Proceedings of the 1984
Winter Simulation Conference
S. Sheppard, U. Pooch, D. Pegden (eds.)

PERTURBATION ANALYSIS METHODS FOR DISCRETE EVENT DYNAMICAL SYSTEMS AND SIMULATIONS

Y.C. Ho
Division of Applied Sciences
Harvard University
Cambridge, MA 02138

ABSTRACT

This paper introduces a new powerful analysis and optimization technique for discrete event dynamic systems, such as, air traffic control systems, discrete part manufacturing processes. It can also be used for enhancement of the simulation results of or the monitoring of the operations of such systems in real time.

I. INTRRDUCTION

It is certainly not necessary to argue the importance of simulation to the study of discrete event dynamic systems. The continued development of new simulation languages and the publication of books in this area are eloquent testimonies to the vitality of the subject matter. Traditionally, simulations like all experiments are used to collect statistics which in term re used to predict system behavior under well defined environments. The problem with simulation, as with all numerical efforts, is the relative lack of insight generated with numerical results. Brute force parametric analysis using simulation is very difficult and expensive . These limitations are well known.

II. WHAT IS PERTURBATION ANALYSIS?

Perturbation Analysis is a newly developed analysis technique that can be used to enhance simulation results. By doing some analysis on the outputs of a simulation, one can literally squeeze ouyt a lot more information than traditionally thought possible. The quickest way to explain what perturbation analysis can do is by way of a paradigm. Consider a DEDS as illustrated in Fig.1.

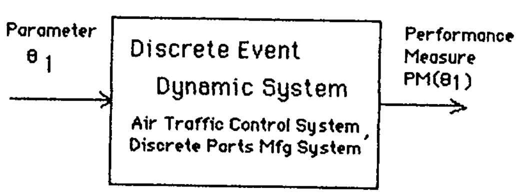

Fig.1

The DEDS has certain parameter, θ_1 , which is a parameter of interest; it also has an output which is usually a performance measure PM, such as, the throughput. We observe the behavior and the PM of the DEDS over an period of time, say, [0,T]. Now imagine that we have an identical twin to the DEDS in Fig.1 in the sense that this twin duplicates EXACTLY the behavior of the original DEDS down to the last detail in lock-step timing (In the case of simulation, we can create this twin by using a second identical computer running the same simulation program with the same random seeds.). However, for this identical twin DEDS we are endowed with the additional capability of being able to modify the input parameter θ_1 by a small amount $\Delta\theta_1$. Clearly, the output or PM of the twin DEDS in this case will in general become different from the original DEDS by some (usually small) amount. We denote the perturbed PM as PM($\theta_1 + \Delta\theta_1$) and computed the sensitivity as

$$\frac{PM(\theta_1 + \Delta\theta_1) - PM(\theta_1)}{\Delta\theta_1} = \text{sensitivity of PM to } \theta_1 \triangleq s_1$$

This is illustrated in Fig. 2.

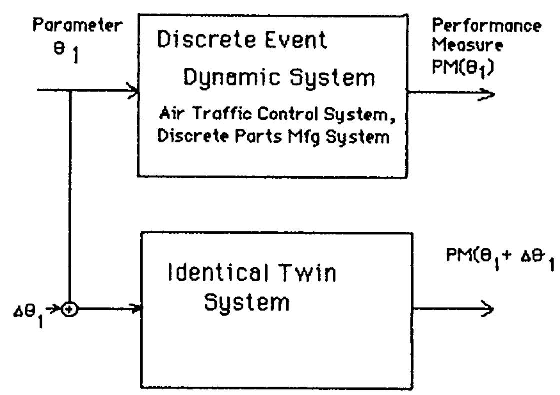

Fig.2

Note becuase of the presence of the twin DEDS, this sensitivity is available continuously and in real time . Finally, imagine that instead of a single twin DEDS, we have identical triplets, quadruplets, in fact, as many identical DEDS as there are parameters of interest. In other words we can compute in real time s_1 , s_2 ,..., s_M , or the gradient or sensitivity vector $\partial PM/\partial\theta = [\ \partial PM/\partial\theta_1, \ldots, \partial PM/\partial\theta_M\ ]$. One cannot emphasis too strongly the significance of this capability for the fining tuning of the design of or the operations management of a DEDS. Many heretofore unsolved problems in DEDS will now be amenable for iterative solution with this capability . The catch here is that ordinarily the cost of such twin systems is prohibitive.

PERTURBATION ANALYSIS IS A TECHNIQUE WHICH IN EFFECT SAYS THAT ONE CAN HAVE THIS CAPABILITY OF PRODUCING GRADIENT INFORMATION OR SENSITIVITY VECTOR IN REAL TIME WITHOUT THE NECESSITY OF DUPLICATING THE DEDS M TIMES PROVIDED ONE IS WILLING TO DO SOME RELATIVELY SIMPLE PROCESSING OF THE SYSTEM BEHAVIOR AS IT EVOLVES.

That the above "almost free lunch" is indeed possible is the power of perturbation analysis as developed by the Harvard group during the past years.

Of course, aside from a demand for proof of the above assertion (see below) alert readers at this point may point out that this "free lunch" is not very appetizing since all it does is to answer the "what if" question of what would have happened during the interval [0,T] if the parameter had been 0+ 0. What the user wants is to be able to use the sensitivity information to predict what will happen in the next observation interval [T, 2T]. The answer to this objection is that

(i) the objection is not an attack on the validaty of perturbation analysis.
(ii) there are well known statistical answers to this question of prediction.

Every experimenter has faced the problem of using the result of experiments to predict the outcome of "similar" future experiments. If you observe over a long series of tosses that a coin is biased 60% for "heads", then one is justified to think that one can give 6:4 odds on the average involving a large number of bets using the coin. On the other hand if he observed the coin turn up heads on the 536th toss in a series of 1000 tosses, it will be foolhardy to bet that the same thing will repeat on the next 1000 tosses Just like statistical averaging justifies our first assertion, statistical variation will completely invalidate the second assertion. In other words, perturbation analysis simply gives you certain information, how you use it is an entirely seperate issue. More importantly , however, recent results in fact show that perturbation analysis method can yield strongly consistent estimates [refs. 21,22] under well defined conditions and has statistical advantges in estimation compared to conventional methods [ref.20].

Historically, the new approach started as an "ad hoc method" to solve a particularly long standing problem in production line management and design [refs.1,2]. Gradually, it was realized that the "ad hoc method" is related to some fundamental ideas of dynamical systems [ref. 3], and that in fact it can be applied to general queueing networks when viewed as stochastic dynamical systems. The analytical and experiemental results obtained in late 1981 in [ref. 4] in recognition of this generalization opened the way to systematic study of this approach as applied to DEDS [refs 5-13, 18-20]. Since most of the details, proofs, and experimental results already exist in the open literature, we refer the interested readers to the appropriate references. Breifly at this point, the technique has been rigorously estblished for a class of DEDS and their perturbations; and shown experimentally to work for a much larger class of DEDS. Although rigorous proof does not exist for this larger class, an understanding of the basic idea behind the technique will persuade most reader of the inrinsic generality of the approach. Important further work awaits interested researchers.

III. APPLICATIONS TO SIMULATION.

Perturbation analysis is noninvasive and complementary to any existing simulation effort . Whatever improvements ,such as variance reduction, made to a simulation will be equally applicable to our perturbation analysis; the computational advantage of N:1 remains constant. Thus, the principles of perturbation analysis are independent of issues of statistical significance and experimental design. In fact, ref [20] points out that perturbation analysis is statistically advantageous in addition to being computationally superior.

The most obvious applications of this method of analysis of DEDS is in performance evaluation or optimization. Most performance measure (PM) of a DEDS can be expressed as a function of Throughput (TP), Sojourn Time (ST), and/or Queue Length (QL), the last quantity being related to the first two via Little's Law (ST*TP=QL). Consequently, the gradient of PM with respect to system parameter, $\partial PM/\partial\theta$, can be calculated as a linear combination of the gradients $\partial TM/\partial\theta$ and $\partial ST/\partial\theta$, both of which we can compute efficiently [refs.4,13]. The parameters that have been considered so far are mean service times, queue sizes, no. of customers which in manufacutering technology translates to machine speeds and failure/repair probability, buffer limits, and no. of pallets; while in computer performance evaluation they may mean, disk speed, memory size, and degree of concurrent programming.

With this gradient information, we can use any one of the numerous optimization techniques for the successive improvement of the PM. Again the question of which one of the several techniques for stochastic optimization should one adopt for a given problem is an issue that can be separately addressed. More generally, perturbation analysis can be carried out on a server basis [ref.4] or on a customer basis [ref.13]. The coupling between the two is determined by the obvious fact that service initiation/completion and customer arrival/departure are essentially equivalent events with the possible addition of waiting or blocking times. Since customer and servers are elementary building blocks of queueing systems, [refs.4 and 13] essentially demonstrate the ability to perform in principle all queueing system analysis.

Another unique application of this approach is the on-line monitoring of systems operations. Since we can calculate sensitivities of PM with respect to system parameters in real time, we can provide valuable sensitivity information concerning the system for operations management. This feature is reminiscent of "neighboring optimal control" or "extended Kalman filtering" ideas in control theory where corrective action for a system can be taken based on observed or estimated perturbation about a nominal trajectory. This is a unique feature of perturbation analysis in that it can be used in real time to analysis real system data.

IV. CONCLUSION

Perturbation Analysis is still in its infancy. We believe it is based on a sound and general idea very different from existing approaches. Experimental and theoretical work done so far justify our beliefs. Much more work remains to be done.

V. ACKNOWLEDGEMENT

The work reported in this paper was supported in part by ONR Contract N00014-75-C-0648 and N00014-79-C-0776 and NSF grant NSF-ECS-82-13680.

VI. REFERENCES

1. Y. C. Ho, M. A. Eyler, and T. T. Chien, A Gradient Techniques for General Buffer Storage Design in a Serial Production Line , Int. J. on Production Research, Vol. 17, No. 6, 1979, pp557-580

2. Y. C. Ho, M. A. Eyler, and T. T. Chien, A New Approach to Determine Parameter Sensitivities on Transfer Lines , Management Science, Vol 29, #6, June 1983, pp700-714

3. Y. C. Ho and C. Cassandras Computing Co-state Variables for Discrete Event Systems Proc. of 1980 IEEE Conf. on Decision and Control, Dec. 1980

4. Y. C. Ho and Xiren Cao, Perturbation Analysis and Optimization of Queueing Networks , Journal of Optimization Theory and Applications, Vol 40, #4, August 1983, pp559-582.

5. Y. C. Ho and Christos Cassandras, A New Approach to the Analysis of Discrete Event Dynamic Systems , AUTOMATICA, Vol 19, #2, pp 149-167, 1983.

6. R. Suri Implementation of Sensitivity Calculation on a Monte Carlo Experiment , J. of Optimization Theory and Applications, Vol 40, #4, August 1983.

7. Y. C. Ho Parametric Sensitivity of a Statistical Experiment , IEEE Trans. on Automatic Control, Vol. AC-24, #6, p982 Dec. 1979

8. Y. C. Ho, X. Cao, and C. Cassandras Finite and Infinitesimal Perturbation Analysis of Queueing Networks , AUTOMATICA, Vol 19, #4, pp 439-445, July 1983

9. R. Suri, and X. Cao The Phantom Customer and Marked Customer Methods for Optimization of Multiclass Queueing Networks with Blocking and General Service Times , ACM Performance Evaluation Review, Vol 12, August 1983

10. Y. C. Ho, C. Cassandras, and R. Suri Stochastic Similarity and Statistical Linearity in Perturbation Analysis of Discrete Event Systems , chapter 10 of [ref.18]

11. C. Cassandras and Y. C. Ho An Event Domain Formalism for Sample Path Perturbation Analysis of Discrete Event Dynamic System submitted for review to IEEE Trans. on Auto. Control

12. R. Suri Infinitesimal Perturbation Analysis of Discrete Event Dynamic Systems: A General Theory IEEE Decision and Control Conference, San Antonio, Texas, Dec. 1983

13. Xiren Cao and Y. C. Ho Perturbation Analysis of Sojourn Time in Queueing Networks , IEEE Decision and control Conference, San Antonio, Texas, Dec. 1983

14. K. M. Chandy, U. Herzog, and L. Woo Parametric Analysis of Queueing Networks , IBM J. of Research and Development, Jan. 1975

15. Y.C. Ho Editorial-Is It Application or Experimental Science? , IEEE Trans. on Automatic Control, Vol. AC-27, #6, p.1142 , Dec. 1982

16.J. Kimemia, S. Gershwin, and D. Bertsakis, Computation of Production Control Policies by Dynamic Programming Technique , Proc. of 1982 IEEE Large Scale Symposium, Virginia Beach, Va. Oct. 1982 IEEE Catalog #82CH1741-8

17. S. Stidham, Jr. Optimal Control of Arrival to Queues and Network of Queues Proc. of IEEE Decision and Control conference, Florida, Dec. 1982

18. Y. C. Ho (Ed) SPEEDS - A New Techniques for the Analysis and Optimization of Queueing Networks Technical Report #675 , Division of Applied Sciences, Harvard University, Feb. 1983

19. V. J. Cogliano, P. Sanchez, and L. Schruben "Simulation Sensitivity Analysis: A Frequencyn Domain Approach for Discrete Parameters" Presented at the 1983 ORSA-TIMS Meeting, Orlando, Florida.

20. KUMAR, P.R. " On Estimating Derivatives of Cost Functions from Simulated or Real Data ", IEEE Conference on Decision and Control, Las Vegas, 1984

21. R. Suri, M. Zazanis "Perturbation Analysis Gives Strongly Consistent Estimates for the M/G/1 Queue" Proc. of 1984 IEEE Decision and Control Conference, Las Vegas.

22. X. Cao, "Convergence of Parameter Estimates in a Stochastic Experiment", Proc. of IEEE Decision and Control Conference, Las Vegas