LIMITATIONS OF THEORETICAL AND COMMONLY USED SIMULATION APPROACHES IN CONSIDERING MILITARY QUEUING SYSTEMS

Nebojsa Nikolic
Military Academy, Postgraduate School, Army of Serbia and Montenegro
Ratka Resanovica street, No.1, Belgrade
E-mail: nidzan@ptt.yu

KEYWORDS:
Military queueing systems, Stochastic process simulation, Steady state detection, VV&A&C

ABSTRACT:

This paper is about military queueing systems that are characterized by finiteness, heavy traffic, and even overloading. Queueing theory deals with infiniteness. Simulation methods have serious problems with heavy traffic, and with accuracy and reliability of simulation results. Both do not concern themselves with overloading. Neglecting the fact that military queueing systems performed their missions in a finite time period, and by applying only steady state results, can produce big mistakes in considering their behavior and determining performance measures. Considering above problems leads to a need for an effective solution of the queueing transient phenomenon.

1. INTRODUCTION

Many military situations, processes or systems can be considered as queueing systems. Those can be of various types and sizes, and related to:

- Battle situations;
- Weapon systems and various technical items;
- Military logistic functions;
- Command processes, and so on.

Some very good examples can be found in literature, like [13.], Shephard and others, 1988. Military research studies usually deal with complex situations, but before studying those, are we able to completely solve some relatively simple situations?

Observation on finiteness

Military units in war, are not continually engaged; enemy tanks, rockets and airplanes are not continually in sight through war; one attack or defense operation usually are planed and executed for a limited part of time. Observation about finiteness of reality is keystone that caused all these research efforts. Constellation of applicable queueing theory knowledge (steady state solutions) and observation on finiteness implicate the main hypothesis: “Real system’s behavior, in a finite working time, can differ from its steady state behavior”.

Example

Here, evacuation process of battle damaged tanks in an separate armored brigade will be considered. It is supposed that brigade performs its full battle mission. In general, duration of brigade’s full engagement is limited to about 3 to 5 days; after that time brigade needs some rest. Heavy-damaged tank needs special vehicle-HET (Heavy-Equipment Transporter; this is usually one wheeled wrecker) for its transportation from forward combat zone to rear zone, in order to be repaired. It is supposed that brigade has only one HET.

From practical point of view, relatively simple situation in this example is of triple importance: firstly, service channel represent HET (whose price is high); secondly, clients are heavy-damaged tanks, whose price is also high; thirdly, tanks’ battle importance can be much greater: evacuation, repair, and come-back to the same battle! Example: fantastic score of German maintenance units in battle for Tobruk, North Africa, April 1941, 100 tanks damaged, 88 recovered. Finally, this is approved in many FMs (Field Manuals), such as: the goal is to manage limited resources to return the maximum number of critical items to the battle.

This situation is chosen as an important and concrete enough example, on which the problem stated in title will be demonstrated. Table 1 presents three possible variants of traffic intensities.

Table 1. Input values for queueing system

Variants	Queueing system type is: M/M/1/∞		
	Average times between demands	Average service times	Traffic intensity
1.	240	200	0.833
2.	200	190	0.95
3.	200	240	1.2

Using the language of military reality, solving of this task gives answers on the next questions which can be putt to S4 and/or maintenance officers:
1.) “How much shall I have to wait for my tank to come back repaired?” (battalions’ commanders).

Proceedings 15th European Simulation Symposium
Alexander Verbraeck, Vlatka Hlupic (Eds.)

ISBN 3-936150-28-1 (book) / 3-936150-29-X (CD)

2.) “How many places shall I have to prepare on the collecting point for evacuated items?” (maintenance unit’s commander).
3.) “Do you need support (more HETs) for evacuation in your brigade?” (G4 officer in war; TOE makers in peace (Table of Organization and Equipment)).

2. SOLVING BY QUEUEING THEORY

From theoretical point of view this is a single server queueing system. At the first sight this is very simple case, almost trivial, but this can be true only for its structure (one queue, one server), but not for its behavior (Cohen's “Single server queue“ has about 600 pages!).

Above example can be easily solved using correspondent queueing theory formulas, but they are valid only in case of steady state conditions, and for traffic intensity of up to 1 ($\rho<1$). That means that the task for third variant, cannot be solved, even if that situation is really possible. Table 2 presents a few usually treated measures of performance of such queueing system.

Table 2. Queueing theory steady state results

Variants	Average waiting time in queue $W_q = T_\mu \bullet \rho / (1-\rho)$	Average queue length $L_q = \rho^2/(1-\rho)$	Average server utilization $\rho = \lambda/\mu$
1.	1000	4.17	0.833
2.	3610	18.05	0.95
3.	Inapplicable for $\rho>1$! (that is: ∞, and ∞)		1.2
Queueing system type is: M/M/1/∞			

One of the rare good things in every battle and war as a whole, is the fact that its duration is finite. Like in a sport match, playing and results are only important during the game time; after that it is another story. This fact is taken as one of the crucial moments for studying military queueing systems: their engagement is time limited. Models created for investigating such reality must respect this fact. Usually calculated RESULTS ABOVE ARE QUESTIONABLE, because it cannot be known in advance that our system reached steady state for a finite time engagement (in this example, it is a 5 days).

Queueing theory uses exact mathematical approach, but not for all types and size of queueing systems, and not for all conditions (Larson, Odoni, 1981), [7.], and this is “state of the art” until today. Main reason for this is a simple fact that “queueing theory is hard” (Kleinrock. 1979), [6.], especially if one wants to know more about behavior of queueing system in the period before steady state. Complexity of mathematical analytic approach causes serious difficulties in practical application, even for mathematicians, and even for so-called simpler queueing systems. There is no doubt that dealing with "hard mathematics" takes a care, time and energy of the researcher, and instead of being dedicated to main subject of investigation, he is dedicated to the method.

It can be summarized what the LIMITATIONS are, when QUEUEING THEORY should be applied in solving such real situations:
1.) Treating the whole busy-cycle (transient period and steady state period);
2.) Treating the complex systems (queueing networks);
3.) Treating queuing systems of general type;
4.) Treating overloaded systems (case when $\rho>1$); and
5.) Defining the EFFECTIVE method for solving all above problems; term "effective", here includes: universality, simplicity, reliability, accuracy and cost.

Check-point

In many queueing theory books, problem of practical beginning of steady state was not treated too much, however some results could be found like [10.] by P.Morse, where he suggests (page 67) a very simple formula for relaxation time of queueing system type M/M/1. There is no comment about maximal error when formula is used, but its existence itself can help very much, as it will be shown. For easier application, that formula will be transformed like this:

$$T^* = \frac{T_\mu}{(1-\sqrt{\rho})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1.)$$

where: T^* - relaxation time (steady state beginning);
T_μ - average service time ($T_\mu = 1/\mu$);
ρ - traffic intensity ($\rho = \lambda/\mu$).

By using this formula, steady state practical (approximate) beginning can be easily calculated, for various traffic intensities. Also, it can be expressed in non-dimensional, relative time units ($T^*/\ T_\mu$), so the solution has universal character (it is valid for queueing system where time unit is hour, as well as for another with time units expressed in milliseconds, an so on). Let’s calculate now approximate steady state beginning according to formula (1.), for a set of different traffic intensities. Results are in Table 3.

Table 3. Steady state beginning for type M/M/1/∞

Traffic intensity	Approximate Steady state beginning [expressed in relative units: T_μ]
0.833333	132
0.95	1,560
0.99	39,800
1.2	110

How to use this results: In our example queueing system works for a finite time of 7,200 time units; this value should be divided with T_μ (average service time); if that value is lower than the corresponding one from above table, than our system does not reach steady state during its busy-cycle. If that value is much, much higher, than steady state is practically reached. This is certainly not too much accurate procedure, but it can helps as a good orientation.

For our example it is clear that queueing system, for any variants of traffic intensity does not reached steady state. So, it can be concluded that theoretical solutions (Table 2) are NOT VALID for this example. Our system works all the time only in transient regime. Logical question now, is: if those results are not good, how to get correct solutions? A little poetry can help here:

"Can we wait for steady state,
or we must study the un-steady."

Effective answer will be obtained by another method – simulation modeling.

3. SOLVING BY SIMULATION

Simulation paradigm and crisis

Simulation itself is a phenomenon and deserves a few words more, but not to explore known things, rather specific ones, maybe new:

1.) What is Simulation, is it "art or science?"– Both! It is Science because it must be mathematically founded as a method and verified by the results. It is Art because it include specific "know-how" skills which still can not be exactly expressed: two man painting (modeling, programming!), one is Leonardo, the other can be anybody! Knowledge on Simulation can be presented as vocabulary, but real good doing Simulation is poetry!

2.) Whatever the Simulation is, where it belongs, to what known sub-field of art or science? – Everywhere! There are a lot of various fields of engineering, but other areas too, where simulation takes its place. On the other side, there is no "Faculty of Simulation", or "Simulation Academy", so it is a paradox, but true that simulation is everywhere and nowhere. It might be the destiny, as well as any new discipline– "a new paradigm of science investigation"[15.].

3.) Who deals with simulation? It could be said, simulation community consists of three general groups. First ones- the practitioners are those who should say: What to do (by simulation). Second ones- the mathematicians, they should say: How to do. And third ones- informaticians, they should only: Do it (write the program). One simulation "dream-team", certainly has to include specialists for all three areas. The better case is "dream-team" league, that is a few independent simulation teams, working separately on the same problem. In that case there is a small possibility for monopoly on the science and truth.

4.) Answer the questions linked on: goodness of an simulation model; and accuracy and reliability of simulation results, that is, very shortly- VV&A&C (Validation, Verification, Accreditation, Credibility) questions. One of the first sharp warnings to the simulationists came from B.Gaither [5.], more as an impression but high qualified (he was editor in chief of ACM Performance Evaluation Review): he "...does not know any other field of engineering or Science, where similar liberties are taken with empirical data". This impression is confirmed 10 years later, by an detailed investigation [15.], where was said that more than 70% of simulation papers are of "don't care" type (clearly: don't care on VV&A&C). Their conclusion on this situation was logically marked as Crisis. An aspect of the "Crisis" is mentioned also in papers of M. Neuts [11.]. Thinking about Crisis in simulation, and remembering on Tomas Kuhn's exciting book "*Structure of scientific revolution*", it is logical to conclude that time must come for radical changes in simulation field.

Initial, transient, start-up, ….

Regardless on its generality, previous notations are deeply involved in this investigation. One of the consequences is locating efforts on effective investigation of period before steady state. Depending on point of view, this period is known in a literature as: *initial period; start-up period; transient period; non-stationary period; warm-up period; relaxation time*. It could be a very interesting story about why there are so many names for only one thing, which is, by the way, just known to simulationists and specialized mathematicians! Also, in much literature there is an opinion about fast reaching the steady state, that is neglecting the initial period. Anyway, this is one of the long-lived problems in queueing simulations (from seventies [4.] until today).

Example solving

Simple simulation model (using GPSS) was created for situation described in Example, and in Table 4. are presented simulation results only for first variant of traffic intensities. There is used so-called "one simulation approach", but not "long simulation run", then terminating (fixed time period). Many simulationists, certainly would have objection on this way of solving tasks like this one. But, this was necessary in order to demonstrate inferiority of "one simulation approach". Also, it should be noticed, that this approach, in some local scientific (or "scientific") societies is known as the only method of simulation!

Table 4. shows that for different RNG, results differ from each other, and from theoretical results too. A set

of logical questions arise: Which result is correct? Why do they differ? Why does this happen? The answer is simple: this approach is, principally wrong. Or, to say it in mathematically precise manner: above results are so good, as it can be an estimation of an stochastic variable from sample of size- one element! The ratio of maximal error and confidence level, for "one-element sample size", is entirely un-useable. Arbitrary choosing the RNG which obtains the best results, is not acceptable.

Table 4. One- not long- simulation run solutions for first variant (T_λ=240, and T_μ=200 time units)

type M/M/1/∞	Denotation of used Random Number Generator RNG(i,j))		Average waiting time in queue	Average queue length	Average server utilization
Case	input stream RNG (i)	output stream RNG (j)			
1.	2	6	380	1.66	0.692
2.	3	7	437	1.51	0.740
3.	4	5	300	1.70	0.870
Queueing theory steady state results			*1,000*	*4.17*	*0.833*
NOTES: - simulation run-length: 7200 time units (5 days expressed in minutes)					

One long run results

Commonly used simulation approaches tend to obtain only steady state results. In spite of modeling the reality, the philosophy is to model the useable queueing theory, that is only steady state. There are a lot of sophisticated, but relatively complex statistical procedures, which obtain steady state results.

One of the problems is how to determine the simulation run length. For this example, the simplest procedure is chosen: run-length is increased 10 times, then 100 times, and so on. Also, as we have some theoretical results, let it exclude statistic from first period of length 132* T_μ, and see what happens then. The results are in Table 5.

Table 5. Excluded initial transience, one long simulation run solutions for first variant

Simulation's run-length enlargements	Denotation of used Random Number Generator (RNG(i,j))		Steady state results (excluded statistic from initial period of length: 132*T_μ =132*200 time units=26,400 time units)		
	input stream RNG (i)	Output stream RNG (j)	Average waiting time in queue	Average queue length	Average server utilization
10 times	2	6	456	1.86	0.727
100 times	2	6	743	2.97	0.808
1,000 times	2	6	945	3.92	0.827
Queueing theory steady state results			*1,000*	*4.17*	*0.833*

Again, some poor results are evident. Steady state should begin after 132* T_μ, but results are not good enough. But, it is clear better steady state results for very long simulation run length: the longer- the better.
Let's translate now these modeled situations into the real situations: engagement duration of queueing system, is 50; 500; and 5,000 days respectively! Even entire wars, especially modern ones, do not have such long duration! Also, it cannot be investigated the case when traffic intensity is great than 1 (overloading). For the heavy traffic case ($\rho \rightarrow 1$), it is needed much more increasing of run length.

It can be summarized what the LIMITATIONS are, when COMMONLY SIMULATION APPROACHES are applied in solving such real situations:
1.) Effective treating of the transient period;
2.) Effective treating of the heavy traffic situations;
3.) Investigation of overloaded systems (case: $\rho > 1$); and
4.) Problems of VV&A&C of simulation model.
A little poetry, again (this time by Matthew Arnold, a real poet; taken from Mihram's book [9.]), can describe known simulation, its philosophy and problems:

"We do not what we ought,
What we ought not we do,
And lean upon the thought,
That chance will bring us through."

4. IMPROVED SIMULATION APPROACH

In order to overcome described limitations, an specific simulation approach has been developed, and it is marked here as improved simulation approach. By its nature, it belongs to the statistical methods, exactly it is: simulation modeling of stochastic processes, with implemented possibilities for generating, gathering, displaying and analyzing the statistics of stochastic processes.

Finally, in the Table 6, there are correct results for considered example, for all three cases of traffic intensities. Sample size of 100,000 independent replications of considered situations, obtain high level of reliability and accuracy of simulation results.

Table 6. Improved simulation concept results

- type of queueing system: M/M/1/∞; - battle duration: 5 days;		Solving method	
		Queueing theory (only steady state)	Simulation's results (sample size: 100,000)
Variant 1 (T_μ=200 min., ρ=0.833)	Average waiting time in queue	1000	411
	Average queue length	4.17	1.82
	Probability of server idle time	0.17	0.27
Variant 2 (T_μ=190 min., ρ=0.95)	Average waiting time in queue	3610	526
	Average queue length	18.05	2.81
	Probability of server idle time	0.05	0.20
Variant 3 (T_μ=200 min., ρ=1.2)	Average waiting time in queue	(∞)	854
	Average queue length	(∞)	4.66
	Probability of server idle time	(0)	0.12

Some experience from real situations can be useful in understanding presented results: "Our service channel is really heavy-loaded and we work almost all the time, but the queue is not so long, nor waiting is so great!". But, these results are so much different, and maybe surprising, that method which generates them, deserve a few words more. Or, to say it sharply: what are the guaranties, that method is good.

Verification of the method

Concept of the method completely corresponds to statistic of stochastic processes. Steady state detection is in a complete accordance with corresponding theoretical approximations. Performance measures of queueing system, obtain by simulation, have also good agreement with theoretical ones.

Primary performance measures. From basic queueing theory postulates, it can be raised that states probabilities are performance measures of highest importance, and here, they are marked as primary measures of performances. All other performance measures (queue length, waiting time, ...), depend on system states probabilities, and because of that, they can be called secondary measures of performance. But in practice, one wants to know just some of the secondary measures, and this fact is probably one of those which caused that primary measures are not studied too much. Another fact, maybe more important, is problem of creating one effective method for generating (solving, getting) states probabilities as time-dependent variables.

Comparison. Here, main idea for this problem is next: generate states probabilities as time-dependent variables. Then compare simulation results with those ones obtained by "stronger" (analytic or numerical mathematic) method. It should be clear, this is possible only for relatively simple queueing systems which can be solved completely by "stronger" method. One simple queueing system of type M/M/1/7, is considered for a finite time engagement (6,000 time units). Average service time is 100, and average time between clients in input stream is 120 time units. This system is solved by simulation and by numerical mathematics. Results (states probabilities) are presented on Figure 1.

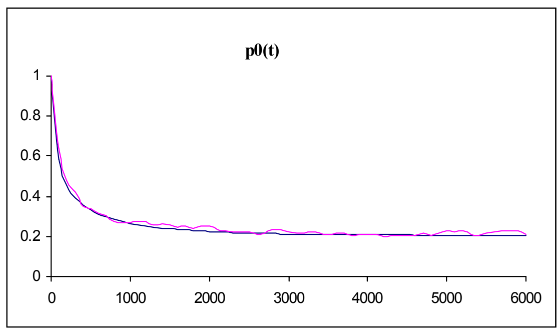

Figure 1. Comparison of numerical and simulated $p_0(t)$

Quantification of differences. Besides the fact that good concordance is clear, it is possible to apply appropriate statistical tests to quantify differences. Simulation results based on sample size of 1075 elements (independent replications of stochastic process). There is only one system state probability presented, but the others are similar. For this sample size: maximal possible error of simulations results is 11.6% for confidence level of 95% (confidence coefficient for this confidence level is: z_c=1.96) for $p_0(t)$. And, this can be tested by chi-square test.

How much (sample size) is enough. Created simulation approach obtains full controllability of ratio: desired accuracy of simulation results, and corresponding level of statistical confidence, for defined sample size. To be concrete, if we want to determine how large should the sample size be, for specified maximal error (distinction) and for desired level of confidence, in proportions (probabilities) estimation process, than the next formula can be used:

$$N = \frac{q}{p}\left(\frac{100}{\varepsilon}\right)^2 z_c^2 \quad \text{................} \quad (2.)$$

where: N - sample size;
p - proportion (probability) to estimate;
q - complement to proportion (q = 1 - p);
ε - percentage maximal error of estimation;
z_c – confidence coefficient.

Above formula can be generated from elementary statistical claims explained in basic courses for statistics, like in [14.]. As the matter of fact, this form is very rare in literature. Some authors even give wrong formulas, or conclude that it is not possible to determine the exact sample size.

How small (probabilities) are enough. Importance of considering rare events, grows in cases of heavy traffic. Then the question arises: how small probabilities are enough to consider? For answering this question, exceptional book [2.] can be of great help. The book suggests next scale for orders of magnitudes: human (10^{-6}), earthly (10^{-15}), cosmic (10^{-50}), universal (10^{-1000}). For the purpose of queueing research, only first level is enough. Exactly, it is enough to consider set of system states probabilities, which consists 99.99 % of possible system's states. States' probabilities inside this border, are up to 10^{-6} order of magnitude, for one of the worst cases: 0.99 traffic intensity for M/M/1 queueing system.

5.Conclusions

Trustworthy modeling of military queueing systems declares a set of specific demands (finiteness, heavy traffic, over-loading), to the generally used methods of solving queueing problems. Those demands could not be satisfied by known methods, so the new method is created. One of the central problems was effective studying of queueing system's transient phenomenon.

In the sense of initial goals, whole study produced a set of collateral effects, all of which are very positive and important. Famous, long-lived problems in queueing simulations: start-up problem, or steady state detection problem, accuracy of simulation results, variance reduction, reliability of simulation results, and heavy traffic situations, can be easily solved using this suggested method.

REFERENCES

[1.] Cohen J.W., "Single server queue", Nort Holand, Amsterdam, 1969.

[2.] De Finetti B., "Theory of probability: A critical introductory treatment", John Wiley&Sons, London, 1974.

[3.] Gross D., Harris M.C., "Fundamentals of queueing theory", John Wiley & Sons, New York, 1974.

[4.] Gafarian A.V., Ancker C.J., Morisaku T., "Evaluation of commonly used rules for detecting "steady state" in computer simulation", NAVAL RESEARCH LOGISTICS, 1978. page 511-529

[5.] Gaither B., "Empty empiricism", ACM Performance Evaluation Review, Vol.18, #2, august 1990. page 2-3

[6.] Kleinrock L., "Power and deterministic rules of thumb for probabilistic problems in computer communications", Conference Record, *International Conference on Communications*, Boston, Massachusetts, pp. 43.1.1 to 43.1.10, June 1979.

[7.] Larson R., Odoni A., "Urban Operations Research", Prentice Hall, 1981.

[8.] Law A., Kelton D., "Simulation modeling and analysis", McGraw Hill, New York, 1982.

[9.] Mihram G.A., "Simulation: statistical foundations and methodology", Academic Press, 1972. New York

[10.] Morse P.M., "Queues, inventories and maintenance", John Wiley & Sons, New York, 1958.

[11.] Neuts M., "Probability modeling in the computer age", International Conference on Stochastic and Numerical Modeling and Applications, January 1997

[12.] Odoni A., Ruth E., "An Empirical Investigation of the Transient Behavior of Stationary Queueing Systems", Operations Research, Vol.31, No3, May-Jun 1983. page 432-455

[13.] Shephard R.W., and others, "Applied operations research - Examples from defense assessment", Plenum Press, New York, 1988.

[14.] Spiegel M.R., "Theory and problems of statistics", McGraw Hill, New York, 1961

[15.] Pawlikowski K., Jeong H.D.J., Ruth Lee J.S., "On credibility of simulation studies of telecommunication networks", IEEE Communications Magazine, January 2002., page 132-139

AUTHOR BIOGRAPHIES

NEBOJSA NIKOLIC is an assistant at Military Academy Postgraduate School, Army of Serbia and Montenegro. He graduated in 1988., at Military Technical Academy (Zagreb, ex-Yugoslavia), as the best student in generation. First 9 years in troops service, on various duties: from platoon and company commander, to S3, S1 and S4 officer. In year 2000. got Master of Science title in military-technical sciences. Last 7 years, interest fields: military logistic, simulation modeling, statistics; logistic in foreign armed forces. Current occupation: preparing PhD thesis.