

Internet QoS Measurement and Traffic Modelling

Haitham Cruickshank

Centre for Communication Systems Research,
University of Surrey,
U.K.

Tel: +44 (0) 1483 69 9493

Tel: +44 (0) 1483 69 6011

z.sun@surrey.ac.uk





Outline of presentation

- Introduction.
- Issues on QoS
- Nature of Internet traffic.
- Traffic characteristics and modelling
- Traffic measurement
- Packet size and inter arrival analysis
- Experimental Results on Internet applications
- Conclusion.



Introduction

- **Why QoS measurements and traffic modelling?**
 - To find out what is the acceptable QoS that network should provide for users
 - To efficiently dimension the network to support the traffic efficiently meeting the QoS requirements
- **What are the challenges?**
 - Convergence of different technologies (still evolving)
 - Development of new applications and services (which have different traffic profile)
 - Different Requirements for QoS
- **The end-to-end approach is presented here.**



Traf

-T)

- Constant bit rate (CBR) or variable bit rate (VBR)
- The timing relationship requirements (maximum Cell Transfer Delay and peak-to-peak Cell Delay Variation)
- The type of connection necessary to satisfy the above requirement (connection oriented or connectionless)



Nature of traf

- Elastic – historically for computer data services including Telnet, FTP, NFS, Email and today WWW. They worked well on different type networks from slow dialup links to high speed broadband networks. Correctness is important though high throughput may make users feel quicker
- Inelastic – Timing and throughput were important parameters to meet QoS requirements for real time services (voice and video) and data service response time



Voice QoS (E-model, ITU-T G.109)

- $R = R_0 - I_s - I_d - I_e + A$

R_0 - The voice quality generated at the terminals

I_s - The impairment caused by digitisation

I_d - The effect of delay on the voice quality

I_e - Equipment Impairment factor, such as the effect of codec characteristics and packet loss ratio

A - Advantage factor. This parameter shows how much bonuses that user will be willing to give to the quality rating otherwise the user will not be able to get the service. This means that customers may accept some degradation in quality for access advantage, such as satellite links



TCP QoS and ef

According to the recent theory and measurement, the effective bandwidth can be estimated for TCP as the following:

- $B = \min\{W_{\max}/RTT, [RTT (2bp/3)^{1/2} + T_0 \min[1, 3(3bp/8)^{1/2}]p(1+32p)^2]\}$

where, **Wm** is the maximum advertised receive window, **T₀** initial timeout value, **RTT** roundtrip time, **b** the number of packets acknowledged by 1 acknowledgement, **p** packet loss probability, **B** the effective bandwidth of (or throughput achieved by) a TCP flow.

- In the future, multimedia applications have to be taking into account



Traf

- Numerous models have been developed to capture and represent the randomness/stochastic behaviour of traffic sources.
- Some key characteristics of a traffic source include:
 - Timing information w.r.t the arrival process (extracted either by inter-arrival time or a counting process).
 - Packet sizes, burstiness, duration of a connection.
 - Correlation between successive arrivals or between arrivals from different sources.
- User behaviour, with the growth of Internet-related traffic, is becoming an important factor effecting the overall characteristic of the traffic.



Traf

- We used Software for the purpose of cost effectiveness and flexibility based on the WinPcap library, though there were commercial equipment available for the measurement
- Measured parameters were Packet Inter-arrival Time and Packet Size
- Measured Applications were WWW and FTP
 - Two flow directions: upstream and downstream
 - Packet length distribution analysis
 - Packet interarrival time was fitted with CDFs)



WWW Packet Length Analysis (1/2)

- Most of the upstream packets, around 82%, have the same size of 64 bytes, which is the minimum Ethernet packet size.
 - simple requests or acknowledges of received packets
- The most interesting thing is that around 17% packets' length fell into the range from 295 to 641 bytes in the upstream.
 - Interaction operation between users and servers
- Around 80% of the downstream packets are either 64 bytes or 1518 bytes long.
 - 64 bytes long packets are acknowledgements from Internet servers.
 - 1518-bytes packets are used to transport objects located in each web page.
 - The downstream has bigger data blocks than the uplink traffic.
- Around 20% of the downstream packets locate in the wide area between the two extreme points.



WWW Packet Length Analysis (2/2)

- The upstream packet length distribution in the range between 295 bytes to 641 bytes is white noise like.
- To simplify the model, we choose the average possibility to approximately represent the probability of packet length falling to any point in this particular range.
- For all of the upstream traces, almost all of the packet has a length either 64 bytes or between 295 bytes to 641 bytes.
- For the downstreams, packet length has noise like distribution between 65 bytes to 1517 bytes, which can be treat as the same as the upstream.



WWW Packet Interarrival Time Analysis

- The packet interarrival time is defined as the difference of the arrival times of the i^{th} packet and the $(i-1)^{\text{th}}$ packet

$$\textit{ArrivalInterval} = \textit{ArrivalTime}(i) - \textit{ArrivalTime}(i - 1)$$

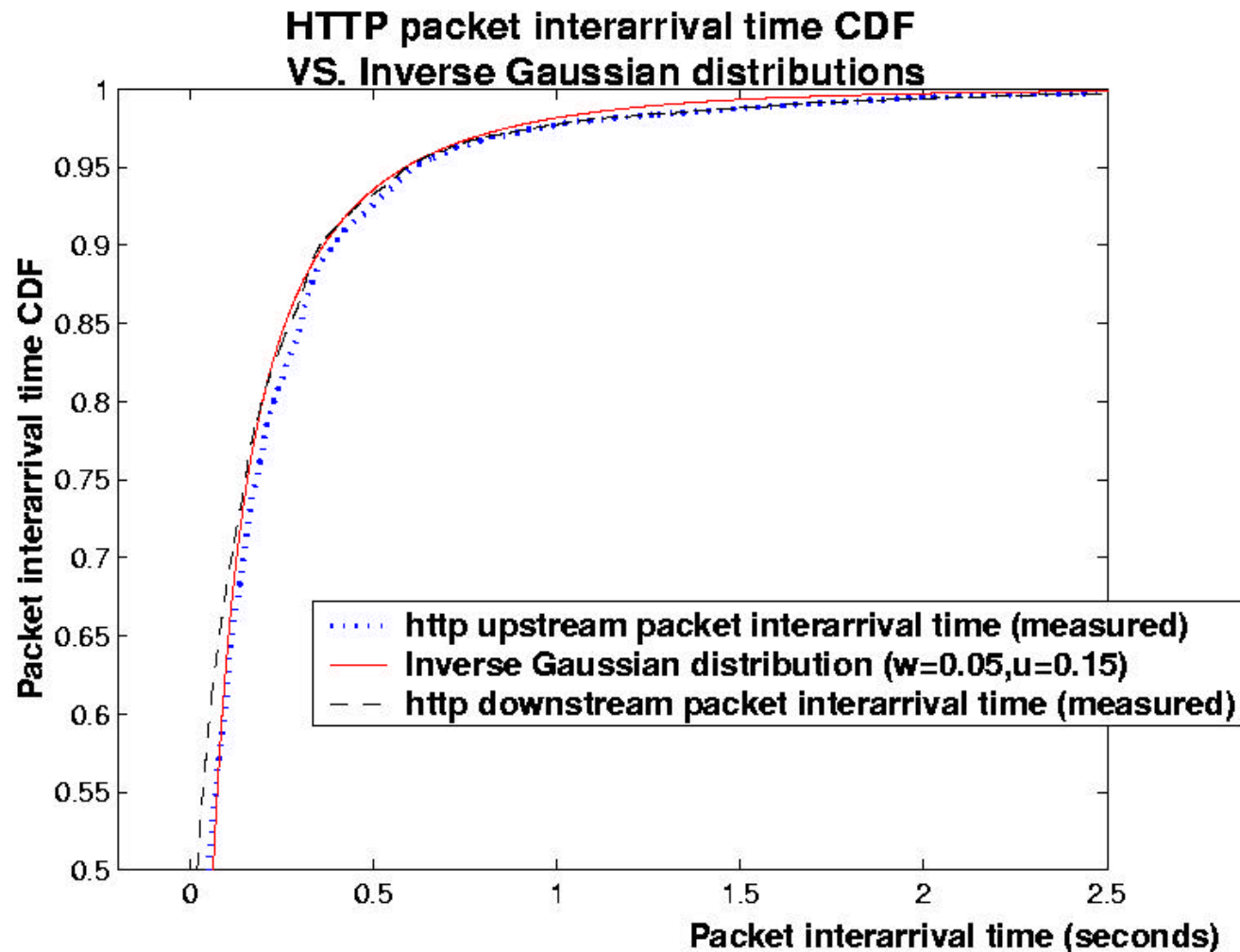
- The measured traces have shown a long-tailed distributions
- The HTTP requests and responses for one web page that contain several objects are typically captured with very short interarrival times.
- Between each web page loading, a user needs reading time or thinking time, which may last up to tens of minutes and cause the tail of the distribution to be significant.
- The round-trip time and the interarrival time between each object in one page also contribute to the shape of the curve



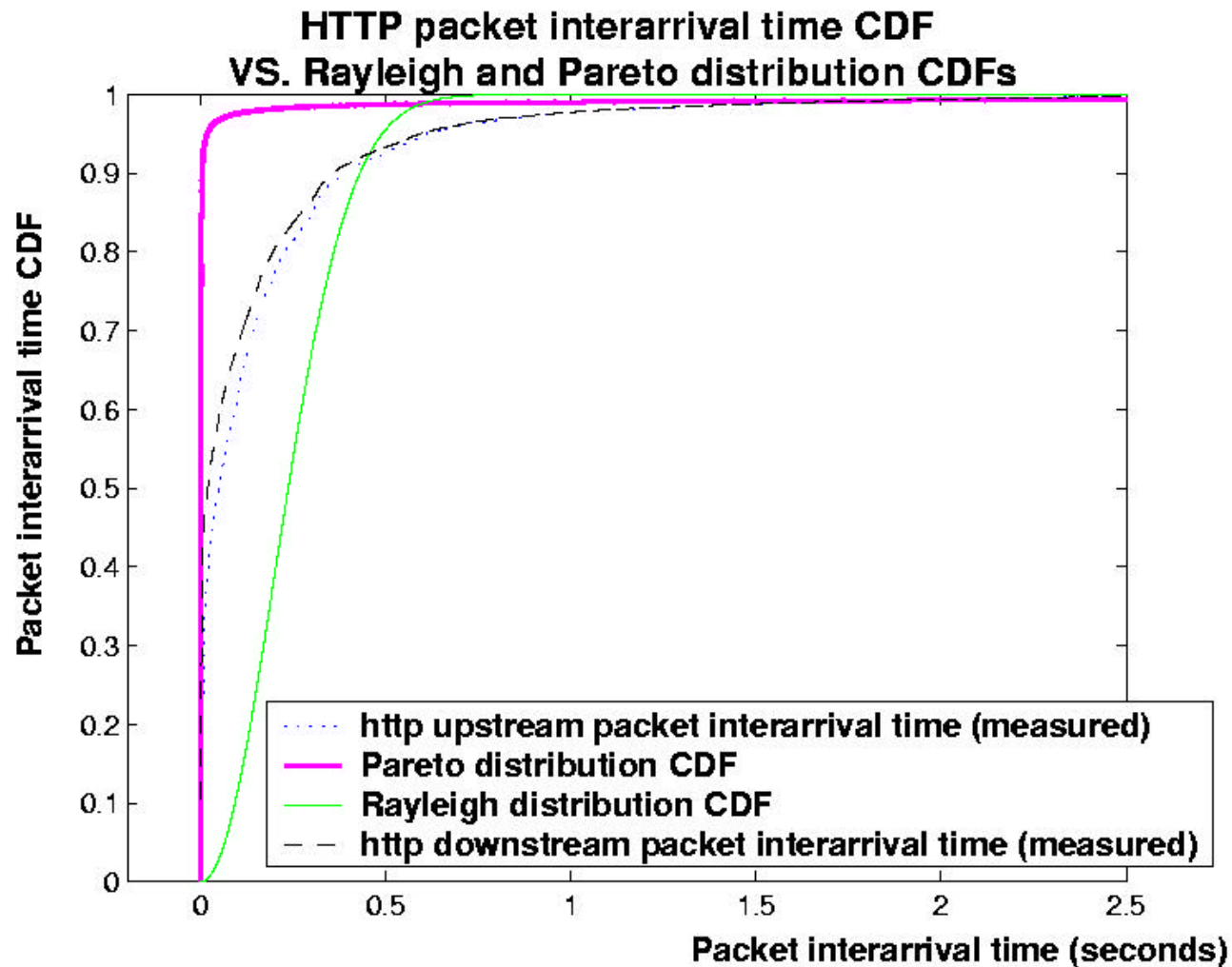
WWW Packet Interarrival Time Modelling

- Many heavy-tailed distributions were tested by varying relevant parameters.
 - Chi-squared distribution, exponential distribution, inverse Gaussian (Wald) distribution, lognormal distribution, Pareto distribution and Rayleigh distribution
- Both Probability Density Functions (PDFs) and Cumulative Distribution Functions (CDFs) of these distributions were used to compare with the measurement plots of each of the traces.
 - The CDF is a better way to present all of the theoretical curves and the measurement curves as well as being easier to use.
- Inverse Gaussian distribution is the best fit to the measurements and other distributions are much worse .

Fitting Using Inverse Gaussian Distribution



Compare with Rayleigh and Pareto Distributions





The Inverse Gaussian Formula

- Thus, the Inverse Gaussian distribution was chosen to model both uplink and downlink WWW packet interarrival time.
- The mathematical representation of PDF of the curve is:

$$f_T(t) = \left[\frac{l}{2t^3 p} \right]^{\frac{1}{2}} \exp \left\{ \frac{-l(t-m)^2}{2m^2 t} \right\}$$

- However, the CDF of the Inverse Gaussian distribution is not given explicitly, and we used the observation that the CDF is the integral of the PDF to approximate it:

$$F_T(t) = \int_{\min}^{\max} f_T(t) dt \approx \sum_{\min}^{\max} f_T(\Delta t) \Delta t$$



FTP Packet Length Analysis

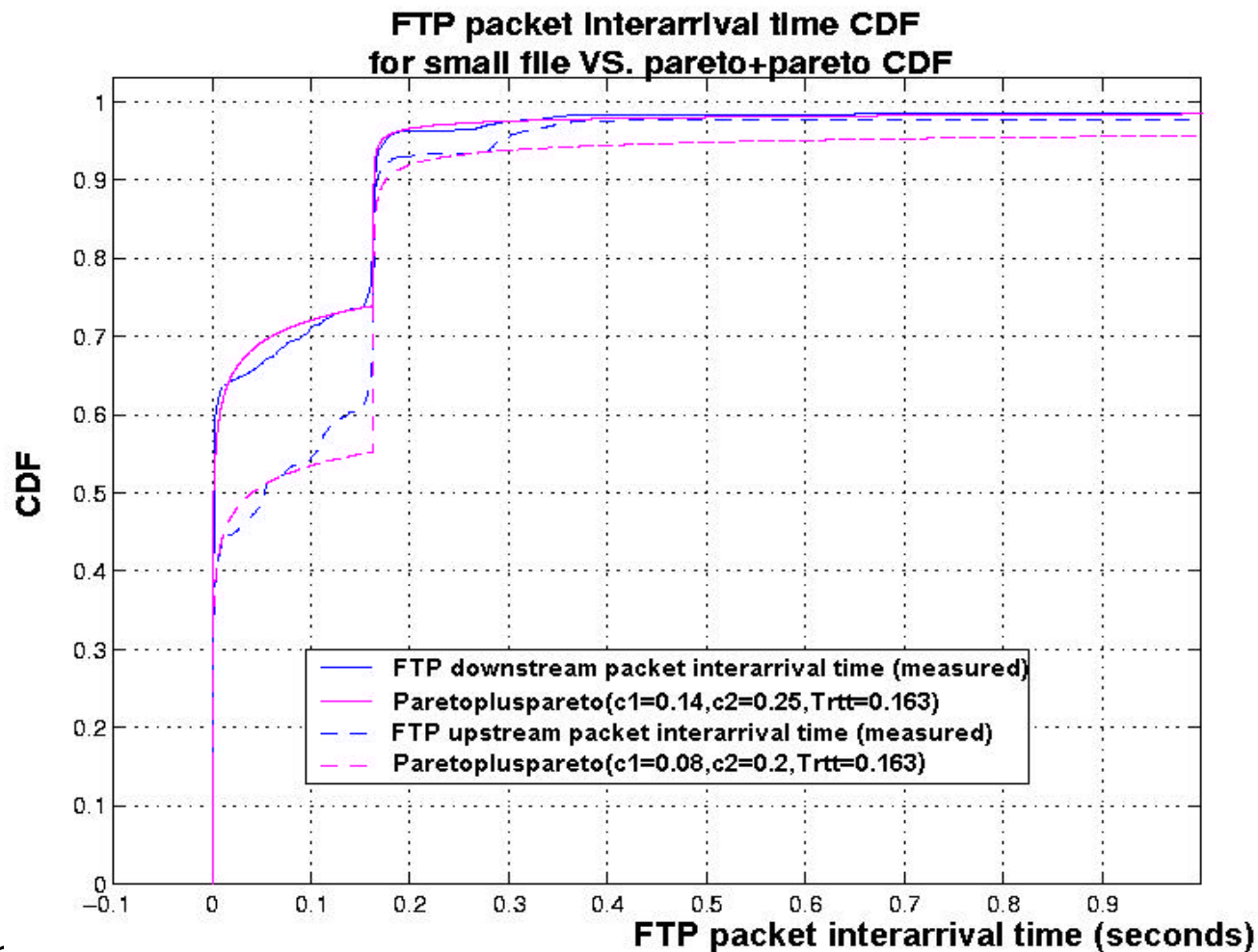
- All FTP traces were based on file downloads from servers to the workstation.
- The variability of packet length for downloading small files is high than the one of big files
 - The small file downloading generated more upstream packets larger than 64 bytes than big file does, and more small downstream packets around 64 bytes.
- By investigating the load of the packets, we believe that those FTP servers transmitted files in block mode.
- The FTP packet size distribution is principally based on the two main factors:
 - Downloaded file size, which reflect the users' operation behaviours.
 - Transmission mode



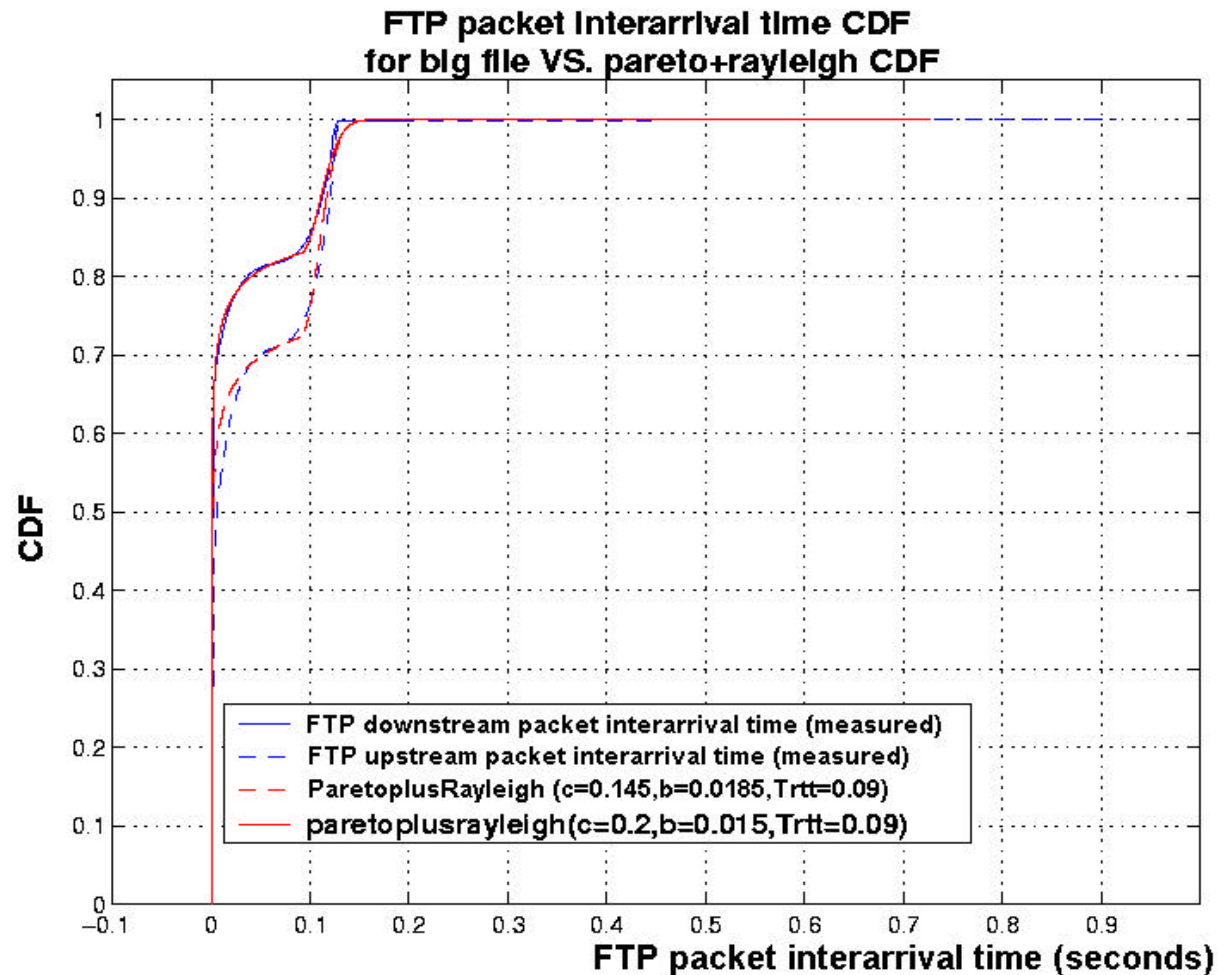
FTP Packet Interarrival Time Analysis

- Downloading files always produces very small interarrival time.
- Either for downloading small file or big file, the RTT has significant effect on the packet interarrival time comparing to typical web browsing HTTP traffic.
- The difference of the presentation of RTT in both cases is decided by user's operation.
- It's necessary to separate these two conditions when we try to establish the FTP interarrival time mathematic models.

Fitting Using Pareto+Pareto Distribution



Fitting Using Pareto+Rayleigh Distribution





FTP Packet Interarrival Time Formula (1/3)

- It has been found that there is no standard distribution can fit well to the measured distributions of the interarrival time for both small and big file download.
 - Pareto distribution fits the measurement curve very well around 0 second
 - Sharp rise cuts off the distribution around the RTT point
- Two different standard distributions were combined to model this kind of cut-off distributions.
 - It should guarantee the final distribution $f_X(x)$ has a CDF:

$$F_X(x) = \int_k^{\infty} f_X(x) dx = 1 \quad x \geq k$$



FTP Packet Interarrival Time Formula (2/3)

- For the small file download, the rise is very sharp. To model this distribution, we chose Pareto+Pareto distribution as the ideal model.

$$f_T(t) = \begin{cases} \frac{c_1 T_{\min}^{c_1}}{t^{c_1+1}} & T_{\min} \leq t < T_{RTT} \\ \mathbf{b} \frac{c_2 T_{RTT}^{c_2}}{t^{c_2+1}} & T_{RTT} \leq t \leq T_{\max} \end{cases}$$

and
$$\mathbf{b} = 1 - \int_{T_{\min}}^{T_{RTT}} \frac{c_1 T_{\min}^{c_1}}{t^{c_1+1}} dx$$

where T_{RTT} is the cut-off point. T_{\min} and T_{\max} is the minimum and maximum value of the FTP packet interarrival time respectively.



FTP Packet Interarrival Time Formula (3/3)

- It was found that Pareto+Rayleigh distribution could model the packet interarrival time very well for big file case.

$$f_T(t) = \begin{cases} \frac{cT_{\min}^c}{t^{c+1}} & T_{\min} \leq t < T_{RTT} \\ \mathbf{b} \frac{t}{b^2} \exp\left(\frac{-t^2}{2b^2}\right) & T_{RTT} \leq t \leq T_{\max} \end{cases}$$

$$\text{and } \mathbf{b} = 1 - \int_{T_{\min}}^{T_{RTT}} \frac{c_1 T_{\min}^{c_1}}{t^{c_1+1}} dx$$

where T_{RTT} , T_{\min} and T_{\max} are the same as previous page.



Experiment results

- A network emulator was developed to introduce end-to-end delay and bit error rate (BER)
- The effects on voice, video and data services were evaluated all based on standard Internet protocols
- The QoS of each service was presented in different form
- These should reflect what users would get over Internet for different QoS of the network in terms of packet loss and end-to-end delay



Voice QoS

BER	Audio Quality
10^{-9} to 10^{-6}	Audio quality not affect by the emulator. This is expected as there is hardly any error being artificially injected into the packets.
10^{-5}	Occasionally, a slight distortion can be detected.
10^{-4}	Slight distortion can be detected regularly.
10^{-3}	High distortion; audio often become indecipherable.

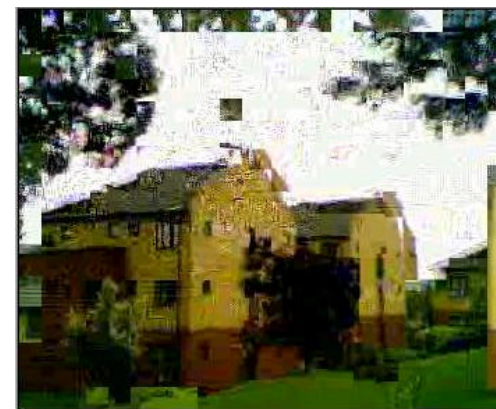
Video QoS



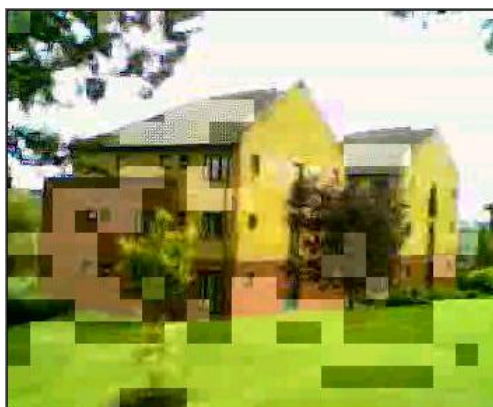
Original



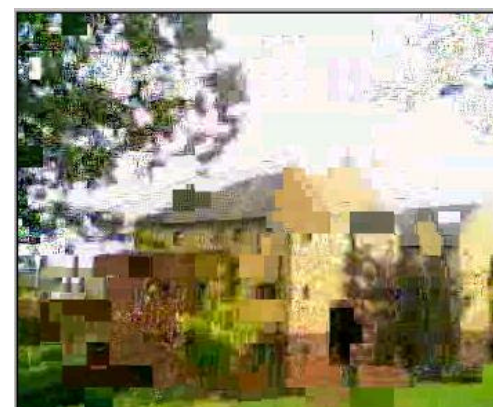
10^{-6}



10^{-5}



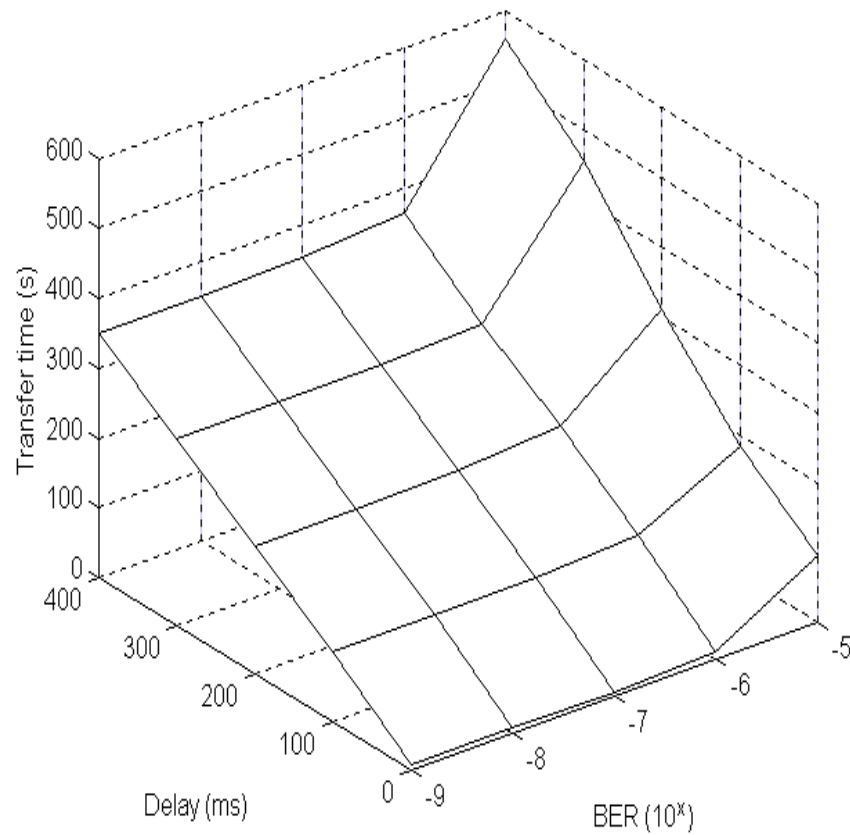
10^{-4}



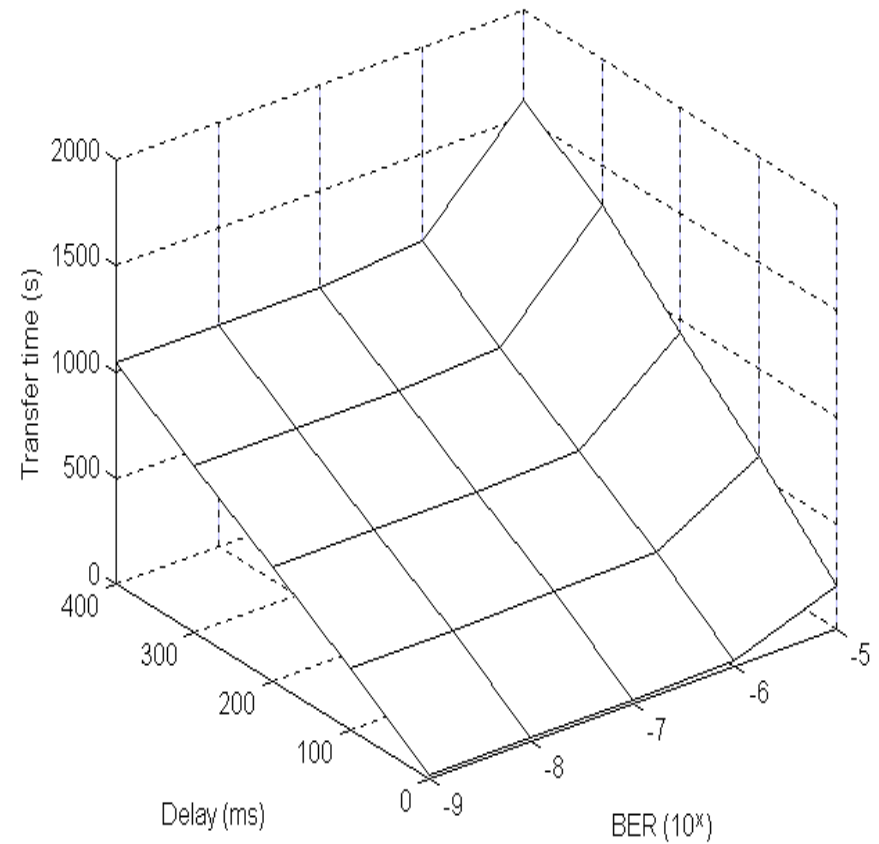
10^{-3}

TCP Data QoS

TCP Transfer for File Size of 1MB



TCP Transfer for file size 3MB





Conclusion

- This paper presented the studies of Internet QoS measurement and traffic modelling and the effects of network QoS parameters (delay and BER) on Internet applications
- It also discussed the key issues on Internet QoS and traffic modelling.
- Traffic measurement were carried out on WWW and FTP applications, and analysis carried out for packet length and packet interarrival time distributions
- Further works should be carried out to find out the relationship between the QoS parameters and user perceived QoS



Acknowledgement

- The authors gratefully acknowledge the support from the European Union 5th Framework IST Programme and the ICEBERGS and GEOCAST projects