C. Alcaraz, "Resilient Industrial Control Systems based on Multiple Redundancy", *International Journal of Critical Infrastructures IJCIS*, vol. 13, pp. 278-295, 2017.
http://doi.org/10.1504/IJCIS.2017.10009287
NICS Lab. Publications: https://www.nics.uma.es/publications

Resilient Industrial Control Systems based on Multiple Redundancy

Cristina Alcaraz

Department of Computer Science, University of Malaga,

Campus de Teatinos s/n, 29071, Malaga, Spain

alcaraz@lcc.uma.es

Abstract

The incessant search for cost-effective recovery solutions for *structural controllability* has led to one of the most challenging research areas within the field of critical infrastructure protection. The resilience of large heterogeneous distributions, like industrial control scenarios, is proving to be a complicated mission due to the inherent non-locality problems of structural controllability and its susceptibility to advanced threats. To address these issues, this paper proposes a new repair approach based on multiple redundant pathways and the lessons learnt from the work presented in [2]. From [2], we have adapted the local measures, to combine them with each of the five strategies of remote reconnection described in this paper. To validate the sustainability of the combined approaches, two practical case studies are presented here, showing that a local dependence on a brother driver node together with remote dependence is enough to reach optimal states in linear times.

Industrial Control Systems; Control; Resilience; Restoration; Structural Controllability; Community Structures.

1 Introduction

Cost-effective automated recovery approaches capable of restoring control capacities in linear times have been the subject of numerous research studies. A clear example of this effort is our own, earlier work [2]. In this paper, we addressed the restoration problem of *structural controllability* [14] by finding optimal or suboptimal solutions mainly based on local redundancy, where the resilience of the underlying system is sustained by a node in the immediate surroundings. However, and although a narrow local dependence on redundant pathways is one of the best options to guarantee recovery in linear states, this may not properly converge due to the current complexities and dynamic functionalities of today's control industry [3, 5], and its susceptibility to specialised attacks [24, 22, 4, 1]. These systems are increasingly demanding more in terms of performance, security and resilience, and the control must always be operative from anywhere, at any time and anyhow [3, 5]. For this reason, we extend the study in

[2] to explore not only the effectiveness of local resources but also the suitability and convenience of using external assets for self-healing.

To model effective restoration solutions, we require the technical capacity of graph theory, structural controllability and the power dominance theory [11], to conceptually represent a specific scenario based on similar structures to the real monitoring systems [18]. Most of the solutions that have been proposed to date primarily concern the extraction and management of tree-based structures [16, 15, 21] and simple redundancy-based approaches [25]; but beyond this, more research is necessary to provide more automated preservation solutions, considering the constraints and critical features of the environment. These technical deficiencies have been the trigger that has encouraged us to return to the work we covered in [2], and extend the resilience capacity to accommodate the redundancy measures in a selective set of driver nodes (the controllers), all of which are located outside the vicinity of an affected node. Thus, when local areas are completely compromised and isolated from the network, the system is still able to solve the situation in optimal times.

For the selection of drivers, five strategies have been defined according to the type of grouping of the driver nodes (e.g. by distance, strength or density), and whose capacities are combined with the local ability for reconnection. To delimit the chief differences with [2], we highlight the main contributions of this work and include two case studies: (i) to maximise the redundant resources through five different strategic approaches; (ii) redesign the power dominance properties taking the new redundant measures into account; and (iii) determine the ideal number of redundant measures for restricted scenarios.

The remainder of this paper is structured as follows: Section 2 outlines preliminary assumptions related to the concepts of structural controllability and power dominance, and also defines the contextual and adversarial model applied throughout the paper. Section 3 describes the five approaches proposed for selecting driver nodes from a strategic point of view, and specifies a new structural controllability and dominance version together with the recovery scheme. All these approaches are then analysed in Section 4 through two practical case studies, and Section 5 concludes the paper and presents future work.

2 Preliminar Assumptions: Contextual and Adversarial

As defined in [2], our approaches are principally based on directed weighted graphs with cycles of type $\mathcal{G}_w(V,E)$. This graph specifies the topological structure of a control network composed of V nodes and E links. This construction concerns those smart nodes (e.g. mobile hand-held interfaces, robots, sensors, and actuators) capable of dynamically gaining access to the control network, and whose links are able to interact with diverse kinds of technologies and transport control loads related to commands, measurements and alarms. To extract these loads, the concept of the *control load capacity* (CLC) can be applied [2] whose value can be computed through the concept of

edge betweenness centrality (EBC) as also detailed in [17]:

$$E_{BC}(e) = \sum_{s,t \in V} \delta(s,t \mid e)\delta(s,t) \tag{1}$$

where $\delta(s,t)$ comprises the number of shortest (s,t)-paths and $\delta(s,t \mid e)$ the number of paths passing through edge e. The result is a new matrix E_{BC} containing the sum of the fraction of the shortest paths that pass through a given edge, e. In this way, the edges with the highest centrality participate in the maximum capacity to drive the 'main' control loads between two peers in the network. The use of E_{BC} results in a new $\mathcal{G}_w(V,E)$ with the interaction strength of each node whose weight corresponds to the sum of weights defined for each e_i such that $\sum_{e_i \in E} E_{BC}(e_i)$ (see Equation 1). The characterisation of this interaction can be addressed through a weighted adjacency matrix $\mathbf{A}$ $(n \times n)$ [17] whose structure also forms part of the formulation given by Kalman [12] for linear time-invariant dynamical systems:

$$\dot{x}(t) = \mathbf{A}x(t) + \mathbf{B}u(t), \qquad x(t_0) \; = \; x_0 \tag{2}$$

where: $\dot{x}(t)$ models the vector $(x_1(t), \ldots, x_n(t))^T$ holding the current state of n nodes at time t; $\mathbf{A}$ showing the topology of the network; and $\mathbf{B}$, an *input* matrix ($n \times m$, $m \leq n$) containing the set of driver nodes under control of a time-dependent input vector $u(t) = (u_1(t), \ldots, u_m(t))$. This vector is responsible for forcing the system to reach a desired configuration state, q, in finite time, t, such that $\dot{x}(t) \; = \; q, \forall \, t \; \in \; [t_0, \, t_n]$ [23]. This also means that if any node of $\dot{x}(t)$ cannot be influenced by $u(t)$, then the system is said to be uncontrollable. However, the computation of Equation 2 can become quite restrictive for large scenarios [2, 4, 6], thus we apply the concept of *structural controllability* given by Lin in [14] and the fulfilment of power dominance initially introduced by Haynes *et al*. in [11].

Conceptually, structural controllability is defined through a digraph $\mathcal{G}(A,B) = (V,E)$ where $V = V_A \cup V_B$ such that V_B includes the driver nodes in B of Equation 2 and $E = E_A \cup E_B$, the set of edges in $\mathcal{G}(V,E)$. To extract the minimum set of driver nodes (denoted here as N_D), we adapt the two observation rules (see Figure 1) simplified by Kneis *et al*. in [13] from the original formulation given in [11]:

OR1 *A vertex in N_D observes itself and all its neighbours*. The result is a new set of nodes based on the dominance problem known as the dominating set (DS).

OR2 *If an observed vertex v of outdegree $d^+ \geq 2$ is adjacent to $d-1$ observed vertices, the remaining un-observed vertex becomes observed as well*. The resulting set, known as the power dominating set (PDS), originates a new concept of dominance where **OR1** is part of the **OR2** such that $DS \subseteq PDS$. Without loss of generality and considering that the PDS problem was initially introduced for observability, we apply it here to its dual problem related to controllability [8].

The application of both rules results in a new N_D, the specifications of which are available in [4] and redesigned in [2] to consider the following four redundancy rules, **RR***x* (see Figure 1):

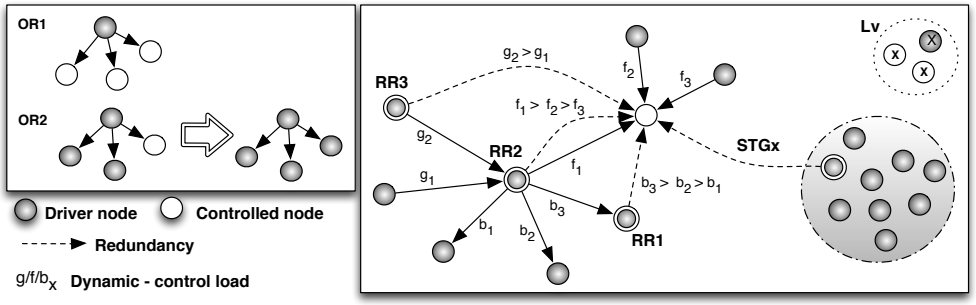

Figure 1: **OR1** - **OR2** and the multiple redundancy strategies, $\mathbf{RR}_x$ [2] and $\mathbf{STG}_x$

[RR1] n_d corresponds to a brother (a node that shares the same father node within its hierarchy with n_d) with the ability to re-link v_i, where n_d represents a driver node $\in N_D$ with the capacity to observe a node v_i in V.

[RR2] n_d is a father driver located at 1-hop, with the ability to reconnect $(n_d, v_i) \in E$.

[RR3] n_d is a grandfather driver located at 2-hops, and with the means to offer redundancy.

[RR4] n_d corresponds to a remote node, located at n-hops, with the capacity to re-link v_i.

The result is a new graph $\mathcal{G}_w^r(V,E')$ of size $n \times n$ and equivalent to $\mathcal{G}_w(V,E)$ but with $|E'| \geq |E|$, and through which it is possible to compute the maximum load capacity that the edges of $\mathcal{G}_w^r(V,E')$ can support [17]:

$$H_{i,j} = (1+\alpha) \times L_{i,j}^0 \tag{3}$$

where α includes a tolerance factor with value $\alpha > 0$ and $L_{i,j}^{t\geq 0} \leq H_{i,j}$, such that $L_{i,j}^{t\geq 0}$ is associated with the load capacity of $\mathcal{G}_w(V,E)$ at the state t. According to Nie *et al.* in [17], this formulation detects the redistribution of shortest paths when one or some edges are removed from network, probably causing an alteration on the initial loads assigned to edges. Therefore, through H it is also possible to map the entire network and visualise the cascading effects of the control dynamics.

It is also necessary to specify the contextual features of the application scenario. Concretely, the network model we apply in our approaches follows topological characteristics of type $y \propto x^{-\alpha}$ such as scale-free and power-law networks [7]. According to Pagani and Aiello in [18], control networks produce similar topological structures to power-law distributions. In these structures, a subset of N_D concentrates the maximum power within the network (e.g. servers, remote terminal units, gateways) and transfers control loads to the rest of components (e.g. sensors, actuators).

As stated in [2], the studies discussed here are also based on the Power-Law out-degree (PLOD) [19] capable of obeying a power-law sequence. Basically, the strategy

randomly assigns an outdegree using $\beta x^{-\alpha}$ to each node and arbitrarily establishes the links taking into account this outdegree. On the other hand, it is assumed that a subset of nodes may be targeted by adversaries with access to an important sub-part of the network. Namely, the threats are limited to δ such that $1 \leq \delta \leq \frac{|V|}{2}$ nodes. Within this weak adversarial model, the goal is to subvert the operative capacities of target nodes or violate their availability by isolating them or removing specific links. In other words, our studies are characterised by combined attacks where adversaries are able, in just one attempt, to produce several types of attacks, like so: (i) isolate a random set of nodes in $\mathcal{G}_w$; (ii) isolate those driver nodes with the maximum degree, i.e. the hubs; (iii) isolate those driver nodes with the highest strength within the network; (iv) arbitrarily remove a few edges from a set of nodes in $\mathcal{G}_w$; (v) remove those edges with the highest edge centrality within the network; and (vi) arbitrarily add a few edges to the network.

3 Local and Remote Redundancy-based Recovery Strategies

Once the context and the adversarial model have been defined, we can specify a set of strategies to find an optimal recovery solution that ensures reconnection from any (local and/or remote) location, at any time and in an acceptable time. Concretely, we specify five strategies which concentrate, through communities, those driver nodes in N_D capable of providing a desirable redundant connectivity (see Figure 1). A community in graph theory corresponds to a community structure in which the nodes of the network can be grouped into sets of nodes such that each set is closely related according to a series of joining policies. In our case, the selection strategies of the driver nodes will be constrained according to the network's type of density, the maximum degree, the minimum diameter and the strength of each driver node. For the sake of clarity, we define five ways to group and select distant driver nodes:

[STG1] This strategy, known as *k-core*, aims to obtain the largest sub-network comprising those driver nodes of degree at least k. To do this, the algorithm recursively checks and removes drivers with degrees lower than k; i.e. $\forall\ n_{d_i} \in N_D$, $(d^+(n_{d_i}) + d^-(n_{d_i})) \geq k$, being d^+ the outdegree of n_{d_i} and d^- the indegree.

[STG2] It identifies those driver nodes with the maximum capacity to inject control, i.e. those driver nodes with the ability to transport control from one peer to another in the network. Technically speaking, the procedure is based on the *s-core* structure (structural property of a tournament digraph [20]), holding the largest sub-network with those driver nodes with strength at least s.

An easy way to apply the technique in our context it is to use the maximum value contained in E_{BC} as a reference. That is, $\forall\ n_{d_i} \in N_D$, $\sum_{e_i \in E} E_{BC}(e_i)) \geq s$ that corresponds to the sum of weights related to E_{BC} of each outdegree of n_{d_i}.

[STG3] This strategy aims to extract those driver nodes contained in dense clusters, computed though the edge betweenness method introduced by *Girvan and Newman* in [10]. The method basically segregates the network into modules, where

the process is based on: (i) calculating the centrality of all edges in the network (i.e. the score in E_{BC}), (ii) removing the edge with the highest betweenness, and (iii) repeating the process until no edges remain to be controlled.

[STG4] Unlike **STG3**, this strategy selects those drivers located at the intersection points between communities and whose nodes have the highest centrality by which all the shortest-paths pass through them. Therefore, the selection procedure is equivalent to **SGT3**, but studying the border nodes.

[STG5] This strategy identifies those driver nodes with the minimum diameter in the network (as proved in [2]). As the network graph follows a digraph structure with cycles, the *breadth-first search* (BFS) method is applied to compute the minimum diameter for each n_{d_i} in $\mathcal{G}_w(V,E)$.

The result of these selection strategies is a new set of driver nodes, N_D^{stg}, with the capacity to provide remote assistance in adverse situations. In this way, the control is no longer only dependent on the local links as stated in [2] but also on distant edges.

3.1 Initial Conditions: Control and Redundancy

Before formalising the five approaches mentioned above, it is also necessary to define a set of initial conditions related to the control restoration and its associated structural controllability. That is, the underlying infrastructure of the control follows a power-law distribution, in which a subset of nodes, the hubs, contains the highest degree value (d^+ and d^-) in $\mathcal{G}_w(V,E)$. Any topological change and its degree may seriously impact on the power-law value, decomposing the general structure of the network and the monitoring services as discussed in [1]. This also means that resilience mechanisms adopted for the protection must also respect the degree conditions before and after their application and the dynamism of the network in terms of mobility (e.g. handheld interfaces, robots). One way of handling this dynamism would be through two fundamental sets Lv and Jn. The former representing those nodes that (temporarily or definitively) decide to leave the system, and the latter, those nodes that wish to enter the system as new members.

Likewise, it is also essential to fulfill the two observation rules (**OR1** and **OR2**) defined in Section 2. Any violation of **OR1** and/or **OR2** may entail an imbalance in the control processes: "*nodes that were controlled by a specific n_{d_i} in a time t will no longer be controlled at t+1, meaning that control signals are completely lost*". This also means that the inclusion of new links from promising remote drivers should not corrupt the normal behaviour of the rules. In a nutshell, let a node v_i be re-linked by a remote driver; N_D^{rmt} to the set of children driver nodes of a $n_{d_i} \in N_D^{stg}$; and O^{rmt} the set of observed nodes children of a $n_{d_i} \in N_D^{stg}$ such that $O^{rmt} \leftarrow O^{rmt} \setminus N^{rmt}$, the following two principles are fundamental to respect the conditions given by **OR2**:

[P1] If v_i is a driver node included in N_D, ($—N_D^{rmt}— \geq 0$ and $—O^{rmt}— \geq 2$) or ($—N_D^{rmt}— \geq 1$ and $—O^{rmt}— = 0$) or ($—N_D^{rmt}— = 0$ and ($—O^{rmt}— = 0$).

[P2] If v_i is not a driver node in N_D, then ($—N_D^{rmt}— \geq 0$ and $—O^{rmt}— \geq 1$) or ($—N_D^{rmt}— = 0$ and ($—O^{rmt}— = 0$).

To comply with the three redundancy principles described in [2] (and related to the fulfillment of **OR1** and **OR2**, and considering the new redundant links), **P1** and **P2** must be applied in the commissioning phase of the network. In this phase, all nodes have to be supported by at least a local driver located in the vicinity (a brother, a father or a grandfather n_{d_i} [2]) and by a distant driver in N_D^{stg}, such that $N_D^{stg} \leq N_D$. Both drivers additionally have to comply with the condition of strength so as to ensure that the greater part of the control load passes through them. Therefore, the initial conditions to be considered in the remainder of this paper are as follows:

[C1] Redundant mechanisms must not perturb the power-law nature of the underlying infrastructure. In addition, so as to ensure these resilience measures and at all times, the driver nodes in N_D^{stg} should not include the hubs of the system as well as those included in Lv. Thus, if hubs are targeted through denial of service attacks, the extra measures will still remain active and reachable from another strategic point of the system.

[C2] Make sure that redundant links are established at those driver nodes containing the highest strength of N_D^{stg}.

[C3] **OR1** and **OR2** must remain active at all times. For this to be guaranteed, it is vital to comply with all the principles of redundancy described both in this paper and in [2], where the control and the redundant links reside in $\mathcal{G}_w^r(V,E')$. This means that any structural variation in $\mathcal{G}_w^r(V,E')$ directly impacts on the two observation rules.

3.2 Five Recovery Approaches: Remote and Local Structural Controllability

Algorithm 3.1 compiles the five strategies defined above, the approaches of which provide a new set of driver nodes with the most promising nodes for the reconnection to each v_i in V. At this point and regardless of the modus operandi and the functionality of each strategy $\mathbf{STG}_x, x = \{1,2,3,4,5\}$, the core of each approach is centred on selecting first those driver nodes in DS (obtained from **OR1** defined in [4]), where the selection is constrained to the highest strength in $\mathcal{G}_w(V,E)$ (known as the *score*, s), such that $\sum_{e_i \in E} E_{BC}(e_i) \geq s$ (**C2**). To make sure that condition **C1** is reached, the system must also extract, from the largest subset N_D^{stg}, those drivers with the highest degree, discarding from N_D^{stg} those nodes that are no longer part of the network and are not part of the hubs.

Algorithm 3.1: COMMUNITIES($\mathcal{G}_w(V,E), E_{BC}, Lv, STG, hubs, k, s, RR_x$)

local $N_D^{stg} \leftarrow \emptyset, DS \leftarrow \emptyset; N_D \leftarrow \emptyset;$
output (N_D)

$DS \leftarrow \mathbf{OR1}_{v1}(\mathcal{G}_w(V,E))$; **comment:** Algorithm specified in [4];
if $STG = 1$
  then
$\{N_D^{stg} \leftarrow$ K-CORE$(\mathcal{G}_w(V,E), DS, k, E_{BC})$;

else
- **if** $STG > 1$
  - **then**
  - **if** $STG = 2$
    - **then**
    - $\{N_D^{stg} \leftarrow$ S-CORE$(\mathcal{G}_w(V,E), DS, s, E_{BC})$;
    - **else**
    - **if** $STG = 3$
      - **then**
      - $\{N_D^{stg} \leftarrow$ DENSECOMMUNITY$(\mathcal{G}_w(V,E), DS, E_{BC})$;
      - **else**
      - **if** $STG = 4$
        - **then**
        - $\{N_D^{stg} \leftarrow$ NON-DENSECOMMUNITY$(\mathcal{G}_w(V,E), DS, E_{BC})$;
        - **else**
        - $\{N_D^{stg} \leftarrow$ MINIMUMDIAMETER$(\mathcal{G}_w(V,E), DS, E_{BC})$;
  - $N_D^{stg} \leftarrow$ MAXDEGREE$(\mathcal{G}_w(V,E), N_D^{stg})$;

$N_D^{stg} \leftarrow (N_D^{stg} \setminus Lv) \setminus hubs$;
$\{\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS\} \leftarrow \mathbf{OR1}_{v3}(\mathcal{G}_w(V,E), RR_x, DS, N_D^{stg}, Lv, E_{BC})$;
$\{\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), N_D\} \leftarrow \mathbf{OR2}_{v2}(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS)$; **comment:** Algorithm specified in [2];
return (N_D)

However, obtaining N_D^{stg} is not enough to ensure a robust control in the face of unexpected events. It is also necessary to ascertain that the redundancy principles properly fit the application context. For example, a candidate in N_D^{stg} breaches one of the basic principles of structural controllability, because $\mathcal{G}_w^r(V,E')$ has not been modelled according to the new conditions of the context and the implicit conditions in **C3**. For this reason, in Algorithm 3.2 we define a new version of **OR1** (note that previous releases are defined in [2] and [4]), in which those promising driver nodes from N_D^{stg} that are able to restore a critical situation from any location of the system are identified. This restorative skill is also reflected in the updating of $\mathcal{G}_w^r(V,E')$, responsible for maintaining the entire topological map of the control system visible and respecting the resilience measures defined in **C3**. Concretely, this feature is contemplated in Algorithm 3.3, which includes the fulfilment of **P1** and **P2**, and returns $\mathcal{G}_w^r(V,E')$ with the new pathways. Launching Algorithm 3.2 results in a new subset DS, whose components are essential to produce the final subset N_D and the definitive matrix $\mathcal{G}_w^r(V,E')$ by executing **OR2** defined in [2]. Note that with N_D, the commissioning phase and the initial start-up are solved.

Algorithm 3.2: $\mathbf{OR1}_{v3}(\mathcal{G}_w(V,E), RR_x, DS, N_D^{stg}, Lv, E_{BC})$

local $relink \leftarrow \emptyset, N \leftarrow V; \mathcal{G}_w^r(V,E') \leftarrow \mathcal{G}_w(V,E)$;
output $(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS)$

while $(N \neq \emptyset)$
do
- *Randomly choose one* $v_i \in N$;
- $\{\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E')\} \leftarrow$ **LOCALREDUNDANCY**$(RR_x, \mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS, v_i, Lv)$; **comment:** as stated in [2];
- $\{\mathcal{G}_w^r(V,E'), relink\} \leftarrow$ **REMOTEREDUNDANCY**$(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS, Lv, v_i, N_D^{stg}, E_{BC})$;
- **if** $v_i \in DS$
  then
  $N \leftarrow N \setminus \{v_i\}$;
  else
  - **if** $relink = \emptyset$
    then
    $DS \leftarrow DS \cup \{v_i\}$;
    else
    $N \leftarrow N \setminus \{v_i\}$;

return $(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS)$

Algorithm 3.3: **REMOTEREDUNDANCY**$(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS, Lv, v_i, N_D^{stg}, E_{BC})$

local $relink \leftarrow \emptyset, N_D^{rmt} \leftarrow \emptyset, O^{rmt} \leftarrow \emptyset, maxE_{BC} \leftarrow \emptyset, promisingN_D \leftarrow \emptyset$;
output $(\mathcal{G}_w^r(V,E'), relink)$

$maxE_{BC} \leftarrow$ **MAXSTRENGTH**$(\mathcal{G}_w(V,E), E_{BC}, N_D^{stg}, v_i)$;
while $(N_D^{stg} \neq \emptyset)$
do
- *Randomly choose one* $d_i \in N_D^{stg}$;
- **if** $E_{BC}(d_i, N_D^{stg}) \geq maxE_{BC}$
  then
  $N_D^{rmt} \leftarrow$ **CHILDREN**$(\mathcal{G}_w^r(V,E'), d_i) \cap DS$;
  $O^{rmt} \leftarrow$ **CHILDREN**$(\mathcal{G}_w^r(V,E'), d_i) \setminus N_D^{rmt}$;
  - **if** $(v_i \in DS)$ **and** $((|N_D^{rmt}| \geq 0$ **and** $|O^{rmt}| \geq 2)$ **or** $(|N_D^{rmt}| \geq 1$ **and** $|O^{rmt}| = 0)$ **or** $(|N_D^{rmt}| = 0$ **and** $(|O^{rmt}| = 0))$
    then
    $promisingN_D \leftarrow promisingN_D \cup \{d_i\}$;
  - **else**
    - **if** $(v_i \notin DS)$ **and** $((|N_D^{rmt}| \geq 0$ **and** $|O^{rmt}| \geq 1) or (|N_D^{rmt}| = 0$ **and** $(|O^{rmt}| = 0))$
      then
      $\{promisingN_D \leftarrow promisingN_D \cup \{d_i\}$;

if $promisingN_D \neq \emptyset$
then
- *Randomly choose one* $d_i \in promisingN_D$;
- $\mathcal{G}_w^r(V,E')\{d_i, v_i\} \leftarrow 1$;
- $E' \leftarrow E' \cup \{(d_i, v_i)\}$;
- $relink \leftarrow relink \cup \{d_i\}$;

return $(\mathcal{G}_w^r(V,E'), relink)$

As for the restoration mechanism in a time t after bootstrapping, we consider the following three states, initially defined in [2] but adapted to this paper:

- *optimal state*, when the re-link mechanisms are successfully launched;
- *suboptimal state*, there is no suitable driver node for the re-link and the system identifies the driver with the minimum diameter to re-connect an unobserved node complying with **C1**, **C2** and **C3**; and
- *non-optimal state*, there is no suitable driver node for the re-link and the system is not able to find a driver with the minimum diameter complying with **C1**, **C2** and **C3**. Thus, this state resolves the problem by including the unobserved node in N_D, and verifying the fulfillment of the observation rules to update the new states of $\mathcal{G}_w^r(V,E')$.

This way of extending the resilience mechanisms improves upon the approaches proposed in [2] so that they can incorporate new and diverse local and remote resources. More specifically, Algorithm 3.4 extends our current research goals to not only comprise the local redundant mechanisms but also distant ones, where the system can also require checking for the existence of new candidates in non-optimal states. Note that E_a in Algorithm 3.4 represents the set of active edges in $\mathcal{G}_w(V,E)$ with coverage to the rest of nodes in V such that $E_a \subseteq E'$.

Algorithm 3.4: DYNAMIC RECOVERY$_{v2}(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), N_D, Lv, E_a, RR_x)$

local $nonOptimal \leftarrow false$;
output $(\mathcal{G}_w^r(V,E'), N_D)$

for $v_i \leftarrow 1$ **to** $\mid V \mid$
  do
  $\{\mathcal{G}_w^r(V,E'), N_D, nonOptimal\} \leftarrow$ **DYNAMIC RECOVERY**$_{v1}(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), N_D, Lv, E_a, RR_x)$;
  comment: Dynamic Recovery$_{v1}$ is specified in [2];
  if *nonOptimal*
    then
    $E_{BC} \leftarrow$ **NEWE**$_{BC}(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), E_{BC})$;
    $\{\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E')\} \leftarrow$ **LOCALREDUNDANCY**$(RR_x, \mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), DS, v_i, Lv)$; **comment:** as stated in [2];
    $\{\mathcal{G}_w^r(V,E'), relink\} \leftarrow$ **REMOTEREDUNDANCY**$(\mathcal{G}_w(V,E), \mathcal{G}_w^r(V,E'), N_D, Lv, v_i, (N_D^{stg} \setminus Lv), E_{BC})$;
    if $relink == \oslash$
      then
      $\{N_D \leftarrow N_D \cup \{v_i\}$;
return $(\mathcal{G}_w^r(V,E'), N_D)$

Both the functional and spatial complexity (i.e. $\mid N_D \mid$) will depend on the frequency that the system enters through the different states: optimal, sub-optimal or non-optimal. If the system is able to find a predefined n_{d_i} in $\mathcal{G}_w^r(V,E')$, the cost of connection will be $O(n)$. Otherwise, the cost will depend on the capacity of the system to identify: (i) a driver node with the minimum diameter complying with **C1**, **C2**, and **C3**; or (ii) a suitable set of drivers for a wide coverage in $\mathcal{G}_w$. Assuming that $nd \approx n$ in the worst case, and $\mid V \mid = n$, $\mid E \mid = e$, $\mid N_D \mid = nd$, the cost of computing Algorithm 3.4 in the worst case is: $O(n^2 log(n))$ for the suboptimal case, and $O(n^3)$ both for the non-optimal solution and for the processes developed in the commissioning phase. Therefore, all studies are analogous to those discussed in [2], but here taking into account the complexity for Algorithm 3.3 is $O(n^2) \subseteq O(n^3)$. As regards spatial costs, the deviation of $\mid N_D \mid$ will depend on the non-optimal states, in which the number of driver nodes increases, at least, by one unit.

To validate the approaches **STG***x* and Algorithm 3.4, in the remainder of this paper we analyse two specific case studies. The former verifies the level of optimisation of the solutions proposed, whereas the latter explores the effectiveness and convenience of the approaches when a significant number of external links appear in the network.

4 Experimentation: Results and Discussions

The experiments have been modelled using the distribution PLOD [19] with cycles and with a connectivity range $\alpha = 0.1$, so as to illustrate more realistic scenarios where

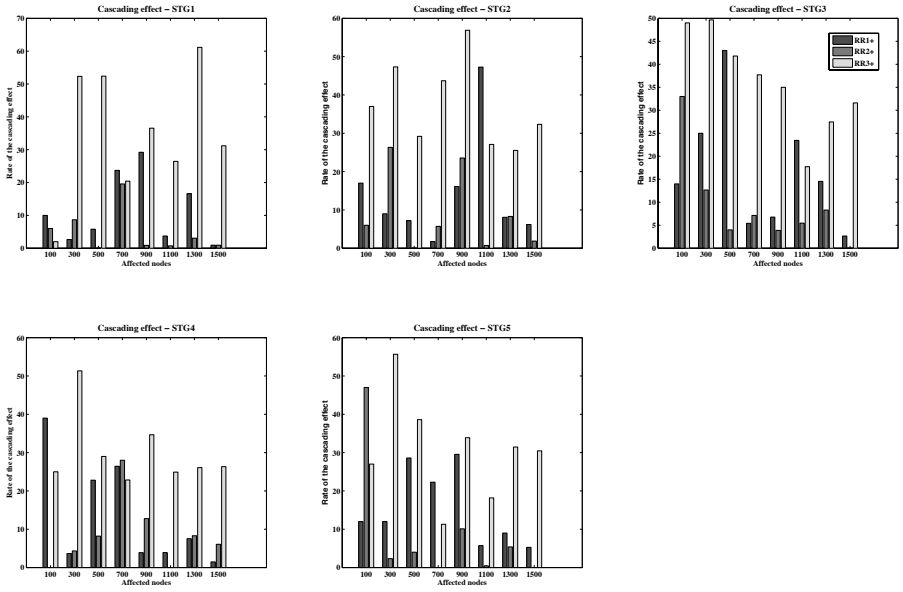

Figure 2: Cascading effect of the approaches: **RR1+**, **RR2+** and **RR3+**

determined nodes serve as hubs. With this type of distribution, we have experimented on Matlab networks in the order of $\simeq$ 100-500 nodes, $\simeq$ 500-1000 nodes and $\simeq$ 1000-1500 nodes, and have planned variable disturbances targeting a random ($\delta \leq V/2$) number of nodes.

4.1 Case Study 1: RR1+, RR2+ and RR3+

The degradation of the networks can be observed in both Figure 2 and Figure 3, where deterioration rates reach minimal values, between 0-60%. Specifically, Figure 2 illustrates the cascading effect after the perturbations in which the disintegration degree of the shortest paths exceeds the minimal values (i.e. more than 20% in all the strategies) and whose value is computed according to the number of nodes that surpass the maximum capacity of the network. As mentioned and in [17], H in Equation 3 can be considered as a suitable indicator to detect variations on the control dynamics and their cascading due to the redistribution of shortest paths.

From this study, we also observe that a duplication of links through **RR3** guarantees less degradation of the network than a network dependent only on father drivers. To the contrary, Figure 3 shows the global efficiency of the entire network before and after a perturbation. Global efficiency can be defined as the average inverse shortest path length in $\mathcal{G}_w(V,E)$ and is inversely related to the characteristic path length [9]. In other words, it corresponds to the capacity of the system to transport control loads from one point to another, and from a global standpoint. Its value is computed as follows:

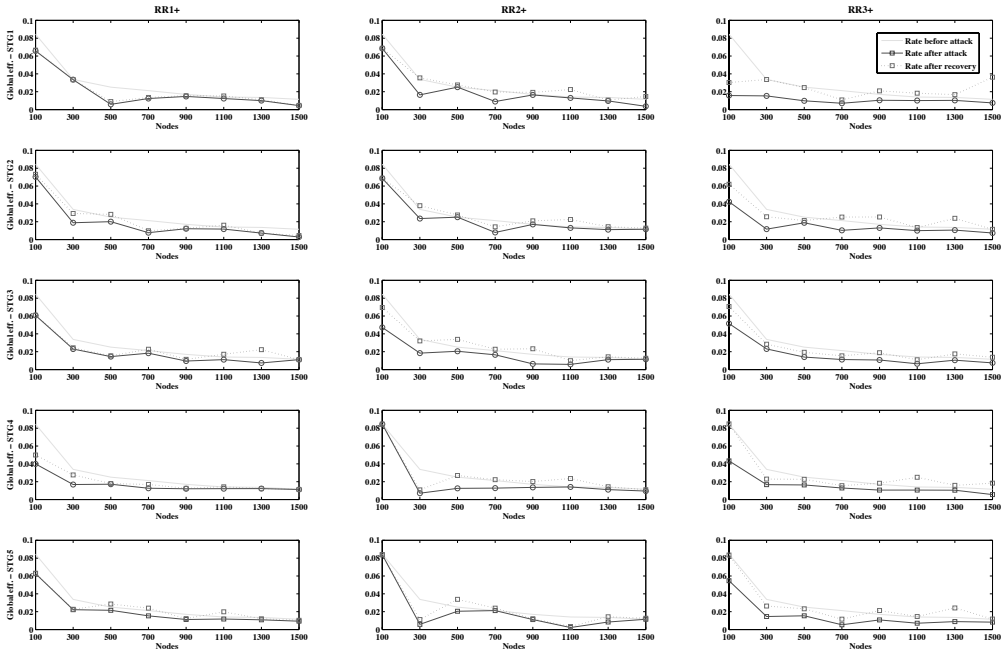

Figure 3: Global efficiency of the approaches: **RR1+**, **RR2+** and **RR3+**

$$E_g = 1|V|(|V|-1) \sum_{v_i,v_j \in V, v_i \neq v_j} 1 d_{v_i,v_j} \tag{4}$$

where d_{v_i,v_j} is the shortest path length between a node v_i and v_j in V.

On the other hand, the representations given in Figure 2 and Figure 3 are based on the strategical combination of a set of redundant pathways, **RR**x+. With **RR**x, we show the local cases specified in [2] such that $x = \{1,2,3\}$ (see Section 2), and with '+', the remote cases of the type **STG**x such that $x = \{1,2,3,4,5\}$ (see Section 3). From Figure 3, it is also possible to discern the devastating effect of the perturbations, in which a random and/or a selective set of nodes, such as the hubs or the nodes with the highest strength within the network, are the principal targets (see Section 2). In these cases, the number of shortest paths and the diameters of the network tend to vary depending on the threat and the decomposition of the network. The restoration of these scenarios and the optimisation of the different **STG**x are depicted in Figure 4, showing the individual efficacy of each approach. From this figure, we note that the most optimised cases reside in those approaches with local dependence on the brother drivers. Namely, the system becomes more efficient and robust when the redundant measures trust in a brother driver; whilst solutions dependent on father or grandfather drivers are in general less effective for critical environments; a behaviour that was also detected in [2].

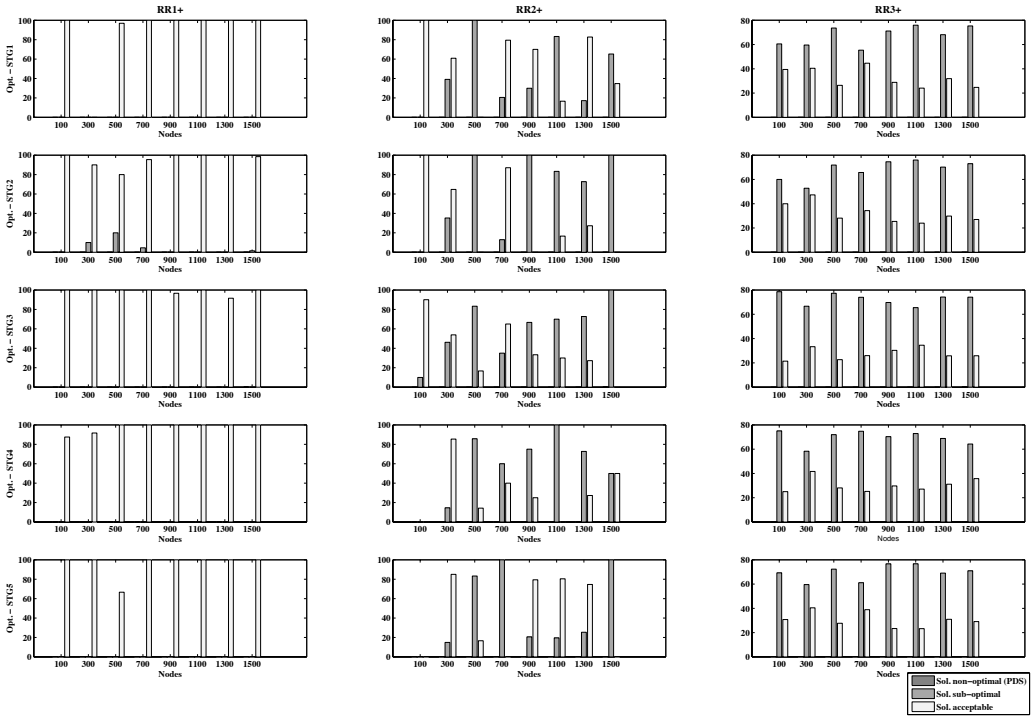

Figure 4: Optimisation of the approaches: **RR1+**, **RR2+** and **RR3+**

Moreover, despite the fact that the five approaches proposed seem to be technically suitable for combined solutions of the type **RR**1+, the approach based on s-core (**STG2**) is not as optimal as expected. For small distributions, the approach tends to enter via suboptimal solutions, regardless of whether or not our study amplifies the searching range of promising n_{d_i} in N_D^{stg} (in the commissioning phase). Namely, our simulations in **STG2** are arranged for $s = s/2$ (equivalent for k-core with $k = k/2$) so as to extend the search range, of at least, a moderate subset of remote links during the simulations. In relation to the optimisation, Figure 5, shows the spatial complexity, in terms of the number of driver nodes in each simulation and whose increase is due to the implications associated with non-optimal states. However, and surprisingly, the devastating effect of the perturbations does not cause a greater variation of N_D in any **STG**x, since, in most cases, they enter via the suboptimal option.

4.2 Case Study 2: RR12+, RR13+ and RR23+

Thus far, we know that the best option is to offer self-healing services based on local dependencies with direct links in brother drivers. Now, we need to determine if this dependency also occurs in those complex contexts in which there is more than one local redundant link in each v_i in V. We have therefore simulated the cases: **RR12+**, **RR13+**

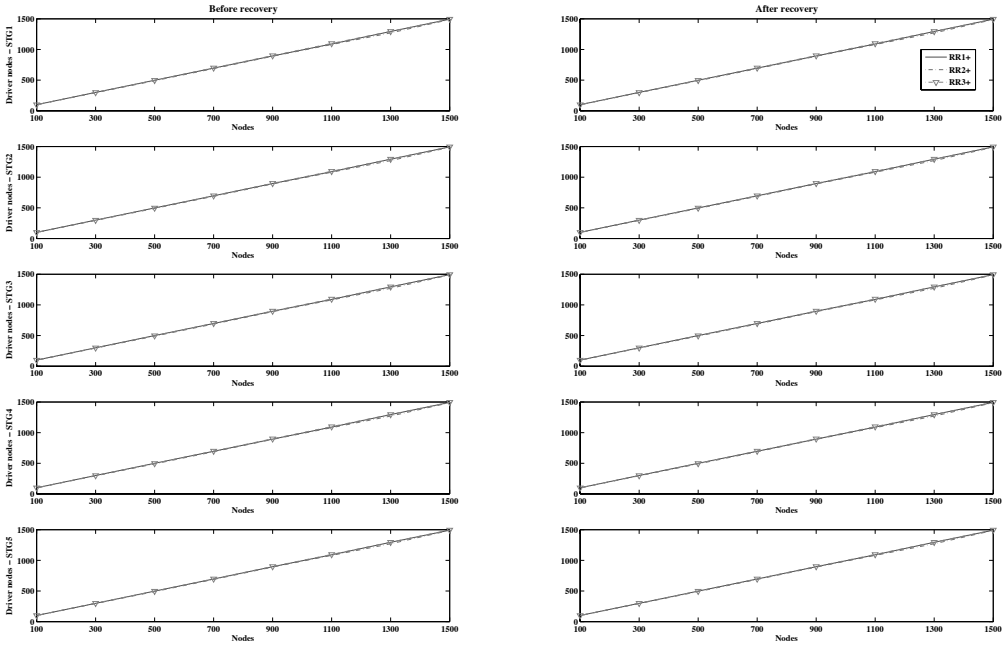

Figure 5: Spatial complexity of the approaches: **RR1+**, **RR2+** and **RR3+**

and **RR23+**, whose results are provided in Figure 6. The figure indicates that an increase of local links (brother-father, brother-grandfather, father-grandfather) does not help achieve the restoration processes in linear times, and a simple local dependence on a brother driver and on a remote driver is enough to take over adverse situations in optimal times. This feature further presumes a low cost of installation and maintenance as the redundancy rate for each device is limited to two. Moreover, **STG2** continues to be susceptible to threats, probably from targeted attacks that aim to isolate the nodes with the highest strength. Note that if, in addition, these nodes correspond to a promising driver in N_D^{stg}, then these can no longer offer their reconnection services, pushing the system to pass over the suboptimal or non-optimal cases.

5 Conclusion

Technological convergence in the industry of control is resulting in complex and dynamic systems, where the structural controllability is distributed across all their devices. This feature makes the underlying infrastructures and their topologies vulnerable to unplanned perturbations, thereby demanding efficient restoration mechanisms capable of working at linear times, as discussed in [2]. Unfortunately, the work done in [2] principally reconnects the most affected parts from a local perspective without looking beyond distant reconnection strategies. So we have extended the approach to study five particular remote reachability-based strategies, which deal with locating

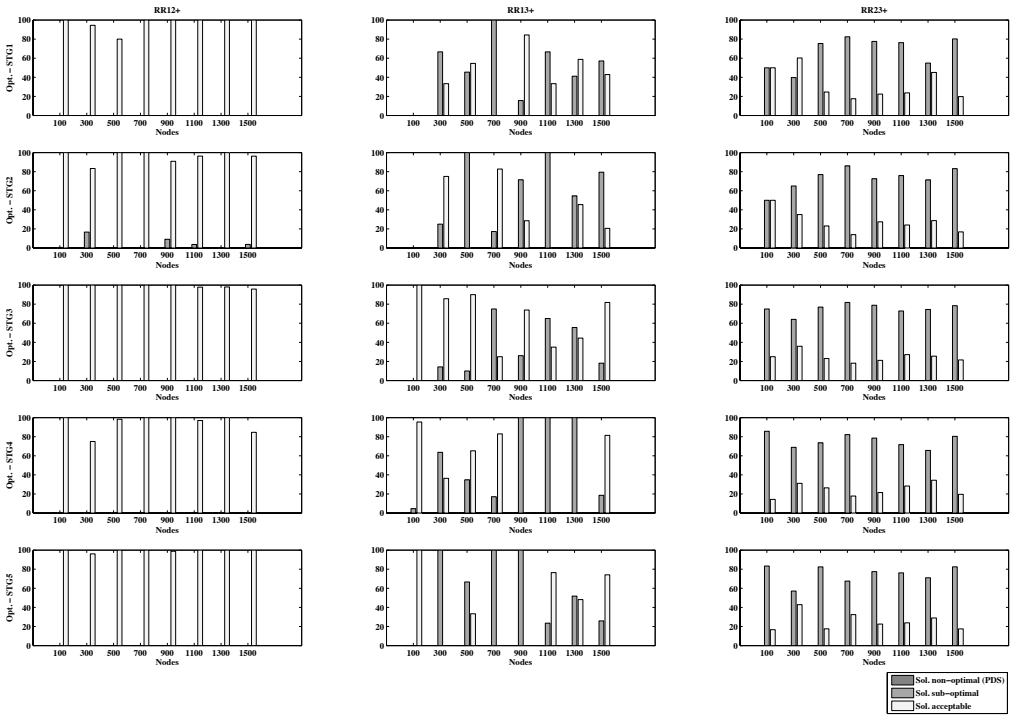

Figure 6: Optimisation of the approaches: **RR12+**, **RR13+** and **RR23+**

those drivers with the highest degree, strength and control capacity, and with the ability to address restore processes from anywhere, anyhow, and at any time. For these processes to be effective, a new version of **OR1** has also been designed to comprise all the reconnection cases, including both the local and remote redundant pathways.

Further to this, we have provided two case studies. The first analyses the effectiveness of the five approaches combined with the ones developed in [2], and the second analyses the ideal redundancy combination, taking into account: the optimal states, and the costs of installation and maintenance. The results indicate that a dependence on a brother driver and on a remote node is enough to reach desirable states in linear times. With these findings, we sucessfully close our initial research goals as well as the future work given in [2], and in the future, expect to probe other optimal restoration mechanisms without support in redundant resources.

6 Acknowledgment

This work has been partially supported by the research projects PERSIST (TIN2013-41739-R) and SADCIP (RTC-2016-4847-8), both financed by the Ministerio de Economía y Competitividad, as well as by the project NECS (H2020-MSCA-ITN-2015) financed by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 675320.

References

[1] C. Alcaraz and J. Lopez. Wide-area situational awareness for critical infrastructure protection. volume 46, pages 30–37. IEEE Computer Society, 2013 2013.

[2] C. Alcaraz and J. Lopez. Safeguarding structural controllability in cyber-physical control systems. In *The 21st European Symposium on Research in Computer Security (ESORICS 2016)*, volume 9879, pages 471–489, Crete, Greece, 2016. Springer, Springer.

[3] C. Alcaraz and J. Lopez. Secure interoperability in cyber-physical systems. In *Security Solutions and Applied Cryptography in Smart Grid Communications, IGI Global, USA*, pages 139–159, USA, 2017. IGI Global, IGI Global.

[4] C. Alcaraz, E. Etcheves Miciolino, and S. Wolthusen. Structural controllability of networks for non-interactive adversarial vertex removal. In *8th International Conference on Critical Information Infrastructures Security*, volume 8328, pages 120–132. Springer, 2013.

[5] C. Alcaraz, R. Roman, P. Najera, and J. Lopez. Security of industrial sensor network-based remote substations in the context of the internet of things. volume 11, page 1091–1104. Elsevier, 2013 2013.

[6] C. Alcaraz and S. Wolthusen. Recovery of structural controllability for control systems. In *Eighth IFIP WG 11.10 International Conference on Critical Infrastructure*, volume 441, pages 47–63. Springer, 2014.

[7] A.-László Barabási. Network science the scale-free property. Cambridge University Press, 2016.

[8] Noah J. Cowan, Erick J. Chastain, Daril A. Vilhena, James S. Freudenberg, and Carl T. Bergstrom. Nodal Dynamics, Not Degree Distributions, Determine the Structural Controllability of Complex Networks. *PLoS ONE*, 7(6):e38398+, June 2012.

[9] Bryan Ek, Caitlin VerSchneider, and Darren A. Narayan. Global efficiency of graphs. *International Journal of Graphs and Combinatorics*, 12(1):1 – 13, 2015.

[10] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, 99:7821–7826, 2002.

[11] T. Haynes, S. M. Hedetniemi, S. T. Hedetniemi, and M. A. Henning. Domination in graphs applied to electric power networks. *SIAM Journal on Discrete Mathematics*, 15(4):519–529, 2002.

[12] R. E. Kalman. Mathematical description of linear dynamical systems. *Journal of the Society of Industrial and Applied Mathematics Control Series A*, 1:152–192, 1963.

[13] J. Kneis, D. Mölle, S. R., and P. Rossmanith. Parameterized power domination complexity. *Information Processing Letters*, 98(4):145–149, 2006.

[14] C.-T. Lin. Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208, 1974.

[15] M. Marchese and M. Mongelli. Simple protocol enhancements of rapid spanning tree protocol over ring topologies. *Computer Network*, 56(4):1131–1151, 2012.

[16] K. Nakayama, N. Shinomiya, and H. Watanabe. An autonomous distributed control method for link failure based on tie-set graph theory. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 59(11):2727–2737, 2012.

[17] S. Nie, X. Wang, H. Zhang, Q. Li, and B. Wang. Robustness of controllability for networks based on edge-attack. *PLoS ONE*, 9(2):1–8, 2014.

[18] G. A. Pagani and M. Aiello. The power grid as a complex network: A survey. *Physica A: Statistical Mechanics and its Applications*, 392(11):2688–2700, 2013.

[19] C. Palmer and J. Steffan. Generating network topologies that obey power laws. In *Global Telecommunications Conference (GLOBECOM '00)*, volume 1, pages 434–438, 2000.

[20] S. Pirzada and U. Samee. Mark sequences in digraphs. In *Snaire Lotharingien de Combinatoire, 55, B55c*, pages 1–13, 2006.

[21] W. Quattrociocchi, G. Caldarelli, and A. Scala. Self-healing networks: Redundancy and structure. *PLoS ONE*, 9(2):e87986, 2014.

[22] A. Sadeghi, C. Wachsmann, and M. Waidner. Security and privacy challenges in industrial internet of things. In *Proceedings of the 52Nd Annual Design Automation Conference*, DAC '15, pages 54:1–54:6, New York, NY, USA, 2015. ACM.

[23] B Southall, B F Buxton, J A Marchant, B Southall , B F Buxton and J A Marchant Controllability and observability: Tools for kalman filter design, 1998.

[24] M. Waidner and M. Kasper. Security in industrie 4.0 - challenges and solutions for the fourth industrial revolution. In *2016 Design, Automation & Test in Europe Conference & Exhibition, DATE 2016, Dresden, Germany, March 14-18, 2016*, pages 1303–1308, 2016.

[25] B. Wang, L. Gao, Y. Gao, and Y. Deng. Maintain the structural controllability under malicious attacks on directed networks. *EPL (Europhysics Letters)*, 101(5):58003, 2013.