

## Estimating the Average Latency of Model Running Time Stimulated by Network Stimulation

### Author's Details:

<sup>(1)</sup>Thi Hoang Yen Tran <sup>(2)</sup>Diep Thang Cao <sup>(3)</sup>Binh Minh Chu <sup>(4)</sup>Van Bang Pham

<sup>(1)</sup> <sup>(2)</sup> <sup>(3)</sup> <sup>(4)</sup>University of Economics - Technology for Industries, Vietnam

Correspondence: Thi Hoang Yen Tran, 456 Minh Khai, Hai Ba Trung, Ha Noi

### Abstract:

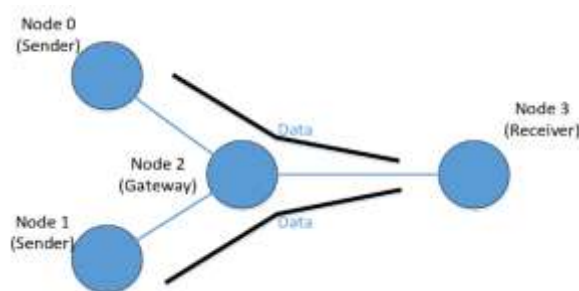
The application of learnt knowledge in past subjects to solve real-life problems has always been an important objective of tertiary education. The present article will introduce some implementations of knowledge that was introduced in Statistics and Probability subject to estimate confidential interval of average latency when stimulating network model using NS2 Network Stimulation application. Mathematical processing and formulae are programmed using Awm programming language and are described by images.

**Keywords:** Confidence Interval, probability, statistics, awk, ns2.

### 1. Introduction

Today, the application of software to build virtual models for modeling real-world models has become popular for economic, intuitive and flexible reasons. In the field of higher education, the simulation of experimental processes, practice, real models on softwares to help students perceive problems visually, improve practical skills such as In reality. Furthermore, this simulation helps educational institutions reduce a large amount of costs for building laboratories.

In recent academic years, the University of Economics-Engineering Industry (University of Economics and Technology) has used many software to simulate real models to give and teach widely to students. In particular, NS2 Network Simulation (NS2) software has been used to model problems in the information technology (IT) major and has obtained some successes. NS2 is an open source software, created for modeling large and complex network systems, making the study and learning of these systems simpler (see [2,3]). . Figure 1 illustrates a simulation of a network topology using NS2.



**Figure 1: A network topology in the NS2 simulation**

When the user applies a simulation model to perform specific calculations, numerical data will be generated. Therefore, it is necessary to have tools to analyze these data to give the necessary information. This is very complicated, it requires an understanding of both the specialized knowledge of the model and knowledge of mathematics, in which, especially the tools for evaluating the accuracy in probability-statistics. . For example, when using NS2 to simulate to evaluate the parameters affecting the quality of network service, it is important to determine the packet transmission delay parameter. The data for each simulation run using NS2 will be saved in an ascii file. Therefore, we need a tool to process the data stored in this file to give the necessary information such as the confidence interval of average latency, the number of simulation runs required to make an estimate for accuracy and confidence. To do these things, the user needs to have some knowledge of statistical

probability and master a programming language to write the processing program using the knowledge of probability. Statistical.

The purpose of this paper is to estimate the average packet transmission latency of the network models simulated by NS2. The data when running these models will be processed by using the knowledge of statistical probability and programming in AWK language to provide the necessary information. The paper consists of four parts, which are presented specifically as follows: Part 2 will introduce some necessary statistical knowledge such as estimation problem, sample size problem. Part 3 presents an overview of the NS2 simulation tool and the AWK analysis language. Part 4 will describe a specific application. Finally, the conclusion will be presented in Section 5.

## 2. BASIC PROCEDURE CONCEPTS

This section will introduce some knowledge of the statistical probability applied to the estimation problem of packet transmission delay.

Set the time for packet transmission delay during a run of the simulation. Then  $E(X) = \mu$  is the average packet transmission delay time during a simulation run. The interval  $(a-\varepsilon, a + \varepsilon)$  is called the confidence interval of  $\mu$  if

$$P(a - \varepsilon < \mu < a + \varepsilon) = 1 - \alpha, \quad (1)$$

where  $1-\alpha$  is called reliability and  $\varepsilon$  is called accuracy.

The meaning of formula (1) is that with  $1-\alpha$  probability, the average packet transmission delay time in a simulation run will be in the range  $(a-\varepsilon, a + \varepsilon)$ . The quality of this estimate is assessed in terms of  $\varepsilon$  and  $1-\alpha$ . The smaller  $\varepsilon$ , the more accurate the estimate and the larger the  $1-\alpha$  (close to 100%), the higher the confidence of the estimate, the lower the probability of a mistake of the estimate. Therefore, it is desirable for each estimation problem to make  $\varepsilon$  small and  $1-\alpha$  large. However,  $\varepsilon$  and  $1-\alpha$  are co-variables, ie want to increase the accuracy of the estimate (decrease  $\varepsilon$ ), it reduces the reliability ( $1-\alpha$  decreases). In fact, depending on the importance of each problem, reliability is always given, usually very large (95%; 99%; ...), then the accuracy of the estimate will be adjusted. by changing the template element number. In the case of  $X \sim N(\mu, \sigma^2)$ , the formula for calculating the probability of being validated in a range is as follows.

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (2)$$

In this article, we are interested in the problem:

With the simulation run times data, estimate the average delay of each packet transmission with  $1-\alpha$  confidence.

To solve this problem, we run the simulation  $n$  times independently and call  $X_i$  the packet transmission delay of the  $i$ th simulation run,  $i = 1, 2, \dots, n$ . When  $n$  is big enough ( $n \geq 30$ ) then according to the Central Limit Theorem, we have

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad (3)$$

with

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (4)$$

is the average transmission delay of the  $n$  simulation runs. The meaning of (3) is that when the number of simulation runs  $n$  is large enough, the average delay time will obey the normal distribution law. The average delay time of  $n$  simulation runs will be expected to be equal to the expectation of the delay time of one run and have a variance of  $n$  times lower than the variance of the delay time of one run.

From (3), we see that  $E(\bar{X}) = \mu$  so  $\bar{X}$  is a non-biased estimate of  $\mu$ . Therefore, we find the confidence interval for the average packet transmission delay time in a simulation of  $\mu$  as  $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$ . The problem is restated in the following format.

With the data of the simulations run and a given  $1 - \alpha$  confidence, find  $\varepsilon$  such that

$$P(\bar{X} - \varepsilon < \mu < \bar{X} + \varepsilon) = 1 - \alpha.$$

From (3) deduced

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Hence, for a given  $1 - \alpha$ , there exists  $u_{\alpha/2}$  such that

$$P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < u_{\alpha/2}\right) = 1 - \alpha$$

Deduced

$$P\left(\bar{X} - \frac{\sigma \cdot u_{\alpha/2}}{\sqrt{n}} < \mu < \bar{X} + \frac{\sigma \cdot u_{\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha.$$

So the accuracy of the required estimate is

$$\varepsilon = \frac{\sigma \cdot u_{\alpha/2}}{\sqrt{n}}$$

and the confidence interval for the average packet transmission delay time during a simulation is

$$\left(\bar{X} - \frac{\sigma \cdot u_{\alpha/2}}{\sqrt{n}}, \bar{X} + \frac{\sigma \cdot u_{\alpha/2}}{\sqrt{n}}\right), \quad (5)$$

where  $u_{\alpha/2}$  is the norm critical value  $\alpha/2$ .

Note. In the case of the unknown population variance  $\sigma^2$ , we use  $S^2$  to estimate no bias for  $\sigma^2$ . Then the confidence interval for the average packet transmission delay time of one simulation in Equation (5) is

$$\left(\bar{X} - \frac{S' \cdot t_{\alpha/2}^{(n-1)}}{\sqrt{n}}, \bar{X} + \frac{S' \cdot t_{\alpha/2}^{(n-1)}}{\sqrt{n}}\right), \quad (6)$$

where  $t_{\alpha/2}^{(n-1)}$  is the critical value student with  $n-1$  degrees of freedom and level  $\alpha/2$ .  $S'^2$  is the corrected sample variance and is calculated by the formula

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (7)$$

### 3. NS2 SOFTWARE AND AWK LANGUAGE

This section will introduce NS2 software to simulate real problems and how to use the AWK programming language in NS2.

#### 3.1. NS2 software [2, 3]

NS2 is an open source software suite developed in 1986 and is popularly used in learning and research to emulate network systems, compatible with the Linux operating system and can be installed via virtual machines on the operating system. Windows operating system. NS can simulate many different types of networks, it simulates the implementation of network protocols such as TCP, UDP; data sources FTP, Telnet, Web, CBR

and VBR; There are router queue management mechanisms such as DropTail, RED and CBR, with route selection algorithms such as Dijkstra ... NS2 also adds multicasting with some MAC layer protocols to support network emulation.

NS2 simulation can create traces in ascii file corresponding to events registered in the network, in Figure 2 is a 12-field structure of the trace file that records event results when performing NS2 simulation.

event	time	from node	to node	pkt type	pkt size	flags	fid	src addr	dst addr	seq num	pkt id
r	:	receive	(at to_node)								
+	:	enqueue	(at queue)					src_addr	:	node.port (3.0)	
-	:	dequeue	(at queue)					dst_addr	:	node.port (0.0)	
d	:	drop	(at queue)								
r	1.3556	3	2	ack	40	-----	1	3.0	0.0	15	201
+	1.3556	2	0	ack	40	-----	1	3.0	0.0	15	201
-	1.3556	2	0	ack	40	-----	1	3.0	0.0	15	201
r	1.35576	0	2	tcp	1000	-----	1	0.0	3.0	29	199
+	1.35576	2	3	tcp	1000	-----	1	0.0	3.0	29	199
d	1.35576	2	3	tcp	1000	-----	1	0.0	3.0	29	199
+	1.356	1	2	cbr	1000	-----	2	1.0	3.1	157	207
-	1.356	1	2	cbr	1000	-----	2	1.0	3.1	157	207

Figure 2. Trace file structure

### 3.2. About AWK [4, 5]:

AWK is a powerful word processing tool created by scientists Alfred Aho, Peter Weinberger, and Brian Kernighan in AT&T Bell Lab since 1977 and is widely used today with many advantages such as: Support for filter filters with regular expressions, working with numbers; supports the ability to split fields on lines of text; supporting programming syntax like C language; get data from: file, redirect, pipe, stdin, ...

The awk utility allows us to perform simple computations on the \*.tr trace file data generated from a simulation such as the average of the column, sum or multiply the terms between several columns. , all data reconstruction tasks, ...

The awk commands are placed in the BEGIN { } { .. } END block as shown in Figure 3.

```

BEGIN {action}

pattern {action}.
pattern {action}

...

END {action}
    
```

Figure 3. Awk file structure

Awk is very powerful when dealing with files that track the results in data simulation. The following program uses awk to calculate the average iteration delay in equation (4).

#### Average.awk

```

BEGIN { FS = "\t" }

      {nl++ }

      {s=s+$4}

END {print "average:" s/nl}
    
```

We can easily execute the awk commands in ubuntu terminal environment with the following syntax:

```
awk -f program_file input_files
```

where program\_file is the program written in Awk and input\_files is the file containing the computation data. For example, if the latency of the simulations is saved in the mp30.dat file, then the average latency is done using the Average.awk file with the following syntax:

```
awk -f Average.awk mp30.dat
```

#### 4. APPLICATION OF RELIABILITY CALCULATION IN SIMULATION

Problem. Using NS2 to simulate the network data transmission model for 10 network nodes and some entity pairs we have the model as shown in Figure 4.

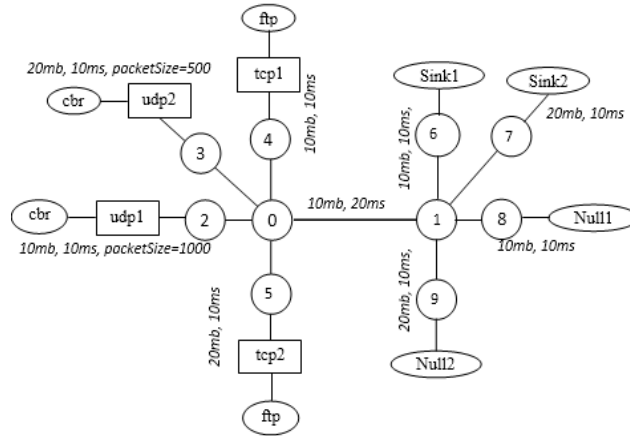


Figure 4. Simulated network structure

Inside:

- \* 0, 1, 2, 3, 4, 5, 6, 8 and 9: are network nodes (nodes);
- \* tcp1 / sink1, tcp2 / sink2: are the send / receive entity pairs of the tcp protocol;
- \* udp1 / null1, udp2 / null2: is the send / receive entity pair of the udp protocol;
- \* ftp: are the traffic sources for file transfer applications - FTP;
- \* cbr: a source of traffic at a constant rate (Constant Bit Rate);

Q: is the queue at the input of link01.

Table 1. Parameters used in the NS2 simulation scenario

ftp1	0.4 s: start, 10.4 s: stop
ftp2	0.6 s: start, 10.6 s: stop
Cbr1	0.8 s: start, 10.8 s: stop
Cbr2	7.0 s: start, 8.0 s: stop, rate = 1,5 Mbps
Tcp1	window_ =32
Tcp2	window_ = 64

Suppose we perform an independent simulation n times (n = 30) and compute the latency times on the cbr application flow. Estimate the confidence interval of mean latency with 95% confidence.

Solution.

Step 1. Run the simulation. With the network simulation model in Figure 4, we follow the diagram in Figure 5, n = 30 times.

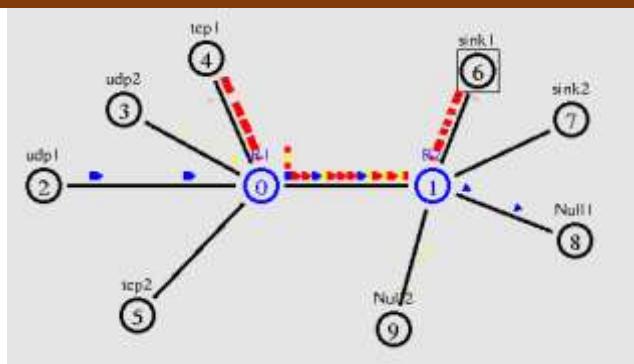


Figure 5. Male visual interface when running the NS2 simulation.

Call the simulation data file *outni.tr* ( $i = 1, \dots, 30$ ). We have file *mp30.dat* which is the file that stores cbr latency average data corresponding to the simulations (see Table 2).

Table 2. Latency of packets in  $n=30$  simulations.

Times	Delay (ms)	Times	Delay (ms)
1	0.032440	16	0.032458
2	0.032436	17	0.032445
3	0.032447	18	0.032443
4	0.032444	19	0.032448
5	0.032485	20	0.032444
6	0.032447	21	0.032440
7	0.032428	22	0.032459
8	0.032441	23	0.032439
9	0.032430	24	0.032441
10	0.032446	25	0.032446
11	0.032448	26	0.032439
12	0.032446	27	0.03245
13	0.032441	28	0.032445
14	0.032485	29	0.032445
15	0.032485	30	0.032449

Using the *xgraph* command to graph the average latency over 30 simulations, the results are shown in the figure.

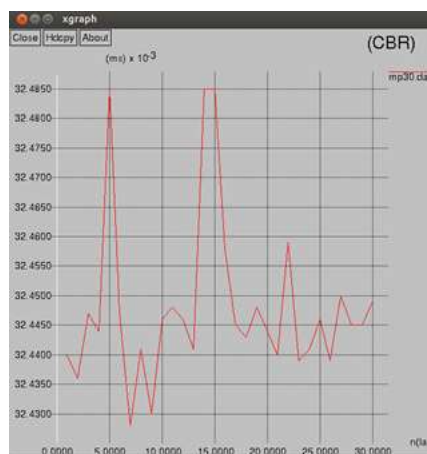


Figure 6. Latency of cbr packets performing 30 simulations

Step 2. Calculate the sample average delay. Let  $\bar{X}$  be the average delay of cbr stream packets corresponding to the simulations.  $\bar{X}$  according to (4) is calculated by calling the program Average.awk

awk -f **Average.awk mp30.dat**

We have  $\bar{X} = 0.032448$  ms

Step 3. Calculate the confidence interval of the average delay.

Program to calculate the confidence interval using Awk. With the number of iterations  $n = 30$  and a confidence level of  $1 - \alpha = 95\%$ , we have

$$t_{\alpha/2}^{(n-1)} = t_{0,025}^{(29)} \approx 2,045.$$

The confidence interval of mean latency in Equation (6) is programmed by awk by the function

**ConfInt.awk**

```
BEGIN { FS = "\t" }
{ ln++ }
{ d = $3 - t }
{ s2 = s2 + d*d }
END {
    s=sqrt(s2/(ln-1));
    print "sample variance: "s
    Conf. Int. 95%: " t "+/-" 2.045*s/sqrt(ln)
}
```

In the terminal call the ConfInt.awk interval calculator

awk -v t=0.032448 -f **ConfInt.awk mp30.dat**

We get results as shown in Figure 7 with sample mean 0.032448 ms and accuracy 0.0118099 ms. So, using the data of 30 simulation runs, with a 95% confidence interval, the average confidence interval of the model is

(0.020639, 0.044259)



```
average:0.032448
chang@ubuntu:~/ns-allinone-2.35/ns-2.35$ awk -v t=0.032448 \
-f ConfInt.awk mp30.dat
sample variance: 0.0330027 Conf. Int. 95%: 0.032448 +/- 0.0118099
```

**Fig. 7. The result uses awk to calculate the confidence interval**

Comment. The calculation steps applied on simulation data analysis to check the calculation of confidence interval according to the theory of statistical probability for the ventricle consistent with the simulation results represented on the graphs.

We can apply the above method to calculate and test other values such as variance, standard deviation, ... to evaluate important parameters affecting network performance when using tissue techniques. simulation NS2.

## 5. CONCLUSION

With the combination of the basic knowledge of the statistical probability module and the NS simulation technique that was taught to the students of UEB, we have built the basis to evaluate the confidence interval for latency. average of simulation runs. The computation is done using Awk programs efficiently. Building models



with NS2, then using the knowledge of statistical probability to analyze and finally apply awk to computation not only trains NS2 simulation skills, awk programming thinking but also helps generate Students deeply understand the statistical probability module, thereby training statistical thinking to apply in other specific problems in practice.

## **REFERENCES**

- i. *Ngo Tu Thanh, "Simulation methods in teaching technical disciplines", Science & Technology Development, Vol 11, No.10, (2008)*
- ii. *Teerawat Issariyakul, Ekram Hossain, "Introduction to Network Simulator NS2", Second Edition, Springer New York, (2016).*
- iii. *Ajit Kumar Nayk, Satyanada Champati rai, Rajit Bemon, "Computer network simulations using NS2" Taylor & Francis, CRC Press, (2016).*
- iv. *Tran Thien Thanh, "Curriculum of statistical probability", Hue University of Science (2011).*
- v. *Le Ba Long, "Lecture on Statistical Probability", Academy of Posts and Telecommunications, (2019).*