

Toward Developing Genetic Algorithms to Aid in Critical Infrastructure Modeling

**2007 IEEE Conference on Technologies
for Homeland Security**

May Robin Permann

May 2007

The INL is a
U.S. Department of Energy
National Laboratory
operated by
Battelle Energy Alliance



This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint should not be cited or reproduced without permission of the author. This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the United States Government or the sponsoring agency.

Toward Developing Genetic Algorithms to Aid in Critical Infrastructure Modeling

May Robin Permann, Idaho National Laboratory, Idaho Falls, Idaho, may.permann@inl.gov

Abstract—Today's society relies upon an array of complex national and international infrastructure networks such as transportation, telecommunication, financial and energy. Understanding these interdependencies is necessary in order to protect our critical infrastructure. The Critical Infrastructure Modeling System, CIMS[®], examines the interrelationships between infrastructure networks. CIMS[®] development is sponsored by the National Security Division at the Idaho National Laboratory (INL) in its ongoing mission for providing critical infrastructure protection and preparedness. A genetic algorithm (GA) is an optimization technique based on Darwin's theory of evolution. A GA can be coupled with CIMS[®] to search for optimum ways to protect infrastructure assets. This includes identifying optimum assets to enforce or protect, testing the addition of or change to infrastructure before implementation, or finding the optimum response to an emergency for response planning. This paper describes the addition of a GA to infrastructure modeling for infrastructure planning. It first introduces the CIMS[®] infrastructure modeling software used as the modeling engine to support the GA. Next, the GA techniques and parameters are defined. Then a test scenario illustrates the integration with CIMS[®] and the preliminary results.

1. INTRODUCTION

National Infrastructures are composed of networks and systems of networks. These networks do not exist in isolation, but are joined together by their complex interactions. Interdependent relationships refer to the dependent relationships or influences between infrastructures. Not well understood, the interdependencies of seemingly separate systems such as transportation, health care and the power grid are often overlooked until an interruption in one occurs, causing a cascade of events in the other sectors. The challenge for infrastructure owners and operators is that when a critical system is lost, the inevitable consequences are not always easy to predict. For example, when electrical power is lost, a cascade of events occurs like a series of falling dominos. People may leave work en masse when traffic lights are out, creating congestion, causing accidents, and requiring emergency response or repair crews. In turn, such problems may impede power restoration. The injured may be taken to hospitals already stressed by the power outage. What are the effects if power remains out for a day, a week, or longer, or in the middle of a heat wave? The interactions of multiple systems and

complex, multi-level system failures create significant consequences and potentially disastrous results.

Thus, given the interconnectivity between infrastructure sectors and components, decision makers need to look beyond the initial impact of an event. The same must be considered for the response to an event. These downstream effects that can occur from the initiating event or subsequent actions are not always obvious. Unfortunately, the secondary and follow-on effects can be many times more consequential than the initial impact. A growing area of research is that of effects analysis to examine not only the immediate and direct effects, but to also examine and understand the latent effects.

Planning and response to an emergency is at some level a resource allocation problem. Resources are allocated against competing needs and limited supplies to achieve a desirable outcome. When making such allocation decisions however, consideration must be given not only to the immediate impact but to downstream effects. This paper describes research that couples infrastructure interdependency modeling and genetic algorithm optimization to rank alternative options for infrastructure preparedness planning and effects based operations (EBO).

This paper first describes effects based operations, interdependency modeling and genetic algorithms (GAs) in Section 2. While the context presented is that of a warfighter, the same analysis is valid for Homeland Defense in natural disaster and malicious events. Section 3 describes the Critical Infrastructure Modeling System (CIMS[®]) as the modeling and simulation framework for evaluation infrastructure interdependencies. Finally, a Genetic Algorithm is described as an analysis approach to evaluate course of action against potential outcomes in Sections 4 and 5.

2. BACKGROUND

Effects-based Operations (EBO)

EBO is a methodology for planning, executing and assessing operations to attain the effects required to achieve desired national security objectives. "EBO explicitly and logically links the effects of individual tactical actions directly to desired military and political outcomes. By

focusing on effects—the full range of outcomes, events, or consequences that result from a specific action—commanders can concentrate on meeting objectives instead of managing target lists. Effects-based actions or operations are those designed to produce distinct, desired effects while avoiding unintended or undesired effects [1].”

EBO is an approach to planning, executing and assessing military operations with an explicit focus on effects as opposed to targets or even objectives. The future of EBO is automated tools to build and assess plans that link objectives to effects (including direct, indirect, physical effects, behavioral effects and the mechanisms through which effects are achieved), and then link the effects and mechanisms to specific actions which need to be taken.

EBO supports all mission types from Humanitarian Relief Operations all the way to Major Theater War. It encompasses all approaches from lethal attacks to utilizing information warfare. This includes a new concept termed forth generation warfare (4GW). Retired Colonel Thomas Hammes, U.S. Military Complex, described this as the ability to defeat a stronger economic and military power by using the society’s networks to carry on its fight and attack the minds of the enemy’s decision makers to destroy its political will. This attacks the adversary through its available political, economic, social, and military networks [2].

Operational Net Assessment (ONA) is the integration of people, processes, and the tools that use multiple information sources and collaborative analysis to build shared knowledge of the adversary, the environment, and ourselves in understanding and effectively employing EBO. ONA analytical products are based on a system-of-systems analysis and understanding of key relationships,

dependencies, strengths, and vulnerabilities within and between the political, military, economic, social, information, and infrastructure (PMESII) elements. Figure 1 shows the dual nature of the effects-based approach.

Although the concept of EBO was conceived for use against an adversary, it follows that it can be used to analyze one’s self to identify weakness and the effects of attack. This includes identifying optimum assets to enforce or protect, testing the addition of or change to infrastructure before implementation, or finding the optimum response to an emergency for response planning.

Interdependency Modeling

As previously stated, it is imperative for leadership and decision makers to understand the relationships and influences of the PMESII dimensions. In order to better understand these interdependencies/relationships/influences and protect our critical assets, simulations can be used as tools to identify key or weak links resulting from infrastructure interconnectivity that otherwise may not be readily apparent. Multiple scenario runs evaluating the loss and sequence of loss of different assets can illuminate the assets most critical to the sustainability of predefined critical assets. This involves randomly or methodically iterating through a wide range of asset state combinations and asset loss scenarios in order to understand the cascading effects and identify the assets supporting the defined critical assets.

Infrastructure asset ranking is routinely conducted by government and private industry in order to prioritize and designate key assets which require the most protection and reinforcement. This identification and ranking of critical infrastructure assets does not always take into consideration the importance of the assets required to sustain critical

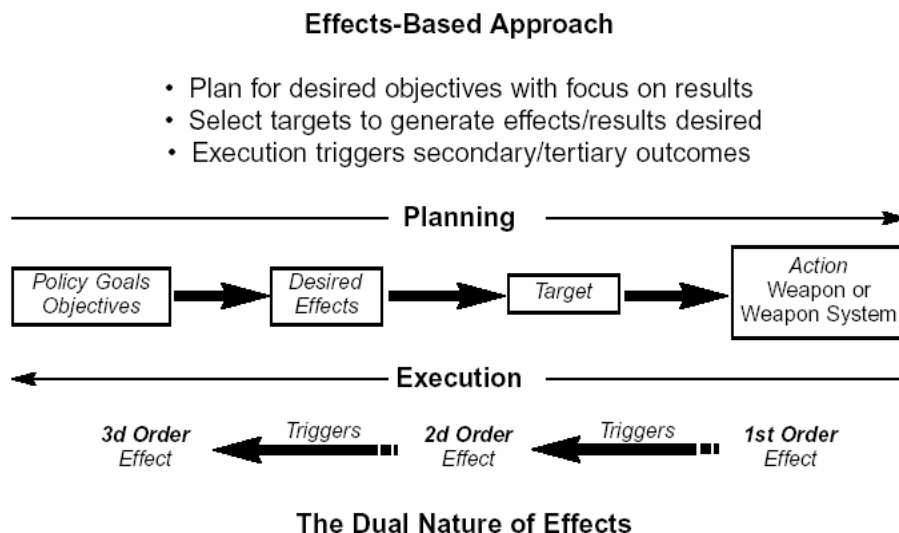


Figure 1 - The effects-based approach

function. For example, a hospital may be deemed a critical infrastructure asset, but does that also include its entire supporting infrastructure? Consider the loss of power to the hospital due to the flooding of an electrical substation. Backup emergency generators can restore power, but have only limited duration if flooding of the transportation routes prevents refueling. Situations similar to this resulted from the flooding from Hurricane Katrina and the interplay of multiple infrastructures became painfully evident in recovery activities. This concept of supporting infrastructure to critical assets is defined as a critical sub-network. An infrastructure critical sub-network of a network G is defined as a sub-graph of G whose elements include only those assets required to sustain a set of critical assets or specific functionality.

This EBO analysis uses infrastructure interdependency modeling to examine potential initial states and course of actions to provide the full range of simulated direct and indirect effects of that action. These effects, of course, are dependent on the scope and accuracy of the model. The outcome of any action is therefore shaped by the intricate interactions between them.

Decision Support Systems

In an emergency or time critical event, decision makers are faced with the task of evaluating potential courses of action against current resources (i.e. costs), effectiveness at reaching desired end results, and the potential for causing undesired effects. As stated earlier, an understanding of these follow-on and potentially unwanted effects is critical to making informed decisions. The goal of the Genetic Algorithm presented in this paper is to provide decision makers with insight into the cost of different course of action along with the potential direct and indirect effects of such action. Here a Genetic Algorithm analytical capability is added to the infrastructure interdependency simulation environment.

Genetic Algorithms

Using techniques based on the Darwin theory of evolution, GAs can be used to evolve solutions to problems with large search spaces, such as interdependent infrastructure network optimization. They “form a class of probabilistic optimization methods that are inspired by some presumed principles of organic evolution [3]”. GAs mimic the natural process of evolution to “evolve” solutions.

As in nature, a GA uses a population which is followed by generations of variations of the initial population. Fitter individuals are selected for producing offspring, so that the population becomes more “robust” as time goes on. In nature, the animals with the superior survival abilities live long enough to reproduce. In many cases, selection is also performed when the stronger males fight the others for the

right to reproduce.

This selection process can be mimicked in a genetic algorithm by representing the desired trait in each individual through a fitness function. Individuals with the best fitness scores are chosen as parents with some randomness. Mutation adds additional randomness to the combination of the two parents’ traits. The theory is that each new generation inherits the better traits and is therefore, in general, more fit than the last.

3. CIMS[®]

This research takes advantage of and expands upon work performed at the Idaho National Laboratory (INL), the Critical Infrastructure Modeling System (CIMS[®]) [4-5]. CIMS[®] was developed to examine the interrelationships between infrastructure networks and more specifically, the emergent systems behaviors that develop when one or more nodes within the system are perturbed.

CIMS[®] provides a highly visual and interactive environment for observing the cascading effects and consequence of infrastructure perturbations. CIMS[®] software simulates the interaction among infrastructures and enables multi-tier cause-effect analysis through a graphical (3D) representation of key infrastructure elements and the associated relationships. Through this visualization, a greater understanding of the emergent behaviors is achieved.

Developed with a simulation and gaming approach to modeling, CIMS[®] software gives decision makers the ability to:

- Model and visualize interdependencies;
- Quickly construct infrastructure models using map images, satellite photos, and other electronic images;
- Drill down and extract/change properties of individual infrastructure elements;
- Tie node behavior directly to live sensor input;
- Link active information to simulated entities (i.e., web page links, text documents, video streams, and custom programs);
- Visualize consequence and damage effects of events.

Given the potential size and complexity of networks, however, visual analysis methods may not suffice. Additional search and analysis methods are required to identify event–effect relationships especially across multiple infrastructures. Therefore the INL and University of Idaho are using artificial intelligence (AI) techniques to help refine the search space and identify subsets of possible interactions. This phase of research is just commencing.

4. INFRASTRUCTURE ANALYSIS GA

Infrastructure Protection

This research has integrated a GA with the infrastructure modeling software, CIMS®, for infrastructure planning. This integration was achieved by allowing the GA to access and effect the simulation agents' attribute and state values [6].

This GA is being developed for integration with CIMS© to determine the optimal infrastructure assets to protect from attack or restore in a disaster situation. This will define the critical sub-network for the infrastructure of concern given information such as:

- List of critical assets;
- Relative importance of each infrastructure asset;
- Cost to protect the individual assets;
- Cost to repair the individual assets;
- Cost to destroy the individual assets; and
- Time to repair the individual assets.

The GA uses this information to evaluate the resilience of infrastructure configurations by using methods such as disabling arbitrary assets and letting the infrastructure stabilize through the CIMS© simulation. This can help determine the optimum (or ranking of) assets to restore or protect from attack or other disaster.

The goal of this research is to address the following areas:

1. Find the critical sub-network(s).
2. Find ways to mediate damages (which nodes can be protected before attack/accident/natural disaster, or which nodes should be restored first after the event).
3. Identify weaknesses in the network.

The GA algorithm is as follows:

- a. Create initial population of individuals with assets randomly destroyed.
- b. Initialize state of asset agents in simulation as state in GA individual i.
- c. Run simulation to steady state or predetermined number of time steps.
- d. Evaluate fitness of infrastructure configuration based on operational asset values before and after simulation.
- e. Select individuals out of the population to perform crossover and mutation for reproduction.
- f. Repeat until steady state is reached (convergence), an acceptable fitness value is obtained, or for predetermined number of generations.

This automates the infrastructure analysis process by allowing the GA to randomly test different infrastructure

state combinations to identify harmful combinations that may not have been apparent.

By selecting infrastructure combinations that have individuals with higher importance values and lower protection costs identified, along with individuals with low priority and low recovery costs; the solution will be an individual with the most crucial and cost effective assets selected for protection/reinforcement.

Each GA individual bit array is randomly assigned a state value that corresponds to each infrastructure asset being modeled in CIMS. A 0 indicates that the asset is nonfunctional; a 1 indicates that it is operational. A population of bit arrays with these asset combinations is initially generated and then evolved to promote the reproduction of bit arrays with the most crucial assets selected. The GA should identify select over multiple generations an asset combination that contains the most critical assets (as defined) along with those assets necessary to support them.

EBO

This GA could also support a DSS for infrastructure planning and decision analysis among multiple criteria multiple objective searches. The goal of the DSS system is to provide the decision maker with a list of course of action options that achieve strategic goals while minimizing/avoiding undesired side effects of such actions.

The CIMS framework is integrated as a means to explore and examine potent causal effects that may not be readily evident. Thus the system supports an effects based analysis of the action.

The input information differs from the above in that the strategic objectives must be defined by associating (possibly weighted) values to the infrastructure assets, i.e.:

- Desired end state of each infrastructure asset
- Importance of each end state

The simulation can then be run against the COAs. The GA will optimize COAs based on the positive and the negative effects of each action.

The GA supports EBO by promoting the infrastructure combinations that have the desired effect initially and after the simulation run. In this case, the GA favors is the infrastructure combinations that have individuals that contain the target state values. Importance values and costs can also be associated with each asset. This can be used to identify an initial infrastructure state that will cascade to the desired state. This could be used in emergency response to identify which assets to restore to most quickly restore the most widespread and critical assets.

5. TEST SCENARIO

In order to test the usefulness and validity of the DSS GA for infrastructure planning, an initial test scenario was created. The first questions to answer are:

1. Given the visual nature of the CIMS[®] output, is a DSS even necessary? (Are answers unintuitive?)
2. If so, what functionality best complements the CIMS[®] output?
3. How accurate are the GA's recommendations?
4. Can the GA produce useful results in a reasonable time frame?
5. Which GA parameters produce the best results?

The test scenario needed to be complex enough to obtain unintuitive effects from given actions and yet have known consequences. A test scenario based on data from the city of New Orleans fit these criteria because enough data was available to generate a complex simulation and the events around the Hurricane Katrina provide actual consequences to compare results. The test simulation included major buildings, water, and power infrastructure in New Orleans gathered from openly available data.

GA Parameters for Initial Testing

The GA parameters used for initial testing are defined in table 1.

The fitness function is the weighted sum of the functional infrastructure asset's importance value minus the cost to restore it. It evaluates the fitness of the individual, feeds the individuals' state values to an infrastructure modeling program as an initial state of the simulation and then evaluates the fitness of the new infrastructure state and adds these two fitness values together.

This fitness function takes into account the importance value that has been placed on each critical infrastructure asset. It also includes a weight for the asset importance attribute, the cost to protect or reinforce it, and the recovery cost portions of the function. This allows the focus to either be biased toward asset criticality or cost, if desired. If the asset agent's state is 1, it contributes its importance value to the fitness value with a penalty for the cost of protecting it. If the asset agent's state is 0, its importance value multiplied by its cost to repair is subtracted from the overall fitness value for the individual. This is calculated before the simulation is run and again after the simulation has run for a specified time period in order to take into account both the state of the infrastructure assets immediately after an incident and after it has had a chance to recover. This gives infrastructure asset sub-networks that initially contain the most important assets, and those that are able to quickly

Table 1 – GA Parameters

Parameter	Implementation
Representation	Bit string – 1 denotes item is operational
Fitness Function	Weighted sum of the functional infrastructure asset's importance value minus the cost to protect it
Algorithm	Generational
Generations	100
Selection	Tournament
Selective Pressure	80 %
Elitism	4 individuals
Population Size	50
Crossover	Two-point with 90% chance
Mutation	Probability of 1/(number of items)
Assets Modeled	665

recover, a higher rank in the GA population. Penalties for asset attributes such as a high cost to recover can be included, depending on the goal of the exercise. This function will be tested and fine tuned.

This fitness function is more elaborate than that of a typical GA and may therefore have too high a time a running time for some applications.

The method for evaluating the fitness of each chromosome, j , in the population is as follows:

1. Evaluate fitness value of each asset represented by the chromosome bit array:

$$fitness(asset) = \begin{cases} sb = 1 : \alpha * i - \beta * c \\ sb = 0 : -(\chi * r) \end{cases} \quad (1)$$

Where:

sb = corresponding bit value

i = importance value assigned to the asset

c = cost associated with protecting the asset

r = cost associated with recovering the asset that could include time and money

$\alpha\beta\chi$ = independent weight values that can be adjusted to affect results by favoring or downplaying the associated attribute

2. Sum the fitness values for all assets:

$$FitnessBeforeSim = \sum_{i=1}^{\#assets} fitness(asset_i) \quad (2)$$

3. Set the state of the assets in the simulation to the corresponding value in chromosome j :

$$\sum_{i=1}^{\#assets} State = bit_i \quad (3)$$

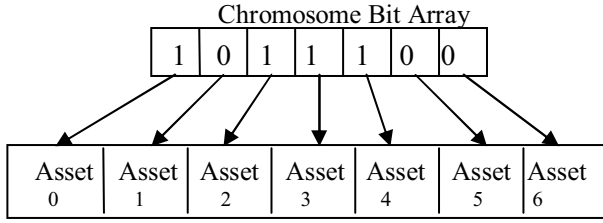


Figure 2. Simulation initialized by GA chromosome bit array values

4. Run simulation n time steps.
5. Set chromosome bit to state of assets after the simulation run. See Figure 3.

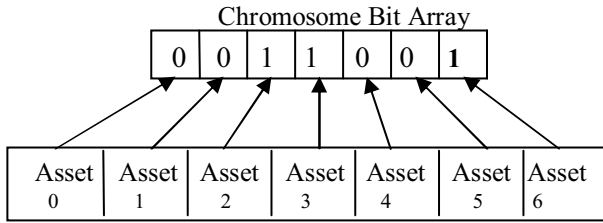


Figure 3. GA chromosome bit array values set by Simulation state.

6. Sum fitness for all assets:

$$FitnessAfterSim = \sum_{i=1}^{\#assets} fitness(asset_i) \quad (4)$$

7. Sum fitness of chromosome before and after the simulation:

$$Fitness = FitnessBeforeSimulation + FitnessAfterSimulation \quad (5)$$

Preliminary Results

Initial results proved that the integration with CIMS[®] was successful. The GA was able to evaluate different infrastructure configurations and suggest a set of assets to protect. Further tuning of the GA parameters and fitness values is needed to provide better results, however.

Analysis time to calculate fitness values was not unreasonable. Further testing will include larger infrastructure and more generations for better analysis. For large infrastructure tests, this application is well suited for parallelism because the simulations for each individual can be spread across multiple processors with only the chromosomes (bit array values) and asset state values

needing to be transferred.

6. CONCLUSION

This application of GAs to Infrastructure modeling and analysis can be applied to the problem of national infrastructure protection. The GA can be configured to optimize COA to achieve desired effects, such as strengthening the Nation's most critical assets or responding to disaster situation.

Initial testing of the infrastructure analysis GA proved to be promising. The next step in this research is to test and tune the GA parameters to achieve the best results and compare these answers to other methods and proven outcomes.

REFERENCES

- [1] Air Force Doctrine Document 1, "AF Basic Doctrine", 17, Sept. 2003, pp. 18.
- [2] Lieutenant Colonel Thomas X. Hammes, "The Evolution of War: The Fourth Generation," Marine Corps Gazette, Vol. 78, No. 9, Sept. 1994, pp. 35-44.
- [3] Beyer, Hans-Georg and Schwefel, Hans-Paul and Wegener, Ingo. "How to analyse evolutionary algorithms", Theoretical Computer Science 287, (2002) pp. 101-130.
- [4] Dudenhofer, Donald D., Permann, May R. and Manic, Milos, "CIMS: A Framework for Infrastructure Interdependency Modeling and Analysis", Proceedings of the 2006 Winter Simulation Conference, IEEE, December 2006, pp 478– 485.
- [5] Pederson, P., D. Dudenhofer, S. Harley, and M. Permann, "Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research", https://www.pcsforum.org/library/files/1159904563-TSWG_INL_CIP_Tool_Survey_final.pdf, INL/EXT-06-11464, Idaho Falls ID, August 2006.
- [6] Permann, May, "Genetic Algorithms for Agent-Based Infrastructure Interdependency Modeling and Analysis", presented at the Spring Simulation Multiconference, Norfolk, VA, March 25–29, 2007, Paper 28.