Management and Content Delivery for Smart Networks: Algorithms and Modeling

Academic Year 2024-2025

L1: Network system simulation

packets arrive randomly based on a poisson process
they are qeued in a buffer if the server is busy
when the server is available it transmits the packet (service)
we simulate both lossless (infinite buffer)and losly (finite buffer) scenario

Summary

The purpose of this lab is to simulate the operation of the output link of a router that is modeled as a queuing system and investigate its performance under variable configurations, to understand how different parameter settings may affect the system behaviour and how modifying some parameters of the network system may impact on performance metrics.
In the considered queuing system, representing the output link of a router, the customers represent the packets that arrive to that output link, whereas the service represents the transmission on the channel. Finally, the waiting line represents the buffer where packets are stored before transmission.

Overview

In order to investigate the performance of the router output link, you should focus on the following aspects:

a. Derive the metrics describing the system performance in the baseline scenario, considering both the cases with and without packet losses.

b. Modify the baseline scenario varying the size of the buffer and considering a multi-server system, assuming that two transmitters are present in the system. Test the system behaviour under different buffer sizes (infinite waiting line, medium to short finite waiting line, no buffer at all) and compare its performance with the single server scenario.

You can start from the network simulator code available in the course material, that represents the operation of an M/M/1 queuing system with infinite waiting line, either choosing an event scheduling approach or a process interaction based approach. The simulator is provided in two versions:

- *queueMM1-ES.py*: this version is based on an event scheduling approach;
- *queueMM1-PI.py*: this version is based on process interaction.

You can make your modifications and integration to the code, in order to investigate the system performance via simulation. Different metrics can be useful to understand the system operation and evaluate its performance, like:

- **number of transmitted packets**
- **number of dropped packets**
- **average number of packets** in the system
- **average queuing delay**, i.e. the average time spent in the system
- **distribution of queuing delay**
- **average waiting delay** per packet, assuming that the packet waiting delay is the time elapsed from the moment at which a packet enters the system to the instant at which the service begins.
  It is useful to compute both of the following types of waiting delay:
  - the average waiting delay experienced by any packet (**averaging over all packets**);
  - the average waiting delay **considering only those packets that actually experience some delay**, since they enter the system while the server(s) is (are) busy and are hence buffered.
- **average buffer occupancy**, i.e. the average number of packets in the buffer.
- **loss probability**, i.e. the fraction of packets that cannot be transmitted since the server(s) is (are) currently busy and no buffer is present in the system or, in case of finite buffer, it is full.
- **busy time**, i.e. the cumulative time that during the simulation each server spends in a busy state serving packets requests. In a single-server system, you can observe how it varies with the buffer size. In a multi-server system, based on the busy time analysis per each server, you can examine the distribution of the load among servers.

You can think of other metrics that may be helpful in your analysis to highlight specific aspects or critical issues in the system operation. When running your simulations, pay attention on setting appropriate values of the simulation duration.

Tasks

In your analysis of the network operation and performance, you should perform the following tasks:

1. Investigate the system performance under **different service rates**, keeping a **fixed value for the average arrival rate**:
   (a) how is the system affected? Highlight your findings by showing the variation of relevant performance metrics;

   (b) set a desired confidence level and, for one of the metrics reported in the graphs produced in task 1.a, show the confidence intervals of the performed measurements;
   (c) are the results of your analysis consistent with theoretical expected values?

2. Consider a multi-server system, assuming two transmitters and a shared finite buffer:
   (a) compare its performance with the single-server case;
   (b) test different values of the buffer size for the multi-server system: how do they affect the system performance with respect to an infinite buffer?
   (c) focusing on the multi-server system, compare the performance in the case of finite buffer shared among the two servers and in the case of two separate finite buffers, one per each server.

3. Considering a multi-server scenario, investigate the load distribution among the N servers. Test different algorithms to assign each new request to an available server (random assignment; round robin; assignment to the fastest servers, assuming for example that each server features a different service rate...) and compare the system performance in the various cases. How is load distribution affected?

4. Try to vary the distribution of the service time, i.e. considering the case of M/G/1, and observe how the system performance changes, assuming one or more different distribution types for the service time instead of exponential distribution.

Discuss in your report the main findings emerged from the various investigated scenarios. Support your claims and observations by plotting graphs that report the most significant performance metrics from different cases and help to highlight the relevant findings under variable configuration settings. Remember to always specify the unit of measure for the parameters and metrics represented in the graphs. For each graph, check that the main configuration settings that have been adopted to obtain the corresponding simulation results are reported.

Task 1.C: Theorical Background

↳ For an M/M/1 System:

$\lambda = 0.2$

$\mu = 0.14, 0.20, 0.33$

→ The key formulas:

$$L = \frac{\lambda}{\mu - \lambda} \quad \text{if } \mu > \lambda$$

↳ e.g. $= \frac{0.2}{0.14} \approx 1.5$

→ Average number of Packets

$$W = \frac{1}{\mu - \lambda} \quad \text{if } \mu > \lambda$$

→ Avg delay per packet (W)

$$\rho = \frac{\lambda}{\mu} \longrightarrow \text{utilization}$$