Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2015) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

[S1]The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015)

Cost-Efficient Risk Management with Reserve Repair Crew Planning in CLOUD Computing

Mehmet Sahinoglu[a,*], Sharmila Ashokan[a], Preethi Vasudev[a]

[a*]*Informatics Institute, Auburn University Montgomery (AUM), P.O. Box 144023, Montgomery AL 36124-4023 USA*

Abstract

This article addresses a discrete event CLOUD simulator, namely CLOURAM (CLOUD Risk Assessor and Manager) to estimate the risk indices in large CLOUD computing environments, comparing favorably with the intractably theoretical Markov solutions or hand calculations that are limited in scope. The goal is to optimize the quality of a CLOUD operation and what countermeasures to take to minimize threats to the service quality by reserve planning of reserve crew members. Cost and benefit analysis is examined after the solutions.

Peer-review under responsibility of organizing committee of The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015).

Keywords—CLOUD Computing; discrete event simulation; reserve planning; repair (maintenance) crew; cost and benefit; breakeven.

1. Introduction

CLOUD computing, a relatively new form of computing using services provided through the largest network (Internet) has become a promising and lucrative alternative to traditional in-house IT computing services, and provides computing resources (software and hardware) on-demand. Why big firms delay using CLOUD? The response lies in the common belief that companies are hesitant to switch to CLOUD for applications that manage key parts of their businesses due to lack of trust as they believe there is still too much risk associated with potentially unreliable Internet connections and dependence on third parties to manage computer servers (See Wall Street Journal, July 17, 2014; Marketplace: “Why Big Firms Delay Using Cloud” by S. Norton and C. Boulton). Consequently, a quantitative assessment of the quality of service (QoS) in such enterprises is critically needed. The quality of CLOUD computing services can be difficult to measure, not only qualitatively but most importantly

1877-0509
Peer-review under responsibility of organizing committee of The 2015 International Conference on Soft Computing and Software Engineering (SCSE 2015).

quantitatively. An algorithmic discrete event simulation accompanied by related statistical inference is conducted to estimate the availability indices in a CLOUD computing environment of small or large service-based systems to mimic real life operations. However, as users (companies, organizations and individual persons) turn to CLOUD computing services for their businesses and commercial operations, there is a growing concern from the security, privacy and availability perspectives as to how those services actually rate. Moreover, the federal government has approved commercial products to operate on a defense CLOUD, marking the industry's first online offerings with this level of security accessible to the military via such an environment [1].

If the reserve capacity (or margin) is less than a zero margin, then we have an undesired deficiency or loss of service. Note that LOLP: Loss of Load Probability. Once these hours (or cycles) of negative margin are added, it will constitute the expected number of hours of loss of load, LOLE. Divided by the total number of exposure units such as 8760 hours (NHRS) for a year, it will give the LOLP = LOLE/NHRS. Once the LOLE is known, and its frequency (f=number of such occurrences of deficiencies per annum), then the average duration, d=LOLE/f, will indicate how long on the average a loss of service is expected to last. What are some of most significant scenarios on "What If" will be studies in subsequent sections, such as what happens when the cloud managers reduce the size of the maintenance crew and at what level should they stop discounting the crew size for best ROI (Return on Investment).

2. CLOUD Resource Management Planning for Employment of Repair (Maintenance) Crews

A most popular example to a what-if query as frequently executed in simulation engineering practices is the resource allocation, which is one of the most vulnerable and softest (weakest) points of the entire CLOUD computing process. We will study the effect of the number of maintenance crews from full to lower. Figure 1 below in the Appendix has originally displayed an unreliability index of 5.44% for 348 units.

The total number of production units is 348 available. In a new analysis, we will simulate (10000 runs or years) for a reduced resource of 100 crews to see the impact. That is how much less reliability we will have to suffice with if we save money by eliminating 248 repair (recovery) crews. This time we are expecting wait times and more unreliability from our CLOUD operation. Now see Figure 2 below in the Appendix. The unavailability index is unfavorably upped to 7.36%, a negligible difference when you take into account the savings you will accrue from employing 248 less repair crews. See Figure 3 below in the Appendix. We now will reduce even further 50 more repair crews, 298 less than originally assumed. Finally so, we will employ solely 50 crews to see if it is economical profitable to do so in Figure 3 below in the Appendix. This time we hit the rock! Outcomes are disastrous. We saw the catastrophic results of a skyrocketing unreliability of 85.71% (while employing only 50 repair crews) from merely 7.36% when we had only 50 more crews. The difference is the breaking point [2]. Therefore this CLOUD management should not lower their repair crews to less than around 100. More detailed studies can be conducted by trying 99….90….70….60 etc. to see the drastic jump. This cost-benefit portfolio of crew-planning analysis as one of the most crucial "What-If" scenarios could save millions of dollars for a CLOUD Resource Management to plan ahead.

3. Step by Step Algorithm CLOUD Management Planning for Employment of Repair Crews

The following software program (CLOUD Management for the manager part of the CLOURAM to follow up with the assessment part) will show you how to do this in a systematic and algorithmic order. Here we will study how to implement the effect of maintenance crews on unavailability index. In this analysis we will simulate for a reduced resource of Crew Intervals *(for example 50, 100, 150 …).* Consider an example, where the total number of production units available is 443. Now let us simulate for the reduced resource of 50 crews, as in Figure 3 below in the Appendix to see the impact that how unavailability will vary. See Figure 4 below in the Appendix for the input dialogue box. Initially the unavailability varies slightly, at one point it increases in a drastic manner which is called breaking point. The CLOUD should not lower the repair crews to less than the breaking point. We can observe, as in Figure 5 below in the Appendix, the crew break-even point is 143. This cost-benefit portfolio of Reserve Crew Planning analysis will save millions of dollars' worth for a CLOUD Resource Management. In Environment

Parameters Section, three buttons are added which indicates type of execution that is Normal (only risk assessment without management), Crew Planning and Product Planning.

First begin with selecting one among the three types of execution. Normal button is selected to implement a normal execution of Cloud. Crew planning button is selected to implement Reserve Crew Planning and to obtain Crew Plot so as to estimate the optimal crew value to halt at for ensuring certain availability (lack of risk). Product planning button is further selected to implement Reserve Product Planning and to obtain Product Plot so as to find the optimal capacity value to halt at so to ensure a desired availability. Follow the below algorithmic steps to implement Maintenance Crew Reserve Planning:

Step 1: Radio button **Crew Planning** must be selected as shown in Figure 4 below in the Appendix.

Step 2: **# of Crew Increments** in default is given as 10, which indicates the number of crew intervals required to plot a graph.

Step 3: **Crew Intervals** in default is given as 50, which indicates the difference between two crew (working personnel employed) values.

Step 4: **ΔLOLP Crew** in default is given as 0.3, which indicates that if the difference between the two consecutive LOLP values is greater than 0.3, then the crew addition stops at the lowest LOLP value.

Step 5: **Starts at Crew** is the total number of maintenance crews initially, for data2000.txt there are 443 crews. *Step 6*: **Starts at LOLP Value** is the LOLP value for the initial number of maintenance crews.

Step 7: **Stops at Crew** is the optimal stopping crew value.

Step 8: **Stops at LOLP Value** is the optimal stopping LOLP value.

Step 9: **Final Crew Value** is the number of crews that are remaining at the end of the crew planning implementation.

Step 10: **Final LOLP Value** is the LOLP value for number of crews that are remaining at the end of the crew planning implementation.

Step 11: **Cost** in default given as $1,000,000.00 for a placeholder, which indicates the dollar amount loss to the entire Cloud operation annually as accrued by one percent increase in Loss of Load resulting from sparing the determined number of extra crew members.

Step 12: **Benefit** in default given as $30,000.00, which indicates the dollar amount of investment spent for each new crew member employed per year, hence amount saved when a crew is released.

Now let us see how to implement the effect of maintenance crews on unavailability index. In this analysis we will simulate for a reduced resource of Crew Intervals *(for example 50, 100, 150 …).* Consider a large cyber CLOUD example, where the total number of available units served by the same number of serving crew units available is 443. Now let us simulate for the reduced resource of 50 crews, as in Figure 5 below in the Appendix to see the impact that how unavailability will vary with the reduction of crew. Initially the unavailability varies slightly; at one point it increases in a drastic manner which is called break-even point. The CLOUD should not lower the repair crews to less than the break-even point. Figure 6 below in the Appendix shows the output for the input dialogue box in Figure 5 below in the Appendix for which using data 2000.txt, the stopping crew value is 143.

Profit/Loss: indicates whether there is profit or loss by stopping at the optimal point. If cost of the LOLP increasing is greater than benefit of sparing (saving) repair crew members, then it is Loss (negative value). If cost is less than benefit then it is Profit (positive value).

Break Even Value: indicates the cost per percent for LOLP index required to have neither profit nor loss.

*Step 13: **Solution*** is as follows. Now 300 x $30K = $9,000,000 gained. ΔLOLP= 0.0126 lost, i.e., 0.0126 x 100% x $1M= $1,260,000. $9M-$1.26 = $7.74M or (+) $7,739,029 exact.

Figure 5's cost-benefit portfolio of Reserve Crew Planning analysis will show to save or lose millions of dollars' worth for a CLOUD Resource Management. It also indicates what the break-even value would be if RHS=LHS in terms of cost and benefit so that the difference (profit or loss) is zero at balance. See Figure 6 below in Appendix for plotting the process. Further, Figure 7 below in Appendix depicts the triplet of operating modes (up, down, waiting) where additionally waiting time is now a reality due to imperfect number of repair crews [3].

4. Some Further Validations of Reserve Crew Planning

When stopping crew value and stopping LOLP value is 0, a message indicates to simulate again with different ΔLOLP crew constraint as red-flagged in Figure 8 shown below in Appendix. This message that the analyst needs to enter a new set of feasible input values which is also shown in Figure 9 in Appendix.

5. Discussions and Conclusion

The CLOUD computing as the new century's paradigm change of doing business at a larger scale of computing supported by the internet of all things, must be measured, monitored and managed to see where things are leading to rather than letting it to take charge by the oft-practiced cruise control alternative. The crucial problem with CLOUD computing is its occasional, though dramatic lack of desired availability and security [4]. Once assessed, the next question becomes as to how to manage, rather mitigate the lack of availability. One of the most predominant measures to do that is the leverage of maintenance crews. Where do we halt recruiting the maintenance the repair crews and what is the breakeven value of the number of crews that need to be kept ready for service to monitor the desired availability at best? Once can certainly not achieve this cost optimization by sheer guessing or hand calculating. This research shows through the related software how best to achieve cost optimal reserve planning for the maintenance crews, a process which can cost CLOUD managing or facilitating companies millions of dollars if not rightfully estimated in advance [5].

Appendix:

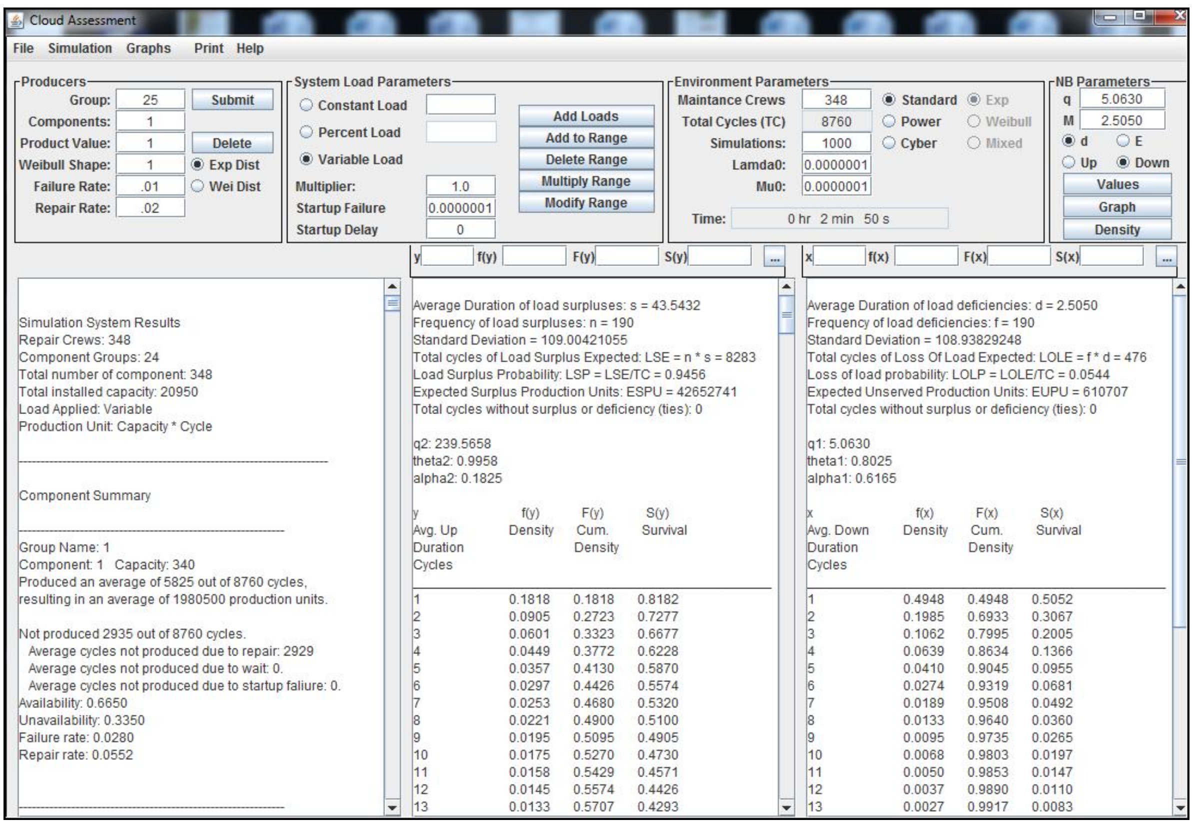

Figure 1. The simulation results (1,000 years) for a large CLOUD with 348 units. Run time: 2 min 5s.

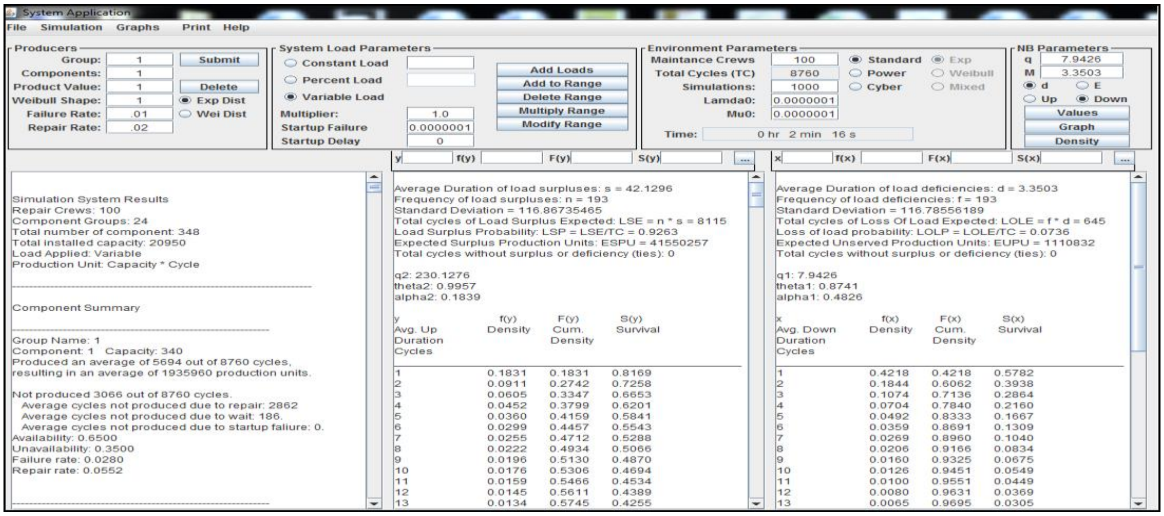

Figure 2. The simulation results (1,000 years) for a large CLOUD with 348 units for 100 repair crews.

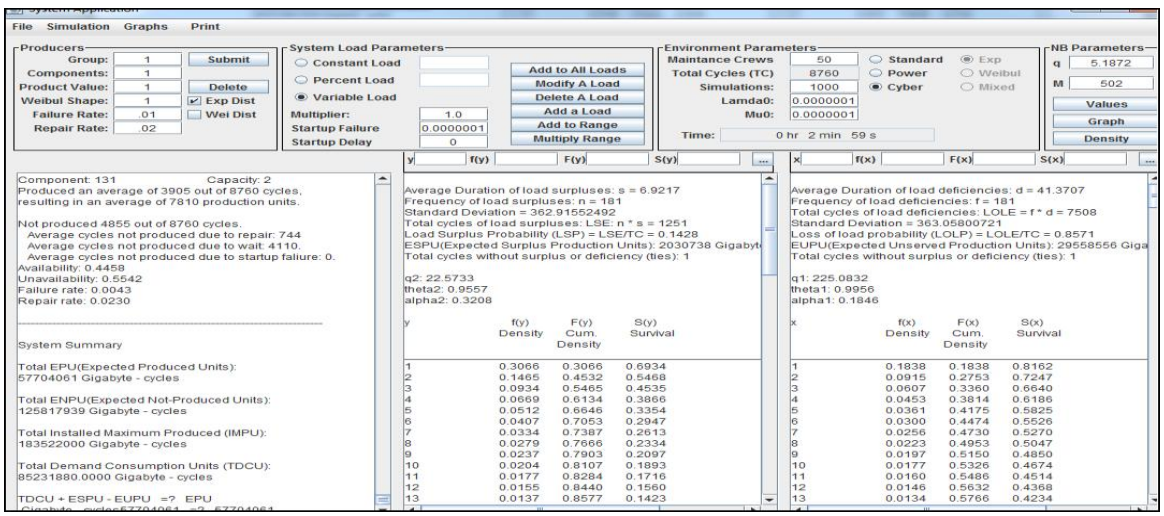

Figure 3. The simulation results (1,000 years) for a large CLOUD with 398 units for 50 repair crews.

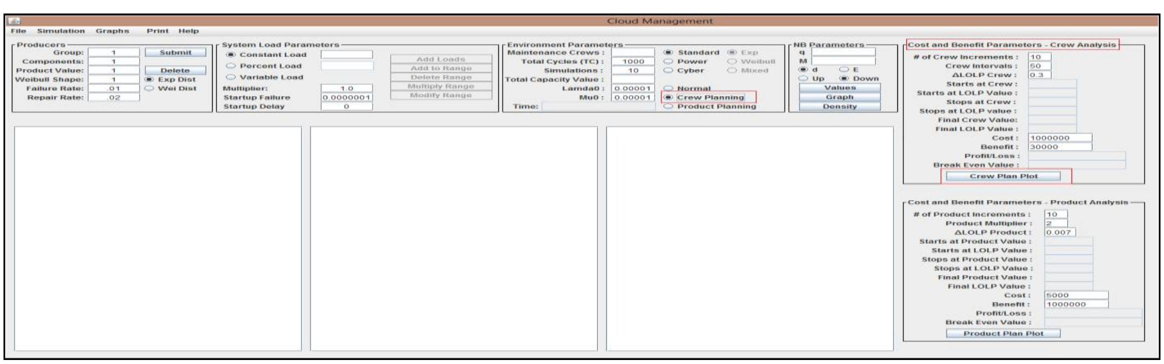

Figure 4. The dialogue box input data example for specifically crew reserve planning .

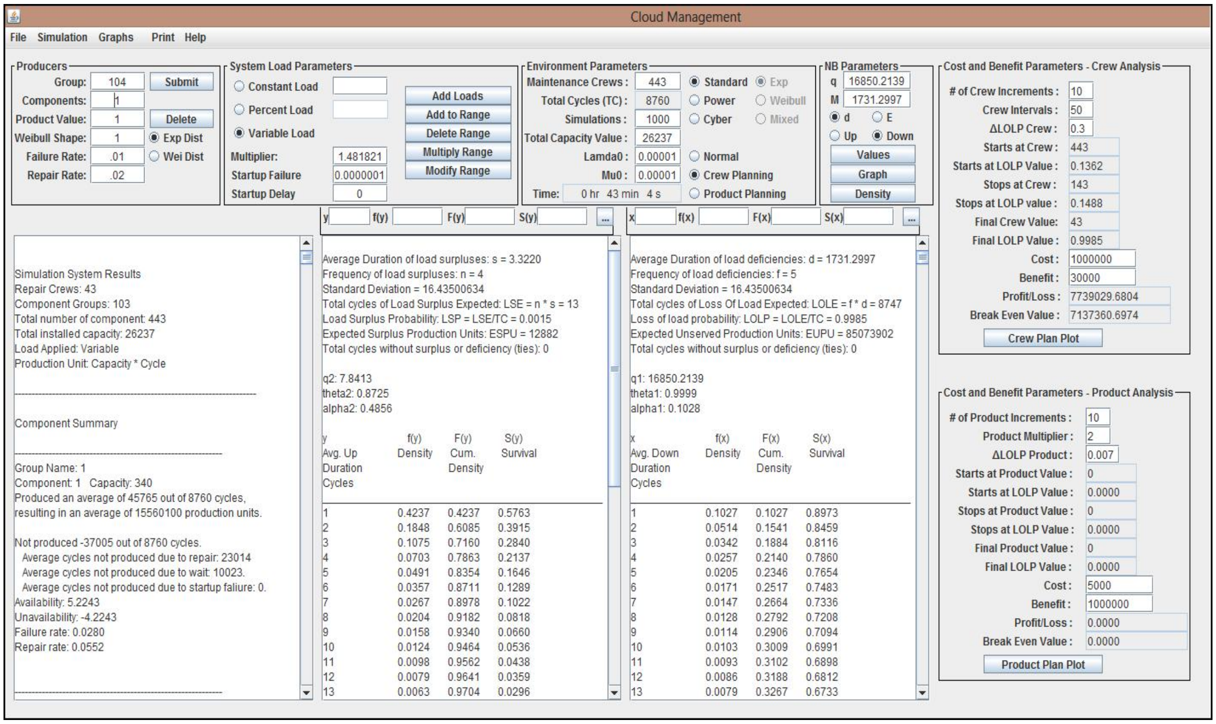

Figure 5. Output dialogue box for Input Data in Figure 4 for 1000 simulations with 443 repair crews.

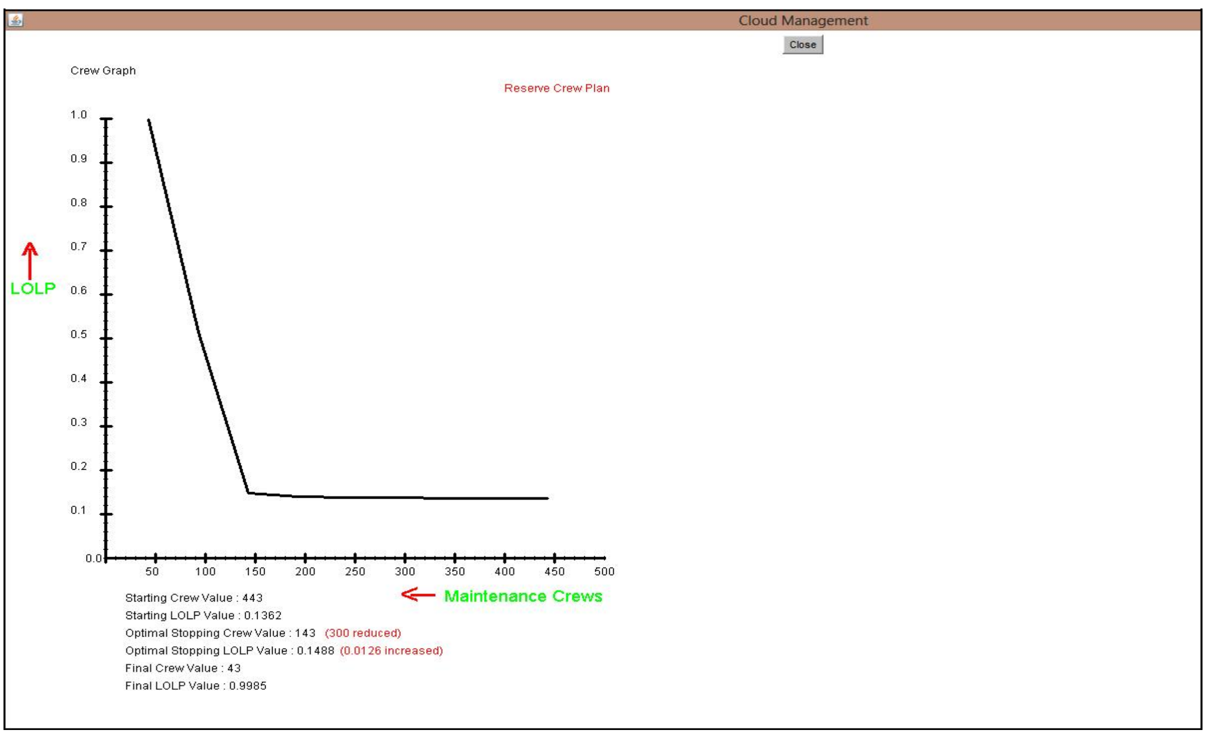

Figure 6. Output Graph in response to Input of Figure 4 for 1000 simulations with 443 repair crews.

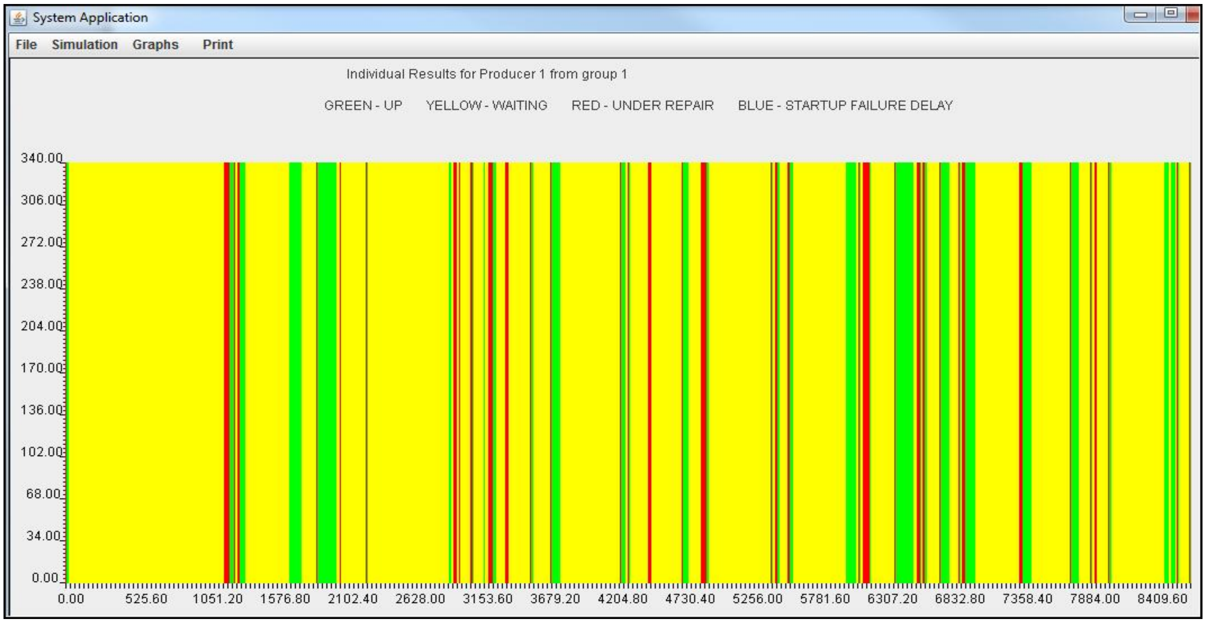

Figure 7. The sequences of up (green), down (red) and yellow (waiting) for Group1's Unit1 history.

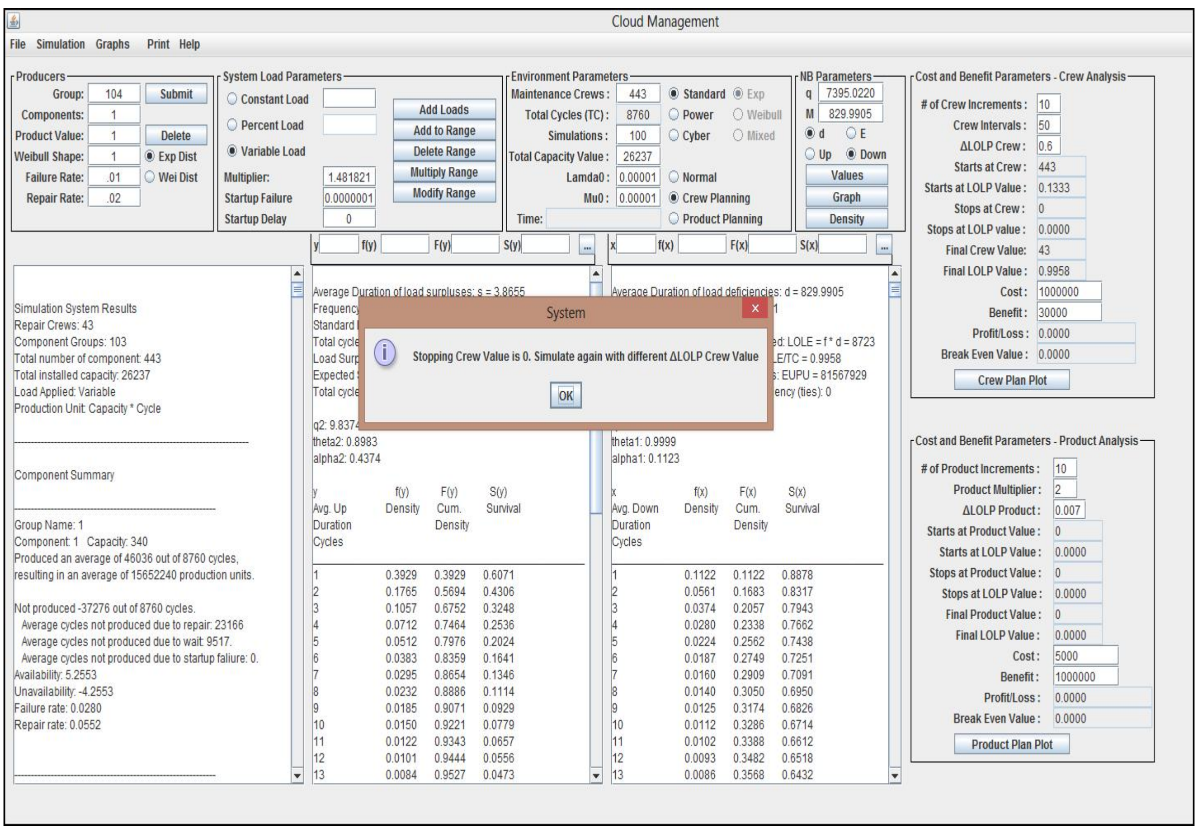

Figure 8. Output dialogue box for Figure 4's input for 100 runs when a new set of input for "ΔLOLP Crew" required as above.

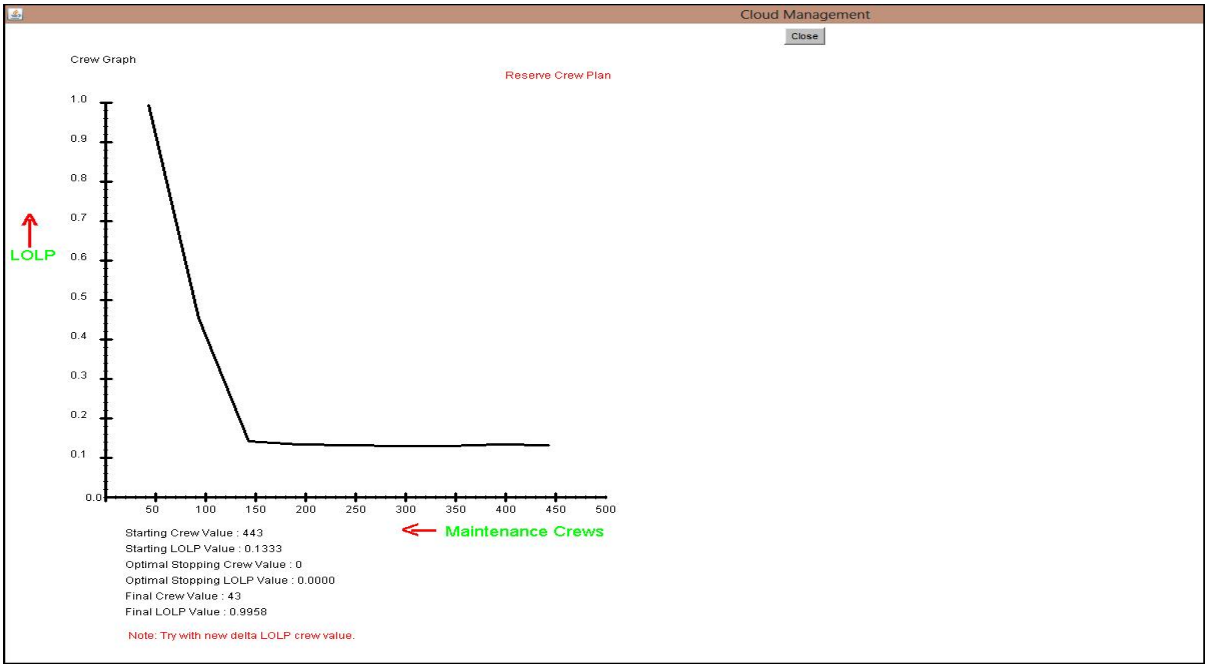

Figure 9. Plot of the LOLP (risk) vs. # Maintenance Crews requiring a new set of feasible input values.

References

[1] N. Leavitt, “Is Cloud Computing Really Ready for Prime Time”, *IEEE Computer*, January issue, 15-20, 2009.

[2] M. Sahinoglu, L. Cueva-Parra, “CLOUD Computing,” Invited Author (Advanced Review) for *Wiley Interdisciplinary Reviews: Computational Statistics*,

New Jersey, Ed.-in-Chief: E. Wegman, Yasmin H. Said, D. W. Scott, Vol. 3, Number 1, pp. 47-68, March 2011. http://authorservices.wiley.com/bauthor/onlineLibraryTPS.asp, DOI=10.1002/wics.139&ArticleID=771921

[3] M. Sahinoglu, CLOUD Computing Risk Assessment and Management, Book Chapter, (Risk Assessment and Management) Academy Publish, November http://www.academypublish.org/book/show/title/risk-assessment-and-management 2012

[4] G. Worthen; J. Vascellaro, Wall Street Journal, E Mail Glitch Shows Pitfalls of Online Software – Photo: Services like Gmail run on vast computer farms.

A Google center in Lenoir, N.C., B 4-5 Media and Marketing, February 26, 200

[5] M. Sahinoglu, S. Morton; Cloud Risk-O-Meter: An Algorithm for Cloud Risk Assessment and Management," Conference of Society of Design and

Process Science (SDPS), Session X1 Cloud Computing: Security and Reliability, 1:30-3:15pm, Berlin, Germany, 13 June 2012.