Estimation of the run length for simulation of an ATM Switch

Prof.P.K.Suri, Prof.K.D.Sharma and Brijesh Kumar

Dept. of Computer Sc. & Soldicore Inc., New Delhi Applications, KUK
Lingaya's Institute of Mgmt & Tech., Faridabad, India

Summary
The exemplary revolution in the area of computing has led to the adoption of computer simulation, as the most popular tool in testing, and acceptance of new techniques and methodologies, especially in the area of network simulation. Research in the direction of network simulation has not only led to the testing of alternative techniques but also, in the direction of proving the methodology of simulation and proving the credibility of results with certain level of confidence. In the current paper an effort has been done to summarize the measures to be used in performance analysis of an ATM switch and accepting the results with certain confidence interval width.

Keywords:
Performance, Confidence Interval, Transient, Queue, Mean, Variance.

1. Introduction

The performance of networks, especially ATM network mainly depends on the performance of interconnecting switching elements, Channey (1997). An ATM switch can boost or degrade the performance of any network, depending on the way the buffers are managed and the context switching is done for forwarding packets from input ports to output ports. This led to an extensive research in the area of various buffer management techniques, scheduling policies and validating the output of simulation results.

For the simulation of any aspect of an ATM switch, once the model has been identified and the program is ready to run. The source data is generated with the help of well though of probability distribution. The results seem to indicate that if the system design in your program is actually put into practice, it would perform well. The statistical analysis of the output of the simulation experiment is mandatory. Otherwise, *"... computer runs yield a mass of data but this mass may turn into a mess<if the random nature of such output dta is ignored , and then>...instead of an expensive simulation model, a toss of the coin had better be used"* Kleijen(1979)

Many simulations include randomness, which can arise in a variety of ways. This randomness may lead to random results. But these errors may be reduced with the help of some statistical methods. In fact, every simulation is basically a random experiment and the results produced are nothing more than statistical samples. Thus the statistical analysis is an absolute necessarily. However two problems are encountered in the analysis of simulation results; one- the results are highly correlated; second-the source data do not satisfy the requirements of statistical independence.

The simulation of even moderately complex switch, is often computation intensive and may require very long runs in order to obtain reliable final results. Statistical errors associated with the final results of simulations are commonly measured by relative statistical error, defined as the ratio of the half width of the confidence interval(CI) and the point estimate of an analyzed performance measure.

In the current paper section 2 derives the formula for estimation of run length and section Section 3 describes the ways to remove the transients, Section 4 and 5 describes the ways to remove the effect due to autocorrelation of data and the strategy to reduce the its effect so that variance and autocorrelation have minimal effect on the output and the results can be achieved with a given limit of significance. Section 6 includes the summarization and conclusion.

2. Length of Simulation Runs

A simulation run is an uninterrupted recording of a system's behavior under a specified combination of controllable variables. One of the important questions that must be resolved in all simulation experiments involving randomness is how long to run a simulation experiment so that we have a reasonable degree of confidence in the numerical results of the considerably simpler question in classical statistics. Deo(1991) explained the calculation of simulation run length with in a given limit of confidence as

$$n = \frac{(y_{1-\alpha/2})^2 \sigma^2}{t^2} \qquad (2.1)$$

t is the tolerance limit we are willing to accept; σ2 is the variance of the parent population and y1- α/2 is the two tailed standardized normal static for the probability (1-α). Typically the confidence level (1-α) might be 90 percent, for which y1- α/2 =1.65; or it might be 95

Manuscript received December 5, 2007
Manuscript revised December 20, 2007

percent for which y1- $\alpha/2$ = 1.96; or 99 percent , for which y1- $\alpha/2$ = 2.58.

Equation-2.1 is very commonly used for formulae for computing the sample size in statistics. In order to use it in determining the sample size we need to know the variance σ^2 (of the distribution being sampled), which in general is not known in advance. It can, however be estimated as

$$\sigma_{est}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (2.2)$$

where x_i is the avg of parameter being measured during i^{th} run of the simulation program. E.g. in case of an ATM Switch this parameter can be average queue length (AvgQL), average waiting time of pkts in queue before getting forwarded on output port (AvgWT) or average ideal time of the server/processor for waiting for the packets to arrive (AvgIDT). When the estimation 2.2 is used in place of the true population, variance σ^2 , the normalized random variable

$$z = \frac{\bar{x} - \mu}{\sigma_{est}}\sqrt{n} \qquad (2.3)$$

is no longer distributed according to the standardized normal distribution; instead it follows a student-t distribution. However, if n is sufficiently large (>50) the difference between the two distributions becomes negligible. Fortunately, in most simulation experiments (especially network related), the run length is large enough to satisfy this condition.

The equation number 2.1 for determining the run length in a simulation experiment is valid provided the two conditions are met

(i) the distribution is stationary, i.e. the simulation has reached a steady state before we start observing $x_1,x_2,...x_m$ *(independent of initial transients)* and

(ii) the samples $x_1,x_2,...x_m$ are not correlated(that is they are statistically independent).

3. Elimination of Transients

Since the transients are due to initial bias, different initial conditions will produce transients of different lengths and magnitude. Primarily, there are three methods of removing the effect of transients.

(i) Ignore an initial section of the simulation run. The run is started from an empty state and stopped after a certain period (when the system is considered to have settled down to a steady state). The state of the system at that time is left intact. The run is then restarted and statistics gathered up to a certain time from the start. The initial cut-off period, is often decided by making some pilot runs to see how long the initial bias persists.

(ii) Another method of reducing the effect of transients in statistics being gathered is to start the system in an initial state which is close to the steady state. Since the transients are due to the difference between the steady state and the initial state, the smaller the difference , the shorter would be the duration of transients.

(iii) The third strategy , which may be used for reducing the effect of the initial bias is to ensure that the runs have been made long enough that the initial bias becomes negligible.

4. Auto correlated Observations

The use of traditional models in networks characterized by self similar processes can lead to incorrect conclusions about the performance of analyzed networks. Before we can compute the required sample size in such a case, we must first determine the degree to which the data is correlated. In a sequence of observations $x_1,x_2,...x_m$ the extent to which values separated by m units affect each other can be measured by

$$r_m = \frac{1}{n-m}\sum_{i=1}^{n-m}(x_i - \bar{x})(x_{i+m} - \bar{x}) \quad (4.1)$$

where x_i is the ith observation and x is the mean value of x_i's as given by equation 4.1. The quantity rm is called an autocorrelation coefficient with lag m. For the specific case m=0, r_0 is nothing but the estimate σ2est of the variance of the distribution from which $_{xi's}$ are drawn as given in equation 2.2.

Using the equation 4.1, one can compute all of these coefficients $r_1,r_2,\ldots$ In all physical systems as m increases coefficient rm will decrease, because the effect of one value on another becomes weaker as the distance between two observations becomes longer. Thus after a certain number M, these coefficient may be considered to have become zero, i.e.,

$$r_{M+1}=r_{M+2}=0 \quad (4.2)$$

This cut-off point M must be large enough to include coefficients that are significant, but it must be much smaller than n, the sample size. An accepted rule of thumb is to keep M≤ n/10, if each of these r_m's is significantly different from zero. (selecting an exact value of M involves a decision that can be made on the basis of a few trial runs.)

The effect of all nonzero autocorrelation coefficients, are included in the following expression for the estimate of the variance of $\bar{x}$:

$$\sigma_x^2 = \frac{\sigma^2}{n}\left\{1 + 2\sum_{k=1}^{M}(1 - \frac{k}{M+1})r_k\right\} \qquad (4.3)$$

Substituting 4.3 in 8.6 in place of σ^2/n , we get the following expression for the run length n for the auto correlated case:

$$n = \frac{(y_{1-\alpha/2})^2 \sigma^2 \left[1 + 2\sum_{k=1}^{M}\left(1 - \frac{k}{M+1}\right)r_k\right]}{t^2}$$

(4.4)

where the autocorrelation coefficient rk's are given by equation 4.1. Using formula 4.4 instead of 2.1, we can determine the sample size needed in an auto correlated case.

Consider an example. A sequence of 1500 observations was made and found to be serially correlated. The autocorrelation coefficients were estimated using eqn. 4.1 as r1=0.33, r2=0.25 and r3=0.15. Others were not significantly different from zero. The mean (of the 500 samples) was found to be 20.6 and the variance as 1021. The calculation of minimum sample size to assume that the estimate lies within ±2 units of the true mean with confidence level of (1-α)=0.95, can be done as following

$$n = \frac{(1.96)^2(1021)}{2^2}\left\{1 + 2\left[\left(1 - \frac{1}{4}\right)(0.33) + \left(1 - \frac{2}{4}\right)(0.25) + \left(1 - \frac{3}{4}\right)(0.15)\right] - \right\}$$

$$= 980.5684 * 1.82 \cong 1785$$

Thus the existence of autocorrelation made the sample size 19% larger.

5. Blocking Methods

The only difficulty one may encounter in using autocorrelation approach is the amount of computation time required in evaluating autocorrelation coefficients. Several output analysis methods have been proposed, to overcome this limitation. It has been observed that Batch means and probabilities tend to be more nearly normally distributed than the raw outputs(due to central limit theorm). So X_i's can be divided into different kind of batch arrangements.. e.g. NBM(Non overlapping Batch Means Method) by Conway(1963), Overlapping Batch Means (OBM) by Meketon et al(1984), and Standerdized Time series(STS) by Schruben(1983). NBM has some advantages over other output analysis methods. In addition to being easy to understand and easy to implement, batch means can be extended by analogy to estimators other than the sample mean, for example standard deviation, Schmeiser et al, (1990).

In case of NBM, the n observations $x_1, x_2, \ldots x_n$ are grouped into *b* consecutive blocks, each of length p=*n/b*. Then let the block averages be denoted by

$$X_b = \frac{x_{p(b-1)+1} + x_{p(b-1)+2} + \ldots + x_n}{p} \qquad (5.1)$$

Schmeiser(1982), in considering non overlapping batches for confidence intervals on the mean, advocates choosing $10 \leq k \leq 30$, even when the run length n is quite large. This is reasonable (even for the general batch statistics). Song et al(1995) discuss the optimal mean-squared-error(mse) batch size. Goldsman et al (1997) Show that NBM requires O(n) Computation and O(1) storage for any fix p.

6. Summary and Conclusion

The conventional way to measure the quality of simulation is mean square error. Considering this, bias and variance are two ways that a simulation experiment can fail.

Bias can arise from at least six sources. First, the pseudorandom numbers. At best only appear to be independent and uniformly distributed on the unit interval. Second, the distribution of the random variates X can be made to differ from the known input model, often for convenience or some specific prupose. Third, initial transients and stopping rules and bias point estimators. Fourth, some good point estimators are inherently biased, such as using order statistics to estimate quantiles, Fifth, the computer number system is only an approximation to the real number system, e.g. all computations have round off error. Sixth, modeling error, generated because of error in the input model, which is often estimated from real world data or from error in the logical model, which is often intentional to simplify coding. Sensitivity analysis can provide a sense of the effect of the un known modeling error.

The effect of the six sources of bias depends on the run length (i.e. sample size of the output data). The long runs of simulators make this bias error minimum but need extensive sources in terms of space & time and some stopping rule. The equation given in 2.1 can be used to calculate the run length of the simulation experiment initially. If the number of runs comes out to be an achievable figure, the improvised formula using eqn. 4.4 can be used to reduce the effect of autocorrelation. But, initially if the number of samples comes out to be in more than 5-6 digits, blocking methods can be used.

References

[1] Chaney, T., Fingerhunt, Flucke, M., Turner, J.S.,(1997), " Design of gigabit ATM switch", Proc. Of INFOCOM'97, pp.2-11, April'1997.
[2] Conway, R.W.(1963), Some tachtical problems in digital simulation, Management Science6:92-110.
[3] Deo, N. " System Simulation with Digital Computer", " Prentice-Hall of India, New Delhi,1991, ch. Design and Evaluation of Simulation Experiements, pp.144-153.
[4] Goldsman, D.M. and B. Schmeiser(1997), "Computational efficiency of batching methods", Prof. of Winter Simulation Conf. ed. S.Andradottir, K Healy, D.Withers and B. Nelson, pp.202-207
[5] Kleijnen, J.P.C,(1979), " The Role of statistical methodology in simulation" . In Methodology in system modeling and simulation, B.P.Zeigler et al(eds), North-Holland, Amsterdam, 1979.
[6] Meketon, M.S. and B.Schmeiser(1984), Overlapping batch means: something for nothing?", Prof. of Winter Simulation Conf., ed. S.Sherppard. U. Pooch and C.D.Pegden, pp.227-230
[7] Schmeiser, B. (1982), "Batch size effects in the analysis of simulation output", Operations Research, 30:pp.556-568.
[8] Schmeiser, B. T.N. Avramidis and S.Hashem,(1990), Overlapping batch statistics", Proc. Of the Winter Simulation Conf. ed. O.Balci, R.P. Sadowski, and R.E.Nance, pp.395-398.
[9] Schruben, L.W(1983), "Confidenec interval estimation using Standardized time series", Operations Research 31,pp.1090-1108
[10] Song, W.T. and Schmeiser B. (1995), " Optimal mean squared error batch sizes", Management Size, 41, pp.110-123