On Hardening Problems in Critical Infrastructure Systems

Joydeep Banerjee, Kaustav Basu, Arunabha Sen

School of Computing, Informatics, & Decision Systems Engineering Arizona State University Email: jbanerje@asu.edu, kbasu2@asu.edu, asen@asu.edu

Abstract

The power and communication networks are highly interdependent and form a part of the critical infrastructure of a country. Similarly, dependencies exist within the networks itself. It is essential to have a model which captures these dependencies precisely. Previous research has proposed certain models but these models have certain limitations. The limitations of the aforementioned models have been overcome by the Implicative Interdependency Model, which uses Boolean Logic to denote the dependencies. This paper formulates the Entity Hardening problem and the Targeted Entity Hardening problem using the Implicative Interdependency Model. The Entity Hardening problem describes a situation where an operator, with a limited budget, must decide which entities to harden, which in turn would minimize the damage, provided a set of entities fail initially. The Targeted Entity Hardening problem is a restricted version of the Entity Hardening problem. This problem presents a scenario where, the protection of certain entities is of higher priority. If these entities were to be nonfunctional, the economic and societal damage would be higher when compared to other entities being nonfunctional. It has been shown that both problems are NP-Complete. An Integer Linear Program (ILP) has been devised to find the optimal solution. A heuristic has been described whose accuracy is found by comparing its performance with the optimal solution using real-world and simulated data.

Keywords: Critical Infrastructure, Entity Hardening, Targeted Entity Hardening, Power Network, Communication Network, Dependency, Interdependency.

1. Introduction

Critical Infrastructures of a nation like Power, Communication, Transportation Networks etc. exhibit strong intranetwork and inter-network dependencies to drive their functionality. The symbiotic relationship that exists between Power and Communication Network provides an example of the inter-network dependency. To elaborate this further, consider entities in either network. The Supervisory Control and Data Acquisition System (SCADA) is an integral entity in the power network controlling the electricity generation and power flow in the power grid. These controls are essentially carried out by signals generated from the communication network. Similarly, every entity in the communication network requires power to be operational. These dependencies cause failure in any of these two networks to have its impact on the other which may eventually lead to a cascade of failures.

Additionally, intra-network dependencies exist as well in a critical infrastructure. In an abstract level, a power network is composed of the following entities — Generation Bus, Load Bus, Neutral Bus and Transmission Lines. When a transmission line trips, the power flowing through the transmission lines needs to be redirected to satisfy load demand of the load buses. This may cause the power flow in some other transmission line to go beyond its line capacity causing it to trip. Eventually, these failures might result in a cascade of trippings/failures resulting in a blackout. Cascading failures in power and/or communication network due to intra/inter dependencies have disastrous effects as seen in power blackouts which occurred in New York (2003) [1], San Diego (2011) [2] and India (2012) [3]. Thus modeling these dependencies is critical in understanding and preventing such failures which might be triggered by natural as well as man-made attacks.

As noted, modeling these complex dependencies and analysis of failure in these infrastructures are considered to be highly important. A number of models have been proposed that capture these kind of dependencies [4], [5], [6], [7], [8], [9], [10], [11]. However, each of these models has their own shortcomings in bringing out the complex nature

of the dependencies that might exist [12]. Authors in [13] bring out the need to address the complex dependency which can be explained through the following example. Let a_x (which can be a generator, substation, transmission line etc.) be a power network entity and b_w , b_y , b_z (which can be a router, end system etc.) a set of communication network entities. Consider the dependency where the entity a_x is operational if (i) entities b_w and (*logical AND*) b_y are operational, or (*logical OR*) (ii) entity b_z is operational. Models in [4], [5], [6], [7], [8], [9], [10], [11] fails to capture this kind of dependency. Motivated by these findings and limitations of the existing models, the authors in [13] proposed a *Boolean logic based dependency* model termed as *Implicative Interdependency Model* (IIM). For the example stated above, the dependency of a_x on b_w , b_y , b_z can be represented as $a_x \leftarrow b_w b_y + b_z$. This equation representing the dependency of an entity is termed as *Interdependency Relation* (IDR). Using this model a number of problems were studied on interdependent power and communication infrastructure system[13], [14], and [15].

In this paper, we restrict to intra-network dependencies in Power Network and inter-network dependency in Power-Communication network and utilize the IIM to analyze and solve two important problems pertaining to critical infrastructures. For an existing critical infrastructure system, an operator would have the capability to measure the extent of failure when a certain set of entities fail initially. Consider a scenario where an operator identifies a set of critical entities which when failed initially would cause the maximum damage. In an ideal case, there would be enough resources available to support those critical entities from initial failure. However, if the availability of resources is a constraint, then an operator might have to choose entities which when supported would minimize the damage. We define the entities to support as the entities to harden and the problem as the Entity Hardening Problem. An entity x_i when hardened is resistant to both *initial* and *induced* failure (failing of entities in the cascading process after the initial failure). In the physical world, an entity can be hardened with respect to cyber attacks (say) by having a strong firewall. Similarly some entities can be hardened by -(a) strengthening their physical structures for protection from natural disaster, (b) placing redundant entity a' for an entity a which can operate when a fails, (c) increasing physical limits of the entity (maximum power flow capacity of the transmission line, maximum generation capacity of a generator bus). There exist multiple such ways to harden an entity from different kind of failures. Even though there may be circumstances under which an entity cannot be hardened, in this paper we relax such possibilities and assume that there always exist a way to harden a given entity. Hardening entities can prevent cascading failures caused by some initial failure. Thus this results in protecting a set of entities including the hardened entities from an initial failure trigger. Using these definitions the Entity Hardening Problem finds a set of k entities that should be hardened (with k being the resource constraint) in an intra-network or inter-network critical infrastructure system that protects the maximum number of entities from failure when a set of K entities fail initially.

The second problem, *Targeted Entity Hardening*, discussed in this article is a restricted version of the *Entity Hardening Problem*. For an intra-dependent power network or interdependent power and communication network, certain entities might have higher priority to be protected. There might exist entities whose non-functionality poses higher economic or societal damage as compared to other entities. For example, power and communication network entities corresponding to office buildings running global stock exchanges, the U.S. White House, transportation sectors like airports etc. presumably are more important to be protected. Let *F* denote the failed set of entities (including initial and induced failure) when a set of *K* entities fail initially. We define a set *P* (with $P \subseteq F$) of entities which have a higher priority to be protected. The *Targeted Entity Hardening* problem finds the minimum set of entities which when hardened would ensure that none of the entities in set *P* fail.

This paper is more inclined towards finding and analyzing the solution of the two problems discussed. Even though procedures are described to generate dependency equations, they are primarily intended to perform a comparative analysis of the provided solutions to the problems. For an intra-dependent or inter-dependent critical infrastructure(s) which can be modeled through IIM, the broader idea is to use the solutions for different decision-making tasks. The paper is structured as follows. The motivation behind IIM along with a formal description is provided in Section 3. The two hardening problems are more formally defined along with their Decision and Optimization Versions in Section 4. The computational complexity of the problems along with the solutions to some restricted cases is provided in Section 5. As both the problems are NP-complete, we provide an optimal Integer Linear Program (ILP) solutions to them along with sub-optimal Heuristics in Section 6. In Section 7 we describe a procedure to generate the dependency equations of the IIM model for intra-dependent power network along with a rule defined approach to generate the same for interdependent power-communication network. For power network, different bus system data are used to generate the dependency equations which are obtained from MatPower [16]. For interdependent Power-Communication network we used real world data of Maricopa County, Arizona, USA obtained from Geotel (*http://www.geo-tel.com*) for communication network and Platts (*http://www.platts.com*) for power network. In the

same section, we provide comparative studies of the heuristic to optimal ILP solutions for both the problem using the generated dependency equations.

2. Related Work

In the last few years, there has been considerable activity in the research community to study Critical Infrastructure Interdependency. One of the earliest studies on robustness and resiliency issues related to Critical Infrastructures of the U.S. was conducted by the Presidential Commission on Critical Infrastructures, appointed by President Clinton in 1996 [17]. Rinaldi et al. are among the first group of researchers to study interdependency between Critical Infrastructures and to propose the use of complex adaptive systems as models of critical infrastructure interdependencies [18], [19]. Pederson et al. in [20], provided a survey of Critical Infrastructure Interdependency modeling, undertaken by U.S. and international researchers. Motivated by the power failure event in Italy 2003, Buldyrev et al. in [4], proposed a graph-based interdependency model, where the number of nodes in the power network was assumed to be the same as the number of nodes in the communication network, and in addition there existed a one-to-one dependency between a node in the power network to a node in the communication network. The authors opine in a subsequent paper [6] that the assumption regarding one-to-one dependency relationship is unrealistic and a single node in one network may be dependent on multiple nodes in the other network. Lin et al. presented an event driven co-simulation framework for interconnected power and communication networks in [21], [22]. A game theoretic model for a multilayer infrastructure networks using flow equilibrium was proposed in [8]. Security of interdependent and identical Networked Control System (NCS) was studied in [23], where each plant was modeled by a discrete-time stochastic linear system, with systems controlled over a shared communication network. The importance of simultaneously considering power and communication infrastructures was highlighted in [24]. The results of a systematic study of human initiated cascading failures in critical interdependent societal infrastructures were reported in [25]. Focusing on the blackout of the Polish power grid, the authors in [26] studied the impact of the order of tripping of overhead lines on the severity of the failure. Analyzing failure in smart grid under targeted initial attack was studied in [27]. The effect of cyber (communication) and power network dependencies in smart grid was studied in [28] for reliability assessments. Recovery of information of the failed entities in a power grid after a failure event was studied in [29]. As described in Section 1, the models used by each of the papers have shortcomings to analyze different aspects of vulnerability in critical infrastructures. IIM is used to overcome such limitations and is utilized to address the Entity Hardening and Targeted Entity Hardening problem in this paper.

3. Implicative Interdependency Model

The need for a model to capture the complex intra and inter network dependencies is elaborated through a descriptive example of interdependent power and communication network. Consider the system shown in Figure 1 where the power network entities such as generators, transmission lines and substations are denoted by a_0 through a_{11} and communication entities such as GPS transmitters and satellites are denoted by b_0 through b_4 . The Smart Control Center (SCC) is represented by the variable c_0 as it is a part of both the power and the communication network. For the SCC to be operational, it must receive electricity either from the generator via the different power grid entities, or from the battery. Similarly, the functioning of the generator will be affected if it fails to receive appropriate control signals from the SCC. The mutual dependency between the generator and the SCC can be expressed in terms of two implicative dependency relations — (i) $a_{11} \leftarrow b_4 c_0$, (ii) $c_0 \leftarrow (b_0 b_3 (b_1 + b_2))(a_0 a_1 + a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11})$. It may be noted that the SCC will not be operational if it does not receive electric power produced at the generating station and carried over the power grid entities to the SCC and its battery backup also fails. This dependency can be expressed by the implicative relation $c_0 \leftarrow a_0 a_1 + a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11}$ implying that c_0 will be operational (i) if entities a_0 and a_1 are operational, or (ii) if entities a_2 through a_{11} are operational. However, the SCC will also not be operational if it does not receive data from the communication system (IEDs, satellites, etc.). This dependency can be expressed by the relation $c_0 \leftarrow (b_0 b_3 (b_1 + b_2))$. This implies that c_0 will be operational (i) if entities b_1 or b_2 is operational, and (ii) if entities b_0 and b_3 are operational. Combining the dependency of the SCC on the power grid and the communication network, the consolidated dependency relation can be expressed as $c_0 \leftarrow (b_0b_3(b_1+b_2))(a_0a_1+a_2a_3a_4a_5a_6a_7a_8a_9a_{10}a_{11})$. Likewise, the dependency relation for the generating station can be expressed as $a_{11} \leftarrow b_4 c_0$, implying that the generating station will not be operational unless it receives appropriate signals from the SCC c_0 , carried over wired or wireless link b_4 . These two implicative relations demonstrate that dependency (or interdependency) is a complex combination of conjunctive and disjunctive terms. We term the model capturing this complex dependencies and interdependencies as *Implicative Interdependency Model*.



Figure 1: Example of Power - Communication Infrastructure Interdependency

In the IIM an intra-network or inter-network critical infrastructure system is represented by $I(E, \mathcal{F}(E))$, where *E* is the set of entities and $\mathcal{F}(E)$ is the set of dependency relations. Throughout this paper, an intra-dependent critical infrastructure or interdependent critical infrastructure is termed as *system* denoted by $I(E, \mathcal{F}(E))$. The dynamics of the model is explained through an example. Consider sets *A* and *B* (with $E = A \cup B$) representing entities in power and communication network (say) with $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3, b_4\}$ respectively. The function $\mathcal{F}(E)$ giving the set of dependency equations are provided in Table 1. In the given example, an IDR $b_3 \leftarrow a_2 + a_1a_3$ implies that entity b_3 is operational if entity a_2 or entity a_1 and a_3 are operational. In the IDRs each conjunction term e.g. a_1a_3 is referred to as *minterms*.

Power Network	Comm. Network
$a_1 \leftarrow b_2$	$b_1 \leftarrow a_1 + a_2$
$a_2 \leftarrow b_2$	$b_2 \leftarrow a_1 a_2$
$a_3 \leftarrow b_4$	$b_3 \leftarrow a_2 + a_1 a_3$
	$b_4 \leftarrow a_3$

Table 1: IDRs for the constructed example

Entities	Time Steps (<i>t</i>)						
	0	1	2	3	4	5	6
<i>a</i> ₁	0	0	1	1	1	1	1
a_2	1	1	1	1	1	1	1
<i>a</i> ₃	1	1	1	1	1	1	1
b_1	0	0	0	1	1	1	1
b_2	0	1	1	1	1	1	1
b_3	0	1	1	1	1	1	1
b_4	0	1	1	1	1	1	1

Table 2: Failure cascade propagation when entities $\{a_2, a_3\}$ fail at time step t = 0. A value of 1 denotes entity failure, and 0 otherwise

Initial failure of entities in $A \cup B$ would cause the failure to cascade until a steady state is reached. As noted earlier, the event of an entity failing after the initial failure is termed as *induced failure*. Failure in IIM proceeds in unit time steps with *initial failure* starting at time step t = 0. Each time step captures the effect of entities killed in all previous time steps. We demonstrate the cascading failure for the interdependent network outlined in Table 1 through an example. Consider the entities a_2 and a_3 fail at time step t = 0. Table 2 represents the cascade of failure in each

subsequent time steps. In Table 2, for a given entity and time step, '0' represents the entity is operational and '1' non operational. In this example a steady state is reached at time step t = 3 when all entities are non operational. IIM also assumes that the dependent entities of all failed entities are killed immediately at the next time step. For example at time step t = 1 entities a_2, a_3, b_2, b_3 and b_4 are non operational. Due to the IDR $a_1 \leftarrow b_2$ entity a_1 is killed immediately at time step t = 2. At t = 3 the entity b_1 is killed due to the IDR $b_1 \leftarrow a_1 + a_2$ thus reaching the steady state.

As noted earlier The model captures the cascading failure that propagates through the entities on an event of *initial failure*. Consider $E = A \cup B$ with A and B representing entities in two separate critical infrastructures. The cascading failure process is shown diagrammatically in Figure 2a with sets $A_d^0 \subset A$ and $B_d^0 \subset B$ failing at t = 0. Accordingly, cascading failure in these systems can be represented as a *closed loop* control system shown in Figure 2b. The steady state after an initial failure is analogous to the computation of *fixed point* of a function $\mathcal{G}(.)$ such that $\mathcal{G}(A_d^p \cup B_d^p) = A_d^p \cup B_d^p$, with steady state reached at t = p. It can be followed directly that for an interdependent system with |E| = m, any initial failure would cause the system to reach a steady state within m - 1 time steps.



(a) Cascading failures reach steady state after p time steps

(b) Cascading failures as a fixed point system

Figure 2: Cascading Failures in Multi-layered Networks

We note some of the challenges in generating the IDRs. The main challenge is an accurate formulation of the IDRs. Two possible ways of doing this would be (i) careful analysis of the underlying infrastructures as in [11], (ii) consultation with domain experts. In Section 7 we provide techniques to generate IDR for power network and interdependent power-communication network. However, the underlying assumptions behind the technique pose some limitations on its applicability to the real world problems. The formulation of IDRs from the interdependent network is an ongoing research and the problem is solved under the reasonable assumption that these IDRs can be developed.

4. Problem Formulation

As discussed in Section 1, an entity when hardened, is protected from both initial and induced failures. With this understanding we formally describe the two hardening problems — Entity Hardening and Targeted Entity Hardening. It is to be noted both the problems return a set of entities to harden. The approach that should be taken to physically harden the entities rest upon the properties of the entities.

4.1. Entity Hardening Problem

Before stating the problem formally, a brief understanding of entity hardening is provided. Consider the system with set of dependency relations given by Table 1. With an initial failure of entities a_2, a_3 the subsequent cascading failures is shown in Table 2 which fails all the entities in the system. We note three instances where entities a_1, a_2 and a_3 are hardened separately with a_2, a_3 failing initially. The failure cascade propagation when a_1, a_2 and a_3 are hardened are shown in Tables 3, 4, and 5 respectively. In the tables the cascading failure is shown till t = 3 because with initial failure of entities a_2, a_3 the cascade propagation stops at t = 3 as seen in Table 2. Hardening entity a_1 protect entities a_1, b_1 from failure. Similarly, when a_2 is hardened it protect a_1, a_2, b_1, b_2, b_3 and hardening a_3 protect entities a_3, b_4 . If the hardening budget is 1 the operator would clearly harden the entity a_2 as it protects the maximum number of entities from failure. We now describe the entity hardening problem formally.

The Entity Hardening (ENH) problem

INSTANCE: Given:

(i) A system $I(E, \mathcal{F}(E))$, where the set E represent the set of entities, and $\mathcal{F}(E)$ the set of IDRs.

- (ii) The set of *K* initially failing entities E', where $E' \subseteq E$
- (iii) Two positive integers k, k < K and E_F .

Entities	Tiı	ne S	teps	<i>(t)</i>
	0	1	2	3
a_1	0	0	0	0
a_2	1	1	1	1
<i>a</i> ₃	1	1	1	1
b_1	0	0	0	0
b_2	0	1	1	1
b_3	0	1	1	1
b_4	0	1	1	1

Table 3: Failure cascade propagation when entities $\{a_2, a_3\}$ fail at time step t = 0 and a_1 is hardened

Entities	Time Steps (t)						
	0	1	2	3			
a_1	0	0	0	0			
a_2	0	0	0	0			
a_3	1	1	1	1			
b_1	0	0	0	0			
b_2	0	0	0	0			
b_3	0	0	0	0			
b_4	0	1	1	1			

Table 4: Failure cascade propagation when entities $\{a_2, a_3\}$ fail at time step t = 0 and a_2 is hardened

Entities	Time Steps (t)							
	0	1	2	3				
a_1	0	0	1	1				
a_2	1	1	1	1				
a_3	0	0	0	0				
b_1	0	0	0	1				
b_2	0	1	1	1				
b_3	0	1	1	1				
b_4	0	0	0	0				

Table 5: Failure cascade propagation when entities $\{a_2, a_3\}$ fail at time step t = 0 and a_3 is hardened

DECISION VERSION: Is there a set of entities $\mathcal{H} = E'', E'' \subseteq E, |\mathcal{H}| \leq k$, such that hardening \mathcal{H} entities results in no more than E_F entities to fail after entities in E' fail at time step t = 0.

OPTIMIZATION VERSION: Find a set of k entities to harden which would maximize the number of protected entities with entities in E' failing initially.

Definition 1. KillSet(S): For an initial failure of set S, the set of entities that fail due to induced failure in the cascading process including the entities in set S is denoted by KillSet(S).

The following points are to be noted regarding the ENH problem — (a) the condition k < K is assumed as with $k \ge K$ hardening the K initially failing entities would ensure that there are no induced and initial failure. (b) with E' entities failing initially, the entities to be harden are to be selected from *KillS et*(E'). Hardening entities outside *KillS et*(E') would not result in protection of any non-hardened entity.

4.2. Targeted Entity Hardening Problem

Qualitatively, for a system $I(E, \mathcal{F}(E))$ the objective of the Targeted Entity Hardening problem is to choose a minimum cardinality set of entities to harden, with a set of initially failing entities, such that all entities in a given set P are protected from failure. We use the example with dependency equations outlined in Table 1 to describe the Targeted Entity Hardening Problem with $P = \{b_4\}$. With $\{a_2, a_3\}$ being the two entities failing initially, hardening entity a_2 (with a_3 failing) would prevent failure of entities a_1, a_3, b_1, b_1, b_3 . Similarly, hardening the entity a_3 (with a_2 failing) would prevent the failure of entity b_4 . Even though hardening a_2 prevent failure of more entities than hardening a_3 , owing to the problem description a_3 has to be hardened which is a solution to the Targeted Entity Hardening problem in this scenario. It is to be noted that other entities might also be protected from failure when a set of entities are hardened to protect a given set of entities. The Targeted Entity Hardening problem is formally stated below accompanied with a descriptive diagram provided in Figure 3 (in the figure direct failure means initial failure) —

The Targeted Entity Hardening (TEH) problem

INSTANCE: Given: (i) A system $I(E, \mathcal{F}(E))$, where the set *E* represent the set of entities, and $\mathcal{F}(E)$ is the set of IDRs. (ii) The set of *K* entities failing initially *E'*, where $E' \subseteq E$. (iii) The set $F \subseteq E$ contains all the entities failed due to initial failure of *E'* entities i.e. *KillS et*(*E'*) (iv) A positive integer *k* and *k* < *K*. (v) A set $P \subseteq F$.

DECISION VERSION: Is there a set of entities $H = E'' \subseteq E$, $|H| \leq k$, such that hardening H entities would result in protecting all entities in the set P after entities in E' fail at the initial time step.

OPTIMIZATION VERSION: Find the minimum set of entities in E to harden that would result in protecting all entities in the set P after entities in E' fail at the initial time step.



Figure 3: Pictographic description of the Targeted Entity Hardening problem

5. Computational Complexity Analysis

The computational complexity for both the problems are provided in this section. The problems are proved to be NP-complete. Additionally, approximate and polynomial solutions to few subcases are provided. The subcases impose restrictions on the IDRs and the solutions can be applied to interdependent systems whose dependency equations fall within the definition of the restriction.

5.1. Entity Hardening Problem

We prove that the ENH problem is NP-complete in Theorem 2. Using the results of Theorem 2 an in-approximability bound of the problem is provided in Theorem 3.

Theorem 2. The ENH Problem is NP Complete

Proof. The Entity Hardening problem is proved to be NP complete by giving a reduction from the Densest *p*-Subhypergraph problem [30], a known NP-complete problem. An instance of the Densest *p*-Subhypergraph problem includes a hypergraph $G = (V, E_V)$, a parameter *p* and a parameter *M*. The problem asks the question whether there exists a set of vertices $|V'| \subseteq V$ and $|V'| \leq p$ such that the subgraph induced with this set of vertices has at least *M* hyperedges that are completely covered. From an instance of the Densest *p*-Subhypergraph problem we create an instance of the ENH problem in the following way. Consider a system $I(E, \mathcal{F}(E))$ with $E = A \cup B$, where *A* and *B* are entities of two separate critical infrastructures dependent on each other. For each vertex v_i and each hyperedge e_j entities b_i and a_j are added to the set *B* and *A* respectively. For each hyperedge e_j with $e_j = \{v_m, v_n, v_q\}$ (say) an IDR of form $a_j \leftarrow b_m b_n b_q$ is created. It is assumed that the value of *K* is set of |V|. The values of *k* and E_F are set to *p* and $|V| + |E_V| - p - M$ (where |A| = |V| and |B| = |E|) respectively.

In the constructed instance only entities of set *A* are dependent on entities of set *B*. Additionally the dependency for an entity a_i consists of conjunction of entities in set *B*. Hence for an entity $a_i \in A$ to fail, either it itself has to fail initially or any one of the entity that a_i depends on has to fail. It is to be noted that the entities in set *B* has no induced failure i.e., there is no cascade. Following from this assertion, with K = |V'|, failing entities in *B* would fail all entities in set $A \cup B$. For this created instance E' is set to B'

If an entity in set *A* is hardened then it would have no effect in failure prevention of any other entities. Whereas hardening an entity $b_m \in B$ might result in failure prevention of an entity $a_i \in A$ with IDR $a_j \leftarrow b_m b_n b_q$ provided that entities b_n, b_q are also defended. With k = p (and $K \le |V| = |B|$) it can be ensured that entities to be defended are from set *B'*.

To prove the theorem, consider that there is a solution to the Densest *p*-Subhypergraph problem. Then there exist *p* vertices which induces a subgraph which has at least *M* hyperedges. Hardening the entities $b_i \in B'$ for each vertex v_i in the solution of the Densest *p*-Subhypergraph problem would then ensure that at least *M* entities in set *A* are protected from failure. This is because the entities in set *A* for which the failure is prevented corresponds to the hyperedges in the induced subgraph. Thus the number of entities that fail after hardening *p* entities is at most $|V| + |E_V| - p - M$, solving the ENH problem. Now consider that there is a solution to the ENH problem. As previously stated, the entities to be hardened will always be from set *B'*. So defending *p* entities from set *B'* would result in failure prevention of at least *M* entities in set *A* such that $E_F \leq |V| + |E_V| - p - M$. Hence, the vertex induced subgraph would have at least *M* hyperedges completely covered when vertices corresponding to the entities hardened are included in the solution of the Densest *p*-Subhypergraph problem. Hence proved.

Theorem 3. For a system $I(E, \mathcal{F}(E))$ with n = |E| and $\mathcal{F}(E)$ having IDRs of form in the created instance of Theorem 2, the ENH problem is hard to approximate within a factor of $\frac{1}{2\log(n)^{\lambda}}$ for some $\lambda > 0$.

Proof. The ENH problem with IDRs of form in the created instance of Theorem 2 is a special case of the densest *p*-subhypergraph problem is proved to be inapproximable within a factor of $\frac{1}{2\log(n)^4}$ ($\lambda > 0$). The same result applied to the ENH problem as well. Hence the theorem follows.

5.1.1. Restricted Case I: Problem Instance with One Minterm of Size One

The IDRs of this restricted case have a single minterm of size 1. This can be represented as $e_i \leftarrow e_j$, where e_i and e_j are entities of a system $\mathcal{I}(E, \mathcal{F}(E))$. Algorithm 1 solves the ENH problem with this restriction optimally in polynomial time utilizing the notion of *Kill Set* defined in Definition 1 with proof of optimality given in Theorem 4.

Algorithm 1: Entity Hardening Algorithm for systems with Restricted Case I type of dependencies
Data: A system $\mathcal{I}(E, \mathcal{F}(E))$, set of K entities failing initially $E', E' \subseteq E$, hardening budget k
Result: Set of hardened entities \mathcal{H}
1 begin
For each entity $e_i \in E'$ compute the set of kill sets and store it in a set $C = \{C_{e_1}, C_{e_2},, C_{e_K}\}$, where $C_{e_i} = KillSet(e_i)$;
3 Set $\mathcal{H} = \emptyset$;
4 for $(i=1; i \leq \mathcal{K}; i++)$ do
5 Choose the set C_{e_k} having the highest cardinality from C ;
$6 \qquad \qquad \text{Update } C \leftarrow C \setminus C_{e_k};$
7 for $C_{e_i} \in C$ do
8 Update $C_{e_j} \leftarrow C_{e_j} \setminus C_{e_k}$;
9 Update $\mathcal{H} \leftarrow \mathcal{H} \cup \{e_k\};$
10 If all Kill Sets are empty then break ;
11 return \mathcal{H}

Theorem 4. Algorithm 1 solves the Entity Hardening problem for the Restricted Case I optimally in polynomial time.

Proof. It is shown in [13] that the kill set for all entities in the interdependent network can be computed in $O(n^3)$ where n = |E|. Thus computing the kill sets of K entities would have a time complexity of $O(Kn^2)$. Each update in line 8 would take O(n) time and hence the total computation of the inner for loop can be done in O(Kn). The outer for loop iterates for K times thus the time complexity of lines 4 - 9 is $O(K^2n)$. Hence Algorithm 1 runs in $O(Kn^2)$.

For two kill sets C_{e_i} and C_{e_j} , it can be shown that either $C_{e_i} \cap C_{e_j} = \emptyset$ or $C_{e_i} \cap C_{e_j} = C_{e_i}$ or $C_{e_i} \cap C_{e_j} = C_{e_j}$ [13]. Using this assertion the set E' can be partitioned into disjoint subsets $E_{X_1}, E_{X_2}, ..., E_{X_m}$ where kill sets of two entities e_a, e_b have no elements in common with $e_a \in E_{X_i}$ and $e_b \in E_{X_j}$ and $i \neq j$. Additionally, for any given subset of entities E_{X_z} there exist an entity $e_k \in E_{X_z}$ whose kill set is a super set of kill sets of all other entities in E_{X_z} . Thus each of the disjoint subset has an entity whose kill set is the super set among all other entities in that subset. Algorithm 1 chose such an entity in line 5 for every iteration and updates in line 8 would make the kill set of all the remaining entities in the partition to be empty and hence would not be hardened in future iterations. Clearly choosing these entities would globally maximize the total number of protected entities from failure. Hence the Algorithm 1 is proved to be optimal.

5.1.2. Restricted Case II: Problem Instance with an Arbitrary Number of Minterm of Size One

The IDRs of this restricted case have arbitrary number of minterm of size 1. This can be represented as $e_i \leftarrow \sum_{q=1}^{p} e_q$, where e_i and e_q are entities of a system $I(E, \mathcal{F}(E))$ and the number of minterms are p. The ENH problem with respect to this restricted case is NP-complete and is proved in Theorem 5. We provide an approximation bound for this restricted case of the problem in Theorem 8 using the results of Theorem 5. The approximation bound uses the notion of *Protection Set* (Definition 6). The Protection Set of an entity can be computed in $O((n)^2)$ where n = |E| and m are number of minterms.

Theorem 5. The ENH problem for Restricted Case II is NP Complete

Proof. The ENH problem for case III is proved to be NP complete by giving a reduction from the Set Cover Problem. An instance of the Set Cover problem is given by a set $S = \{x_1, x_2, ..., x_n\}$ of elements, a set of subsets $S = \{S_1, S_2, ..., S_m\}$ where $S_i \subseteq S$ and a positive integer M. The decision version of the problems finds whether there exist at most M subsets from set S whose union would result in the set S. From an instance of the set cover problem we create an instance of the ENH problem in the following way. Consider a system $I(E, \mathcal{F}(E))$ with $E = A \cup B$, where A and B are entities of two separate critical infrastructures dependent on each other. For each element x_i in set S we add an entity a_i in set A. For each subset S_i in set S we add an entity b_i in set B. For all subsets in S, say S_p, S_m, S_n , which has the element x_i an IDR of form $a_i \leftarrow b_m + b_n + b_l$ is added to $\mathcal{F}(E)$. The values of positive integers k and E_F are set to M and m - M respectively. It is assumed that the value of K = m and E' = B.

The constructed instance ensures that the entities to be hardened are from set *B*. This is because hardening an entity $a_i \in A$ would only result in prevention of its own failure whereas hardening an entity $b_j \in B$ would result in failure prevention of its own and all other entities in set *A* for which it appears in its IDR.

Consider there exists a solution to the Set Cover problem. Then there exist M subsets whose union results in the set S. Hardening the entities in set B corresponding to the subsets selected would ensure that all entities in set A are prevented from failure. This is because for the dependency of each entity $a_i \in A$ there exist at least one entity (in set B) that is hardened. Hence the number of entities that fails after hardening is m - M which is equal to E_F , thus solving the ENH problem. Now, consider that there is a solution to the ENH problem. As discussed above the entities to be hardened should be from set B'. To achieve $E_F = m - M$ with k = M, no entities in the set A must fail. Hence for each entity $a_i \in A$ at least one entity in set B that appears in its IDR has to be hardened. Thus, it directly follows that the union of subsets in set S is equal to the set S, solving the Set Cover Problem. Hence proved.

Definition 6. For an entity $e_i \in E$ the Protection Set is defined as the entities that would be prevented from failure by hardening the entity e_i when all entities in E' fail initially. This is represented as $P(x_i|E')$.

Theorem 7. For two entities $e_i, e_j \in A \cup B$, $P(e_i|E') \cup P(e_j|AE') = P(e_i, e_j|E')$ when IDRs are in form of Restricted *Case II.*

Proof. Assume that defending two entities e_i and e_j would result in preventing failure of $P(e_i, e_j|E')$ entities with $|P(e_i|E') \cup P(e_j|E')| < |P(e_i, e_j|E')|$. Then there exist at least one entity $e_p \notin P(e_i|E') \cup P(e_j|E')$ such that it's failure is prevented only if e_i and e_j are protected together. So two entities e_m and e_n (with $e_m \in P(e_i|E')$ and $e_n \in P(e_j|E')$ or vice versa) have to be present in the IDR of e_p . As the IDRs are of restricted Case II so if any one of e_m or e_n is protected then e_p is protected, hence a contradiction. On the other way round $P(e_i, e_j|E')$ contains all entities which would be prevented from failure if e_i or e_j is defended alone. So it directly follows that $|P(e_i|E') \cup P(e_j|E')| > |P(e_i, e_j|E')|$ is not possible. Hence the theorem holds.

Theorem 8. There exists an $1 - \frac{1}{e}$ approximation algorithm that approximates the ENH problem for Restricted Case II.

Proof. The approximation algorithm is constructed by reducing the problem for this restricted case to *Maximum Coverage* problem. An instance of the maximum coverage problem consists of a set $S = \{x_1, x_2, ..., x_n\}$, a set $S = \{S_1, S_2, ..., S_m\}$ where $S_i \subseteq S$ and a positive integer M. The objective of the problem is to find a set $S' \subseteq S$ and $|S'| \leq M$ such that $\bigcup_{s_i \in S} S_i$ is maximized. Consider a system $I(E, \mathcal{F}(E))$ with $E = A \cup B$, where A and B are entities of two separate critical infrastructures dependent on each other. For a given initial failure set $E' = A' \cup B'$ with $|A'| + |B'| \leq K$, let $P(e_i|A' \cup B')$ denote the protection set for each entity $e_i \in A \cup B$. We construct a set $S = A \cup B$ and for each entity e_i a set $S_{e_i} \subseteq S$ such that $S_{e_i} = P(e_i|A' \cup B')$. Each set S_{e_i} is added as an element of a set S. The conversion of the problem to Maximum Coverage problem can be done in polynomial time. By Theorem 7 defending a set of entities $X \subseteq S$ would result in failure prevention of $\bigcup_{e_i \in X} S_{x_i}$ entities. Hence, with the constructed sets S and S and a positive integer M (with M = k) finding the Maximum Coverage would ensure the failure protection of maximum number of entities in $A \cup B$. This is same as the ENH problem of Restricted Case II. As there exists an $1 - \frac{1}{e}$ approximation algorithm for the Maximum Coverage problem hence the same algorithm can be used to solve this restricted case of the ENH problem using this transformation.

5.2. Targeted Entity Hardening Problem

In this subsection we prove the computational complexity of the Targeted Entity Hardening Problem to be NPcomplete in Theorem 9.

Theorem 9. The TEH problem is NP-complete

Proof. We proof that the Targeted Entity Hardening is NP complete by a reduction from Set Cover problem. An instance of the Set Cover problem consists of (i) a set of elements $U = \{x_1, x_2, ..., x_n\}$, (ii) a set of subsets $S = \{S_1, S_2, ..., S_m\}$ with $S_i \subseteq U \forall S_i \in S$, and (iii) a positive integer M. The problem asks the question whether there is a subset S' of S with $|S'| \leq M$ such that $\bigcup_{S_k \in S'} S_k = U$. From an instance of the Set Cover problem we create an instance of the Targeted Entity Hardening Problem as follows. Consider a system $I(E, \mathcal{F}(E))$ with $E = A \cup B$, where A and B are entities of two separate critical infrastructures dependent on each other. For each element x_j in U we add an entity a_j in set A. Similarly for each subset S_i in set S we add an entity b_i in set B. For each element $x_i \in U$ which appears in subsets $S_m, S_n, S_p \in S$ (say) we add an IDR $a_i \leftarrow b_m + b_n + b_p$ to $\mathcal{F}(E)$. There are no IDRs for entities in $A \cup B$. The set of P entities to be protected is set to A and k is set to M.

Consider there exists a solution to the Set Cover problem. Then there exist a set S' of cardinality M such that $\bigcup_{S_k \in S'} S_k = U$. For each subsets $S_k \in S'$ we harden the entity $b_k \in B$. So in each IDR of the A type entities there exist a B type entity that is hardened. Hence all A type entities will be protected from failure thus solving the Targeted Entity Hardening problem.

On the other way round consider there is a solution to the Targeted Entity Hardening problem. This ensures either that for each entity $a_j \in A$ (i) a_j itself is hardened, or (ii) at least one entity from set B in a_j 's IDR is hardened. For scenario (i) arbitrarily select an entity b_p in a_j 's IDR and include it in set C. For scenario (ii) include the hardened entities in the IDR of a_j into set C. This is done for each entity $a_j \in A$. For each entity in set C select the corresponding subset in set S. The union of these set of subsets would result in the set U. Thus solving the set cover problem. Hence the theorem is proved.

5.2.1. Restricted Case I: Problem Instance with One Minterm of Size One

This restriction imposed on the IDRs is the same as that of restricted case I of the ENH problem. Using the definition of *Protection set* (Definition 6) and the result in Theorem 10 we design an algorithm (Algorithm 2) that solves the problem for this restricted case optimally in polynomial time (proved in Theorem 11).

Theorem 10. Given a system $I(E, \mathcal{F}(E))$ with IDRs of form restricted case I and $E' \subset E$ entities failing initially, for any entity e_i and e_j with $e_i \neq e_j$ either (a) $PS(e_i|E') \subseteq PS(e_j|E')$, (b) $PS(e_j|E') \subseteq PS(e_i|E')$, or (c) $PS(e_i|E') \cap PS(e_j|E') = \emptyset$.

Proof. Consider a directed graph $G = (V, E_D)$. The vertex set V consists of a vertex for each entity in E. For each IDR of form $y \leftarrow x$ there is a directed edge $(x, y) \in E_D$. In this proof the term vertex and entity is used interchangeably as an entity is essentially a vertex in G. It can be shown that G is either (a) Directed Acyclic Graph (DAG) with maximum in-degree of at most 1 or, (b) contain at most one cycle with no incoming edge to any vertex in the cycle and maximum in-degree of at most 1, or (c) collection of graphs (a) and/or (b). Consider a vertex $x_i \in V$. Let $G' = (V', E'_D)$ be a subgraph of G with V' consisting of x_i and all the vertices that has a directed path from x_i . Moreover, the edge set E'_D consists of all edges $(x, y) \in E_D$ with $x, y \in V'$ except for any edge (y, x_i) with $y_i \in V'$. Such a subgraph G' would be a directed tree with (i) one or more entities in $V' \setminus \{x_i\}$ is in $A' \cup B'$. Let X denote the set of such entities which satisfy this property, or (ii) no entities in $V' \setminus \{x_i\}$ is in E'. If the entity x_i is hardened then for case (i) all the entities in $V' \otimes (x_i|E') \cap PS(e_j|E') = \emptyset$. For case (ii) for any entity $x_j \in V'$ the condition $PS(e_j|E') \subseteq PS(e_i|E')$ always holds (the equality holds for graphs of type (b) as stated above). This property holds for all entities in the entity set E. Hence proved.

Theorem 11. Algorithm 2 solves the Targeted Entity Hardening problem with IDRs having single minterms of size 1 optimally in polynomial time.

Algorithm 2: Algorithm for TEH problem with IDRs in form of Restricted Case I

Algorithm 2. Algorithm for TET problem with DAs in form of Resulted Case 1	
Data: A system $I(E, \mathcal{F}(E))$, set E' with $ E' = K$ entities failing initially and the set P of entities to be protected from	
failure.	
Result: A set of entities <i>H</i> to be hardened.	
1 begin	
2 For each entity $e_i \in (E)$ compute the Protection Sets $PS(e_i E')$;	
3 Initialize $H = \emptyset$;	
4 while $P \neq \emptyset$ do	
5 Choose the Protection Set with highest $ PS(e_i E') \cap P $;	
6 Update $H \leftarrow H \cup \{e_i\}$;	
7 Update $P \leftarrow P \setminus PS(e_i E');$	
s for all $d_i \in E$ do	
9 $PS(e_{j} E') = PS(e_{j} E') PS(e_{i} E');$	
10 return H ;	

Proof. The *Protection Sets* of the entities can be found in a similar way as that of computing *Kill Sets* defined in [13]. It can be shown that computing these sets for all entities in *E* can be done in $O(n^3)$ where n = |E|. The while loop in Algorithm 2 iterates for a maximum of *n* times. Step 5 can be computed in $O(n^2)$ time. The for loop in step 8 iterates for *n* times. For any given e_j and e_i , $PS(e_j|E') = PS(e_j|E') \setminus PS(e_i|E')$ can be computed in $O(n^2)$ time with the worst case being the condition when $|PS(e_i|E')| = |PS(e_j|E')| = n$. As step 9 is nested in a for loop within the while loop this accounts for the most expensive step in the algorithm. The time complexity of this step is $O(n^4)$. Thus Algorithm 2 runs polynomially in *n* with time complexity being $O(n^4)$.

In Algorithm 2 the while loop iterates till all the entities in *P* are protected from failure. In step 5 the entity e_i with protection set $PS(e_i|E')$ having most number of entities belonging to set *P* is chosen to be hardened. Correspondingly the entity e_i is added to the hardening set *H*. The set *P* is updated by removing the entities in $PS(e_i|E')$. Similarly all the protection sets are updated by removing the entities in $PS(e_i|E')$.

We use the result from Theorem 10 to prove the optimality of Algorithm 2. An entity e_i is selected to be hardened at any iteration of the while loop has maximum number of entities in $PS(e_i|E') \cap P$. All entities e_j with $PS(e_j|E') \subseteq$ $PS(e_i|E')$ would have $PS(e_j|E') \cap P \subseteq PS(e_i|E') \cap P$. Moreover there exist no entity e_k for which $PS(e_i|E') \subset PS(e_k|E')$ otherwise e_k would have been hardened instead. Hence there exist no other entity that protect other entities in Pincluding $PS(e_i|E') \cap P$. So Algorithm 2 selects the minimum number of entities to harden that protects all entities in P.

5.2.2. Restricted Case II: IDRs having arbitrary number of minterm of size 1

For instance created in Theorem 9 the IDRs were logical disjunctions of minterms with size 1. We consider this restriction to design an approximation algorithm for the TEH problem and is shown in Theorem 12.

Theorem 12. The Targeted Entity Hardening Problem is O(log(|P|)) approximate when IDRs are logical disjunctions of minterms with size 1.

Proof. We first compute the protection set $PS(e_i|E')$ for all entities $e_i \in E$. Each protection set is pruned by removing entities that are not in set P. Now the Targeted Entity Hardening Problem can be directly transformed into Minimum Set Cover problem by setting U = P and $S = \{PS(e_1|E'), PS(e_2|E'), ..., PS(x_i|E||E')\}$. Selecting the corresponding entities of the protection sets that solve the Minimum Set Cover problem would also solve the Targeted Entity Hardening problem. There exists an approximation ratio of order O(log(n)) (where n is the number of elements in set U) for the Set Cover problem. Therefore using the approximation algorithm that solves the Set Cover problem, the same ratio holds for the Targeted Entity Hardening problem with n = |P|. Hence proved.

6. Optimal and Heuristic Solution to the Problems

Owing to the problems being NP-complete, we provide optimal solutions to them by formulating Integer Linear Program (ILP). For both the problems we also provide sub optimal heuristic that runs in polynomial time.

6.1. Solutions to the Entity Hardening Problem

6.1.1. Optimal Solution using Integer Linear Programming

We propose an Integer Linear Program (ILP) that solves the ENH problem optimally. For a system $I(E, \mathcal{F}(E))$ let $G = \{g_1, g_2, ..., g_n\}$ be variables denoting entities in set E. Given an integer K, G is a array of K 1's and n - K 0's where $g_i = 1$ if the entity $e_i \in E$ fails at t = 0 and $g_i = 0$ if the the entity is operational at t = 0. Thus the array G gives the set of K entities failing initially. Additionally for each entity $e_j \in E$ a set of variables x_{jd} with $0 \le d \le n - 1$ and $d \in I^+ \cup \{0\}$ are created. The value of $x_{jd} = 1$ denotes that the entity x_j is in failed state at t = d and $x_{jd} = 0$ denotes it is operational. As noted earlier for |E| = n the cascade can proceed till n - 1 so the range of d is [0, n - 1]. Using these definitions the objective of the ENH problem is as follows —

$$\min\left(\sum_{i=1}^{n} x_{i(n-1)}\right) \tag{1}$$

The constraints guiding the problem are as follows:

Constraint Set 1: $\sum_{i=1}^{n} q_{x_i} \le k$, with $q_{x_i} \in [0, 1]$. If an entity x_i is hardened then $q_{x_i} = 1$ and 0 otherwise.

Constraint Set 2: $x_{i0} \ge g_i - q_{x_i}$. This constraint implies that only if an entity is not defended and $g_i = 1$ then the entity will fail at the initial time step.

Constraint Set 3: $x_{id} \ge x_{i(d-1)}, \forall d, 1 \le d \le n-1$, in order to ensure that for an entity which fails in a particular time step would remain in failed state at all subsequent time steps.

Constraint Set 4: Modeling of constraints to capture the cascade propagation for IIM is similar to the constraints established in [13] with modifications to capture the hardening process. A brief overview of this constraint is provided here. Consider an IDR $e_i \leftarrow e_j e_p e_l + e_m e_n + e_q$. The following steps are enumerated to depict the cascade propagation with respect to this constraint:

Step 1: Replace all minterms of size greater than one with a variable. In the example provided we have the transformed minterm as $e_i \leftarrow c_1 + c_2 + e_q$ with $c_1 \leftarrow e_j e_p e_l$ and $c_2 \leftarrow e_m e_n$ $(c_1, c_2 \in \{0, 1\})$ as the new IDRs.

Step 2: For each variable *c*, a constraint is added to capture the cascade propagation. Let *N* be the number of entities in the minterm on which *c* is dependent. In the example for the variable c_1 with IDR $c_1 \leftarrow e_j e_p e_l$, constraints $c_{1d} \ge \frac{x_{j(d-1)}+x_{p(d-1)}+x_{l(d-1)}}{N} \forall d \in [1, n-1]$ are introduced (N = 3 in this case). If IDR of an entity is already in form of a single minterm of arbitrary size, i.e., $e_i \leftarrow e_j e_p e_l$ (say) then constraints $x_{id} \ge \frac{x_{j(d-1)}+x_{p(d-1)}+x_{l(d-1)}}{N} - q_{x_i}$ and $x_{id} \le x_{j(d-1)} + x_{p(d-1)} + x_{l(d-1)} \forall d \in [1, n-1]$ are introduced (with N = 3). These constraints satisfies that if the entity e_i is hardened initially then it is not dead at any time step.

Step 3: Let *M* be the number of minterms in the transformed IDR as described in Step 1. In the given example with IDR $e_i \leftarrow c_1 + c_2 + e_q$ constraints of form $x_{id} \ge c_{1(d-1)} + c_{2(d-1)} + x_{q(d-1)} - (M-1) - q_{x_i}$ and $x_{id} \le \frac{c_{1(d-1)} + c_{2(d-1)} + x_{q(d-1)}}{M} \forall d \in [0, 1]$ are introduced. These constraints ensures that even if all the minterms of e_i has at least one entity in dead state then it will be alive if the entity is hardened initially.

With objective (1) along with the constraints minimize the number of entities failed at the end of the cascading failure with a hardening budget of k and K entities failing initially. The ILP gives an optimal solution to the ENH problem, however its run time is non-polynomial.

6.1.2. Heuristic Solution

In this subsection we provide a greedy heuristic solution to the Entity Hardening problem. For a given system $I(E, \mathcal{F}(E))$ with set of entities E'(|E'| = K) failing initially and hardening budget k, a heuristic is developed based on the following two metrics — (a) *Protection Set* as defined in Section 5, (b) *Cumulative Fractional Minterm Hit Value (CFMHV)* (Definition 14).

Definition 13. The Fractional Minterm Hit Value of an entity $e_j \in E$ in a system $I(E, \mathcal{F}(E))$ is denoted as $FMHV(e_j, X)$. It is calculated as $FMHV(e_j, X) = \sum_{i=1}^{m} \frac{1}{|s_i|}$. In the formulation m are the minterms in which e_j appears over all IDRs except for the IDRs of entities in set X. The parameter s_i denotes i^{th} such minterm. If entity e_j is hardened (or protected from failure) then the computed value provides an estimate of the future impact on protection of other non operational entities.

Definition 14. The Cumulative Fractional Minterm Hit Value of an entity $e_j \in E$ is denoted as $CFMHV(e_j)$. It is computed as $CFMHV(e_j) = \sum_{\forall x_i \in PS(e_j|E')} FMHV(x_i, PS(x_i|E'))$. This gives a measure of the future impact on protecting non functional entities when the entity e_j is hardened and entities $PS(e_j|E')$ are protected from failure.

Using these definitions a heuristic is formulated in Algorithm 3. For each iteration of the while loop in the algorithm, the entity having highest cardinality of the set $PS(x_i|A' \cup B') \cap P$ is hardened. This ensures that at each step the number of entities protected is maximized. In case of a tie, the entity having highest Cumulative Fractional Minterm Hit Value among the set of tied entities is selected. This causes the selection of an entity that has the potential to protect maximum number of entities in subsequent iterations. Thus, the heuristic greedily maximizes the number of entities protected when an entity is hardened at the current iteration with metric to measure its impact of protecting other non operational entities in future iterations. Algorithm 3 runs in polynomial time, more specifically the time complexity is $O(|P|k(n + m)^2)$ (where n = |E| and m = Number of minterms in $\mathcal{F}(E)$).

Algorithm 3: Heuristic Solution to the ENH Problem

Data: A system $I(E, \mathcal{F}(E))$, set of entities E' failing initially with |E'| = K and hardening budget k. **Result:** Set of hardened entities \mathcal{H} . 1 begin Initialize $\mathcal{H} \leftarrow \emptyset$ and $\mathcal{D} \leftarrow \emptyset$; 2 3 Update $\mathcal{F}(E)$ as follows — (a) let Q be the set of entities that does not fail on failing K entities, (b) remove IDRs corresponding to entities in set Q, (c) update the minterm of remaining IDRs by removing entities in set Q; Update $E \leftarrow E \setminus Q$; 4 while ($|\mathcal{H}|$ is not equal to k) do 5 For each entity $e_i \in E \setminus \mathcal{D}$ compute the Protection Sets $PS(e_i | E')$; 6 For each entity $e_i \in E \setminus \mathcal{D}$ compute $CFMHV(e_i)$; 7 if There exists multiple entities having same value of highest cardinality of the set $PS(e_i|E')$ then 8 Let e_p be an entity having highest $CFMHV(e_p)$ among all e_p 's in the set of entities having highest cardinality 9 of the set $PS(e_i|A' \cup B')$; If there is a tie choose arbitrarily; 10 Update $\mathcal{H} \leftarrow \mathcal{H} \cup \{e_p\}$; 11 Update $\mathcal{D} \leftarrow \mathcal{D} \cup PS(e_p|E');$ 12 Update $\mathcal{F}(E)$ by removing entities in $PS(e_p|E')$ both in the left and right side of the IDRs; 13 14 else Let e_i be an entity having highest cardinality of the set $PS(e_i|E')$; 15 Update $\mathcal{H} \leftarrow \mathcal{H} \cup \{e_i\}$; 16 Update $\mathcal{D} \leftarrow \mathcal{D} \cup PS(e_i|E');$ 17 Update $\mathcal{F}(E)$ by removing entities in $PS(e_i|E')$ both in the left and right side of the IDRs; 18 19 return \mathcal{H} ;

6.2. Solutions to the Targeted Entity Hardening Problem

6.2.1. Optimal solution using Integer Linear Program

The ILP formulation of the TEH problem is similar to that of ENH problem. The only difference being there is no hardening budget in TEH problem and additionally there is a set $P \subset E$ of entities that should be protected from failure. We use the same notations as of the ILP that solves the ENH problem. Using this the objective of the TEH problem is formulated as follows:

$$\min\left(\sum_{i=1}^{n} q_{x_i}\right) \tag{2}$$

The constraint sets 2,3, and 4 of the ENH problem is employed in the TEH problem as well along with an additional constraint set as described below:

Additional Constraint Set: For all entities $e_i \in P$, $x_{i(n-1)} = 0$. This ensures that all the entities in set P are protected from failure at the final time step.

With these constraints, the objective in (2) minimizes the number of hardened entities that results in protection of all entities in set P.

6.2.2. Heuristic Solution

In this subsection we provide a greedy heuristic solution to the TEH problem. For a given system $I(E, \mathcal{F}(E))$ with set of entities as E'(|E'| = K) failing initially and set of entities to protet being *P*, a heuristic is developed based on the following two metrics — (a) *Protection Set* as defined in Section 5, (b) *Prioritized Cumulative Fractional Minterm Hit Value (PCFMHV)* (Definition 16).

Definition 15. The Prioritized Fractional Minterm Hit Value of an entity $e_j \in E$ in an interdependent network $I(E, \mathcal{F}(E))$ is denoted as $FMHV(e_j, X)$. It is calculated as $PFMHV(e_j, P) = \sum_{i=1}^{m} \frac{1}{|s_i|}$. In the formulation *m* are the minterms in which e_j appears over IDRs in non operational entities in set P. The parameter s_i denotes i^{th} such minterm. If the e_j is hardened (or protected from failure) the value computed provides an estimate future impact on protection of other non operational entities in set P.

Definition 16. The Prioritized Cumulative Fractional Minterm Hit Value of an entity $e_j \in E$ is denoted as $PCFMHV(e_j)$. It is computed as $PCFMHV(e_j) = \sum_{\forall x_i \in PS(e_j|E')} PFMHV(x_i, PS(x_i|E'))$. This gives a measure of future impact on protecting non functional entities in P when the entity e_j is hardened and entities $PS(e_j|E')$ are protected from failure.

Algorithm 4: Heuristic solution to the TEH problem	
Data: A system $I(E, \mathcal{F}(E))$, set of K vulnerable entities and the set P of entities to be protected from	n failure.
Result: A set of entities <i>H</i> to be hardened.	
1 begin	
2 Initialize $\mathcal{D} = \emptyset$ and $H = \emptyset$;	
3 Update $\mathcal{F}(E)$ as follows — (a) let Q be the set of entities that does not fail on failing K entities,	(b) remove IDRs
corresponding to entities in set Q , (c) update the minterm of remaining IDRs by removing entit	ties in set Q ;
4 while $P \neq \emptyset$ do	
5 For each entity $e_i \in E \setminus \mathcal{D}$ compute the Protection Sets $PS(e_i E')$;	
6 For each entity $e_i \in E \setminus \mathcal{D}$ compute $PCFMHV(e_i)$;	
7 if There exists multiple entities having same value of highest cardinality of the set $PS(e_i E')$	$\cap P$ then
8 Let e_p be an entity having highest $CFMHV(e_p)$ among all e_p 's in the set of entities have	ing highest cardinality
of the set $PS(e_i A' \cup B')$;	
9 If there is a tie choose arbitrarily;	
10 Update $H \leftarrow H \cup \{e_p\}$;	
11 Update $\mathcal{D} \leftarrow \mathcal{D} \cup PS(e_p E');$	
12 Update $P \leftarrow P \setminus PS(e_p E');$	
13 Update $\mathcal{F}(E)$ by removing entities in $PS(e_p E')$ both in the left and right side of the IDI	Rs;
14 else	
15 Let e_i be an entity having highest cardinality of the set $PS(e_i E') \cap P$;	
16 Update $H \leftarrow H \cup \{e_p\}$;	
17 Update $\mathcal{D} \leftarrow \mathcal{D} \cup PS(e_i E');$	
18 Update $P \leftarrow P \setminus PS(e_i E');$	
19 Update $\mathcal{F}(E)$ by removing entities in $PS(e_i E')$ both in the left and right side of the IDF	ts ;
20 return H ;	

Using these definitions, the heuristic is formulated in Algorithm 4. For each iteration of the while loop in the algorithm, the entity having highest cardinality of the set $PS(x_i|A' \cup B') \cap P$ is hardened. This ensures that at each step the number of entities protected in set P is maximized. In case of a tie, the entity having highest Prioritized Cumulative Fractional Minterm Hit Value among the set of tied entities is selected. This causes the selection of the entity that has the potential to protect maximum number of entities in updated set P in subsequent iterations. Thus, the heuristic greedily minimizes the set of entities hardened which would cause protection of all entities in P. The

heuristic overestimates the cardinality of *H* from the optimal solution. Algorithm 4 runs in polynomial time, more specifically the time complexity is $O(|P|n(n + m)^2)$ (where n = |E| and m = Number of minterms in $\mathcal{F}(E)$).

It is to be noted if Algorithm 3 and 4 returns H (or \mathcal{H} as in ENH problem) with $|H| \ge K$ then we harden entities belonging to the set of K initially failing entities. This is because hardening these K initially failing entities would protect all entities in the interdependent network from failure.

7. Generating IDRs and Experimental Results

7.1. Generating Dependency Equations for Power Network

In this subsection, we describe a strategy to generate dependency equations of an intra-dependent power network. We restrict to load bus, generator bus, neutral bus and transmission line as the entities in the power network. For a given power network, AC power equations are solved to determine the direction of flow in the transmission lines. We use the power flow solver available in MatPower software for different bus systems [16]. For a given set of buses and transmission lines, the MatPower software uses load demand of the bus, impedance of the transmission lines etc. to solve the power flow and outputs the voltage of each bus in the system. We restrict to real power flow analysis. For a given solution, the real part of generation is taken as the power generated by a generator bus. Similarly, the real part of the load demand is taken as demand value of a load bus. For two buses e_1 and e_2 connected by a transmission line e_{12} the power flowing through the transmission line is calculated as $P_{12} = Real(V_1 * (\frac{V_1 - V - 2}{I_{12}})^*)$, where V_1 is the voltage at bus e_1 , V_2 is the voltage at bus e_2 , I_{12} is the impedance of the transmission line e_{12} and $(\frac{V_1 - V - 2}{I_{12}})^*$ denotes the complex conjugate of $(\frac{V_1 - V_2}{I_{12}})$. P_{12} is the real component of the power flowing in the transmission line e_{12} . Power flows from bus e_1 to e_2 if P_{12} is positive and from bus e_2 to e_1 otherwise.

The generation of the dependency equation is explained through a nine bus system shown in Figure 4. The figure represents a system $I(E, \mathcal{F}(E))$ with set E consisting of generator buses from G_1 through G_3 , load buses L_1 through L_4 , neutral buses $\{N_1, N_2\}$ and transmission lines T_1 through T_9 . The values in the red blocks denote the amount of power a generator is generating, the green block being the load requirements and blue neutral (value of 0). The value in the grey blocks correspond to power flow in the transmission lines. The transmission lines don't have any IDR. The IDRs for a bus b_1 is constructed by the following — (a) let b_2 , b_3 be the buses and b_{12} (between b_1 and b_2) and b_{13} between (b_1 and b_3) be the transmission lines for which power flows from these buses to b_1 , (b) the dependency equation for the bus b_1 is constructed as conjunction of minterms of size 2 (consisting of the bus from which power is flowing and the transmission line) with each conjunction corresponding to bus that has power flowing to it. For this example the dependency equation $b_1 \leftarrow b_{12}b_2 + b_{13}b_3$ is created. Using this definition the dependency equations for the buses in Figure 4 are created and is shown in Table 6.

The following points are to be noted regarding the generation rule — (a) The transmission lines can only fail initially due to a man made attack or natural disaster. Hence it entails the underlying assumption that the transmission lines would have enough capacity to transmit any power that is required to flow in it, (b) The generator bus is also only susceptible to initial failure and is assumed to have a generation capacity that is enough to supply the power demanded by a instance of power flow, (c) Neutral and Load buses are prone to both initial and induced failure. For example consider the failure of transmission lines T_9 and T_1 at t = 0. Owing to this the load bus L_1 and neutral bus N_2 fails at t = 2. At t = 3 load bus L_2 fails due to the failure of buses L_1, N_2 . It is to be noted that load bus L_3 does not fails as it still receives power from N_1 as transmission line T_4 is expected to have a capacity that can support a power flow equal to the demand of L_3 .

Owing to the underlying assumptions in the the creation of dependency equations, there is a limitation to its applicability to real world problems. However, with respect to power network, creating dependency equations like the one discussed is a preliminary step. Further research is required to be done to have a more accurate abstract representation of the dependency equations that can have widespread applicability to real world problems. The purpose of this subsection is — (1) presenting a preliminary way the dependency equations can be generated for power network, (2) larger data sets that can be used to measure the performance of the optimal solution to the heuristic.

7.2. Generating Dependency Equations for Interdependent Power-Communication Network

In this subsection, we describe rules to generate dependency relations for interdependent power and communication network infrastructure as used in [13]. Real world data of Maricopa County, Arizona , USA was taken.

Dependency Equations
$L_1 \leftarrow T_1 G_1$
$L_2 \leftarrow T_2 L_1 + T_7 N_2$
$L_3 \leftarrow T_3 L_1 + T_4 N_1$
$L_4 \leftarrow T_6 N_1 + T_8 N_2$
$N_1 \leftarrow T_5 G_3$
$N_2 \leftarrow T_9 G_2$

Table 6: IDRs of the buses in Figure 4



Figure 4: Example of Power Network Dependency

This county is one of the most densest populated region of Arizona with approximately 60% residents. Specifically, we wanted to measure the amount of resource required to protect entities in particular regions of the county when these regions have a set of entities failing initially. The data for power network was obtained from Platts (http://www.platts.com/) that contains 70 generator buses (including solar homes that generate minuscule unit of power) and 470 transmission lines. The communication network data was obtained from GeoTel (http://www.geotel.com/) consisting of 2, 690 cell towers, 7, 100 fiber-lit buildings and 42, 723 fiber links. Figures 5a and 5b displays the snapshot of power network and communication network for a particular region of Maricopa county. In Figure 5a the orange dots represent the generator buses and continuous yellow lines represent the transmission lines. In Figure 5b fiber-lit buildings are represented by pink dots, cell towers by orange dots and fiber links by continuous green lines.

The *load* of the power network are assumed to be cell towers and fiber-lit buildings. There exist other entities that draws electrical power. Since it is not relevant for the comparative analysis of the heuristic and the ILP such entities are ignored. The interdependent power-communication system is represented mathematically as $I(E, \mathcal{F}(E))$ with $E = A \cup B$. A and B consist of the entities in the power network and communication network respectively. With respect to this data the power network consist of three type of entities — generating stations, load (which are cell towers and fiber-lit buildings) and transmission lines (denoted by a_1, a_2, a_3 respectively). The communication network comprises of the following type of entities — cell towers, fiber-lit buildings and fiber links (denoted by b_1, b_2, b_3 respectively). It is to be noted that the fiber-lit buildings and cell towers are considered as both power network entities as well as communication network entities. From the raw data the dependency equations are constructed using the following rules.

Rules: We take into consideration that an entity in the power network is dependent on a set of entities in the communication network for either being operational and vice-versa. To keep things uncomplicated, we consider the dependency equations with at most two minterms. For the same reason we consider the size of each minterm is at most two.

Generators $(a_{1,i}, 1 \le i \le p)$, where p is the total number of generators): We assume that every generator $(a_{1,i})$ is, i) dependent on the closest Cell Tower $(b_{1,j})$, or, ii) closest Fiber-lit building $(b_{2,k})$ and the corresponding Fiber link $(b_{3,l})$ connecting $b_{2,k}$ and $a_{1,i}$. Hence, we have $a_{1,i} \leftarrow b_{1,j} + b_{2,k} \times b_{3,l}$.

Load $(a_{2,i}, 1 \le i \le q, where q is the total number of loads)$: The power network loads do not depend on any entities in communication network

Transmission Lines $(a_{3,i}, 1 \le i \le r, where r is the total number of transmission lines): The transmission lines in the power network do not depend on any entities in communication network.$

Cell Towers $(b_{1,i}, 1 \le i \le s, where s is the total number of cell towers)$: The Cell Towers depend on two components, i) the closest pair of generators, and, ii) corresponding transmission line, connecting the generator to the cell tower. Thus we have $b_{1,i} \leftarrow a_{1,j} \times a_{3,k} + a_{1,j'} \times a_{3,k'}$.

Fiber-lit Buildings $(b_{2,i}, 1 \le i \le t, where t is the total number of fiber-lit buildings): The Fiber-lit Buildings depend$ on two components, i) the closest pair of generators, and, ii) corresponding transmission line, connecting the generator $to the fiber-lit buildings. Thus we have <math>b_{2,i} \leftarrow a_{1,i} \times a_{3,k} + a_{1,i'} \times a_{3,k'}$.

Fiber Links $(b_{3,i}, 1 \le i \le u$, where u is the total number of fiber links): The Fiber Links aren't dependent on any power network entity. These links require power only for the amplifiers connected to them. The amplifiers are required if the length of the fiber link is above a certain threshold. We consider only those fiber links which are 'quite long', need

power. The fiber links depend on the closest pair of generators and the transmission lines connecting the generators to the fiber link under consideration. Thus we have $b_{3,i} \leftarrow a_{1,j} \times a_{3,k} + a_{1,j'} \times a_{3,k'}$. We do not consider that these fiber links need any power as we cannot determine the length of the fiber links or the exact threshold value due to the lack of data.



(a) Snapshot of Power Network

(b) Snapshot of Communication Network

Figure 5: Snapshots of the real world data corresponding to power and communication network

7.3. Comparative Study of the ILP and Heuristic for the Problems

A comparative study of the ILP and heuristic solution for both the problems is done in this subsection. A machine with intel i5 processor and 8 GB of RAM was used to execute the solutions. The coding was done in *java* and a student licensed *IBM CPLEX* external library file is used to execute the ILP. 8 different bus systems available from MatPower with number buses 24, 30, 39, 57, 89, 118, 145, 300 were used to generate the dependency equations for power network (using the rules described in Section 7.1). The time to generate the dependency equations were less than 2*ms*. Within the Maricopa county 4 disjoint regions were considered labeled as Region 1 through 4. Dependency equations for the interdependent power-communication network were generated for these regions using the rules described in Section 7.2. The java codes along with data files of the generated dependency equations are open sourced and is available in the following url *https://github.com/jbanerje1989/HardeningProblem*.

The number of entities in each of the 12 data sets are enumerated in Table 7. To determine the initially failing entities we used the ILP solution of \mathcal{K} most vulnerable entities in [13]. The \mathcal{K} most vulnerable entities problem finds a set of \mathcal{K} entities in a system $I(E, \mathcal{F}(E))$ which when failed at t = 0 causes the maximum number of entities to fail. For a given data set representing a system $I(E, \mathcal{F}(E))$, for both the problems the initially failing entities was taken as a set E' (|E'| = K) such that — (a) The set E' constitutes the K most vulnerable entities in the system, (b) Failing the entities in set E' would cause failure of at least |E|/2 entities in total. The cardinality of the set E' along with the total number of entities failed are enumerated in Table 7.

In comparing the ILP and heuristic solution of the ENH problem we considered 5 distinct hardening budgets for each data set. With K being the number of initially failing entities in a data set the hardening budgets were chosen between [1, K - 1] (with value of K obtained from Table 7). It is also ensured that the hardening budgets chosen had a high variance. Figures 6 - 17 shows the total number of entities protected from failure for each data set using the ILP and heuristic solution. The run-time performance of the solutions are provided in Table 8 (in the table 'Heu' refers to the heuristic solution and *Hi* refers to the hardening budget corresponding to the *i*th budget from left used in the bar graph plots). From Figures 6 - 17 it can be seen that the heuristic performs almost similar to that of the ILP solution in terms of quality. The maximum percent difference of the total number of entities protected in the ILP when compared to the heuristic solution occurs for a hardening budget of 39 in the 145 bus system (Figure 12) with the percent difference being 3.1%. In terms of run-time, heuristic outperforms the ILP with the heuristic computing solutions nearly 200 times faster in larger systems (as seen for the 300 bus system in Table 8). Hence it can be reasonably argued that the heuristic produces fast near optimal solutions for the ENH problem.

A similar kind of experimental analysis is performed for the TEH problem. 5 distinct protection sets P were considered for each data set. Let F denote the set entities failed in total when K entities fail initially. The cardinality of set F and the value of K was taken from Table 7 for each data set. The cardinality of the protection set for a given data set was chosen between [1, |F| - 1] ensuring that the chosen values have high variance. For a given cardinality C the protection set P was constructed by choosing C entities from the set F corresponding to a particular data set. Figures 18 - 29 shows the comparison of the Heuristic solution with the ILP in terms of total number of entities

hardened for a given cardinality of protection budget. The run-time comparison of the solutions are provided in Table 9. A maximum percent difference of 25% (ILP compared with Heuristic) in the number of entities hardened can be seen in Region 2 for a |P| value of 13 (Figure 27). However, for most of the cases the heuristic produces near optimal or optimal solution. The heuristic also compute the solutions nearly 200 times faster than the ILP for larger systems as seen in Table 9. Hence it can be claimed that the heuristic solution to the TEH problem produces near optimal solution at a much faster time compared to the ILP solution.

DataSet	Num. Of Entities	K	Num. of Entities Killed
24 bus	58	8	29
30 bus	71	13	36
39 bus	84	17	42
57 bus	135	26	68
89 bus	295	78	147
118 bus	297	89	149
145 bus	567	191	284
300 bus	709	145	355
Region 1	48	6	26
Region 2	46	8	23
Region 3	48	6	24
Region 4	53	8	27

 Table 7: Number of entities, K value chosen and number of entities failed when the K vulnerable entities are failed initially for different data sets



Figure 6: Comparison of ILP solution with Heuristic for 24 bus system (ENH)



Figure 9: Comparison of ILP solution with Heuristic for 57 bus system (ENH)



Figure 7: Comparison of ILP solution with Heuristic for 30 bus system (ENH)



Figure 10: Comparison of ILP solution with Heuristic for 89 bus system (ENH)



Figure 8: Comparison of ILP solution with Heuristic for 39 bus system (ENH)



Figure 11: Comparison of ILP solution with Heuristic for 118 bus system (ENH)



Figure 12: Comparison of ILP solution with Heuristic for 145 bus system (ENH)



Figure 15: Comparison of ILP solution with Heuristic for Region 2 (ENH)



Figure 13: Comparison of ILP solution with Heuristic for 300 bus system (ENH)



Figure 16: Comparison of ILP solution with Heuristic for Region 3 (ENH)



Figure 14: Comparison of ILP solution with Heuristic for Region 1 (ENH)



Figure 17: Comparison of ILP solution with Heuristic for Region 4 (ENH)

	Running time (in sec)									
DataSet	H1		H2		H3		H4		H5	
	ILP	Heu	ILP	Heu	ILP	Heu	ILP	Heu	ILP	Heu
24 bus	0.45	0.01	0.25	0.01	0.72	0.01	0.23	0.01	0.21	0.01
30 bus	2.44	0.01	0.38	0.01	0.38	0.01	0.35	0.01	0.34	0.01
39 bus	0.80	0.01	0.50	0.01	0.49	0.01	0.48	0.01	0.47	0.01
57 bus	2.67	0.03	1.73	0.01	2.21	0.01	2.27	0.01	1.68	0.01
89 bus	23.2	0.05	14.6	0.03	14.6	0.03	14.5	0.03	14.7	0.75
118 bus	20.9	0.04	16.2	0.06	17.2	0.09	17.1	0.02	17.1	0.02
145 bus	85.2	0.05	71.0	0.10	71.3	0.18	68.4	0.06	78.3	0.07
300 bus	282	0.15	222	1.56	217	0.85	253	0.39	264	0.40
Region 1	0.53	0.01	0.36	0.01	0.34	0.01	0.36	0.01	0.36	0.01
Region 2	13.8	0.01	12.9	0.01	12.8	0.01	13.1	0.01	13.2	0.01
Region 3	1.92	0.01	1.36	0.01	1.29	0.01	1.31	0.01	1.44	0.01
Region 4	1.48	0.01	1.43	0.01	1.10	0.01	1.06	0.01	1.05	0.01

Table 8: Run time comparison of Integer Linear Program and Heuristic for different Data Sets (ENH)



Figure 18: Comparison of ILP solution with Heuristic for 24 bus system (TEH)



Figure 19: Comparison of ILP solution with Heuristic for 30 bus system (TEH)







Figure 21: Comparison of ILP solution with Heuristic for 57 bus system (TEH)



Figure 24: Comparison of ILP solution with Heuristic for 145 bus system (TEH)



Figure 27: Comparison of ILP solution with Heuristic for Region 2 (TEH)



Figure 22: Comparison of ILP solution with Heuristic for 89 bus system (TEH)



Figure 25: Comparison of ILP solution with Heuristic for 300 bus system (TEH)



Figure 28: Comparison of ILP solution with Heuristic for Region 3 (TEH)



Figure 23: Comparison of ILP solution with Heuristic for 118 bus system (TEH)



Figure 26: Comparison of ILP solution with Heuristic for Region 1 (TEH)



Figure 29: Comparison of ILP solution with Heuristic for Region 4 (TEH)

	Running time (in sec)									
DataSet	P1		P2		P3		P4		P5	
	ILP	Heu	ILP	Heu	ILP	Heu	ILP	Heu	ILP	Heu
24 bus	0.42	0.01	0.22	0.01	0.20	0.01	0.19	0.01	0.19	0.01
30 bus	0.57	0.01	0.37	0.01	0.34	0.01	0.34	0.01	0.31	0.01
39 bus	0.82	0.01	0.49	0.01	0.49	0.01	0.46	0.01	0.47	0.01
57 bus	2.23	0.02	1.69	0.02	2.00	0.02	2.07	0.02	1.91	0.01
89 bus	17.3	0.05	14.4	0.08	14.4	0.16	14.0	0.11	14.1	0.07
118 bus	17.6	0.13	17.3	0.15	16.9	0.10	16.3	0.08	17.0	0.07
145 bus	79.0	0.06	76.6	0.26	77.2	0.27	75.7	0.24	74.9	0.25
300 bus	302	0.18	241	1.01	234	1.41	229	1.03	230	1.20
Region 1	0.55	0.01	0.40	0.01	0.43	0.01	0.35	0.01	0.34	0.01
Region 2	14.5	0.01	13.5	0.01	13.4	0.01	13.4	0.01	13.3	0.01
Region 3	1.58	0.01	1.40	0.01	1.29	0.01	1.29	0.01	1.29	0.01
Region 4	1.36	0.01	1.12	0.01	1.21	0.01	1.09	0.01	1.03	0.01

Table 9: Run time comparison of Integer Linear Program and Heuristic for different Data Sets (TEH)

8. Conclusion

In this paper, we have studied the Entity Hardening problem and the Targeted Entity Hardening problem in Critical Infrastructure network(s). We have used the IIM model to capture the interdependencies and dependencies that exist between power-communication network and power network in isolation respectively. Using such a model, we have formulated the Entity Hardening and the Targeted Entity Hardening problems. Both problems are proved to be NP-Complete. For both the problems, the optimal solution, obtained from the ILP, is compared with the developed heuristic solution using dependency equations generated from power-communication network data of the Maricopa County, Arizona and power network data of different bus systems obtained from MatPower. As noted in the Experimental Analysis for both the problems the performance of the heuristics are comparable to that of ILP and solutions are produced in much lesser time.

References

- [1] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, et al., Causes of the 2003 major grid blackouts in north america and europe, and recommended means to improve system dynamic performance, Power Systems, IEEE Transactions on 20 (4) (2005) 1922–1928.
- [2] F. E. R. Commission, the North American Electric Reliability Corporation, Arizona Southern California Outages on September 8, 2011, Causes and Recommendations, Tech. rep. (04 2012).
- [3] Y. Tang, G. Bu, J. Yi, Analysis and lessons of the blackout in indian power grid on july 30 and 31, 2012, in: Zhongguo Dianji Gongcheng Xuebao(Proceedings of the Chinese Society of Electrical Engineering), Vol. 32, Chinese Society for Electrical Engineering, 2012, pp. 167–174.
- [4] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature 464 (7291) (2010) 1025–1028.
- [5] J. Gao, S. V. Buldyrev, H. E. Stanley, S. Havlin, Networks formed from interdependent networks, Nature Physics 8 (1) (2011) 40–48.
- [6] J. Shao, S. V. Buldyrev, S. Havlin, H. E. Stanley, Cascade of failures in coupled network systems with multiple support-dependence relations, Physical Review E 83 (3) (2011) 036116.
- [7] V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. Porcellinis, R. Setola, Modelling interdependent infrastructures using interacting dynamical models, International Journal of Critical Infrastructures 4 (1) (2008) 63–79.
- [8] P. Zhang, S. Peeta, T. Friesz, Dynamic game theoretic model of multi-layer infrastructure networks, Networks and Spatial Economics 5 (2) (2005) 147–178.
- [9] M. Parandehgheibi, E. Modiano, Robustness of interdependent networks: The case of communication networks and the power grid, arXiv preprint arXiv:1304.0356.
- [10] D. T. Nguyen, Y. Shen, M. T. Thai, Detecting critical nodes in interdependent power networks for vulnerability assessment.
- [11] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, G. Zussman, Power grid vulnerability to geographically correlated failures-analysis and control implications, arXiv preprint arXiv:1206.1099.
- [12] J. Banerjee, A. Das, A. Sen, A survey of interdependency models for critical infrastructure networks, NATO Science for Peace and Security Series -D: Information and Communication Security 37 (2014) 1–16.
- [13] A. Sen, A. Mazumder, J. Banerjee, A. Das, R. Compton, Identification of k most vulnerable nodes in multilayered network using a new model of interdependency, in: Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on, IEEE, 2014, pp. 831–836.
- [14] A. Das, J. Banerjee, A. Sen, Root cause analysis of failures in interdependent power-communication networks, in: Military Communications Conference (MILCOM), 2014 IEEE, IEEE, 2014, pp. 910–915.
- [15] A. Das, C. Zhou, J. Banerjee, A. Sen, L. Greenwald, On the smallest pseudo target set identification problem for targeted attack on interdependent power-communication networks, in: Military Communications Conference, MILCOM 2015-2015 IEEE, IEEE, 2015, pp. 1015–1020.
- [16] R. D. Zimmerman, C. E. Murillo-Sánchez, R. J. Thomas, Matpower: Steady-state operations, planning, and analysis tools for power systems research and education, IEEE Transactions on power systems 26 (1) (2011) 12–19.
- [17] W. J. Clinton, Executive order eo 13010 critical infrastructure protection, Tech. rep. (July 1996).

- [18] S. M. Rinaldi, Modeling and simulating critical infrastructures and their interdependencies, in: System sciences, 2004. Proceedings of the 37th annual Hawaii international conference on, IEEE, 2004, pp. 8–pp.
- [19] S. M. Rinaldi, J. P. Peerenboom, T. K. Kelly, Identifying, understanding, and analyzing critical infrastructure interdependencies, Control Systems, IEEE 21 (6) (2001) 11–25.
- [20] P. Pederson, D. Dudenhoeffer, S. Hartley, M. Permann, Critical infrastructure interdependency modeling: a survey of us and international research, Idaho National Laboratory (2006) 1–20.
- [21] H. Lin, S. Sambamoorthy, S. Shukla, J. Thorp, L. Mili, Power system and communication network co-simulation for smart grid applications, in: Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES, IEEE, 2011, pp. 1–6.
- [22] H. Lin, S. S. Veda, S. S. Shukla, L. Mili, J. Thorp, Geco: Global event-driven co-simulation framework for interconnected power system and communication network, IEEE Transactions on Smart Grid 3 (3) (2012) 1444– 1456.
- [23] S. Amin, G. A. Schwartz, S. Shankar Sastry, Security of interdependent and identical networked control systems, Automatica 49 (1) (2013) 186–192.
- [24] S. C. Müller, H. Georg, J. J. Nutaro, E. Widl, Y. Deng, P. Palensky, M. U. Awais, M. Chenine, M. Kuch, M. Stifter, et al., Interfacing power system and ict simulators: Challenges, state-of-the-art, and case studies, IEEE Transactions on Smart Grid.
- [25] C. Barrett, K. Channakeshava, F. Huang, J. Kim, A. Marathe, M. V. Marathe, G. Pei, S. Saha, B. S. Subbiah, A. K. S. Vullikanti, Human initiated cascading failures in societal infrastructures, PloS one 7 (10) (2012) e45406.
- [26] R. Pfitzner, K. Turitsyn, M. Chertkov, Controlled tripping of overheated lines mitigates power outages, arXiv preprint arXiv:1104.4558.
- [27] S. Ruj, A. Pal, Analyzing cascading failures in smart grids under random and targeted attacks, in: Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on, IEEE, 2014, pp. 226–233.
- [28] B. Falahati, Y. Fu, L. Wu, Reliability assessment of smart grid considering direct cyber-power interdependencies, Smart Grid, IEEE Transactions on 3 (3) (2012) 1515–1524.
- [29] S. Soltan, M. Yannakakis, G. Zussman, Joint cyber and physical attacks on power grids: Graph theoretical approaches for information recovery, in: ACM SIGMETRICS Performance Evaluation Review, Vol. 43, ACM, 2015, pp. 361–374.
- [30] M. Hajiaghayi, K. Jain, K. Konwar, L. Lau, I. Mandoiu, A. Russell, A. Shvartsman, V. Vazirani, The minimum k-colored subgraph problem in haplotyping and dna primer selection, in: Proceedings of the International Workshop on Bioinformatics Research and Applications (IWBRA), Citeseer, 2006.