model applied to the PLB-model for dispersity routed systems of 2 – 6 paths. Characteristics for which the E-model cannot compute MOS estimates are depicted as dashed curve regions.*

Chapter 4 • Accuracy of Quality Improvement Quantification

This chapter comprises three experiments that along with analyses collectively establish the accuracy of the QCE-model in estimating deliverable quality. For two of these experiments, simulations (computed as in section 2.4) that use the measured VOIP characteristics serve as the standard against which corresponding estimates by the QCE-model adopting the same VOIP characteristics are tested.

The first of these two experiments adopts the measured VOIP characteristics as observed, and, thus, establishes how dispersity routing may improve deliverable VOIP quality in an actual setting. In contrast, the second experiment extracts observed packet loss bursts from the measurements and synthesises call profiles that represent worse conditions than those observed by condensing the extracted packet loss bursts in the synthesised call profiles. Finally, the third experiment illustrates the accuracy of the model for all possible combinations of a system of 2 paths communicating a stream of 17 packets.

While there are discrepancies between the simulated and modelled estimates, causes for which are identified, in the first two experiments they are so small as to be all but beyond human discern. In the third experiment the discrepancies are much more pronounced, but in situations that are unlikely to occur in reality, and when they do occur dispersity routing may be used to reduce the discrepancies.

4.1 EXHAUSTIVE SIMULATION

The first experiment compares the modelling results against the simulation results for all combinations of a 2-path dispersity routing system that adopts the call profiles measured (see section 2.1 above). Since a dispersity routing system that does not experience packet loss on any one of its paths when communicating a message will deliver that message without packet loss, only call profiles with packet loss are selected for this experiment. Of the 6265 measured call profiles (see section 2.2), 1714 call profiles contain packet loss. Therefore, there are $\binom{1714}{2} = 1\,468\,041$ combinations of a 2-path dispersity routing system adopting these measurements for their paths. The remaining $\binom{6265}{2} - \binom{1714}{2} = 18\,153\,939$ combinations that contain at least one call profile without packet loss always result in the same perfect and

readily quantifiable outcome. Combined, all call profiles exhaustively demonstrate how a 2-path dispersity routing system would have improved quality in the call centre with the measured VOIP telephone calls. However, because the discrepancies between the modelling and simulation results for scenarios with at least one lossless call profile are known to be zero, they are excluded from this experiment.

Figure 19 depicts cumulative distributions of the differences between simulated and modelled quality estimates computed as the simulated MOS minus the modelled MOS. Note that just as in section 3.2.1, state transition probability $p_{i,j}$ of state transition matrix **P** for a VOIP telephone call is $p_{i,j} = \frac{c_{i,j}}{L}$ where $c_{i,j}$ is the count of transitions from state i to state j and L is the total number of lost and received packets in the call profile of that call. The first cumulative distribution (shown as the solid blue line) is the difference for all combinations of the measured VOIP telephone calls in a dispersity routing system of 2 paths. 98% of the differences are within 1.40E−02, a difference so small as to be all but beyond human discern.

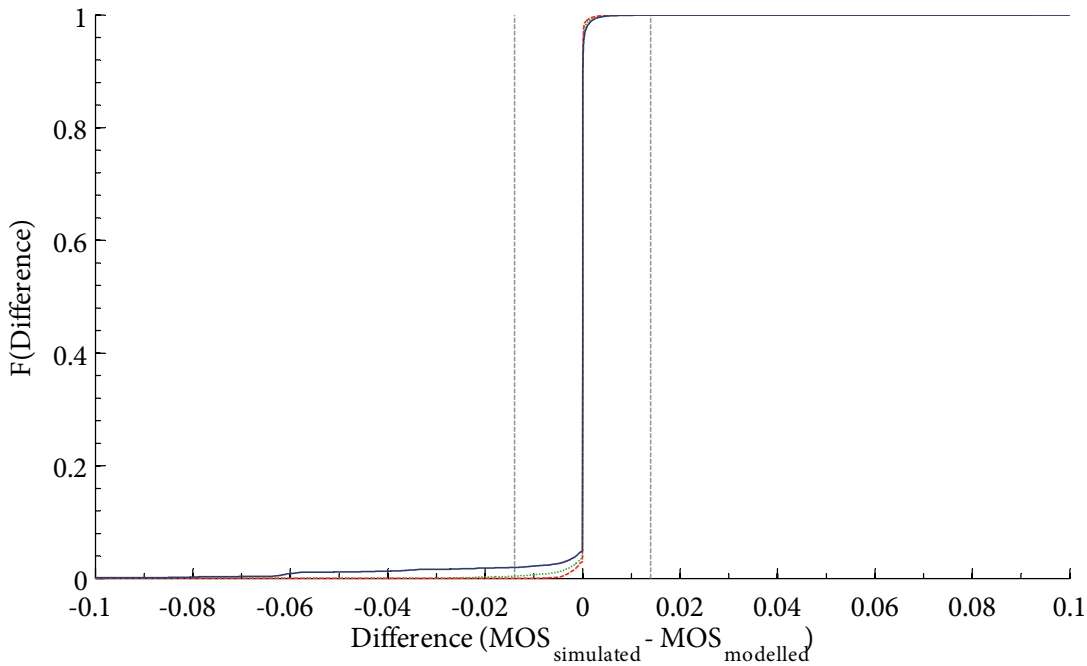

Figure 19: *Cumulative distribution of differences between modelled and simulated MOS estimates for* every *2-path combination of lossy call profiles measured. 98% (the area bounded by the vertical dashed lines) are within 1.40E−02. The dotted and dashed distributions exclude calls under 5 and 30 seconds respectively.*

Indeed, 50% of the differences are within 1.33E−07 MOS, which is even less discernible. Also shown in figure 19 are the cumulative distributions (as the dotted green line) of the differences using measurements taken for VOIP telephone calls with a duration of at least 5 seconds only, and (as the dashed red line) of the differences using measurements taken for VOIP telephone calls that are at least 30 seconds only.

The extrusion into the negative in the bottom left of figure 19 of the solid blue distribution for the differences using measurements for all VOIP telephone calls is caused by the simulations computing a much lower MOS estimate than the model for some combinations. For these combinations, the model overestimates the MOS; that is, the actual MOS will be no higher than the estimate computed by the model. The extrusion becomes smaller when excluding scenarios that use measurements for VOIP telephone calls of a short duration, as depicted by the dotted green and dashed red distributions. The dotted green distribution excludes VOIP telephone calls less than 5 seconds in duration and the dashed red distribution those less than 30 seconds.

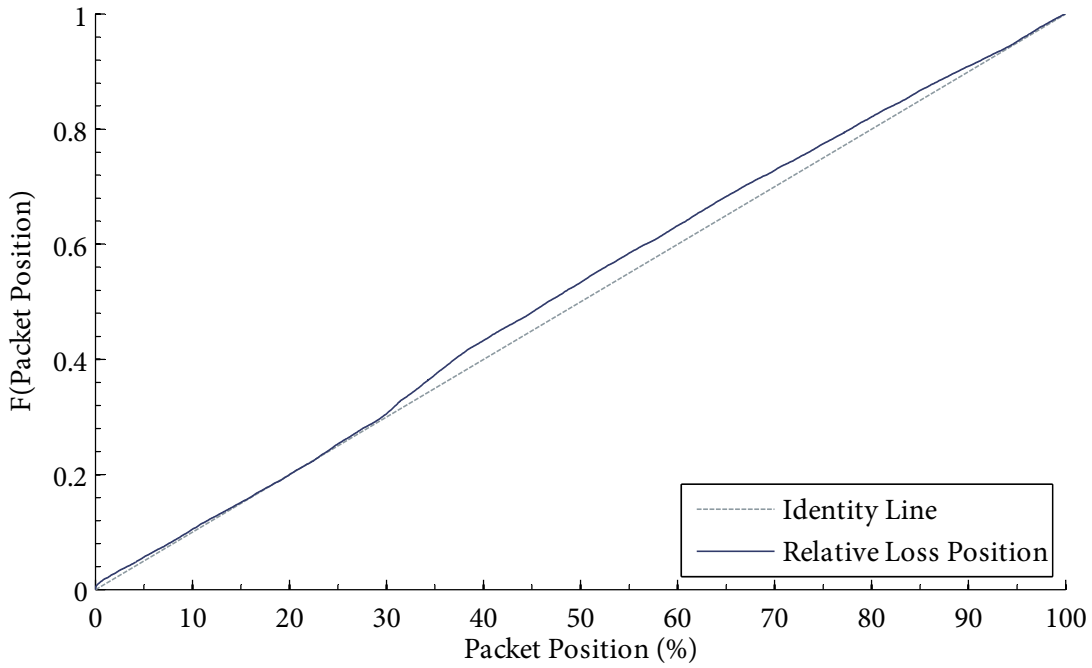

Figure 20: *Cumulative distribution of the relative position (where 0% is at the beginning of the VOIP telephone call and 100% is at the end) of the packets observed as lost in the 6265 measured VOIP telephone calls. Also shown for reference is the identity line.*

Figure 20 depicts the cumulative distribution of the relative position of every lost packet observed in the measured VOIP telephone calls, where 0% is at the beginning of the call and 100% is at the end of the call. While packet loss is fairly evenly distributed over the VOIP telephone call durations, there are nevertheless some biases. The most prominent bias that may be seen in figure 20 occurs between 30% and 40% of VOIP telephone call durations, which last to between 70% and 95%. However, another bias exists at the beginning of VOIP telephone calls, depicted clearly in figure 21. While that bias does not rise as high, it does rise more steeply, that is, loss occurring at the beginning of VOIP telephone calls is biased to occur in a more confined portion of the calls. Indeed, the first percentile of packet loss occurs in the first 0.32% of VOIP telephone calls, and the first half percentile within the first 0.05%. Intuitively, a data flow needs to "settle in" along its path, while the various components along the path accommodate the new data flow.

This bias in the measurements for loss to occur at the beginning of a call is more pronounced for shorter VOIP telephone calls, and possibly the reason for these VOIP

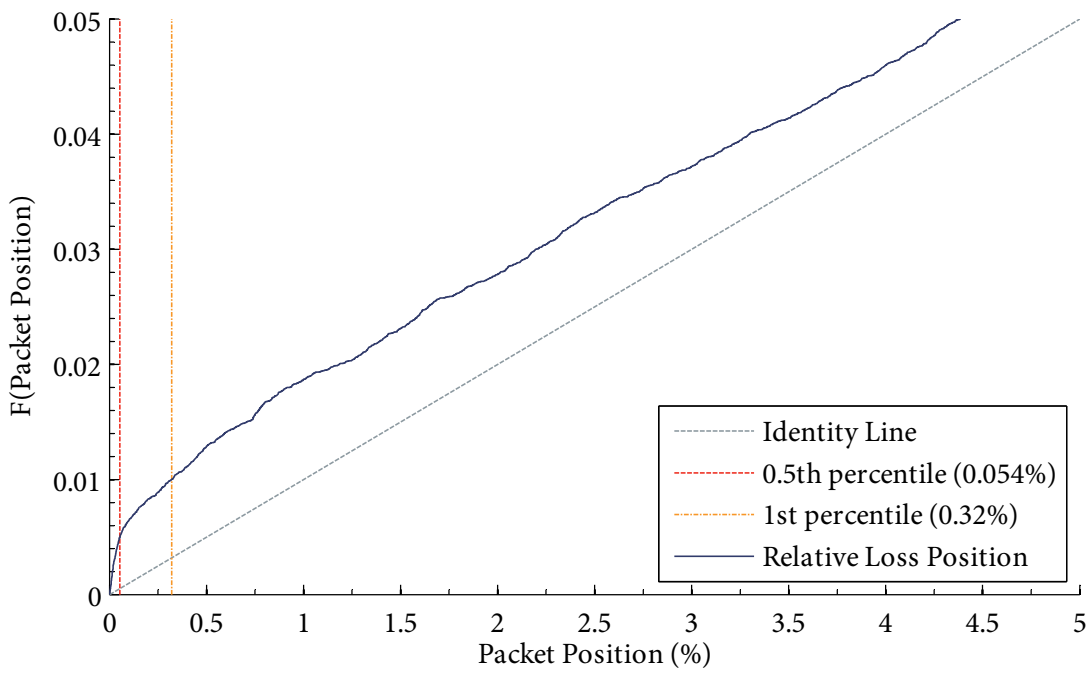

Figure 21: *Cumulative distribution of relative position of packet loss occurring within first 5% of VOIP telephone calls. Also shown are the identity line and the 0.5th and 1st percentile of all packet loss. The 1st percentile of packet loss occurs within the first 0.32% of VOIP telephone calls.*

telephone calls to be that short in the first place. For example, for measured calls under 30 seconds, the first 20th percentile of packet loss occurs within the first 1.6% of these calls, and for VOIP telephone calls under 5 seconds, the first 50th percentile within the first 1.3%. Clearly, this bias causes discrepancies between simulation and model results, because the simulation computes the exact outcome whereas the model computes the most likely outcome based on the probability of packet loss, from which a MOS estimate is then computed.

As an aside, note also the *absence* of a significant bias at the *end* of telephone calls. If telephone calls tended to be ended when packet loss occurs, that bias would be reflected with a corresponding bias for packet loss to occur at the end of telephone calls. However, the slight bias that occurs in figure 20 between 94.6% and 97.8% of VOIP telephone call durations suggests only that telephone calls are ended sometimes when packet loss occurs.

Removing short VOIP telephone calls from the experiment reduces the extrusion in figure 19, as illustrated by the dotted green and the dashed red distributions, which exclude

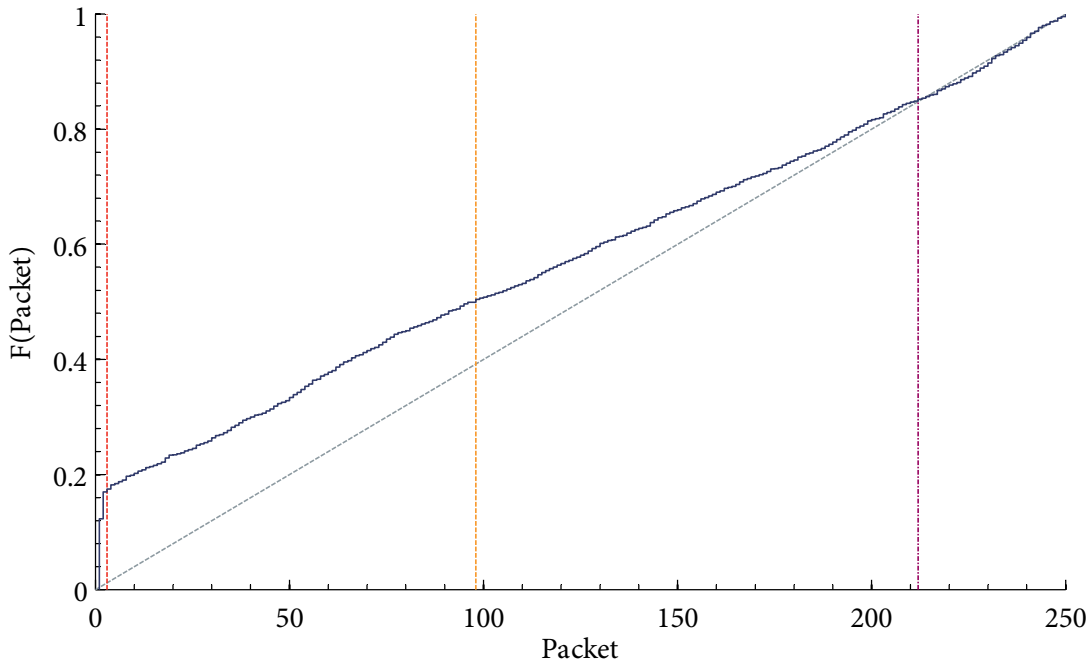

Figure 22: *Cumulative distribution of absolute position of packet loss occurring within first 250 packets of VOIP telephone calls of at least 250 packets. Also shown are the identity line and (left to right) the first 17th, 50th and 85th percentiles of packet loss occurring in these first 250 packets.*

calls under 5 and 30 seconds respectively. For VOIP telephone calls that are 30 seconds or longer in duration, 99.7% of model estimates are within 7.56E−02 MOS of the corresponding simulation estimates, 98% within 1.77E−03 MOS, and 50% are within 1.92E−07 MOS, which are even less discernible than the discrepancies between the estimates for all measurements. Note that only short calls were excluded from these distributions; packet losses occurring at the beginning of included VOIP telephone calls were *not* excluded.

This correlation in bias for packet loss at the beginning of VOIP telephone calls is a problem for dispersity routing, which depends on at least one path delivering a packet to mask any loss of that packet on other paths. If paths are correlated in a bias to lose packets at the beginning of a data stream, then dispersity routing may not be able to mask loss at the beginning of calls without additional paths during that portion of the call. However, in practice the beginning of a call typically communicates ring tones rather than voice conversation.

4.2 SHUFFLED SIMULATION

The second experiment compares the modelling results against the simulation results of 25 sets of simulations of a dispersity routing system using 2 paths. Each simulation synthesises 250 random call profiles as described below of at least 3000 packets from the 6265 call profiles measured (see section 2.1 above), resulting in $\binom{250}{2} = 31\,125$ scenarios in each simulation set. Just as in sections 3.2.1 and 4.1, state transition probability $p_{i,j}$ of state transition matrix $\mathbf{P}$ for a synthesised call profile is $p_{i,j} = \frac{c_{i,j}}{L}$ where $c_{i,j}$ is the count of transitions from state i to state j and L is the total number of lost and received packets in that call profile.

By synthesising call profiles from the measurements, not only are the observed characteristics randomised, but threats to quality may also be condensed to quantify the performance of dispersity routing and the accuracy of the QCE-model under worse conditions than those observed. This differs from the first experiment which establishes how dispersity routing performs in observed conditions. In contrast, this experiment establishes how dispersity routing performs in worse conditions than those observed by extracting observed packet loss bursts from the measurements and synthesising call profiles from a condensing of these extracted packet loss bursts.

4.2.1 SYNTHESISING SCENARIOS WITH CONDENSED QUALITY THREATS

Let $\mathcal{E}$ be the ordered set of loss bursts (loss bursts as defined in section 3.1) extracted from all measured call profiles, where $e_i \in \mathcal{E}$ is the ith loss burst in $\mathcal{E}$. There are 21 430 loss bursts in the 6265 call profiles measured (see section 2.1). Furthermore, let the length of loss burst e_i be $|e_i|$, and the number of losses in loss burst e_i be E_i. Given that function ψ quantifies the impact of loss burst e_i as

$$\psi(i) = E_i \cdot \frac{E_i}{|e_i|} = \frac{(E_i)^2}{|e_i|}, \tag{19}$$

$\mathcal{E}$ is, therefore, ordered such that higher impact losses, as defined by ψ, occur before lower impact losses, that is, $\psi(i) \geq \psi(i+1)\, \forall i \in \{1,2,\ldots,|\mathcal{E}|-1\}$. That is, given that the impact of a loss burst is quantified by equation (19) as the product of the number of losses in the loss burst and the *density* of the loss burst (the proportion of the number of losses in relation to the length of the loss burst), higher impact loss bursts occurs before lower impact loss bursts in $\mathcal{E}$. Impact is quantified in equation (19) as the product of the number of losses in the loss burst and the *density* of the loss burst, because, informally, the impact on telephonic speech quality of a packet loss burst is greater the more packet loss occurs in that packet loss burst and the denser it is in that packet loss burst.

A call profile of at least L packets is synthesised as an initial gap of consecutively received packets followed by a sequence of loss bursts separated by gaps of at least $\mathrm{G}_{\min}$ consecutively received packets. The initial gap has a length selected as a uniformly distributed random value from the integer interval $[1 \,..\, \mathrm{G}_{\min}]$. In this experiment, subsequent gaps always have a length of exactly $\mathrm{G}_{\min}$. As in section 3.1, $\mathrm{G}_{\min}$ is fixed to 16 as recommended by [80]. Loss bursts are selected from $\mathcal{E}$ by a gamma-distributed random value to give preference to higher impact bursts. The gamma distribution has shape, k, of 1, and a scale, θ, of $\frac{|\mathcal{E}|}{2k}$ to give the distribution a mean of half the number of bursts.

Let α_0 be the initial gap, $|\alpha_0|$ the (uniformly random) length of α_0, and α_x (where $x > 0$) any subsequent gap of length $|\alpha_x| = \mathrm{G}_{\min} = 16$ packets. Furthermore, let the loss burst β_x, where $x > 0$, be of length $|\beta_x|$. Let T, then, be a sequence of truth values, where a truth value is either $\top$ (that is, true) or $\bot$ (that is, false), such that $\top$ represents a packet loss and $\bot$ a

packet receipt. The truth value sequence T, as depicted in figure 23, is constructed by concatenating α_0 with β_i followed by α_i, where $i \in \{1,2,\dots\}$, until the sequence contains at least L truth values subject to the last truth value in T being $\perp$. That is, when the Lth truth value is $\top$, continue to append truth values into the sequence until a $\perp$ is appended into the sequence such that the last truth value in T is $\perp$. Furthermore, let $t^{(i)}$ be a subsequence of T that corresponds to the ith loss burst β_i. Subsequence $t^{(i)}$ begins in T with position $|\alpha_0| + \sum_{j=1}^{i-1}\left(|\beta_j| + |\alpha_j|\right) + 1$, and ends with position $|\alpha_0| + \sum_{j=1}^{i}|\beta_j| + \sum_{j=1}^{i-1}|\alpha_j|$. Loss burst β_i defines the truth values for subsequence $t^{(i)}$ such that $t_j^{(i)}$ is $\perp$ when the jth packet in loss burst β_i is received and $\top$ when it is lost. Clearly, all truth values in any subsequence of T that correspond to a gap in this experiment are $\perp$.

Packet characteristics for the call profile being synthesised are read from the pool of measured call profiles. Upon commencing synthesis of a call profile, select a call profile from the pool using a uniformly distributed random variable. Let $\mathcal{M}$ be the set of received packets from that first selected call profile. Synthesise the call profile by adopting the characteristics of the packets in $\mathcal{M}$ for the packets that are not lost in the call profile being synthesised. That is, for all instances of truth value $\perp$ in T, adopt the packet characteristics of a packet, such that the first $\perp$ in T adopts the characteristics of the first packet in $\mathcal{M}$, the second $\perp$ in T adopts the characteristics of the second packet in $\mathcal{M}$, and so on. When $\mathcal{M}$ is exhausted, select

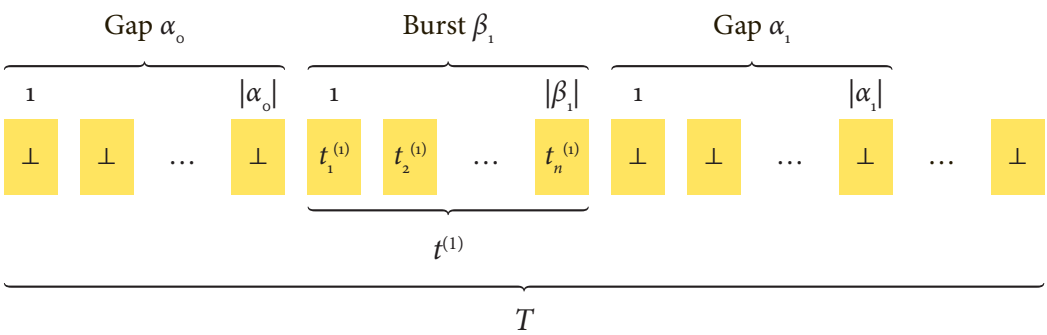

Figure 23: *Composition of sequence T from loss bursts selected from $\mathcal{E}$ at random. Loss bursts are separated by gaps, with the first gap of a length in the integer interval $[1\,..\,G_{min}]$. As the sequence must end with a value of $\perp$, the length of T may exceed L.*

another call profile from the pool using a uniformly distributed random variable, and let $\mathcal{M}$ be the set of received packets from that call profile.

4.2.2 ACCURACY OF MODEL FOR SYNTHESISED SCENARIOS

Figure 24 depicts the cumulative distributions of the differences between simulated and modelled quality estimates for the 25 sets of simulations (as the dotted grey distributions) and all simulations (as the solid blue distribution). Just as in section 4.1 above, differences are computed as the simulated MOS minus the modelled MOS. Note that the solid blue distribution for all simulations is slightly skewed and too acute to be Gaussian, as illustrated by the best-fitting Gaussian distribution depicted as the dashed red distribution. However, the 50th percentile (the single vertical green dashed line) of the distribution for all simulations is 8.95E−04, which is close to 0.

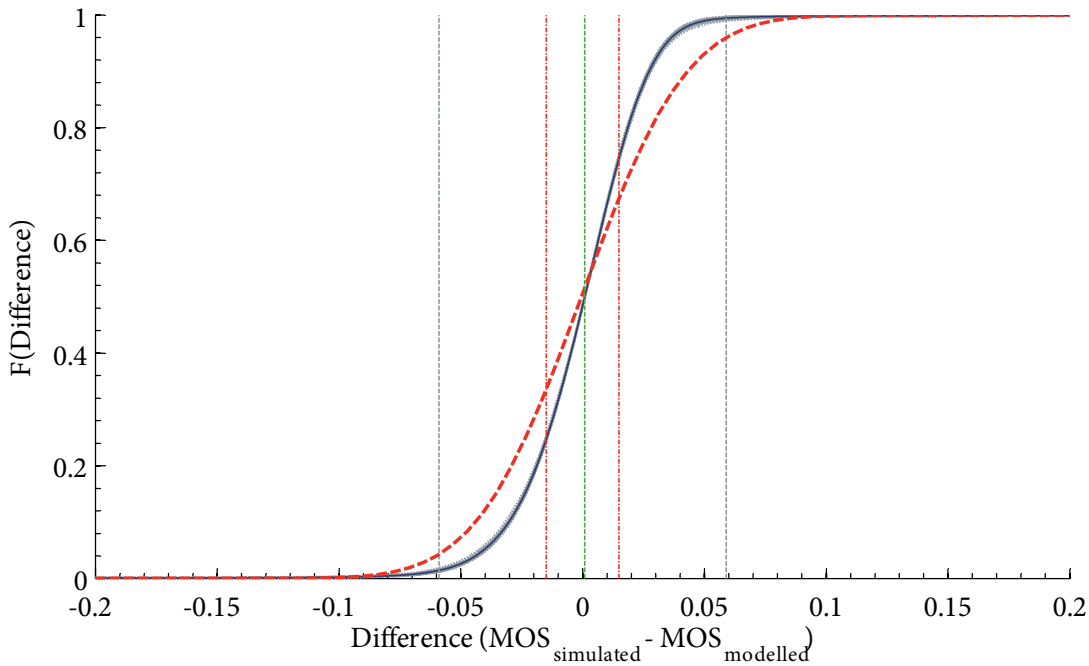

Figure 24: *Cumulative distribution of differences between modelled and simulated MOS estimates for every 2-path combination of 250 synthesised call profiles in 25 sets of simulations. 98% (the area bounded by the vertical dashed lines) are within 5.90E−02, and 50% (the area bounded by the dash-dotted lines) are within 1.50E−02 of the MOS estimate.*

As can be seen, 98% of the differences (marked by the pair of vertical dashed grey lines) are within 5.90E−02, and 50% (marked by the two vertical dashed-dotted red lines) are within 1.50E−02. These are larger than the values of 1.40E−02 for 98% and 1.33E−07 for 50% of the differences in figure 19. However, the 98% value is still smaller than a tenth of the smallest distance between the numerical scores assigned to the opinions in most of the scales (such as the conversation opinion scale) used to assess telephonic speech quality (see section 1.1.2 above). Therefore, the modelled estimates may be considered to be in fairly good agreement with the simulated estimates. Furthermore, these values are for synthesised call profiles that comprise densely packed loss bursts with a preference given to higher impact loss bursts and which, overall, represent conditions much worse than those observed. The consistently higher threat to quality is also the reason for the significant increase for the 50% value from 1.33E−07 in figure 19 to 1.50E−02 in figure 24, as explained in section 4.3.2 below.

4.2.3 DISPERSITY ROUTING PERFORMANCE FOR SYNTHESISED SCENARIOS

As well as quantifying the accuracy of the model in estimating the quality that dispersity routing may deliver for the scenarios synthesised as described in section 4.2.1 above, the quality deliverable under those conditions itself is quantified. Figure 25 depicts the cumulative distributions of the quality, expressed as a MOS estimate, deliverable under those conditions, as computed through simulation and by the model. The solid blue distribution depicts the distribution of the MOS estimates computed through simulation, whereas the dashed red distribution depicts the distribution of the MOS estimates computed by the model. Also shown are the 5th percentiles for (1) comparison of the simulated and the modelled distributions and for (2) comparison of the deliverable quality for these synthesised scenarios with that for the empirical scenarios discussed in section 2.5 above. For reference, the maximum possible MOS for a lossless output is also marked, and the minimum MOS values for the top two user satisfaction interpretations (see table 3).

Besides illustrating that the modelled and simulated distributions match fairly well, figure 25 also quantifies the quality that dispersity routing may deliver with just two paths despite facing the threats to quality elevated significantly in this experiment. Figure 26 depicts the cumulative distributions of the probabilities of loss of the synthesised call profiles, both for

each of the 25 sets of simulations (as the dashed grey distributions) and for all synthesised call profiles overall (as the solid blue distribution).

The similarity in the distributions shown in these two figures highlights the significance of the impact of loss on perceived quality. About 10% of synthesised call profiles have an elevated loss rate, visible as the extrusion towards the top right of figure 26. This elevated loss rate manifests itself in figure 25 as the extrusion visible towards the bottom left. The elevated loss rate, resulting from a condensing of threats to quality while synthesising the call profiles used in this simulation (see section 4.2.1), causes a corresponding decrease in MOS estimates for about the same proportion of scenarios. The same applies to the small proportion of synthesised call profiles with a lower loss rate (visible as the small curving of the distribution towards the left at the bottom of figure 26), which causes a corresponding proportion of scenarios with a higher MOS estimate (visible as the small curving of the distribution towards the right at the top of figure 25). Finally, the large proportion of synthesised call

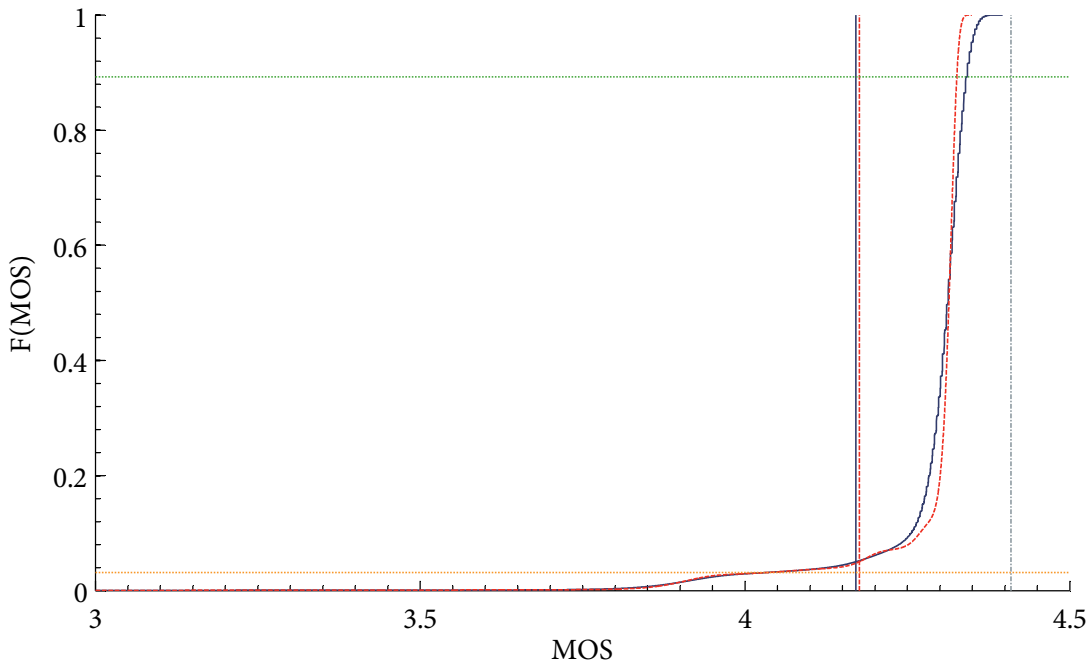

Figure 25: *Distribution of MOS estimates by simulation (solid blue distribution) and by model (dashed red distribution). Also shown are 5th percentiles for simulation (solid blue vertical line) and model (dashed red vertical), maximum MOS (dash-dotted grey vertical) and simulated lower thresholds for* very satisfied *(dotted top horizontal) and* satisfied *(dotted bottom).*

profiles with a loss probability of around 0.1 is visible as the large proportion of scenarios with a MOS estimate of around 4.31. The higher loss rate prevents the MOS from reaching the maximum possible MOS of 4.4094 for a lossless output, as was observed in sections 2.2 and 2.5 above.

While only 10.77% of simulated scenarios adopting the synthesised call profiles deliver a call quality that may be interpreted to be very satisfying using the provisional MOS interpretation guide shown in table 3 above, 86.10% may be interpreted as satisfying. The threshold between very satisfying and satisfying scenarios is marked in figure 25 by the dotted green (top) horizontal line and the threshold between satisfying and less than satisfying scenarios by the dotted orange (bottom) horizontal line.

4.3 COMBINATORIAL SIMULATION

The third experiment compares modelling results against simulation results for a 2-path dispersity routing system communicating a stream of 17 packets for every possible

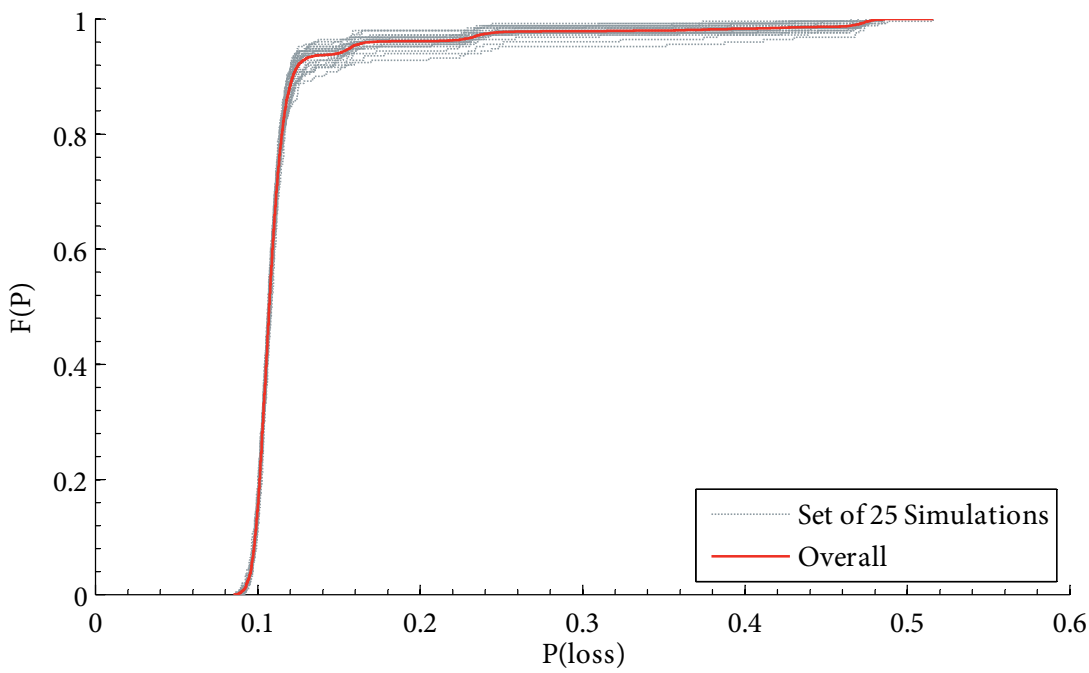

Figure 26: *Cumulative distributions of packet loss probabilities for synthesised call profiles in each of the 25 sets of simulations (as the dotted grey distributions), and for all synthesised call profiles (as the solid blue distribution). Mean packet loss probability overall is 0.1177.*

combination of packet loss on these two paths. Since there are $2^{17} = 131\,072$ possible packet loss combinations for a stream of 17 packets, for 2 paths communicating a stream of 17 packets there are $2^{17^2} = 17\,179\,869\,184$ possible combinations. Just as in sections 3.2.1, 4.1 and 4.2, state transition probability $p_{i,j}$ of state transition matrix **P** for a combinatorial call profile is $p_{i,j} = \frac{c_{i,j}}{L}$ where $c_{i,j}$ is the count of transitions from state i to state j and L is the total number of lost and received packets in that call profile.

This experiment quantifies the accuracy of the model for all combinations possible in the arrangement selected for this experiment, rather than observed scenarios and scenarios that may be interpolated and extrapolated from observed measurements. All packet loss probabilities that are possible, from 0 to 1, are covered in this experiment.

4.3.1 ACCURACY OF MODEL FOR COMBINATORIAL SIMULATION

Figure 27 depicts the cumulative distribution of the differences between the simulated and the modelled MOS estimates for all 2^{17^2} combinations as the solid blue distribution. Also shown

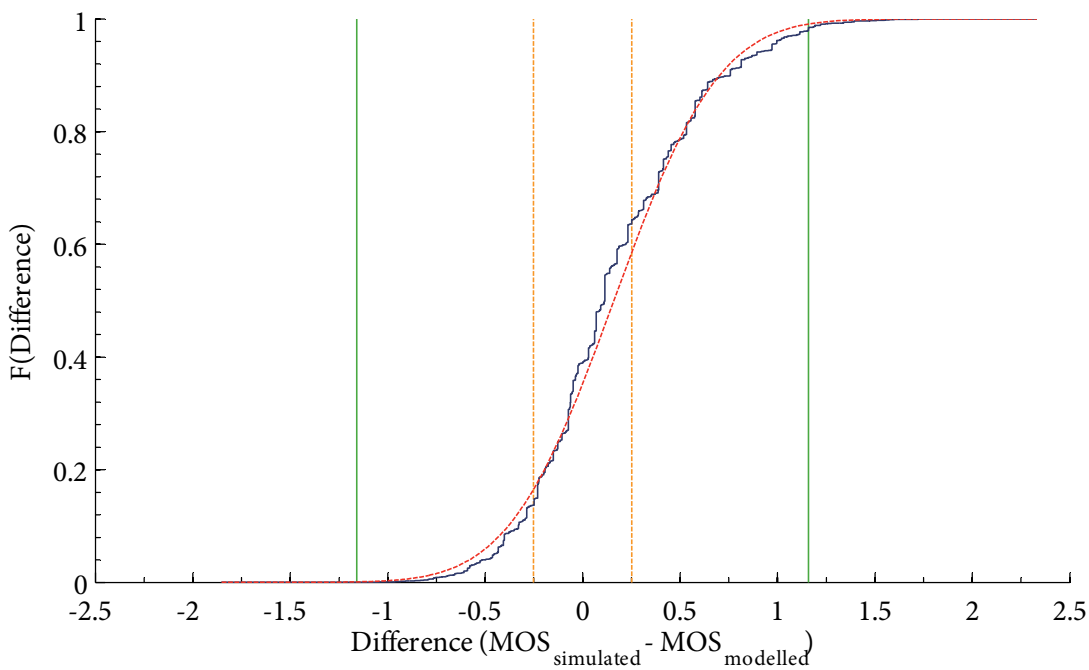

Figure 27: *Distribution of differences between modelled and simulated MOS estimates for every combination possible for stream of 17 packets. 50% are within 0.25, 98% within 1.16 MOS. Mean of 1.13E−01 suggests model underestimates MOS. MOS not computable for 75.58% of scenarios.*

for reference, as the red dashed distribution, is the best fitting Gaussian distribution. 98% of the differences between simulated and modelled MOS estimates, marked by the area bounded by the outer solid green vertical lines, are within 1.16. This difference is *not* beyond human discern. Indeed, a discrepancy of this magnitude, with a range of 2.32 MOS, covers 46.32% of the entire MOS scale. While 50% of the differences (marked by the area bounded by the inner dashed orange vertical lines) are within a lesser 0.25 MOS, this range still covers 10.05% of the MOS scale. Clearly, the QCE-model has limitations.

The 50th percentile of 1.13E−01 suggests that the QCE-model underestimates the MOS for the 24.42% of scenarios for which both simulated and modelled MOS estimates are computable. That is, in practice the MOS is actually slightly higher, making the estimate by the QCE-model a worst-case estimate.

Figure 28 and figure 29, respectively, depict the distributions of the differences in the packet loss probabilities and packet loss burst probabilities between the same modelled and simulated scenarios as in figure 27. These two probabilities are computable for every scenario,

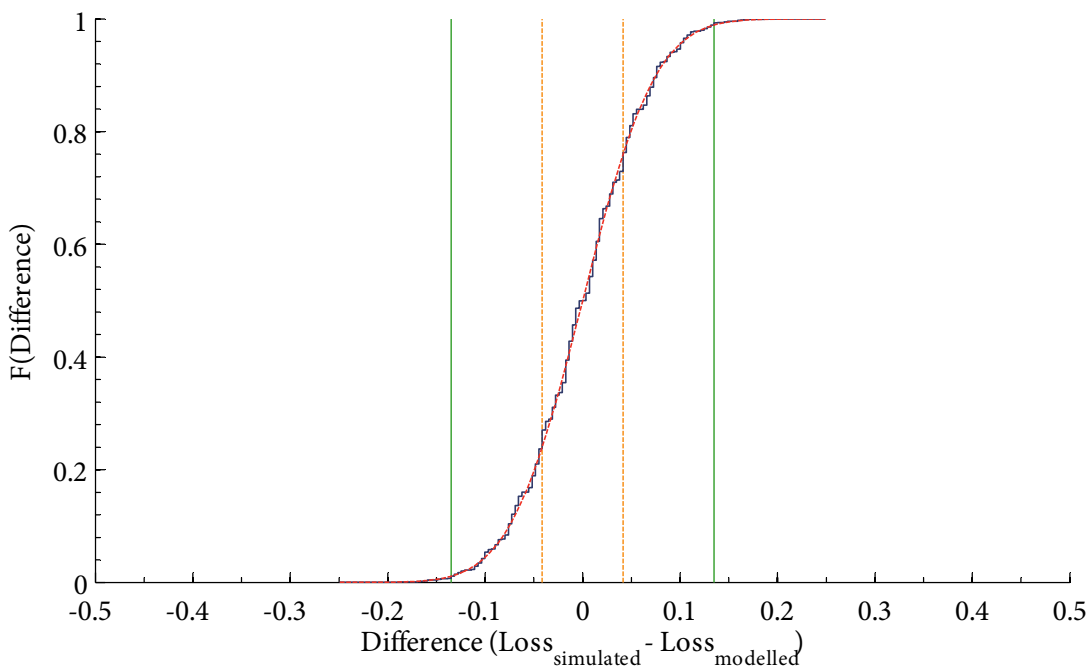

Figure 28: *Distribution of differences between modelled and simulated loss estimates for every combination possible for a stream of 17 packets. 50% of differences are within 4.15E−02, and 98% within 1.35E−01 packet loss probability. The mean is 0.*

unlike the MOS which is computable for limited parameter ranges only (for example, the E-model's permitted range for the parameter that quantifies the packet loss probability is 0 to 0.2). Because these probabilities are computable for every scenario, the distributions of the differences are more Gaussian with arithmetic means that are closer to 0 than the distribution of the differences for the MOS estimates. However, most notably, the ranges of discrepancies are much smaller than the range of discrepancies in the MOS shown in figure 27.

4.3.2 QUANTIFYING DISCREPANCIES

These discrepancies are to be expected, since the (continuous) mathematical QCE-model computes the single most likely estimate as the *expected value* while other outcomes are possible and which the (discrete) simulations compute. As seen in figure 27, figure 28, and figure 29, these other outcomes cluster around that most likely estimate computed by the QCE-model.

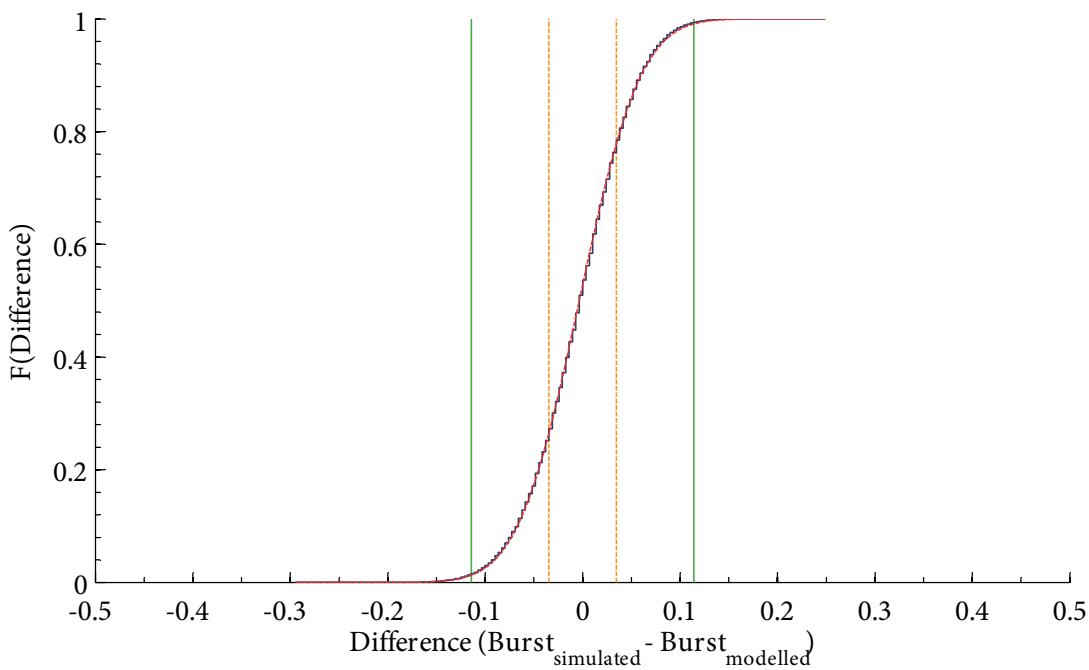

Figure 29: *Distribution of differences between modelled and simulated packet loss burst probability estimates for every combination possible for stream of 17 packets. 50% of differences are within 3.46E−02, 98% within 1.142E−01 packet loss burst probability. Mean is −2.77E−02.*

In explaining this discrepancy, consider an example involving two common six-sided dice. The most frequently occurring sum of throwing two such dice is 7 (with a probability of $\frac{6}{36}$), but there are 10 other possible sums ranging from 2 to 6 (with probabilities $\frac{1}{36}, \frac{2}{36}, \ldots, \frac{5}{36}$ respectively) and 8 to 12 (with probabilities $\frac{5}{36}, \frac{4}{36}, \ldots, \frac{1}{36}$ respectively). Similarly, the QCE-model computes an estimate as the expected value whereas the simulations compute actual values.

Applying the example above to packet loss since that is, while not the sole, the most significant factor in the estimation as described in section 3.1 of telephonic speech transmission quality as a MOS, let n be the number of packets to be communicated. Furthermore, let $\mathcal{C}$ be the set of lost packet counts on each path such that c_i, the ith element in $\mathcal{C}$, describes the number of packets lost on path r_i. The number of ways in which this packet loss can be arranged is $\prod_{i=1}^{|\mathcal{C}|} \binom{n}{c_i}$. Since a packet must be lost on all paths for a dispersity routing system to also lose that packet, the highest possible packet loss that a dispersity routing system may deliver is

$$b_{\text{worst}} = \min(\mathcal{C}) \,. \tag{20}$$

That is, the packet loss delivered by a dispersity routing system cannot be worse than that of its best path. Conversely, the lowest possible packet loss that a dispersity routing system may deliver is the minimum overlap possible with (1) the best path and (2) the worst path. If there is no overlap, then the best case is no loss. Therefore, the lowest possible packet loss is

$$b_{\text{best}} = \max(\min(\mathcal{C}) + \max(\mathcal{C}) - n, 0) \tag{21}$$

The discrete probability distribution of the packet loss between these two extremes that a dispersity routing system delivers may be computed readily. Let $k \in \{b_{\text{best}}, b_{\text{best}} + 1, \ldots, b_{\text{worst}}\}$ be the packet loss count for which to compute the number of ways in which that packet loss can be arranged. For each value of k, partition the stream as in figure 30 into (1) the k packet losses occurring on paths $r_1, r_2, \ldots, r_N$ at the same time (the right partition marked *loss*) and (2) the $n - k$ packets remaining in the stream (the left partition marked *non-loss*). Since for k only the k packet losses occur at the same time, any combination that includes packet loss in the $n - k$ remaining packets (the left partition marked *non-loss*) is excluded for k. A combination includes packet loss when a packet is lost (depicted in figure 30 as $\top$) on all N

paths. Assuming $N = 3$, $n = 6$ and $k = 3$, in the example depicted in figure 30, packet $n - k = 6 - 3 = 3$, which is in the non-loss partition, would be lost because all N paths lose that packet. This combination is excluded, along with other combinations that result in packet loss in the non-loss partition, because the k packet losses occur in the loss partition, leaving no packet loss to occur in the non-loss partition for k. Therefore the probability density function of the discrete distribution describing the discrepancy in packet loss is given by

$$\mathrm{p}(k; \mathcal{C}, n) = \begin{cases} \dfrac{\omega(k; \mathcal{C}, n)}{\prod_{i=1}^{|\mathcal{C}|} \binom{n}{c_i}}, & \text{if } b_{\text{best}} \leq k \leq b_{\text{worst}} \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

where $\omega(k; \mathcal{C}, n)$ is defined as

$$\omega(k; \mathcal{C}, n) = \binom{n}{k} \cdot \left(\prod_{i=1}^{|\mathcal{C}|} \binom{n-k}{c_i - k} - \sum_{j=1}^{\min(\mathcal{C})-k} \omega(j; \{c - k : c \in \mathcal{C}\}, n - k) \right). \tag{23}$$

For the purposes of clarity, $\binom{n}{k}$, the left factor in equation (23), quantifies the number of ways that the loss partition in figure 30 can be arranged in the n packets of the stream. The number of ways in which the non-loss partition can be arranged without resulting in a packet loss (because the k lost packets are already quantified in the loss partition) is quantified by the right factor in equation (23). This non-loss partition comprises the minuend $\prod_{i=1}^{|\mathcal{C}|} \binom{n-k}{c_i - k}$ which

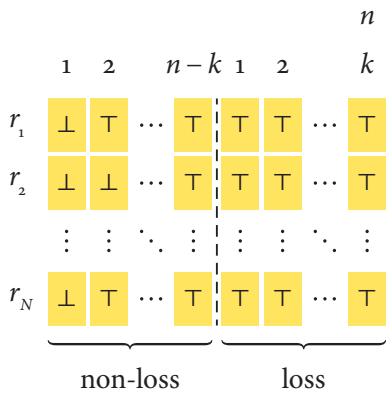

Figure 30: *Example of computing packet loss discrepancy distribution for a dispersity routing system of N paths communicating n packets. Density at k packet losses is given by (1) the count of arranging the* loss *partition in the stream's n packets, times (2) the count of arranging* non-loss *from which (3) the count of arranging potentially overlapping* non-loss *has been subtracted.*

quantifies the number of ways that the rest of the stream can be arranged, and the subtrahend $\sum_{j=1}^{\min(\mathcal{C})-k} \omega(j; \{c-k : c \in \mathcal{C}\}, n-k)$ which quantifies, recursively, the number of ways in which any potential loss in the non-loss partition may be arranged and which must be excluded since it is not included in k.

Figure 31 plots with equation (22) the probability density functions of the packet loss for 4 dispersity routing systems that use 2 paths to communicate a data stream of 17 packets. Plotted in solid blue, the second system from the right loses 9 packets on each path. The (continuous) expected value for deliverable packet loss of $\left(\frac{9}{17}\right)^2 \cdot 17 = 4.76$ packets by the dispersity routing system is closest to the most frequently occurring packet loss count of 5 packets in the (discrete) probability density functions. In contrast, the second dispersity routing system from the left (plotted in dashed green) loses 5 packets on each path, with an expected value for deliverable packet loss of $\left(\frac{5}{17}\right)^2 \cdot 17 = 1.47$ packets. Again, this is closest to the most frequently occurring packet loss count for this system of 1 packet.

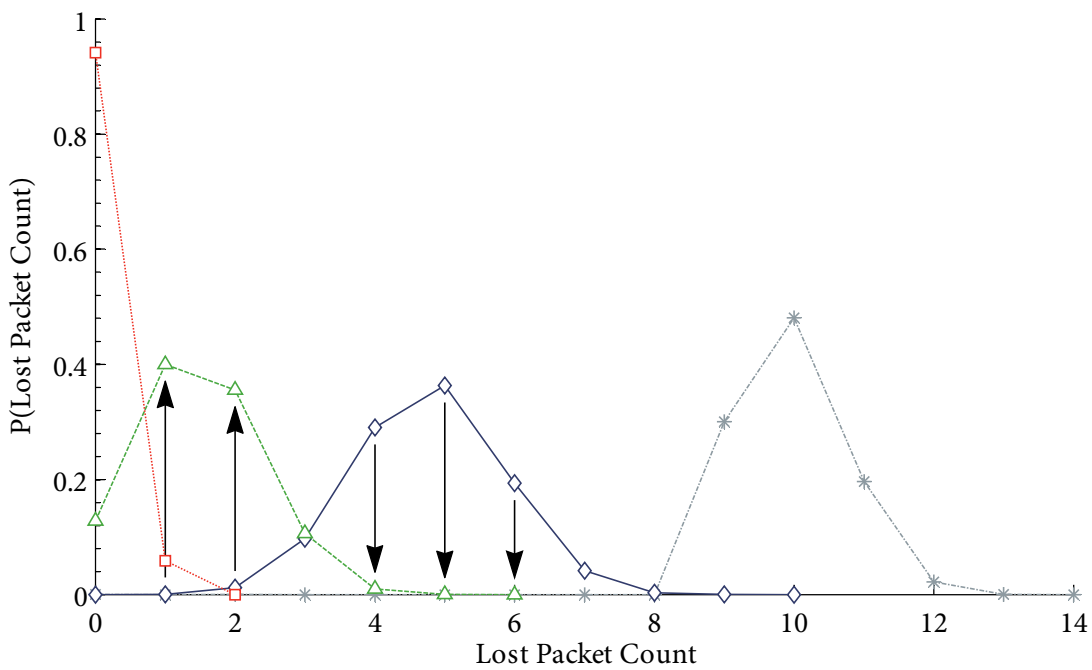

Figure 31: *Probability density functions for 4 dispersity routing systems using 2 paths to communicate 17 packets. The paths of the dispersity routing systems, from right to left, each experience 13, 9, 5, 1 packet losses respectively. Decreasing packet loss pushes the curves to the left.*

Reducing the packet loss count by 4 packets on each path pushes the probability density function to the left by reducing the probability of larger packet loss counts and increasing the probability of smaller packet loss counts, as indicated by the arrows in figure 31 that illustrate the transformation of the solid blue function to the dashed green function. Further decreasing the packet loss count by 4 packets on each path to 1 packet pushes the probability density function further to the left, as shown by the dotted red function. This reduces the range of packet loss from $b_{\text{worst}} - b_{\text{best}} + 1 = 9$ for the second-right system to $b_{\text{worst}} - b_{\text{best}} + 1 = 1$ for the leftmost system. Shown also, as the rightmost probability density function, is a system experiencing 13 packet losses on each path. As can be seen, the probability density functions for systems in the midrange are slightly more spread out than those at the extremes. Note also the diminishing returns gained by reducing packet loss on each path from 13 to 9, to 5 packets, and then to 1 packet.

Increasing the numbers of paths that a dispersity routing system uses has, unsurprisingly, a similar effect on the probability density function as reducing packet loss does. Figure 32

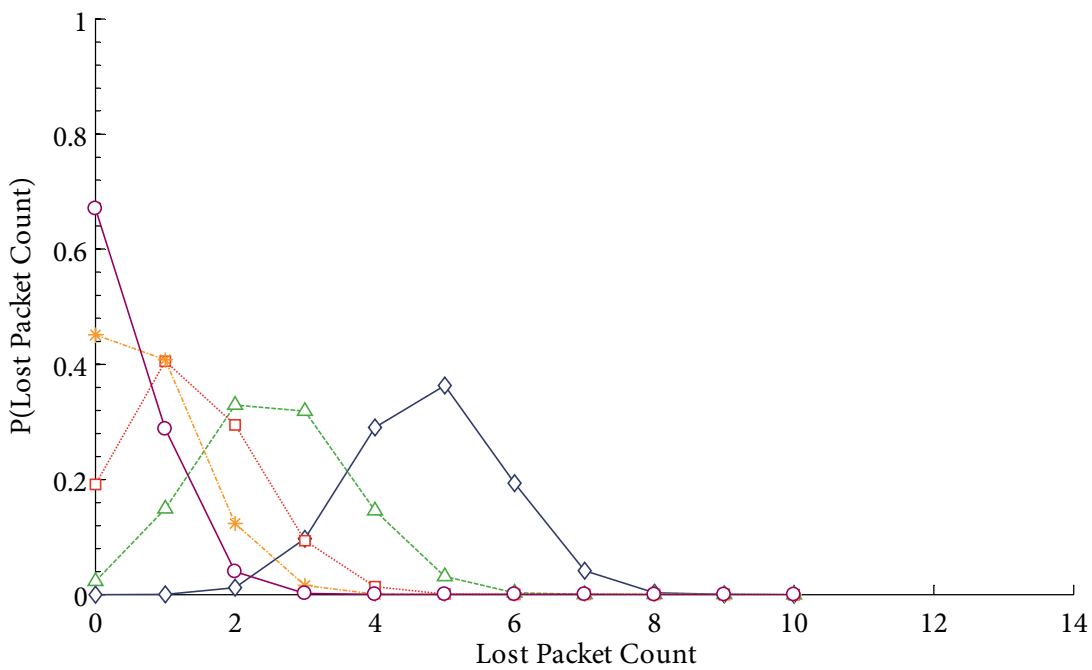

Figure 32: *Probability density functions for dispersity routing systems using, from right to left, 2 – 6 paths to communicate 17 packets. The packet loss rate is fixed to 9 packets for each path. Each additional path pushes the probability density function further to the left.*

depicts the probability density functions for 5 dispersity routing systems communicating a stream of 17 packets. The systems employ, from right to left, 2 – 6 paths, with the second curve from the right in figure 31 corresponding to the rightmost curve in figure 32. Each path is fixed at losing 9 packets. Additional paths decrease the deliverable packet loss rate, each pushing the probability density function further to the left. Also visible are the diminishing returns of additional paths. The gain of going from 2 to 3 paths, for example, is greater than that of going from 3 to 4 paths.

4.3.3 LIMITATIONS OF THE QCE-MODEL

The magnitude of the discrepancies depicted in figure 27 between the modelled and simulated MOS estimates may be explained by the combination of a number of factors. First, the clustering of the actual packet loss computed by the simulations around the expected packet loss computed by the QCE-model, as described in section 4.3.2, causes a discrepancy. This discrepancy has a slightly larger spread in the midrange than at low or high packet loss probabilities.

Next, this combines with the significantly larger number of combinations being simulated in this experiment that are in the midrange rather than in the low and high ranges of packet loss. Because every possible combination is being explored once, the packet loss distribution over these combinations on each path alone may be described by a binomial distribution. These paths are then combined exponentially, resulting in a considerable number of combinations in the midrange, which is also where the discrepancies have that slightly larger spread than at the low or high packet loss probabilities.

Then the packet loss probability is mapped to a MOS estimate using the E-model. However, a packet loss probability discrepancy in the midrange results in a larger discrepancy in the resulting MOS estimate than an equal packet loss probability discrepancy in the low range does. Consider a packet loss probability discrepancy of 0.1 between packet loss probabilities 0.3 and 0.4. Referring to figure 15, for 2 paths this packet loss probability discrepancy results in a MOS estimate between 2.45 and 3.34, a range of 0.89. For packet loss probabilities in the low range the converse is true, that is, the same packet loss probability discrepancy at the low range, such as between packet loss probabilities 0 and 0.1, results in a

much smaller MOS discrepancy. In figure 12, such a packet loss probability discrepancy results in a MOS estimate between 4.33 and 4.41, a range of 0.08, less than a tenth of the discrepancy of 0.89 in the first example above.

That these factors combine in this manner is, thus, a limitation of the QCE-model. Put simply, quality estimates by the QCE-model are more accurate for lower loss rates than for the midrange, with the parameter limitations of the E-model precluding quality estimates for high loss rates. While the QCE-model does offer the quality estimate that is the most likely to be delivered, in the midrange the discrepancy between the estimated and actually achieved quality may be much greater than for low packet loss probabilities. This greatly reduces the accuracy of the QCE-model in that midrange.

However, first, this limitation is readily quantifiable. Indeed, using equations (20) – (23) a defined quality goal, such as achieving a particular MOS with a given confidence using paths of known packet loss characteristics, may be sought.

Secondly, in practice the packet loss probabilities of (fixed line) paths are in the low range. For the 6265 VOIP telephone calls measured (see section 2.1) over 6 calendar months, the arithmetic mean packet loss probability is 1.328E−03. Only 4 VOIP telephone calls measured (that is, 0.064%) were observed with a packet loss probability exceeding 1.000E−01. Therefore, the QCE-model limitation described above may not be that much of a problem in practice. Indeed, in section 4.2, call profiles where synthesised with condensed quality threats to quantify dispersity routing performance and model accuracy in conditions much worse than observed. The mean packet loss probability in these synthesised call profiles was elevated to 1.177E−01. Even at this artificially elevated rate, for just 2 paths the discrepancies due to the QCE-model limitation is not significant, with additional paths reducing the significance of the discrepancies even further.

Finally, approaches for mitigating the QCE-model limitation exist as those already outlined in section 3.2.3 for meeting particular quality goals. These approaches, presented in section 3.2.3 for the purposes of improving deliverable quality, transform the characteristics delivered by the dispersity routing system such that when mapped to a MOS estimate using the E-model, the discrepancies are minimised.

An example that illustrates this is the packet loss probability discrepancy of 0.1 between packet loss probabilities 0.3 and 0.4 described earlier, and which as per figure 15 results in a MOS estimate in the range of 0.89 (that is, between 2.45 and 3.34 MOS). Adopting the approach of adding a path, the resulting MOS estimate is of a smaller range — in addition to being of a higher value. For the sake of simplicity in this example, the path being added is assumed to be of the same packet loss and packet loss burstiness characteristics as the existing paths in the dispersity routing system. By adding such a path, the delivered MOS estimate is between 3.70 and 4.15 MOS, that is, in the range of 0.45. Not only has the range almost halved, but the expected value for the deliverable MOS estimate has increased, raising the probability that the particular quality goal is meet or even exceeded.

The same principles apply to the other approaches outlined in section 3.2.3, including adopting a combination of the approaches. By improving the deliverable quality, the characteristics delivered by the dispersity routing system are transformed so that when mapped using the E-model to a MOS estimate, discrepancies in the resulting MOS estimate are minimised.

Consequently, while the QCE-model does have a limitation, that limitation occurs in situations that tend not to occur, but when these situations do occur they may be managed. In this experiment, though, that limitation impacted on a significant number of scenarios, causing the results depicted in figure 27 to suggest that the model is inaccurate overall, which contradicts with the conclusions that may be drawn from the experiments described in sections 4.1 and 4.2. However, the inaccuracies are the result of the QCE-model limitation, and the nature of this experiment causes this limitation to assert itself a significant number of times for conditions that (1) are unlikely and (2) when they do occur may be managed.

4.3.4 EFFECTS OF E-MODEL LIMITATION

The E-model is applicable to well-defined parameter ranges only, such as packet loss probabilities in the range 0 to 0.2 [58]. Consequently, a MOS estimate cannot be computed for all scenarios in this experiment. A result of this limitation in combination with the discrepancies in the MOS estimates is a mean of 1.13E−01 in the distribution in figure 27. This suggests that in this experiment the QCE-model under-estimates the MOS. However, figure

28 and figure 29 show that the QCE-model itself does *not* over-estimate the packet loss probability or the burstiness of that packet loss which would result in a MOS under-estimate. Intuitively, the difference in nature between the (continuous) mathematical QCE-model and the (discrete) simulations scatters the discrepancies as described by equation (22), and the limitations of the E-model then truncate an upper portion of that distribution preventing it from balancing the arithmetic mean to 0, causing the simulation to appear under-estimated.

That the limited parameter ranges to which the E-model is applicable causes the mean of 1.13E−01 may be shown by applying equation (23) to a dispersity routing system of two paths communicating 17 packets. While this approach considers packet loss only, and largely ignores other path characteristics such as packet loss burstiness, it does show that the non-zero mean of 1.13E−01 can be explained by the E-model parameter limitations. Indeed, in doing so it shows the importance of considering packet loss burstiness when estimating quality, since that yields a more accurate estimate than when not considering it.

To show how the E-model parameter limitations cause the non-zero mean, equation (23) is applied using algorithm 1. This algorithm uses algorithm 2 to compute the discrepancies between modelled and simulated MOS estimate approximations for a dispersity routing system of N paths communicating a stream of n packets. The discrepancies and their frequencies are accumulated into set Δ. Both (1) MOS estimate approximations assuming independent loss (that is, random, or non-bursty, or independent loss) and (2) MOS estimate approximations assuming maximum burstiness (that is, packet loss occurring as a single burst) are computed. Consequently, element $\delta_x^y \in \Delta$ quantifies the frequency of MOS discrepancy $x \in [-4,4]$ assuming $y \in \{\text{bursty,random}\}$ loss occurring.

Modelled MOS estimate approximations are based on packet loss probabilities estimated simply as the product of the packet loss probabilities of the paths used by the dispersity

Algorithm 1: *Compute MOS discrepancies from packet loss discrepancies.*

```
function Δ ← ComputeMosDiscrepancies(n, N )
    Δ ← {δ_x^y : ∀y ∈ {bursty,random}, ∀x ∈ [−4,4], δ_x^y = 0}
    Δ ← ComputeMosDiscrepancies(n, N, ∅, Δ )
end function
```

routing system. This packet loss probability is adopted as an approximation of the packet loss probability computed by the QCE-model, and from which MOS estimates are then computed assuming both random and maximally bursty packet loss.

In contrast, simulated MOS estimate approximations are based on packet loss probabilities derived from the packet loss discrepancy distribution expressed by equation (23) for a stream of n packets. For any given scenario, this distribution quantifies the frequency of a given packet loss count occurring. Therefore, the simulated MOS estimate approximations are computed as MOS estimates for all possible packet loss counts in a stream of n packets assuming both random and maximally bursty packet loss. The discrepancies, then, are computed as the simulated MOS estimate approximation less the modelled MOS estimate approximation, with the computed frequency quantifying the occurrences of this discrepancy for that packet loss count for that scenario, for both random and maximally bursty packet loss.

Algorithm 2 traverses recursively through the combinations of lost packet counts for the N paths communicating n packets. For each combination of lost packet counts, algorithm 2 uses algorithm 3 to compute and accumulate into Δ the MOS discrepancies for that combination. Algorithm 3 computes MOS estimates as $\mathrm{MOS}(b, e)$ using the E-model [58] for (1) packet loss probability b and (2) packet loss burst probability e, assuming (3) the same codec as observed in the call centre measurements (see section 2.1) and (4) an absolute delay

Algorithm 2: *Traverse the set of lost packet count combinations for the N paths.*

```
1  function Δ ← ComputeMosDiscrepancies(n, N, 𝒞, Δ )
2      if |𝒞| < N
3          for i ∈ {0,1, … , n} do
4              Δ ← ComputeMosDiscrepancies(n, N, {c ∈ ℤ: c ∈ 𝒞 ∨ c = i}, Δ )
5          end for
6      else
7          Δ ← ComputeMosDiscrepancies(n, 𝒞, Δ )
8      end if
9  end function
```

not exceeding 100 ms. $\mathrm{MOS}(b, e)$ evaluates to ∞ when the E-model cannot compute a MOS estimate because the parameters are outside of their defined ranges [58]. Note that the frequency of k packet losses occurring in a scenario is computed as

$$\mathrm{f}(k; \mathcal{C}, n) = \begin{cases} \omega(k; \mathcal{C}, n), & \text{if } b_{\text{best}} \leq k \leq b_{\text{worst}} \\ 0, & \text{otherwise} \end{cases}. \tag{24}$$

Figure 33 depicts two cumulative distribution functions. The solid blue distribution shows the MOS discrepancies assuming maximum bursty loss, as the MOS discrepancies

Algorithm 3: *Compute MOS discrepancies for a particular scenario.*

```
1   function Δ ← ComputeMosDiscrepancies(n, 𝒞, Δ )
2       l ← ∏_{c∈𝒞} c/n
3       m^random ← mos(l, l)                    // modelled independent loss assumed
4       m^bursty ← mos(l, 1/n)                  // modelled maximum burstiness assumed
5       for k ∈ {0,1, … , n} do
6           f ← f(k; 𝒞, n)
7           if f > 0
8               s^random ← mos(k/n, k/n)        // simulated independent loss assumed
9               s^bursty ← mos(k/n, 1/n)        // simulated maximum burstiness assumed
10              if m^random ≠ ∞ ∧ s^random ≠ ∞
11                  x ← s^random − m^random
12                  δ_x^random ← δ_x^random + f, where δ_x^random ∈ Δ
13              end if
14              if m^bursty ≠ ∞ ∧ s^bursty ≠ ∞
15                  x ← s^bursty − m^bursty
16                  δ_x^bursty ← δ_x^bursty + f, where δ_x^bursty ∈ Δ
17              end if
18          end if
19      end for
20  end function
```

$\left\{x \in [-4,4]: \forall \delta_x^{\text{bursty}} \in \Delta, \delta_x^{\text{bursty}} > 0\right\}$ and their corresponding frequencies $\{\delta_x^{\text{bursty}} \in \Delta: \delta_x^{\text{bursty}} > 0\}$. In contrast, the dashed red distribution shows the discrepancies assuming random loss, as the MOS discrepancies $\left\{x \in [-4,4]: \forall \delta_x^{\text{random}} \in \Delta, \delta_x^{\text{random}} > 0\right\}$ and their corresponding frequencies $\{\delta_x^{\text{random}} \in \Delta: \delta_x^{\text{random}} > 0\}$. The vertical lines show the 50th percentiles of these distributions, where the right (solid blue) vertical line at 0.17 marks the 50th percentile of the MOS discrepancies assuming maximum bursty loss, and the left (dashed red) vertical line at 0.11 marks the 50th percentile of the MOS discrepancies assuming random loss. Also shown fitted to the MOS discrepancies distributions are Gaussian distributions. The dashed orange Gaussian distribution is the best fit to the distribution of MOS discrepancies assuming maximum bursty loss, whereas the dotted green distribution is the best fit to the distribution of MOS discrepancies assuming random loss.

Not only does the mean of 1.13E−01 observed in figure 27 fall between the 50th percentiles of the distributions of 0.11 and 0.17 in figure 33, showing that the E-model parameter

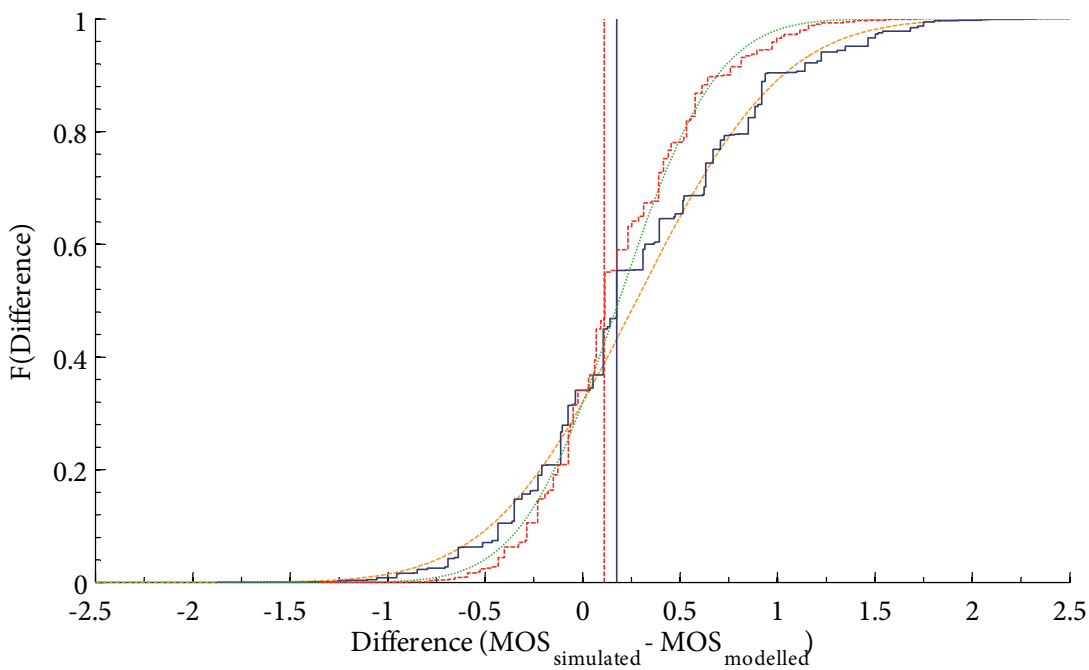

Figure 33: *Cumulative distribution functions of MOS discrepancies for all combinations of packet loss. The solid blue distribution shows discrepancies assuming maximum bursty loss, while the dashed red distribution shows discrepancies assuming random loss.*

limitations account for the non-zero mean. The shape in figure 27 of the distribution in relation to the fitted Gaussian distribution also matches the shapes in figure 33 of the distributions in relation to their fitted Gaussian distributions. All are below their fitted Gaussian distribution up to approximately the 20th percentile, above from approximately the 20th to approximately the 75th percentile, and then again below from approximately the 75th percentile upwards. This corresponds to the shape of the MOS curve (see figure 6), which is flatter in the first 20th and last 25th percentiles (approximately) than in the middle (between the 20th and 75th percentile, approximately).

4.4 SUMMARY

In this chapter, three experiments establish the accuracy of the QCE-model collectively. A detailed analysis shows that the QCE-model is in good agreement with the simulations in the first two of the experiments. In the first experiment, which adopts conditions measured over six calendar months in a commercial call centre, 98% of the differences in E-model MOS estimates computed with the QCE-model and by simulation are within 1.40E−02. Such a discrepancy would be difficult for a human to discern.

Agreement between the QCE-model and the simulations, while still good, is not quite as good in the second experiment, which adopts conditions synthesised to be much worse than those observed in the call centre by condensing the observed threats to quality. In this experiment, 98% of the differences in E-model MOS estimates computed with the QCE-model and by simulation are within 5.90E−02, which is greater than in the first experiment, but which would still be difficult for a human to discern.

In the third experiment, which tests the accuracy of the QCE-model for a dispersity routing system of two paths adopting every possible combination of packet loss in a stream of 17 packets, three factors combine to cause notable discrepancies between the E-model MOS estimates computed with the QCE-model and by simulation. That this combination occurs is a limitation of the QCE-model. In this third experiment, 98% of the differences between the E-model MOS estimates computed with the QCE-model and by simulation are within 1.16. However, detailed analysis shows that these notable discrepancies occur under conditions that are unlikely to occur in reality. Indeed, the probability of these conditions occurring may

be reduced with dispersity routing, thus reducing the likelihood that these notable discrepancies, which are readily quantifiable, occur in the first place. Collectively the experiments show that the QCE-model is useful for estimating the quality that may be expected from a given dispersity routing system with known path characteristics.

Chapter 5 • Deploying Dispersity Routing to Improve Quality

This chapter outlines how dispersity routing may be deployed to improve deliverable quality of a real-time service such as VOIP. Other, more complex and technologically more challenging systems are of course possible, however this chapter not only purposes to show that dispersity routing may be used to improve the deliverable quality of a real-time service, but also that a system devised to pursue that goal may be relatively simple. Another objective of this chapter is to convey the design elements of such as system in the context of a feasible example. This provides a context for activities such as intelligently choosing paths (see section 2.3) and selecting an appropriate value for Q (see section 3.1).

5.1 DEPLOYMENT MOTIVATIONS

Dispersity routing may be used to improve the quality of real-time services such as VOIP on the Internet. However, the following is required to make it possible for dispersity routing to do so. First, paths are needed that are as uncorrelated as possible in their failure and delay variation behaviour but that have similar delay characteristics. Second, when any level of service guarantee is sought, the paths need to be chosen appropriately, and they need to be managed on an ongoing basis to ensure that they continue to be appropriate. Third, the dispersity routing service needs to be delivered in a form that can be used, such as a router that implements dispersity routing.

Providing these as a value-added service may provide a VOIP telephony service provider with an advantage over other VOIP telephony service providers that do not provide such a service. As shown in this thesis, dispersity routing can improve the quality of a VOIP service to be more on par with traditional telephony. Such a service may be attractive to consumers of VOIP services, who would be getting a VOIP service with quality that is more on par with traditional telephony than VOIP is currently, but at the possibly much lower cost of dispersity routed VOIP telephony. Especially when service level guarantees may be offered, such as the quality estimates offered by the QCE-model described in section 3.1 as shown in section 3.2 and with the accuracy as shown in section 4.3.2.

5.2 DELIVERING DISPERSITY ROUTING

Dispersity routing may be made available for use by real-time services in various ways. For example, it may be provided by dedicated software that presents a tunnel interface to a dispersity routing protocol, and which is then deployed on the points in the network between which dispersity routing is to occur. Indeed, in section 2.3 dispersity routing is described in terms of a system connecting two points on the Internet.

Another approach is to incorporate a dispersity routing protocol directly in routers, such as the subsystems of the various operating systems and the dedicated hardware devices that provide routing services [84]–[85]. Clearly, integrating a dispersity routing protocol into routers has a number of advantages, such as performance and interoperability with other services like *Network Address Translation* (NAT) and *Application Level Gateway* (ALG) [86]–[87], *Security Architecture for IP* (IPsec) [88], and SIP proxy servers [89]. However, in the interests of simplicity and clarity, the next section assumes that dispersity routing is provided by a dedicated software subsystem that presents a tunnel interface to a dispersity routing protocol.

5.3 CONNECTING TWO SUBNETWORKS WITH DISPERSITY ROUTING

In this section, two subnetworks are connected using dispersity routing. These subnetworks are assumed to be (1) in different geographical locations, (2) connected through the Internet by multiple diverse paths with uncorrelated failure behaviours, and (3) located in the same (possibly private) network address space. Subnetworks may be connected through the Internet by multiple diverse paths with uncorrelated failure behaviours by, for example, using diverse service providers that collectively provide diverse paths. A dedicated software subsystem that implements dispersity routing connects any two points using dispersity routing by providing a tunnel interface to that point. This is similar to the approaches taken by [23] and [36], who also provide tunnel interfaces to their systems that implement forms of partially-redundant and non-redundant dispersity routing systems respectively.

Figure 34 illustrates an example architecture that connects two subnetworks at two points (that is, tun_{s1} and tun_{d1}) using dispersity routing. The two points, tun_{s1} and tun_{d1}, are interfaces

to *virtual network devices* (as found in operating systems such as Linux and FreeBSD) that (1) write data to a user space process instead of a network device, and (2) read data from a user space process instead of a network device. Note that in this section, performance is not a concern. Rather, the architecture depicted in figure 34 and adopted in this section primarily serves to illustrate how dispersity routing may be achieved.

Any data routed to the tun_{s1} interface in the left of figure 34 is read by the user space process in the left of the figure labelled Dispersity Routing, referred to as *Dispersity Router* for the remainder of this chapter for the sake of brevity. This process is configured to use N destinations to dispersity route any packet that it reads from tun_{s1}, where, in the example adopted in this section, these N destinations are configured by IP routing tables to be routed to network interface eth_{s1}, eth_{s2}, …, eth_{sN}. Each packet is encapsulated [14] with a header (see section 2.3) that includes a 32-bit sequence number similar to the 16-bit sequence numbers found in RTP packets [69]. That is, the first packet is assigned a uniformly distributed random number in the integer interval $[0 .. 2^{32} - 1]$, with each successive packet being

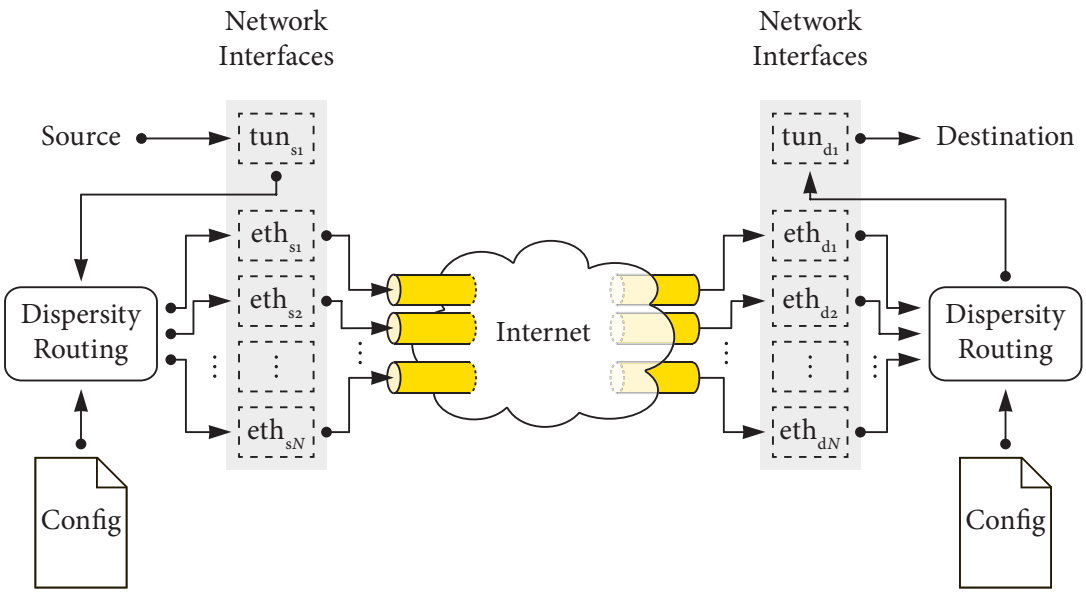

Figure 34: *Overview of elements collaborating in connecting two points using dispersity routing. In this example, packets written to the virtual network device* tun_{s1} *(on the top left) are sent over multiple paths by a user space dispersity routing process. The receiving dispersity routing process de-disperses the packets and writes them to virtual network device* tun_{d1} *where they may be read.*

assigned a sequence number that is an increment of exactly 1 (one), modulo 2^{32}, of the last sequence number assigned. Also included in the header is a 32-bit timestamp similar to the 32-bit timestamps found in RTP packets [69], and which is used to schedule delivery of packets from the de-dispersion buffer when a delay $Q > 0$ is adopted by the dispersity routing system for the de-dispersion buffer. The encapsulated packet is then sent to the N destinations using the protocol (such as UDP and TCP) appropriate for the data being communicated and configured for that destination, and which are then routed to the network interfaces eth_{s1}, eth_{s2}, …, eth_{sN}.

At the receiving end, the Dispersity Router in the right of figure 34 is similarly configured to use N destinations, where these N destinations are configured by IP routing tables to be routed to network interface eth_{d1}, eth_{d2}, …, eth_{dN}. Any packet that arrives from any of these destinations is read, decapsulated, and the sequence number found in the header is used to write to the tun_{d1} interface the first instance that arrives of each packet, with any remaining instances of that packet that arrive being discarded. Packets written to tun_{d1} may then be read by anyone reading packets from a destination configured by IP routing tables to be routed to tun_{d1}.

The architecture depicted in figure 34 may be symmetric, in that any packets routed to tun_{d1} may be read by the Dispersity Router in the right of the figure, which behaves in the same fashion as the left Dispersity Router. That is, any packets routed to the tun_{d1} interface (in the right of the figure) are communicated to the tun_{s1} interface (in the left of the figure) by sending them to the N destinations that the right Dispersity Router has been configured to use. Note that the two directions (that is, left to right, and right to left) need not adopt the same value for N, although in this example they are.

Connecting two subnetworks in the manner described above, where the two subnetworks share the same network address space, as is the case in this section, enables communication between these two subnetworks to be dispersity routed. In addition, by presenting as a tunnel interface, any communication may be dispersity routed, including real-time communications such as VOIP.

The scenario considered in this section is a simple one. Since the two subnetworks exist in the same, although possibly private, address space, the complexities introduced by address

translation (as exemplified by NAT and SIP ALGs) are not a concern in this section. Security considerations are not examined, and neither is the connecting of more than two subnetworks or the IP fragmentation that may be caused by the addition of the header to facilitate dispersity routing. Nevertheless, the scenario exemplifies one way in which elements may be brought together to provide dispersity routing.

5.4 SECURITY CONSIDERATIONS

The architecture described in section 5.3 does not explore security considerations. As described, packets are simply encapsulated with a header that includes a sequence number, and multiple copies of that packet are then sent over multiple paths. That is, instead of sending the data along a single path which may be vulnerable, it is instead sent along multiple paths which may be vulnerable. This increases the exposure of the data to vulnerabilities, since it now exists in more places that may be vulnerable than it would have existed had it just been sent along a single path using ordinary, non-dispersity, routing. In contrast, non-redundant dispersity routing (see section 1.1.1) only sends a part of the message along each path, that is, no one path is given the entire message, which may be considered as increasing security [90]–[91].

Since dispersity routing is provided in section 5.3 by componentry that is a part of the networking subsystem, approaches that have been devised elsewhere may be helpful here in alleviating security concerns. For example, it may be possible to apply IPsec [92] or tunnelling mechanisms such as VPNs [93] to any of the network interfaces (such as eth_{s1}, eth_{s2}, …, eth_{sN}) in figure 34 that have not already been secured and are thus vulnerable.

Different paths may be secured with different mechanisms, subject to these mechanisms meeting the needs of the communications that these paths facilitate. Real-time communications such as VOIP for example, have time constraints. Clearly, a mechanism that delays the data beyond its time constraints is of no use. Similarly, the mechanism must accommodate the data to be delivered. When delivering VOIP comprising SIP, RTP and RTCP using UDP, a security mechanism such as the tunnelling mechanism provided by the Secure Shell (SSH) [94] that only transports TCP may not be appropriate. This may be true even when the UDP may be encapsulated as TCP for delivery through the SSH tunnel,

because TCP may not be appropriate for the real-time communication being dispersity routed [95]. For VOIP, the security mechanism chosen is likely to be confronted with UDP packets that are lost, duplicated or arrive out of order. A security mechanism is appropriate only when it is able to tolerate these happenings in a manner that satisfies requirements.

5.5 ADJUSTING DISPERSITY ROUTING SESSION PARAMETERS

A dispersity routing system of the architecture described in section 5.3 may also be constructed for a single application, for the lifetime of that application. This contrasts with the dispersity routing system described in section 5.3 which is constructed to connect two subnetworks for any communication between those two subnetworks that is routed over the dispersity routing system. For example, to improve the deliverable quality of a VOIP telephone call, a dispersity routing system may be constructed specifically for that call, and destroyed when the call ends. Doing this has the advantage that the dispersity routing system may be customised to meet the specific requirements of that one VOIP telephone call, because the system exists specifically for the VOIP telephone call.

Two ways in which the dispersity routing system used to improve the deliverable quality of a VOIP telephone call may be customised is in setting the numbers of paths used (that is, N; see section 1.1.1) and the delay adopted by the de-dispersion buffer (that is, Q; see section 3.1). Both of these parameters have the capacity to impact on the quality delivered by the dispersity routing system. For VOIP, increasing N affects the perceived quality (as depicted in figure 15) by employing additional paths, and increasing Q increases delay in exchange for decreasing delay variation and the number of packets arriving out of order.

The simplest way for a person participating in a 2-way VOIP telephone call to set these parameters is to express them directly as a change in N and Q. For example, when using SIP, by adding an attribute, such as "`a=dispersity:N Q`", into the *Session Description Protocol* (SDP) that describes the VOIP telephone call, where `N` is the number of paths requested, and `Q` is the delay adopted by the de-dispersion buffer. The SDP would need to be parsed by a SIP ALG to receive the setting of these parameters.

A change in Q by one person affects the delay adopted by the de-dispersion buffer for that person, that is, the de-dispersion buffer local to the person. This de-dispersion buffer de-

disperses the packets delivering the media consumed by that person. Besides altering the delay experienced by the media originating from the other person, it also decreases the delay variation in the media, which may affect any de-jitter buffer that processes media for the person that changed Q [32]. Consequently, any impact on such a de-jitter buffer may need to be considered when choosing a value for Q.

In contrast, changing N affects the number of paths used to send packets to the other person, and so affects the quality perceived by the other person. In order for a person to affect the quality perceived by them by altering N, the (remote) dispersity router that sends the packets containing media to the person that has changed N must be instructed to implement the altering of N.

However, it may be difficult for a person to decide on and set the number of paths directly to experience a desired quality. Rather, a person may simply wish to express quality expectations. For example, when initiating an important call, the person may wish to state that the call is expected to be at least of a very satisfying quality with some probability C. Using (1) the characteristics of available paths as observed by the person thus far, (2) the QCE-model, and (3) a quantification of QCE-model accuracy for those characteristics as described in section 4.3.2, it may be possible to (1) choose a number of paths from the set of available paths to satisfy the quality expectation of the person with probability C. When the stated quality expectation with probability C is unlikely to be satisfied, the person may be informed *before* placing the VOIP telephone call that currently it is unlikely the stated quality expectation may be met with probability C. They may then make an informed decision as to whether (1) the VOIP telephone call cannot be made, or (2) they lower their quality expectation and proceed with the call.

5.6 GETTING AND MANAGING PATHS FOR DISPERSITY ROUTING

Dispersity routing requires paths that ideally (1) have comparable delays and (2) are uncorrelated [96]–[99] in their packet loss and delay variation characteristics. Once paths that satisfy these criteria have been chosen, they may need to be monitored and — should they no longer meet these criteria — replaced. Chosen paths may also be replaced when more appropriate paths become available.

While it may not be possible to define the exact path taken by packets being communicated between two points over the public Internet, it may suffice to exploit the observation that paths that diverge earlier tend to converge later than paths that diverge later [54]–[55][100]. That is, routing packets to a particular point may suffice in suggesting the path that the packets are to traverse. To dispersity route a packet over multiple diverse paths, it may suffice to route a copy of that packet to multiple points, each point suggesting a diverse path. Another approach would be to use overlay networks [4][35][52].

Though getting N diverse paths is likely to be increasingly difficult as N increases, as shown in sections 2.5 and 3.2 the largest gain in deliverable quality happens when changing from no dispersity routing to dispersity routing with just two paths. However, while the gains in quality earned by that first additional path are significant (see section 2.5.1), each additional path returns increasingly smaller gains in quality. Therefore, not only may it be unnecessary in practice to use a large number of paths in a dispersity routing system because a small number of paths already contribute significantly to improving quality, doing so also gains comparatively very little.

The suitability of paths in meeting set quality goals may be assessed by measuring their packet loss and packet loss burstiness characteristics and estimating the deliverable quality with the QCE-model described in section 3.1 as shown in section 3.2. Besides the uncorrelated packet loss and delay variation characteristics, other considerations may impact on the selection of paths. For example, the physical proximity of the paths to each other may be a concern [101]. Two paths that physically are close to another may be vulnerable to the same geospatial events, such as a fire occurring in a tunnel that the two paths may traverse [70]. Likewise, the paths that pass through a particular region may all be affected when that region is subjected to fire — or any number of other natural phenomena. Similarly, business considerations may be of concern when selecting paths, such as financial or organisational arrangements. It may not be appropriate to source all paths from the same service provider in order to select paths that are as uncorrelated in their packet loss and delay characteristics as possible. Avoiding these correlations may be difficult, for example when different service providers sublease services from the same third-party service provider. While the paths may present as diverse, they may not actually be diverse.

Clearly, the selection process of diverse paths must be driven by the kinds of failures to be tolerated. For example, if the system is to continue delivering a service despite a given geospatial event occurring, the paths must be chosen accordingly. That is, the paths must be chosen such that not all paths are affected by the geospatial events to be tolerated.

In the contrived example depicted in figure 35, three paths diverge at node 3, and do not converge again until node 21. In contrast, the two paths that diverge at node 7 converge again at node 19. The general observation then that earlier diverging paths tend to converge later than later diverging paths [54]–[55][100] may be useful in the search for diverse paths. It also illustrates that diverse paths may not need to be completely diverse. For example, paths with a shared beginning and ending path segment but a diverse middle segment may suffice when the shared segments offer high-quality communications that satisfy requirements. This might be the case when these shared segments are provisioned appropriately and reserved for the sole use of the real-time communications being diversity routed to improve its quality, which may be possible when these segments are not in the public Internet.

Therefore, in this thesis the selection or construction of paths is assumed to be a manual process. This manual process calls for one or more persons to select paths intelligently after

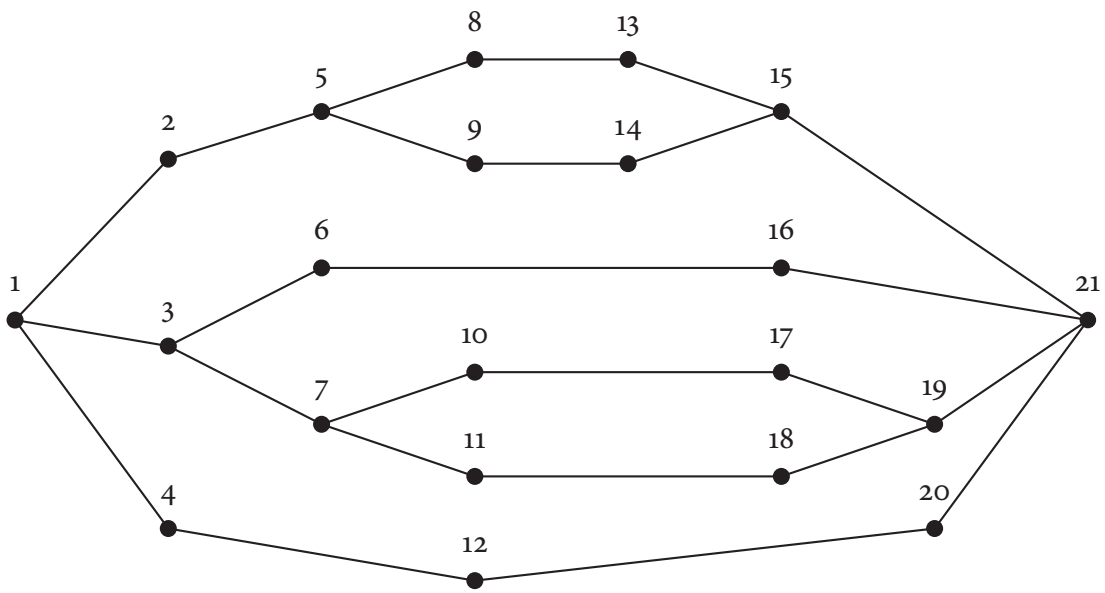

Figure 35: *The paths connecting two nodes may include shared segments. In the network that connects nodes 1 and 21, three paths share the segment from node 1 to node 3. Similarly, two paths share the segment from node 19 to node 21.*

making considerations such as those above. Selecting paths manually may be sufficient when the points being connected with dispersity routing are fairly static, as is the case when connecting two subnetworks such as in section 5.3 or a home user to a service provider, none of which typically relocate frequently.

5.7 SUMMARY

In summary, dispersity routing may be delivered in many different ways. Section 5.3 presents architecture for connecting two subnetworks, as might be the case when connecting two branch offices of an organisation. Another example would be to provide dispersity routing to a home user as *customer premises equipment* (CPE) in the form of a single household appliance that provides dispersity routing. Delivering dispersity routing with a single CPE has a number of advantages, such as performance and increased integration.

Alternatively, access to a service could be provided, where the VOIP telephony device connects through a point very close to the device, but then uses dispersity routing to connect to the rest of the network. A combination of these may also be possible.

Chapter 6 • Conclusion

This thesis proposes the use of fully redundant dispersity routing (referred to simply as *dispersity routing* in this thesis for brevity) to improve the quality and reliability of real-time services, focusing in particular on VOIP. Besides illustrating how dispersity routing improves the quality and reliability of real-time services on the Internet, this thesis shows how effective dispersity routing is in achieving this goal using mathematical tools developed in this thesis.

First, simulations that adopt measurements of real VOIP traffic taken in a commercial call centre for this thesis quantify the quality that dispersity routing would have delivered for those observed conditions, and which may be compared to the quality achievable without dispersity routing. Second, a mathematical model called the *Qualitative Characteristics Estimation Model* (QCE-model) is developed in this thesis for estimating the most likely quality deliverable by a given dispersity routing system, even for conditions that have not yet been observed. This model is applied to a mathematical model called the *Packet Loss and Packet Loss Burstiness Model* (PLB-model) constructed in this thesis from the measurements to quantify the quality deliverable by dispersity routing in general. Third, the accuracy of the QCE-model in estimating the quality that dispersity routing is most likely to deliver is established. This chapter concludes this thesis by summarising the research, itemising the main research contributions and discussing further research.

6.1 SUMMARY

In the first stage of the research, real VOIP traffic in a commercial call centre, an environment maintained in optimum condition by professional staff, was measured as described in chapter 2. The measurements both (1) establish a baseline of current (that is, non-dispersity routed) VOIP performance, and (2) are a source of realistic path characteristics. Chapter 2 then shows how dispersity routing may be used for real-time communications in a packet-switched network, and details how such a system is simulated in this thesis. A single simulation using measurements selected specifically for the task then details how dispersity routing improves quality by lowering delay, delay variation, packet loss

and packet loss burstiness. Next, three sets of simulations show how VOIP quality may be improved with dispersity routing. It is found that for the characteristics measured in the call centre, the proportion of VOIP telephone calls that may be interpreted to be of a quality with which humans would be less than very satisfied improves significantly. That is, from slightly more than 1 in 23 for non-dispersity routed VOIP telephone calls, to — with just two paths — slightly less than 1 in 10 000 for dispersity routed VOIP telephone calls.

Chapter 3 then develops and presents the QCE-model for estimating the delay, packet loss and packet loss burstiness characteristics that a given dispersity routing system is most likely to deliver. From these characteristics, and with knowledge of the manner in which the media being communicated is encoded, a quality estimate for VOIP may then be computed using the E-model. Chapter 3 also constructs and presents the PLB-model of path characteristics from the measured VOIP traffic characteristics.

The QCE-model is applied to the PLB-model for dispersity routing systems of 2 – 6 paths, and together with (1) quality estimates for a non-dispersity routed system derived from the PLB-model using the E-model and (2) an analysis of the results, achieves two objectives. First, the quality possible with dispersity routing under various conditions is illustrated. Second, the relationships between packet loss, packet loss burstiness, numbers of paths and deliverable quality are demonstrated. Collectively, they show how dispersity routing may improve quality under what conditions, representing a planning tool that may be used in the pursuit of VOIP telephony quality goals.

Chapter 4 quantifies the accuracy of the QCE-model by comparing the quality estimates obtained by simulation to quality estimates obtained with the QCE-model for three sets of simulations. The first set is for a dispersity routing system of two paths adopting every combination of the measured VOIP telephone calls. Simulated and modelled quality estimates are found to be in good agreement. However, an increased probability of packet loss at the beginning of the measured VOIP telephone calls is found to increase discrepancies, and suggests that additional paths may be needed at the beginning of a VOIP call to counter the increased probability of packet loss during this portion of the call.

In the second set, the accuracy for a dispersity routing system of two paths is quantified in 25 sets of simulations adopting path characteristics with elevated threats to quality, but with

the increased probability of packet loss at the beginning of the VOIP telephone call removed. The QCE-model estimates are found to be in fairly good agreement with the simulation estimates, even for these conditions of artificially increased quality threats. Besides quantifying accuracy, this simulation also illustrates that dispersity routing is effective at increasing quality in worse conditions than those observed.

The third set quantifies the accuracy of the QCE-model for a dispersity routing system of two paths adopting every possible combination of a stream of 17 packets. Unlike the other two sets, the discrepancies in this set show that the QCE-model has limitations. The discrepancies are quantified as a probability distribution, and the limitations of the QCE-model and the impact of the limitations of the E-model are analysed.

The final part of this thesis details considerations for deploying dispersity routing. Discussions of motivations and mechanisms for deploying dispersity routing are followed in chapter 5 by a presentation of an example that connects two points in the Internet using dispersity routing. Security considerations show how dispersity routing may fit into a network infrastructure, and suggestions for configuring dispersity routing parameters for VOIP are offered. The chapter concludes with a discussion of how paths are obtained and maintained for dispersity routing, and the forms that dispersity routing deployment may take.

6.2 RESEARCH CONTRIBUTIONS

The research conducted for this thesis makes a number of contributions. This section summarises the main contributions.

1 Real VOIP traffic in a commercial call centre, where professional staff maintains the environment in an optimum condition, was measured. These measurements contribute in two ways. First, they establish a baseline of the current performance of VOIP in a real environment. Second, the measurements may be used as a source of path characteristics in simulations and in the PLB-model. The measurements show that in the call centre, slightly more than 1 in every 23 of the VOIP telephone calls measured is likely to be perceived to be of a quality with which humans would be less than very satisfied according to established quality interpretations (see table 3).

2 Simulations of dispersity routing systems that adopt the measurements as path characteristics illustrate how dispersity routing improves quality, by enabling comparison of the quality deliverable with dispersity routing to the quality deliverable without dispersity routing. It is shown that, while increasing numbers of paths yields decreasing gains in quality, with just two paths adopting the measurements taken in the call centre, dispersity routing may increase deliverable quality significantly. With two paths adopting the measurements, the proportion of VOIP telephone calls likely to be perceived to be of a quality with which humans would be less than very satisfied decreases from slightly more than 1 in every 23 to slightly less than 1 in every 10 000 VOIP telephone calls.

3 A mathematical model called the *Qualitative Characteristics Estimation Model* (QCE-model) is developed that estimates the most likely delay, and packet loss and packet loss burstiness characteristics delivered by a dispersity routing system with known path characteristics. Along with knowledge of the codec with which the media being communicated is encoded, a VOIP quality estimate may be computed from these estimates with the E-model as a MOS estimate. The QCE-model considers both packet loss and packet loss burstiness. As shown (in figure 15 and figure 33 for example), quality estimates assuming random packet loss may vary significantly to quality estimates assuming bursty packet loss. Since packet loss is bursty, as confirmed by the measurements taken in the call centre, the estimates by the QCE-model may be expected to be more accurate than estimates using a model that does not consider packet loss burstiness.

4 A *Packet Loss and Packet Loss Burstiness Model* (PLB-model) is constructed from the measured packet loss and packet loss burstiness characteristics. Using this model, the QCE-model may be applied to conditions that have not yet been observed, but that may be extrapolated and interpolated from the measured characteristics. Besides illustrating the relationships between numbers of paths, packet loss and packet loss burstiness probabilities, and the quality that may be expected, such an application of the QCE-model may be used as a planning tool.

5 The accuracy of the QCE-model is quantified using simulations that adopt the measurements, and discrepancies between the quality estimates computed by the QCE-model and the simulations are analysed and quantified. For example, for simulations adopting the measurements as observed, 98% of the MOS estimates are within 1.40E−02, a discrepancy so small as to be all but indiscernible to humans. However, the discrepancies for 98% increases to 5.90E−02 for simulations adopting the same measurements but with quality threats condensed to represent worse conditions than those observed. Finally, for a dispersity routing system of two paths adopting all possible combinations for a small stream of 17 packets, 98% of the MOS estimates are within 1.16, a discrepancy that is *not* beyond human discern. This is due to a limitation of the QCE-model, which causes estimates to be most accurate at zero packet loss and complete packet loss, and least accurate at the deliverable midpoint between these two extremes. However, (1) the probability distribution of the discrepancies are readily quantifiable, (2) in reality, packet loss is not around that midpoint (for example, only 0.064% of measured VOIP telephone calls had a packet loss rate exceeding 10%), and (3) accuracy can be increased using the same mechanisms available for increasing quality.

6 Considerations for deploying dispersity routing to improve the quality of real-time services, and in particular VOIP, are explored, beginning with a discussion of the motivations for using dispersity routing. Deployment options for dispersity routing are outlined, and security considerations are investigated that show how dispersity routing may fit into a network infrastructure. An example is presented that connects two points in the Internet using dispersity routing, and suggestions for configuring dispersity routing parameters for VOIP are offered.

6.3 FURTHER RESEARCH

This section enumerates suggestions for further research that continue the research in this thesis.

1 This thesis uses fully redundant dispersity routing to improve the deliverable quality of real-time services, focusing in particular on VOIP. In addition to VOIP arguably

being the most popular real-time service currently, mature tools exist for quantifying VOIP quality, both subjectively and objectively. The research could be extended into the use of fully redundant dispersity routing to improve the quality of other real-time services such as video conferencing.

2 The QCE-model does not consider the effect of packet loss caused by excessive delay variation [102]. In the measurements taken, 9.98E−02% of received packets experience delay variation in excess of 40 milliseconds, and which may be lost if a de-jitter buffer considers such packets as arriving too late and discards them. However, dispersity routing also has a *competitive de-jittering effect* that occurs when packets with lower delay variation outcompete packets with higher delay variation (see section 2.4.2). The research could be extended to investigate the impact on deliverable quality of the effect of delay variation in combination with the effect of dispersity routing on delay variation.

3 In quantifying the accuracy of the QCE-model, the probability distribution of the discrepancies between simulated and modelled quality estimates is quantified using packet loss. The research could be extended to include packet loss burstiness as well as the effect of delay variation.

4 As shown in this thesis, fully redundant dispersity routing can improve the quality of real-time communications. However, combining this form of dispersity routing with non-redundant dispersity routing and path switching techniques may yield further improvements in quality and is, therefore, a worthwhile research topic. A suggestion for combining these approaches into a hybrid would be to use M paths, where $M > n$. Each packet is then communicated over n of these M paths, employing path switching approaches with the QCE-model to switch the n paths used to communicate a given packet among the M paths being used collectively. Besides distributing the load over M paths and potentially decreasing packet loss burstiness, this approach facilitates the use of additional paths (that is, varying n) during phases with increased packet loss probabilities, like at the beginning of using a particular path. The effectiveness of this hybrid approach in improving the deliverable quality should be investigated.

5 In this thesis, the first instance of each packet to arrive in the de-dispersion buffer is delivered, and any subsequent instances that arrive are discarded. However, when the de-dispersion buffer adopts a positive delay Q, and multiple instances of a packet arrive before the packet is scheduled for delivery, a possibility for exploiting this redundancy exists. For example, if the first instance of a packet arrives with a corrupted payload, rather than discarding it as lost, it may be possible to combine it with subsequent instances of that packet that also have corrupted payloads to recover the packet. Clearly, if a non-corrupted packet instance arrives before the packet is scheduled for delivery, recovery is not necessary and the non-corrupted instance may simply be delivered. However, if multiple corrupted instances of a packet arrive before the packet is scheduled for delivery, and they can be combined to recover the packet, then their combination allows delivery of a packet that would otherwise have been lost. It may be warranted to investigate the effectiveness and limitations of this approach in improving real-time communications.

6 The use of fully redundant dispersity routing is proposed in this thesis for improving the deliverable quality of real-time services, in particular VOIP. This routing technique uses the path diversity available in the Internet to actively replicate the data over multiple paths, allowing the effect of a failure on one path to be reduced or even masked completely by the other paths. Having shown that fully redundant dispersity routing is an effective approach for improving the quality of VOIP, a protocol that selects programmatically diverse paths with comparable delays should be developed. The kinds of diversity available (such as path, service provider, and geographical), and the means of expressing diversity in these terms, should be studied.

References

[1] N. F. Maxemchuk, "Dispersity routing in store-and-forward networks," Ph.D. dissertation, Univ. Pennsylvania, USA, 1975.

[2] N. F. Maxemchuk, "Dispersity routing," in *Proc. ICC '75*, San Francisco, USA, 1975, pp. 41.10–41.13.

[3] N. F. Maxemchuk, "Dispersity routing in high-speed networks," *Comput. Networks and ISDN Systems*, vol. 25, no. 6, pp. 645–661, 1993.

[4] N. F. Maxemchuk, "Dispersity routing on ATM networks," in *Proc. 12th Annu. Joint Conf. IEEE Computer and Communications Societies (IEEE INFOCOM '93)*, 1993, pp. 347–357.

[5] N. F. Maxemchuk and S. Lo, "Measurement and interpretation of voice traffic on the internet," in *Proc. 1997 IEEE Int. Conf. on Communications (ICC '97)*, Montreal, 1997, pp. 500–507.

[6] N. F. Maxemchuk, "Dispersity routing: past and present," in *Proc. 2007 IEEE Military Communications Conf. (MILCOM 2007)*, 2007, pp. 1–7.

[7] S. Bettermann and Y. Rong, "Effects of fully redundant dispersity routing on VOIP quality," in *Proc. 2011 IEEE Int. Workshop Technical Committee on Communications Quality and Reliability (CQR 2011)*, Naples, USA, 2011.

[8] S. Bettermann and Y. Rong, "Estimating the deliverable quality of a fully redundant dispersity routing system," in *Proc. 17th Asia-Pacific Conf. on Communications (APCC 2011)*, Kota Kinabalu, Malaysia, 2011.

[9] B. Boehm and R. Mobley, "Adaptive routing techniques for distributed communications systems," *IEEE Trans. Commun.*, vol. 17, pp. 340–349, Jun. 1969.

[10] G. L. Fultz and L. Kleinrock, "Adaptive routing techniques for store-and-forward computer-communication networks," in *Proc. IEEE Int. Conf. on Communications*, Montreal, Canada, 1971, pp. 39/1–39/8.

[11] E. Gustafsson and G. Karlsson, "A literature survey on traffic dispersion," *IEEE Network*, vol. 11, no. 2, pp. 28–36, 1997.

[12] A. Banerjea, "Simulation study of the capacity effects of dispersity routing for fault tolerant realtime channels," in *1996 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '96)*, Stanford, CA, USA, 1996, pp. 194–205.

[13] N. Gogate, C. Doo-Man, S. S. Panwar and W. Yao, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 777–792, Sep. 2002.

[14] R. Kokku, A. Bohra, S. Ganguly and A. Venkataramani, "A multipath background network architecture," in *Proc. 26th IEEE Int. Conf. on Computer Communications (INFOCOM 2007)*, Anchorage, AK, USA, 2007, pp. 1352–1360.

[15] B. Kao, H. Garcia-Molina and D. Barbará, "Aggressive transmissions of short messages over redundant paths," *IEEE Trans. Parallel Distrib. Syst.*, vol. 5, pp. 102–109, Jan. 1994.

[16] T. T. Lee and S. C. Liew, "Parallel communications for ATM network control and management," in *Proc. 1993 IEEE Global Telecommunications Conf. (GLOBECOM '93)*, Houston, TX, USA, 1993, pp. 442–446.

[17] J. Duncanson, "Inverse multiplexing," *IEEE Commun. Mag.*, vol. 32, pp. 34–41, 1994.

[18] H. Adiseshu, G. Parulkar and G. Varghese, "A reliable and scalable striping protocol," in *Proc. ACM SIGCOMM '96*, Palo Alto, California, USA, 1996, pp. 131–141.

[19] C. B. S. Traw and J. M. Smith, "Striping within the network subsystem," *IEEE Network*, vol. 9, pp. 22–32, Jul./Aug. 1995.

[20] M. Zhang, J. Lai, A. Krishnamurthy, L. Peterson and R. Wang, "A transport layer approach for improving end-to-end performance and robustness using redundant paths," in *Proc. 2004 USENIX Annu. Technical Conf.*, Boston, MA, USA, 2004, pp. 99–112.

[21] H. Sivakumar, S. Bailey and R. L. Grossman, "PSockets: the case for application-level network striping for data intensive applications using high speed wide area networks," in *Proc. ACM/IEEE 2000 Supercomputing Conf.*, Dallas, TX, USA, 2000.

[22] E. Vergetis, R. Guérin and S. Sarkar, "Realizing the benefits of user-level channel diversity," *ACM SIGCOMM Comput. Commun. Review*, vol. 35, no. 5, pp. 15–28, Oct. 2005.

[23] T. Kawamoto and K. Goto, "Design and evaluation of IP multipath transmission with feedback," in *Proc. 19th Int. Conf. on Systems Engineering (ICSENG '08)*, 2008, pp. 294–299.

[24] S. J. Lee and M. Gerla, "Split multipath routing with maximally disjoint paths in ad hoc networks," in *Proc. 2001 IEEE Int. Conf. on Communications (ICC 2001)*, Helsinki, Finland, 2001, pp. 3201–3205.

[25] H. Levy and H. Zlatokrilov, "The effect of packet dispersion on voice applications in IP networks," *IEEE/ACM Transactions on Networking*, vol. 14, pp. 277–288, 2006.

[26] Y. Cai, "A thin-layer protocol for utilizing multiple paths," in *Proc. 2009 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, St. Louis, USA, 2009, pp. 1049–1054.

[27] D. G. Andersen, A. C. Snoeren and H. Balakrishnan, "Best-path vs. multi-path overlay routing," in *Proc. 3rd ACM SIGCOMM Conf. on Internet Measurement (IMC '03)*, Miami Beach, FL, USA, 2003, pp. 91–100.

[28] S. Jain, M. Demmer, R. Patra and K. Fall, "Using redundancy to cope with failures in a delay tolerant network," *ACM SIGCOMM Comput. Commun. Review*, vol. 35, no. 4, pp. 109–120, Oct. 2005.

[29] H. Garcia-Molina, B. Kao and D. Barbará, "Aggressive transmissions over redundant paths," in *Proc. 11th Int. Conf. on Distributed Computing Systems*, 1991, pp. 198–207.

[30] P. Ramanathan and K. G. Shin, "A multiple copy approach for delivering messages under deadline constraints," in *Proc. 21st Int. Symp. on Fault-Tolerant Computing (FTCS-21)*, Montréal, Canada, 1991, pp. 300–307.

[31] L. Chung-Sheng and C. J. Georgiou, "Implementation and performance analysis of congestion-tolerant isochronous communication in ATM networks using diversified routing," in *Proc. 1994 IEEE Int. Conf. on Communications (SUPERCOMM/ICC '94)*, New Orleans, LA, USA, 1994, pp. 1341–1345.

[32] A. Markopoulou and D. R. Cheriton, "The case for redundant arrays of internet links (RAIL)," *Computing Research Repository (CoRR)*, vol. abs/cs/0701133, 2007.

[33] V. Bui, Z. Weiping, A. Botta and A. Pescapé, "A markovian approach to multipath data transfer in overlay networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, pp. 1398–1411, Oct. 2010.

[34] Y. J. Liang, E. G. Steinbach and B. Girod, "Multi-stream voice over IP using packet path diversity," in *Proc 2001 IEEE 4th Workshop on Multimedia Signal Processing*, Cannes, France, 2001, pp. 555–560.

[35] Y. J. Liang, E. G. Steinbach and B. Girod, "Real-time voice communication over the internet using packet path diversity," in *Proc. 9th ACM Int. Conf. on Multimedia*, 2001, pp. 431–440.

[36] L. Yi, Z. Yin, Q. Lili and S. Lam, "SmartTunnel: achieving reliability in the internet," in *Proc. 2007 IEEE 26th Int. Conf. on Computer Communications (INFOCOM 2007)*, 2007, pp. 830–838.

[37] J. Apostolopoulos, T. Wong, W.-t. Tan and S. Wee, "On multiple description streaming with content delivery networks," in *Proc. 21st Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM 2002)*, New York, NY, USA, 2002, pp. 1736–1745.

[38] J. G. Apostolopoulos, T. Wai-tian and S. J. Wee, "Performance of a multiple description streaming media content delivery network," in *Proc. IEEE 2002 Int. Conf. on Image Processing (ICIP 2002)*, Rochester, NY, USA, 2002, pp. II/189–II/192.

[39] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. Visual Communications and Image Processing (VCIP)*, San Jose, CA, USA, 2001, pp. 392–409.

[40] J. G. Apostolopoulos and S. J. Wee, "Unbalanced multiple description video communication using path diversity," in *Proc. 2001 IEEE Int. Conf. on Image Processing (ICIP 2001)*, Thessaloniki, Greece, 2001, pp. 966–969.

[41] J. G. Apostolopoulos and M. D. Trott, "Path diversity for enhanced media streaming," *IEEE Commun. Mag.*, vol. 42, pp. 80–87, Aug. 2004.

[42] M. Kurant, "Exploiting the path propagation time differences in multipath transmission with FEC," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1021–1031, May 2011.

[43] A. Banerjea, "On the use of dispersity routing for fault tolerant realtime channels," *European Transactions on Telecommunications*, vol. 8, pp. 393–407, 1997.

[44] L. Rizzo, "Effective erasure codes for reliable computer communication protocols," *ACM SIGCOMM Comput. Commun. Review*, vol. 27, pp. 24–36, Apr. 1997.

[45] M. Podolsky, C. Romer and S. McCanne, "Simulation of FEC-based error control for packet audio on the internet," in *Proc. 17th Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM 1998)*, 1998, pp. 505–515.

[46] E. Altman, C. Barakat and V. M. Ramos, "Queueing analysis of simple FEC schemes for IP telephony," in *Proc. 20th Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM 2001)*, 2001, pp. 796–804.

[47] J.-C. Bolot, S. Fosse-Parisis and D. Towsley, "Adaptive FEC-based error control for internet telephony," in *Proc. 18th Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM 1999)*, New York, USA, 1999, pp. 1453–1460.

[48] X. Yu, J. W. Modestino and X. Tian, "The accuracy of gilbert models in predicting packet-loss statistics for a single-multiplexer network model," in *Proc. 24th Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM 2005)*, 2005, pp. 2602–2612.

[49] I. Cidon, A. Khamisy and M. Sidi, "Analysis of packet loss processes in high-speed networks," *IEEE Trans. Inf. Theory*, vol. 39, pp. 98–108, 1993.

[50] E. Altman and A. Jean-Marie, "Loss probabilities for messages with redundant packets feeding a finite buffer," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 778–787, 1998.

[51] J. Wenyu and H. Schulzrinne, "Modeling of packet loss and delay and their effect on real-time multimedia service quality," in *Proc. 10th Int. Workshop Network and Operating System Support for Digital Audio and Video (NOSSDAV 2000)*, Chapel Hill, NC, USA, 2000.

[52] S. Tao, K. Xu, Y. Xu, T. Fei, L. Gao, R. Guérin, J. Kurose, D. Towsley and Z.-L. Zhang, "Exploring the performance benefits of end-to-end path switching," in *Proc. 12th IEEE Int. Conf. Network Protocols (ICNP 2004)*, Berlin, Germany, 2004, pp. 304–315.

[53] S. Tao, K. Xu, A. Estepa, T. Fei, L. Gao, R. Guérin, J. Kurose, D. Towsley and Z.-L. Zhang, "Improving VOIP quality through path switching," in *Proc. 24th Annu. Joint*

Conf. IEEE Computer and Communications Societies (INFOCOM 2005), Miami, FL, USA, 2005, pp. 2268–2278.

[54] S. Tao, "Improving the quality of real-time applications through path switching," Ph.D. dissertation, Univ. Pennsylvania, USA, 2005.

[55] T. Fei, S. Tao, L. Gao and R. Guérin, "How to select a good alternate path in large peer-to-peer systems?," in *Proc. 25th IEEE Int. Conf. Computer Communications (INFOCOM 2006)*, Barcelona, Spain, 2006, pp. 1–13.

[56] W. Cui, I. Stoica and R. H. Katz, "Backup path allocation based on a correlated link failure probability model in overlay networks," in *Proc. 10th IEEE Int. Conf. Network Protocols (ICNP 2002)*, Paris, France, 2002, pp. 236–245.

[57] C. Tang and P. K. McKinley, "Improving multipath reliability in topology-aware overlay networks," in *Proc. 25th IEEE Int. Conf. Distributed Computing Systems Workshops (ICDCS 2005 Workshops)*, Columbus, OH, USA, 2005, pp. 82–88.

[58] ITU-T, "The E–model, a computational model for use in transmission planning," Recommendation G.107, 2008.

[59] ITU-T, "Methods for subjective determination of transmission quality," Recommendation P.800, 1996.

[60] ITU-T, "Mean Opinion Score (MOS) terminology," Recommendation P.800.1, 2006.

[61] ITU-T, "Vocabulary for performance and quality of service," Recommendation P.10, 2006.

[62] ITU-T, "Definition of categories of speech transmission quality," Recommendation G.109, 1999.

[63] ITU-T, "Application of the E-model: a planning guide," Recommendation G.108, 1999.

[64] ITU-T, "Transmission impairments due to speech processing," Recommendation G.113, 2007.

[65] R. G. Cole and J. H. Rosenbluth, "Voice over IP performance monitoring," *ACM SIGCOMM Comput. Commun. Review*, vol. 31, pp. 9–24, Apr. 2001.

[66] ITU-T, "Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Recommendation P.862, 2001.

[67] ITU-T, "Perceptual objective listening quality assessment," Recommendation P.863, 2011.

[68] A. P. Markopoulou, F. A. Tobagi and M. J. Karam, "Assessing the quality of voice communications over internet backbones," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 747–760, 2003.

[69] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: a transport protocol for real-time applications," IETF RFC 3550, Jul. 2003.

[70] M. R. Carter, M. P. Howard, N. Owens, D. Register, J. Kennedy, K. Pecheux, *et al.*, "Effects of Catastrophic Events on Transportation System Management and Operations, Howard Street Tunnel Fire, Baltimore City, Maryland – July 18, 2001," Baltimore City, Maryland, USA, 2001.

[71] J. Sliwinski, A. Beben and P. Krawiec, "EmPath: tool to emulate packet transfer characteristics in IP network," in *Proc. 2nd Int. Workshop on Traffic Monitoring and Analysis (TMA '10)*, Zürich, Switzerland, 2010.

[72] E. N. Gilbert, "Capacity of a burst-noise channel," *The Bell System Technical Journal*, vol. 39, pp. 1253–1265, Sep. 1960.

[73] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *The Bell System Technical Journal*, vol. 42, pp. 1977–1997, Sep. 1963.

[74] A. Mukherjee, "On the dynamics and significance of low frequency components of internet load," Univ. Pennsylvania, USA, Rep. MS-CIS-92-83/DSL-12, 1992.

[75] J.-C. Bolot, "End-to-end packet delay and loss behavior in the internet," *SIGCOMM Comput. Commun. Review*, vol. 23, no. 4, pp. 289–298, Oct. 1993.

[76] W. E. Naylor and L. Kleinrock, "Stream traffic communication in packet switched networks: destination buffering considerations," *IEEE Trans. Commun.*, vol. 30, pp. 2527–2534, 1982.

[77] F. Ishizaki, "Analysis of performance improvement with packet dispersion," in *Proc. 2005 IEEE Region 10 Conf. (TENCON 2005)*, Melbourne, Australia, 2005, pp. 1–6.

[78] A. D. Clark, "Modeling the effects of burst packet loss and recency on subjective voice quality," in *Proc. IP Telephony Workshop*, New York, USA, 2001.

[79] ITU-T, "Performance parameter definitions for quality of speech and other voiceband applications utilizing IP networks," Recommendation G.1020, 2006.

[80] T. Friedman, R. Caceres and A. Clark, "RTP control protocol extended reports (RTCP XR)," IETF RFC 3611, Nov. 2003.

[81] ETSI, "Quality of service (QOS) measurement methodologies," ETSI TS 101 329–5 V1.1.1 (2000-11), Nov. 2000.

[82] A. Raake, "Short- and long-term packet loss behavior: towards speech quality prediction for arbitrary loss distributions," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, pp. 1957–1968, 2006.

[83] M. Yajnik, M. Sue, J. Kurose and D. Towsley, "Measurement and modelling of the temporal dependence in packet loss," in *Proc. 18th Annu. Joint Conf. IEEE Computer and Communications Societies (INFOCOM '99)*, New York, NY, USA, 1999, pp. 345–352.

[84] A. Bianco, R. Birke, D. Bolognesi, J. M. Finochietto, G. Galante, M. Mellia, M. L. Prashant and F. Neri, "Click vs. linux: two efficient open-source IP network stacks for software routers," in *Proc. 2005 Workshop on High Performance Switching and Routing (HPSR 2005)*, 2005, pp. 18–23.

[85] I. Kofler, R. Kuschnig and H. Hellwagner, "Evaluating the networking performance of linux-based home router platforms for multimedia services," in *Proc. 2011 IEEE Int. Conf. Multimedia and Expo (ICME)*, 2011, pp. 1–6.

[86] P. Srisuresh and M. Holdrege, "IP network address translator (NAT) terminology and considerations," IETF RFC 2663, Aug. 1999.

[87] M. Holdrege and P. Srisuresh, "Protocol complications with the IP network address translator," IETF RFC 3027, Jan. 2001.

[88] S. Kent and K. Seo, "Security architecture for the internet protocol," IETF RFC 4301, Dec. 2005.

[89] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley and E. Schooler, "SIP: session initiation protocol," IETF RFC 3261, Jun. 2002.

[90] W. Lou and Y. Fang, "A multipath routing approach for secure data delivery," in *Proc. IEEE 2001 Conf. Military Communications (MILCOM 2001)*, 2001, pp. 1467–1473.

[91] L. Huawei, M. Yinghua and L. Zhongcheng, "Evaluation of dispersity routing strategies in anonymous communication," in *Proc. 2004 Joint Conf. 10th Asia-Pacific Conf. on Communications and 5th Int. Symp. on Multi-Dimensional Mobile Communications (APCC/MDMC '04)*, Beijing, China, 2004, pp. 534–538.

[92] S. Kent, "IP encapsulating security payload (ESP)," IETF RFC 4303, Dec. 2005.

[93] B. Gleeson, A. Lin, J. Heinanen, G. Armitage and A. Malis, "A framework for IP based virtual private networks," IETF RFC 2764, Feb. 2000.

[94] T. Ylonen and C. Lonvick, "The secure shell (SSH) protocol architecture," IETF RFC 4251, Jan. 2006.

[95] H. V. Balan, L. Eggert, S. Niccolini and M. Brunner, "An experimental evaluation of voice quality over the datagram congestion control protocol," in *Proc. 26th IEEE Int. Conf. on Computer Communications (INFOCOM 2007)*, Anchorage, AK, USA, 2007, pp. 2009–2017.

[96] R. Teixeira, K. Marzullo, S. Savage and G. M. Voelker, "Characterizing and measuring path diversity of internet topologies," in *Proc. 2003 ACM SIGMETRICS Int. Conf. Measurement and Modeling of Computer Systems (SIGMETRICS '03)*, New York, NY, USA, 2003, pp. 304–305.

[97] R. Teixeira, K. Marzullo, S. Savage and G. Voelker, "In search of path diversity in ISP networks," in *Proc. 3rd ACM SIGCOMM Conf. on Internet Measurement (IMC '03)*, Miami Beach, FL, USA, 2003, pp. 313–318.

[98] A. Orda and R. Rom, "Routing with packet duplication and elimination in computer networks," *IEEE Trans. Commun.*, vol. 36, pp. 860–866, 1988.

[99] Y. Xiaowei and D. Wetherall, "Source selectable path diversity via routing deflections," in *Proc 2006 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '06)*, Pisa, Italy, 2006, pp. 159–170.

[100] T. Fei, S. Tao, L. Gao, R. Guérin and Z.-l. Zhang, "Light-weight overlay path selection in a peer-to-peer environment," in *Proc. 25th IEEE Int. Conf. on Computer Communications (INFOCOM 2006)*, Barcelona, Catalunya, Spain, 2006, pp. 1–6.

[101] M. T. Gardner, C. Beard and D. Medhi, "Avoiding high impacts of geospatial events in mission critical and emergency networks using linear and swarm optimization," in *Proc. 2012 IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2012, pp. 264–271.

[102] L. Ding and R. A. Goubran, "Speech quality prediction in VOIP using the extended E-Model," in *Proc. 2003 IEEE Global Telecommunications Conf. (GLOBECOM '03)*, vol. 7, pp. 3974–3978, Dec. 2003.

Appendix

Figure 36 below depicts a subset of the contents of a call profile measured on 27 December 2011. For brevity, in the figure all but the first two RTP packets have been removed and replaced with a vertical ellipsis. Furthermore, the peer has been anonymised.


```xml
<?xml version="1.0" encoding="utf-8"?>
<profile xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:noNamespaceSchemaLocation="CallProfile-2.0.0.0.xsd">
  <context>
    <category>incoming</category>
    <host>truffle</host>
    <start>2011-12-27T15:11:00.951494+0800</start>
    <uniqueId>2367820539</uniqueId>
  </context>
  <media length="1">
    <audio length="2">
      <rtpMap>
        <payloadType>8</payloadType>
        <encodingName>PCMA</encodingName>
        <clockRate>8000</clockRate>
      </rtpMap>
      <rtpMap>
        <payloadType>101</payloadType>
        <encodingName>telephone-event</encodingName>
        <clockRate>8000</clockRate>
      </rtpMap>
    </audio>
  </media>
  <packetCollection>
    <packet>
      <direction>read</direction>
      <loggedTimestamp>2011-12-27T15:11:02.001407+0800</loggedTimestamp>
      <payloadLength>160</payloadLength>
      <payloadType>8</payloadType>
      <rtpTimestamp>1773402246</rtpTimestamp>
      <sequenceNumber>20286</sequenceNumber>
      <serialNumber>1</serialNumber>
      <ssrc>451248190</ssrc>
    </packet>
    <packet>
      <direction>read</direction>
      <loggedTimestamp>2011-12-27T15:11:02.022339+0800</loggedTimestamp>
      <payloadLength>160</payloadLength>
      <payloadType>8</payloadType>
      <rtpTimestamp>1773402406</rtpTimestamp>
      <sequenceNumber>20287</sequenceNumber>
      <serialNumber>2</serialNumber>
      <ssrc>451248190</ssrc>
    </packet>

    ⋮

  </packetCollection>
  <delay>55898</delay>
  <peer>192.0.2.12</peer>
</profile>
```


Figure 36: *Contents of a call profile with all but the first two RTP packets removed from the figure for the sake of brevity, and with an anonymised peer.*

The elements that may be found in each call profile are enumerated in table 6, along with a brief description of each element. Other information about the VOIP telephone call described by a call profile, such as the length of the call, which packets were lost, or the amount of delay variation (that is, jitter) experienced by a packet in relation to another packet, may be derived from the data contained in the call profile.

Table 6: *Elements that may be contained in each call profile, and a description for each element. Other information, such as the length of the VOIP telephone call, may be derived from the data contained in the call profile for that call.*

Element	Description
profile	Call profile for a VOIP telephone call.
context	Context of the call.
category	Category of the call. The following values are used: 1 String "incoming" for calls from remote hosts to local hosts. 2 String "outgoing" for calls from local hosts to remote hosts. 3 String "simulated" for simulated calls.
host	Name of the host on which the call profile was created.
start	Date and time that the call started.
uniqueId	Unique identifier of the call.
media	Description of media that may be contained in the call. The length attribute, when present, quantifies the number of child elements.

Element	Description
audio	Description of audio media that may be contained in the call. The length attribute, when present, quantifies the number of child elements.
rtpMap	One or more dynamic mappings of an RTP packet payload type to a description of the media in the payload of that packet.
payloadType	RTP packet payload type being mapped.
encodingName	Name of the media encoding for a payload type.
clockRate	Clock rate of the media encoding for a payload type.
packetCollection	RTP packets read for the call.
packet	A single RTP packet read for the call.
direction	Direction of RTP packet. The following values are used: 1 String "read" for packets sent by remote hosts to local hosts. When both are local or both are remote, for packets sent by called hosts. 2 String "write" for packets sent by local hosts to remote hosts. When both are local or both are remote, for packets sent by calling hosts.
loggedTimestamp	Date and time that RTP packet was read, as determined by pcap.
payloadLength	Length, in octets, of the payload of the RTP packet; that is, excluding the header. Computed as the length, in octets, of the RTP packet as read by pcap, less the length, in octets, of the RTP packet header.

Element	Description
payloadType	Payload type of the RTP packet, as read from the PT field of the RTP packet header [69].
rtpTimestamp	Timestamp of the RTP packet, as read from the timestamp field of the RTP packet header [69].
sequenceNumber	Sequence number of the RTP packet, as read from the sequence number field of the RTP packet header [69].
serialNumber	Serial number of the RTP packet, where the first RTP packet read for a call has serial number 1, the next RTP packet read for that call has serial number 2, and so on.
ssrc	Synchronization source of the RTP packet, as read from the SSRC field of the RTP packet header [69].
delay	Estimate of delay for the call, in microseconds, as a moving average of the halved round-trip propagation delays computed from sender and receiver reports contained in RTCP packets sent to the local host in the call [69].
peer	A unique identifier of the remote host participating in the call.