The Computational Complexity of Probabilistic Interdependent Network Design Problems

Andrés D. González  
Graduate Student, Dept. of Civil & Environmental Engineering, Rice Univ., Houston, USA  
Dept. of Civil & Environmental Engineering, Univ. de los Andes, Bogotá, Colombia

Leonardo Dueñas-Osorio  
Associate Professor, Dept. of Civil & Environmental Engineering, Rice Univ., Houston, USA

Mauricio Sánchez-Silva  
Professor, Dept. of Civil & Environmental Engineering, Univ. de los Andes, Bogotá, Colombia

Andrés L. Medaglia  
Professor, Dept. of Industrial Engineering, Univ. de los Andes, Bogotá, Colombia

ABSTRACT: We present a rigorous study of the computational complexity of an Interdependent Network Design Problem (INDP) solution model, which is at the core of future infrastructure restoration studies. The paper details how each constraint in the INDP formulation adds to the overall complexity, while also revealing strategies to tame the computational demands for large interdependent networks. It is shown how some algorithms that are used to enhance the Network Design Problem (NDP) (Poss 2011), including decomposition techniques, can be cleverly adapted to the INDP model. Computational examples are based on diverse yet idealized topologies, to illustrate the sensitivity of the INDP model to input parameters and constraints, as well as the underlying topological properties of the studied systems. The present study paves the way for future approaches to improve the INDP and its Mixed-Integer Programming (MIP) model, particularly in terms of its efficiency and capability to handle large size instances, and by extension, its ability to support practical infrastructure decision making and resilience analyses.

1. INTRODUCTION

Recent approaches to study infrastructure networks have highlighted the importance of interdependency among them, especially when studying their resilience. Buldyrev et al. (2010) and Dueñas-Osorio et al. (2007) showed theoretically and computationally that a system of interdependent networks has increased vulnerability and is susceptible to cascading failures. Ouyang & Dueñas-Osorio (2011) presented an approach to design and retrofit infrastructure systems to reduce the likelihood of cascading failures, by properly accounting for interdependencies. From a recovery perspective, González et al. (2014) defined the Interdependent Network Design Problem (INDP) as the problem of finding the least-cost recovery strategy for a partially destroyed system of interdependent infrastructure networks. To solve the INDP, González et al. (2014) also proposed a Mixed Integer Programming (MIP) model, which seeks the minimum-cost recovery strategy while considering operational constraints, limited availability of resources, and interdependencies between the networks. Such model showed high applicability in studying realistic scenarios, such as for optimizing the recovery strategies of the gas, power, and water networks in Shelby County, TN, after an earthquake. This paper studies the computational complexity of the INDP in a probabilistic context, related to the likelihood of
diverse disaster scenarios. The paper starts by presenting a re-formulation of the INDP-MIP model based on the former work by González et al. (2014), specifying now its structure and how it can be used in a probabilistic framework. Then, it assesses the complexity class of the INDP, and analogous problems related to finding optimal flows through a system of networks. Afterwards, the paper presents an approach to analyze the INDP-MIP model complexity from a numerical approach, studying the dependence between its running time and parameters such as the size and density of the system of networks, its level of interdependency, and the availability of resources. Computational experiments are shown to also study the INDP-MIP model complexity under different topologies. Then, conclusions and ideas for future work follow.

2. THE INTERDEPENDENT NETWORK DESIGN PROBLEM (INDP)-MIP MODEL

To solve the INDP, González et al. (2014) proposed an MIP model that accounts for the costs associated to component reconstruction, flow of commodities, unsatisfied demands, and geographical preparation of component repair sites. This INDP-MIP model considers constraints related to flow balance, limited flow capacities, limited availability of resources for the reconstruction process, and interdependencies. The latter includes physical interdependency between networks (implying that certain components should be functional in a system in order for others to be functional in different systems as well), and geographical interdependency (allowing for costs reductions by simultaneously repairing co-located components).

We propose the following re-formulation of the INDP-MIP model, in order to make the structure of the optimization problem more transparent for computational complexity analyses, by using generalized sets of network elements (to comprise arcs and nodes). The use of sets of elements reduces by half the original sets of constraints, and facilitates associating them with properties of the model relevant to its time-complexity (to be analyzed in section 3.1). Without loss of generality, we assume that only one commodity can flow through each network. To formulate this more transparent problem, let us define the following sets, parameters, and variables as:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{K}$</td>
<td>Set of infrastructure networks</td>
</tr>
<tr>
<td>$R$</td>
<td>Set of limited resources to be used in the reconstruction process</td>
</tr>
<tr>
<td>$S$</td>
<td>Set of geographical spaces (spatial distribution of the area that contains the infrastructure networks)</td>
</tr>
<tr>
<td>$N_k$</td>
<td>Set of nodes before a destructive event</td>
</tr>
<tr>
<td>$N_k'$</td>
<td>Set of nodes destroyed in network $k \in \mathcal{K}$</td>
</tr>
<tr>
<td>$A_k$</td>
<td>Set of arcs before a destructive event</td>
</tr>
<tr>
<td>$A_k'$</td>
<td>Set of arcs destroyed in network $k \in \mathcal{K}$</td>
</tr>
<tr>
<td>$E_k$</td>
<td>Set of elements in network $k \in \mathcal{K}$</td>
</tr>
<tr>
<td>$E_k'$</td>
<td>Set of destroyed elements in network $k \in \mathcal{K}$</td>
</tr>
<tr>
<td>$f_e$</td>
<td>Reconstruction cost of element $e \in E$</td>
</tr>
<tr>
<td>$c_e$</td>
<td>Flow unitary cost through arc $e \in A$</td>
</tr>
<tr>
<td>$M^+_e$</td>
<td>Cost of oversupply in node $e \in N$</td>
</tr>
<tr>
<td>$M^-_e$</td>
<td>Cost of unsatisfied demand in node $e \in N$</td>
</tr>
<tr>
<td>$g_s$</td>
<td>Cost of preparing geographical space $s \in S$</td>
</tr>
<tr>
<td>$h_r$</td>
<td>Total availability of resource $r \in R$</td>
</tr>
<tr>
<td>$u_e$</td>
<td>Capacity of arc $e \in A$</td>
</tr>
<tr>
<td>$v_r$</td>
<td>Total availability of resource $r \in R$</td>
</tr>
<tr>
<td>$\beta(es)$</td>
<td>Zero-one parameter indicating whether or not space $s \in S$ has to be prepared when reconstructing element $e \in E$</td>
</tr>
<tr>
<td>$\gamma(e\bar{e})$</td>
<td>Zero-one parameter indicating whether or not element $e \in E$ depends physically on element $\bar{e} \in E$, where $e \neq \bar{e}$</td>
</tr>
<tr>
<td>$x_e$</td>
<td>Flow through arc $e \in A$</td>
</tr>
<tr>
<td>$\delta^+_e$</td>
<td>Oversupply in node $e \in N$</td>
</tr>
<tr>
<td>$\delta^-_e$</td>
<td>Unsatisfied demand in node $e \in N$</td>
</tr>
<tr>
<td>$\gamma_e$</td>
<td>Zero-one variable indicating whether or not element $e \in E$ becomes functional</td>
</tr>
<tr>
<td>$z_s$</td>
<td>Zero-one variable indicating whether or not geographical space $s \in S$ is prepared.</td>
</tr>
</tbody>
</table>
Based on the previous definitions, the proposed MIP re-formulation for the INDP is:

\[
\begin{align*}
\text{minimize} & \quad \sum_{k \in \mathcal{K}} \left( \sum_{e \in \mathcal{E}'_k} f_e y_e + \sum_{e \in \mathcal{A}_k} c_e x_e \right) \\
& \quad + \sum_{e \in \mathcal{N}_k} (M^+_e \delta^+_e + M^-_e \delta^-_e) \\
& \quad + \sum_{s \in \mathcal{S}} g_s z_s
\end{align*}
\]  

(1)

subject to,

\[
\sum_{j: (i,j) \in \mathcal{A}_k} x_{(i,j)} - \sum_{j: (j,i) \in \mathcal{A}_k} x_{(j,i)} + \delta^+_k - \delta^-_k = b_i, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}_k, \\
-x_{(i,j)} + u_{(i,j)} y_e \geq 0, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}_k, \forall e \in \{i,j,(i,j)\}, \\
\sum_{k \in \mathcal{K}} \sum_{e \in \mathcal{E}'_k} h_{er} y_e \leq v_r, \quad \forall r \in \mathcal{R}, \\
\sum_{e \in \mathcal{E}_k} y_{e \bar{e}} y_e - y_{\bar{e}} \geq 0, \\
\forall k, \bar{k} \in \mathcal{K}, \forall \bar{e} \in E_{\bar{k}} | \sum_{e \in E_{\bar{k}}} y_{e \bar{e}} > 0,
\end{align*}

(2)

(3)

(4)

(5)

\[
\begin{align*}
z_s - \beta_{es} y_e & \geq 0, \quad \forall k \in \mathcal{K}, \forall e \in \mathcal{E}'_k, \forall s \\
& \quad \in \mathcal{S} | \beta_{es} > 0, \\
\delta^+_e & \geq 0, \quad \forall k \in \mathcal{K}, \forall e \in \mathcal{N}_k, \\
\delta^-_e & \geq 0, \quad \forall k \in \mathcal{K}, \forall e \in \mathcal{N}_k, \\
x_e & \geq 0, \quad \forall k \in \mathcal{K}, \forall e \in \mathcal{A}_k, \\
y_e & \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall e \in \mathcal{E}_k, \\
z_s & \in \{0,1\}, \quad \forall s \in \mathcal{S}.
\end{align*}
\]

(6)

(7)

(8)

(9)

(10)

(11)

Note that each \( e \in \mathcal{A}_k \) is uniquely defined by a pair of nodes \((i,j)\), where \( i \in \mathcal{N}_k \) represents the starting node and \( j \in \mathcal{N}_k \) the ending node. The objective function (1) minimizes the recovery costs associated to reconstructing damaged components (first term), the total flow costs (second term), the unbalance costs related to over or under supply (third term), and the total cost of preparing the different areas for the reconstruction process (fourth term). The set of constraints (2) guarantee the commodity flow balance at each node, for each infrastructure network. Constraints (3) ensure that flows through an arc are limited by a given capacity, and can exist only if the arc is functional, along with its starting and ending nodes. Constraints (4) limit the use of each resource. Constraints (5) model the physical interdependence between networks, by guaranteeing that for a given element to be functional, at least one from a set of required elements should be functional as well. If the support components (set of required components) are more than two, the element is called multiply supported. On the other hand, if there is only one support component, the element is called simply supported. Constraints (6) ensure that when any element is reconstructed, the area that contains it has to be prepared. Note that constraints (5) and (6) take place only if a component is interdependent to others, i.e., if \( \sum_{e \in \mathcal{E}_k} y_{e \bar{e}} > 0 \) and \( \beta_{es} > 0 \) respectively. Finally, constraints (7)-(11) are related to the nature of the variables.

Note that the INDP-MIP formulation is fully deterministic, as all the sets and parameters are assumed to be known. Nevertheless, the INDP-MIP model could be used in a probabilistic framework, if we consider the uncertainty related to the functionality of the system components and the topology of the networks. Given a particular hazard, each component is assumed to have a specific probability of destruction. Such a probability will depend on the specific system and hazard under study, and should consider the fragility of the components and the magnitude of the studied damaging events. To use the INDP-MIP formulation in a probabilistic scenario, one could use methods such as simulation, stochastic and robust optimization, among others. For this study, we used Monte-Carlo simulation, based on generating random disaster scenarios according to a given destruction probability distribution. For each disaster scenario, the INDP-MIP model would indicate the optimal recovery strategy, and will have different running times.
The following section will focus on studying the running times of the INDP-MIP model, and its dependence on parameters such as the size and density of the system and the availability of resources for the reconstruction. The effect of network topology is also studied afterwards.

3. COMPUTATIONAL COMPLEXITY OF THE INDP

In general, MIP models based on linear constraints belong to the NP-hard computational complexity class. Nevertheless, an in-depth analysis should be performed on the proposed INDP-MIP formulation, as it can be considered an extension or generalization of well known problems in literature, such as the Minimum Cost Flow Problem (MCFP) (Vaidyanathan 2010; Hamacher et al. 2007; Tomlin 1966), and the classical Network Design Problem (NDP) (Poss 2011; Dionne & Florian 1979).

First, consider a context where there is no set of limiting resources, and the geographical preparation costs are negligible. Then, under this context variable \( z_s \) would not be relevant, and constraints (4), (6), and (11) would not exist.

Now, assume the simple case where there is only a single network. If such a network is fully functional (i.e., there are no destroyed components), then \( y_e = 1, \forall e \in E \), which in turn implies that \( z_s = 0, \forall s \in S \). Then, only constraints (2) and (3) would take place and the problem would not be MIP anymore, as all the integer variables are known beforehand. Then, under this configuration, the INDP-MIP model would be identical to the MCFP, which is solvable in polynomial time (Orlin 1997). On the other hand, if we assume a single partially destroyed network, only constraints (2) and (3) would take place again. Nevertheless, \( y_e \) would be unknown. Then, under this configuration, the INDP-MIP model would be identical to the classical NDP, and would be NP-complete (Johnson et al. 1978).

By studying multiple networks simultaneously, the INDP-MIP model would be solving simultaneous NDPs for each network. Nevertheless, the more interdependent the networks are, the more constraints (5) will take place. This coupling constraints would increase the complexity of the overall model, as they depend on the binary design variables \( y_e \). Likewise, including a set of limiting resources would also increase the complexity, as constraints (4) also depend on \( y_e \). Finally, including geographical preparation costs into the analysis would also increase the complexity of the MIP model, as it depends on the additional binary variables \( z_s \). Then, from the previous line of reasoning, we can conclude that the full INDP-MIP model, which considers interdependence between networks and also geographical preparation costs related to binary variables \( z_s \), would be in general NP-hard. Nevertheless, the average computational complexity of the INDP-MIP model could be improved by applying enhancement techniques based on the previously mentioned structure and properties. For example, Benders decomposition (Benders 1962; Poss 2011), could solve for the binary design variables separately from the flow variables. Also, Dantzig-Wolfe decomposition (Dantzig et al. 1979) could be used to solve the coupling constraints (in this case the ones related to interdependence) separately from the uncoupled constraints. Then, even though the INDP-MIP model is NP-hard in its full form, it could be highly enhanced to be computationally tractable for practical scenarios.

3.1. Numerical study on the complexity of the INDP-MIP model

The structure and computational complexity of the INDP-MIP model are highly dependent on certain parameters, such as the probability of failure, the size of the system, the level of interdependency between networks, and the availability of resources.

In particular, we study the behavior of the INDP-MIP model running time, as a function of the following analysis parameters:

- Failure probabilities: The more components are initially destroyed, the larger the number of constraints, since constraints (5) and (6) are defined for each destroyed component. Also,
the more destroyed components, the more terms related to binary variables in the objective function and in constraints (6).

- Nodes in the system of networks: The number of nodes studied determine the number of variables and equations considered, independently of the type of constraints.
- Links density: This term is defined as the ratio between the number of links in the system, and the maximum number of links for the studied network topology. The more links, the larger the number of flow variables.
- Interdependency density: This term is defined as the ratio between the components that depend on other components, and the total number of components in the system. The more dependent components, the larger the amount of constraints (5).
- Interdependency strength: This term is related to constraints (5), and refers to the average number of support components per dependent element. The more support components there are, the easier to enable the functionality of dependent components.
- Resource availability: Lower availability of resources imply less components that can be recovered. Then, it is expected that the less resources the more difficult to find the optimal recovery strategy.

The study that follows assumes that there is a given system of interdependent networks, with fully defined sets and parameters according to the INDP-MIP formulation. Then, the exploration consists of varying each of the mentioned analysis parameters, such that it is possible to measure their impact on computational complexity and associated running times.

4. COMPUTATIONAL EXPERIMENTS

In order to obtain generalizable conclusions about the INDP computational complexity, this study does not focus on a single particular case, but studies different idealized topologies, some of which constitute the building blocks of realistic systems. In particular, this analysis focuses on the study of a multilayered system of two interdependent networks, with similar sizes and topologies. We selected three different topologies:

- Grid: Frequently found substrate in city-level transportation and utilities distribution networks.
- Wheel: Topology that is studied in communication networks (Mackenzie 1966; Callander & Plott 2005), where there is one driving node connected to the others. A wheel network with n nodes is composed by a single node connected to all nodes of a cycle of size n-1, resembling distribution within service areas.
- Erdős–Rényi: Related to graphs where all the links are randomly distributed. In this particular case, we generated these networks to have the same number of links of the grid, without ensuring connectivity.

Without loss of generality, the INDP input parameters where generated according to the following uniform distributions, considering that this study is intended to be general, where no prior information about the probabilistic behavior of the parameters: $b_e \sim U(-10,10)$, $u_e \sim U(0,10)$, $c_e \sim U(1,10)$, $f_e \sim U(1,100)$, $g_s \sim U(1,200)$, $M_e^{+} = 1000$, $h_e = 1$. Such distributions describe a system with large variability, without bias or asymmetry. The upper and lower bounds are chosen such that in average the demands and capacities have similar magnitudes. Also, such that the cost of preparing a reconstruction jobs for multiple co-located components is in average greater that the cost of repairing a single component. Note that the costs are never zero (or less), implying that all flows and repairing processes would increase the objective function. This analysis was based on a multilayered system (Boccaletti et al. 2014), where each component from one network has exactly one correspondent component in the other network. Thus, it was assumed that repairing a component implied repairing a unique space $s \in S$ specific to each component. For all topologies, the system
analyzed had the following default analysis parameters: Failure probabilities = 0.5, order of the system = 32 nodes, links density = 0.5, interdependency density = 0.5, interdependency strength = 0.5, and resource availability = 6 units. We used 1000 Monte-Carlo simulations per configuration, such that the mean running time of a system with default parameters had an interval of ±1.3 milliseconds, with 95% confidence.

To give a sense of the proportion of constraints and variables of each type in an average system, Figure 1 shows the matrix representations of the INDP constraints for a sample of each of the studied topologies. Observe that these matrices have almost identical sizes and structure, implying that the INDP-MIP model does not depend strongly on the topology of the studied networks, as long as the size and order of the networks are similar. For the Erdős–Rényi topology, there are 467 constraints and 412 variables in total. The set of constraints (2) has 32 equations (rows 1-32), constraints (3) has 360 inequalities (rows 33-392), constraint (4) has one inequality (row 393), constraints (5) has 11 inequalities (rows 394 to 404), and constraints (6) has 63 inequalities (rows 405 to 467). Regarding the variables, for network 1 there are 60 flow variables (columns 1-60), 76 binary design variables (columns 61-136), and 32 variables for unmet demand (columns 137-168). Similarly, for network 2 there are 60 flow variables (columns 169-228), 76 binary design variables (columns 229-304), and 32 variables for unmet demand (columns 305-336). Finally, there are 76 spaces to be prepared (columns 337-412).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Constraints</th>
<th>Grid</th>
<th>Wheel</th>
<th>Erdős-Rényi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 200 400</td>
<td>0 200 400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>St.Dev.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1. INDP coefficient matrices for the studied topologies, with standard configuration](image)

Figure 2 shows the running times for the studied topologies as a function of the failure probabilities. As expected, the mean running times increased with the probability of failure. Nevertheless, such increment appears to be linear. Similarly, the running times variability is also higher for larger failure probabilities. This was expected, as the probability of failure is directly linked with the number of destroyed components.

Similarly, Figure 3 shows the mean running times for diverse system sizes. As predicted, the times (and their variability) increase linearly with size. Figure 4 shows that, contrary to what one could expect, the link density of the undestroyed system appears to have no major impact on the running times. This is explained by the fact that the constraints that include binary variables (which are the ones that drive most of the complexity) are related not to the number of links, but to the number of destroyed links and nodes.

![Figure 2: Mean running times and standard deviations as a function of failure probability](image)

![Figure 3: Mean running times and standard deviations as a function of the number of nodes in the system of networks](image)

![Figure 4: Mean running times and standard deviations as a function of link density (with respect to the original topology)](image)
Figure 5 shows that the larger the interdependency density, the larger the running time. This was also expected, as the interdependency density determined the number of constraints related to the physical interdependence. Similarly, Figure 6 shows that the running times tend to increase with the interdependency strength. Nevertheless, observe that the change virtually does not exist for the grid topology. Such a behavior is related to the resilience of the system, since the grid topology is in general highly redundant, whereas the other topologies may get disconnected easier due to failing components.

Figure 5: Mean running times and standard deviations as a function of interdependency density

Figure 6: Mean running times as a function of interdependency strength

Figure 7: Mean running times and standard deviations as a function of the resource availability

Finally, Figure 7 shows that the running times highly depend on the resource availability. Nevertheless, as opposed to the other scenarios, the relation between the running times and the analysis parameter is not monotonic. In this case, the complexity increases vastly as the resources increase, but only for resource levels up to a value of approximately six. From that value, the running times slightly decrease with the resource availability. This is a most interesting behavior, related to the ratio between the available resources $v_r$ and the resource consumption $h_{er}$. Given that $h_{er}$ was assumed to be one in this experiment, $v_r$ could be traduced as the number of components that could be fixed. Note that to fix one link that is dependent on another, you need to ensure functionality of at least the parent link with its starting and ending nodes, and the dependent link with its starting and ending nodes. This corresponds to six components, which is precisely the found threshold.

In general, it was possible to observe that the time-complexity of the INDP-MIP model increased with the size of the networks, the density and strength of interdependencies, the resource availability (when resources are too limited), and the probability of failure. Nevertheless, it was possible to observe that the time-complexity was independent from the studied topology, except when studying the interdependency strength, where a less reliable topology implied a larger negative impact on the running times.

5. CONCLUSIONS
This paper proposed an MIP re-formulation for the Interdependent Network Design Problem (INDP) (González et al. 2014), which seeks the optimal recovery strategy for a partially destroyed system of physically and geographically interdependent networks. Even though the INDP is NP-hard, the paper shows that the INDP can be seen as a generalized version of the Minimum Cost Flow Problem (MCFP), which itself can be solved in polynomial time. However, INDP is also a generalization of the classical Network Design Problem (NDP), which is NP-complete. Thus, algorithmic enhancement techniques such as Dantzig-Wolfe and Benders decompositions are applicable, although an understanding of the
problem objective function and the structure of constraints is of paramount importance first.

Using numerical approaches, this study explores the structure of the INPD-MIP model, and shows that its time efficiency is sensitive to the size of the network, and the probabilities of failure of individual components. High levels of interdependency between networks also showed to have a major impact, but mostly for systems that have little connectivity redundancy. Similarly, using low resource availability showed to highly reduce the time complexity of the INDP-MIP model.

From studying different topologies, it was possible to observe that the reliability of the studied system does not impact the behavior of the INPD-MIP model running times, except when studying different interdependency strengths. In this case, the more reliable the system, the less the influence of the interdependency strength.

A similar study should be performed for real infrastructure systems under hazard events. Similarly, it is desirable implement improved solution techniques to practical decision support and resilience analyses.

ACKNOWLEDGEMENTS
This publication was partially funded by “Research Program 2012” Grant from the Office of the Vice President for Research - Universidad de los Andes (Bogotá, Colombia). The authors also gratefully acknowledge the support by the U.S. Department of Defense (Grant W911NF-13-1-0340), and FICO for providing the Xpress-MP licenses used in the computational experiments.

6. REFERENCES