Resilience to cascading failures: a complex network approach for analysing the Croatian power grid

Zlatan Sićanica\(^1\) and Ivan Vujaklija\(^2\)

\(^1\)Končar – Power Plant and Electric Traction Engineering Inc. Zagreb, Croatia (zlatan.sicanica@koncar-ket.hr)
\(^2\)Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia (ivan.vujaklija@fer.hr)

Abstract – A transmission power grid system is arguably the single most important utility of any nation since it provides a basic infrastructure for a wide variety of interconnected and interdependent systems of vital importance. Therefore, its resilience to intentional attacks or random failures is of critical importance. This is especially true with respect to cascading failures which can be initiated by seemingly minor events but can lead to economic losses worth billions of dollars, as recent examples show. In order to gain insight into the resilience of the Croatian transmission power grid we investigated its resilience to three different types of failures, including both intentional attacks and random failures. Our analyses are based on the complex network approach. Extensive numerical simulations were carried out based on three different discrete-time cascading failure models. The first two models were described previously, whereas we substantially modified the third model.

Keywords - cascading failure, network efficiency, transmission power grid resilience, complex network analysis

I. INTRODUCTION

A transmission power grid comprises a system of high-voltage transmission lines which connect power generating stations and distribution substations. Since electrical energy provides a basic infrastructure for a wide variety of interconnected and interdependent systems of vital importance (such as transportation and communication infrastructure, heating, gas and water supply to name just a few) power grids are arguably the most important infrastructural resource of any country. Power grid systems of developed countries are highly resilient and most power outages are of local character and short duration. However, although rare, large scale blackouts can cause tremendous economic damage and as such it is of vital importance to understand their dynamics.

One of the early examples of a major blackout is the New York, July 1977 blackout which lasted for hours and caused economic damages estimated to be well over $1 billion in today's money. Some of the more recent wide scale blackouts include the August 2003 blackout which affected large parts of Northeastern and Midwestern United States as well as Ontario (with an estimated cost of over $10 billion), the September 2003 Italian blackout (over $1 billion), the July 2012 Indian blackout ($6 billion) and the March 2015 Turkish blackout ($700 million) \(^1\). The key factor in such major blackouts are cascading failures which are characterized by initial failure which causes redistribution of power flow which triggers further overloads and failures which spread through the entire network or large parts of the network \(^1\). The causes of initial (trigger) failures vary from component failure (e.g. sagging power lines, Italy 2003 blackout), human error (e.g. US & Canada 2003 blackout), lightning strikes (e.g. New York 1977 blackout), to system failures (e.g. Indian 2012 blackout). Whatever the cause, it is clear that such large scale blackouts have tremendous economic impact.

During the past two decades numerous studies have analysed power grid systems from the complex network perspective (\([2], [3], [8]\)). Among others, national power grid systems of US, Italy, Spain, France, UK, Germany, South Korea, Iran, India and China have been analysed using the complex network approach (\([9], [2], [7], [10], [11], [3], [12], [13], [14]\)). Although modelling a highly complex power grid system as a simple undirected/directed, weighted/unweighted network, without taking into account their electrical and physical properties, is obviously a major simplification and as such has major limitations (\([4], [5]\)), the simplicity of the network approach on the other hand makes the study of their complex dynamics (e.g. cascading failures) more tractable. Indeed, many highly complex systems have been successfully modelled as networks. Examples include metabolic networks, biochemical networks, protein interaction networks, ecological networks, transportation networks, internet, social networks, spread of epidemics and genetic regulatory networks (\([5], [3]\)).

Two modelling approaches based on complex networks have been investigated in the literature: static analysis and dynamic analysis \([11]\). Static analysis is a simple approach where failure of a node has no influence on neighbouring nodes. Namely, nodes are removed one by one (either according to random failure or intentional attack model) and after each removal, the network efficiency and size of the giant component \([5]\) are computed. A major limitation of the static models however is their inherent inability to model cascading failures. In dynamic analysis on the other hand only the initial node (or edge) is removed. After the initial removal, the system is left to behave according to its own dynamics. The goal of dynamic models is to capture the more realistic behaviour where the initial failure of a single node or edge causes a dynamic redistribution of loads throughout the network. This is in line with the more realistic cascading type of failures \([7]\).
In this study we analysed the resilience of the Croatian Power Grid (CPG for short) to cascading failures based on the complex network analysis approach. Both random failures and targeted attacks on particular nodes were investigated using numerical simulations based on three different discrete-time dynamical models, two of which have been described previously ([7], [15]). The third model is based on [16] and [17], however we introduce an important modification.

II. METHODS AND RESULTS

In the complex network analysis approach the power grid is modelled as a graph \( G = (N, E) \). Mathematically, a graph consist of a finite set of nodes \( N \) (usually called vertices in the mathematical literature) and a finite set of edges \( E \), where each edge connects two nodes. In the power grid context, generators and distribution substations are represented by nodes and transmission lines by edges ([1, 5, 7]). Edges can be weighted or unweighted and directed or undirected depending on the particular model. We follow the same approach as in [7] and distinguish two types of nodes: generator nodes and distribution nodes. Generator nodes are sources of electrical energy whereas distribution nodes distribute electrical energy to end consumers and are thus “sinks of energy”. According to this model there are 75 generator nodes and 190 distribution nodes in the CPG, which are shown by red and blue circles respectively (Fig.1).

![Fig. 1. The model of the Croatian Power Grid (CPG). Generators are represented by red circles and distribution stations by blue circles.](image)

Importantly, we assume that electricity is transmitted from all generators to all distribution nodes along the most efficient paths. The efficiency of the path between generator node \( i \) and distribution node \( j \), \( \epsilon_{ij} \), is defined simply as the inverse of their (shortest) distance, i.e. \( \epsilon_{ij} = 1/d_{ij} \). Note that efficiency \( \epsilon_{ij} \) and hence the distance \( d_{ij} \) will depend on the particular model being used. We measure the resilience of a network \( \mathcal{N} \) by the network efficiency \( E(\mathcal{N}) \), defined as in [7]:

\[
E(\mathcal{N}) = \frac{1}{2 n_g n_d} \sum_{i \in N_G} \sum_{j \in N_D} \frac{1}{d_{ij}} = \frac{1}{n_g n_d} \sum_{i \in N_G} \sum_{j \in N_D} \epsilon_{ij} \tag{1}
\]

Here \( n_G \) is the number of generator nodes, \( n_D \) is the number of distribution nodes, \( N_G \) is the set of all generator nodes, \( N_D \) is the set of all distribution nodes, \( d_{ij} \) is the distance (i.e. the most efficient path) between a generator node \( i \) and a distribution node \( j \) and \( \epsilon_{ij} \) is the efficiency of this path. Note that only paths between generator nodes and distribution nodes are taken into account by the formula (1). We also scale (1) by 1/2 since we double-count each edge (i.e. both \( \epsilon_{ij} \) and \( \epsilon_{ji} \) are counted although they are equal and represent the same edge).

Generally speaking there are two possible node removal strategies, namely, random removal: whereby each node is equally likely to be removed and targeted removal where nodes are removed based on their “importance”. Random removal simulates random failures whereas targeted removal simulates intentional attack scenarios [9]. Although several targeted removal strategies for both node removal and edge removal (e.g. [14]) have been investigated we follow two standardly used targeted attack approaches based on a node’s degree and a node’s betweenness centrality [16]. Node degree (also called degree centrality) is simply a number of edges connected to it. Betweenness centrality of the node is defined as the number of shortest paths (i.e. most efficient paths) between any pair of nodes which pass through it, including the path between the node itself and other nodes [5].

A. Crucitti-Latora-Marchiori model

First we analysed the behaviour of the CPG using the Crucitti-Latora-Marchiori (CLM for short) model and assuming either intentional attack or random failure. The CLM model originally described in [6] is generic. Later it has been applied to analyse the resilience of the Italian [7] and Korean [9] power grid. The model comprises two types of nodes, generators and distribution substations with transmission lines (i.e. edges) between them. All edges are assigned the so-called edge efficiency, with values between zero and one where one represents optimal working condition and zero indicates line failure. As mentioned, it is assumed that a given generator transmits power to all distribution substations in the network along most efficient paths. Given path \( P \), consisting of edges \( i = 1, \ldots, K \) with their corresponding efficiencies \( \epsilon_i \), the path efficiency \( E(P) \) is defined as:

\[
E(P) = \left[ \sum_{i=1}^{K} \left( \frac{1}{\epsilon_i} \right) \right]^{-1} \tag{2}
\]

Furthermore, each node \( i \) is assigned a maximum capacity \( C_i \) proportional to its initial load \( L_i(0) \); i.e. \( C_i = \alpha \cdot L_i(0) \); where \( \alpha > 1 \) is the tolerance coefficient (essentially a safety margin). This is quite intuitive since all nodes should have the capacity to sustain disturbances or increased demand (hence \( \alpha > 1 \)). The initial load of node \( i \), \( L_i(0) \) corresponds
to standard operating conditions and is defined as its betweenness centrality, assuming all edges having optimal efficiency, $e_{ij} = 1$, at $t = 0$.

At $t = 1$ a single node $k$ and its connecting edges are removed from the network. In the intentional attack or worst case scenario, $k$ is taken to be either the node with the highest degree or node with the highest betweenness centrality, since both measures are commonly taken to represent the node’s relative importance ([9], [7]). In the standard malfunction scenario a random node is removed. Either way, shortest paths which traverse removed edges will have to be reassigned to new edges. Note that the length of the reassigned path can be only greater or equal to the old length. Consequently network efficiency will decrease and is updated according to (1), while still assuming all edges have their initial efficiency of one. Next, the flow of electricity is redistributed along the new most efficient paths. This in turn can cause an overloading of some nodes and a consequent efficiency decrease of their connecting edges. Namely,

$$e_{ij}(t + 1) = \begin{cases} 1 & \text{if } L_j(t) \leq C_i \text{ and } L_i(t) \leq C_j \\ e_{ij}(0) \cdot \frac{C_i}{L_i(t)} & \text{if } L_j(t) > C_i \text{ and } L_i(t) \leq C_j \\ e_{ij}(0) \cdot \frac{\min(C_i, C_j)}{L_i(t) + L_j(t)} & \text{if } L_j(t) > C_i \text{ and } L_j(t) > C_j \end{cases}$$

Note that the case where both nodes $i$ and $j$ are overloaded was not described in the original work [7], so we extended the original definition to also account for this case. In order to simplify the notation we assume it is clear that $L_j(t) > C_i$ and $L_i(t) \leq C_i$ case is defined analogously to the $L_j(t) > C_i$ and $L_i(t) \leq C_j$ case with the simple change of subscripts. After computing new edge efficiencies (3) for all edges, new most efficient paths are found for all pairs of generator-substation nodes based on the updated edge efficiencies, and a new network efficiency (1) is computed. Next, the flow of electricity between generator-substation pairs is redistributed along the new most efficient paths. This redistribution of flow can change loads of nodes again, which in turn can cause edge efficiencies to change again, thus decreasing network efficiency further and the whole process is repeated. To summarize, the process is described by:

- Remove an initial node and its edges at $t = 1$
- Repeat:
  - find new most efficient paths for all pairs of nodes
  - compute network efficiency (1) based on new paths
  - redistribute the flow along the new paths
  - update node loads
  - update edge efficiencies according to (3)

We applied the CLM model to CPG and results are shown in Figure 2. In case of random failure simulation we randomly selected 50 different nodes to be removed during each simulation round (i.e. for each value of $\alpha$, starting from one and with step $\Delta \alpha = 0.5$). The same sample was used during each simulation round in order to avoid fluctuations of network efficiency due to resampling. This was done since we are interested in the relationship between the network efficiency $E(N)$ and $\alpha$ and resampling would just add unwanted noise. As can be seen in the figure, targeted attack scenarios (blue and green curves) cause greater loss of network efficiency, as expected. Interestingly, there are clear curve elbows\(^2\) for all three curves although their values vary with failure type. In cases of highest degree node failure and random node failure the threshold is $\alpha = 1.5$, whereas in the highest load node failure case there are two thresholds, at $\alpha = 1.5$ and at $\alpha = 4$.

B. Modelling catastrophic failures

The second model we used to analyse the CPG resilience was proposed in [15] and was later used to investigate the behaviour of the Iranian power grid under the cascading failure type of events [11]. The node loads in this model are defined in the same way as in the CLM model (i.e. as node betweenness centralities). Node capacity is again defined by $C_i = \alpha \cdot L_i(0)$, where $\alpha$ is the tolerance parameter as before. The major difference in comparison to the CLM model is that according to this model nodes are removed from the network if the load exceeds their capacity. This also removes all edges connected (directly) to them. Note that in the CLM model only a single node with its connected edges was initially (at $t = 1$) removed from the network. Subsequent redistribution of electricity in the CLM model caused deterioration of edge efficiencies and consequently network efficiency but no further complete failures of nodes and their connecting edges from the network. Removing nodes from the network in the case of overload is more in line with “catastrophic failure” where an overload causes complete breakdown of the node and its connecting edges as opposed to only diminished functionality. The algorithm itself proposed in [15] is very simple. Namely, removal of the initial node causes betweenness centralities of some nodes to change. If the updated betweenness centrality exceeds capacity of any node in the network, that node is removed from the network, new shortest distances are computed,

\(^2\) assuming $t = 1$ initially

\(^2\) Sometimes also called the knee of a curve
power flow is redistributed along the new shortest paths and the whole process is repeated either until new equilibrium is reached or until a pre-specified number of simulation steps are exceeded. Note that edges in this model have no efficiencies, i.e., they are either present or have been removed from the network due to failure of one of the two nodes connected to them.

We investigated the CPG network efficiency using the above described model. The same three failure models were used. Namely, two intentional attack scenarios (initial removal of the highest degree node or the node with highest betweenness centrality) and random node removal (as described in the previous subsection), with step $\Delta x = 5$. The results of the simulations are shown in the figure below.

As can be seen, the same general pattern can be observed. Namely, targeted attack scenarios (blue and green curves) cause greater loss of network efficiency than random failures. However, a couple of notable differences can also be observed. Firstly, curve elbow values are much higher here. Secondly, in contrast to the CLM model, initial failure of the highest load (i.e., highest betweenness centrality node) causes greater loss of network efficiency than the failure of the highest degree node. Finally, network efficiencies are much lower than in the CLM model for the same range of $\alpha$ values. This is understandable as already mentioned node overload in this model causes full blown failure of the node and its connecting edges as opposed to simply diminishing functionality like in the CLM case.

C. Local load redistribution model

The local load redistribution model (LLR for short) used in our analysis is partially based on [16] and [17] although we introduce some important modifications. Unlike the previous two models which define the initial load of the node to be equal to its betweenness centrality, thus taking into account topology of the whole network, the LLR model is local in the sense that the initial load of each node depends only on its local topology ([16], [17]). In [16] the initial load of node $i$ at $t = 0$ is defined by $L_i(0) = \mu \cdot k_i^\theta$, where $k_i$ is the degree of node $i$ and $\mu$ and $\theta$ are model parameters. In order to give it an intuitive interpretation we set $\mu = \theta = 1$. The capacity of node $i$ equals $C_i = \alpha \cdot L_i(0) = \alpha \cdot k_i$ where $\alpha > 1$ is the tolerance coefficient, as before ([16], [17]). Initial conditions are the same as before. Namely, at $t = 1$ a single node fails and is removed from the network. Again, we simulate the same three failure scenarios.

An important difference between the LLR model and the previous two models is that in the LLR extra loads generated by the initial node failure spread “outward” from the failed node $i$ and at the next time step affect only those nodes which are in direct contact with $i$. Namely, assuming that node $i$ which is connected to node $j$ fails at time $t$, the proportion of its load redistributed to $j$, $\Delta L_i$ is in the original LLR model ([16], [17]) defined by (where $I_i$ is the “neighbourhood” of $i$):

$$\Delta L_i(t) = L_i(0) \cdot \frac{L_i(t)}{\sum_{m \in I_i} L_m(0)} \quad (4)$$

The new load of $j$ at $t + 1$ is given by

$$L_j(t + 1) = L_j(t) + \Delta L_i(t) \quad (5)$$

Instead of using (4), we define the extra load to be

$$\Delta L_j(t) = L_i(t) \cdot \frac{C_j - L_j(t)}{\sum_{m \in I_i} (C_m - L_m(t))} \quad (6)$$

The reason for using (6) instead of (4) is twofold. Firstly, (6) assumes that the present load of $i$, $L_i(t)$ is redistributed to its neighbours (due to its failure) as opposed to its initial load $L_i(0)$. Namely, taking the sum over $j \in I_i$ in (4), yields $\sum_{j \in I_i} \Delta L_j = L_i(0)$, which is the load of $i$ at $t = 0$. Taking the same sum in (6), yields $\sum_{j \in I_i} \Delta L_j = L_i(t)$ which is the load of $i$ at time $t$. We argue the latter is a more realistic assumption since the present load accounts for the present state of the network. Namely, it accounts for the initial node failure and all subsequent load redistributions which took place. Secondly, extra load redistributed to $j$ in (6) is not proportional to its normalized initial load $L_j(0) / \sum_{m \in I_i} L_m(0)$, as in (4), but rather to its currently available normalized capacity $C_j - L_j(t) / \sum_{m \in I_i} (C_m - L_m(t))$. This is in effect a mitigation strategy which models human intervention. Intuitively, it makes sense to redirect the extra load to neighbouring nodes in proportion to their currently available capacity as opposed to their initial load. As in the original model we also use (5) to update the weights. Note that if more neighbours of $i$ fail at the same time, the contribution of all of them has to be taken into account. This is done simply by taking the sum $\sum_k \Delta L_k(t)$ in (5) where $k$ denotes all nodes which fail at time $t$.

We simulated the behaviour of CPG based on the above described model (6) and the same three failure scenarios as before (with step $\Delta x = 0.025$). Results are shown in Figure 4 below. Noteworthy, it is inevitable that neighbouring nodes will fail for very small values of $\alpha$, since in such cases all nodes are running at the very limit of their capacity. Hence, even a very small extra load a node receives from its failed neighbour will inevitably exceed its capacity and will cause it to fail. Indeed, such behaviour can be clearly seen in Figure 4 where network efficiencies for $\alpha = 1$ equal zero, thus indicating total network collapse. An obvious solution to this problem is to simply increase the tolerance coefficient $\alpha$, which is obviously the case in practice. Interestingly, even a
small increase of tolerance coefficient of about 50% substantially increases the network resilience, as shown in Figure 4.

![Figure 4. LLR model. Simulation results for CPG (Δα = 0.025)](image_url)

**III. CONCLUSION**

In this study we analysed the resilience of the Croatian power grid based on the complex network analysis approach. Three different models were used. Major limitation of all three models is that they are discrete-time models. Also the assumption that a generator node transmits power to all distribution nodes in the network is unrealistic. Moreover generators and distribution stations as well as transmission lines are treated as simple nodes and edges respectively, without modelling their physical and electrical properties. That being said however, the expectation of developing both analytically tractable and accurate models of the highly complex cascading failure processes in power grids is at present unrealistic [18]. Therefore, despite all of the above mentioned limitations of simple models described here and other similar models described elsewhere in the literature, they should be viewed as an important stepping stone towards more realistic models.


