University of Pisa
MSc in Computer Engineering
Performance Evaluation of Computer Systems and Networks
2018 - 2019

Performance Evaluation of an Exam Session

Chiara Bonsignori
Alessio Ercolani
Sara Lotano

Contents

1 Introduction

This study analyzes the performances of an exam session.
The exam committee can be organized as follows:

- *One student per teacher*, from now on it will be called *Parallel.*
- *Pipelined.*

In both systems, in order to pass the exam, each student has to answer a certain number N of questions which corresponds to the number of teachers.
In the first system, each student is examined by a single teacher which asks him the sequence of all N questions. Instead, in the second system, teachers are sorted in a pipeline and each teacher asks the student only one question.
In both systems there is always one student ready to be examined: in the parallel case all teachers examine a new student as soon as they end the previous one, whereas, in the pipelined case a new student enters in the system only when the first teacher is not busy.

2 Modeling

2.1 Assumptions

The developed model works under the following assumptions:

- The communication between teachers and students is considered instantaneous.
- Each teacher does not need any time to think about a new question.
- All student movements within the system are considered null (e.g. in the *Pipelined* case the time spent by a student to move from one teacher to another).

2.2 Implementation

The simulator built for this project is developed using the framework *OMNeT++ v5.*
Each system is implemented by a different *NED Network* and each student is a *cMessage (Student.msg).*

Parallel system The network used is defined in *ParallelExamSession.ned.* To simulate the beginning of an exam, each teacher creates a new student and asks him all the N questions one at a time. Teacher's structure and behaviour are contained in *Teacher.ned, Teacher.h* and *Teacher.cc.*

Pipelined system The network used is defined in *PipelinedExamSession.ned.* In this case teachers (described in *PipelinedTeacher.ned, PipelinedTeacher.h* and *PipelinedTeacher.cc*) can communicate with each other using an handshake protocol, created to permit the exchange of students. The pipeline is implemented as a *blocking system*, i.e. each teacher can send the student to the following colleague only when the latter is available.

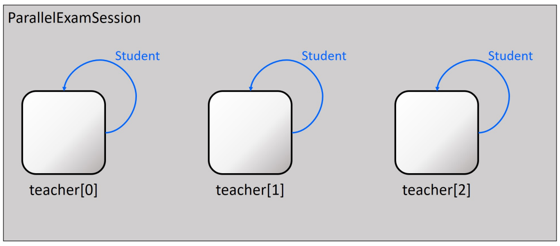

Figure 1: Parallel system (N = 3)

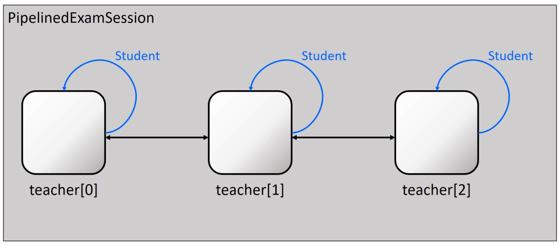

Figure 2: Pipelined system (N = 3)

2.3 Performance indices

Parallel system analyzes:

- *Examination Time* for each student, computed as the sum of the times needed to answer N questions.

Pipelined system considers:

- *Examination Time* for each student, computed as above but including also the *Waiting Time*.
- *Waiting Time* (Figure 3), i.e. time spent by each student inside the pipeline without being examined.
  It is computed as the time interval between the instant a student ends to answer a question and the instant he moves to the following teacher.

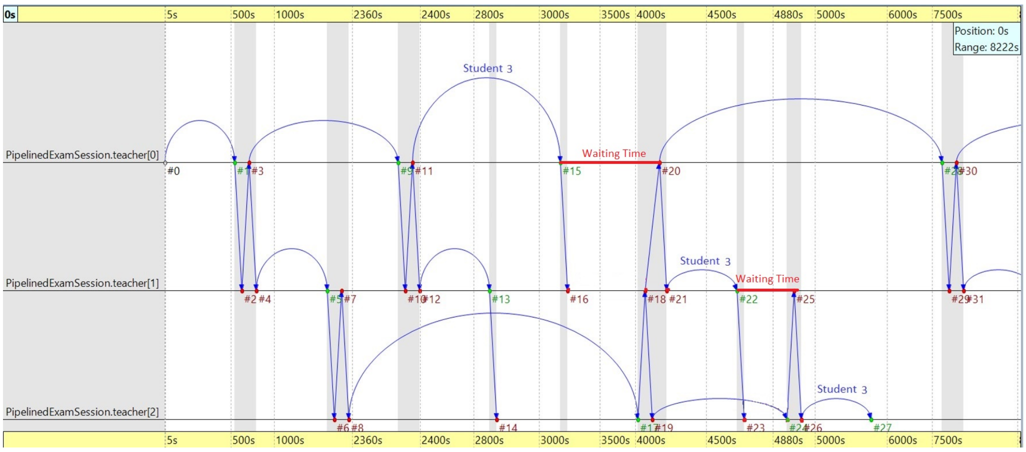

Figure 3: Waiting time example for Student 3 (N = 3)

- *Idle Time*, i.e. time wasted by each teacher. It may occur in the following cases:
  - from the beginning of the simulation to the instant a teacher receives his first student (Figure 4, type 1);
  - when a teacher finishes to examine a student and has not yet received a new one from the previous colleague (Figure 4, type 2);
  - when a teacher finishes to examine a student but the next colleague is still busy and cannot handle him (Figure 4, type 3).

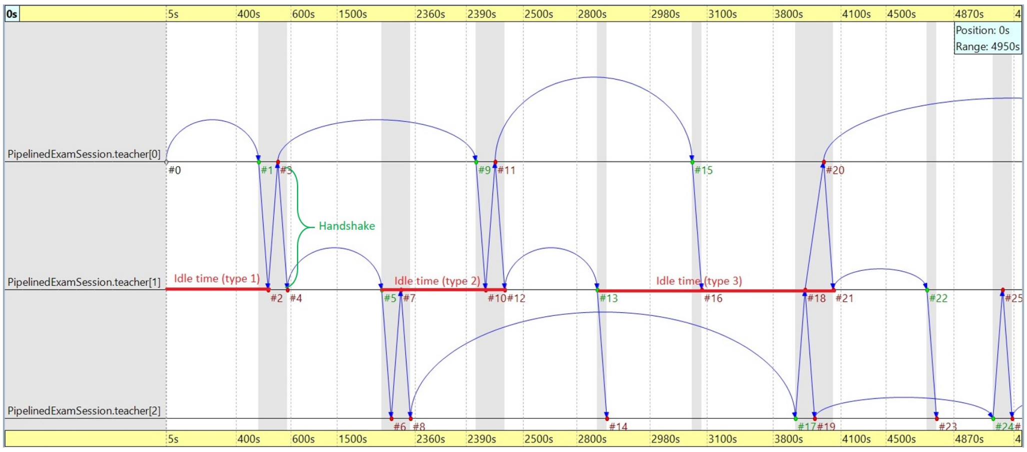

Figure 4: Idle time example for teacher[1] (N = 3)

2.4 Factors and parameters

The analysis is performed varying the number N of teachers in the exam committee. The interval includes values from 1 to 7 to be coherent with real cases (the upper bound 7 is chosen as limit because it is the number of teachers in Italian high school final exams).

The *Answer Time*, i.e the time spent by each student to answer one question, is a value generated randomly according to two different scenarios with *Uniform* and *Lognormal* distributions. Since there is only this random variable, only one RNG is used in the *OMNeT* implementation.

3 Calibration

3.1 Uniform distribution

In the uniform scenario the *Answer Time* is t $\sim U(300, 900)$ where the parameters are specified in seconds and correspond respectively to 5 and 15 minutes. With these values the mean for *Answer Time* is 600s (10 minutes) that is a good estimation of an answer time duration during a real exam.

3.2 Lognormal distribution

In order to make comparable the two scenarios, the mean of the lognormal distribution is the same of the uniform one.
The value for the standard deviation is estimated using the PDF and the CDF of the lognormal distribution (plotted with *plotLognormal.m*). This value is estimated, after some trials, equals to 300s and permits to have a distribution that can fit the real behaviour of an exam. In this way, according to the CDF, only the 10% of the students answer within 5 minutes and only the 15% need more than 15 minutes.

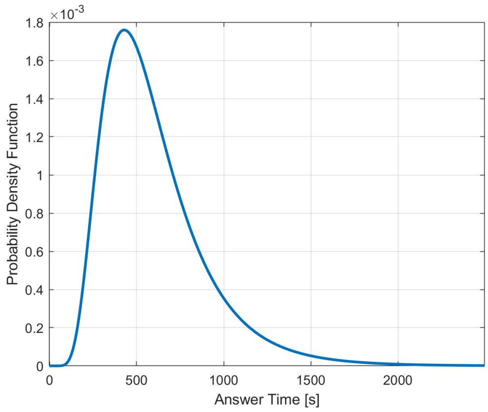

Figure 5: Lognormal PDF

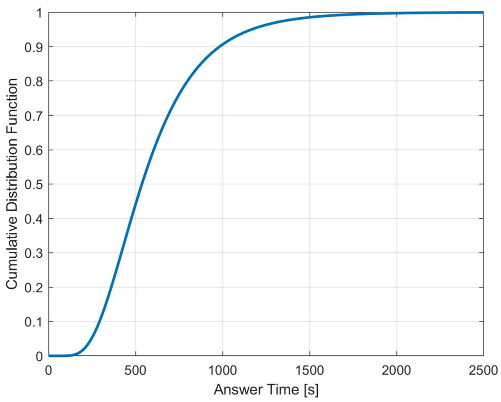

Figure 6: Lognormal CDF

Lognormal and normal distributions are related to each other according to the following relationship: *X is a random variable with a Lognormal distribution if* $Y = \ln(X)$ *is Normally distributed.*[1]
In *OMNeT* lognormal parameters are the mean and the standard deviation of the corresponding normal distribution. Thus, the following formulas are used to compute these values:

$$\begin{cases} \mu_{Normal} = 2 \cdot \ln(\mu_{Lognormal}) - \frac{1}{2} \cdot \ln(\mu^2_{Lognormal} + \sigma^2_{Lognormal}) \\ \\ \sigma_{Normal} = \sqrt{-2 \cdot \ln(\mu_{Lognormal}) + \ln(\mu^2_{Lognormal} + \sigma^2_{Lognormal})} \end{cases}$$

After computations, *Answer Time* is set equals to t $\sim LN(6.2854, 0.4724)$.

[1]Teoria dei fenomeni aleatori - Gaspare Galati, Gabriele Pavan

4 Verification

All tests are performed with $Simulation\ Time$ = 18000s (5 hours) and with N = $\{1, 3, 7\}$.

4.1 First test - Deterministic answer time

To evaluate the correctness of the developed simulator, the first test is performed considering a deterministic *Answer Time* equals to 600s (that corresponds to the mean values of uniform and lognormal distributions). With this deterministic test is possible to compute results by math and compare them with those obtained by the simulator. If they are equals, the simulator can be considered validated.

4.1.1 Parallel system

Table (1.a) contains results obtained by the simulator and table (1.b) contains results computed by math with the following equations:

$$\begin{cases} Examination\ Time = N \cdot Answer\ Time \\ Number\ of\ examined\ students = N \cdot \left\lfloor \frac{Simulation\ Time}{Examination\ Time} \right\rfloor \end{cases}$$

(a) Simulator results

N	Examination Time	Number of examined students
1	$600s$	30
3	$1800s$	30
7	$4200s$	28

(b) Results computed by math

N	Examination Time	Number of examined students
1	$1 \cdot 600s = 600s$	$1 \cdot \left\lfloor \frac{18000s}{600s} \right\rfloor = 30$
3	$3 \cdot 600s = 1800s$	$3 \cdot \left\lfloor \frac{18000s}{1800s} \right\rfloor = 30$
7	$7 \cdot 600s = 4200s$	$7 \cdot \left\lfloor \frac{18000s}{4200s} \right\rfloor = 28$

Table 1: Parallel system

4.1.2 Pipelined system

Table (2.a) contains results obtained by the simulator and table (2.b) contains results computed by math. The *Examination Time* is computed as in the previous case while, for the *Number of examined students*, are used the following formulas:

$$\begin{cases} Number\ of\ examined\ students = N \cdot \left\lfloor \frac{Simulation\ Time}{Examination\ Time} \right\rfloor & \text{if } N = 1 \\ \\ Number\ of\ examined\ students = 1 + \left\lfloor \frac{(Simulation\ Time - ExaminationTime_N)}{Answer\ Time} \right\rfloor & \text{if } N = 2 \dots 7 \end{cases}$$

In the first N·*Answer Time* seconds only one student is examined while, after this time, the pipeline is full and students finish their exam every *Answer Time* seconds.

(a) Simulator results

N	Examination Time	Number of examined students
1	$600s$	30
3	$1800s$	28
7	$4200s$	24

(b) Results computed by math

N	Examination Time	Number of examined students
1	$1 \cdot 600s = 600s$	$1 \cdot \left\lfloor \frac{18000s}{600s} \right\rfloor = 30$
3	$3 \cdot 600s = 1800s$	$1 + \left\lfloor \frac{18000s-1800s}{600s} \right\rfloor = 28$
7	$7 \cdot 600s = 4200s$	$1 + \left\lfloor \frac{18000s-4200s}{600s} \right\rfloor = 24$

Table 2: Pipelined system

In this deterministic scenario there is no *Waiting Time* because all the answer times are equals, thus each student always finds next teacher available.
The *Idle Time* is only due to the initial transient when the pipeline is not yet full. It is 0 only for the first teacher and $i \cdot$ *Answer Time* for the other teachers, according to their i^{th} position in the pipeline.

4.2 Second test - Introduction of randomness

This test evaluates the correctness of the simulator with the introduction of randomness and it is performed for both distributions. However, with the lognormal scenario is more difficult to compare results obtained from the simulator with those computed by math due to the high value of the standard deviation. For this reason the uniform scenario is considered and reported below and the *Answer Time* is $t \sim U(300, 900)$ [s].

4.2.1 Parallel system

Since *Examination Time* is now the sum of N uniform RVs, its mean and standard deviation are computed using the CLT, with $\mu_{Answer\ Time} = 600$s and $\sigma_{Answer\ Time} = 173$s and they are compared with the corresponding values obtained by the simulator. Considering that in this testing phase the simulation is done with only one repetition, results are pretty close to theoretical ones as shown in the table:

	E [Examination Time]		**StdDev (Examination Time)**	
N	**Simulation results**	**Theoretical results**	**Simulation results**	**Theoretical results**
1	$654s$	$600s$	$136s$	$173s$
3	$1863s$	$3 \cdot 600s = 1800s$	$319s$	$\sqrt{3 \cdot 173^2}s = 300s$
7	$4238s$	$7 \cdot 600s = 4200s$	$497s$	$\sqrt{7 \cdot 173^2}s = 458s$

Table 3: Simulation and theoretical results for parallel system

To compute by math the number of examined students, the formula is:

$$Number\ of\ examined\ students = N \cdot \left\lfloor \frac{Simulation\ Time}{E[Examination\ Time]} \right\rfloor$$

and it is verified by simulation results.[2]

[2] Computations are reported here: Project/data/verificationResults.xlsx

4.2.2 Pipelined system

In this system the *Examination Time* is the sum of the *Answer Time* and the *Waiting Time*. The latter is a random variable of an unknown distribution and obtain mathematical formulas to explain simulation results is not possible. However, to evaluate the correctness of the system behavior, *OMNeT* event logs are used. Thanks to the graphical interface it is possible to see that what happens during the simulation is what expected from reality by doing computations manually (see Figure 3).
Same considerations are done for the *Idle Time* (see Figure 4).

4.3 Final considerations

In the **Parallel** system, even if the number of teachers in the exam committee increases (and consequently also the number of questions needed to complete the exam), the total number of examined students is quite the same.
This can be explained with the following formula, where t = Answer Time:

$$\begin{aligned} Number\ of\ examined\ students &= N \cdot \left\lfloor \left(\frac{Simulation\ Time}{E\left[Examination\ Time\right]} \right) \right\rfloor \\ &= N \cdot \left\lfloor \left(\frac{Simulation\ Time}{E[\sum_{i=1}^{N} t_i]} \right) \right\rfloor \\ &= N \cdot \left\lfloor \left(\frac{Simulation\ Time}{\sum_{i=1}^{N} E[t_i]} \right) \right\rfloor \\ &= N \cdot \left\lfloor \left(\frac{Simulation\ Time}{N \cdot E[t_i]} \right) \right\rfloor \end{aligned}$$

In the **Pipelined** system the total number of examined students:

- is lower than in the parallel case due to the *Waiting Time*;
- decreases when the number of teachers increases. This happens because, with a long pipeline, each student can stop more times going from the beginning to the end of the pipeline. This diminution is coherent with the growth of the teachers' *Idle Time* and students' *Waiting Time*.[2]

5 Warm-up time estimation

Warm-up time is evaluated considering the main performance indices (*Examination Time* and *Waiting Time*) in both systems and scenarios for all possible combinations of factors. A simulation time of 604800s (1 week) and 10 repetitions are used to find this value.
The *Waiting Time* in the pipelined system with lognormal scenario and N equals to 7 teachers represents the worst case and it is chosen as the warm-up time. It corresponds to 60000s ($\approx$ 16 hours).

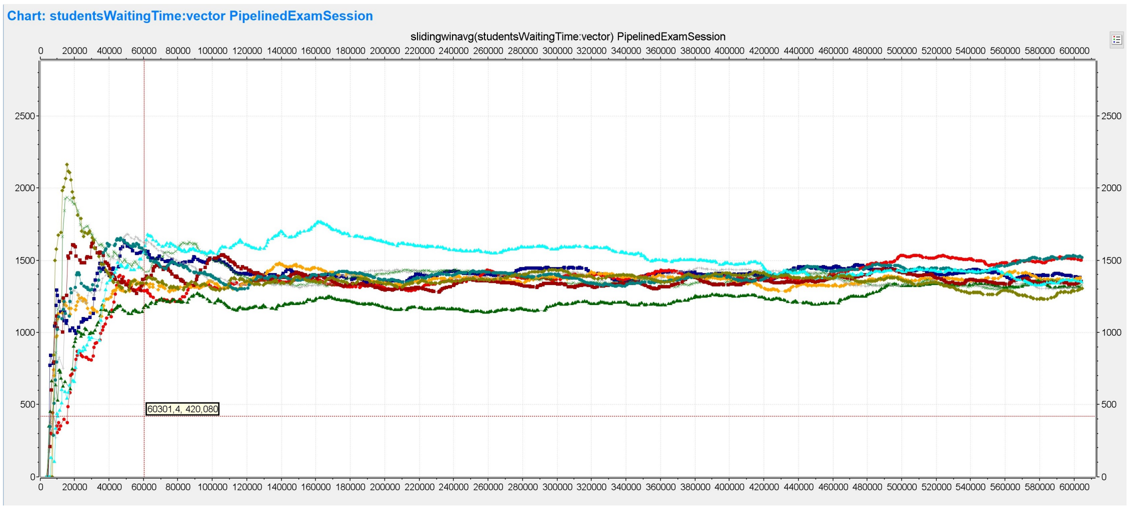

Figure 7: Warm-up time estimation

After the estimation of the *Warm-up time*, *Simulation Time* is set equals to 350000s ($\approx$ 4 days); this leads to get, in the worst case, a sample of about 300 students.
The number of *repetitions* for each simulation is set to 30.

6 Data Analysis

This data analysis is performed computing the 95% *Confidence Intervals* for all the performance indices.

6.1 Examination Time

The first analysis regards the variation of the *Examination Time* with respect to the number N of teachers in the exam committee and is performed for both scenarios.
The *Examination Time* $\overline{Y}$ plotted on the y-axis of the graphs is computed as follows:

1. Average values $\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_{30}$ are computed for each repetition R_i; the latter contains values of examination time experienced by all students.

2. $\overline{Y}$ is computed as the mean of $\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_{30}$.

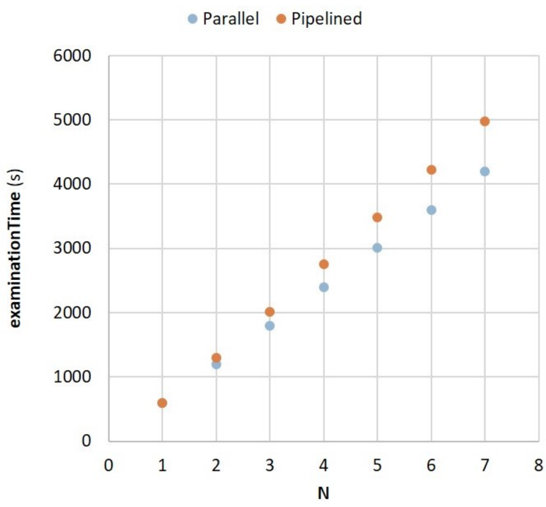

Figure 8: Uniform scenario

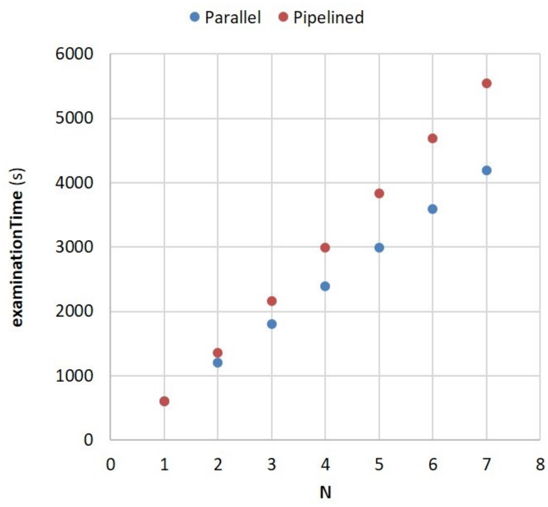

Figure 9: Lognormal scenario

As shown in the graphs, in each single scenario:

- *Examination Time* grows when N increases for both systems and it is always smaller in the parallel case.

- The difference between examination times of pipelined and parallel systems grows with N (for simplicity, it will be called "gap").

Comparison between scenarios:

- In parallel systems, the *Examination Time* is quite the same and this means that the chosen distribution does not influence it.

- The increase of the "gap" is due to the fact that the pipelined lognormal examination times grow faster than the corresponding ones in the uniform case.

Confidence intervals reported in both graphs are too small to be visible ($2 \div 3$ orders of magnitude fewer) as shown in the following tables:

(a) Uniform scenario

N	1	2	3	4	5	6	7
Parallel (s)	600 ±3	1198 ±4	1798 ±4	2397 ±5	3002 ±6	3601 ±7	4202 ±8
Pipelined (s)	600 ±3	1301 ±4	2014 ±8	2756 ±9	3478 ±11	4223 ±11	4973 ±12

(b) Lognormal scenario

N	1	2	3	4	5	6	7
Parallel (s)	602 ±5	1197 ±7	1797 ±9	2394 ±9	2994 ±10	3595 ±10	4195 ±11
Pipelined (s)	602 ±5	1358 ±11	2161 ±14	2994 ±18	3830 ±28	4685 ±29	5548 ±27

Table 4: Examination Time ± Confidence Interval

This means that, in an exam session with a fixed number of teachers, the examination time experienced by students is quite the same. This allows to estimate the mean *Examination Time* with an high level of precision.

Number of examined students Another interesting metric that can be derived from the same data used for the examination time, is the *Number of examined students*.
It is computed as the mean value of the number of examined students in each repetition.

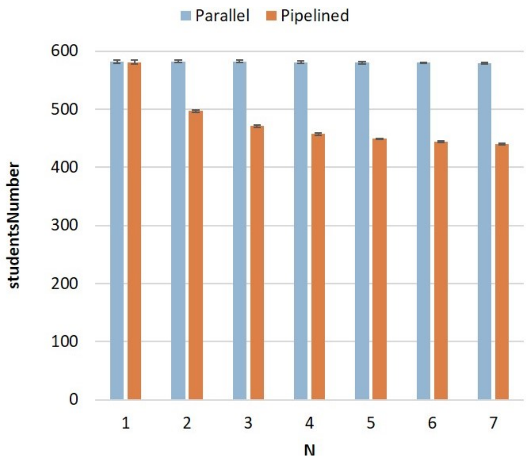

Figure 10: Uniform scenario

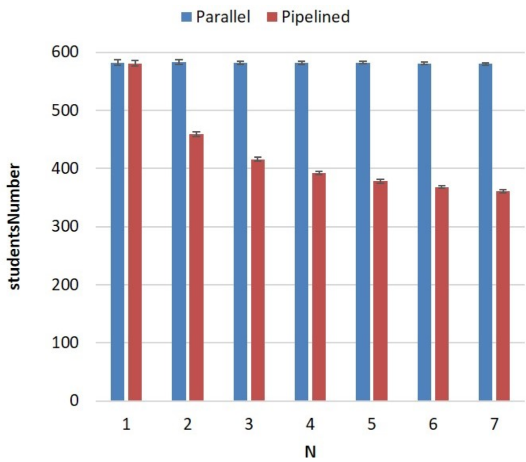

Figure 11: Lognormal scenario

For both scenarios, in the parallel system the *Number of examined students* is pretty constant independently from N and this confirms what shown in the verification phase (see Verification 4.3).
Instead, in the pipelined system, the *Number of examined students* decreases when N increases and this is coherent with the growth of the mean values of examination time. This reduction is higher in the lognormal scenario due to the fact that the *Answer Time* can assume higher values in the lognormal distribution with respect to the uniform one.

6.2 Waiting Time

Waiting Time is one of the metrics used to evaluate performances of the pipelined system, comparing different scenarios.
Data obtained from the simulator are used in the following ways:

- Students' waiting times within each repetition are aggregated using the *median*. The use of the median, instead of the mean, is justified by the high variability within the samples.
  These values are used to plot Lorenz Curves (Figure 12) and ECDFs (Figure 13, Figure 14) in order to evaluate the characteristics of *Waiting Time* distributions.
- Medians are further aggregated using the *mean* operator to obtain a single value for each possible N (Figure 15).
- Percentages of students with no waiting time are computed for each repetition and additionally aggregated using the *mean* (Figure 16).
- An average value for the percentage of waiting time over the total examination time is evaluated computing the *mean* across all the repetitions (Figure 17, Figure 18).

6.2.1 Uniform scenario

The study of Lorenz curves and ECDFs points out how the fairness of the system changes with the number of teachers in the exam committee.

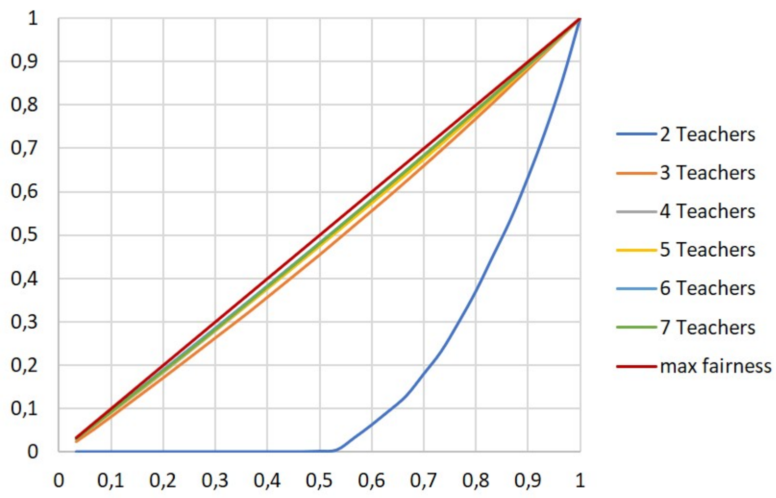

Figure 12: Lorenz curves

The case with N = 2 is the most unfair, indeed, approximately 50% of students experience waiting time, whereas the others find always both teachers available.
For N $\geq$ 3 the system is fair and tends to the *max fairness line*.

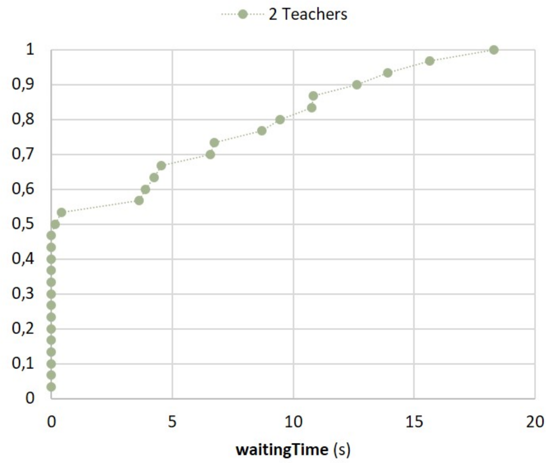

Figure 13: ECDF for N = 2

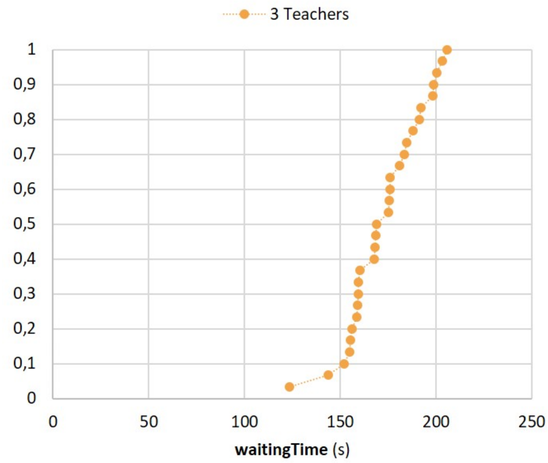

Figure 14: ECDF for N = 3

As previously underlined, the system completely changes its behaviour when the number of teachers in the exam committee goes from N = 2 to N $\geq$ 3.
This becomes clear observing ECDFs in the previous graphs.
For N = 2 ECDF has two separate regions and, thus, students can be treated by the system in two different ways. When N $\geq$ 3, ECDFs are "S-shape" functions with a narrow base on the x-axis and this confirms that all students, within the same exam session, have to wait more or less for the same amount of time.
ECDFs for N $\geq$ 3 are placed at different offsets on the x-axis according to their mean *Waiting Time* values and they maintain their shapes. For the sake of brevity, it is reported only the ECDF when N = 3.

6.2.2 Lognormal scenario

The same conclusions reached for the uniform scenario hold for the lognormal one.
In the following table are shown the comparisons between LCGs of the two scenarios:

N	**2**	**3**	**4**	**5**	**6**	**7**
LCG$_{Uniform}$	0,5391	0,0457	0,0247	0,0267	0,0179	0,0196
LCG$_{Lognormal}$	0,7034	0,0450	0,0338	0,0296	0,0290	0,0249

Table 5: LCGs for both scenarios

If N = 2 the LCG in the lognormal distribution is higher than in the uniform one while for N $\geq$ 3 there are no significant differences.
This means that the distribution does not influence the fairness that hence depends only on the number of teachers in the pipeline.

6.2.3 Other metrics

The following analysis evaluates for all possible values of N the average *Waiting Time* experienced by students and the *percentage* of them with zero waiting time.
The first graph (Figure 15) shows that the average *Waiting Time* has an increasing trend, in particular in the lognormal case.
The second graph (Figure 16) plots the percentages of students that experience zero *Waiting Time* with respect to the total.
The two graphs are coherent with each other because when waiting time increases there is a smaller proportion of students that does not experience it.

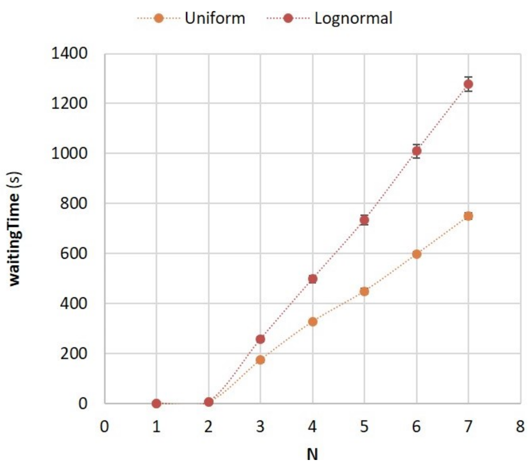

Figure 15: Waiting time

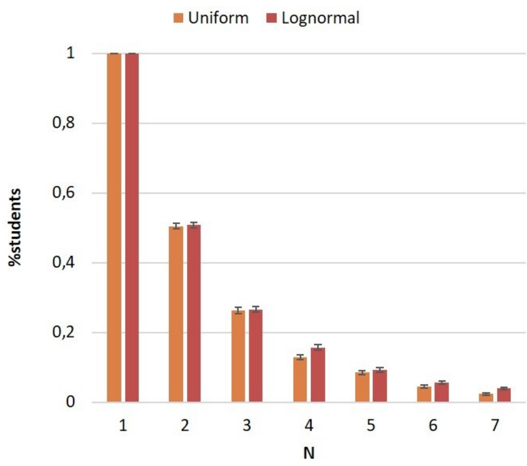

Figure 16: Students with no waiting time

The trend of the *Waiting Time* may be misunderstood. Indeed, if it is related with total examination time, its increase results to be not so "wild": in the worst case, in the uniform scenario the percentage is $\leq 20\%$ and in the lognormal one is $\leq 25\%$ (as shown in graphs of Figure 17 and Figure 18).

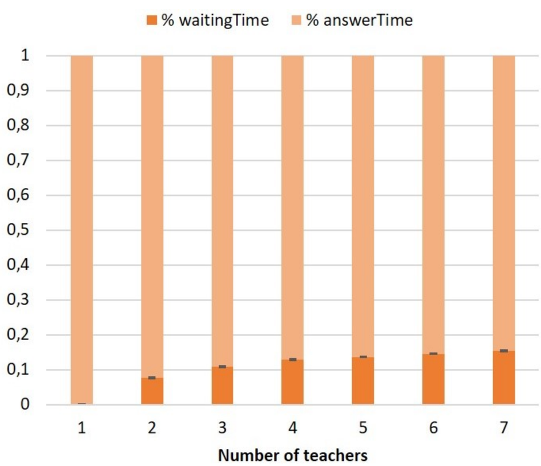

Figure 17: Percentage of waiting time
Uniform scenario

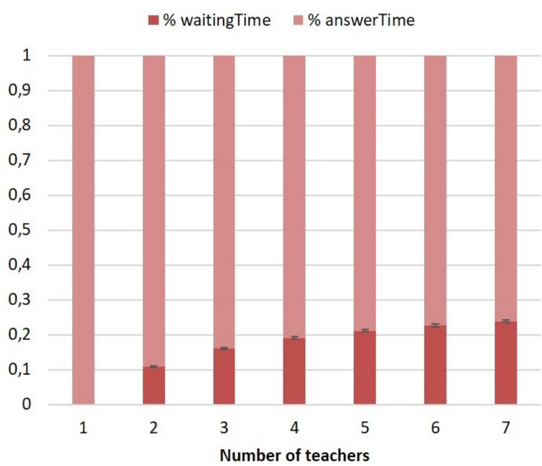

Figure 18: Percentage of waiting time
Lognormal scenario

6.3 Idle Time

The last metric used to evaluate pipelined system is the *Idle Time* computed in two ways: mean idle time for exam committee and mean idle time for each teacher.

6.3.1 Mean idle time for exam committee

All the following computations are done for each possible number of teachers in the committee and for both scenarios.
The mean *Idle Time* $\overline{Y}$ for exam committee is computed as follows:

1. Average values $\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_{30}$ are computed for each repetition R_i; the latter contains values of idle time for each teacher in the exam committee.

2. $\overline{Y}$ is computed as the mean of $\overline{R}_1, \overline{R}_2, \ldots, \overline{R}_{30}$.

The graph of Figure 19 shows that the *Idle Time* for the exam committee increases with N: at the beginning the growth is more striking and at the end it "seems" to stabilize. Indeed, seven points are not enough to claim that idle time stabilizes after a certain value of N.

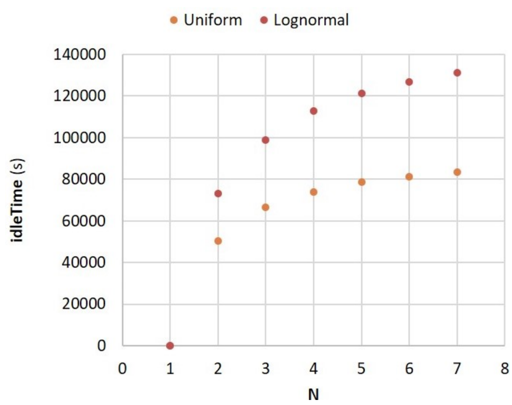

Figure 19: Idle time for exam committee

N	1	2	3	4	5	6	7
Uniform	0%	14%	19%	21%	22%	23%	24%
Lognormal	0%	21%	28%	32%	35%	36%	37%

Table 6: Percentage of idle time over the total simulation time

Since values computed so far are evaluated considering a simulation time of $\approx$ 4 days, a more interesting thing is to compute percentages with respect to the total.
As shown in the above table, when N increases, the percentage of teachers' wasted time grows a lot. For instance, in the lognormal case with N = 7, the exam committee spends about the 37% of the exam session time in doing anything.

6.3.2 Mean idle time for each teacher in the pipeline

In this last analysis idle time is computed as an average value for each teacher to verify if there is any relation with teacher's position. Each of them is obtained as the mean value across repetitions of idle times of the same teacher (identified by its position within the pipeline).

All plotted graphs contain a number of points equals to the number N of teachers in the pipeline, indeed in the x-axis are represented teachers' positions and on y-axis the corresponding idle times. From all the graphs is evident that all the idle times can be considered equals (because of confidence intervals), thus they are independent from the teacher's position in the pipeline.

Idle times are always higher in the lognormal scenario, coherently with students' waiting time behaviour.

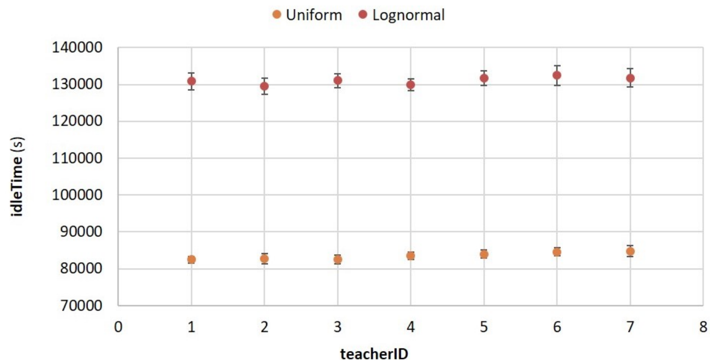

Figure 20: Idle time for each teacher in the pipeline (case N = 7)

7 Conclusions

Examination Time The main metric used for the comparison between parallel and pipelined systems is the *Examination Time* and it turns out to be smaller in the parallel case (here, smaller is better).
In more detail, up to N = 2 pipelined and parallel systems can be considered more or less equivalent for both distributions (indeed, in this case the "gap" (see 6.1) is about 3 minutes and can be considered perfectly negligible).
These gaps between the examination times of the two systems become more and more evident when N increases (e.g. for N = 7 this gap is 12 minutes in the uniform case and 22 in the lognormal one).
A further consideration can be done. This analysis assumes that the time spent by students moving from one teacher to another is equal to 0. Introducing this overhead in a real system, the pipelined case will be even worse because it will have a number of movements that increases with the number of teachers; while, in the parallel case, students will always move only one time (at the beginning of the exam when they sit down) and this system will continue to be better.

Number of examined students In the parallel system the number of examined students does not depend on the number of teachers in the exam committee; instead, in the pipelined case, this number decreases with the increase of N. This means that, in the parallel case, the examination committee can choose the number of questions for each student, as long as it is composed by an equal number of available teachers, and this choice does not affect the number of students examined in the exam session.

Waiting Time In both distributions, the exam experience for each student changes with N. For N = 2 the system offers different treatments: some lucky students do not experience waiting time at all, instead, others have to wait for teachers' availability.
From the students' point of view, system behaviour can be considered convenient for those who do not have to wait, but this condition makes the system unfair for the others.
At the contrary, when N $\geq$ 3 exam experience becomes more homogeneous among the various students, because the majority of them experience the same amount of delay.

Idle Time The pipelined system is not convenient for teachers' productivity; the average values of teachers' idle times can reach the 37% of the total exam session duration.
Fixing the pipeline length, the average idle time is the same for all teachers and therefore does not depend on their position. This is consistent with the fact that parameters chosen for the answer times are the same for all the teachers. Probably, using different parameters for each of them (in order to model the different techniques to examine students) idle times should result different.