

Report on Exam Task
Simulation and Modelling of Communication Networks

Nicolás Chopitea Kober, Sebastian Lindner

Summer Term 2016

Contents

1	Overview	3
1.1	Description	3
1.2	Requirement Analysis	3
1.2.1	Network Description	3
1.2.2	Statistical Web Browsing Model	4
1.2.3	Open Questions	5
2	Evaluation	6
2.1	Simulation	6
2.1.1	Configurations	6

1 Overview

1.1 Description

We were tasked by a university to analyze the usefulness, practicability and limitations of a remote university building's direct radio link to the main university campus.

In this chapter we will attempt to fully capture the scenario at hand, abstract it into a model and extract the requirements that have to be fulfilled in order to have an applicable solution to connecting the remote building with the main campus via radio.

1.2 Requirement Analysis

1.2.1 Network Description

The aim is to connect a remote building's network to the main university campus network. The fast cabled connection is endangered due to a construction site in close proximity to the building. That is why a direct radio link could be employed to maintain connectivity throughout the construction process.

The following network usage cases could be identified:

CCTV A CCTV camera is connected to the remote building's router.

It is continuously transmitting a livestream of its video.

Wireless A wireless access point operating on the *802.11g* standard is connected to the remote building's router.

Connected to this access point are users engaged in the following three activities.

FTP Upload A single file upload can be observed that is ongoing throughout the simulation of the network.

Video Livestream A remotely located professor is livestreaming his or her lecture to a video conference laptop.

This connection is bidirectional so that participating students can ask questions.

Web Browsing A variable number of students browsing the web is connected to the access point at any time.

This describes the network inside the remote building that we are trying to connect. The main campus' network can currently be simplified to three connections:

Remote Professor A fast connection to the remotely located professor's laptop is established via the Deutsches Forschungsnetz.

It will communicate with the video conference laptop inside the remote building.

Porter's Office A CCTV monitoring station sits in the porter's office.

The livestream of the remote building's CCTV camera will travel here.

Internet A VDSL connection to the Internet is present.

Both web browsing and FTP upload traffic will have this destination.

A graphical representation of the scenario is as follows and includes technical details such as channel delay and bandwidth:



Figure 1.1: A model of the scenario depicted

1.2.2 Statistical Web Browsing Model

The student's web browsing behavior is difficult to capture. We are going to model it by assuming a student issues an HTTP request, receives a response and then spends some time reading the response that is exponentially distributed. Missing at this point is the size of the response that follows a request. To model this we have analyzed a trace file containing 1000 response size values.

We have chosen the *chi squared goodness of fit test* to evaluate how well a distribution fits the observed data. As a first step we investigated the density of values within request size intervals. These intervals initially were of equal size and the observed values were associated to an interval that they lied in. To remove intervals with no values associated to them, these were merged with neighboring intervals. In this way we obtained 160

intervals of which most have observed values associated to them. This is a graphical representation of what we found:



Figure 1.2: Request size value density in intervals

We have the highest density of observed values in regions of low response sizes, and the number of observations decreases exponentially for increasing response sizes. This looks closely related to values coming from a *negative exponential distribution*, which is why we decided to apply the test against this distribution.

The test states that the observed values follow an Exponential distribution with mean $\lambda = 580390$ Byte at a significance level of 99.95%.

1.2.3 Open Questions

We have now modeled the scenario at hand. What remains unclear is

1. What number of students can the network handle?
2. What are the bottlenecks of the network?
3. How is the lecture livestream's error rate correlated to the number of students?
4. What impact does the CCTV livestream have on the network?
5. What impact does the FTP upload have on the network?

In the next chapter we will try and answer these questions.

2 Evaluation

2.1 Simulation

We are going to answer the open questions from section 1.2.3 through simulation. For this we have modeled the scenario using the *omnet++* simulator.

2.1.1 Configurations

We have set up the simulations in the following ways.

Simulation Time Deciding upon a reasonable simulation time is essential. There are several different protocols in use here. The *802.11g WLAN standard* has, for example, a certain time that it takes until handshakes are complete and regular transmission can start.

That is why we have simulated the scenario for an extensive period of 5000s to see how throughput fluctuates in which time frames. Figure 2.1 shows the throughput of the remotely located Professor's laptop that livestreams his lecture to the remote building's conference laptop. We chose this use case as it represents the most time critical of all:

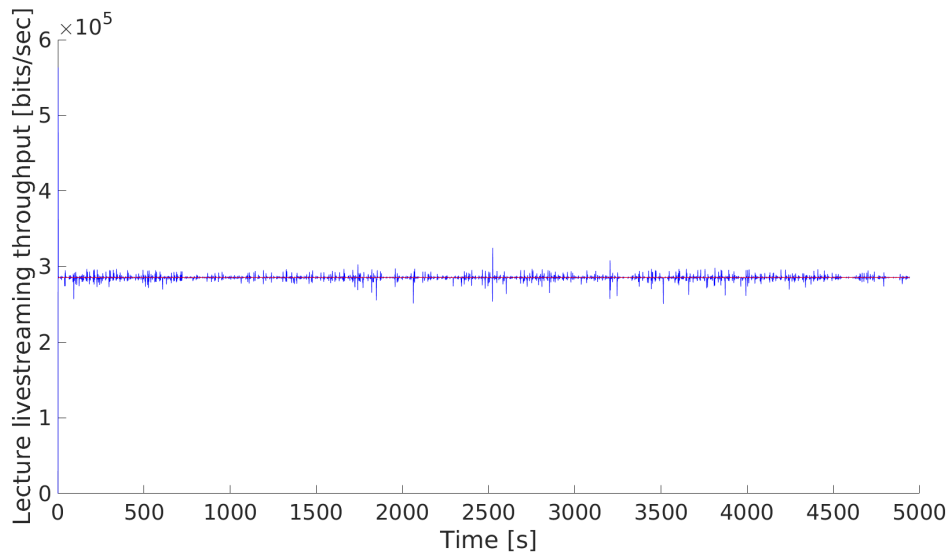


Figure 2.1: Throughput over 5000s

Our approach to this problem looks like this:

We want to find a simulation time m so that if we went through the throughput depicted in figure 2.1 and sliced it into time intervals of length m - in a sliding window-like manner, and then averaged the throughput values inside the sliding windows, we want these means to be relatively close to each other. This would mean that our m is large enough so that it does not matter much which sliding window we took, the result would be more or less the same. This is equivalent to looking at the variance of these mean values. We applied this procedure on the data from the 5000s run for all $m \in \{250s, 350s, 500s, 750s, 1000s\}$. The result can be seen in figure 2.2.

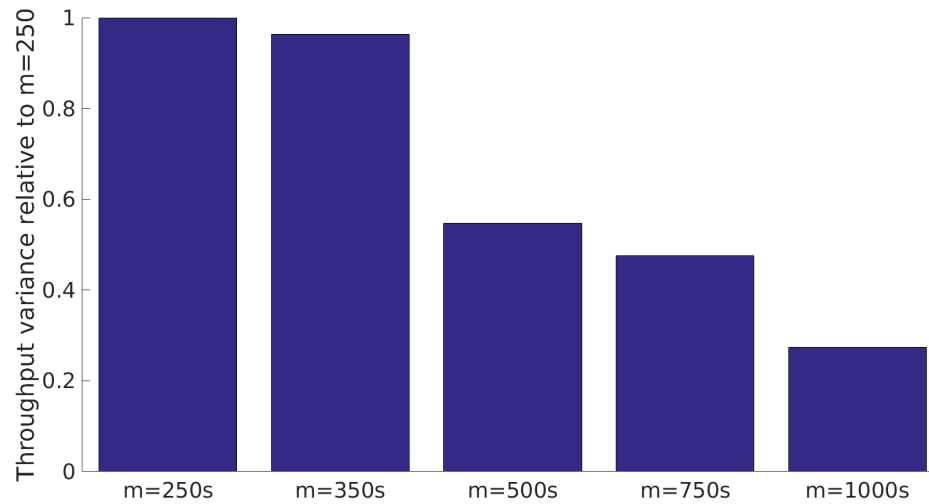


Figure 2.2: Relative variance of the throughput

One can observe that doubling m from $m = 250s$ to $m = 500s$ roughly halves the variance. Increasing it further greatly increases the time required to run these simulations while gaining relatively little.

This led us to the decision of setting the simulation time $m = 500s$.

Number of Repetitions As simulations rely on random events and those build upon pseudo-random numbers that we generate, it is important to run each and every simulation several times with new seeds for the random number generators in order to cover a meaningful number of possibilities with the goal of approximating real world behavior.

Each run's performance is recorded and denoted as a *sample*. Following the *Central Limit Theorem* we can assume a sample mean to stem from a *Normal distribution*.

A sample mean estimate

$$\mu_m = \frac{1}{m} \sum_{j=1}^m y_j$$

where m = number of samples

and $y_i = i^{\text{th}}$ sample

has a confidence interval of

$$\mu_m - \frac{z \cdot \sigma}{\sqrt{m}} \leq \mu \leq \mu_m + \frac{z \cdot \sigma}{\sqrt{m}}$$

where σ = Normal distribution's variance

and z is found using a Normal distribution's table using the required level of confidence $1 - \alpha$, $\alpha \in [0, 1]$.

So apparently the confidence interval depends on $\frac{1}{\sqrt{m}}$, which looks like this:

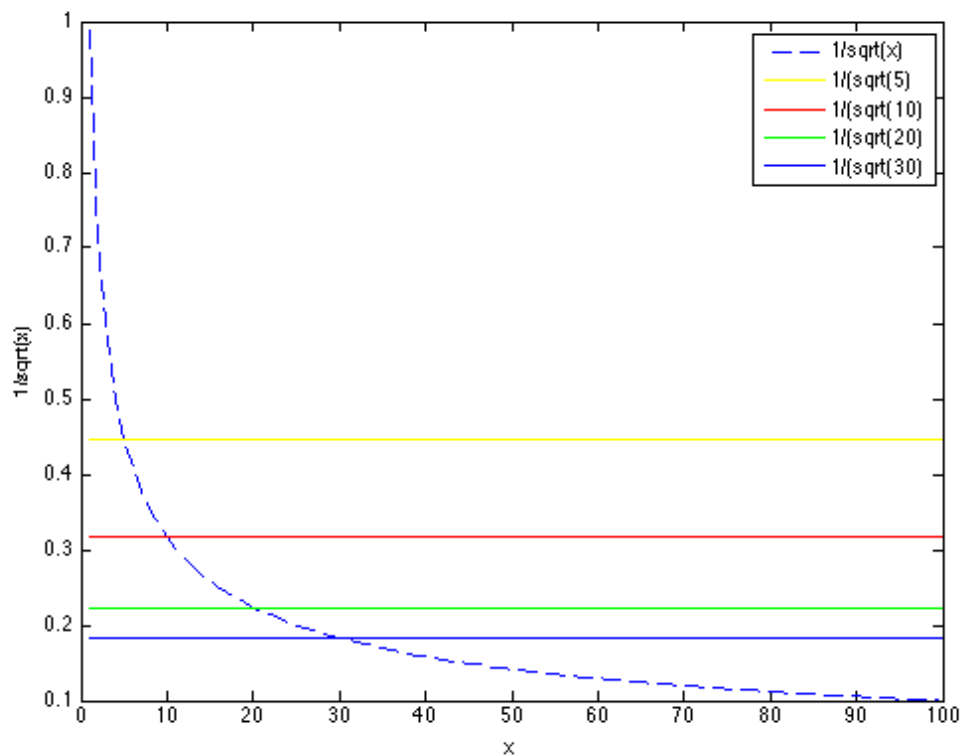


Figure 2.3: Behavior of $\frac{1}{\sqrt{x}}$

In general it should hold that the more repetitions your simulation goes through, the closer its behavior gets to real world conditions. But as with many things, we

want to find a good balance between *correctness* and the *required effort*, or a good efficiency in other words. From investigating figure 2.3 we concluded that setting $m = 20$ seems like a good trade-off as at that point increasing $m > 20$ further does not yield much better results anymore, relative to values of $m < 20$ where performance gains are more significant.

Number of students As the number of students is variable we simulated for a number of values, narrowing down the maximum number of students that could be supported when taking all requirements into account.

CCTV Camera It was communicated to us that turning off the CCTV camera at the remote building location is an option. Thus we simulated every scenario once with the camera *on* and once with it being *off*.