Optimizing Investment for Recovery in Interdependent Infrastructures

Ningxiong Xu
Linda K. Nozick
Mark A. Turnquist
School of Civil and Environmental Engineering
Cornell University
Hollister Hall
Ithaca, NY 14853
emails: nx22@cornell.edu, lkn3@cornell.edu and mat14@cornell.edu

Dean A. Jones
Sandia National Laboratories[1]
Albuquerque, NM 87185
email:dajones@sandia.gov

June 2006

Abstract

Reliable operation of complicated interdependent infrastructures, (including transportation, electric power, oil, gas, telecommunications and emergency services) is vital to developed economies. This paper develops a method to estimate the "time to recover" from a disruption in such interdependent infrastructures. It also develops a mathematical model and solution procedure to optimize investments in interconnected infrastructures to achieve improvements in "time to recover" subject to a budget constraint. These methods are illustrated on an example gas-electric infrastructure network.

1. Introduction

Rinaldi *et al.* [1] 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 They observe the need for more research to consider infrastructure interdependencies, development of metrics to describe the performance of infrastructures, and creation of models for designing and operating these systems.

Two leading approaches for reliability analysis in interdependent infrastructures are agent-based simulation and input-output analysis. The core idea behind the development of agent-based simulations for this application is that individual components and subsystems can be represented as agents and by letting them evolve and interact, emergent behaviors (i.e. interdependencies) can be identified (e.g. [2], [3], [4] and [5]). Agent-based simulation is also being used to investigate the markets for electric power and natural gas (e.g. [6], [7], [8], and [9]). The emphasis in market-level models is on trading behavior of economic agents in these industries.

Input-output analysis has traditionally been used to model the interactions of sectors of the economy and forecast the effects of changes in one part of the economy on performance in other sectors. Haimes and Jiang [10] suggest that the same modeling paradigm may be useful to model the interactions and interdependencies within and across infrastructures. Whereas much of the agent based simulation for reliability analysis contains very detailed system representation, input-output modeling is likely to be very aggregate.

A third type of analysis is based on network models. Nozick *et al.* [11] develop a mathematical representation for interconnected infrastructures based on a network representation (graphs of nodes

[1] 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.

and arcs). The arcs represent components or subsystems in an infrastructure or the connections between infrastructures. The arcs have capacities and these capacities may be uncertain and evolve over time.

Changes in arc capacity over time include both random failures (that reduce arc capacity) and repair actions of uncertain duration (that restore capacity). The capability of one component or subsystem may be correlated with the performance of several other components or subsystems.

In this paper we extend the mathematical framework in Nozick *et al.* [11] in two significant dimensions. First, we allow for storage within these systems. This analysis emphasizes the effective use of storage as a mechanism to improve service reliability. Second, we develop a mathematical formulation and solution procedure for investment planning to achieve improvements in "time to recover" from disruptions.

The next section describes the illustrative example we will use to illustrate the modeling tools. This example is a modification of that used in Nozick *et al.* [11]. The third section describes the analysis developed to estimate the key transient performance metric, "time to recover". The fourth section develops the formulation and solution procedure to optimize system performance with respect to this measure subject to a budget constraint. The fifth section gives conclusions and directions for future research.

2. Illustrative Example

We 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.

We 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., [12], [13]).

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.

There is a tank in the network which is used to store extra gas in a period for use in future periods. The key idea here is to use the tank for protecting the network against disruption.

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*. The gas from node *DS1* can either be used in a period or stored in the tank, *ST*, if there is sufficient space in the tank for the additional gas. The gas entering the tank in a period can supply the network for the next period and thereafter (until it is consumed). Similarly, the gas that enters the distribution network from supplier *S2* enters via node *DS2*.

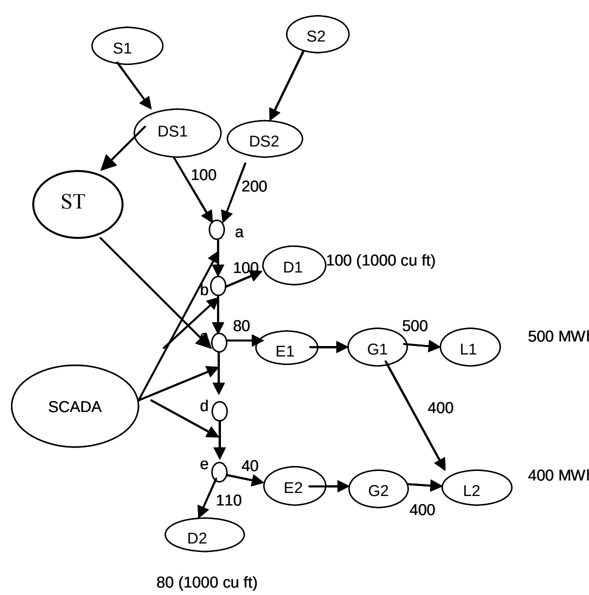


Figure 1. An example network for interdependent gas and electric infrastructures

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.

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 $a\rightarrow b$, $b\rightarrow c$, $c\rightarrow d$ and $d\rightarrow e$. We will assume (for the purposes of this example) that the SCADA has two core subsystems. One subsystem supports links $a\rightarrow b$ and $b\rightarrow c$ and the second supports $c\rightarrow d$ and $d\rightarrow e$. 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 process to represent state transitions over time. 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 $S1\rightarrow DS1$ can have a capacity of 90, 95, 100 or 105. The evolution of capacities on all links is assumed to be a Markov process.

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 $a\rightarrow b$ and $b\rightarrow c$ 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 $c\rightarrow d$ and $d\rightarrow e$. Since the capacities on links $a\rightarrow b$, $b\rightarrow c$, $c\rightarrow d$ and $d\rightarrow e$ 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 $a\rightarrow b$ and $b\rightarrow c$ is in diminished condition the highest capacity state is 250 on link $a\rightarrow b$ instead of 300.

Link: $S1\rightarrow DS1$
States: (90 ,95, 100, 105)

$$P_{S1\rightarrow DS1} = \begin{bmatrix} 0.03 & 0.03 & 0.03 & 0.91 \\ 0.05 & 0.3 & 0.15 & 0.5 \\ 0.005 & 0.005 & 0.01 & 0.98 \\ 0.005 & 0.005 & 0.05 & 0.94 \end{bmatrix}$$

Link: $S2\rightarrow DS2$
States: (100,160,200)

$$P_{S2\rightarrow DS2} = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.005 & 0.5 & 0.495 \\ 0.005 & 0.005 & 0.99 \end{bmatrix}$$

Links: *SCADA (SUB1, SUB2)*

States: (0 0, 0 1, 10, 11)

$$P_{SCADA} = \begin{bmatrix} 0.8 & .04 & .04 & .12 \\ 0.1 & 0.4 & 0.3 & 0.2 \\ 0.1 & 0.3 & 0.4 & 0.2 \\ .0006 & 0.0002 & 0.0002 & 0.999 \end{bmatrix}$$

Link: $E1\rightarrow G1$
States: (600,800)

$$P_{E1\rightarrow G1} = \begin{bmatrix} 0.1 & 0.9 \\ 0.01 & 0.99 \end{bmatrix}$$

Link: $E2\rightarrow G2$

States: (0,250,400)

$$P_{E2\rightarrow G2} = \begin{bmatrix} 0.2 & 0.1 & 0.7 \\ 0.1 & 0.2 & 0.7 \\ 0.005 & 0.005 & 0.99 \end{bmatrix}$$

Links: $a\rightarrow b$ & $b\rightarrow c$
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)

$$P_{a\rightarrow b,b\rightarrow c} = P_{c\rightarrow d,d\rightarrow e} = \begin{bmatrix} 0.05 & 0.05 & 0.05 & 0.05 & 0.2 & 0.2 & 0.4 \\ 0.05 & 0.05 & 0.05 & 0.2 & 0.05 & 0.2 & 0.4 \\ 0.05 & 0.05 & 0.05 & 0.2 & 0.2 & 0.05 & 0.4 \\ 0.05 & 0.20 & 0.05 & 0.2 & 0.05 & 0.05 & 0.4 \\ 0.05 & 0.20 & 0.05 & 0.2 & 0.05 & 0.05 & 0.4 \\ 0.05 & 0.20 & 0.05 & 0.2 & 0.05 & 0.05 & 0.4 \\ 0.0020 & 0.002 & 0.002 & 0.003 & 0.003 & 0.004 & 0.984 \end{bmatrix}$$

Links: $c\rightarrow d$ & $d\rightarrow e$
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)

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

Figure 2 – Parameters for the uncertain links

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.

3. Estimating the "time to recover"

There is considerable debate over how to assess the performance for infrastructure systems. Some studies focus on steady-state performance and others focus on performance when the system is operating in a degraded state. Nozick *et al.* [11] focus on the probability distribution for the product delivered to each of the customers in steady state. Xu *et al.* [14] considers multiple measures including the average time to restore power to all customers, and the time required to restore power to either 90% or 95% of customers, when determining the order to repair components of an electric power transmission system after an earthquake. In a transportation context, Sun *et al.* [15] use the probability distribution for "network capacity" which is defined as the total volume which can be accommodated across all customers when all facilities are operating "normally" and when some facilities are in "degraded" condition.

After an event there is great incentive to restore services as quickly as possible. Dahlhamer *et al.* [16], through a study of financial losses in the Northridge earthquake, found that business' loss increases dramatically when an electric power outage lasts longer than 24 hours. In this paper we focus on "the time to recover" from a disruption. For this illustrative analysis we concentrate on the following two core questions:

1. If the capacity on each link (as understood through the associated stochastic process) is as low as possible, on average how long does it take the system to "recover" and therefore be capable of satisfying all demands?
2. If the capacity on each link (as understood through the associated stochastic process) is as low as possible, what is the probability that it will take at least a pre-defined number of periods for the system to "recover" and satisfy all demands (e.g. at least 6 periods)?

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*. Denote $c_t(i,j)$ as 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 discrete time Markov process with initial state distribution $\mu=(\mu(x): x\in E)$ and transition matrix $P=(P(x,y): x,y\in E)$. Let *D* be the demands at each demand node in each period. The performance measure defined on the Markov Chain under the probability distributions μ and *P* is $\alpha=E_{(\mu,P)}(h(C_0,\dots,C_T))$ for some suitable function h defined on the Markov chain *C*, where *T* is a stopping time.

Observe that this formulation includes our two cases. Recall that the first case is the average time when the network first meets all demands and the other is the probability that the time when the network first meets all demands is at least T_0 periods. To see how these two performance measures fit within the formulation, let C_t be the vector of link capacities in the network in time period *t* and let *S* be the set of all feasible link flows. Let *F* be the subset of *S* that corresponds to flows that do not meet demands at the demand nodes. Let *T* be the first time when the flows meet all demands, i.e., $T=\min\{t: C_t\in S-F\}$. The stopping time *T* represents the "recovery time" of the system from the disruption.

In the first case, let $h(C_0,\dots,C_T)=T$. Then $\alpha=E_{(\mu,P)}(h(C_0,\dots,C_T))$ is the average time when the network first meets all demands. For the second case, if T_0 is a constant, let $h(C_0,\dots,C_T)=I(T\ge T_0)$. I is an indicator function, i.e., $I(T\ge T_0)=1$ if $T\ge T_0$ and 0 otherwise. Then $\alpha=E_{(\mu,P)}(h(C_0,\dots,C_T))$ is the probability that the first time when the network meets demands is at least T_0 periods.

The performance measure can be estimated by using simulation as follows:

1. Choose a sample size n
2. Simulate the Markov chain *C* up to time period *T* based on the probability measures (μ, *P*).
3. Calculate $h(C_0,\dots,C_T)$
4. Repeat steps 2 and 3 n times and obtain the average and standard deviation of the performance measure.

Steps 2 and 3 in the algorithm can be specified for the two performance measures as follows:

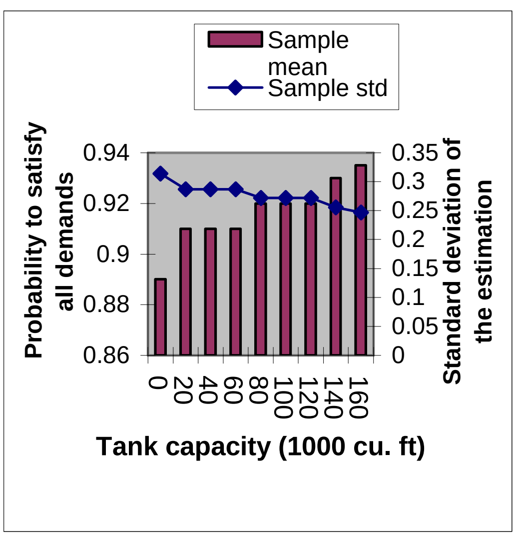
a. Let $i=1$; for each link assume the capacity has just reached the lowest feasible state.
b. Identify the capacity on each link, and solve a generalized flow problem to determine the demand satisfied at each location, assuming all demands are equally important.
c. If all demands are satisfied stop. i is an observation of the number of periods

required to recover. If all demands are not satisfied, let $i=i+1$ and go to b.

d. If the performance measure is based on the average "time to recover," then let $h(C_0,\ldots,C_T)=i$. If the performance measure is based on the probability that the "time to recover" is at least T_0 periods, then $h(C_0,\ldots,C_T)=1$ if $i \ge T_0$ and $h(C_0,\ldots,C_T)=0$ otherwise.

The simulation procedure to estimate the performance measure can be implemented by using the importance sampling technique. The core idea behind using importance sampling in this application is to select alternative transition matrices that are computationally advantageous, and then "correct" the results using the relative likelihood of seeing the observations under the original parameters.

Figure 3 shows the performance of the illustrative network in steady-state. There is a probability of approximately 0.89 that all demand is satisfied if no tank is present. The standard deviation of the estimate is 0.3. The steady-state probability of meeting all demand increases to 0.91 if a tank with a capacity of 20,000 cu ft is available, and to approximately 0.94 if the tank capacity is increased to 160,000 cu ft.



\

Figure 3 – Impact of tank capacity on the probability of satisfying all demands in steady-state

The steady-state demand in these experiments is 270,000 cu ft. per period, and the average per-period deliveries in steady-state increase from 259,000 cu ft. with no tank to approximately 265,000 cu ft. as the tank capacity increases. Thus, even with no storage tank, most of the demand is being met. As the storage capacity increases, the net increase in delivery capacity is not large, but the reliability of the system improves.

Figure 4 illustrates the impact of the tank on the probability that it will take at least six periods for all the demands to be satisfied. If no tank is available, the chance it will require at least 6 periods to satisfy all demands is 0.6. The coefficient of variation in this estimate is about 2.9%. That is, there is about 40% of chance that at least one customer does not receive the entire product they have requested within 5 periods. The existence of the tank reduces the probability it will take at least 6 periods to about 20%. The coefficient of variation in this estimate is about 7.8%. Again the capacity of the tank plays an important role in the ability of the system to satisfy demands quickly after a disruption.

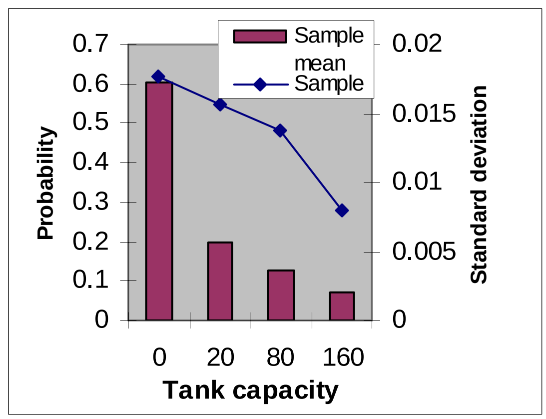


Figure 4 – Impact of tank capacity on the probability that at least six periods are required to satisfy all demands

4. Optimizing the "time to recover"

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 both the initial distribution and transition matrix for the Markov chain $C = (C_t : t \geq 0)$ depend on a parameter θ, the initial distribution and transition matrix can be written as $\mu(\theta) = (\mu(x,\theta) : x \in E)$ and $P(\theta) = (P(x,y,\theta) : x, y \in E)$ for $\theta \in \Theta$. Suppose that the performance measure $h(C_0,\ldots,C_T)$ is a function of $(C_0,\ldots,C_T)$ only. Note that our performance measures in this paper satisfy this form. Then the performance measure is $\alpha(\theta) = E_{(\mu(\theta),P(\theta))}(h(C_0,\ldots,C_T))$ and the decision problem is to choose $\theta \in \Theta$ that minimizes $\alpha(\theta)$.

Solving the optimization problem requires evaluation of the performance measure $\alpha(\theta)$ for different values of θ. This is difficult because $\alpha(\theta)$ does not have an analytic form for most complex systems. One way to overcome this difficulty is to approximate the function based on a sample mean. Let $\alpha_n(\theta)$ be the sample mean with sample size n. The approximated problem is to choose $\theta_n \in \Theta$ that minimizes $\alpha_n(\theta)$. Suppose that θ_n^* minimizes $\alpha_n(\theta)$ subject to $\theta_n \in \Theta$ and θ^* minimizes $\alpha(\theta)$ subject to $\theta \in \Theta$. Under some conditions, $\theta_n^* \to \theta^*$ as $n \to \infty$ (see [17]). Thus, $\theta_n^* \approx \theta^*$ for large n under some conditions. In this paper, we use θ_n^* to approximate θ^* in our optimization problem.

An effective way to estimate the sample mean is to use the importance sampling technique. That is, if we choose an initial distribution $\nu(x)$ and a transition matrix $Q = (Q(x,y) : x, y \in E)$ such that $\nu(x) > 0$ whenever $\mu(x,\theta) > 0$ and $Q(x,y) > 0$ whenever $P(x,y,\theta) > 0$ for all $\theta \in \Theta$, then the performance measure can be written as follows:

$$\alpha(\theta) = E_{(\nu,Q)}(L(C_0,\ldots,C_T,\theta)h(C_0,\ldots,C_T)) \quad (1)$$

where $L(C_0,\ldots,C_T,\theta)$ is the likelihood function.

If the evaluation of function h for the given sample path $(\widetilde{C}_0, \widetilde{C}_1, \ldots,)$ is "expensive," the transformation is a very efficient way to create the sample mean function $\alpha_n(\theta)$. This is because function h only needs to be evaluated based on the sample path generated by the probability distribution (ν, Q). That is, it is not necessary to evaluate function h for the sample paths generated by the probability measure dependent on each parameter θ value. In our case, this is important because estimating the "time to recover" in a given simulation sample requires solving a network flow problem in each period for given link capacities. This can be computationally intensive.

The next question is how to find the minimum θ_n^* in the optimization problem. This can be very difficult if the feasible set Θ is a large discrete set. In this paper, we develop a Genetic Algorithm (GA) heuristic to find a solution for the investment optimization.

The string used in the GA to solve the optimization has one entry for each link, representing the amount of investment on that link. We use one point crossover with 5% probability of mutation, simple ranking selection, 20 individuals in a generation, and 30 generations. We evaluate the fitness of an individual using simulation with importance sampling.

To illustrate the information that can come from this type of analysis, 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 transition probability to enter the lowest state decreases by 80%. The amount of the decrease is then added evenly to the remainder of the transition probabilities to the other states.

For example, consider Link *S1*$\rightarrow$*DS1*. The new transition matrix for an investment of 100K would then be as follows (rows may not sum to 1 as a result of the number of significant figures displayed).

$$P_{S1\rightarrow DS1} = \begin{bmatrix} 0.006 & 0.038 & 0.038 & 0.918 \\ 0.01 & 0.313 & 0.163 & 0.513 \\ 0.001 & 0.006 & 0.011 & 0.98 \\ 0.001 & 0.006 & 0.051 & 0.94 \end{bmatrix}$$

For a second investment increment, the probability of transitions from any state to the second lowest state can be reduced. For a third investment increment, the probability of transitions from any state to the third lowest state can also be reduced, etc. For each successive improvement, we assume the cost is \$150K, \$200K, \$250K, \$300K, \$350K, and \$400K respectively. For example, for an additional \$150K invested in a link, the transition probability to the second lowest capacity state decreases by 80%. This reduction is then assumed to be distributed evenly to transitions to higher states.

Illustrating the investment of an additional \$150K on the link from $S1 \rightarrow DS1$ would result in the following transition matrix (rows may not sum to 1 as a result of the number of significant figures displayed).

$$P_{S1\rightarrow DS1} = \begin{bmatrix} 0.006 & 0.0076 & 0.053 & 0.933 \\ 0.01 & 0.063 & 0.289 & 0.639 \\ 0.001 & 0.001 & 0.014 & 0.984 \\ 0.001 & 0.001 & 0.054 & 0.944 \end{bmatrix}$$

The investments on a link must be done in order. For example, to ensure that the link ES1 $\rightarrow$ S1 has the previous transition matrix, \$250K must be invested.

This scheme has been used to create an illustration of the methods. In the experiments in this paper the percent reduction used is 20% rather than the 80% illustrated in the examples above. In practice each potential investment on a link could map to a unique transition matrix that would be used in the optimization.

Figure 5 illustrates the impact of investment on "the mean time to recover" if there is no tank available. If the investment is \$100K, the estimated mean time to recover drops to about 5.7 periods. The recommended investment is to improve the performance of SCADA. This is because SCADA controls links $a \rightarrow b$, $b \rightarrow c$, $c \rightarrow d$ and $d \rightarrow e$. The improvement of the performance of the SCADA system would lead to the improvement of the performance of all of these links.

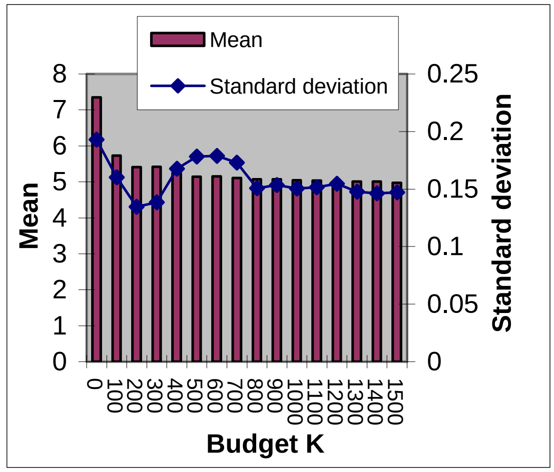


Figure 5 – Relationship between budget and average number of periods to recover

Figure 6 illustrates the impact of investments on the mean time to recover, given the tank capacities are 20,000 cu ft. and 80,000 cu ft., respectively and a budget of \$100K. Unlike the case where there is no the tank, the recommended investments are focused on improving the performance of link S1 $\rightarrow$ DS1 because the tank receives gas via DS1. The improvement in link S1 $\rightarrow$ DS1 allows for more gas to be available to be stored in the tank and therefore used to aid in the recovery.

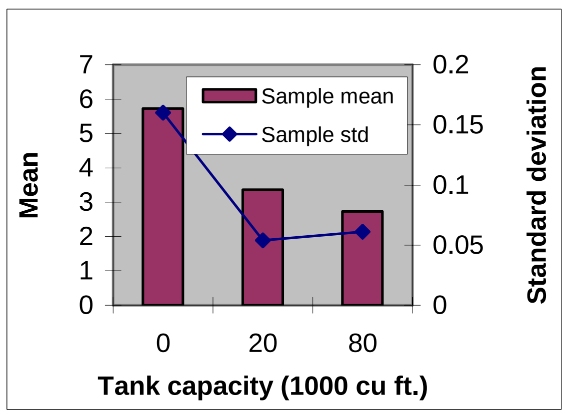


Figure 6 – Relationship between tank capacity and the mean number of periods required to recover (for an investment of \$100K)

Figure 7 illustrates the probability that at least six periods are needed to recover given the tank capacities are 20,000 cu ft. and 80,000 cu ft., respectively when the budget is \$100K. The existence of the tank reduces the probability of long recovery times quite dramatically.

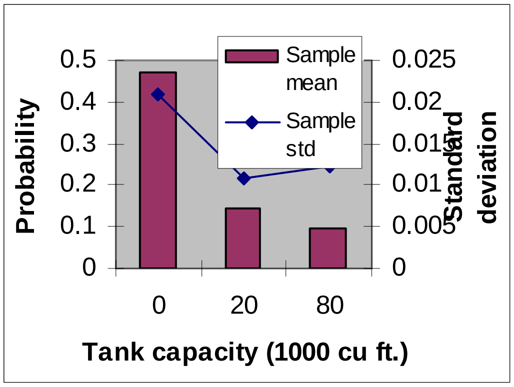


Figure 7 – Relationship between the tank capacity and the probability that six or more periods are required to recover (for an investment of $100K)

5. Conclusions

This paper focuses on representing interdependent infrastructure networks using network flow models and Markov models. The network flow models are used to represent the flow of commodities through the system and the storage of products within the system, whereas the Markov models are used to represent the evolution in the capacity of the links over time. The Markov-based approach allows analysis of both transient and steady-state estimates of service quality. It also enables representation of correlations (both spatial and temporal) between the conditions of network links, as well as the effect of the uncertain condition of the supporting information infrastructure (e.g., SCADA systems) on the performance of the controlled physical system. A gas-electric network 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 and the transition matrices. A discrete optimization problem has been formulated and solved using a genetic algorithm to find the set of investments that should be chosen to optimize a performance measure subject to a budget constraint. This has also been illustrated using a gas-electric network example.

The modeling framework described here suggests several important areas for further re-search. Our current analysis treats the demand at each location as fixed over time which is not typical of these types of systems. Hence, an important extension to this framework is the development of metrics to characterize the performance of systems with demands that vary over time under both steady-state and transient conditions.

In the example developed in this paper there is a single tank and its connection to the gas network is fixed. Further, the capacity of the tank itself is fixed. Storage can serve as an effective hedge against uncertainty. However, how much storage to provide and where to place it within the system is an important question. Hence an important extension of this analysis is to consider the optimization of storage size and location within the system.

Finally, the model structure is clearly dependent on having good estimates of parameters (transition matrices, 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 for making the approach really useful in practice.

Acknowledgements

The authors would like to thank the National Science Foundation (CMS -00408577) for financial support of this research. This support is gratefully acknowledged, but the authors take sole responsibility for the content of the paper.

References

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

[2] Y. Tomita, C. Fukui, H. Kudo, J.Koda, and K. Yabe, "Cooperative Protection System with an Agent Model", *IEEE Transactions on Power Delivery*, Vol. 13, No. 4, 1060-1066, 1998.

[3] A. M. Wildberger, "Modeling with Independent Intelligent Agents for Distributed Control of the Electric Power Grid", *Proceedings of the American Power Conference*, vol. 59-1, Proceedings of the 1997 59th Annual American Power Conference, Chicago, IL 361-364, 1997.

[4] A.M. Wildberger, "Modeling the Infastructure Industries as Complex Adaptive Systems", in M. Ades & R. Griebenow (eds.), *Proceedings of Simulation International XV*, 168-173, San Diego, CA, 1998.

[5] M. Amin, "Toward Secure and Resilient Interdependent Infrastructures", *Journal of Infrastructure Systems*", Vol 8 No. 3, 2002, 67-75.

[6] M. North, *SMART II: the Spot Market Agent Research Tool version 2.0 plus Natural Gas*, ANL/DIS/CP-103635, Argonne National Laboratory, US DOE, Illinois, 2000.

[7] M. North, *SMART II+: the Spot Market Agent Research Tool version 2.0*, ANL/DIS/CP-103634, Argonne National Laboratory, US DOE, Illinois, 2000.

[8] L.H. Tsoukalas and Ö. Uluyol, "Anticipatory Agents for the Deregulated Electric Power System", C. Macal and D. Sallach (eds.), *Proceedings of the Workshop on Agent Simulation: Applications, Models, and Tools*, University of Chicago, 114-123, October 1999.

[9] R. Thomas, T.D. Mount, R. Zimmerman, W.D. Schulze, R.E. Schuler and L.D. Chapman, "Testing the Effects of Price Responsive Demand on Uniform Price and Soft-Cap Electricity Auctions", *Proceedings of the 35th Hawaii International Conference on Systems Sciences*, 2002.

[10] Y. Haimes and P. Jiang, "Leontief-Based Model of Risk in Complex Interconnected Infrastructures", *Journal of Infrastructure Systems*, Vol. 7, No. 1, March 2001.

[11] L. Nozick, M. Turnquist, D. Jones, J. Davis, and C. Lawton. "Assessing the Performance of Interdependent Infrastructures and Optimizing Investment", *Proceedings of the 37th Hawaii International Conference on Systems Sciences,* Jan. 2004**.**

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

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

[14] N. Xu,, S. Guikema, S., R. Davidson, L. Nozick, Z. Çağnan, and K. Vaziri. "Optimizing Scheduling of Post-Earthquake Electric Power Restoration Tasks", working paper, Cornell University, 2006.

[15] Y. Sun, M.A. Turnquist, and L. Nozick. "Estimating Freight Transportation System Capacity, Flexibility and Degraded-Condition Performance", to appear in the *Transportation Research Record* 2006.

[16] J.M. Dahlhamer, G.R. Webb, K.J. Tierney, "Predicting business financial losses in the 1989 Loma Prieta and 1994 Northridge Earthquakes: Implications for loss estimation research," *Preliminary Paper No. 282*, Disaster Research Center, University of Delaware, 1999.

[17] G. Gürkan, Y. Ozge and S. Robinson, "Sample Path Optimization in Simulation", *Proceedings of the 1994 Winter Simulation Conference*, December 11-14 1994, 247-254.