A Unified Approach for Reliability Assessment of Critical Infrastructures Using Graph Theory and Entropy

Mohammadreza Iranpour, *Student Member, IEEE*, Maryam A. Hejazi, *Member, IEEE*, and Mohammad Shahidehpour, *Fellow, IEEE*

***Abstract*—The growth in the critical infrastructure (CI) complexity, interdependency, and data flows has required enhanced reliability assessment methods. This paper proposes a reliability assessment model which considers various sources affecting the CI performance, including internal and external fault sources including natural disasters and malicious attacks. The paper proposes a hybrid data and model-based approach for a comprehensive CI reliability assessment, which considers fault source effects on CI subsystems and their interdependencies. The Shannon entropy concept is proposed for detecting any variations in the system information fault sources. The graph concept is used to determine interdependencies of CI subsystems. We have presented two case studies to illustrate the effectiveness of the proposed method, and compare the results with similar studies.**

***Index Terms*—Critical infrastructures, reliability modeling, entropy, big data, information measure, Shannon entropy, interdependency, Monte Carlo simulation.**

I. Introduction

CRITICAL infrastructures (CIs), such as power, gas, water, transportation and communication systems represent systems which are interdependent [1]. Recently, additional CI integrations have converted them into complex system of systems, which are characterized by large dimensionality, significant nonlinearity, and strong interactions [2]–[4]. Information and Communication Technology (ICT) and Integrated Energy System (IES) are two examples of system of systems which were studies extensively in the last decade.

In recent years, new challenges have been introduced in CI reliability assessment, which relate to natural disasters and malicious attacks, growing interdependences, complexity of data, and dynamic CI behavior. The inclusions of these challenges have resulted in several new indices and concepts pertaining to reliability, resilience, robustness, survivability, and stability in order to consider these challenges simultaneously. Although several studies have been done on these topics, a unified model has yet to be introduced in order to consider these concurring features simultaneously.

A CI is a highly interconnected and interdependent network of network that consists of thousands of devices, services, and processes with complex relationships [5]. These devices, services, and processes contain a large sum of information. Based on the information-theoretical approach, the occurrence of an event (such as a cyber-attack) can change the complexity of the information in a CI network. The CI reliability assessment is usually based on three main challenges [6] including the network structure (modeling of interdependences), network dynamic (data stream) and failure modeling and analyses. The first challenge in the CI performance assessment is the inclusion of CI model. Different studies on CI models have classifies the CI modeling and simulation approaches into six types [7], [8], including empirical, agent-based, dynamic-based, economic-based, network-based, and general. CI modeling based on its physical and functional characteristics allows CI to be partitioned into multiple layers (subsystems) in which appropriate modeling methods can be applied to each layer to represent CI behaviors and functionalities.

Considering the first challenge, [9] presented a multilayer network model that can be used to represent most types of complex systems such as CIs. Hybrid modeling approaches have proved to be the most appropriate for capturing CI complex behaviors and assessing their performances under normal and degraded conditions [2]. The idea of mapping a CIs to multiple layers was presented in [2] which considered three steps for the implementation of multilayer hybrid modeling framework. These steps include decompose CIs into different levels, develop subsystem-specific models individually, and model interactions among subsystems. The interdependency generally indicates that a proper operation of one element will depend on the proper function of some other elements [10]. The modeling of infrastructure interdependency is a relatively new area of research and analyses which is necessary for the CI reliability assessment.

The second challenge in CI performance assessment is the evaluation of the big data, which are taken from the data sources. The growth in infrastructure complexity results in information growth as CI devices, services, and processes

Manuscript received May 14, 2019; revised October 19, 2019, January 23, 2020, and May 7, 2020; accepted June 6, 2020. Date of publication June 30, 2020; date of current version October 21, 2020. Paper no. TSG-00676-2019. *(Corresponding author: Maryam A. Hejazi.)*

Mohammadreza Iranpour and Maryam A. Hejazi are with the Department of Electrical and computer Engineering, University of Kashan, Kashan 87317-53153, Iran (e-mail: mreza.iranpourm@gmail.com; mhejazi@kashanu.ac.ir).

Mohammad Shahidehpour is with the Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: ms@iit.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2020.3005862

contain a vast amount of information. This level of complexity obstructs the traditional approaches for CI reliability assessment, failure prediction, and maintenance planning. ICT solutions collect the big data, but extensive models and tools for creating knowledge from these data are still lacking required. Data mining methods should be used to predict CI failure events and calculate correct CI performance. The third challenge in CI performance evaluation and reliability assessment is the modeling and analyses of fault sources such as natural disasters, malicious attacks, and technical fault in each subsystem.

Therefore, a hybrid data-based approach is effective in a comprehensive assessment of CI reliability which considers subsystem interdependencies, big data, and various fault sources affecting the CI performance. Accordingly, traditional deterministic methodologies for stability, security and reliability assessments cannot represent the exact CI performance [11] and the corresponding challenges and advantages of hybrid model-based approaches are reviewed in [12].

Review of the Related Work: In this part, we review the previous reliability work which is categorized into three parts. First, we review the reliability modeling and present the related shortcomings when facing with a large sum of data. Second, we review the research on interdependency modeling for CI reliability assessment and present their weaknesses in large-scale cases. Third we review the research on failure sources and mechanism which affect the CIs reliability.

Common methods for CI reliability assessment include input/output modeling, network theory, decision analysis, simulation-based, agent-based, and game theory. For example, in [13], modeling and reliability evaluation of IES is introduced which is based on Smart Agent Communication (SAC). In [14], electricity and natural gas systems were combined for the reliability evaluation of electric power distribution system. Reference [15] evaluates the network reliability for electric power, railway, telecommunication, and computer networks using the game theory. These methods may be inefficient in large information cases. In such cases, we ally a data-based model for specific problems, which lead to more generic and adaptable methods such as machine learning, which often require less expert knowledge.

A variety of modeling approaches have been used for the modeling of CI interdependencies. Interdependencies are typically classified into four categories [5] which include physical, geographical, cyber, and logical. The CI failure behavior are affected by interactions among topological, physical, and operational characteristics which pose significant challenges to the application of standard risk assessment tools, which will be inadequate in assessing the levels of CI vulnerability, reliability, and risk. Reference [16] provides a systematic perspective of CI vulnerability and risk analysis. Challenges that affect the CI reliability are divided into internal and external physical sources. Here, a physical source represents elements, devices, services, or processes which are exposed to natural disasters and malicious attacks based on their occurrence probabilities. The impact of physical sources on CI reliability has been studied in different references.

In [10], based on cyber network failures, four types of cyber-power external interdependencies are introduced which are Direct/Indirect Element-Element Interdependencies (DEEI/IEEI) and Direct/Indirect Network-Element interdependencies (DNEI/INEI). By using P-table, new cyber-power reliability algorithms for considering direct and indirect interdependencies are presented in [10], [17]; but the effects of these types of interdependencies have not considered, simultaneously. Also, the implementation of this method is cumbersome in complex CI reliability assessments. Reference [18] investigates cyber-physical power system interdependences by using a Cyber-Physical Interface Matrix (CPIM) and developing reliability evaluation methods by considering external interdependencies. However, the method presented in [18] may not be inefficient in large-scale systems, which is due to the large sizes of the cyber-physical interface matrix and the resulting examined states. Reference [19] provides an overview of interdependencies and reliability in a combined ICT and power system. In addition, the effects of cyber-attacks and natural disasters on CI models and challenges for solving respective hierarchies have been presented in several studies [20]–[23].

Contributions of the Proposed Work: Considering the challenges of network structure (modeling of interdependences), network dynamic (data stream) and failure mechanism (modeling and analyses) in large-scale systems, signify that traditional deterministic methodologies for reliability assessments may not be applicable to CI performance. Moreover, there are different models and indices which are introduced in various references for representing the reliability of CI subsystems. These indices often determine the reliability of each subsystem by considering the effects of other subsystems; for example, the reliability of power systems by considering the effects of cyber or natural gas systems; But the indices, do not provide a unified reliability index of cyber-power-natural gas systems.

There are two approaches for assessing the CI reliability including the system level and the topological approach. In the first approach, the system-level reliability assessment consists of three stages: 1) assessing the probability of each component failure; 2) determining the system behavior stemming from component failures; and 3) a hybrid method for combining the first two approaches to obtain an overall system reliability index [2]. The second approach uses topological methods for describing the system behavior by ignoring physical constraints such as failure rates of systems elements [24].

Hybrid methods will combine the strengths of both approaches to yield models that accurately reflect the CI reliability. The proposed method in this paper seeks to provide a new hybrid approach for accurately calculating the CI reliability which combines the topological approach for considering different interdependencies and system-level approach for considering all affecting sources on the system reliability index. We have considered these challenges by introducing a unified approach for evaluating the CI reliability. The graph concept and information theory (specifically entropy concept) are used in the proposed framework for CI reliability assessments.

TABLE I
INTERNAL FAULT SOURCES IN CI SUBSYSTEMS

CI Subsystems	Internal Fault Sources in CI Subsystems
Power system	Random failure of power system elements (generation unit, transmission line, transformer, and circuit breaker)
	Variable power of renewable sources (wind turbines and solar generator units)
	Load forecast errors considering uncertain loads such as electric vehicles
Cyber system	Physical and cyber losses of communication and control devices.
Gas system	Random failure of gas system elements (pipelines, liquefied natural gas terminals, and compressor stations) [13]
	Gas load forecast errors.

The contributions of this paper are listed as follows:

1) A multilayer hybrid modeling method is proposed for representing three types of CI system interdependences in three cyber, gas and power subsystems. 2) A generic system performance measure is proposed for the CI reliability assessment. A graph theory is used to measure the system entropy and then a binary model is used to calculate the probability of reducing the distance between different entropy levels to zero 3) The proposed method considers the CI performance using internal failure sources in each subsystem and external failure sources such as natural disasters and malicious attacks. 4) The proposed CI reliability assessment indices are applicable to CI subsystems with their specific characteristics.

This paper is organized as follows. In Section II, first fault sources in CIs and then suitable model for hybrid reliability modeling have been introduced. In Section III, hybrid reliability modeling of CIs, based on graph and entropy concepts has been presented. Simulation results for a general graph to show the ability of the newly introduced model for CIs reliability assessment presented in Section IV and Section V concludes this paper.

II. CI Fault Sources and Hybrid Reliability Model

We propose three steps for the CI reliability assessment: 1) identifying CI fault sources, 2) providing a suitable model for expressing the relationship among fault sources, 3) selecting an appropriate method for analyzing the faults. Fig. 1 shows different fault sources that affect the reliability of the cyber, gas and power networks.

Fault sources that affect CI reliability are divided into internal and external groups. Internal faults are originated from internal sources inside each subsystem and external faults are originated from interactions among CI subsystems, including natural disasters and malicious attacks. Different internal fault sources in CI subsystems are listed in Table I.

After identifying the CI fault sources, the second step for the CI reliability assessments is to provide a suitable model for expressing the relationship among fault sources. The graph concept is used to determine and analyze interdependencies and relationships among different system parts.

Third step for the CI reliability assessment is to select an appropriate method for analyzing the information obtained from fault sources and interdependencies. In the CI reliability assessment, we consider two up/down states for each component expressed in one bit with the random information. So, there are many bits for the CI reliability assessment. Any system failures will cause changes in the complexity or the amount of data streams in CIs. In this paper, we use an information measures for the CI reliability assessment [25].

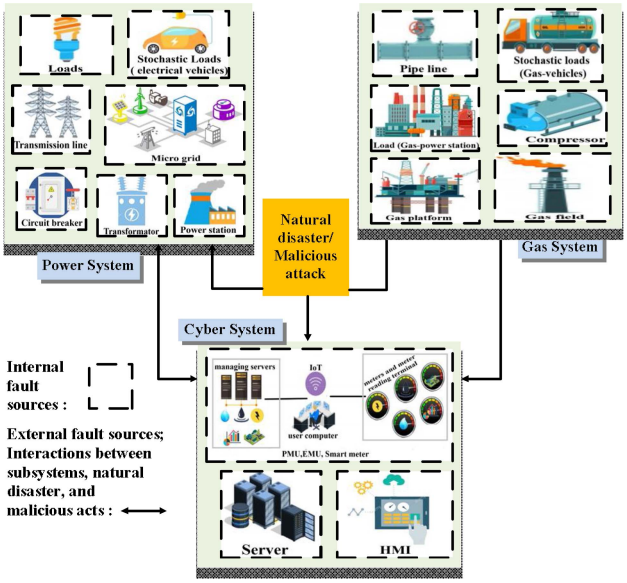

Fig. 1. Fault sources in and between cyber, gas and power networks.

In this paper, the Shannon entropy concept is applied for data mining [26] to assess information changes originated from different CI events. By defining failures as information sources, the Shannon Entropy is obtained as a probability distribution function (PDF) of information source. Other potential entropy methods for our proposed data mining include Approximate Entropy (ApEn), Sample Entropy (SampEn), Fuzzy Entropy (FuzzyEn), and Topological Entropy (FiniteTopEn) which are reviewed in [26].

The proposed reliability framework in this paper offers a new approach to decompose the CIs characteristics into smaller subsystems in which each subsystem is modeled as a graph. The proposed approach connects the smaller graphs to model the entire system. The graph concept is chosen to model the relationship among information sources based on interdependencies. The use of entropy concept will estimate the information in each CI subsystem by considering sources that affect the CI reliability. The proposed graph and entropy method has three essential steps:

Step 1: Draw the CI graph.

Step 2: Compute the CI information using the graph and entropy concept.

Step 3: Use the Monte Carlo simulation to calculate CI reliability indices.

III. Hybrid CI Reliability Model Using Graph and Entropy Concepts

A. Modeling of CI Subsystems as Graph

Graph theory has been used earlier in reliability analysis [27]. An undirected graph $G = (V; E)$ is

a mathematical structure consisting of two sets V and E, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes and $E = \{e_1, e_2, \ldots, e_m\}$ is the set of edges. A CI or any complex system can be represented as an undirected graph and each subsystem (power, cyber and gas network) is modeled by a subgraph which in it every element of the system has been considered as a link and the nodes are the connections between elements according to the topology of the system. Also, interdependencies between subsystems are modeled as specific links between end nodes of the two dependent elements or subsystems. The CI graph model is considered with n terminals (nodes) and m link.

Let m be the number of system components. Links include components and interdependencies between subsystems in CIs. Each component $i \in [m]$ is assumed to have two states: available and unavailable. In this paper, the steady-state system behavior under random component outages is studied. The steady-state probability of failure is the probability that the system is unavailable, which means that the system fails to connect all n terminals. Any collection of unavailable components which results in the system unavailability is referred to as a cut set. A cut set is minimal if it does not contain any other cut sets. The probability of failure of a cut set is equal to the probability that all cut set components are unavailable.

Let $C = \{C_1, \ldots, C_{nc}\}$ be the set of all minimal cut sets in a CI. The number of minimal cut sets (n_c) is generally exponential in the number of terminals (n). If a specified weight ($w_i = log(p_i)$) is considered for each component, a specified weight for each cut set, $w(C_i) = \sum_{i \in C_i} w_i$, can be calculated. For any constant $\alpha \geq 1$, let C_α be the set of all minimal cut sets in a system with a weight that is less than or equal to α, and N_α is the number of cut set in C_α. The cut sets in C_α is considered as α-min cut set. Using the cut set approach [28], the probability of system failure is equal to:

$$P = \sum_{i=1}^{N} P(C_i) \tag{1}$$

where $P(C_i)$ is failure probability of *i-th* cut set and P is failure probability of the system.

In this paper, minimal cut sets (C_i), are introduced as information sources as shown in Fig. 2 in which CI is modeled as interconnected graphs and component failure uncertainties in each cut set are information in each source. Three types of interdependencies are considered among CI subsystems including (i) Source-Source Interdependencies (SSI), (ii) Network-Source or Source-Network Interdependencies (stocktickerNSI, SNI) and (iii) Network-Network Interdependencies (NNI). Fig. 2 shows the CI graph with SSI and NSI in each subsystem.

The three cases are discussed as follows.

(i) Source-Source Interdependencies (SSI): This interdependency means that the changes in one information source cause changes in other sources. Equipment faults cause changes in information. For example failure in a gas pipeline feeding a power plant can affect the outage of a generator.

(ii) Network-Source or Source-Network Interdependencies (NSI, SNI): This interdependency means that changes in one information source in a specified network/source causes changes in the information of other source/network. In this research, the network means a group of cut sets. For example, a power blackout can stop operation of water pumps in a water network. So, each type of interdependency (SSI, NSI, and SNI) is considered a link which interconnects cut sets (i.e., information sources).

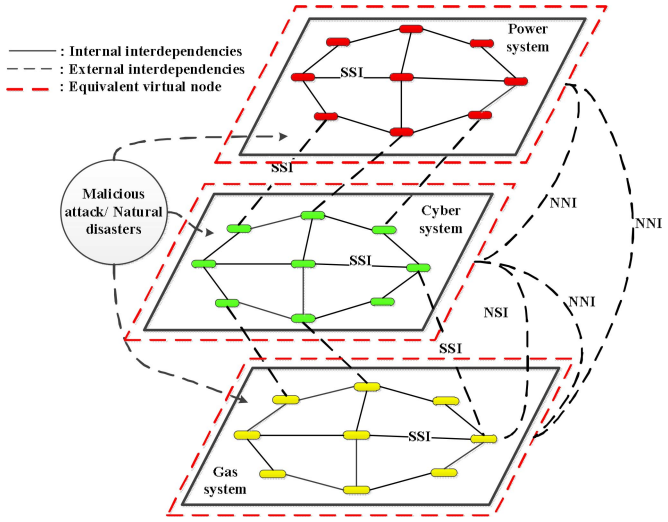

Fig. 2. CI model using interconnected graphs.

In Fig. 2, the unavailability of each subsystem or subnetwork is considered as an equivalent node shown by red dotted line in graph model. This virtual node is used for considering NSI interdependences and introducing NNI as the new type of interdependency.

(ii) Network-Network Interdependency (NNI): The NNI considers that the performance of one network causes or changes in the specification of other networks. For example, a failure in the gas network can reduce the gas required for the operation of power plants.

After modeling of the system as a graph, the next step is to introduce the entropy concept for measuring the CI information for reliability assessment.

B. Measuring CI Graph Information by Using the Entropy Concept

Reference [29] has introduced the use of information measures for the detection of events in network streams. Two widely used information measures are based on complexity and entropy approaches. This approach takes samples from the stream and estimates the information content of these samples by computing the number of steps required to build the sample string with elementary steps (complexity measures [30]), or computing the rate at which new patterns appear in the sample string (entropy measures [31]). The complexity measurement technique assumes that samples contain a roughly constant amount of information during normal operation. Also, in this approach the information carried in a string x is regarded as the minimum number of elementary steps required to construct x. The central concept in this area is the Kolmogorov-Chaitin complexity.

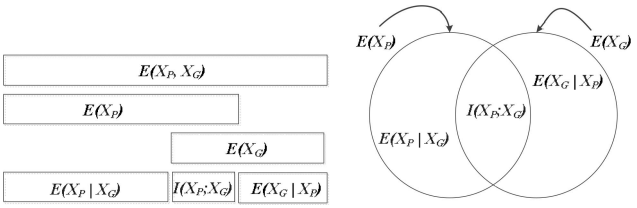

Fig. 3. Venn diagram of Entropy of a pair of random variables.

The entropy measurement technique assumes that the rate with which new patterns appear in the samples is roughly constant. The Shannon entropy merely is a function of the probability distribution of an information source, rather than a function of the strings generated by the source which considered in Kolmogorov-Chaitin complexity. In the CI reliability assessment, random component failures are sources of information. Therefore, the Shannon entropy is used in the CI reliability assessment as an appropriate information measure to assess information changes that are originated from different CI events, such as faults in devices, services and processes [25].

Let X be a discrete random variable with probability distribution function $P(X)$. The Shannon entropy $E(X)$ of X which is a measure of its uncertainty is defined as [25]:

$$E(X) = -\sum_{i=1}^{N} P(X_i)\log_2 P(X_i) \tag{2}$$

Let X be a binary Bernoulli random variable,

$$X = \begin{Bmatrix} 1 & with & probability & P \\ 0 & with & probability & 1-P \end{Bmatrix} \tag{3}$$

Then

$$E(P) = \big(-P\log_2(P) - (1-P)\log_2(1-P)\big) \tag{4}$$

$E(P)$ is 0 when $P = 0$ or 1 and the uncertainty is maximum when $P = \frac{1}{2}$ which is the maximum value of entropy [25]. We present the failure probability as a function of time and entropy.

The information of random variables is calculated for calculating cut sets. Consider a cut set with two links (two random variables), in which one of them is from the power system (X_{Po}) and the other is from the gas system (X_G). We extend the definition of entropy, for measuring the information, to a pair of random variables. The entropy $E(X_{Po}, X_G)$ of a pair of discrete random variables (X_{Po}, X_G) with a joint distribution $p(x_{po}, x_g)$ is defined as [25]:

$$E(X_P, X_G) = -\sum_{x_p \in X_P,\, x_g \in X_G} p\big(x_p, x_g\big)\log_2 p\big(x_p, x_g\big) \tag{5}$$

Based on the Venn diagram presented in Fig. 3:

$$E(X_{Po}, X_G) = E(X_{Po}) + E(X_G|X_{Po}) \tag{6}$$

The mutual information $I(X_{Po}; X_G)$ between the random variables X_{Po} and X_G is given by:

$$I(X_{Po}; X_G) = E(X_{Po}) - E(X_{Po}|X_G) \tag{7}$$

$$E(X_{Po}, X_G) = E(X_{Po}) + E(X_G) - I(X_{Po}; X_G) \tag{8}$$

If two random variables are independent, i.e., $(p(x_{po}, x_g) = p(x_{po}).p(x_g)$ or $I(X_{Po}; X_G) = 0)$, then

$$E(X_{Po}, X_G) = E(X_{Po}) + E(X_G) \tag{9}$$

So,

$$E(X_{Po}, X_G) \le E(X_{Po}) + E(X_G) \tag{10}$$

The entropy of a collection of random variables (different links in each cut set) is the sum of the conditional entropies. If $X_1, X_2, \ldots, X_n$ can be drawn according to $p(x_1, x_2, \ldots, x_n)$, then:

$$E(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} E(X_i|X_{i-1}, \ldots, X_1) \tag{11}$$

If variables are independent:

$$E(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} E(X_i) \tag{12}$$

Here, (1) can be used for measuring the information level in information sources. Various system states can be considered by use of conditional entropy in this equation. The Shannon entropy is the expected value of information associated to a single event. By examining the entropy associated with different system states, we can estimate the system events. As mentioned in (12), if variables are independent, the entropy in a cut set is the sum of the entropies of all variables in the cut set. In this paper, the upper limit of information (entropy) in a cut set will first be calculated without considering any interdependency. The first case study has analyzed a general graph based on this assumption. Then, using (8) and (11), we study the effects of interdependencies on CI reliability by introducing the information on interdependency. The second case study has analyzed these interdependencies.

Suppose that the system graph model is generated with n elements. The *i-th* element is expressed by m different independent random variables $X_i = [X_i^1, X_i^2, \ldots, X_i^m]$ for each failure. Each X_i^j is a Bernoulli distribution function with $P_i^j(\mathrm{j} = 1, \ldots, m)$ representing the probability of being in the down states because of physical failures of each element, malicious attacks, natural disasters, human errors, or unavailability of renewable sources. Therefore, because X_i^j is independent random variable, an overall probability of failure for *i-th* element (link), P_i, can be calculated based on probability theory as follows:

$$P_i = \prod_{j=1}^{m} P_i^j \tag{13}$$

We have the probability of failure for each element (link) and the whole system. By applying the Shannon entropy to each graph link, the information on failure uncertainty of each element is calculated as follows:

$$E_e = \big(P_i\log_2 P_i + (1-P_i)\log_2(1-P_i)\big) \tag{14}$$

We calculate the changes in the graph information level (i.e., the system performance measure). In the first step, the graph information level is calculated. Based on (1), the system

failure probability is dependent on the failure of all cut sets. Therefore, the system failure information is related to information of all cut sets. Accordingly, if the upper limit of entropy in each cut set is considered, the information in each cut set is calculated based on (15) which uses (12),

$$E_{ci}(E_1, E_2, \ldots, E_{n_e}) = \sum_{k=1}^{n_e} E_k \tag{15}$$

where E_{ci} is the entropy of the i-th cut set, n_e is the number of elements in this cut set, and E_k is the entropy of elements in the cut set. If the upper limit of system entropy is considered, the system entropy will be calculated based on (16) which uses (12). E_s is the information on the system failure uncertainty in which the whole system information is calculated based on (14) and (15). The approach use the following equation where J is the minimum cut set of the system and E_r is the entropy of each cut set in the system,

$$E_s(E_{c1}, E_{c2}, \ldots, E_{cJ}) = \sum_{r=1}^{J} E_{cr} \tag{16}$$

The change in information is equivalent to an event occurrence. After calculating the system information for the reliability assessment, we will discuss in the next section the changes in the level of information and entropy. Here, we analyze the interdependency based on its effect on information sources (cut sets).

Two random variables are independent if $p(x_{po}, x_g) = p(x_{po}).p(x_g)$ or $I(X_{Po}; X_G) = 0$. Two random variables be dependent if $I(X_{Po}; X_G)$ is non-zero. The mutual information index introduces the level of dependency of variables. Each minimum cut set in subsystems is considered an information source in order to calculate the system information level. So, based on these sources, interdependencies in and between subsystems, which affect the system performance, are determined as follows:

The mutual information among system cut sets is for different types of interdependencies in and between subsystems.

The mutual information between two minimum cut set is SSI. The mutual information between a minimum cut set and a group of cut sets (in one subsystem) is SNI or NSI. The mutual information between two groups of cut sets is NNI.

C. System Reliability Assessment Based on Graph-Entropy

For a CI, different levels of entropy can be obtained based on (16) and the changes in CI information level are monitored for events detection. Suppose function *E(t, p)*, which is drawn based on failure probability of different states, and different plates are shown in Fig. 4. Plate E_0 shows the zero entropy level. The system entropy is zero for $P = 0$ and $P = 1$. At E_0 and $P = 0$, the system has a correct performance, and in $P = 1$, the system certainly has as an incorrect performance.

The distance between E_0 plate and *D(E0, E)* plates related to different system states, which describes the system performance, is proposed for the CI reliability assessment. When CIs is in its normal operation, the probability of system failure is close to zero. So, the system is operated in region *A*.

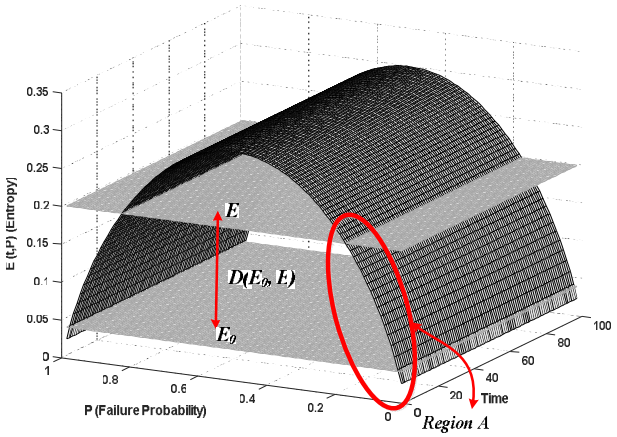

Fig. 4. Reliability and entropy distance relation.

In this region, adding any uncertainty will increase the system unreliability. In region *A*, CI elements have low failure probabilities, and limiting the failure probability to zero causes the distance between the *E* plate and the normalized entropy (information about uncertainty) of the whole system, and plate E_0, closer to zero with a *high* probability. Based on this description, the *reliability* of the system based on performance measure of system can be defined as follow:

The reliability of a system modeled by a graph is the probability of approaching the distance between E and E_0 to 0.

$$reliability = \left\{ \Pr(D(E_0, E) \to 0 \;\; In \;\; region \;\; A \right\} \tag{17}$$

The calculation of *D(E0, E)* as the distance between *E* and E_0 and probability of approaching *D(E0, E)* to zero are two important points in (18). *D(E0, E)* is calculated by using the Kullback-Leibler concept which measures the distance between two probability distributions based on (18) [25].

$$D(E_0, E) = \sum_{x \in X} E_0(x) \log_2 \frac{E_0(x)}{E(x)} \tag{18}$$

Based on (18), $D(E_0, E) = 0$ if and only if $E_0(x) = E(x)$. Because of the stochastic system behavior, the probability of approaching *D(E0, E)* to zero has been calculated by the Monte Carlo method which is discussed as follows.

In the following, conventional reliability indices are presented and then redefined with the graph and entropy concept for calculating the system reliability.

The system reliability indices range from network performance to network connectivity reliability indices [32]. Examples of network performance reliability indices include Average Service Availability Index (ASAI), System Average Interruption Frequency Index (SAIFI) and etc. The network connectivity reliability merely considers network topology and introduces the 'probability of connectivity achieved by network' as a reliability criterion, which includes two-terminal connectivity, K-terminal connectivity, and all-terminal connectivity reliability [33].

A general trend for system modeling and the network connectivity reliability index assessment by applying the graph and entropy concept is expressed as follows.

Step 1: Specify elements that influence CI subsystem reliabilities (power *(Po)*, gas *(G)* and cyber *(Cy)* system).

Step 2: Based on the system topology and relations among subsystem elements (internal interdependencies), draw an equivalent subsystem graph such that any elements that perform as a set is considered a link and any element input/output is modeled as a node in the graph.

Step 3: Based on external interdependencies of power, gas and cyber system draw an equivalent system graph such that any interdependencies is considered a link.

Step 4: Determine minimal cut sets in the system graph.

Step 5: Generate P_s as a failure probability vector for system elements,

$$P_s = \begin{bmatrix} P_{Po} & P_G & P_{Cy} \end{bmatrix} \tag{19}$$

The probabilities are calculated based on (13) by considering affecting failure sources for each element including random system failures, malicious attacks, natural disasters, human errors, and unavailability of renewable sources.

The failure probability of power, gas and cyber system elements is calculated as,

$$P_{Po/G/Cy} = \begin{bmatrix} P_{Po/G/Cy,1}, P_{Po/G/Cy,2}, \ldots, P_{Po/G/Cy,i}, \ldots, \\ P_{Po/G/Cy,/n_{Po}/n_G/n_{Cy}} \end{bmatrix} \tag{20}$$

where, n_{Po}, n_G and n_{Cy} are the number of elements in power, gas and cyber system.

Based on (13), $P_{Po/G/Cy,i}$ is stated as,

$$P_{Po/G/Cy,i} = \prod_{j=1}^{m} P_i^j \tag{21}$$

where, P_i^j represents any failure source for an element. Considering that there are multiple scenarios for system reliability studies, several failure probability terms is considered.

Step 6: Calculate the entropy of each element *(E_i)* and cut set *(E_c)* by using (14) and (15).

Step 7: Calculate the system entropy based on (16) *(E_s)*.

Step 8: Calculate *$D(E_0, E_s)$ for* E_s and E_0 based on (18).

Step 9: The probability that *$D(E_0, E_s)$* approaches *0* is the system reliability. This probability is an index for network connectivity reliability assessment which is calculated by the Monte Carlo method as,

$$R = \frac{1}{N} \sum_{f=1}^{N} I_{\{D(E_s(X_f), E_0) \le \varepsilon\}} \tag{22}$$

where, I is the identity matrix, ε is a small number, and vector X is a Bernoulli random variable defined in (3) with P calculated from (19). After any change in the system state (change in information level), P is redefined as in (1).

In this paper, graph theory is used to measure the system entropy and then a binary model is used to calculate the probability of reducing the distance between different entropy levels to zero. The method is summarized in Fig. 5.

Accordingly, input data requirements for the proposed algorithm are probability distribution functions which influence the availability of system elements such as malicious attacks, natural disasters, human errors, and unavailability of renewable sources. These probabilities of failures are updated in every time window. The accuracy of the method is enhanced by increasing N which is the number of iterations in the Monte Carlo algorithm.

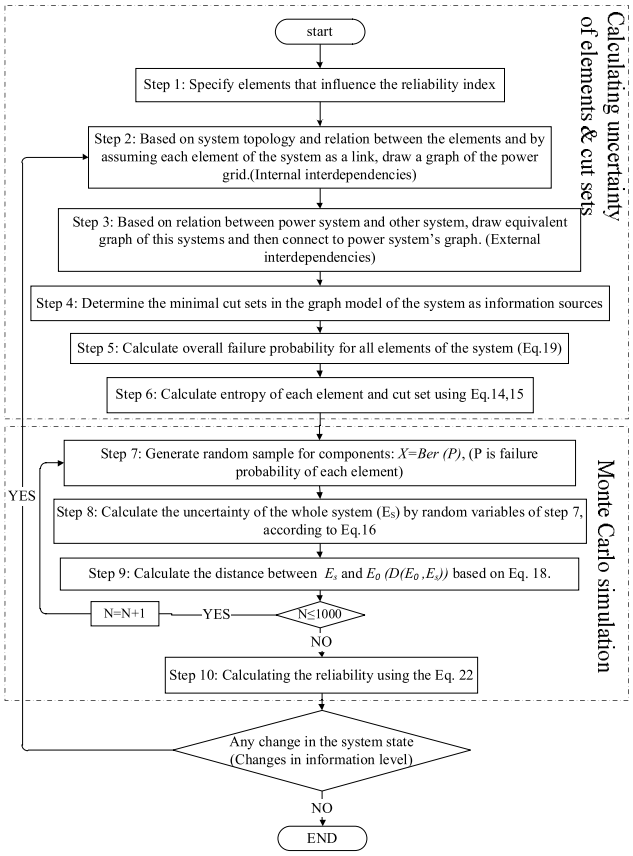

Fig. 5. Modeling and reliability assessment of a system.

IV. Case Studies

The proposed network reliability indices range from the network performance to network connectivity reliability indices. To illustrate the effectiveness of the proposed method for assessing the two kinds indices, two scenarios are considered in this section.

First Scenario: In the first scenario, the failure probability is calculated by selecting a general graph and the results are compared with those in [34] for the all-terminal connectivity reliability assessment scenario. In this example, the system graph is modeled by specifying links and interdependencies. In this scenario, the failure probability is considered for a system whose terminals are connected through components that experience random failure and repair processes over time.

Fig. 6 depicts a simple system with $n = 9$ nodes and $m = 12$ links. The unavailability of all components is p for values listed in Table. II. For six section time (t), Table II lists the maximum failure probability of a cut set (P^*), α, and the number of α-min cut sets being enumerated (N_α) [34]. In this case, for the comparison of our simulation results with others, only

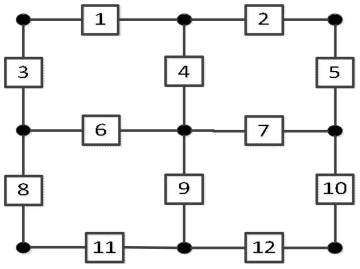

Fig. 6. A system with 9 terminals and 12 components [34].

TABLE II
PROBABILITIES OF COMPONENT FAILURE AND CUT SETS IN FIG. 6

T	P_D	P*	α	N_α
1	1e-2	1e-4	2.93	53
2	1e-2.6	6.31e-6	2.22	37
3	1e-2.8	2.51e-6	2.09	37
4	1e-3	1e-6	2	20
5	1e-3.2	3.98e-7	1.92	20
6	1e-3.4	1.58e-7	1.84	20

TABLE III
PROBABILITIES OF SYSTEM FAILURE IN FIG. 6

T	E	$D(E_0,E)$	P		Error (%)
			Proposed	Ref [34]	
1	0.0360	0.0036	0.0051	0.0053	3.0189
2	0.0056	0.0058	2.3200e-4	2.3347e-4	0.6296
3	0.0044	0.0015	9.2000e-5	9.287e-5	0.9638
4	0.0041	0.0014	2.1000e-5	2e-5	5
5	0.0040	0.0010	8.0000e-6	7.96e-6	0.3769
6	2.8014e-5	1.2725e-4	3.0000e-6	3.16e-6	5

the probability of being in the down state ($P_D = \lambda/\lambda + \mu$) is considered for each element and then the failure probability of each cut set is calculated. The entropy of each state of system *(E)*, distance between E and E_0 plates *(D(E, E_0))* and failure probability of a system which shown in Fig. 6, for various failure rates of components in different section times are listed in Table II and simulation results of the newly introduced method have been compared with [34]. For this purpose, first minimal cut sets (information sources) have been determined. Next, information about failure uncertainty of each element and cut set based on steps 5-8 and eventually, the failure probability of a system based on step 10, have been calculated. In Table II, the failure probability of system has been listed in column P (Proposed and Ref columns). The failure probability of each cut set is presented in [34] which have been expressed as P* in Table II. Based on these values, the system failure probability is calculated using (1) and the corresponding results are shown as [34] in Table III. Our simulations were done using MATLAB and the results shown in Table III indicate less than 5% error for different states in comparison to [34].

In Table I, three-time sections are considered and the corresponding failure probability of the system is depicted in Fig. 7 *(P (t, E))*. Fig. 7 shows that an increase in downstate probability will increase the system entropy and the failure probability.

Second Scenario: In the second scenario, the Swiss high-voltage Electric power supply system (EPSS) is introduced as CI by considering power and cyber subsystems. The comprehensive results are provided in https://faculty.kashanu.ac.ir/mhejazi/en/page/15312/cis-illustrative-example.

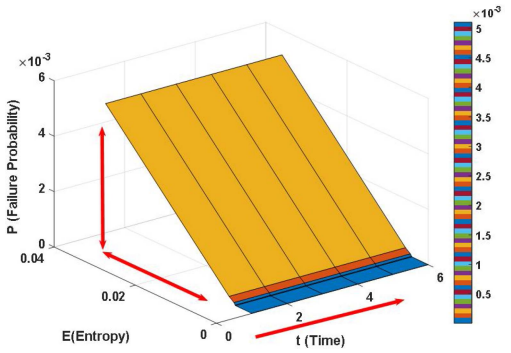

Fig. 7. Failure probabilities in Fig. 6 as a function of time and entropy.

V. CONCLUSION

In comparison to previous works on the CI reliability assessment, different challenges signify that traditional deterministic methodologies for reliability assessments cannot represent the comprehensive CI performance. The modeling of network structure (interdependences), network dynamic (data stream), failure mechanism, and the introduction of proper indices for the reliability assessment of CI subsystems are examples of these challenges. These challenges have been addressed in this paper by introducing a new performance measure for the CI reliability assessment, which consists of the following components: 1) a multilayer hybrid model for representing interdependences in and between CI subsystems. Three types of interdependencies are introduced using the multilayer model for considering simultaneous sources of uncertainty in subsystems and interdependencies between subsystems. 2) A performance metric for reliability assessment is introduced which can consider various sources affecting the system performance, including internal fail sources in each subsystem and external fail sources such as natural disasters and malicious attacks. Also, 3) Reliability indices are provided for CI subsystems without regard to the nature of the subsystem. To illustrate the effectiveness of the proposed method, failure probability of a general system is presented by selecting a general graph. The proposed results corroborate the accuracy of the proposed method in comparison to other studies. However, we believe that considerably more work is to be done to demonstrate the effectiveness of the proposed algorithm on practical cases which include power, gas and cyber systems. In addition, the consideration of resilience, robustness, survivability, and stability will be imperative for assessing the merits of the method proposed in this paper.

REFERENCES

[1] I. Eusgeld, C. Nan, and S. Dietz, "System-of-systems' approach for interdependent critical infrastructures," *Rel. Eng. Syst. Safety*, vol. 96, no. 6, pp. 679–686, 2011.
[2] C. Nan and G. Sansavini, "Multilayer hybrid modeling framework for the performance assessment of interdependent critical infrastructures," *Int. J. Critical Infrastruct. Protect.*, vol. 10, pp. 18–33, Sep. 2015.

[3] W. Kröger and C. Nan, "Addressing interdependencies of complex technical networks," in *Networks of Networks: The Last Frontier of Complexity*. Cham, Switzerland: Springer, 2014, pp. 279–309.
[4] R. D'Souza, C. Brummitt, and E. Leicht, "Modeling interdependent networks as random graphs: Connectivity and systemic risk," in *Networks of Networks: The Last Frontier of Complexity*. Cham, Switzerland: Springer, 2014, pp. 73–94.
[5] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Syst. Mag.*, vol. 21, no. 6, pp. 11–25, Dec. 2001.
[6] J. Gao, X. Liu, D. Li, and S. Havlin, "Recent progress on the resilience of complex networks," *Energies*, vol. 8, no. 10, pp. 12187–12210, 2015.
[7] M. Ouyang, "Review on modeling and simulation of interdependent critical infrastructure systems," *Rel. Eng. Syst. Safety*, vol. 121, pp. 43–60, Jan. 2014.
[8] P. Pederson, D. Dudenhoeffer, S. Hartley, and M. Permann, "Critical infrastructure interdependency modeling: A survey of U.S. and international research," Idaho Nat. Lab., Idaho Falls, ID, USA, Rep. INL/EXT-06-11464, 2006.
[9] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *J. Complex Netw.*, vol. 2, no. 3, pp. 203–271, 2014.
[10] B. Falahati, Y. Fu, and L. Wu, "Reliability assessment of smart grid considering direct cyber-power interdependencies," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1515–1524, Sep. 2012.
[11] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Power systems big data analytics: An assessment of paradigm shift barriers and prospects," *Energy Rep.*, vol. 4, pp. 91–100, Nov. 2018.
[12] R. Preece and J. V. Milanović, "Assessing the applicability of uncertainty importance measures for power system studies," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2076–2084, May 2016.
[13] G. Li, Z. Bie, Y. Kou, J. Jiang, and M. Bettinelli, "Reliability evaluation of integrated energy systems based on smart agent communication," *Appl. Energy*, vol. 167, pp. 397–406, Apr. 2016.
[14] J. Munoz, N. Jimenez-Redondo, J. Perez-Ruiz, and J. Barquin, "Natural gas network modeling for power systems reliability studies," in *Proc. IEEE Bologna Power Tech Conf.*, vol. 4. Bologna, Italy, 2003, p. 8.
[15] X. Zhang and S. Mahadevan, "A game theoretic approach to network reliability assessment," *IEEE Trans. Rel.*, vol. 66, no. 3, pp. 875–892, Sep. 2017.
[16] E. Zio, "Challenges in the vulnerability and risk analysis of critical infrastructures," *Rel. Eng. Syst. Safety*, vol. 152, pp. 137–150, Aug. 2016.
[17] B. Falahati and Y. Fu, "Reliability assessment of smart grids considering indirect cyber-power interdependencies," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1677–1685, Jul. 2014.
[18] H. Lei, C. Singh, and A. Sprintson, "Reliability modeling and analysis of IEC 61850 based substation protection systems," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2194–2202, Sep. 2014.
[19] I. A. Tøndel, J. Foros, S. S. Kilskar, P. Hokstad, and M. G. Jaatun, "Interdependencies and reliability in the combined ICT and power system: An overview of current research," *Appl. Comput. Informat.*, vol. 14, no. 1, pp. 17–27, 2018.
[20] Y. Wang, C. Chen, J. Wang, and R. Baldick, "Research on resilience of power systems under natural disasters—A review," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1604–1613, Mar. 2016.
[21] X. Liu, M. Shahidehpour, Z. Li, X. Liu, Y. Cao, and Z. Bie, "Microgrids for enhancing the power grid resilience in extreme conditions," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 589–597, Mar. 2017.
[22] A. Kovacevic and D. Nikolic, "Cyber attacks on critical infrastructure: Review and challenges," in *Handbook of Research on Digital Crime, Cyberspace Security, and Information Assurance*. Hershey PA, USA: IGI Global, 2015, pp. 1–18.
[23] U. P. D. Ani, H. He, and A. Tiwari, "Review of cybersecurity issues in industrial critical infrastructure: Manufacturing in perspective," *J. Cyber Security Technol.*, vol. 1, no. 1, pp. 32–74, 2017.
[24] S. LaRocca, "Modeling the reliability and robustness of critical infrastructure networks," Ph.D. dissertation, Dept. Doctor Philos., Johns Hopkins Univ., Baltimore, MD, USA, 2014.
[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Somerset, U.K.: Wiley, 2012.
[26] A. Holzinger *et al.*, "On entropy-based data mining," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Heidelberg, Germany: Springer, 2014, pp. 209–226.
[27] S. Xu, Y. Qian, and R. Q. Hu, "On reliability of smart grid neighborhood area networks," *IEEE Access*, vol. 3, pp. 2352–2365, 2015.
[28] R. Billinton and R. N. Allan, *Reliability Evaluation of Engineering Systems*. New York, NY, USA: Springer, 1992.
[29] R. E. Eimann, "Network event detection with entropy measures," Ph.D. disseration, Dept. Comput. Sci., Univ. Auckland, Auckland, New Zealand, 2008.
[30] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 75–81, Jan. 1976.
[31] C. Shannon, "A mathematical theory of communication," *Bell Syst. Techn. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
[32] Y.-N. Jiang, R.-Y. Li, N. Huang, and R. Kang, "Survey on network reliability evaluation methods," *Comput. Sci.*, vol. 39, no. 5, pp. 9–13, 2012.
[33] L. Wu, Q. Tan, and Y. Zhang, "Network connectivity entropy and its application on network connectivity reliability," *Physica A, Stat. Mech. Appl.*, vol. 392, no. 21, pp. 5536–5541, 2013.
[34] A. Heidarzadeh, A. Sprintson, and C. Singh, "A fast and accurate failure frequency approximation for *k*-terminal reliability systems," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 933–950, Sep. 2018.

Mohammadreza Iranpour (Student Member, IEEE) was born in Isfahan, Iran, in 1989. He received the B.Sc. degree (Hons.) in electrical engineering and the M.Sc. degree (Hons.) in power electrical engineering from the University of Kashan, Kashan, Iran, in 2012 and 2015, respectively. His major research interests include reliability and resilience assessment of critical infrastructure systems, smart grids, and complex networks.

Maryam A. Hejazi (Member, IEEE) was born in Isfahan, Iran, in 1980. She received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Amir-kabir University of Technology in 2011, 2006, and 2003, respectively. She has been an Academic Member with the Department of Power Engineering, Faculty of Electrical and Computer Engineering, University of Kashan, since September 2011. Her research interests include monitoring, power transformers, distributed resources, reliability and resilience assessment of critical infrastructure systems, and smart grids.

Mohammad Shahidehpour (Fellow, IEEE) received an Honorary Doctorate degree in electrical engineering from the Polytechnic University of Bucharest, Bucharest, Romania. He is the Bodine Chair Professor and the Director of the Robert W. Galvin Center for Electricity Innovation, Illinois Institute of Technology, Chicago, IL, USA. He is a Fellow of the American Association for the Advancement of Science, and the National Academy of Inventors. He is a member of the U.S. National Academy of Engineering.