Assessing the Performance of Interdependent Infrastructures and Optimizing Investments

Linda K. Nozick and Mark A. Turnquist
School of Civil and Environmental Engineering
Cornell University
Hollister Hall
Ithaca, NY 14853
emails: lkn3@cornell.edu and mat14@cornell.edu

Dean A. Jones, Jennifer R. Davis and Craig R. Lawton
Sandia National Laboratories
Albuquerque, NM 87185
emails: dajones@sandia.gov, jbechd@sandia.gov and crlawto@sandia.gov

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Abstract

Our nation's security as well as the quality of life of its citizenry depends on the continuous reliable operation of a collection of complicated interdependent infrastructures including transportation, electric power, oil, gas, telecommunications and emergency services. A disruption in one infrastructure can quickly and significantly impact another, causing ripples across the nation. Our infrastructures are increasingly reliant on new information technologies and the Internet to operate, often being connected to one another via electronic, informational links. While these technologies allow for enormous gains in efficiency, they also create new vulnerabilities. The focus of this paper is the development of a unifying mathematical framework to represent these "mega infrastructures" and a collection of algorithms that can be used to estimate performance and optimize investment. We include a small computational example that focuses on the delivery of gas and electric services, including the underlying SCADA system that supports the gas network, to illustrate the operation of the algorithms.

1. Introduction

Our nation's security as well as the quality of life of its citizenry depends on the continuous reliable operation of a collection of complicated interdependent infrastructures including transportation, electric power, oil, gas, telecommunications and emergency services. A disruption in one infrastructure can quickly and significantly impact another, causing ripples across the nation. Our infrastructures are increasingly reliant on new information technologies and the Internet to operate, often being connected to one another via electronic, informational links. While these technologies allow for enormous gains in efficiency, they also create new vulnerabilities. The same technology that allows us to transmit information around the globe at the click of a mouse can be used to disrupt our vital systems including the flow of electric power or water, and the dispatch of emergency services.

Rinaldi *et al*. [9] describe infrastructures as complex adaptive systems (CASs) and create a conceptual framework for addressing infrastructure interdependencies. They identify six dimensions that can be used to describe infrastructure interdependencies, and apply this framework to analyze the role other infrastructures play in supporting the operation of the electric power system.

In this paper, we develop a mathematical framework to represent interconnected infrastructure networks and a collection of algorithms that can be used to estimate performance and optimize investment. This framework treats uncertainty in the capacity of links in the networks as a key characteristic and uses measures like those suggested by Aven [1,2] to reflect network performance. Algorithms based on both simulation and analytic methods relate the evolving condition of individual components and subsystems (including cyber services) to the ability of the interconnected infrastructures to deliver services to customers. This structure is then extended to answer questions related to optimal investments by optimizing the choices for the stochastic processes that describe the evolution of component and subsystem conditions.

2. Mathematical Representation for Interdependent Infrastructures

We begin by asserting that we can represent interconnected infrastructures through the use of networks (graphs of nodes and arcs). The arcs represent components or subsystems in an infrastructure or the connection between infrastructures. The arcs have capacities and these capacities may be uncertain and evolve over time. Further, the capability of one component or subsystem may be correlated with the performance of several other components or subsystems. Correlations stem from a number of causes. For instance, links in similar geographic locations are likely to suffer similar fates, either as a result of contamination from one infrastructure to another or because they fall victim to the same event. Also, the topology of some networks implies that when certain links experience diminished capacity all downstream links encounter the same or more severe difficulties.

Nodes and links can also represent the information infrastructure. The state of the information system links affects the capacity of the links they support. For example, when the signaling system in the telephone system experiences difficulties, calls may be dropped, reducing capacity in the circuit switched network.

To illustrate these ideas consider a small example of two interconnected infrastructures – a natural gas distribution network and an electricity generation/distribution network. The gas distribution network is supported by a supervisory control and data acquisition (SCADA) system. This example has been created to be small enough to make it easy to understand but complex enough to contain illustrations of the types of relationships that would be found in much larger real networks.

The combined gas and electric network is shown in Figure 1. It contains two separate suppliers, labeled *S1* and *S2*. The gas that enters the distribution network from supplier *S1* enters via node *DS1*. Similarly, the gas that enters the distribution network from supplier *S2* enters via node *DS2*. There are four distinct demands for gas served by this network, two of which are electric power generation stations (*E1* and *E2*). Either generating station can serve the electric load at *L2* but only one of the generators can serve the electric load at *L1*. By separating the generating facilities into two nodes and a connecting arc (e.g., *E1* and *G1*) we can represent a "node failure" (partial or complete loss of a generator) as a capacity loss on the connecting arc. Values of demands for gas and electricity (per period) are noted at nodes *D1*, *D2*, *L1* and *L2*. The numbers alongside links in the network represent the nominal capacities of those links.

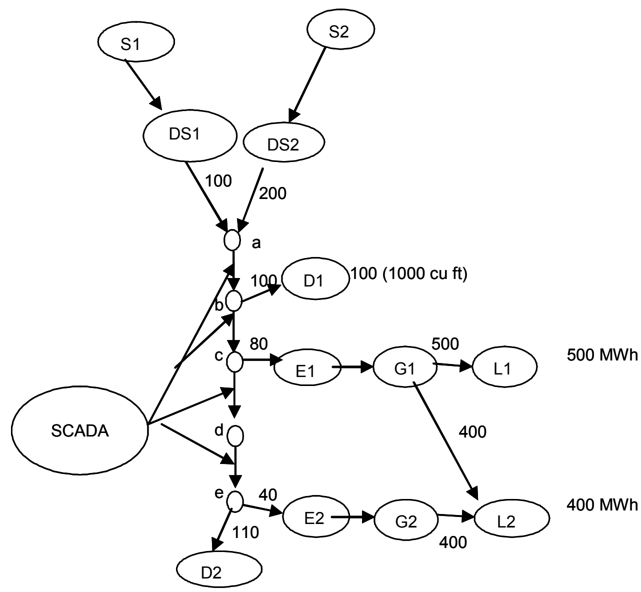

Figure 1. An example network for interdependent gas and electric infrastructures.

In joining these two networks together, we have also created a transformation of the "commodity" flowing through the network at the *E1* and *E2* nodes, where gas is transformed into electricity. For this example, we will assume that this transformation occurs with a constant coefficient (e.g., 100 cu. ft. of natural gas produces 1 MWh of electricity). However, the efficiency of the conversion could also be represented as uncertain.

SCADA equipment monitors volumes, pressures, and temperatures as well as the status of pipeline facilities. It can be used to remotely start and stop compressors, thereby changing flow volumes. A SCADA system controls the flow of gas in links *a→b, b→c, c→d* and *d→e*. We will assume (for the purposes of this example) that the SCADA has two core subsystems. One subsystem supports links *a→b* and *b→c* and the second supports *c→d* and *d→e*. In this example, we assume that the SCADA system always has access to the necessary electric power, but the structure described here can be generalized to the case where that may not be true.

Changes in link capacity over time may include both random failures (that reduce arc capacity) and repair actions of uncertain duration (that restore capacity). We will define states on a link corresponding to different capacity levels, and use Markov and semi-Markov processes to represent state transitions over time. Use of Markov and semi-Markov processes to model evolving system and component conditions is consistent with several previous models (e.g., some recent works include [4,7]).

Figure 1 gives the capacities on links that are considered to be deterministic. Figure 2 defines the stochastic processes for those links that are treated as having uncertain capacity. For example, link *S1→DS1* can have a capacity of 90, 95, 100 or 105. The evolution of capacities on the gas supply links are assumed to be semi-Markov processes; the remaining links are assumed to be characterized by Markov processes. The holding time distributions for the semi-Markov process (*T* matrices) are all assumed to be Normal (and therefore defined by a mean and standard deviation – denoted in Figure 2 by two values separated by a comma within the *T* matrix specifications). Observations from these distributions are rounded to determine the number of periods the process holds. The use of semi-Markov processes on the gas supply links and Markov processes on the remaining links in this example is for illustration only – there is no requirement in the modeling approach for any specific type of process on any given type of link, and Markov and semi-Markov processes can be mixed as appropriate to represent the specific situation being modeled.

Link: *S1→DS*
States: (90 ,95, 100, 105)

$$P_{S1\to D1}=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.000 & 0.000 & 0 & 0.99 \\ 0.000 & 0.000 & .99 & 0 \end{bmatrix}$$

$$T_{S1\to D1}=\begin{bmatrix} - & 4,1 & - & - \\ - & - & 4,1 & - \\ 1,3 & 1,3 & - & 2,2 \\ 1,3 & 1,3 & 2,2 & - \end{bmatrix}$$

Link: *S2→DS2*
States: (100,160,200)

$$P_{S2\to d2}=\begin{bmatrix} 0 & 1 & 0 \\ 0.0 & 0 & 0.9 \\ 0.0 & .9 & 0 \end{bmatrix}$$

$$T_{S2\to D2}=\begin{bmatrix} - & 4,1 & - \\ 1,3 & - & 4,1 \\ 1,3 & 2,2 & - \end{bmatrix}$$

Link: *E1→G1*
States: (600,800)

$$P_{E1\to G2}=\begin{bmatrix} 0.1 & 0.9 \\ 0.001 & 0.999 \end{bmatrix}$$

Links: *SCADA (SUB1, SUB2)*
States: (0 0, 0 1, 10, 11)

$$P_{SCAD}=\begin{bmatrix} 0.8 & .0 & .0 & .1 \\ 0.1 & 0.4 & 0.3 & 0.2 \\ 0.1 & 0.3 & 0.4 & 0.2 \\ .000 & 0.000 & 0.000 & 0.99 \end{bmatrix}$$

Link: *E2→G2*
States: (0,250,400)

$$P_{E2\to G2}=\begin{bmatrix} .2 & .1 & .7 \\ 0 & .2 & .8 \\ .000 & .000 & .99 \end{bmatrix}$$

Links: *a→b & b→c*
State Definitions
If SUB1=0: (0 0, 40 40, 90 60, 130 80, 170 100, 210 120, 250 140)
If SUB1=1: (0 0, 50 70, 100 90, 150 110, 200 130, 250 150, 300 170)

$$P_{a\to b,b\to c}=\begin{bmatrix} .0 & .0 & .0 & .0 & .2 & .2 & .4 \\ .0 & .0 & .0 & .2 & .0 & .2 & .4 \\ .0 & .0 & .0 & .2 & .2 & .0 & .4 \\ .0 & .2 & .0 & .2 & .0 & .0 & .4 \\ .0 & .2 & .0 & .2 & .0 & .0 & .4 \\ .0 & .2 & .0 & .2 & .0 & .0 & .4 \\ .0 & .0 & .0 & .002 & .002 & .003 & .9 \end{bmatrix}$$

Links: *c→d & d→e*
State Definitions
If SUB2=0: (0 0, 10 10, 20 20, 40 40, 60 60, 80 80, 90 90)
If SUB2 =1: (0 0, 20 20, 30 30, 50 50, 70 70, 90 90, 100 100)

$$P_{c\to d,d\to e}=P_{a\to b,b\to c}$$

Figure 2 – Parameters for the uncertain links.

The condition of each of the two SCADA subsystems is represented by a binary random variable where 0 indicates diminished condition and 1 indicates fully functional. Since links *a→b* and *b→c* are controlled by a single SCADA subsystem, changes in their capacity determined by the condition of the SCADA system occur together, creating a correlation between them. This is also the case for links *c→d* and *d→e*. Since the capacities on links *a→b, b→c, c→d* and *d→e* are affected by the condition of the SCADA system, the state definitions depend on the condition of the relevant SCADA subsystem. For example, if the portion of the SCADA system that provides support to links *a→b* and *b→c* is in diminished condition the highest capacity state is 250 on link *a→b* instead of 300.

3. Understanding System Performance

Both steady-state and transient behavior are important to understanding system performance. For this illustrative analysis we will concentrate on two core questions, one focused on transient behavior and the other on steady-state performance. The questions are as follows:

1. If the capacity on each link (as understood through the associated stochastic process) is as low as possible, how long does it take the system to "recover" and satisfy all demands?
2. In steady state, what are the probability distributions for the product delivered at *D1, D2, L1,* and *L2?*

Estimating probability distributions for delivered gas at *D1* and *D2* and electric power delivered to *L1* and *L2* is critical to understanding the service quality that can be offered to customers. Understanding the "time to recover" provides insight into system robustness.

The general problem of which our example is an instance can be described as follows. Consider an infinite horizon generalized network flow problem with the node set N and the arc set A. Suppose $c_t(i,j)$ be the capacity of arc $(i,j)\in A$ in period t. Let $C_t = (c_t(i,j)) \in E$ and $C = \{C_t\} \in E^\infty$ where E is the state space for the capacity on all links in period *t*. Assume that C is a semi-Markov process with probability measure $\mu(C,\theta)$. Let D be the demands at each demand node in each period. Suppose that f is a performance measure defined on E^∞. In the transient analysis f is a distribution of the time to "recover". In the steady state analysis, f is the probability distributions for the product delivered to each demand node (in the example, f contains four probability distributions).

We can estimate the probability distribution for the time to recover using simulation. The procedure to create an observation from this distribution is as follows and by iteratively employing it, we can estimate the distribution.

1. Let $i=1$; for each link assume the capacity has just reached the lowest feasible state.
2. Given the capacity on each link, solve a generalized flow problem to determine the demand satisfied at each location assuming all demands are equally important.
3. Let $i=i+1$.
4. If all demands are satisfied stop. i is an observation of the number of periods required to recover.
5. Update each link state based on the associated stochastic process; go to step 2.

Since some of the stochastic processes have transition probabilities that are quite small and holding time distributions that are quite long, many replications are likely to be required. To overcome this difficulty, importance sampling can be used. The core idea behind using importance sampling in this application is to select alternative transition matrices and holding time distributions which are more computationally advantageous but to "correct" the results using the relative likelihood of seeing the observations under the original parameters. Jeneja and Shahabuddin [6] and Cai, *et al*. [3] begin to give some insight into how to do this; however, there are still significant challenges, especially when the holding time distributions are very long.

If each link is at its lowest capacity, no product can be delivered to any of the customers. Figure 3 presents the probability distribution for the time to recover based on 1000 replicates. The average time to recover is 10.6 periods, but there is about a 5% chance that it will take 20 or more periods, and in one sample experiment, it took 36 periods for recovery to occur. Because of the structure of the analysis, it is easy to determine the conditions that give rise to each observation for recovery time. This type of information is likely to be particularly valuable to decision-makers seeking to improve system performance.

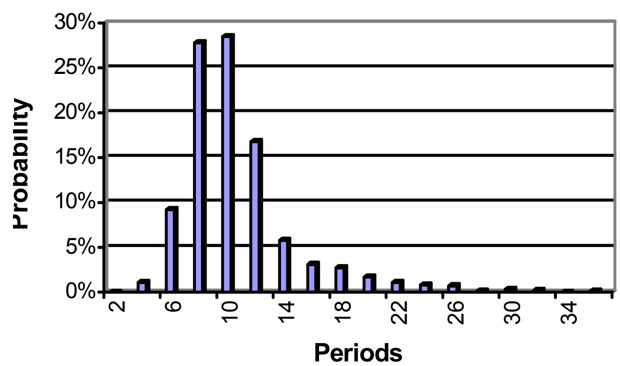

Figure 3 – Distribution of time to recover.

We can estimate the steady-state probability distributions for the product delivered to each demand location using the core ideas in the previous algorithm. Figure 4 illustrates the probability distribution for product delivered to each demand node (when the storage tank is not available) based on 1000 replicates of the steady-state sampling scheme. The proportion of periods in which demands at the various "load nodes" of the system are met varies from about 94% to above 99%. In general, it is harder to meet demand at *D2* then *D1* because of the uncertainty associated with links $b \rightarrow c$, $c \rightarrow d$ and $d \rightarrow e$.

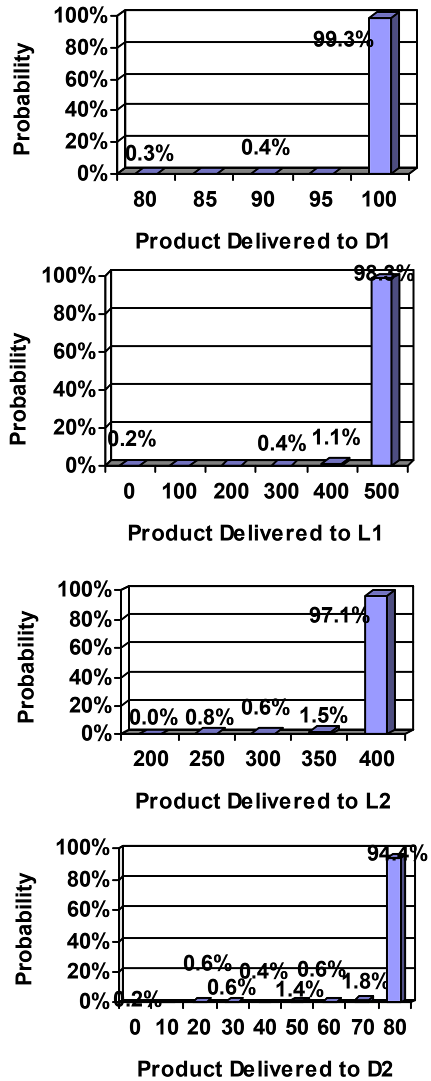

Figure 4 – Probability distribution of product delivered in steady-state.

4. Optimizing Investments

Investment opportunities that can improve performance can be represented in the Markov models as changes in the transition matrices. For example, we could improve the reliability of a piece of equipment, and represent this improvement as reduced probabilities of entering failure states in the Markov model for its capacity. This alternative transition matrix for a link in the network will have an overall effect on the performance of the system as a whole, and this effect can be evaluated via simulation. The substitution of the new transition matrix for the old also implies a cost for making the improvement. The investment optimization problem is then to choose what investments (changes to specific transition matrices) to make so as to have the greatest effect on improving system performance, subject to budget constraints on the total cost incurred.

This optimization problem is straightforward to represent mathematically, but it is quite complicated to solve directly, in part because the evaluation of the benefits for any specific combination of investments requires doing a simulation. In a general mathematical sense, if C is a Markov or semi-Markov process that depends on some parameter θ, it has a probability measure $\mu(C,\theta)$ that determines the probabilities of the system occupying various states. If the system has a performance measure $g(C)$ that is of interest, the simulation model can be viewed as constructing an estimate of the expected performance for a given θ:

$$f(\theta) = E_\theta[g(C)] = \int g(C)\mu(dC,\theta) \qquad (1)$$

The optimization entails choosing $\theta \in \Theta$ to maximize $f(\theta)$ where Θ is the set of all possible choices for the stochastic processes on each link.

Optimization of the investments using (1) is difficult, as the underlying probability measure depends on the parameter θ and the function f needs to be evaluated using simulation for each choice of the parameter θ. However, if we change the underlying probability measure using importance sampling, the optimization is much easier. Now, the evaluation of the function f is totally independent of the parameter θ and can be done beforehand. The results of the evaluation of f simply need to be "corrected" by the relative likelihood as the parameter θ changes in the course of the optimization. Notice that, since the observations are generated and then rescaled based on the relative likelihood with respect to a particular θ, this method inherently uses common random numbers.

Suppose our objective is to maximize the probability that demand is satisfied at all four locations. The performance measure is then as follows:

$$\alpha(\theta) = \sum_C g(C)\pi(C,\theta) \qquad (2)$$

where $g(C)$ is the probability that all the demands are satisfied and $\pi(C,\theta)$ are the steady-state probabilities under θ.

To illustrate the insights that can be obtained through this type of optimization, we use a simple greedy heuristic to estimate the optimal investment strategy. The procedure is as follows:

Step 1: Calculate 1000 sample paths of the system for 1000 periods each, using transition matrices for each link that are similar to those in the base configuration of the system but allow for more "effective simulation." Let θ represent the stochastic processes selected for each links. Calculate the probability all demands are satisfied under θ and let this value be P^*.

Step 2: Identify the links for which there is sufficient funds to make the next incremental investment. If there are no links, stop.

Step 3: For each of the links identified in Step 2, separately calculate the improvement in the probability that all the demands can be satisfied if the next incremental investment on the link is made. Each calculation will require the "correction" of the 1000 sample paths identified in Step 1 based on the "importance function" for the links.

Step 4: Make the additional investment on the link that results in the largest increase in the probability that all the demands can be satisfied provided that increase is positive. If the improvement is positive, update P^*, decrement the budget available given this investment, update the set of stochastic process on the links, θ, and go to step 2; else stop.

Since we assume a probability measure to generate the observations and rescale based on the importance function, many modern heuristic search methods are potentially useful as augmentations of this simple greedy search procedure, such as those given in Rayward-Smith, *et al*. [8] and Hromkovic [5].

To illustrate the type of information that can come from an analysis of this character, assume that investments can be made to improve the reliability of the delivery of gas from both suppliers, the reliability of the SCADA system, the gas transmission lines as well as the generators. For simplicity, assume that all the investments conform to the following pattern as to changes in the stochastic processes. Suppose that for \$100K invested on a link, the lowest capacity state is removed and the transition probabilities to that state are added to those for the next lowest state. For each successive state removed, the cost is \$150K, \$200K, \$250K, \$300K, \$350K, and \$400K respectively. The investments on a link must be done in order. For example, to ensure that at least 100,000 cu ft. of gas is available from supplier 1, \$250K must be invested.

Figure 5 illustrates the trade-off frontier when the objective is to maximize the steady state probability that all the demands are met. Since the estimates are the result of simulation, 95% confidence intervals are also given. The top of the vertical line represents the upper limit and the bottom the lower limit. The mean is the value at the middle of the line. The order of investments suggested is to first ensure the reliability of gas supply from supplier 2, then invest in gas transmission links $c \rightarrow d$ and $d \rightarrow e$, and then in the electric generation links $E1 \rightarrow G1$ and $E2 \rightarrow G2$. The most significant improvement in overall system reliability is from increasing the reliability of gas supply from supplier 2. Without that supply, increases in system capacity further "downstream" in the network are ineffective. Subsequent investments in the gas transmission links and the electric power generation can increase system reliability modestly, but the optimization points to the gas supply as the critical investment area.

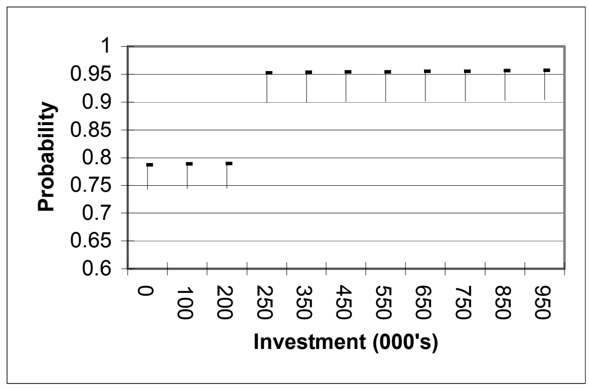

Figure 5 –Level of investment vs. steady-state probability that all demands are satisfied.

5. Conclusions

This paper focuses on representing interdependent infrastructure networks using Markov and semi-Markov processes to reflect uncertain capacity on network links. The Markov-based approach allows analysis of both transient and steady-state concerns regarding availability of service. It also enables representation of correlations (both spatial and temporal) between the conditions of network links, as well as the effects of uncertain condition of supporting information infrastructure (e.g., SCADA systems) on the performance of the controlled physical system. A small-scale example has illustrated the structure of the analysis approach.

In this framework, investments that would improve the performance of selected system components are represented as changes in the stochastic processes governing link capacities. These changes can include changes in the state-space, the transition matrices, or both. A discrete optimization problem can then be formulated to find the set of investments that should be chosen to maximize a performance measure subject to a budget constraint. This has also been illustrated using a gas-electric network example. In the example contained here, the optimization was solved heuristically, using a greedy search procedure.

The modeling framework described here suggests several important areas for further research. One direction for investigation is improved use of importance sampling for the simulation portion of the analysis. The semi-Markov models are likely to have very long transition time distributions from some states, and the transition matrices in both Markov and semi-Markov models of infrastructure networks have very small transition probabilities to some low-capacity states. This implies that the simulation will only rarely enter those states. We have implemented some basic ideas for importance sampling to improve the estimates of system performance in those states, but further effort in this direction is important.

The computational procedures inside the simulation of the network also could be enhanced to represent a broader range of types of infrastructure interconnections more effectively, and to reflect storage of commodities (e.g., natural gas) within the network. These are natural extensions to the simple kinds of examples we have worked with so far.

There is substantial room for more work on the search process for optimal investments. Our initial experiments have used a simple greedy search process. Although this has worked relatively well, there are several other avenues that ought to be explored, and that could yield more effective optimization methods.

There are also likely to be several different choices for system performance measures, and those choices will have effects on the types of investments that are considered to be most beneficial in the networks. This is an important area related to the optimization.

Finally, the model structure is clearly dependent on having good estimates of parameters (transition matrices, time-to-transition distributions, etc.), and these estimates have to be constructed from empirical data. Because the infrastructure systems of interest are typically highly reliable, there may be relatively little data on transitions to some potentially "interesting" states that are very rarely entered. This is an important empirical issue in making the approach really useful in practice.

6. References

[1] T. Aven, Availability Evaluation of Oil/Gas Production and Transportation Systems, *Reliability Engineering*, 18, 35-44, 1987.

[2] T. Aven, Availability Evaluation of Flow Networks with Varying Throughput-Demand and Deferred Repair, *IEEE Transactions on Reliability*, 38, 499-505, 1989.

[3] W. Cai, M.H. Kalos, M. deKoning, and V.V. Bulatov, "Importance Sampling of Rare Transition Events in Markov Processes," *Physical Review E*, 66, 046703, 2002.

[4] J.A.M. Hontelez, H.H. Burger and D.J.D Wijnmalen, "Optimum-condition Based Maintenance Policies for Deteriorating Systems with Partial Information", *Reliability Engineering & System Safety*, Vol. 51, No. 3, 267-274, 1996.

[5] J. Hromkovič, "*Algorithmics for Hard Problems: Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics*", Springer, Germany, 2001.

[6] S. Juneja and P. Shahabuddin, "Fast Simulation of Markov Chains with Small Transition Probabilities", *Management Science*, Vol. 47, No. 4, 547-562, 2001.

[7] N. Limnios and G. Oprisan, *Semi-Markov Processes and Reliability*, Birkhäuser, Boston, 2001.

[8] V.J. Rayward-Smith, I.H. Osman, and C.R. Reeves and G.D. Smith, *Modern Heuristic Search Methods*, John Wiley & Sons Ltd., New York, 1996.

[9] S. Rinaldi, J. Peerenboom and T. Kelly, "Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies", *IEEE Control Systems Magazine*, 11-25, December 2001.