Do Not Trust All Simulation Studies Of Telecommunication Networks[⋆]

Krzysztof Pawlikowski

Department of Computer Science, University of Canterbury
Christchurch, New Zealand
{k.pawlikowski@cosc.canterbury.ac.nz}

Abstract.
Since the birth of ARPANET and the first commercial applications of computer networks, through explosion of popularity of the Internet and wireless communications, we have witnessed increasing dependence of our civilization on information services of telecommunication networks. Their efficiency and reliability have become critically important for the well-being and prosperity of societies as well as for their security. In this situation, the significance of performance evaluation studies of current and future networks cannot be underestimated. Increasing complexity of networks has resulted in their performance evaluation studies being predominantly conducted by means of stochastic discrete-event simulation. This paper is focused on the issue of credibility of the final results obtained from simulation studies of telecommunication networks. Having discussed the basic conditions of credibility, we will show that, unfortunately, one cannot trust the majority of simulation results published in technical literature. We conclude with general guidelines for resolving this credibility crisis.

1 Introduction

Since the birth of ARPANET and the first commercial applications of computer networks, through explosion of popularity of the Internet and wireless communications, we have witnessed increasing dependence of our civilization on information services of telecommunication networks. Their efficiency and reliability have become critically important for well-being and prosperity of societies as well as for their security. In the United States, the Department of Defense has listed Network Modeling and Simulation as one of the seventeen most important research areas of Information Processing; see www.darpa.mil/ipto/.

Increasing complexity of modern networks has resulted in their performance studies being predominantly conducted by means of computer simulation. A survey of over 2246 research papers on networks published in Proceedings of IEEE INFOCOM (1992-8; in total 1192 papers), IEEE Transactions on Communications (1996-8; in total 657 papers), IEEE/ACM Transactions on Networking

---

[⋆] Invited talk at the opening of ICOIN'2003, 2003 International Conference on Information Networking, Jeju Island, Korea, Feb. 2003

(1996-8; in total 223 papers), and Performance Evaluation Journal (1996-8; in total 174 papers) has shown, see Figure 1, that over 51% of all publications on networks' performance reported results obtained by means of simulation, with the rest of the papers relying on two other paradigms of science: theory and experimentation. Such reliance on simulation studies of telecommunication networks raises the question of credibility of the results they yield.

The main credibility issues of quantitative simulation are discussed in the next section. This is followed by a discussion of the results of a survey conducted for showing how much researchers, who use simulation as the tool of their scientific investigations, are concerned about credibility of the results they produce. The paper concludes with general guidelines for conducting fast and credible simulations of telecommunication networks.

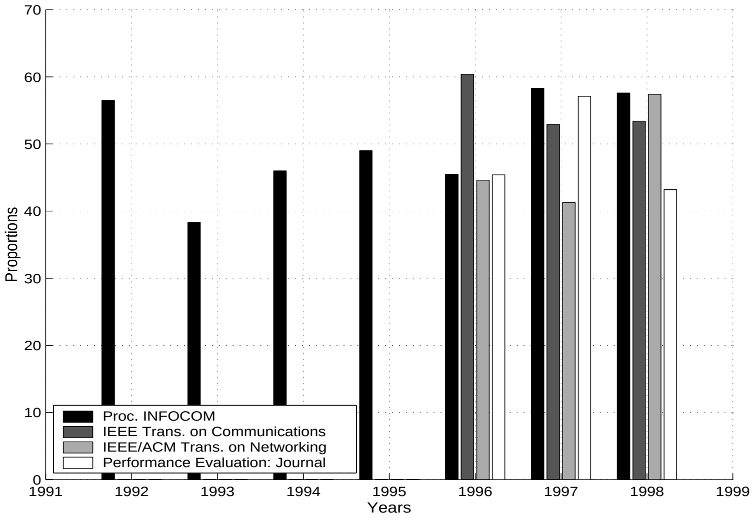


Fig. 1. Proportion of papers that reported results obtained by means of stochastic simulation; from a survey of 2246 papers published in the Proc. IEEE INFOCOM (1992-8), IEEE Trans. on Communications (1996-8), IEEE/ACM Trans. on Networking (1996-8) and Performance Evaluation J. (1996-8).

2 Credibility of Simulation Studies of Networks

The first necessary condition of any trustworthy performance evaluation study based on simulation is to use a *valid simulation model*, with an appropriately chosen level of detail. Some experts assess that the modeling phase of a system for computer simulation consumes about 30-40% of the total effort of a typical simulation study [1]. In the case of telecommunication networks, it means a conceptually correct model of the network, based on correct assumptions about the network's internal mechanisms, their limitations, appropriate characteristics of simulated processes etc. Next, having implemented the model in software, one needs to verify this implementation, to ensure that no logical errors have been introduced. Validation and verification have been generally recognized as important stages of any credible simulation study. A good discussion of general guidelines for correct and efficient execution of these processes in simulation practice can be found, for example, in [2]. However, these are only the first steps for ensuring credibility of the final results of a simulation study, since *"succeeding in simulation requires more than the ability to build useful models ..."*, [3].

The next step is to ensure that the verified software implementation of a given valid simulation model is used in a *valid simulation experiment*. In the overwhelming majority of simulation studies of telecommunication networks, networks or their components are modeled as stochastic dynamic systems. In such a stochastic simulation-based experiment, two primary issues which have to be addressed when trying to ensure its final validity are: (i) application of appropriate source(s) of randomness, and (ii) appropriate analysis of simulation output data. In this updated and extended version of [4], having discussed the last two credibility issues in more detail, we will produce an evidence that one cannot unfortunately trust the majority of simulation studies of telecommunication networks.

It is common practice today to use algorithmic generators of (pseudo-random) uniformly distributed numbers as sources of the basic randomness for stochastic computer simulation. Such a pseudo-random number generator (PRNG) generates numbers in cycles, i.e. having generated whole cycle of numbers, it begins to repeat generation of the same sequence of numbers. Using the same pseudo-random numbers again and again during one simulation is certainly a dangerous procedure since it can introduce unknown and undesirable correlations between various simulated processes. *"Results* <of a stochastic simulation can be very> *misleading when correlations hidden in the random numbers and in the simulated system interfere constructively ..."* [5]. Thus, the practical advice is to use PRNGs that generate numbers in such long cycles that the generated numbers would not be repeated during even the longest simulation. In 2002, using a workstation equipped with a CPU operating at 2.2 GHz, I could generate 10^6 pseudo-random numbers in less than 0.14 second. It meant that, since PRNGs that are still the most frequently used today generate numbers in cycles of length of order 2^{31}, whole cycle of numbers would be generated in about 4.8 minutes. Assuming that the process of random number generation takes, say, 1% of the total simulation time, such a PRNG could be safely used on a 2.2 GHz CPU in a simulation lasting up to about 8 hours.

However, PRNGs with much longer cycles are required if we take seriously the fact that the primary pseudo-random numbers should pass statistical tests for being uniformly distributed. Namely, it can be shown that some tests will always reject the hypothesis about distributional uniformity of pseudo-random numbers if more than a fraction of the cycle is tested. For example, if we are concerned with two-dimensional uniformity of pseudo-random numbers, then only $O(\sqrt[3]{L})$ numbers from a linear congruential PRNG with the cycle of length L can be used. Longer sequences cannot pass a test of uniformity considered in [6]. Empirical analysis of some popular PRNGs reported in [6] has specified that limit as $16\sqrt[3]{L}$. This restricts the number of pseudo-random numbers available from a PRNG with the cycle of $2^{31}-1$ to just about 20 000, and to about 1 000 000 in the case of PRNGs with the cycle of $2^{48}-1$. Assuming as previously that the process of random number generation takes 1% of the total simulation time, a "statistically safe" PRNG, allowing to run a simulation on a 2.2 GHz CPU for up to 8 hours, would need to have the cycle of at least 2^{81} long. Practically during any simulation, one needs two (or more) dimensional pseudo-random vectors that should be uniformly distributed ....

As the computing technology continues advancing according to Moore's law and CPUs operating with clock frequencies well over 2.2 GHz are expected to be commercially available soon[1], we need PRNGs with cycles much longer than 2^{81}, to be able to run simulation experiments over a reasonably long time intervals. Fortunately, such PRNGs have already been proposed. For example, a PRNG known as Mersenne Twister, within a class of Generalized Feedback Shift Register PRNGs, with a super astronomical cycle of $2^{19937}-1$, and good pseudo-randomness in up to 623 dimensions (!) for up to 32-bit accuracy, has been proposed in [7]. Such a PRNG will remain "statistically safe" for any practical simulation experiment executed even on an all-optical computer, a technology that some say can be available is 10 years. And ... its portable implementation in C, on 32-bit machines, is much faster than a standard PRNG used in the ANSI C rand() function[2]; see www.math.keio.ac.jp/matumoto/emt.html for the latest news regarding the Mersenne Twister.

Thus, at this stage, there exist PRNGs that can be used as quite reliable sources of elementary randomness in stochastic simulations. We only need to use them. Unfortunately, uncontrolled distribution of various computer programs has resulted in uncontrolled proliferation of really poor PRNGs, of clearly unsatisfactory or unknown quality. Thus, the advice given by D. E. Knuth in 1969 is even more important today, in the era of Internet: *"... replace the random generators by good ones. Try to avoid being shocked at what you find ..."* [8].

2.1 Analysis of Simulation Output Data

Any stochastic computer simulation, in which random processes are simulated, has to be regarded as a (simulated) statistical experiment and, because of that, application of statistical methods of analysis of (random) simulation output data 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 data is ignored, and then> *... instead of an expensive simulation model, a toss of the coin had better*

---

[1] Written in December 2002

[2] The ANSI C rand() function uses a linear congruential PRNG with modulus of 2^{31}, 1103515245 as the multiplier, and 12345 as the additive constant

be used" [9]. John von Neumann, having noticed a similarity between computer simulators producing random output data and a roulette, coined the term *Monte Carlo simulation.*

Statistical error associated with the final result of any statistical experiment or, in other words, the degree of confidence in the accuracy of a given estimate (point estimate), is commonly measured by the corresponding interval estimate, i.e. by the *confidence interval* (CI) expected to contain an unknown value with the probability known as the *confidence level.* In any correctly implemented simulation, the width of a CI will tend to shrink with the number of collected simulation output data, i.e. with the duration of simulation.

Two different scenarios for determining the duration of stochastic simulation exist. Traditionally, the length of simulation experiment was set as an input to simulation programs. In such *fixed-sample-size scenario*, where the duration of simulation is pre-determined either by the length of the total simulation time or by the number of collected output data, the magnitude of the final statistical error of results is a matter of luck. This is no longer an acceptable approach !

Modern methodology of stochastic simulation offers an attractive alternative solution, known as the *sequential scenario* of simulation or, simply, *sequential simulation.* Today, the sequential scenario is recognized as the only practical approach allowing control of the error of the final results of stochastic simulation, since *"... no procedure in which the run length is fixed before the simulation begins can be relied upon to produce a confidence interval that covers the true* $<$ value $>$ *with the desired probability level"* [2]. Sequential simulation follows a sequence of consecutive checkpoints at which the accuracy of estimates, conveniently measured by the *relative statistical error* (defined as the ratio of the half-width of a given CI and the point estimate), is assessed. The simulation is stopped at a checkpoint at which the relative error of estimates falls bellow an acceptable threshold.

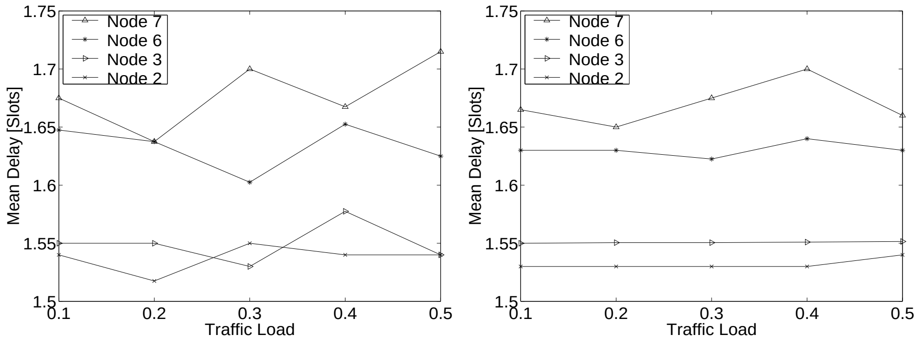
There is no problem with running simulation sequentially if one is interested in performance of a simulated network within a well specified (simulated) time interval; for example, for studying performance of a network during 8 hours of its operation. This is the so-called *terminating* or *finite time horizon simulation.* In our example, one simply needs to repeat the simulation (of the 8 hours of network's operations) an appropriate number of times, using different, statistically independent sequences of pseudo-random numbers in different replications of the simulation. This ensures that the sample of collected output data (one data item per replication) can be regarded as representing independent and identically distributed random variables, and confidence intervals can be calculated using standard, well-known methods of statistics, based on the central limit theorem; see, for example, [2].

When one is interested in studying behavior of networks in steady-state, then the scenario is more complicated. First, since steady-state is theoretically reachable by a network only after an infinitely long period of time, the problem lies in execution of *steady-state simulation* within a finite period of time. Various methods of approaching that problem, mostly in the case of analysis of mean values and quantiles, are discussed for example in [10]. Each of them involves some approximations. Most of them (except the so-called method of regenerative cycles) require that output data collected at the beginning of simulation, during initial warm-up periods, are not used to calculate steady-state estimates. If they are included in further analysis, they can cause a significant bias of the final results. Determination of the lengths of warm-up periods can require quite elaborate statistical techniques. When this is done, one is left with a time series of (heavily) correlated data, and with the problem of estimation of confidence intervals for point estimates obtained from such data. However, although the search for robust techniques of output data analysis for steady state simulation continues ([11]), reasonably satisfactory implementations of basic procedures for calculating steady-state confidence intervals of, for example, mean values and quantiles have been already available; see, for example, [10] and [12].

There are claims that sequential steady-state simulation, and the associated with it problem of analysis of statistical errors, can be avoided by running simulation experiments sufficiently long, to make any influence of the initial states of simulation negligible. While such *brute force approach* to stochastic steady-state simulation can sometimes lead to acceptable results (the author knows researchers who execute their network simulations for a week, or longer, to get the results that, they claim, do represent steady-state behavior of simulated networks), one can still finish with very statistically inaccurate results. It should be remembered that in stochastic discrete-event simulation collecting of sufficiently large sample of output data is more important than simply running the simulation over a long period of time. For example, when analyzing rare events, the time during which the simulated network is "idle", i.e. without recording any event of interest, has no influence on the statistical accuracy of the estimates of the event. What matters is the number of events of interest recorded.

Stopping stochastic simulation too early can give misleading, or at least inconclusive, results. Figure 2 shows the final results from sequential steady-state simulation of a MAC protocol in a unidirectional bus LAN from simulation stopped when the relative error dropped below 15% (Figure 2.a) and 5% (Figure 2.b). Even more significant influence of the level of statistical error on clarity of results can be found in [4]. On the basis of this evidence, one can question the sense of drawing conclusions on the basis of results with high statistical errors, or results for which statistical errors were not measured at all !

Unfortunately, sequential stochastic simulation is still not popular among designers of simulation packages, with overwhelming majority of them advocating analysis of output data only after the simulation is finished. This makes the final statistical errors of results the matter of luck. Very few commercial packages can execute simulations sequentially. Among a few packages designed at universities and offered as freeware for non-profit research activities one should mention Akaroa2 ([14]), designed at the University of Canterbury in Christchurch, New Zealand. Recently, Akaroa2 has been linked with Network Simulator NS2, allowing sequential simulation with models developed in NS2;



(a) Statistical errors of 15% or less (b) Statistical errors of 5% or less

Fig. 2. Example showing influence of statistical errors on the final simulation results. The assumed confidence level=0.95. Evaluation of a MAC protocol in a unidirectional bus LAN considered in [13].

see www.cosc.canterbury.ac.nz/research/RG/net_sim/simulation_group.html for more detail.

3 Credibility Crisis

It would be probably difficult to find a computer scientist or telecommunication engineer today who has not been trained how to assess and minimize errors inevitably associated with statistical inference. Nevertheless, looking at further results of our survey of eight recent proceedings of INFOCOM as well as three recent volumes of IEEE Transactions on Communications, IEEE/ACM Transactions on Networking and Performance Evaluation Journal, one can note, see Figure 3, that, on average, about 77% of authors of simulation-based papers on telecommunication networks were not concerned with the random nature of the results they obtained from their stochastic simulation studies and either reported purely random results or did not care to mention that their final results were outcomes of an appropriate statistical analysis. Let us add that Figure 3 was obtained assuming that even papers simply reporting average results (say, averaged over a number of replications), without any notion of statistical error, were increasing the tally of papers “with statistically analysed results”.

While one can claim that the majority of researchers investigating performance of networks by stochastic simulation simply did not care to mention that their final results were subjected to an appropriate statistical analysis, this is not an acceptable practise. Probably everybody agrees that performance evaluation studies of telecommunication networks should be regarded as a scientific activity in which one tests hypotheses on how these complex systems would work if implemented. However, if this is a scientific activity, then one should follow the *scientific method*, generally accepted methodological principle of modern science,

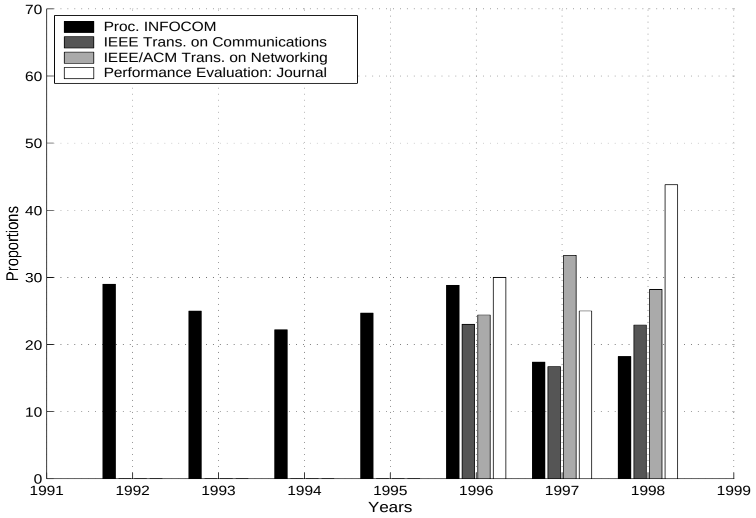


Fig. 3. Proportion of all surveyed papers based on simulation in which results were statistically analysed; from a survey of 2246 papers published in the Proc. IEEE INFOCOM (1992-8), IEEE Trans. on Communications (1996-8), IEEE/ACM Trans. on Networking (1996-8) and Performance Evaluation J. (1996-8).

[15]. This methodology requires that *any scientific activity should be based on controlled and repeatable experiments.*

The real problem is that the vast majority of simulation experiments reported in telecommunication network literature is not repeatable. A typical paper contains very little or no information about how simulation was conducted. Our survey has revealed that authors of almost 77% of papers reporting simulation-based results did not care to inform whether their results came from a terminating or from steady-state simulation.

While the principles of the scientific method are generally observed by researchers in such natural sciences as biology, medicine or physics, this crisis of credibility of scientific outcomes from simulation is not limited to the area of telecommunication networks. It has spanned over whole area of computer science, as well as electronic and computer engineering, despite of such early warnings like that in 1990, by B. Gaither, then the Editor-in-Chief of the ACM Performance Evaluation Review, who, being concerned about the way in which stochastic simulation was used, wrote that he was unaware of *"any other field of engineering or science* < other than computer science and engineering> *where similar liberties are taken with empirical data ..."* [16]. What can be done to change the attitude of writers (who, of course, are also reviewers) of papers reporting simulation studies of telecommunication networks ? Consequences of drawing not fully correct, or false, conclusions about a network performance are potentially huge ...

3.1 A Solution ?

The credibility crisis of simulation studies of telecommunication networks could be resolved if some obvious guidelines for reporting results from simulation studies were adopted. First, the reported simulation experiments should be repeatable. This should mean that information about *the PRNG(s) used during the simulation*, and *the type of simulation*, is provided, either in a given publication or in a technical report cited in the publication. In the case of terminating simulation, its time horizon would need to be specified, of course. The next step would be to specify *the method of analysis of simulation output data*, and *the final statistical errors associated with the results.*

Negligence of proper statistical analysis of simulation output data cannot be justified by the fact that some stochastic simulation studies, in particular those aimed at evaluating simulated systems in their steady-state, might require sophisticated statistical techniques. On the other hand, it is true that in many cases of practical interest, appropriate statistical techniques have not been developed yet. But, if this is the case, then one should not pretend that he/she is conducting a precise quantitative study of performance of a telecommunication network. A more drastic solution of this credibility crisis in the area of computer simulation is to leave computer simulation to accredited specialists [17].

4 Final Comments

We discussed the basic issues related with credibility of simulation studies of telecommunication networks. Then, the results of a survey of recent research publications on performance evaluation of networks were used to show that the majority of results of simulation studies of telecommunication networks published in technical literature unfortunately cannot be classified as credible.

Of course, simulations of telecommunication networks are often computationally intensive and can require long runs in order to obtain results with an acceptably small statistical error. Research on speeding up execution of simulation of telecommunication networks is one of challenging problems which has attracted a considerable scientific interest and effort.

One direction of research activities in this area has been focused on developing methods for concurrent execution of loosely-coupled parts of large simulation models on multi-processor computers, or multiple computers of a network. Sophisticated techniques have been proposed to solve this and related problems. In addition to efficiently managing the execution of large partitioned simulation models, this approach can also offer reasonable speedup of simulation, provided that a given simulation model is sufficiently decomposable.

In the context of stochastic simulation, there is yet another (additional) solution possible for speeding up such simulation. Namely, collecting of output data for sequential analysis can be sped up if the data are produced in parallel, by multiple simulation engines running statistically identical simulation processes. This approach to distributed stochastic simulation, known as Multiple Replications In Parallel (MRIP), has been implemented in Akaroa2 [14], a simulation controller that is offered as a freeware for teaching and non-profit research activities at universities; see www.cosc.canterbury.ac.nz/~krys.

Acknowledgements

The author would like to thank his colleagues from University of Canterbury, New Zealand: Ruth Lee and Joshua Jeong for their involvement in the cited survey of literature, as well as Don McNickle and Greg Ewing, for their contributions in research activities that led to this paper.

References

1. Law, A. M., McComas, M. G.: Secrets of Successful Simulation Studies. Proc. 1991 Winter Simulation Conf., IEEE Press, 1991, 21-27.
2. Law, A. M., Kelton, W. D.: Simulation Modelling and Analysis. McGraw-Hill, New York (1991).
3. Kiviat, P. J.: Simulation, Technology and the Decision Process. ACM Trans. on Modeling and Computer Simulation, **1**, no.2, 1991, 89-98.
4. Pawlikowski, K., Jeong, H.-D. J., Lee, J.-S. R.: On Credibility of Simulation Studies of Telecommunication Networks. IEEE Comms., Jan. 2002, 132-139.
5. Compagner, A.: Operational Conditions for Random-Number Generation. Phys. Review E., **52**, 1995, 5634-5645.
6. L'Ecuyer P.: Software for Uniform Random Number Generation: Distinguishing the Good and the Bad. Proc. 2001 Winter Simulation Conf., IEEE Press, 2001, 95-105.
7. Matsumoto, M., Nishimura, T.: Mersenne Twister: a 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. ACM Trans. on Modeling and Computer Simulation, **8**, no.1, 1998, 3-30.
8. Knuth, D. E.: Art of Programming, Volume 2: Seminumerical Algorithms. Addison-Wesley, Reading (1998).
9. Kleijnen, J. P. C.: The Role of Statistical Methodology in Simulation. In Zeigler B.P. et al. (eds.): Methodology in Systems Modelling and Simulation, North-Holland, Amsterdam (1979).
10. Pawlikowski, K.: Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions. ACM Computing Surveys, **2**, 1990, 23-170.
11. Pawlikowski, K., McNickle, D., Ewing, E.: Coverage of Confidence Intervals in Sequential Steady-State Simulation. Simulation Practice and Theory, **6**, no.2, 1998, 255-267.
12. Raatikainen, K. E. E.: Sequential Procedure for Simultaneous Estimation of Several Percentiles. Trans. Society for Computer Simulation, **1**, 1990, 21-44.
13. Sarkar, N. I.: Probabilistic Scheduling Strategies for Unidirectional Bus Networks. M.Sc. thesis, Computer Science, University of Canterbury, Christchurch, New Zealand, 1996.
14. Ewing, G., K. Pawlikowski and D. McNickle: Akaroa2: Exploiting Network Computing by Distributed Stochastic Simulation. Proc. 1999 European Simulation Multiconf., ISCS, 1999, 175-181.
15. Popper, K. R.: The Logic of Scientific Discovery. Hutchison, London (1972).

16. Gaither, B. Empty Empiricism. ACM Performance Evaluation Review. **18**, 1990, no.2, 2-3.
17. Balci, O, Ormsby, W. F.: Planning for Verification, Validation and Accreditation of Modeling and Simulation Applications. Proc. 2000 Winter Simulation Conf., IEEE Press, 2000, 829-839.