Traffic Modelling and Quality-of-Service in High Performance Multimedia Networks

Professor Geyong Min

Networks and Performance Engineering Research Group
Department of Computing
University of Bradford, U.K.
Email: g.min@brad.ac.uk
Outline

- Introduction
- Heterogeneous Network Traffic Characteristics
- Analytical Model for Priority Queueing Systems
  - Queue length distribution
  - Loss probability
- Analytical Model for Hybrid Scheduling
  - Queue Decomposition
- Resource Management and Optimization
  - Buffer Allocation
  - Minimum Bandwidth Requirement
- Conclusions
Introduction

- Advanced multimedia applications with diverse QoS requirements in modern communication networks
  - IP telephony, IPTV, video conferencing, interactive game, and distant learning

- Traffic scheduling plays a key role in the user-perceived QoS
  - Priority Queueing (PQ)
  - Generalized Processor Sharing (GPS)
  - Weighted Fair Queuing (WFQ) - Packet-based GPS (PGPS)
Priority Queueing (PQ)

- PQ is an important and popular mechanism for provisioning of differentiated QoS.
- Many research efforts have been made on its performance analysis, development, and applications.

![Diagram showing traffic flows and priority queueing](image-url)
Generalized Processor Sharing (GPS)

- Fairness
- Traffic isolation
- Work-conserving

Traffic flow $N$

Guaranteed: $\mu_1 C$

Guaranteed: $\mu_N C$

Traffic flow 1

Traffic flow $N$
Introduction - Scheduling

- Hybrid PQ-GPS Scheduling
  - Priority Queueing (PQ) + Generalized Processor Sharing (GPS)/its variants
  - Cisco: IP Real-time Transport Protocol (RTP) Priority, Low Latency Queuing (LLQ)
Introduction – Network Traffic

- Heterogeneous traffic in multi-service networks
  - Long-Range Dependent (LRD) self-similar traffic
    - Generated by multimedia applications; Strict QoS requirements
    - Scale-invariant burstiness
    - Large-lag correlation
  - Short-Range Dependent (SRD) Poisson traffic
    - Generated by the text communications
Introduction - Methodologies

- Performance evaluation studies can be achieved through either simulation or analytical modelling.
- The convergence of simulation to a steady state in the presence of self-similar traffic is often very slow.
  - Analytical modelling is a cost-effective method.
Introduction - Objectives

- Existing research problem
  - No any analytical model has been developed for the hybrid scheduling scheme due to the interdependent relationship between traffic flows and the high complexity of modelling heterogeneous multimedia traffic.

- Objectives
  - To develop the first comprehensive analytical model that can derive the queue length distribution and loss probability in PQ-GPS under heterogeneous traffic;
  - To use the model as a cost-efficient tool for performance analysis and network resource management.
Modelling LRD Self-Similar Traffic

- Scale-invariant burstiness
- Being ubiquitous
  - LAN, WAN, WWW
  - TCP, FTP, and TELNET
  - VBR video and multimedia systems
  - Wireless LAN, Ad-Hoc network traffic
- Degrade network performance and QoS.
Modelling LRD Self-Similar Traffic

- fractional Brownian motion (fBm)
  - An efficient way for modelling and generating self-similar traffic
  - fBm is Gaussian in essence.
  - Formulation: \( A_f(t) = mt + \sqrt{amZ_f(t)} \)
    - \( A_f(t) \): The cumulative number of packets arriving up to time \( t \)
    - \( m \): Mean arrival rate
    - \( \alpha \): Variance coefficient
  - Variance: \( \nu_f(t) = amt^{2H} \)
  - \( H \) (0.5 ≤ \( H \) ≤ 1): Hurst parameter
Modelling SRD Poisson Traffic

- Burstiness over small time scales
- Poisson traffic
  - Formulation: $A_p(t) = \lambda t + \tilde{Z}_p(t)$
  - Variance: $\nu_p(t) = \lambda t$
  - Gaussian approximation: accurate when the arrival rate is large and the processing time tends to infinity

Analytical Model for PQ:

- **Traffic Flow 1**: High Priority, Multimedia, fBM
- **Traffic Flow 2**: Low Priority, Text, Poisson
  - Total Queue Length Distribution
  - Individual Queue Length Distribution
  - Loss Probability
Analytical Model for PQ: Total Queue Length Distribution

- **Lower and upper bounds** of the total queue length distribution of PQ systems
  - Total queue length:
    \[ Q(t) = Q_f(t) + Q_p(t) = \sup_{s \leq t} \{ A_f(s, t) + A_p(s, t) - C(t - s) \} \]
  - **Large Deviation Principle**
  - **Bounds:**
    \[
    \frac{\exp(-\frac{1}{2}Y(t_x))}{\sqrt{2\pi}(1 + \sqrt{Y(t_x)})^2} \leq P(Q > x) \leq \exp(-\frac{1}{2}Y(t_x))
    \]
    \[ t_x = \arg \min_t Y(t) \text{ and } Y(t) = \left( -x + (C - \sum_{i=1}^{N} \lambda_i)t \right)^2 \sum_{i=1}^{N} v_i(t) \]
  - **Queue length distribution:** (Geometric mean)
    \[ P(Q > x) \approx \frac{\exp\left(-\frac{1}{2}Y(t_x)\right)}{\sqrt[4]{2\pi(1 + Y(t_x))^2}} \]
Analytical Model for PQ: Individual Queue Length Distributions

- **High priority fBm traffic**
  - **Fact:**
    - The high priority traffic in a PQ system is served in a manner as if the low priority traffic does not exist.
  - **Queue length distribution:**
    \[
    \frac{\exp\left(-\frac{1}{2} Y_f(t_x)\right)}{\sqrt{2\pi(1 + \sqrt{Y_f(t_x)})}^2} \leq P(Q_f > x) \leq \exp\left(-\frac{1}{2} Y_f(t_x)\right) \\
    t_x = \arg\min_t Y_f(t) \text{ and } Y_f(t) = \frac{(-x + (C - m)t)^2}{v_f(t)}
    \]
Analytical Model for PQ: Individual Queue Length Distributions

- **Low priority Poisson traffic**
  - The total queue in a PQ system is almost exclusively composed of the low priority traffic.
- **Empty Buffer Approximation**: The queue length distribution of the low priority traffic can be approximated by the total queue length distribution.

**Queue length distribution**:

\[
\frac{\exp\left(-\frac{1}{2} Y_f(t_x)\right)}{\sqrt{2\pi(1 + \sqrt{Y_f(t_x)})^2}} \leq P(Q_f > x) \leq \exp\left(-\frac{1}{2} Y_f(t_x)\right)
\]

\[
t_x = \arg\min_t Y_f(t) \quad \text{and} \quad Y_f(t) = \frac{(-x + (C - m)t)^2}{v_f(t)}
\]
Analytical Model for PQ: Model Validation

- **Methodology:** Comparison between analytical and simulation results

- A simulator of the PQ system: C++ language

- **Two typical scenarios**
  - Hurst parameter \( H_f \in \{0.85, 0.75\} \)
  - Service capacity \( C = 120 \)
  - Mean arrival rates
    - Case I: \( \lambda_f = 90 \) and \( \lambda_p = 20 \)
    - Case II: \( \lambda_f = 55 \) and \( \lambda_p = 55 \)
    - Case III: \( \lambda_f = 20 \) and \( \lambda_p = 90 \)
  - Dominative traffic flow: fBm in Case I; Poisson in Case III
Simulation results **closely match** the corresponding analytical results.
Simulation results **closely match** the corresponding analytical results.
Analytical Model for PQ-GPS

High level decomp.

Low level decomp.
Analytical Model of PQ: Application

Buffer allocation

- Total buffer fixed: 400
- QoS requirements:
  - $P_{L,f}(x_f) < 10^{-8}$
  - $P_{L,p}(x_p) < 10^{-2}$
- How to allocate the buffer to fBm and Poisson traffic flows such that their QoS requirements are satisfied?
- Admissible region:
  - $73 < x_f < 367$ and $33 < x_p < 327$
Conclusions

- Developed and validated the comprehensive analytical model for PQ and the hybrid PQ-GPS systems subject to heterogeneous traffic
- Demonstrated the application of the model to performance analysis and resource management