Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P7627-cd

Game theory-based defense strategies against coordinated attacks on multi-state interdependent critical infrastructures

Maria Valentina Clavijo Mesa*
Energy Department
Politecnico di Milano
Milan, Italy
mariavalentina.clavijo@polimi.it
*Corresponding author

Yanfang Wu
School of Management
Northwestern Polytechnical University
Xi'an, China
wwuyanfang@mail.nwpu.edu.cn

Enrico Zio
MINES Paris - PSL University, Centre de Recherche sur les Risques et les Crises (CRC),
Sophia Antipolis, France
Energy Department, Politecnico di Milano
Milan, Italy
enrico.zio@polimi.it
enrico.zio@mines-paristech.fr

***Abstract*— Critical Infrastructures (CIs) delivering essential services, like energy and water, share vulnerabilities through their interdependences of many kinds, including physical, operational and others. These interdependences generate systems of systems exposed to coordinated attacks that can lead to cascading failures across the infrastructures. This work presents an attack-defense game model to determine optimal defense strategies for multi-state interdependent CIs. The model combines game theory and network theory to assess the topological and operational features of the interdependent infrastructures considered. To estimate the operational impact of disruptions on the CIs, inoperability assessments are developed and critical nodes are identified based on their topological importance in the system of systems. The model considers the risk attitude of both attackers and defenders by evaluating their respective game payoffs with Cumulative Prospect Theory (CPT). A case study regarding a system of systems made of a power grid and a water network is used to illustrate the application of the proposed model.**

Keywords—Critical Infrastructures, Interdependencies, Multiple States, Inoperability, Coordinated Attacks, Attack-defense Game, Cumulative Prospect Theory.

I. Introduction

Critical Infrastructures (CIs), such as energy grids, water systems and transportation networks are increasingly vulnerable to intentional manmade attacks [1]. For example, in 2015, a coordinated cyberattack on Ukraine's power grid caused a massive outage, leaving 230,000 people without electricity and straining critical services, including healthcare and emergency response [2]. Similarly, in 2022, a sabotage of electrical substations in North Carolina affected the U.S. power grid, leaving thousands without power and disrupting water distribution systems reliant on electricity [3].

Game theory, a widely used approach for analyzing attacks on CIs, considers the strategic behavior of multiple players, namely attackers and defenders [4]. For instance, the binary attack-defense game model has been applied to prioritize critical components for defense in urban water distribution networks [5]. Additionally, the resource allocation attack-defense game model has been employed to determine optimal resource allocation strategies for defenders to protect specific components, using a contest success function [6]. Both approaches incorporate valuable analytical insights but assume risk neutrality for attackers and defenders, which limits their applicability to real-word scenarios where risk preferences significantly influence action decision-making.

To address the impact of risk preferences, a game theory-based model has been proposed that incorporates Cumulative Prospective Theory (CPT) to account for differences in how attackers and defenders assess risk [7][8]. One study proposes a framework that employs the Binary Decision Diagram (BDD) method to analyze potential failure states on the CI and demonstrates how risk attitudes influence strategic decisions [7]; however, the proposed framework is limited to small networks due to the computational challenges of BDD, making it unsuitable for large-scale systems. Furthermore, no strategies are provided to identify critical nodes in the CIs. To overcome this latter limitation, another framework combines CPT with genetic algorithms for identifying critical nodes [9]; whereas this improves scalability, it focuses only on a single infrastructure, thus not considering the cascading process across interconnected infrastructures.

In this paper, we propose a framework for developing defense strategies against coordinated attacks on multi-state interdependent CIs, addressing key challenges such as critical nodes identification, cascading effects and resource allocation. First, a critical target identification method is proposed to develop the strategy model for players. Second, a CI inoperability model is introduced to evaluate the performance of CIs considering cascading failures. The inoperability metric is, then, used to construct the payoff model, incorporating the Cumulative Prospective Value (CPV). The framework is validated through a case study involving a power and a water network; interdependencies between the two systems are considered and shown to exacerbate disruptions.

The remainder of the paper is organized as follows: Section 2 presents the proposed attack-defense framework; Section 3 outlines the method for identifying critical targets in CIs; Section 4 describes the case study of the interdependent power grid and water distribution network; Section 5 discusses the results and Section 6 concludes the paper with future research directions.

II. ATTACK-DEFENSE GAME FRAMEWORK

The attack-defense game framework proposed for interdependent CIs is composed of four phases, as shown in Figure 1.

The first phase involves an inoperability assessment model for each CI. This phase considers the topological and operational characteristics of the interconnected infrastructures to obtain the identification of potential target components or systems in each CI. In the second phase, a strategy model is developed using the outputs from the inoperability assessment, including the identified targets, to formulate the strategic interactions between the attacker and defender. The third phase focuses on the development of a payoff model. CPT is introduced to account for the risk preferences of the attackers and defenders, quantifying their strategic utilities based on their resource allocations and outcomes. Finally, the fourth phase applies a two-stage optimization framework to calculate the game equilibrium solution, determining the optimal resource allocation strategies for the players.

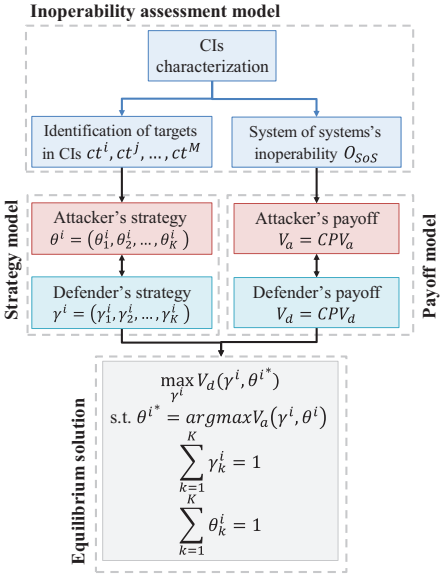

Figure 1. Framework of attack-defense game on CIs

A. Assumptions

Before detailing each phase of the framework, the following assumptions underpin the proposed game model: The model considers one attacker and one defender, each potentially representing teams of multiple players. The defender acts as the leader, selecting defense strategies first, whereas the attacker follows, making attack decisions after observing the defender's actions. Both players allocate resources to influence the state of targets in the targeted CI, where the likelihood of a target reaching a complete failure state is determined by the relative resource investment of the attacker and defender.

B. Inoperability assessment model

This phase evaluates the performance of each CI under various disruption scenarios, considering its internal structure and the interdependencies with the other CIs in the system of systems.

Each CI is first represented as a directed graph $G^i \equiv (N^i, E^{(i,i)})$, where N^i denotes the nodes representing components or systems of the generic i-th CI, and $E^{(i,i)}$ captures the internal connections. Each node $n_z^i \in N^i$ is characterized by a discrete multi-state variable $\bar{x}^{n_z^i}$, representing operational states ranging from fully functional $\left(x_1^{n_z^i}\right)$ to complete failure $\left(x_{K_{(n_z^i)}}^{n_z^i}\right)$. These node states are used to construct a structural function that captures all potential states of the infrastructure. Under the assumption of a perfect state of the infrastructure $\left(X_1^i \middle| x_1^{n_z^i} \; \forall n_z^i \in N^i\right)$, the nominal demand is estimated $D\left(X_1^i\right)$; the inoperability of the infrastructure in a given state (X_l^1) is, then, quantified as:

$$O\left(X_l^i\right) = 1 - \frac{D(X_l^i)}{D(X_1^i)} \tag{1}$$

where $D(X_l^i)$ is the demand met in state X_l^i and $D(X_1^i)$ corresponds to the demand met in the perfect state X_1^i. This metric, ranging from 0 (no inoperability) to 1 (complete inoperability), provides a standardized measure of performance degradation.

To capture the interdependencies in the system of systems, additional edges are introduced between the graphs of the interdependent CIs. For instance, in the case of a generic i-th CI interdependent with a generic j-th CI, edges $E^{(i,j)}$ and $E^{(j,i)}$ are introduced between the graphs G^i and G^j. Nodes associated with these interdependencies are identified in G^i as $N_{i\to j}^i$ and $N_{j\to i}^i$,and in G^j as $N_{i\to j}^j$ and $N_{j\to i}^j$. These nodes represent critical connections through which disruptions in one CI can cascade to another. Cascading effects are modeled using a dependency matrix, as described in [10].

When a disruption scenario occurs where one or more nodes of a CI are directly affected, the failed node(s) and their associated edges are removed, leading to changes in the states of the CI hit by the disruption and to potential cascading failures in the dependent CIs. Once the system of systems stabilizes, the inoperability of each CI is estimated and the overall system of systems inoperability, O_{SoS}, is calculated as the weighted average of the inoperabilities of the individual CIs:

$$O_{SoS} = \sum_{i=1}^{M} w_i \cdot O(X_l^i), \qquad \sum_{i=1}^{M} w_i = 1 \tag{2}$$

where w_i represents the relative importance of the i-th CI and M is the total number of CIs in the system of systems.

For each disruption scenario, generated by removing individual nodes, O_{SoS} is evaluated to quantify the operational impact.

Building on these results, potential target nodes are identified in each CI using the topological metrics described in Section III.

C. Strategy model

The strategy model begins by defining the initial condition of the game, which consists of selecting the CI to be attacked. Assuming that the generic i-th CI is chosen as the target of the attack, the model is constructed based on its critical nodes as potential targets of intelligent attacks. The critical target node set for the generic i-th CI is defined as $ct^i = \left(ct_1^i, ct_2^i, \dots, ct_K^i\right)$ where $ct_k^i = n_k^i \in N^i$. Given this target node set and the resources available to the attacker (R_a) and the defender (R_d), the two players strategically allocate their resources on the K targets.

The allocation ratios for the attacker and defender are represented as $\theta^i = \left(\theta_1^i, \theta_2^i, \dots, \theta_K^i\right)$ and $\gamma^i = \left(\gamma_1^i, \gamma_2^i, \dots, \gamma_K^i\right)$, respectively, where θ_k^i denotes the proportion of the attacker's total resources (R_a) allocated to attacking critical target k, and γ_l^i represents the proportion of the defender's total resources (R_d) allocated to protecting critical target l. It is assumed that a successful attack results in the complete failure of the targeted node; using the traditional Tullock model, the probability of successfully attacking a critical target node k (ct_k^i), denoted as p_k^i, is given by [8]:

$$p_k^i = \frac{R_a \cdot \theta_k^i}{\left(R_a \cdot \theta_k^i + R_d \cdot \gamma_k^i\right)} \tag{3}$$

This formulation captures the strategic interplay between the attacker and defender, where the likelihood of a successful attack is determined by the relative resource investments of both players.

The interaction between the attacker and defender results in 2^K possible outcomes, as each of the K targets in the critical target set ct^i of the target i-th CI can be successfully either attacked or defended. These outcomes are summarized in Table I.

TABLE I. SCENARIOS OF ATTACK-DEFENSE GAME

Num of failed target nodes	Description	Scenarios	Probability of scenario occurrence ($\mathcal{P}$)
0	None target in ct^i is compromised	C_0^i	$\prod_{k=1}^{K}(1-p_k^i)$
1	One target in ct^i is in failure state	C_1^i	$p_k^i \cdot \prod_{f=1, f\neq k}^{K}(1-p_f^i)$
2	Two targets in ct^i are in failure state	C_2^i	$p_k^i \cdot p_f^i \cdot \prod_{h=1, h\neq k,f}^{K}(1-p_h^i)$
⋮	⋮	⋮	⋮
K	The K targets in ct^i are in failure state	C_K^i	$\prod_{k=1}^{K} p_k^i$

Possible outcomes occur with different probabilities and range from no target nodes being in a failure state to all K target nodes being in a failure state. For example, the probability of no target in ct^i being in a failure state $\left(C_0^i\right)$ is calculated as the product of the probabilities of all targets being successfully defended, whereas the probability of all K targets being in a failure state $\left(C_K^i\right)$ is the product of the probabilities of all targets being successfully attacked. Intermediate scenarios (C_1^i to C_{K-1}^i) represent combinations of successful attacks and defenses, each with its related probabilities of occurrence.

D. Payoff model

In this game, O_{SoS} and CPVs are employed to analyze the players' payoffs, incorporating their risk preferences and the performance of the system of systems. The CPV for each player is determined by evaluating the utilities of all potential outcomes and their associated probabilities. For the defender, the central objective is to minimize inoperability by preventing the attacker from compromising critical target nodes. The defender's utility increases when the system of systems remains operational, whereas the attacker's utility increases when it becomes inoperable.

The utilities u_a for the attacker and u_d for the defender in scenario s are mathematically defined as:

$$u_{a(s)} = O_{SoS} \tag{4}$$

$$u_{d(s)} = 1 - O_{SoS} \tag{5}$$

For each scenario s among the 2^K possible outcomes, the utilities represent the respective gains or losses for each player under all scenarios. Specifically, the attacker's utility corresponds to the inoperability of the system of systems, whereas the defender's utility reflects its non-inoperability. These utilities, denoted as $u_{a(1)}, u_{a(2)}, \dots, u_{a(2^K)}$ for the attacker and $u_{d(1)}, u_{d(2)}, \dots, u_{d(2^K)}$ for the defender, capture the full range of possible outcomes in the game.

To better reflect potential losses and gains, the process outlined in [9] is adopted, where utilities are ranked in ascending order. For the attacker, the ranked utilities are represented as $\tilde{u}_{a(1)} \leq \cdots \leq 0 \leq \tilde{u}_{a(s)} \leq \tilde{u}_{a(2^K)}$, where $\tilde{u}_{a(1)}$ denotes the smallest utility value for the attacker. Similarly, for the defender, the ranked utilities are expressed as $\tilde{u}_{d(1)} \leq \cdots \leq 0 \leq \tilde{u}_{d(s)} \leq \tilde{u}_{d(2^K)}$, with $\tilde{u}_{d(1)}$ corresponding to the smallest utility value for the defender.

Each utility is paired with a probability that reflects the likelihood of the associated outcome (Table I). For the attacker, the probabilities are ranked as $\tilde{\mathcal{P}}_{a(1)} \leq \cdots \leq 0 \leq \tilde{\mathcal{P}}_{a(s)} \leq \tilde{\mathcal{P}}_{a(2^K)}$ and similarly for the defender $\tilde{\mathcal{P}}_{d(1)} \leq \cdots \leq 0 \leq \tilde{\mathcal{P}}_{d(s)} \leq \tilde{\mathcal{P}}_{d(2^K)}$.

The CPVs for the attacker (V_a) and defender (V_d) are calculated as:

$$V_a = \sum_{\tau=1}^{s} \pi_{a(\tau)}^{-} \times v(\tilde{u}_{a(\tau)}) + \sum_{\rho=s+1}^{2^K} \pi_{a(\rho)}^{+} \times v(\tilde{u}_{a(\rho)}) \tag{6}$$

$$V_d = \sum_{\tau=1}^{s} \pi^{-}_{d(\tau)} \times v(\tilde{u}_{d(\tau)}) + \sum_{\rho=s+1}^{2^K} \pi^{+}_{d(\rho)} \times v(\tilde{u}_{d(\rho)}) \quad (7)$$

In these equations, $\tau = 1$ to s corresponds to losses (scenarios with a utility of zero or less), whereas $\rho = s+1$ to 2^K represents gains (scenarios with positive utilities). Decision weights ($\pi^{-}_{a(\tau)}, \pi^{-}_{d(\tau)}$ for losses and $\pi^{+}_{a(\rho)}, \pi^{+}_{d(\rho)}$ for gains) reflect how probabilities are subjectively perceived by decision-makers. Specifically, decision-makers tend to overestimate low-probability events and underestimate high-probability events [8].

The utility functions $v(\tilde{u}_B)$ for the attacker (a) and defender (d), whether associated with losses (τ) or gains (ρ), are defined as [9]:

$$v(\tilde{u}_B) = \begin{cases} \tilde{u}_B^{g} & \tilde{u}_B \geq 0 \\ -\lambda(-\tilde{u}_B)^{l} & \tilde{u}_B < 0 \end{cases} \quad (8)$$

where B refers to $a(\tau)$, $a(\rho)$, $d(\tau)$ or $d(\rho)$, based on the player and context (loss/gain). Parameters g and l define risk preferences: $0 < g < 1$ indicates risk aversion over gains, while $0 < l < 1$ reflects risk-seeking over losses. Higher g and l reduce these tendencies. The factor λ represents loss-aversion, with $\lambda > 1$ indicating that individuals are generally more sensitive to losses than to equivalent gains.

The decision weight $\pi^{-}_{(\tau)}$ and $\pi^{+}_{(\tau)}$ are calculated as:

$$\pi^{-}_{(\tau)} = w^{-}\left(\sum_{h=1}^{\tau} \tilde{p}_{(h)}\right) - w^{-}\left(\sum_{h=1}^{\tau-1} \tilde{p}_{(h)}\right) \quad (9)$$

$$\pi^{+}_{(\tau)} = w^{+}\left(\sum_{h=\tau}^{2^k} \tilde{p}_{(h)}\right) - w^{-}\left(\sum_{h=\tau+1}^{2^k} \tilde{p}_{(h)}\right) \quad (10)$$

The weighing functions for losses $w^{-}(p)$ and gains $w^{+}(p)$ are obtained by:

$$w^{-}(p) = \frac{p^{\delta}}{[p^{\delta} + (1-p)^{\delta}]^{1/\delta}} \quad (11)$$

$$w^{+}(p) = \frac{p^{\chi}}{[p^{\chi} + (1-p)^{\chi}]^{1/\chi}} \quad (12)$$

where δ and χ are weighing parameters. Risk parameters such as δ, χ, g, l and λ need to be estimated through experiments [11].

E. Equilibrium solution

In the sequential game, the defender is assumed to act as the leader, committing to a defense strategy, whereas the attacker acts as the follower, taking attack actions after observing the defender's strategies. Both players aim to maximize their CPVs given the available resources. The optimal resource allocation strategies can, then, be searched using a two-phase optimization model as follows:

$$\max_{\gamma^i} V_d(\gamma^i, \boldsymbol{\theta}^{i*}) \quad (13)$$

$$s.t.\ \boldsymbol{\theta}^{i*} = arg\max V_a(\gamma^i, \theta^i) \quad (14)$$

$$\sum_{k=1}^{K} \gamma_k^i = 1 \quad (15)$$

$$\sum_{k=1}^{K} \theta_k^i = 1 \quad (16)$$

In this formulation Equation (13) ensures that the defender optimizes the CPV considering the attacker's best response. Equation (14) ensures the attacker optimizes the CPV, given the defender's strategy. Equation (15) and (16) enforce that the resource allocation for both players are normalized, summing to 1. This two-phase optimization captures the interaction between the players' strategies and ensures that the equilibrium solution reflects their respective objectives.

III. Identification of targets in CIs

Identifying target nodes in CIs requires analyzing operational and structural characteristics. Operational metrics provide insights into the functionality of components or systems but require in-depth knowledge of CI's operations. In contrast, structural information, related to the network topology, is usually readily available and allows topology-based analyses relevant for assessing potential vulnerabilities. For this reason, many studies focus on topological metrics to identify critical nodes in CIs.

Topological metrics such as betweenness centrality, degree centrality and closeness centrality have been used to assess network vulnerabilities. Betweenness centrality, which measures a node's influence by the number of shortest paths passing through it, has been applied to evaluate critical nodes in dynamic Internet of Things networks [12] and gas transmission networks [13], for example. Degree centrality, reflecting the number of direct connections that a node has, is another commonly used metric across various domains. In power system analysis, it has been applied to assess the vulnerability of power transmission networks [14] and Integrated Energy Systems (IES) [15], identifying critical nodes that support system performance by maintaining connectivity. Closeness centrality, which quantifies a node's accessibility by calculating the inverse of the average shortest path length to other nodes, has proven effective in optimizing sensor placement for contamination detection in water distribution networks. By leveraging this metric, critical nodes can be identified to improve system monitoring and resilience [16]. Together, these metrics provide quantitative information for identifying critical target nodes in CIs.

In this way, critical nodes in CIs are identified based on their topological characteristics, recognizing that these attacks are more likely to be due to the minimal information required. For a generic i-th CI, three topological metrics are estimated for each node n_z^i: degree centrality index $I_D(n_z^i)$, betweenness centrality index $I_B(n_z^i)$ and closeness centrality index $I_C(n_z^i)$. These metrics are selected for their ability to capture distinct aspects of a node's role and influence on the network.

The degree centrality index is defined as [17]:

$$I_D(n_z^i) = \frac{degree\,(n_z^i)}{N^i - 1},\ degree(n_z^i) = \sum_{j \in N^i} a_{zj} \quad (17)$$

Here, N^i is the total number of nodes in the generic i-th CI and a_{zj} is the adjacency matrix ($a_{zj} = 1$ if nodes n_z^i and n_j^i are connected, otherwise $a_{zj} = 0$). Higher values of $I_D(n_z^i)$

indicate nodes with more direct connections, suggesting greater influence on network connectivity.

The betweenness centrality index is estimated as [18]:

$$I_B(n_z^i) = \frac{betweenness\ (n_z^i)}{\binom{N^i-1}{2}}, \quad \text{where } betweenness(n_z^i) = \sum_{s\neq z\neq t} \frac{\sigma_{st}(n_z^i)}{\sigma_{st}} \tag{18}$$

σ_{st} is the total number of shortest paths between nodes n_s^i and n_t^i, and $\sigma_{st}(n_z^i)$ is the number of paths passing through node n_z^i. The denominator $\binom{N^i-1}{2}$ represents the total number of shortest paths in the generic i-th CI. Nodes with higher values of $I_B(n_z^i)$ serve as critical bridges in the network, significantly influencing the flow of information or resources between other nodes.

The closeness index evaluates how quickly a node can reach others in the CI, and is calculated as [19]:

$$I_C(n_z^i) = \frac{N^i - 1}{\sum_{k\neq z} d(n_z^i, n_k^i)} \tag{19}$$

where $d(n_z^i, n_k^i)$ represents the shortest path distance between n_z^i and n_k^i. Higher values of $I_C(n_z^i)$ indicate nodes with shorter distances to others, implying greater accessibility and a key role in maintaining network reachability.

Each topological metric provides valuable insights but relying on a single index can overlook a node's overall importance in the CI. For example, a node in a power grid may have high degree centrality due to its numerous connections but score low on betweenness or closeness centrality if it lacks critical shortest paths or is geographically isolated. Thus, using one metric alone could lead to a myopic assessment. This underscores the need to combine metrics for a comprehensive evaluation of node criticality.

To address this, a novel structural index, $I_S(n_z^i)$ is introduced, integrating degree, betweenness and closeness centrality using the Borda count method. This method ranks nodes for each metric in descending order and assigns scores based on their rank [20]. The final structural index is computed as the sum of these scores across all metrics, providing a balanced and holistic evaluation of each node's importance in the CI.

In summary, the operational impacts of disruptions are analyzed through the inoperability of the system of systems, whereas the identification of critical nodes relies exclusively on the network's topology. This approach ensures practicality and scalability by leveraging structural information instead of detailed operational data.

IV. Case Study

The system of systems under study is made of a water network (WN) and a power network (PN): the PN relies on the WN for cooling processes and the WN depends on the PN for pumping operations. The PN is modeled using the IEEE57-bus test system and the WN's topology is derived from the IEEE39-bus system [21,22]. Nodes within these networks are classified as supplier or demand nodes based on their function in each CI [23]. Accordingly, the node sets are defined as $N^{PN} = N_S^{PN} \cup N_D^{PN}$ and $N^{WN} = N_S^{WN} \cup N_D^{WN}$.

In both networks, supplier nodes (sets N_S^{PN} and N_S^{WN}) can exist in four states: (i) fully operational $\left(x_1^{n_z^{PN}}, x_1^{n_z^{WN}}\ \forall n_z^{PN} \in N_S^{PN}, n_z^{WN} \in N_S^{WN}\right)$, (ii) disconnected due to edge failures $\left(x_2^{n_z^{PN}}, x_2^{n_z^{WN}}\right)$, (iii) overloaded when exceeding capacity $\left(x_3^{n_z^{PN}}, x_3^{n_z^{WN}}\right)$ or (iv) completely inoperable $\left(x_4^{n_z^{PN}}, x_4^{n_z^{WN}}\right)$. Similarly, demand nodes (sets N_D^{PN} and N_D^{WN}) can exist in three states: (i) fully operational $\left(x_1^{n_z^{PN}}, x_1^{n_z^{WN}}\ \forall n_z^{PN} \in N_D^{PN}, n_z^{WN} \in N_D^{WN}\right)$, (ii) disconnected due to loss of connectivity with suppliers $\left(x_2^{n_z^{PN}}, x_2^{n_z^{WN}}\right)$ or (iii) inoperable $\left(x_3^{n_z^{PN}}, x_3^{n_z^{WN}}\right)$.

The PN topology (G^{PN}) includes seven supplier nodes (N_S^{PN}) that provide electricity to fifty demand nodes (N_D^{PN}), meeting a demand of $D(X_1^{PN}) = 428.8$ MW under normal conditions (X_1^{PN}). Similarly, the WN topology (G^{WN}) includes ten supplier nodes (N_S^{WN}) that deliver $D(X_1^{WN}) = 5141.03\ m^3/h$ to twenty-nine demand nodes (N_D^{WN}) under normal conditions (X_1^{WN}). Disruptions in one network can cascade to the other, amplifying the inoperability of the system of systems. Figure 2 illustrates the network topologies.

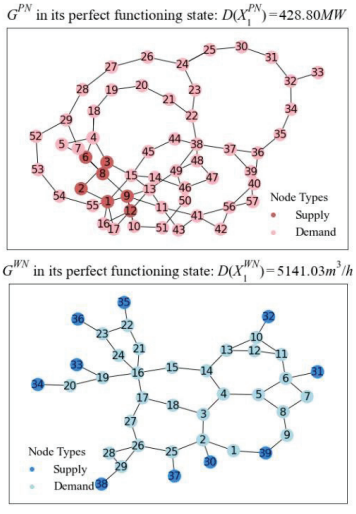

Figure 2. Topologies of G^{PN} and G^{WN}

The interdependencies $PN \leftrightarrow WN$ are illustrated in Figure 3. Each supplier node in the WN depends on power redundantly supplied by two specific nodes in the PN. Similarly, the seven supplier nodes in the PN rely on water redundantly supplied by two demand nodes in the WN. This redundancy ensures that cascading effects in the dependent CI only occur if a node in the dependent CI fails due to the failure of both nodes in the source CI that supply it. In other words, a node in the dependent CI remains operational unless both of its connected nodes in the source CI are in a state other than perfect functioning.

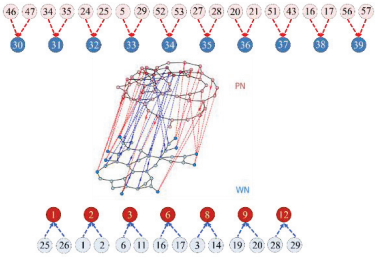

Figure 3. Physical interdependency PN↔WN

V. RESULTS

The proposed methodology was applied to the interdependent PN and WN described in the case study. As shown in Figure 1, the methodology begins with characterizing the CIs, focusing on the identification of critical nodes and their impact on the operation of the system of systems. Results highlight key nodes using topological metrics and evaluate cascading effects through the inoperability model.

Table II and Table III present the top three critical nodes identified in PN and WN, respectively. To estimate O_{SoS}, Equation (2) was applied assuming equal relative importance for each network ($w_{PN} = w_{WN} = 0.5$).

TABLE II. TOP THREE CRITICAL NODES IN THE PN ACCORDING TO I_S

Node ID	$I_D(n_z^{PN})$		$I_B(n_z^{PN})$		$I_C(n_z^{PN})$		O_{SoS}
	Value of index	Ranking	Value of index	Ranking	Value of index	Ranking	
13	0.11	1	0.28	2	0.29	1	0.01
38	0.09	3	0.32	1	0.27	3	0.01
9	0.11	1	0.24	3	0.26	4	0.009

TABLE III. TOP THREE CRITICAL NODES IN THE WN ACCORDING TO I_S

Node ID	$I_D(n_z^{WN})$		$I_B(n_z^{WN})$		$I_C(n_z^{WN})$		O_{SoS}
	Value of index	Ranking	Value of index	Ranking	Value of index	Ranking	
16	0.13	1	0.48	1	0.29	1	0.15
4	0.08	5	0.29	3	0.28	2	0.02
14	0.08	5	0.30	2	0.27	4	0.02

Based on the identified critical nodes, defense and attack strategies can be defined. The structural index, which integrates multiple metrics (I_D, I_B and I_C), effectively identifies critical nodes from multiple topological perspectives. For instance, in the PN, nodes 9 and 13 share the same value for I_D, but the structural index points to node 13 as most critical due to its high influence across other topological metrics.

The inoperability results confirm the robustness of the system of systems under single-node failures, as O_{SoS} remains low. Yet, a comparison of the last column in the Tables shows that WN attacks have a greater impact on the system of systems than attacks on the PN. Notably, attackers targeting multiple nodes, as shown in Figure 4, can significantly disrupt the system of system, with WN attacks causing higher inoperability due to the system-of-systems configuration. This underscores the importance of considering the topologies of the CIs to inform attack-defense strategies.

To advance the analysis, equilibrium strategies and payoffs are evaluated using the attack-defense game model introduced in Section II. The model considers risk preferences to reflect realistic decision-making, with parameters detailed in Table IV based on [8].

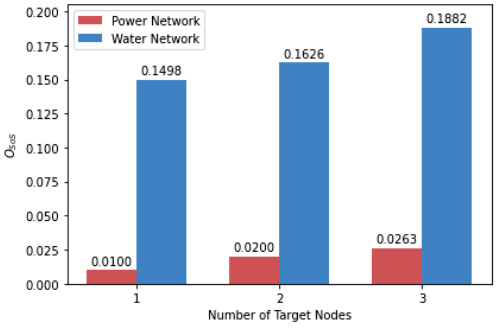

Figure 4. O_{SoS} under direct attacks on PN (red bars) and WN (blue bars)

The expected payoffs for the attacker, considering scenarios where nodes are protected or unprotected, are analyzed for PN and WN under varying numbers of targeted nodes. If a node n_k^i is unprotected, $p_k^i = 1$, indicating that the attacker is guaranteed to succeed. Conversely, if n_k^i is protected, the probability of a successful attack (p_k^i) depends on the resource investment of both players, as described by Equation (3).

TABLE IV. MODEL PARAMETERS SETTING

Contest success function (p_c)	Weighing function $w(p)$		Utility function $v(\tilde{u}_B)$		
	$w^-(p)$	$w^+(p)$	Loss-aversion factor	Risk aversion over gains	Risk seeking over losses
$m = 1$	$\delta = 0.69$	$\chi = 0.61$	$\lambda = 2.25$	$g = 0.89$	$l = 0.92$

Figure 5 shows that as the number of targeted nodes increases, the attacker's payoff rises, with the WN yielding significantly higher payoffs than the PN due to its critical role in the system-of-systems configuration. The disparity between protected and unprotected scenarios is also greater in the WN, where targeting

unprotected nodes offers a higher strategic advantage. For example, attacking three unprotected WN nodes yields a payoff of 0.2263 compared to 0.0756 for protected nodes, showing that protection mechanisms could reduce the attacker's effectiveness by over 66%. In practical terms, this means that implementing effective protection strategies in the WN could drastically limit an attacker's ability to achieve their objectives, particularly in scenarios involving coordinated attacks on multiple nodes.

On the other hand, the PN demonstrates inherent resilience under limited-scale attacks. For instance, the payoff difference is smaller when targeting a single critical node (node 13) or multiple nodes. Figure 6 confirms that defender's payoffs improve with protection measures, especially in the WN. In the PN, differences between protected and unprotected scenarios are minimal. For example, when three nodes are targeted, the defender's payoffs decrease only slightly from 0.9844 (protected) to 0.9766 (unprotected).

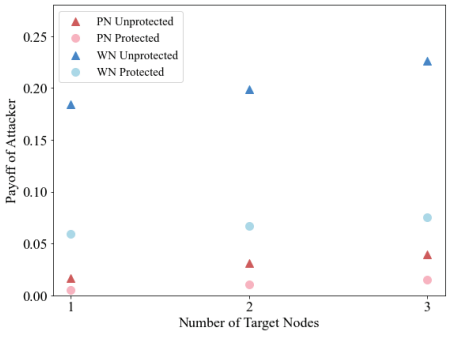

Figure 5. Attacker's expected payoffs for targeting PN (red) and WN (blue) in protected and unprotected scenarios

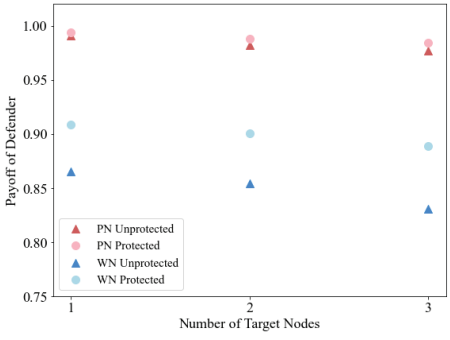

Figure 6. Defender's expected payoffs for targeting PN (red) and WN (blue) in protected and unprotected scenarios

In contrast, WN shows a more significant improvement in the defender's payoff under protection. When three nodes are targeted, the payoff rises from 0.8305 (unprotected) to 0.8889 (protected), representing an approximate 7% increase. This highlights the WN's higher susceptibility to attacks and underscores the importance of protection strategies to mitigate cascading failures. These results reinforce the critical role of targeted defense measures in enhancing system resilience, particularly for the WN.

Table V and Table VI present equilibrium solutions, considering the interactions between the attacker and defender under different target numbers, in the context of attacking the power network and the water network, respectively. It is shown that as the number of targets increases, the expected payoffs for the attacker rise, whereas the expected payoffs for the defender decrease. Both players tend to distribute their resources across multiple targets as the number of targets grows. This can be attributed to the fact that the inoperability of the system of systems grows as the number of simultaneous failed nodes rises, prompting both players to consider all targets.

TABLE V. THE EQUILIBRIUM SOLUTION FOR ATTACKING PN

Target num	Node ID	γ	θ	V_a	V_d
1	[13]	1	1	0.0053	0.9940
2	[13, 38]	0.4996, 0.5005	0.5001, 0.4999	0.0108	0.9882
3	[13, 38, 9]	0.3560, 0.3604, 0.2836	0.3475, 0.3542, 0.2984	0.0153	0.9844

TABLE VI. THE EQUILIBRIUM SOLUTION FOR ATTACKING WN

Target num	Node ID	γ	θ	V_a	V_d
1	[16]	1	1	0.0593	0.9087
2	[16, 4]	0.9087, 0.0914	0.8698, 0.1302	0.0669	0.9007
3	[16, 4, 14]	0.7846, 0.1119, 0.1036	0.6409, 0.1028, 0.2563	0.0756	0.8889

By comparing Table V and Table VI, it can be observed that, when the number of target nodes considered is three, both players tend to distribute their resources evenly across the nodes, specifically [13, 38, 9] when attacking PN because the failure of each of these node has a roughly equal impact on O_{SoS}, as indicated in Table II. In contrast, when attacking WN, the impact of targeting node 16 on O_{SoS} is considerably greater than that of nodes 4 and 14, as demonstrated in Table III: under this situation, both players prioritize the allocation of resources to node 16.

VI. CONCLUSIONS

This paper presents a game theory-based framework to model and analyze coordinated attacks and defense strategies in interdependent CIs. By integrating network theory, topological metrics and CPT, the methodology captures the complexity of intelligent attacks, where attackers strategically target nodes with high structural importance.

The framework incorporates multiple dimensions of failure, including cascading effects and escalating disruptions across interdependent CIs. The case study results demonstrate the greater vulnerability of the WN, where disruptions lead to higher inoperability compared to the PN. The model effectively captures how failures propagate through direct dependencies and amplify across the system of systems, providing insights for CI owners and planners.

The analysis also emphasizes the importance of strategic resource allocation in defense planning. Given limited protection resources, prioritization is necessary to maximize system operability. The results show that protecting critical nodes in the WN significantly reduces the attacker's payoffs, particularly in multi-node attack scenarios, where targeted defense can lower the attacker's effectiveness by over 66%. In contrast, the PN exhibits greater resilience under single-node attacks, suggesting that protection efforts should focus on nodes that contribute most to multi-node disruptions.

Concerning the computational demand, the critical node identification by topological metrics is computationally quite efficient whereas the equilibrium solution search through iterative game-theoretic calculations can be computationally intensive, especially for large-scale networks with numerous interdependent nodes. Then, for scalability of the framework, future research should explore the use of parallel computing and the introduction of heuristic search approaches. Additionally, future work will refine critical node identification by considering operational and interdependent perspectives, incorporating a deeper understanding of the attacker's knowledge and strategies. The model could also be expanded to account for potential edge failures and the costs of protecting nodes and edges.

ACKNOWLEDGMENT

The work of Maria Valentina Clavijo Mesa has been financed by - Next Generation EU, Missione 4, componente 1 "Potenziamento dell'offerta dei servizi di istruzione: dagli asili nido all'Università" – Investimento 3.4 "Didattica e competenze universitarie avanzate" e Investimento 4.1 "Estensione del numero di dottorati di ricerca e dottorati innovativi per la pubblica amministrazione e il patrimonio culturale" CUP D43C22002930001. The work of Yanfang Wu has been financed by the National Natural Science Foundation of China (Grant No. 72171195).

REFERENCES

[1] Liu X, Ferrario E, Zio E. Resilience analysis framework for interconnected critical infrastructures. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering 2017;3. https://doi.org/10.1115/1.4035728.

[2] CISA. Cyber-Attack Against Ukrainian Critical Infrastructure. Cybersecurity & Infrastructure Security Agency 2024.

[3] Zimmerman R. Human-Made Disasters: Electric Power and Transit Linked Outages. Encyclopedia of Security and Emergency Management, Cham: Springer International Publishing; 2021, p. 423–31. https://doi.org/10.1007/978-3-319-70488-3_291.

[4] Ge H, Zhao L, Yue D, Xie X, Xie L, Gorbachev S, et al. A game theory based optimal allocation strategy for defense resources of smart grid under cyber-attack. Inf Sci (N Y) 2024;652:119759. https://doi.org/10.1016/j.ins.2023.119759.

[5] Wu Y, Chen Z, Gong H, Feng Q, Chen Y, Tang H. Defender–attacker–operator: Tri-level game-theoretic interdiction analysis of urban water distribution networks. Reliab Eng Syst Saf 2021;214:107703. https://doi.org/10.1016/j.ress.2021.107703.

[6] Hausken K. Defense and attack of complex and dependent systems. Reliab Eng Syst Saf 2010;95:29–42. https://doi.org/10.1016/j.ress.2009.07.006.

[7] Peng R, Wu D, Sun M, Wu S. An attack-defense game on interdependent networks. Journal of the Operational Research Society 2021;72:2331–41. https://doi.org/10.1080/01605682.2020.1784048.

[8] Wu Y, Guo P, Wang Y, Zio E. Attack-defense game modeling framework from a vulnerability perspective to protect critical infrastructure systems. Reliab Eng Syst Saf 2025;256:110740. https://doi.org/10.1016/j.ress.2024.110740.

[9] Lin C, Xiao H, Peng R, Xiang Y. Optimal defense-attack strategies between M defenders and N attackers: A method based on cumulative prospect theory. Reliab Eng Syst Saf 2021;210:107510. https://doi.org/10.1016/j.ress.2021.107510.

[10] Clavijo-Mesa MV, Di Maio F, Zio E. "unpublished" Dynamic Inoperability Input-Output modeling of a system of systems made of multi-state interdependent critical infrastructures. Reliab Eng Syst Saf 2024.

[11] Abdellaoui M, Bleichrodt H, Paraschiv C. Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. Manage Sci 2007;53:1659–74. https://doi.org/10.1287/mnsc.1070.0711.

[12] Niu Z, Li Q, Ma C, Li H, Shan H, Yang F. Identification of Critical Nodes for Enhanced Network Defense in MANET-IoT Networks. IEEE Access 2020;8:183571–82. https://doi.org/10.1109/ACCESS.2020.3029736.

[13] Zio E, Piccinelli R. Randomized flow model and centrality measure for electrical power transmission network analysis. Reliab Eng Syst Saf 2010;95:379–85. https://doi.org/10.1016/j.ress.2009.11.008.

[14] Cadini F, Zio E, Petrescu C-A. Using Centrality Measures to Rank the Importance of the Components of a Complex Network Infrastructure, 2009, p. 155–67. https://doi.org/10.1007/978-3-642-03552-4_14.

[15] Zhang L, Su H, Zio E, Zhang Z, Chi L, Fan L, et al. A data-driven approach to anomaly detection and vulnerability dynamic analysis for large-scale integrated energy systems. Energy Convers Manag 2021;234:113926. https://doi.org/10.1016/j.enconman.2021.113926.

[16] Nazempour R, Monfared MAS, Zio E. A complex network theory approach for optimizing contamination warning sensor location in water distribution networks. International Journal of Disaster Risk Reduction 2018;30:225–34. https://doi.org/10.1016/j.ijdrr.2018.04.029.

[17] Alipour Z, Monfared MAS, Zio E. Comparing topological and reliability-based vulnerability analysis of Iran power transmission network. Proc Inst Mech Eng O J Risk Reliab 2014;228:139–51. https://doi.org/10.1177/1748006X13501652.

[18] Barthlemy M. Betweenness centrality in large complex networks. The European Physical Journal B - Condensed Matter 2004;38:163–8. https://doi.org/10.1140/epjb/e2004-00111-4.

[19] Wang L, Zheng S, Wang Y, Wang L. Identification of critical nodes in multimodal transportation network. Physica A: Statistical Mechanics and Its Applications 2021;580:126170. https://doi.org/10.1016/j.physa.2021.126170.

[20] Saari DG. Selecting a voting method: the case for the Borda count. Constitutional Political Economy 2023;34:357–66. https://doi.org/10.1007/s10602-022-09380-y.

[21] MATPOWER. Case 57 Power flow data for IEEE 57 bus test case. Https://MatpowerOrg/Docs/Ref/Matpower50/Case57Html 2014.

[22] MATPOWER. Case 39 Power flow data for 39 bus New England system 2014. https://matpower.org/docs/ref/matpower5.0/case39.html (accessed October 26, 2024).

[23] Almoghathawi Y, Barker K, Albert LA. Resilience-driven restoration model for interdependent infrastructure networks. Reliab Eng Syst Saf 2019;185:12–23. https://doi.org/10.1016/j.ress.2018.12.006.