Small Cluster in Cyber Physical Systems: Network Topology, Interdependence and Cascading Failures

Zhen Huang, Cheng Wang, Amiya Nayak, Ivan Stojmenovic, Fellow, IEEE

Abstract—In cyber physical system (CPS), computational resources and physical resources are strongly correlated and mutually dependent. Cascading failures occur between coupled networks, cause the system more fragile than single network. Besides widely used metric giant component, we study small component (small cluster) in interdependent networks after cascading failures occur. We first introduce an overview on how small clusters distribute in various single networks. Then we propose a percolation theory based mathematical method to study how small clusters be affected by the interdependence between two coupled networks. We prove that the upper bounds exist for both the fraction and the number of operating small clusters. Without loss of generality, we apply both synthetic network and real network data in simulation to study small clusters under different interdependence models and network topologies. The extensive simulations highlight our findings: except the giant component, considerable proportion of small clusters exists, with the remaining part fragmenting to very tiny pieces or even massive isolated single vertex; no matter how the two networks are tightly coupled, an upper bound exists for the size of small clusters. We also discover that the interdependent small-world networks generally have the highest fractions of operating small clusters. Three attack strategies are compared: Inter Degree Priority Attack, Intra Degree Priority Attack and Random Attack. We observe that the fraction of functioning small clusters keeps stable and is independent from the attack strategies.

Index Terms—Cyber Physical Systems; Small Cluster; Cascading failure; Interdependent Networks

1 INTRODUCTION

Cyber physical system (CPS) merges monitoring, computer-based control and communication systems. The hardware, software and networking technology are integrated together rather than computation alone. In a typical CPS system, embedded computers and communication networks are sort of computational resources, while actuator and sensors are viewed as physical resources. Nowadays, computational resources and physical resources are strongly coupled. That is, computational resources schedule and control physical resources, on the other hand, the proper functioning of computational resources requires steady operation of physical resources. As one important application of CPS, smart power grid is composed of power grid and communication/control network. Communication network depends on power grid for the electricity. Power grid, from another point of view, has to be controlled by communication network. The two networks are mutually-close connected. Similar to smart power grid, interdependence also exists in the wireless sensor/robot/actuator network. Robot/actuator receives or transmits messages through sensor network, while sensor placement, maintenance and relocation are completed by robot/actuator.

However, this growing interdependence in the coupled networks causes it more fragile than the single network [3], [4]. Cascading failures are big issues in such coupled networks. The unexpected malfunction of nodes initiates the sequence of failures not only within the single network, but also between the networks. The huge blackout happened in Italy in 2003 was caused due to cascading failures [3]: the initial breakdown of power station led to the disconnect of communication network, which further affected the operation of power stations. Understanding the underlying cascading failures patterns to protect interdependent cyber physical systems is quite necessary.

Recent studies [2][8] on interdependent networks assessed system vulnerability. By calculating the size of remaining giant component after cascading failures stop, they studied how the initial failures could damage the entire interdependent networks. Buldyrev et al. [3] indicated a simple ‘one-one’ interdependence model as well as percolation-based method used to analyze the model. Yagan et al. [5] studied how different interdependence...
models affect the network vulnerability. [2] indicated a near-optimal heuristic greedy algorithm to detect the critical nodes that might cause fatal cascading failures.

The common underlying assumptions of these works are that a node can function only if: 1) it belongs to the giant component; 2) it has at least one inter link. While, the first assumption is demanding and incompatible with many engineering reality. In many application scenarios, once the network gets fragmented due to the cascading failures, except the giant component, the network possibly has plenty of autonomous and self-sufficient disconnected clusters.

Fig. 1 gives an example of two networks A and B with ‘one-to-one’ [3] interdependence, where both of them fragment to clusters after the cascading failures. Network A has a giant component a1, and two small clusters a2, a3. B is similar with A. Notice that except the two giant components, the small clusters a2, a3, b2 and b3 do operate because the nodes involved have the supporting inter links. This example illustrates that in the interdependent networks, if we only consider the giant components, some small clusters that could function are neglected. Consequently, the results the mentioned studies obtained are not accurate enough to reflect the real coupled cyber physical systems. It is convincing that such ignored clusters could play important roles in the network of some real-life systems.

In this paper, we aim to study the small clusters caused by cascading failures in interdependent networks. Using real-world network data and extensive simulations, we first introduce the small cluster overviews, including the size distribution and the number of clusters, in single complex network. We show that the clusters in Erdős-Rényi random network are tiny, e.g., size is less than 5. While, the cluster size in scale-free network and small-world network could be much larger.

We indicate a percolation based mathematical analysis method to study how small clusters be affected by the interdependence between two coupled networks. We prove that both the fraction and the number of operating small clusters have upper bounds.

Without loss of generality, in experiment part, we use both synthetic network and real network data to study the impacts of two different interdependence models ‘k-n’ model [4] and ‘one-to-one’ model [3] on small clusters, where the descriptions of these models could be found in Section 4. We find that the fraction of small clusters do not monotonously increase as the growing of compactness of interdependence. The simulation validates our mathematical analysis that an upper bound for the fraction of small clusters exists. We also show that except the giant component and small clusters, the fragmentation of interdependent networks also bring massive tiny pieces, i.e., isolated single vertex.

Considering the initial failure could be either random or targeted, to this end, three different attack strategies are discussed: Inter Degree Priority Attack, Intra Degree Priority Attack and Random Attack. The interesting finding is that the proportion of functioning small clusters keeps stable and is independent from attack strategies, which, at the same time, leads to quite different sizes of the giant components. We observe that the Intra Degree Priority Attack destroys the whole system very effectively, compared with others.

Our main contributions in this article are:

- Show small cluster size distributions in various single networks
- Indicate a percolation based math tool to study the impacts of different interdependence models on small cluster in coupled networks
- Use extensive simulations to give insightful views on small cluster in interdependent networks, under different interdependence models and network topologies

This work explores how interdependence affects the system robustness and reveal the properties of small clusters. The cascading model applied is that only the node satisfying 1) it belongs to the component with sizes over predefined parameter $\Delta$; 2) it has at least one inter link, are considered as functioning. We do not consider
any network physical properties, such as power line transmission limit, electrical features and node capacity, and thus, do not apply any cascading models depending on physical characteristics.

This article is organized as follows: we discuss the existing studies in Section 2. In Section 3, we use examples and simulations to show why small clusters are important, and define the problems we try to study. Section 4 indicates the math tools to analyse small clusters in interdependent networks. The extensive simulation results and comparisons are in Section 5. Conclusions are summarized in Section 6.

2 RELATED WORK

Most of recent work studying interdependent networks are focusing on the interdependence model itself. Percolation theory is widely applied to reconstruct the process of cascading failures. [3] was the very first work in this area. The authors proposed an ‘one-to-one’ interdependence model with bidirectional inter link, where each node functions relying on one unique node in other network, and vice versa. A more general ‘multiple-to-multiple’ interdependence model was proposed in [13], where each node might have multiple inter links. Once at least one inter link is valid, the node is considered as operating.

Yagan et al. [5] introduced a ‘regular allocation’ algorithm to allocate the same number of inter links to each node. They proved that the regular allocation scheme is optimal in the case that the topologies of networks are unknown. [14] studied the percolation of several interdependent networks under random failure, and developed a general analytical framework. [17] discussed the robustness of n interdependent networks with partial support-dependence. Cascading link failures was introduced in [11]. A survey of interdependent networks can be found in [6].

Our previous work [4] pointed out that the nodes in each network have different roles, e.g., nodes in communication network can be roughly divided to relay nodes and control nodes. For those nodes with different roles, different number of inter links are allocated. We proposed a novel ‘k-n’ interdependence model and studied the relation between control cost and system robustness by calculating the fraction of giant component.

Huang et al. [12] discussed the target attack problem in interdependent networks. A mapping algorithm was indicated to transfer target attacks to random failures. The targeted failure in network of networks was recently studied in [15]. Dung T. Nguyen et al. [2] proposed a greedy algorithm based on networks’ interdependence to detect the critical nodes that might cause fatal cascading failures. BakTang-Wiesenfeld sandpile model was applied to study how interdependence affects cascading behaviors [10]. They found the interconnectivity provides benefits which can balance the dangers at stable point. While, too much interconnectivity impacts the system. Huang et al. [18] developed a method for evaluating how clustering within the networks affects the system robustness. [16] proposed an epidemic spreading based theory, which is more straightforward than previous work, to formulate the interdependent networks.

The above research drawn insightful results and conclusions, helped us understand the behaviors of interdependent networks. However, the common underlying assumption that only the node belongs to giant component can operate is demanding in real-life networks. These works ignored the small clusters.

3 SMALL DISCONNECTED CLUSTERS

Recent studies explore network characteristics by using the metric of giant component, or specifically called largest connected cluster [2], which is defined as the largest connected cluster that spans across most of the network. It is reasonable to study giant component of a single network, because each network could be approximately represented by its giant component. However, there is only one giant component in network and in most scenarios it does not span the entire network [9]. Non-giant components are called here small components or small clusters, although their sizes may not be limited when network size grows.

In an interdependent cyber physical system, in addition to the giant component, the small clusters may play an important role in the cascading failure process. In this section, we explore the role of small clusters in different single complex networks. In the sequel, we use giant component and largest cluster interchangeably.

Considering a single network, we denote it as $G = (V, E)$, where $V$ is the set of vertices of $G$, $E$ is the set of edges between each pair in $V$. Once a set of vertices $V_j \subseteq V$ is failure or malfunctioning, the whole network $G$ fragments to a set of disconnected sub networks, which denoted by $S(V_j) = \{G_1, G_2, G_3, ..., G_j\}$, among which there exists one giant component whose size is proportional to the size of $G$ [9]. Without loss of generality, we let $G_1$ represent the unique giant component, and call others small clusters. Previous studies were mostly focusing on exploring the characteristics of $G_1$. While, we give our argument that the remaining elements in $S(V_j)$ are also important, thus cannot be simply ignored.
3.1 Small Clusters in Single Network

Problem Definition: Given a single network $G = (V, E)$ and initial random failure $V_f \subseteq V$, let $|G|$ be the size of $G$. After $V_f$ is applied to $G$, what are the sizes of $\{G_2, G_3, ..., G_j\}$? How many small clusters exist in $S(V_f)$, i.e., the value of $j - 1$?

We use the extensive simulation to illustrate to what extent the small clusters could be important in a single network. We build synthetic random, scale-free and small-world networks using Erdős-Rényi model, Barabási-Albert model and Watts and Strogatz model [9] respectively. For each of them, the network size $|G| = 2000$, with random initial failure $|V_f| = 400$. To obtain the mean value of the small clusters, this construction-failure process was repeated for 60 times. Moreover, we use real Western States Power Grid data [1] and do the same process, as shown in Fig. 2.

After eliminating the largest cluster, one can observe the mean sizes of small clusters for each type of network. In random network, most of the nodes belong to the clusters $\{G_i, G_{i+1}...\}$ whose sizes are less than 5 ($1 \leq |G_i| \leq 5$). In other words, the random network fragments to very tiny pieces. However, for the other three network types, sizes are different. A large proportion of nodes in scale-free and small-world networks locates into the clusters $\{G_i, G_{i+1}...\}$, where $11 \leq |G_i| \leq 400$. Therefore, after suffering from the random failure, the scale-free and small-world networks break into a giant component $G_1$, few relatively smaller clusters with several hundreds of nodes inside, and hundreds of small clusters whose sizes are less than 10. For Western States Power Grid, an important observation is that the small clusters are not so ‘small’. A large proportion of nodes are within the clusters whose sizes may be larger than 400. We also notice that the distribution of small cluster’s size does not follow a monotonically decreasing trend, according to Fig. 2.

We repeat the simulation to calculate the cluster size distribution (the number of small clusters in certain size ranges), as shown in Table 1. The random network has the highest number of small clusters, and consequently the lowest average cluster size, as shown in Fig. 2. We
believe the reason is that the random network constructed by Erdős-Rényi model does not have the small world effect, which is also known as six degrees of separation. Once the random failure occurs, the network fragments to massive tiny disconnected local clusters. For the same reason, the small-world network has fewer number of clusters (195.3) than the others: a few long distance edges connect different local clusters, which consequently merge to the bigger clusters. Fig. 2(c) illustrates that the number of nodes belong to \( |G_i| \leq 5 \) are much less than others.

We also observe that the number of clusters \( G_i \), where \( |G_i| \leq 10 \), dominates the total number of small clusters. Over 90% of small clusters are quite small. Only a few clusters occupy a significant part of the whole network.

### 3.2 Small Clusters in Interdependent Networks

The findings according to Fig. 2 and Table. 1 bring us an overview on how the network behaves once got fragmented. Knowing this behavior directs us understand how to deal with modern systems. In a multiple sink wireless sensor actuator network, if the failure of nodes causes the disconnected of clusters, for each cluster it may has its own sink node. In this case, the original whole system is considered as several independently-working sub-systems. Not only the largest cluster, but also the small clusters work properly in this steady state.

- it has at least one inter link,
- it belongs to a cluster with size greater than a predefined threshold \( \Delta \).

Fig. 3 is a typical working pattern for the interdependent networks: the giant components \( a_1 \) and \( b_1 \) are partially dependent. Small clusters may fully interdependent \( (b_2, a_1) \) or partially interdependent \( (a_2, b_1) \) on the giant components. It is possible that two small clusters may mutually interdependent and thus operate properly, such as \( a_3 \) and \( b_3 \).

The way that the recent studies [3]–[5] used that only considering largest cluster is inappropriate, because significant part of network is ignored deliberately. In the remaining parts of this article, we will study the sizes of small clusters in the interdependent networks, and reveal how the network structure and initial failure impact small clusters.

**Problem Definition**: Given \( G_p = (V_p, E_p) \) and \( G_c = (V_c, E_c) \) representing physical resource network and computational resource network respectively, e.g., power grid and communication network, a set of failure nodes \( V_f \subseteq V_p \) is applied to the network \( G_p \). Due to the interdependence, \( G_p \) and \( G_c \) fragment to two sets of clusters \( S_p(V_f) = \{ G_{p1}, G_{p2}, G_{p3}, \ldots \} \) and \( S_c(V_f) = \{ G_{c1}, G_{c2}, G_{c3}, \ldots \} \) respectively. Besides the largest clusters \( G_{p1} \) and \( G_{c1} \), we are interested in the fractions of operating small clusters \( \{ G_{p2}, \ldots, G_{pI} \} \) and \( \{ G_{c2}, \ldots, G_{cj} \} \), where the two assumptions mentioned above are satisfied that: 1) \( |G_{p2}, \ldots, G_{pi}|, |G_{c2}, \ldots, G_{cj}| > \Delta / 2 \) each node involved has at least one inter link. Moreover, as different interdependence models indicating quite different system vulnerabilities, we will compare our ‘k-n’ model [4] with ‘one-to-one’ [2], [3] to discuss how interdependence affects the small clusters.

### 4 Mathematical Estimation

In this section, we indicate a method to estimate the functioning giant components and small clusters in inter-
dependent networks, under two interdependence models ‘k-n’ and ‘one-to-one’.

Although given the same network model parameters, e.g., power law exponent or mean degree, we may obtain different results. Thus, in this section, we use mean filed to study the ensembles of $G_p$ and $G_c$. Definitions and notations are listed in Table 2. We introduce a notation $F(\phi, G_p)$ from [5] to represent the expected fraction of giant component in the subnetwork which occupies the fraction $\phi$ of the nodes in the entire network $G_p$. Similarly, $F(\phi, G_c)$ is the fraction of giant component for network $G_c$. We begin random failure set $V_f$ with the size of $(1 - \phi) \cdot |G_c|$ on network $G_p$.

We briefly introduce ‘k-n’ and ‘one-to-one’ models to help us understand this section more profoundly.

\textbf{• ‘k-n’ interdependence model} was proposed for smart power grid, where the nodes in $G_c$ are divided into information relay nodes and control nodes. Each node in $G_c$ is supported by only one node in $G_p$, while each node in $G_p$ is monitored and operated by $k$ control nodes. On the other hand, each control node operates exactly $n$ nodes in $G_p$. Two types of unidirectional inter links exist in the model: energy-dependency and control-dependency. An example is given in Fig. 4(a).

\textbf{• ‘one-to-one’ model} is a simple and widely used model, where each node in $G_p$ depends on exactly one unique node in $G_c$, and vice versa. Nodes in each pair connected by one inter link are mutually dependent. Therefore, $|G_p| = |G_c|$ is satisfied in this model. Fig. 4(b) shows an example of ‘one-to-one’ model.

\textbf{Theorem 1:} Recall $\Delta$ is the threshold size above which we consider the small cluster could operate independently. Then, it follows that a smaller value of $\Delta$ leads to a higher robust interdependent system.

\textbf{Proof 1:} Consider the ‘k-n’ interdependence model. Let $S(\phi, G_p, \Delta)$ represent the fraction of operating small clusters whose sizes are greater than $\Delta$ in network $G_p$, with $1 - \phi$ of network failing. We begin the random removal of a fraction $1 - \phi$ of nodes in $G_p$. The fraction of nodes in the remaining network which retain at least one inter link is $\mu_{p1} = \phi$. As required by the second condition, nodes that belong to the giant component and small clusters with sizes greater than $\Delta$ could operate. Thus, the remaining functioning fraction $\mu_{p1}$ is given by

$$\mu_{p1} = \mu_{p1} \cdot (F(\mu_{p1}, G_p) + S(\mu_{p1}, G_p, \Delta)).$$

As the network $G_p$ fragmentation, a part of inter links is removed. Some nodes in network $G_c$ lose inter link and thus stop function. the probability for one energy inter link removal is $1 - \mu_p$. (See details in [4]). The fraction of nodes in $G_c$ that has an inter link is approximate to $\mu_p$.

So, we have $\mu_{c2} = \mu_{p1}$. Then, the fraction of operating giant component and small clusters in $G_c$ is given by

$$\mu_{c2} = \mu_{p1} \cdot (F(\mu_{c2}, G_c) + S(\mu_{c2}, G_c, \Delta)).$$

The failure in $G_c$ causes further failures in $G_p$. As a result, the fraction $1 - \mu_{c2}$ of control centers are removed, given the network size $|G_p|$ is large enough. The probability for each control inter link to be removed is approximated by $1 - \mu_{c2}$, because each control center has the same ability to control $n$ nodes. With this respective, a node in $G_p$ loses all its $k$ control links with the probability of $(1 - \mu_{c2})^k$. The fraction $\mu_{p1}'$ of nodes with at least one inter control link is $(1 - (1 - \mu_{c2})^k) \cdot \mu_{p1}$. By the transformation idea in [4], we consider the effect of removing the fraction of $1 - \mu_{p3}'$ of nodes in $\mu_{p1}'$ has the same effect as taking out the same fraction size from $\mu_{p1}'$. Thereby, from initial network $G_p$ to $\mu_{p3}'$, we have the following equivalent removing process:

$$1 - \phi + \phi \cdot (1 - \mu_{c2})^k = 1 - (\phi - \phi \cdot (1 - \mu_{c2})^k).$$

Thus, the equivalent $\mu_{p3}' = \phi \cdot (1 - (1 - \mu_{c2})^k)$. The fraction of operating clusters is

$$\mu_{p3} = \mu_{p3}' \cdot (F(\mu_{p3}', G_p) + S(\mu_{p3}', G_p, \Delta)).$$
The fractions of operating giant component and small clusters in the final steady state are given by

\[
\begin{align*}
\lim_{j \to \infty} \mu_{p_j} &= \mu_{p_{\infty}} = x \cdot (F(x,G_p) + S(x,G_p,\Delta)), \\
\lim_{j \to \infty} \mu_{c_j} &= \mu_{c_{\infty}} = y \cdot (F(y,G_c) + S(y,G_c,\Delta)).
\end{align*}
\]  

Theoretically, once solving \(x\) and \(y\) in Eq. (4), we reach to the final results. While, no analytical theory is yet available to give the closed-forms of \(F(\cdot)\) and \(S(\cdot)\). But we still can derive some conclusions from these equations.

We use \(\pi_s\) to denote the probability that a random chosen vertex belongs to a small cluster with the size exactly \(s\) in total. The generating function of \(\pi_s\) is

\[
H(z) = \pi_1 \cdot z + \pi_2 \cdot z^2 + \pi_3 \cdot z^3 + \pi_4 \cdot z^4 + \pi_5 \cdot z^5 + \cdots.
\]  

Combining the threshold \(\Delta\), we have

\[
H(z)|_{s \geq \Delta} = \pi_{\Delta} \cdot z^\Delta + \pi_{\Delta+1} \cdot z^{\Delta+1} + \cdots.
\]

Once reducing the size threshold \(\Delta\), the value of left of Eq. (7) increases. In other words, the fraction of operating small clusters increases. Therefore, we can conclude \(S(y,G_c,\Delta)\) and \(S(x,G_p,\Delta)\) on the right side of Eq. (4) would grow. Consequently, the left side \(x\) and \(y\) would increase. Given \(F(\phi,G_p)\) and \(F(\phi,G_c)\) are monotone increasing functions [9], Eq. (5) deterministically grows. Thus, a smaller value of \(\Delta\) leads to a higher robust interdependent system.

**Lemma 1:** In ‘\(k\)-\(n\)’ interdependence model, there exists an upper bound on the fraction of operating small clusters.

**Proof 2:** First of all, we have,

\[
\begin{align*}
0 \leq F(x,G_p) + S(x,G_p,\Delta) &\leq x, \\
0 \leq F(y,G_c) + S(y,G_c,\Delta) &\leq y.
\end{align*}
\]

From the result in [4], we get that for a large value of \(k\), the giant components \(F(x,G_p)\) and \(F(y,G_c)\) approach to \(x\) and \(y\) respectively. In other words, with a strong interdependence, the remaining parts in both networks are almost fully connected. In this case, the fractions of small clusters hold that

\[
\begin{align*}
0 \leq S(x,G_p,\Delta) &\leq x - F(x,G_p) = 0, \\
0 \leq S(y,G_c,\Delta) &\leq y - F(y,G_c) = 0.
\end{align*}
\]

On the other hand, there are no survivals (both giant component and small cluster) if \(k = 0\), according to Eq. (4). Thus, as \(k\) is increasing from 0 to infinite, both \(S(\phi,G_p,\Delta)\) and \(S(\phi,G_c,\Delta)\) start from zero and vanish to zero, which completes the proof.

<table>
<thead>
<tr>
<th>Notations for math approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G_p)</td>
</tr>
<tr>
<td>(G_c)</td>
</tr>
<tr>
<td>(\mu'_{p_1})</td>
</tr>
<tr>
<td>(\mu'_{c_1})</td>
</tr>
<tr>
<td>(\mu'_{p_i})</td>
</tr>
<tr>
<td>(\delta)</td>
</tr>
</tbody>
</table>

| \(\{\}\) sizes over \(\Delta\)ponent and small clusters with sizes over \(\Delta\) in \(\mu'_{c_i}\) |

Once repeat the above steps, we can observe the pattern of these equations as follows:

\[
\begin{align*}
\mu'_{p_1} &= \phi, \\
\mu'_{p_2} &= \mu'_{p_1} \cdot (F(\mu'_{p_1},G_p) + S(\mu'_{p_1},G_p,\Delta)), \\
\mu'_{c_2} &= \mu'_{c_1} \cdot (F(\mu'_{c_1},G_c) + S(\mu'_{c_1},G_c,\Delta)), \\
\mu'_{c_3} &= \mu'_{c_2} \cdot (F(\mu'_{c_2},G_c) + S(\mu'_{c_2},G_c,\Delta)), \\
&\vdots, \quad \ldots, \\
\mu'_{c_{p_2-1}} &= \phi \cdot (1 - (1 - \mu'_{c_{p_2-1}})^k), \\
\mu'_{c_{p_2}} &= \mu'_{c_{p_2-1}} \cdot (F(\mu'_{c_{p_2-1}},G_c) + S(\mu'_{c_{p_2-1}},G_c,\Delta)), \\
\mu'_{c_{p_2+1}} &= \mu'_{c_{p_2}} \cdot (F(\mu'_{c_{p_2}},G_c) + S(\mu'_{c_{p_2}},G_c,\Delta)), \\
\mu'_{c_{p_2+2}} &= \mu'_{c_{p_2+1}} \cdot (F(\mu'_{c_{p_2+1}},G_c) + S(\mu'_{c_{p_2+1}},G_c,\Delta)), \\
&\vdots, \quad \ldots, \\
\mu'_{c_{p_2+j}} &= \mu'_{c_{p_2+j-1}}, \\
\mu'_{c_{p_2+j+1}} &= \mu'_{c_{p_2+j}} \cdot (F(\mu'_{c_{p_2+j}},G_c) + S(\mu'_{c_{p_2+j}},G_c,\Delta)).
\end{align*}
\]

Once the cascading failures stop, at the final steady state, the following equations hold

\[
\begin{align*}
\mu'_{p_{2j+1}} &= \mu'_{p_{2j+3}} = \mu'_{p_{2j-1}}, \\
\mu'_{c_{2j}} &= \mu'_{c_{2j+2}} = \mu'_{c_{2j-2}},
\end{align*}
\]

because both \(G_p\) and \(G_c\) stop fragmenting. Let \(x = \mu'_{p_{2j+1}} = \mu'_{p_{2j+3}} = \mu'_{p_{2j-1}}\) and \(y = \mu'_{c_{2j}} = \mu'_{c_{2j+2}} = \mu'_{c_{2j-2}}\), then we have:

\[
\begin{align*}
x &= \phi \cdot (1 - (1 - y \cdot F(y,G_c) + S(y,G_c,\Delta)))^k, \\
y &= x \cdot (F(x,G_p) + S(x,G_p,\Delta)).
\end{align*}
\]
Theorem 2: In ‘k-n’ interdependence model, an upper bound exists for the number of operating small clusters $\delta$.

Proof 3: A straightforward loose upper bound of $\delta$ is the number of survived nodes in the network. We will derive a tighter bound based on $S(x, G_p, \Delta)$. Loose bound satisfies:

$$0 \leq \delta \leq S(x, G_p, \Delta) / \Delta. \quad (10)$$

The mean size of the small cluster containing a randomly chosen operating node is given by:

$$\bar{s} = \sum_{s \geq \Delta} |G_{p1}| \cdot s \cdot \pi_s / \sum_{s \geq \Delta} |G_{p1}| \cdot \pi_s,$$

where $|G_{p1}|$ is the giant component size of $G_p$. A tight bound of $\delta$ could be obtained by

$$0 \leq \delta \leq \frac{S(x, G_p, \Delta)}{\bar{s}} = \frac{S(x, G_p, \Delta) \cdot \sum_{s \geq \Delta} |G_{p1}| \cdot \pi_s}{\sum_{s \geq \Delta} |G_{p1}| \cdot \pi_s}. \quad (11)$$

As proven in Lemma 1, $S(x, G_p, \Delta)$ has an upper bound, thus, we can conclude that $\delta$ also has an upper bound.

We also study the widely used ‘one-to-one’ interdependence model [3]. Once repeat the steps above, we obtain new transcendental equations:

\[
\begin{align*}
x &= \phi \cdot (F(y, G_c) + S(y, G_c, \Delta)), \\
y &= x \cdot (F(x, G_p) + S(x, G_p, \Delta)).
\end{align*}
\]  

In general, it is difficult to derive the expression for $x$ and $y$. Deriving the closed-form of $F(y, G_c)$ and $S(y, G_c, \Delta)$ is a tough task and still an open problem.

### 5 Experiments

In this section, we use the extensive simulations to reveal how different network topologies and interdependence affect the small clusters.

#### 5.1 Simulation Setup

We build synthetic random, scale-free and small-world networks using Erdős-Rényi model, Barabási-Albert model and Watts and Strogatz model [9] respectively. We also use real Western States Power Grid data [1], where the data set could be found at [19].

We write a specific program using Java to simulate the cascading failure process in interdependent networks. In all the experiments the following approach is employed:

- we construct one or two synthetic networks
- then, we couple two or one synthetic network with Western States Power Grid using ‘k-n’ or ‘one-to-one’ interdependence models.
- we choose nodes in network $G_p$ as failure or attack randomly or depending on the degrees.
- the failure propagates between the two networks.
- Our program simulates each stage, stores and updates the network information.
- finally, our program calculates the number of survived nodes and the number of clusters in both networks.

#### 5.2 Interdependence Impacts

To study how small clusters be affected by the interdependence between coupled networks, we compare the ‘k-n’ interdependence model [4] with the widely used ‘one-to-one’ model [2], [3]. We couple two synthetic small-world networks using both models. The results are given in Fig. 5. Some observations are as follows:

- Generally speaking, we observe that unless the interdependent networks collapse, the small clusters surviving from the cascading failures occupy a big part of the system. In the case of (c), (d) and (e), the fractions of operating small clusters could be greater than 40% of the whole system. In Fig. 5(d), when $|V_f| = 120$, the size of small clusters (41%) even exceeds the size of largest cluster (27%) in $G_p$.

- As shown in (a), (b) and (c), the proportions of small clusters first increase with the growing of $|V_f|$, then decrease to zero once the entire system fails. The sum of largest cluster and small clusters keeps stable (approximate 95%) for $|V_f| < 120$ in (c). This value for case (b) is also 95% if $|V_f| < 60$, after which the size of functioning clusters drops quickly.

- In ‘k-n’ interdependence model, comparing (a), (c) and (b), (d), we find increasing $k$ could increase both largest cluster and small cluster sizes. The reason is that higher value of $k$ means the node in $G_p$ is operated by more control nodes. The hardware redundancy improves the system robustness [4]. Notice that the proportion of small clusters does not grow unlimitedly. As shown in (d) and (e), the small cluster size experiences a slightly reduction even we raise $k$ from 3 to 4. But, at the same time, the largest cluster keeps growing. The reason of this phenomenon is that once $G_p$ and $G_c$ are tightly coupled, the nodes are more likely to survive and stay in the largest cluster, which potentially affects the sizes of small clusters. Therefore, for the small values of $k$, the small cluster proportion is monotonously increasing with $k$. Then it decreases because the increasing largest cluster contests the nodes. The finding here supports our Lemma 1 that an upper bound exists for the fraction of operating small clusters.

- Recall different values of $n$ do not affect the total number of inter links [4], thus do not improve the system.
The impacts of interdependencies on smaller clusters. We use coupled small-world networks for the implementations. For (a), (b), (c), (d) and (e), the sizes of $G_p$ and $G_c$ are 4000 and 6000. For ‘one-to-one’ model (f), both networks have the same size of 5000. We set the small cluster size threshold $\Delta = 20$.

robustness. The findings in (a), (b), (c) and (d) have the similar results. Comparing (a) and (b), one can observe that raising $n$ leads to a sharper transition [4], and thus causes a different distribution of small clusters when $60 < |V_f| < 120$. Same thing occurs in (c) and (d). Therefore, according to our simulations, the value of $n$ has the impacts on small clusters rather than largest cluster.

- (f) reports how ‘one-to-one’ interdependence model affects the small clusters, which only occupy a small fraction of the entire system. One interesting observation is that the sums of largest cluster and small cluster for $G_p$ and $G_c$ are equal. In other words, these two coupled networks have the same number of survivals. As required by ‘one-to-one’ model, each node in $G_p$ depends on exactly one unique node in $G_c$. Each survival node has the unique interdependent node for operating. Thereby, the numbers of survival nodes in $G_p$ and $G_c$ are equal.

### 5.3 Tiny Pieces in Interdependent Networks

We are also interested in revealing how the interdependent networks look like once they get fragmented, in terms of clusters distributions. As percolation theory pointed out, there exists a constant threshold for the proportion of faulty nodes, beyond which the system collapses [6]. Our study [4] found once the whole system collapses, even the small clusters cannot survive, i.e., the networks break to disconnected tiny pieces.

However, if the system does not collapse, the small clusters occupy a significant proportion according to Fig. 6. With increasing $\Delta$, the fraction of small clusters and the size of largest cluster dramatically reduce at the same time. Both $G_p$ and $G_c$ keep the stable sizes of survivals if $\Delta = 2$, but are close to vanish as $\Delta$ grows to 20. This phenomenon indicates that considering small clusters highly affects the final results. The findings in Fig. 6 also support our Theorem 1 that a smaller value of $\Delta$ leads to a higher robust interdependent system. Thereby, only using giant component to study the interdependent networks is not enough to draw the accurate conclusions.

The more important observation here is that the threshold $\Delta$ causes different proportions of small clusters in the final steady state. As shown in Fig. 6(a) and Fig. 6(b), the fraction of small clusters occupies at most 20% of the network. But for Fig. 6(c) and Fig. 6(d), 2% to 5% of the network is consisted by small clusters. This gives us an overview on how $\{|G_{p1}|, |G_{p2}|, ......|G_{p1}||$ and $\{|G_{c1}|, |G_{c2}|, ......|G_{c1}||$ distribute. In our simulation case, except the largest cluster, about 3% of networks are the small cluster whose sizes are over 20 in $G_c$. A significant fraction of $G_c$ is consisted of the tiny clusters with the sizes approximate to 2. While, the remaining parts of the
greater than coupled scale-free, for both $G_p$ and $G_c$. Meanwhile, only 5% of nodes which have small cluster sizes of coupled small-world networks is more or less stable for various sizes of $\Delta$. In other words, only the largest cluster could survive in the final steady state. While, the fraction of small clusters in coupled scale-free networks is more or less stable for various sizes of $V_f$. The interesting thing for coupled small-world networks is that small clusters occupy a large proportion of $G_p$, up to 10%. In the meanwhile, only 5% of $G_c$ is small clusters.

Comparing Fig. 7(b) and Fig. 7(c), one can observe that the small cluster sizes of coupled small-world are greater than coupled scale-free, for both $G_p$ and $G_c$. To explain the reason, we introduce an fundamental concept of scale-free network: that most part of nodes which have few edges are connected by the ‘hub’ whose degree is relatively high. The breakdown of ‘hub’ nodes may lead to disconnected tiny clusters, sometimes even to plenty of isolated vertex.

We notice that the coupled random networks are of the most vulnerable among all three implementations. Its counterpart, coupled small-world networks, has the fastest degradation speed. 120 failure nodes could devastate both of them.

5.5 Number of Functioning Clusters

We are interested in finding how many small clusters could be created after cascading failures terminate. In real world interdependent networks, this problem translates to how many isolated clusters could operate interdependently. These counts are given in Fig. 8. Coupled random networks belong to only the giant component (if any). None of working small clusters could be found. Fig. 7(a) also shows that the proportion of small clusters is zero. Coupled small-world networks have the largest number of functioning clusters, which are up to 15 in $G_p$ and 12 in $G_c$, thus at most 27 different clusters could work in parallel in this system. The number of coupled scale-free networks are in between, having approximate 9
We evaluate the sizes of small clusters of three different critical nodes possibly happen in the real-life networks. Previous sections study the small clusters caused by the random failure $V_f$. While, the targeted attacks to some critical nodes possibly happen in the real-life networks. We evaluate the sizes of small clusters of three different attack strategies:

- **Inter Degree Priority Attack**: the nodes in $V_f$ have the highest degrees of inter links. Those nodes are targeted to be attacked in the beginning.
- **Intra Degree Priority Attack**: the nodes with highest intra degrees are prior to be attacked.
- **Random Attack**: failure nodes are randomly chosen.

We use the real Western States Power Grid and an synthetic scale-free network to construct the interdependent networks. The Western Power has 4941 nodes with 6594 edges, the synthetic scale-free network generated has the power law exponent of 3, with the initial size of 6000. Our previous ‘$k$-$n$’ interdependence model is applied, with $k=2$, $n=3$. Initially, $G_p$ has 4000 nodes, $G_c$ has 6000 nodes. $\Delta = 20$.

Fig. 8 reports the sizes of largest clusters and small clusters caused by different attack strategies. We observe no matter which attack we use, the size of functioning small clusters is almost equal to zero in $G_p$. In other words, if the system survives from the cascading failures, $G_p$ only consists a largest cluster. In the meanwhile, the remaining parts fragment into massive tiny pieces.

One also can observe that the Intra Degree Priority Attack destroys the whole system very effectively. The first 40 nodes with the highest intra degrees could devastate both networks. It needs 50 nodes for Inter Degree Priority Attack strategy to complete the same mission. By contrast, the interdependent networks perform much better under Random Attack, where 40 failure nodes only affect a small part on $G_p$ and $G_c$.

The principle given by these results is that different attack strategies do not affect the proportion of operating small clusters in the interdependent network. On the other hand, they indeed lead to the different vulnerabilities. Targeted attack makes the interdependent networks more fragile, especially when the attacker chooses the candidates with high intra degrees.
Fig. 9. The sizes of small clusters and largest cluster caused by initial failure $V_f$. We couple Western States Power Grid ($G_p$) and an synthetic scale-free network ($G_c$) using our ‘k-n’ interdependence model, with k=2, n=3. Barabási-Albert model is used to generate scale-free network with power law exponent 3. The initial network sizes are 4941 and 6000 for $G_p$ and $G_c$ respectively. The initial failures $V_f$ start from $G_p$. We show the average results of 20 instances for each.

5.7 Discussions

As we found in Section 3, the small-world network has the highest average size of small clusters, compared with random network and scale-free network. What we observed in this section is quite similar. The coupled small-world networks generally have the highest fractions of operating small clusters. One difference is that Fig. 8 tells us the number of operating clusters for coupled small-world networks is higher than the counterparts. While, in the single network, this number should be the lowest, according to Table 1. The reason is that although the number of clusters is not as large as the others, the small-world network has significant part of clusters whose sizes are quite bigger, which are more likely to survive from cascading failures. We also find in interdependent random networks, small cluster is almost non existent. Thus, only studying giant component is enough in such networks.

However, our mathematical and experimental work does not consider any network physical characteristics, such as line transmission limit, electrical feature and node capacity. This work is based on network topology theory and stochastic theory. More studies on various cascading failure models and physical properties are required to understand real interdependent systems.

Also, in this work, the interdependence coupling method is random. There are many different coupling methods and the random interdependency usually does not occur in real interdependent systems. It is much more common that a central computational station depends on a central physical node, and vice versa [20]. Moreover, the coupled networks display some similarity in structure. An overpopulated area has many physical and computational resources.

In ‘one-to-one’ model, we study random, degree-based and preferential coupling schemes. Fig. 10 shows the difference of robustness of these schemes from our simulations. We couple two small world networks by connecting high degree node in $G_p$ with high degree node in $G_c$ in Fig. 10(a). Fig. 5(f) shows that degree-based coupling scheme is significantly more robust to random failure.
We also compare preferential coupling scheme in Fig. 10(b), where the neighbors of coupled pair tend to be coupled. This scheme is also much more robust than random coupling scheme. Our simulation shows that the fractions of operating small clusters in these coupling schemes are also significant.

6 CONCLUSIONS

In interdependent networks, except giant component, the small clusters play important roles. We indicate a percolation based mathematical tool to study the small clusters under different interdependence models and network topologies. Both mathematical and experimental results demonstrate that an upper bound exists for the fraction of operating small clusters. We also discover that the interdependent small world networks generally have the highest fractions of operating small clusters.

This work is helpful to find insightful properties of small clusters in interdependent networks. While, due to the lack of math tools, no analytical theory is yet available.

Also, we need experiments on real world interdependent systems. How to find the inter relations of real system and abstract them to simple models is an important topic for future work.

REFERENCES

Amiya Nayak received his B.Math. degree in Computer Science and Combinatorics and Optimization from University of Waterloo in 1981, and Ph.D. in Systems and Computer Engineering from Carleton University in 1991. He has over 17 years of industrial experience in software engineering, avionics and navigation systems, simulation and system level performance analysis. He is in the Editorial Board of several journals, including IEEE Transactions on Parallel and Distributed Systems, International Journal of Parallel, Emergent and Distributed Systems, International Journal of Computers and Applications, and EURASIP Journal of Wireless Communications and Networking. Currently, he is a Full Professor at the School of Electrical Engineering and Computer Science at the University of Ottawa. His research interests are in the area of fault tolerance, distributed systems/algorithms, and mobile ad hoc networks.

Ivan Stojmenovic was editor-in-chief of IEEE Transactions on Parallel and Distributed Systems (2010-3), and is founder of three journals. He is editor of IEEE Transactions on Computers, IEEE Network, IEEE Transactions on Cloud Computing and ACM Wireless Networks and steering committee member of IEEE Transactions on Emergent Topics in Computing. Stojmenovic is on Thomson Reuters list of Highly Cited Researchers from 2013, has top h-index in Canada for mathematics and statistics, and has over 15000 citations. He received five best paper awards. He is Fellow of IEEE, Canadian Academy of Engineering and Academia Europaea. He received Humboldt Research Award.