The State of the Art in Critical Infrastructure Protection: a Framework for Convergence

Ebrahim Bagheri a,*, Ali A. Ghorbani a

aFaculty of Computer Science, University of New Brunswick, Fredericton, N.B., Canada

Abstract

The protection of critical infrastructure systems has recently become a major concern for many countries. This is due to the effect of these systems on the every day life of all citizens and the high possibility of disruption because of their complex structure and hidden interdependencies, which subsequently attracts the attention of many researchers and scientists. The investigations of researchers have encompassed issues of national security, policy making, infrastructure system organization, and behavior analysis and modeling. In this paper, we look into the latter subject and explore the attempts that have been made. Based on the available schemes and the requirements of this area, we propose a five dimensional framework that introduces the major research necessities in this field. Among the various available schemes, we study ten of the most recently developed and/or influential systems. A comparison of these schemes based on the features of our proposed framework is made. The comparison allows us to conclude our examination with the identification of current research strengths, and guidelines for future work.

Key words: Critical Infrastructure Protection, Modeling and Simulation, CIP Framework

1 Introduction

Critical infrastructures are among the most significant technical systems that influence the ordinary life of any person or the normal operation of any industrial sector. Their importance is mainly due to the type of facilities/utilities that they provide. These facilities (either be in the form of asset supply or service provisioning) serve as the building block for any other simple or complex...
functionality of the society. The outputs of the infrastructures although complex in nature, can be thought of as the essential atomic inputs to other more complex systems. Apparently, without the proper operation of infrastructure systems, the function of other reliant systems would be disrupted.

The very interesting point is that throughout the years, infrastructures themselves have become dependent on each others’ outputs, turning the so vertically integrated systems with only a few points of communication, into horizontally integrated systems with many points of interaction in many of their dimensions [Donzelli and Setola (2006)]. Analogous to the dependency of other systems on infrastructures, it can be observed that infrastructure systems themselves are inter-reliant or in other words tightly coupled. As has been extensively studied in the field of fault tolerant computing, a complex system built from many interacting components is exposed to a high risk of failure derived from the possibility of mal-function in any of its components. The degree of effectiveness of the failing component in the overall architecture (e.g. in a digital circuit it can be thought of as the number of input/output connections of a specific component), suggests an estimate of the degree of damage or harm that its failure will cause.

With the increase in the complexity of the system, the approximation of the degree of effectiveness of each component becomes a time-consuming, cumbersome, complicated, and at times impossible task. Surprisingly, other than the visible mutual effects of system components on each other, there are usually unexpected side-effects from a component failure. A high load on a transistor in a digital circuit, not only raises the possibility of its own failure, but can also be a source for extreme heat generation in the circuit. The diffused heat can melt parts of the circuit wiring, causing a mal-function in a different section of the circuit. This type of failure would have never happened nor anticipated under normal conditions or even normal break down of the components, making the fault detection and recovery process a hard mission.

The high interdependency of infrastructures makes their characteristics some what similar to the explained systems. Their extreme inter-connectedness makes one think of them as different components of a single network. Comparable to digital circuits, a failure in a node of this complex network of interdependent infrastructures, results in catastrophic failures; many of which had not been foreseen. These failures are in many cases the result of the propagation of failure through these interconnected systems. Failure propagation is known as the *cascading effect* or *ripple effect* and has been the inspiration for many fruitful research efforts [Rinaldi (2004)].

Researchers have pursued two main directions of investigation for studying the structure and behavior of critical infrastructures, each of which has been to our belief successful. The researchers in the first group have been mainly
involved with the study, analysis, and understanding of the infrastructures’ current makeup. Their goal has been to identify methods, techniques, tools and schemes for describing the current status of an infrastructure. Based on this understanding, these researchers exploit various vulnerability, risk and/or threat assessment methods to gain more insight into the operation of an infrastructure. A fine deployment of this process reveals many of the possible causes of failure and to a great extent their consequences. It is understandable that the result of this process would only shed light on the types of failure that are a clear result of a breakdown or malfunction of a system. It should also be noted that although many of the possible roots for failure are detected in these approaches, but not all of their consequences are visibly perceived and understood.

From the understanding gained from the study of the infrastructure organization, and the identification of their points of weakness, proper risk mitigation strategies can be proposed and ranked based on three metrics namely costliness, time-consuming, and effectiveness [Turner and Hunsucker (1999)]. Each of these metrics can be weighted and suitable mitigation strategies can be selected to enhance infrastructure safety and protection according to the priorities of the infrastructure management and their strategic directions.

The other direction of research mainly focuses on the understanding of the dynamic behavior of infrastructure systems. In this route, the investigators attempt to explore the many paths of infrastructure process and operation through which they will try to identify any causes for instability. The search mainly focuses on the discovery of the paths that introduce more risk of catastrophic failures into the system.

One of the barriers to the identification of critical incidents is that they are mostly of low probability - high impact nature [Dunn (2005)]. For understanding and identification of the behavior of an infrastructure under these conditions, two main approaches have been taken. The proposal of different mathematical models using differential equations have been one option to simulate the behavior of these systems. The great advantage of these models is their accuracy and preciseness. However, as the complexity of the system grows out of a certain degree, devising a suitable mathematical representation of the system does not seem to be an easy task. To overcome this difficulty, the task of analyzing infrastructure behavior as a whole can be turned into the analysis of the aggregate behavior of many smaller interacting models. Research in this line has exploited the notion of multiagent systems. In these models, the behavior of a set of infrastructures is studied through a multiagent system, where each of the agents represents an infrastructure or an infrastructure system or asset. The collective behavior of all these agents within an integrated environment allows the analysis of infrastructure systems response to extreme events.
In the next section of this paper, ten critical infrastructure modeling or analysis schemes are introduced and briefly explained. Section 3 then introduces our proposed framework in detail. The schemes introduced in Section 2 are compared based on the dimensions of the proposed framework in Section 4. Discussions and future work are presented in Section 5, and the paper is finally concluded in Section 6.

2 Critical Infrastructure Analysis Schemes

During the past few years, many different schemes have been developed with the aim of supporting the process of understanding the behavior of an infrastructure and identify its points of weakness. From among the available schemes, we have chosen ten (however we have considered ASPEN and its descendants as one system. The same has also been applied to the CASCADE and coupled CASCADE models), so that at least one pattern for each paradigm within the field of infrastructure analysis is covered. The most important criteria for selection have been: 1) the scheme should have been in the focus of research and been well cited; 2) the scheme should have been recently developed; 3) the scheme should have a distinguishing feature that makes them significantly different compared with the others; and, 4) the scheme should have been well documented and sufficient amount of literature for understanding its performance should exist. For example we have chosen, ASPEN [Basu et al. (1998)] because it has been well documented and serves as one of the first systems that have been thoroughly studied and cited in the field of infrastructure modeling and simulation. Agent-based Interdependency Modeling and Simulation, AIMS[Ghorbani et al. (2006)], and Critical Infrastructure Simulation by Interdependent Agents, CISIA [Panzieri et al. (2005)], have been introduced due to their recent development (2006 and 2005, respectively) while UML-CI [Bagheri and Ghorbani (2006b), Bagheri and Ghorbani (2007), Bagheri and Ghorbani (2006a)] is being introduced due to the unique features that it provides for profiling the organization and behavior of an infrastructure. Hierarchical Holographic Modeling, HHM [Haimes (1981)], is the representative of the group of models that have been quite extensively used and documented. Table 1 shows the list of the selected schemes in alphabetical order, along with their references. The most prominent features of these schemes are also mentioned in this table.

As it is quite visible from Table 1, although the list of the selected schemes is not at all exhaustive, but they have been chosen such that most of the research in the field would be covered. In the following part of this section, we will briefly visit the main features of these schemes. This would allow us to further compare the selected schemes in our proposed framework.
<table>
<thead>
<tr>
<th>Scheme</th>
<th>References</th>
<th>Prominent Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIMS</td>
<td>[Ghorbani et al. (2006)]</td>
<td>Extendable multi-agent based architecture, pluggable visualization and analysis modules</td>
</tr>
<tr>
<td>ASPEN</td>
<td>[Basu et al. (1998), Barton and Stamber (2000)]</td>
<td>Agent-based simulation, a thorough simulation of the U.S economy with the use of evolutionary learning techniques</td>
</tr>
<tr>
<td>CASCADE</td>
<td>[Newman et al. (2005), Dobson et al. (2004a), Dobson et al. (2004b)]</td>
<td>Probabilistic and dynamic complex system models for cascading failure analysis</td>
</tr>
<tr>
<td>CISIA</td>
<td>[Panzieri et al. (2004), Panzieri et al. (2005)]</td>
<td>Agent based simulation of infrastructures with three different types of interdependency (through incidence matrices)</td>
</tr>
<tr>
<td>GoRAF</td>
<td>[Donzelli and Setola (2006)]</td>
<td>Integrating engineering and business perspectives for identifying risks associated with enterprise interdependencies</td>
</tr>
<tr>
<td>HHM</td>
<td>[Haimes (1981), Haimes and Li (1991)]</td>
<td>Risk identification, ranking and filtering through the integration of multiple perspectives</td>
</tr>
<tr>
<td>IIM</td>
<td>[Haimes et al. (2005b), Haimes et al. (2005a)]</td>
<td>Sector vulnerabilities assessment using inoperability and economic loss impact metrics</td>
</tr>
<tr>
<td>IRAM</td>
<td>[Ezell et al. (2000)]</td>
<td>Focusing on risk modeling for scarce resource allocation to improve system surety</td>
</tr>
<tr>
<td>OGC CIPI</td>
<td>[OGC (2006)]</td>
<td>Sharing of geospatial data for emergency management and response</td>
</tr>
<tr>
<td>UML-CI</td>
<td>[Bagheri and Ghorbani (2006b), Bagheri and Ghorbani (2007), Bagheri and Ghorbani (2006a)]</td>
<td>Providing means for stakeholder and modeler communication, common understanding and knowledge transfer through a common infrastructure metamodel</td>
</tr>
</tbody>
</table>
AIMS AIMS is a multiagent modeling and simulation suite that allows its users to create models of an infrastructure and observe the behavior of the modeled system through simulations. One of the distinguishing features of AIMS is that it provides its users with a set of predefined component templates (e.g. pipes, switches, etc). The organization of an infrastructure can be built through the instantiation of these component templates into actual software agents. AIMS also provides the means for inserting transients into a running simulation for observing the outcome of the occurrence of an unexpected event. The set of these transients called scenarios, control and hence effect the operation of the simulation. Visualization and analysis modules can also be plugged into a running simulation to view and study the structure and behavior of the simulated infrastructure.

ASPEN ASPEN is an agent based micro-analytic simulation model designed specifically at the SANDIA national laboratories to simulate the US economy. Based upon evolutionary learning within the agent architecture, ASPEN has been able to simulate the behavior of simple decision making agents of the economy. The agents that have been modeled in ASPEN have been households, banks, companies and the government. It also provides means for analyzing the impact of various policies and regulations on the economy. Different economic sectors can be modeled independently or as an integrated whole within ASPEN.

CASCADE The CASCADE model depicts a scenario in which \( n \) identical components (either from the same infrastructure or from different infrastructures) operate and interact. Each of the components can tolerate a load between \([L^{\min}, L^{\max}]\), and is bearing an initial load of \( L^j \) (\( j \) denotes the component number and \( L^j \) is a random variable uniformly distributed within \([L^{\min}, L^{\max}]\)). A component in this setting is susceptible to failure if its load exceeds a threshold \( (L^{\text{fail}}) \). The load of a failing component will be evenly distributed amongst the rest of components. To model a cascading failure, an initial load is enforced on one of the components forcing it to fail. The breakdown of the component will transfer an extra burden onto the other components that may result in their collapse (depending on their current load), causing a cascading failure effect. Depending on how the setting is configured (variation in the number of components, their interconnection, degree of load transfer, recovery models, etc), this model provides a very good basis for understanding the behavior of interconnected systems under extreme failure conditions.

CISIA CISIA models the behavior of an infrastructure (or a set of interacting infrastructures) through a set of non-linear interdependent agents. Each of the agents represents a macro component of the modeled system. For the sake of generality, and to be able to encompass more infrastructure systems,
CISIA employs a very high level description of the internal logic of an agent. This description consists of only the specification of the agents' operative level (agents' stamina), requirements (agents' needs), and faults. To model the interaction of the agents (provide mutual requirements or disseminate failure), three types of matrices namely Operative Level Incidence Matrix, Requirement Incidence Matrix, and Fault Incidence Matrices are devised. Fault incidence matrices are further refined to allow the analysis of different types of failure propagation (geographical, physical, and cyber).

**GoRAF** To integrate the business and engineering points of view within a single framework, GoRAF has both employed a goal-based requirement engineering methodology and the CISIA simulator to identify the most critical resources of an infrastructure. Based on the business perspective, the Business Value metric has been devised that represents the economic and strategic losses that result from the malfunction of a resource. The Risk Value metric is calculated through the CISIA simulator and shows the probability of inoperability of an infrastructure resource. According to this framework, criticality of the resources is estimated based on these two metrics.

**HHM** In the hierarchical holographic modeling scheme, different models represent multiple perspectives of a complex system such as an infrastructure. This approach justly claims that no single perspective on a complex system is adequate for its proper understanding; hence it provides the means to identify and integrate multiple complementary decompositions of a system. These decompositions are hierarchically organized and show the subcomponents of the system in different levels. In an attempt using HHM, Sixteen different perspectives were identified which are namely Physical, Scope, Temporal, Maintenance, Institutional, Organizational, Management, Resource Allocation, Supervisory Control and Data Acquisition (SCADA), Systems Configuration, Hydrology, Geography, External Factors, Buffers, Contaminants, and Quality of Surface and Ground Water.

**IIM** The inoperability input/output model is based on the Leontief's input/output model [Leontief (1966)]. It illustrates the interdependency between different sectors of the economy (the framework in which the participating sectors interoperate and affect) by focusing the investigation on the degree of direct or indirect financial reliance between the sectors (according to the examples used in [Haines et al. (2005b)] this reliance has been calculated through the employment of the publications of the Bureau of Economic Analysis). Based on a set of initial disruptions in a sector, the model is able to characterize the cascading effect of failure. A dynamic extension to IIM allows a temporal analysis of the recovery mode.
IRAM The infrastructure risk analysis model is a four phase methodology to identify, rank and manage the extreme risks that threaten an infrastructure system. These four phases are namely identify, model, assess, and manage. In the first phase and through the application of the hierarchical holographic model (HHM), a system is decomposed into its sub-components. The vulnerabilities and threats to the sub-components are identified and rank ordered. In the next phase, appropriate scenarios of the sub-component operations are generated and shown with the help of event trees. The created scenarios are also rank ordered. The assessment of infrastructure surety in the third phase is undertaken through five different risk measures. These measures allow a decision maker to see the expected value of damage, understand the behavior of low probability-high impact events, and assess the surety of the system. Finally in the last phase, based on the analysis of the previous phases, a decision maker can make qualitative and/or quantitative judgments of the current situation and take appropriate steps to reach an ideal state.

OGC CIPI Critical Infrastructure Protection Initiative (CIPI) is a program coordinated by the Open Geospatial Consortium (OGC). The CIPI pilot project explored the possibility of emergency management through data sharing at different levels of the government. It provided a basis to share geographical information and make different standards interoperable. It also developed a draft interface specifications for emergency alert notification. This interface provides the means for geospatial software interaction and hence integration with emergency alert systems. The main goal of CIPI is to improve interoperability between different sources of information (telecommunication, water resources, oil and gas, government, transportation, emergency response, electric power and health services infrastructure) in order for them to collaboratively detect, prevent, plan for, respond to, and recover from natural vulnerabilities and human threats.

UML-CI UML-CI is a model driven architecture-based design [Frankel (2003)] that provides its users with the high level metamodels to create a profile of an infrastructure system. Its development has been based upon the six dimensions of critical infrastructure specification. The advantages that stem from the notion of infrastructure profiling are that it gives initial insight for infrastructure analysis and system identification, provides sound basis for common understanding, communication, and knowledge transfer, it also allows the documentation of best practices and infrastructure metamodels. The models created based on the UML-CI metamodel can be re-used as the base or guideline for the creation of new infrastructure systems. It is also feasible that simulation models of an infrastructure system be built from the models created in this metamodel.
Based upon the understanding gained from the study of many critical infrastructure modeling and analysis schemes among which we have only introduced a subset in this section, it can be perceived that although many of the aspects of an infrastructure organization, behavior, its risks, threats and vulnerabilities are thoroughly studied, but these attempts have not been in a complementary fashion. Without a framework that methodically identifies the critical points of research in the realm of critical infrastructure protection, a complete understanding of infrastructure systems is hard to reach and most results would only partially reveal an insight.

In the following sections of this paper, we attempt to formulate an integrated framework to consolidate current research attempts into a unifying structure. We believe that through the exploitation of the proposed framework a full lifecycle methodology (abstracted away from any technique) for critical infrastructure conception can be created. To operationalize the framework, investigation teams should appropriately exploit the dimensions of the framework through the proper selection of algorithms, methods, and procedure for each of the dimensions. We also introduce and provide some high level guidelines for method selection in each of the phases of this framework.

3 The Framework

Many researchers define critical infrastructures as networks of highly complex adaptive systems that can be classified as socio-technical systems (See Figure 1)\(^1\) [Rinaldi (2004), Dunn (2005)]. This definition at least shows two of their major facets: social, and technical. For this reason, and to fully understand the operation of infrastructure systems, various aspects of them should be comprehended from different perspectives. A multifaceted strategy requires an integration of many different techniques that are in most cases from a totally different background. The consolidation of these methods and schemes can be achieved through a unified framework. In this sense, the framework is responsible for specifying the role of each technique, the degree of their contribution, type of incorporation into the framework, and their temporal involvement. For our purpose, we propose a five dimensional framework that allows a step by step analysis of an infrastructure behavior and the identification of its interdependencies (either visible or hidden) on other infrastructures. The first two layers of the proposed framework are shown in Figure 2.

The first dimension of the framework, system analysis, provides the suitable perspective for understanding the organization of an infrastructure. This per-

\(^1\) The original sketch of this figure has been derived from http://necsi.org/projects/mclemens/ and has been modified for our purpose.
Fig. 1. Critical Infrastructures as Complex Adaptive Systems (CAS)

spective is mainly concerned with the business aspects of an infrastructure such as infrastructure components, how they are connected, their functionality and responsibilities, their visible mutual interdependencies (such as commodity trade, or geographical proximity) and their probable operational risks. The main common feature of the tasks related to the revelation of these sorts of information, is that there is no explicit need for dynamic models (e.g. simulations) to assist their disclosure. They can be identified through ethnographical methods like suitable questionnaires, expert round-tables, thorough brainstorming, and etc.

Although system analysis models are able to give valuable insight into infrastructure operation, they are not capable of identifying hidden interdependencies and/or concealed infrastructure behavior under certain conditions. Gaining this type of knowledge requires an additional approach to the study of infrastructures. These types of information should be either acquired through long-term observation of the infrastructure operation or by the computer simulation of infrastructure functionality. The first approach is easily feasible, but possibly at a great expense. The occurrence of unknown events is rare and may have a very damaging effect on both the infrastructure system and the economy as whole. So, although pure observational methods are feasible, but they are not recommended. On the other hand, computer simulation models, if properly devised, can give good insight into infrastructure behavior.

Monitoring the sequence of actions in a simulation provides the investigators
Fig. 2. The Proposed Five Dimensional CIP Framework

with a wealth of information on how the infrastructure as a whole interoperates with other infrastructures. It also provides valuable information on how the internal components of the infrastructure function. The techniques used in this realm are categorized under *behavior analysis* in our proposed framework.

Relying only on the human inference mechanism for extracting all of the patterns within a simulation is not sufficient for the study of infrastructures behavior and interdependencies. This is because although humans possess the most powerful inference engine that can mine almost any kind of pattern, but the time scale of a simple infrastructure operation simulation can easily span a couple of years (this can in simulation terms be a few hours of real time simulation execution). The large amount of data generated as a result of the simulation requires additional techniques that can use simulation logs to infer appropriate knowledge. These methods can reveal many hidden interdependencies (*knowledge discovery*), and expose much other information such as market behavior, opponent strategies and etc.
Other than the extracted patterns and knowledge through knowledge mining techniques, many patterns cannot be discovered by computer techniques and algorithms. This is simply because these methods are neither fully scalable nor general enough to find all of the hidden information within the logged data. For this reason and to make sure that all of the required knowledge (information\textsuperscript{2}) is provided to the researchers, data visualization techniques should be employed. These techniques would allow the investigators to find any patterns that have not been discovered in the previous phases.

Besides these types of activity that discover most of the threats and vulnerabilities of critical infrastructure systems, and provide the basis for proper decision making and management, suitable information sharing schemes should also be devised. These mechanisms would allow different parties to interact and provide each other with valuable data that results in precious collective information that wouldn’t have been otherwise discovered. This information can help detect and manage unpredictable catastrophic failures and/or provide opportunities for better infrastructure management.

In the following parts of this section, we will more carefully visit each of these dimensions and classify the sort of analysis that can be performed in each dimension. We will also briefly introduce samples of relevant techniques that may be useful in each of the respective dimensions.

3.1 System Analysis

Infrastructure system analysis is mostly involved with the identification of the structural and business aspects of the systems involved in an infrastructure. It attempts to depict a clear picture of the current infrastructure organization through the identification of its different components. The identification of system components facilitates the specification of their requirements and productions which allows an initial recognition of the set of interdependencies. Risk analysis models can also roughly give an estimate of what kind of threats, risks, and vulnerabilities endanger the proper operation of an infrastructure. The aggregate understanding attained in this layer needs to be well documented for consistent re-visits, mitigation strategy selection, guideline extraction, and managerial decision makings.

\textsuperscript{2} We focus more on knowledge than information since information is ‘data’ arranged in a meaningful fashion, while knowledge is the conclusion derived from information.
3.1.1 Infrastructure Organization Elicitation

Understanding infrastructure organization requires an analysis from multiple viewpoints. Other than the technical structure of how the system components are structured, complementary arrangements such as their social settings, human resource hierarchy, enterprise architecture should be realized. To address this issue, the hierarchical holographic modeling scheme, has advocated the fact that only a multidimensional analysis can give enough insight into the operation of a complex system. Figure 3 represents a sample HHM model for a data retrieval and management system.

By exploiting the aggregate knowledge derived from these models, depending on the degree of elaboration in each of the perspectives, one can claim that a correct overall understanding of the infrastructure organization has been reached. For example, to fully understand a telecommunication infrastructure an understanding of how technical systems operate is not sufficient. Other aspects such as regulations, international connectivity strategy, information systems, and models of communication services should also be considered. In each of these aspects more viewpoints can exist according to their nature. In the information systems category, for instance, management information systems, billing software and SCADA systems should be separately analyzed.
Fig. 3. A Sample Hierarchical Holographic Model of a Data Retrieval and Management System [Haimes et al. (1995)]
3.1.2 Apparent Interdependency Identification

According to the classification given by Rinaldi et al [Rinaldi (2004)], critical infrastructure interdependencies can be either physical, cyber, geographical, and/or logical. Among these four types of interdependencies, physical and geographical interdependencies seem to be more apparent and can be more easily identified (although an exhaustive list of interdependencies may never be reached).

The identification of the physical interdependencies can be achieved through the examination of the requirements and productions of each system component. For example, the electricity infrastructure (specifically its generators in this example) is physically dependant on the water and sewage infrastructure for consuming water for cooling purposes. Geographical interdependencies are also easily perceivable, but not so effortlessly understandable.

Any two infrastructures or their systems that are within a geographical proximity have geographical interdependency. By analyzing the geographical coordinates of two different infrastructure systems, their geographical interdependencies can be revealed. In the route of this process two major obstacles exist. Firstly, not all infrastructures in a geographical region are known to every body. There are usually visibility levels that are devised for security concerns. For example, a nuclear power plant close to a radio transmission station may not be visible to the operators of the telecommunication company that own the radio transmission station; however, the reverse may not be the case. In these situations the identification of geographical interdependencies is either not possible or is only achievable through a higher level authority (that receives classified information through information sharing schemes). Secondly, even if all of the geographical interdependencies are well known, still their consequences cannot be easily understood. This is mostly because the internal organization of infrastructures is veiled from outsiders. For this reason, the consequences of an extreme event or failure in a system cannot be perceived by other infrastructure owners and therefore, its consequences on other infrastructure facilities is also unidentified. The only solution to this dilemma is to provide some sort of information sharing mechanism between different infrastructure stakeholders.

Cyber and logical interdependencies seem to be more complicated compared with the other two types of interdependencies. For the exploration of cyber interdependencies, the type of information that is passed between different systems needs to be identified. If all information provisioning and data transfer activities between the infrastructure systems took place through specified information systems, then a similar analysis to the physical interdependency identification over the information systems would reveal cyber interdependencies. But this is currently not the case where no exact port for information
transmission exists; therefore, cyber interdependencies cannot be so clearly identified.

Logical interdependencies are also inherently concealed among the many relationships, resources, and interdependencies of infrastructure systems. We will more carefully visit these concepts in later sections.

### 3.1.3 Risk Management

Risk management is the practice of measuring and evaluating the degree of risk assigned to the normal operation of any process and usually consists of three major steps: risk analysis, assessment, and evaluation [Boehm and DeMarco (1997)]. Risks in the operation of a critical infrastructure can arise from many different sources (have various causes). These causes can be due to natural disaster, human error, terrorist act, and/or the complexity of the operating system [Narich (2005)]. To analyze the risk of a system, the risks assigned to its constituent components (operations and resources) should be analyzed. The most important concern here is to identify the sources of a risk, its associated consequences and allow for the recognition of proper mitigation strategies. Risk management techniques mostly integrate into the techniques that elicit system organization, and based on that information, identify threats and vulnerabilities. These techniques should be able to provide the means for risk identification, ranking, mitigation and filtering. Some of the major techniques of risk analysis are Infrastructure Risk Analysis Model (IRAM), Failure Mode and Effect Analysis (FMEA), Failure Modes, Effects and Criticality Analysis (FMECA), Fault Tree Analysis (FTA), HAZard and Operability Analysis (HAZOP), and Cause Consequence Analysis (CCA) [Aagedal et al. (2002)]. These techniques are not restricted to the field of critical infrastructure protection; rather most of them have been adopted from other fields such as security critical system design, and software reliability analysis.

Some of the recent research has focused on the notion of adaptive risk management [Ulieru and Worthington (2005)]. These sorts of risk management systems claim to be capable of learning, adapting, preventing and responding to dynamic threats that have not been previously defined through legacy risk management techniques. The basic idea arises from the concept of biological defense mechanisms widely visible in nature and borrow techniques from the fields of artificial intelligence, adaptive control theory, and evolutionary algorithms.

### 3.1.4 Infrastructure Profiling

The proper understanding of a critical infrastructure organization, its components, interdependencies and the attached risks, requires a systematic doc-
ocumentation schema (a unique metamodeling structure). The major benefits of this schema would be that it provides a base recognition and an initial infrastructure understanding for the teams that have started to model an infrastructure. It also caters suitable grounds for common understanding of the requirements of the modeling effort among the investigators, infrastructure managers and stakeholders. With the employment of a metamodel and based on the common understanding, enhanced communication among the role players would also be achieved. A unique shared metamodel among different investigators, accommodates easier transfer of current understanding and existing knowledge to other teams, and hence makes the research efforts synergistic. It also provides the possibility to transfer best practices among different research teams.

A created infrastructure profile based on a standard schema allows a consistent documentation of the current infrastructure state, and it also facilitates the re-use of the existing profile through extensions for other purposes. UML-CI is an infrastructure profiling mechanism that provides five major categories of metamodels namely: Ownership and Management, Structure and Organization, TRV, Resource, and Relationships. It provides a comprehensible graphical notation created based on the Unified Modeling Language (UML). One of the major benefits of employing the UML-CI profile is that its metamodels can be later transformed into proper simulation components for infrastructure behavior analysis.

3.2 Behavior Analysis

It was previously discussed that the analysis of the behavior and performance of an infrastructure (or even the collective behavior of a set of interacting infrastructures) can be achieved through either observational techniques or computer simulations. The evident fact is that observational techniques are too cost ineffective to pursue. It is only rational to follow such practice only if historical data on the operation of an infrastructure already exist and have been readily collected. This means that the costs associated with data collection have been paid, so it would be illogical not to use this data. The techniques used in the analysis of this sort of data would be similar to the methods employed in simulation log file analysis. For this reason, we only focus on the aspects of infrastructure behavior analysis that are some how related to infrastructure operation simulation here, and leave the analysis of historical data to later sections.
3.2.1 Simulation Control

A simulation is an analytical model that is meant to imitate the actual behavior of any intended system (or a set of systems). The various types of computer simulation for infrastructure behavior analysis (to the extent of the authors knowledge) have mainly been focused on two different approaches. They have either been in the form of mathematical equations with changeable variants (change can be dependent upon different factors such as context, time, state, or etc), or a set of interacting agents that represent instances of real-world role-players, system components, or resources. The inspection of the simulation flow and paths brings about new understanding of the possible routes that an actual infrastructure system may follow.

The key issues related to the execution of a simulation are related to the suitable selection of the simulation scenarios, and the context variables that are used while the model is been examined. An inappropriate choice of scenarios or unsuitable selection of context variables deteriorates the efficiency of the simulation results. In this section, we will more closely look into the two models that tackle these issues.

Monte Carlo Simulation

Monte Carlo simulation is an approach for repeatedly analyzing a deterministic model through the employment of random input values for the uncertain variables of the model [Rubinstein (1981)]. The technique is widely used for simulating non-linear complex systems that have a range of undecided parameters. Through the employment of random values a stochastic simulation of a deterministic model is created. The aim of the simulations performed through the Monte Carlo technique is to study the effect of variation, uncertainty, fault and/or attack on the performance and/or reliability of the system. The nature of this technique matches well with the character of the problem we are trying to undertake. Understanding the behavior of critical infrastructure systems under unknown conditions can be studied through the employment of Monte Carlo simulations. The random values selected for any parameter of the infrastructure system models in a Monte Carlo simulation is hence based on a predefined probability distribution, within a specified range. The probability distributions are usually selected from amongst the normal, triangular, uniform, and log-normal distributions.

Scenario Building

Scenarios are tools for creating alternative scenes based on the different arrangements of the initial hypotheses about environmental parameters and sys-
tem settings [Zhang and Bose (1989)]. They are usually built in cases where more knowledge or understanding of a situation or system is required. They allow the observation of different perspectives on the same system by giving new viewpoints. In a computer simulation, a scenario specifies the initial setting of the model, and provides guidelines on what paths the simulation has to follow. Similar to the selection of the initial simulation parameter value, creating appropriate scenarios for the simulation of the behavior of an infrastructure is of utmost importance. Since many of the interesting (and at many times harmful) events within an infrastructure system are unpredictable and rare, scenarios should be able to design simulation routes that cover most of these events. A scenario that guides a simulation into the normal operation of an infrastructure does not seem to be so valuable. It is only the scenarios that help the investigators reach new understanding that are precious. Scenario graphs [Mathew et al. (2005)] and event trees [Papazoglou (1998)] are two of the models that can be used to create scenarios. It is obvious that creating a scenario graph or an event tree by hand that incorporates all of the different possibilities of a simulation execution for a complex system such as an infrastructure is both tedious and impossible. Therefore, appropriate predictive techniques should be used to create such graphs or trees based on the initial understanding that has been gained through the employment of the methods introduced in the previous sections.

3.2.2 Simulation Models

For any type of system, understanding its interaction behavior with other systems requires an initial perception of its internal structure. Infrastructure systems are no exception to this matter. To fully understand how an infrastructure would react to an outside stimulus, a thorough knowledge of its internal functions and organization is required. In the following, we will study how each of the two mathematical, and multiagent approaches support the simulation of infrastructure behavior.

Mathematical Models

Among the various infrastructures, electricity infrastructure has been the focal point of mathematical simulation models. These models have mainly focused on modeling the behavior of a certain component or aspect of the infrastructure. For example, different mathematical schemas for predicting the electricity demand level and price have been developed. [Joskow and Kahn (2001), and Nogales et al. (2002)] are just a few to name. Other attempts have aimed at developing mathematical models for eliciting hidden interdependencies between different sectors of the economy (or interdependencies between infrastructures). The inoperability input/output model (IIM) introduced earlier in
this paper, is one of the most prominent examples of these efforts.

Although the results of mathematical models of infrastructure analysis is highly reliable, they lack scalability features [Amin (2001)]. This is due to the fact that infrastructures are fairly complex systems with many intricate details that rapidly change. Designing a monolithic mathematical model that incorporates all these rapidly evolving specifications is almost impossible. This has provoked researchers to look for more dynamic solutions that can integrate with the available mathematical models.

Multiagent systems can incorporate suitable mathematical models into appropriate agent logic components. Each of the agents would then behave based on its internal mathematical representation.

**Multiagent Systems**

Multiagent systems provide a collective understanding of a very complex system through a unified view on the aggregate behavior of their constituent elements (agents). Infrastructure systems can hence be very well modeled by multiagent systems. Depending on the modelers understanding of the structure and organization of an infrastructure, appropriate multiagent systems can be created to mimic the actual behavior of an infrastructure.

There are two main issues to a proper design for a multiagent based architecture: agents internal design, and agents environmental commitments and collaborations (See Figure 4). The first issue addresses the internal representation of each agent within the multiagent architecture. This aspect specifies what functionalities/services an agent requires/provides, what resources it owns/consumes, and what goals it pursues. These specifications would allow an agent to decide on its reactions under various conditions. For example, an electricity generator can be modeled as an agent in a multiagent architecture. If the generator goal is to maximize its profit, it would regardless of its current load, accept new requests, but if its objective is to provide a very high quality of service to its customers, it would prefer to accept a lower number of requests but provide them with a better service.

Investigators can use a very well known multiagent architecture known as BDI, which specifies the internal structure of an agent through three concepts: Belief, Desire and Intention [Dignum (2000)]. An agents belief is its perception of the surrounding environment. Desire is the agents motivational states or goals, and an agents intention is its current decision in order to reach his goals (desires). Agent intentions are realized through intelligent planning [Fraile et al. (1999)].
Following the previous example, an electricity generator can forecast power market’s future, and plan not to forward sale its electricity supplies because of a rise in electricity prices in the near future. As this example portrays, the sort of planning problems in a multiagent system that simulates an infrastructure system is usually complicated and have to be discovered and optimized in a multidimensional space and hence requires complex techniques.

Identifying the internal representation of all of the components of a critical infrastructure is always a time consuming task. We assume that if the proposed framework in this paper is fully observed, the required descriptions would be available in this phase; however for many research teams, it is infeasible to fully implement a complete study based on this framework. In these cases, the internal representation of an agent can either be replaced with a mathematical or conditional equation. The incompleteness of these models does not necessarily suggest that they are inappropriate, rather it implies that they are not as accurate as a full fledged implementation.

An agent’s societal interaction within a multiagent architecture is the other point that should be specified when modeling an infrastructure system. This would include features such as Negotiation, Cooperation, and Coordination [Zlotkin and Rosenschein (1991)]. Negotiation is a social practice through communication that is experienced for regulating power, resources and commitments among different agents.
Agents are autonomous by their definition (as their corresponding infrastructure systems are in the real world), but require interaction to reach their goals. As a result of a successful negotiation, cooperation can be achieved. Negotiation and cooperation procedures appear to be able to reveal a set of infrastructure interdependencies. The negotiation of agents (each agent being an infrastructure system) over each others' resources depicts their interdependencies. This allows the revelation of many physical or cyber interdependencies. However, still logical interdependencies are not revealed.

Logical interdependencies are related to a more sophisticated notion in multi-agent systems: Coordination. Coordination is conceptually different from cooperation, since two coordinated agents may not in fact cooperate with each other, however make coordinated decisions. Suppose that the price of oil is strictly dependant upon the peace process in the Middle East. Therefore, major oil producers and consumers, try to coordinate their decisions with the political developments in that region. It is visible that oil companies have no direct cooperation with the political side of the issue, but they are highly coordinated with their decisions. This coordinated behavior can be considered as logical interdependency between two infrastructure systems.

3.3 Knowledge Discovery

Knowledge discovery is the process of analyzing data which has been derived from various sources and studying their features, correlations, and patterns from different perspectives to reach a new understanding about the sources of those data. In the study of critical infrastructure systems, data may be available from two major sources. They may be either the result of a long term observational process or be synthetically generated through computer simulations. In either case, the procedure for extracting knowledge is very similar. The only difference may be related to the reliability of the data sources. Data collected from the observational process can be deemed to be more trustworthy compared with the results obtained from a computer simulation. This is because a computer simulation may have a very small invisible flaw, which causes a significant difference in the final output. In this section, we investigate two major techniques used for inferring knowledge from the available data sources.

3.3.1 Hypothesis Testing

In hypothesis based knowledge discovery, the investigators make a set of assumptions about the behavior of an infrastructure under certain conditions. For example, a research team may claim that the price of electricity is logically
dependant on the variation of climate changes throughout winter. To verify the correctness of this claim, they will try to find corresponding evidences from the collected data. The techniques used to support this process have to search all of the available data for any clue that matches a set of predefined criteria. The criteria should be precisely described through an understandable, unambiguous notation. Logical notations such as first order logic, modal logic, temporal logic, and others can be used to describe the hypotheses that need to be validated.

Like many other techniques, there are certain advantages and drawbacks associated with this approach. The major deficiency of this method is that the research team should make the hypotheses themselves. This may have two major consequences. Firstly, many of the relationships and correlations between data gathered from the operation of an infrastructure are not easily understandable, and hence guessing all of the probable associations is not feasible. For this reason, there is always the possibility that many undetected relationships exist. Secondly, the process of specifying the associations and making the hypotheses is a time-consuming task. However this approach has its own advantage and that is the fact that simple reasoning models and pattern (data) mining techniques can assist in the process of hypothesis evaluation.

3.3.2 Exploratory Data Analysis

As opposed to the method introduced in the previous section, exploratory data analysis is concerned with the methodical identification of the hidden associations between a set of available data without any prior knowledge over their essence [Tukey (1977)]. In this approach, many of the combinations of data relationships are tested through the exploitation of a wide range of statistical methods and machine learning techniques to find appropriate patterns and relationships (In the case of infrastructure systems, identifying hidden interdependencies, vulnerabilities, probable threats and risks and even business opportunities). The statistical techniques that can be employed in the study of critical infrastructure behavior may vary from simply observing parameter distributions, and variable correlation analysis to multivariate exploratory techniques such as cluster analysis, factor analysis, multidimensional scaling, log-linear analysis, canonical correlation, correspondence analysis, time series analysis, and classification trees.

Aside from the statistical models for pattern identification, machine learning techniques can also be used to make accurate predictions. These models differ from statistical models in that they do not possess any conceptual relationship with the system being modeled and hence have no theoretical understanding of the underlying causal processes. However, these techniques have the capability to generate accurate predictions over a set of patterns. Artificial neural
networks are among the well known examples of this techniques. Machine learning techniques are mainly used to make predictions of future behavior based on a set of currently available data or patterns. The major drawback of techniques such as artificial neural networks is that their predictions cannot be logically verified.

In the context of studying the behavior of a set of interdependent infrastructures, statistical exploratory techniques can be used to discover any correlation (hidden/ logical interdependency) between the infrastructure systems. On the other hand, machine learning techniques can observe the current behavior of the infrastructures and help predict their future reactions under certain conditions. While statistical models seem to be more appropriate for extracting knowledge from the simulation results, machine learning techniques show more suitability for pure prediction reasons and infrastructure behavior modeling purposes.

3.4 Visualization

In the context of infrastructure systems, visualization techniques can be used for two main tasks. They can be employed to reveal the visual structure and organization of an infrastructure (from the information which has been collected in the system analysis layer). Using these representational techniques (that can mostly be used in conjunction with the techniques introduced in the system analysis dimension) many different types of analysis can be performed (e.g. reliability analysis, connectivity, points of failure and bottleneck identification and etc.). The other usage of visualization techniques is that they provide different graphical displays of the data (raw or processed) that have been manipulated in the behavior analysis, and knowledge discovery dimensions. This would allow the investigators to spot any inconsistencies in the results of the previous dimensions or infer new knowledge that has not been extracted by automatic methods.

3.4.1 Infrastructure Organization Representation

There are various techniques that can be used to depict the structure and organization of an infrastructure. Based on our understanding of the available techniques, a suitable technique for representing the structure of a system should provide relevant methods for organization analysis, along side its figurative notations. We have selected two of such schemes namely graph theory, and layered geographical information systems that will be explained in more detail in the following lines.
Table 2
A Brief List of Graph Theory Techniques Applicable to CIP

<table>
<thead>
<tr>
<th>Technique</th>
<th>Contribution to Infrastructure Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph Coloring</td>
<td>Resource Allocation, Task Scheduling</td>
</tr>
<tr>
<td>Visibility Graph</td>
<td>Urban Planning</td>
</tr>
<tr>
<td>Clique Problem</td>
<td>Reliability Analysis</td>
</tr>
<tr>
<td>Graph Reduction</td>
<td>Reliability Analysis, and Fault Tolerance</td>
</tr>
<tr>
<td>Network Flow Analysis</td>
<td>Capacity Planning</td>
</tr>
<tr>
<td>Scale-Free Network Analysis</td>
<td>Fault Tolerance, Vulnerability and Threat Identification</td>
</tr>
<tr>
<td>Graph Drawing and Re-arrangement</td>
<td>Cartography, Organization Planning</td>
</tr>
</tbody>
</table>

**Graph Theory**

In theory, graphs are a set of nodes that are intentionally connected to each other by a set of edges [Cenek (2000)]. The connection between these nodes conveys a meaningful concept. For our purpose, various infrastructures (or their internal systems) can be represented as nodes and their interdependencies can be shown as their respective edges. Through this representation, many of the currently available graph theory techniques can be used to analyze the behavior of the set of interacting infrastructures. Table 2 shows a subset of these techniques and introduces their possible contribution to infrastructure analysis.

Network analysis techniques are a subset of graph theory. Among the very interesting notions in network analysis that has attracted a lot of attention in the recent years has been the concept of scale-free networks [Wang and Chen (2003)]. Scale-free networks are special types of network topology in which only a small subset of the nodes called hubs have an extreme number of connections (high degree of connectivity), while the other nodes do not have such a connection degree. The distribution of the node degrees in such a network follows a power law distribution. The reason why these networks have been so well studied is that they are quite resistant to random failures. For example, an electricity supplies network that has the form of a scale-free network is very rarely disrupted due to random failures because of its fairly tightly connectedness. However, the same network is vulnerable to targeted (purposeful) attacks that aim to disrupt the operation of its hubs. For this reason, the study of infrastructure organization from the perspective of scale-free networks provides the basis to perform a fault tolerance analysis. It is usually recommended that the nodes in an infrastructure network that form its hubs should be well guarded, replicated, and have suitable backups (See
Fig. 5. A Simplified Network of the U.S. Airlines System. The Red Nodes Resemble the Hubs that have a Significant Number of Connection Compared with the Rest of the Nodes. If these Nodes Fail, They will have a Much More Severe Effect on The U.S. Air Traffic [Barabasi and Bonabeau (2003)]

Figure 5).

Layered Geographical Information Systems

A Geographical Information System (GIS) is a software system that provides the functionality to develop, maintain, analyze and handle geo-spatial information [Wolthusen (2005)]. These systems are mainly developed in a multi-layered style, so that various forms of information can be integrated into a single model. The three most important features of a geographical information system that makes it appealing to the study of infrastructure systems are:

- Integrating information from multiple sources: Since a GIS is designed in a layered fashion; it can integrate other sorts of information (or even additional information) from different sources. This has the advantage that there is no need for a central team to collect information which results in more completeness, and accuracy.
- Topological modeling: A GIS can recognize and analyze spatial relationships between geometric entities that have been stored in a digital format. The typical types of information that can be inferred are entity adjacency, proximity and containment. These forms of information can vary depending on the data available within the GIS.
- Map overlay: The integration of various sorts of spatial information onto a single map creates an overlay map. This provides better understanding of the relationships between various types of information (improves the process of interdependency identification).
Fig. 6. Failure Propagation in Both Gas and Electricity Infrastructures have been Shown in Different Colors in a Geographical Information System. [Tolone et al. (2004)]

Other than planning purposes, geographical information systems are the best source of geographical interdependency identification. An overlay map provides the best means to find any geographical proximity between two different infrastructure systems. As has been practiced in [Tolone et al. (2004)], a GIS can also visualize the effect of a rippling failure through picturing the failure propagation in a different color or shape (See Figure 6).

3.4.2 Analysis Data Visualization

Tremendous amounts of information lie within unstructured data that cannot be identified by statistical or machine learning techniques. This information can only be detected by humans through the employment of proper visualization techniques. These techniques would mostly contribute to the identification of correlation between the parameters of a system, or the detection of logical interdependencies between different infrastructure systems. The most important models of data visualization include (but are not restricted to) function fitting and plotting, data smoothing, overlaying and merging of multiple displays, categorizing data, splitting/merging subsets of data in graphs, aggregating data in graphs, identifying and marking subsets of data that meet specific conditions, shading, spectral planes, integrated layered compressions, and projected contours, data image reduction techniques, interactive (and continuous) rotation with animated stratification of 3D displays, and selective highlighting of specific series and blocks of data [StatSoft (2006)].

3.5 Information Sharing

The protection of critical infrastructure systems requires a complete understanding over their internal organization and external links. All of the previ-
ous dimensions of this framework tackle this problem from the perspective of a single protection team that intends to protect a particular infrastructure. Therefore, they try to identify the infrastructures organization, discover the hazards that it faces and depict its external dependencies. But as it turns out to be this is not enough for the complete protection of the infrastructure. Many of the threats posed on an infrastructure do not actually stem from roots in that infrastructure, and have even deeper causes than mutual infrastructure interdependencies. Theses origins of threats to an infrastructure are a collective result of many single bits of invaluable information or events that once summoned together result in precious knowledge. The detection of these events requires the integration of multi-dimensional data that are collected from various sources. The compilation of this type of information necessitates the act of information sharing between all of the involved parties in the operation, management, and monitoring of infrastructure systems. A few information sharing models in the realm of multiagent systems have been proposed such as [Marsh et al. (2003), Carter et al. (2002)]; however their suitability for critical infrastructure information sharing has not yet been investigated.

There are two major barriers for achieving information sharing. Firstly, infrastructure owners are reluctant to share their private information with other parties. This is because the transpiration of their information may threaten their market and business advantages. Secondly, sharing information should be achieved through proper communication channels (e.g. based on integrated information systems) that are in accordance with proper rules, regulations, and policies. To date, this set of conventions does not fully exist and vary from country to country, and since many of the infrastructure systems operate in multiple countries, the variety of regulations impedes proper information sharing. The solution to these problems seems to lie in the hands of international cooperation of governments in order to devise proper worldwide rules and regulations and to establish trustable third party agencies. These agencies can then based on international agreements provide suitable platforms for sharing information (e.g. integrated information bases) between different parties according to their access levels and authorities. The overlay of this aggregate information can result in the prevention of many catastrophic events and even enhance the operation of many critical infrastructure systems.

Although information sharing has been introduced as a separate dimension of our framework, but correct, complete, and consistent results from the other four dimensions fully depends on the degree of information sharing among different parties. To a great extent, the proper integration of the rest of the dimensions depends upon the existence of a platform so that information can be shared and knowledge be transferred. This information sharing platform not only allows multiple infrastructures to collaborate to protect their infrastructure systems, but also provides suitable means for a higher degree of cooperation within a single infrastructure.
4 Comparing CIP Schemes

In Section 2 of this paper, we introduced ten relevant schemes to the field of infrastructure protection. The most outstanding features of these schemes were briefly mentioned in Table 1, and a short summary of their actual capabilities were also given in that section. Based on these descriptions, Section 3 introduced a framework to classify most of the related research and investigation that are currently under progress (or have reached valuable results). In our description of the different dimensions of the framework, we introduced the major concerns, research themes and open problems that are of interest to the critical infrastructure protection field. We also summarized some of the related solutions that have been proposed in the literature accordingly (in the case of availability and to the extent of the authors information). In this section, we further compare the introduced schemes based on our proposed framework.

Table 3 shows a brief comparison of the different studied schemes in this framework. The columns of the table represent framework dimensions, while
the rows correspond to each specific scheme. In every cell of the table, the contribution of the related scheme to that dimension of the framework is described. For example, according to the table, the hierarchical holographic modeling scheme provides means for system analysis. The contribution of HHM to system analysis is that it allows investigators to perform the infrastructure organization elicitation, and apparent interdependency identification tasks.

5 Discussions and Future Work

In their seminal paper [Rinaldi et al. (2001)], Rinaldi et al propose six dimensions that can be used to define the characteristics of an infrastructure. These dimensions consist of various features for specifying infrastructures interdependencies, their environment, their coupling behavior, types of failure, state of operation and characteristics. Even though this classification scheme does not provide any means for actually profiling critical infrastructure characteristics, it offers good insight into understanding the features that should be sought while studying critical infrastructure systems. For this reason, we have employed these six dimensions as a way to evaluate the degree of contribution of each of our proposed framework dimensions to the correct perception of critical infrastructure organization and behavior.

As it can be seen in Figure 8, our analysis involves a rough estimation of the degree of success of each of the dimensions to perform a comprehensive analysis of each of the critical infrastructure specification dimensions.
<table>
<thead>
<tr>
<th>Scheme</th>
<th>System Analysis</th>
<th>Behavior Analysis</th>
<th>Knowledge Discovery</th>
<th>Visualization</th>
<th>Information Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIMS</td>
<td>Agent based Simulation</td>
<td></td>
<td>Supports Machine Learning</td>
<td>Interface to third party software e.g. Matlab</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and Statistical Methods as</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Extensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASPEN</td>
<td>Agent based Simulation</td>
<td></td>
<td>Evolutionary Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASCADE</td>
<td>Mathematical Analysis</td>
<td></td>
<td>Statistical Methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CISIA</td>
<td>Agent based Simulation</td>
<td></td>
<td>Fuzzy Logic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GoRAF</td>
<td>Risk Management</td>
<td>Integration with</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CISIA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHM</td>
<td>Infrastructure Organization Elicitation</td>
<td></td>
<td></td>
<td>Tree Like Infrastructure Organization Representation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apparent Interdependency Identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIM</td>
<td>Apparent Interdependency Identification</td>
<td></td>
<td>Mathematical Dependency</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRAM</td>
<td>Risk Mgmt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGC CIPI</td>
<td>Infrastructure Organization Elicitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UML-CI</td>
<td>Infrastructure Profiling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We select the first layer, system analysis, to explain how this diagram serves our purpose. System analysis aims at clarifying business and organizational perspectives of an infrastructure system. As we discussed it earlier, the methods employed in this layer are only suitable for identifying physical and geographical (and partially cyber) interdependencies. Therefore, the area specified below the type of interdependencies feature is only partially shaded. Although system analysis only partly contributes to interdependency identification, but its techniques are the most important methods for identifying various aspects of the infrastructure environment, their coupling and response behaviors, and their characteristics; therefore, this area has been fully shaded. Since many of the states of infrastructure operation and their related failures are identified through their dynamic behavior, system analysis techniques are not so successful in identifying them and hence have not been fully filled in the diagram. The effectiveness degree of each of the framework dimensions to an infrastructure dimension is shown by a corresponding colored graph in Figure 8. Each of the framework dimensions only partially contribute to the study of critical infrastructures. Their integration allows a complete understanding of infrastructure organization and behavior.

Our analysis of the current schemes, models, techniques, and approaches for understanding critical infrastructure organization and behavior has revealed the fact that although the field is fairly new, but a lot of valuable contributions have been made. The classification of these schemes based on our proposed framework shows that a lot of these attempts have been towards either system analysis, and/or behavior analysis of infrastructure systems. But fewer efforts have been made to specify the type of knowledge discovery and visual-
ization techniques that can be used to analyze the information that have been gathered from the two previous dimensions. It would seem to be a great step forward if these techniques along with their possible contributions could be studied (for example, a classification of machine learning techniques that can contribute to knowledge discovery in infrastructure systems).

The other field that needs great investigation is information sharing. Apparently, due to its complexity, information sharing has attracted less attention. As it is clearly visible from Figures 7 and 8, information sharing serves as one of the most important dimensions for understanding infrastructure systems. Unlike the other dimensions, information sharing is an aspect whose lifetime encompasses all of the other dimensions. Its main role is to facilitate the operation of the other dimensions. To advance the efforts in this field forward, proper information sharing schemes, policies and regulations, models of information integration, and models of international collaboration should be devised.

Besides the efforts that need to be made to further advance the current schemes, other attempts to integrate these techniques should also be made. For example, although techniques for infrastructure organization elicitation such as HHM exist, but they may not be suitable for working with techniques used for risk management. Therefore, suitable integration strategies to bring these techniques together, need to be devised. A first step towards this notion has been taken in [Donzelli and Setola (2006)]. The authors propose a dynamic risk management scheme that integrates an objective based organization structure identification technique into a behavior analysis method (CISIA). Although this is a precious attempt, but still a lot of methodical research need to be undertaken so that unified integration strategies can be created.

6 Conclusions

In this paper, we have proposed a five dimensional framework to study the current efforts in the field of infrastructure protection. We have identified the major points of research in these dimensions and explained ten schemes in our study. Each of the dimensions of the framework, address different issues of critical infrastructure protection. We have also compared these ten schemes based on our proposed framework. Guidelines for the direction of future work have also been provided in this paper.
References


URL citeseer.ist.psu.edu/dignum00towards.html


Tukey, J. W., 1977. Exploratory Data Analysis. Addison-Wesley, USA.


