AIMS Energy, 6(2): 245–260.
DOI: 10.3934/energy.2018.2.245
Received: 29 November 2017
Accepted: 12 March 2018
Published: 19 March 2018

http://www.aimspress.com/journal/energy

Research article

Distribution network topology identification based on synchrophasor

Stefania Conti, Santi A. Rizzo*, Nunzio Salerno and Giuseppe M. Tina

Dipartimento di Ingegneria Elettrica Elettronica e Informatica (DIEEI), University of Catania, viale A. Doria, 6-95125 Catania, Italy

* **Correspondence:** Email: santi.rizzo@dieei.unict.it.

Abstract: A distribution system upgrade moving towards Smart Grid implementation is necessary to face the proliferation of distributed generators and electric vehicles, in order to satisfy the increasing demand for high quality, efficient, secure, reliable energy supply. This perspective requires taking into account system vulnerability to cyber attacks. An effective attack could destroy stored information about network structure, historical data and so on. Countermeasures and network applications could be made impracticable since most of them are based on the knowledge of network topology. Usually, the location of each link between nodes in a network is known. Therefore, the methods used for topology identification determine if a link is open or closed. When no information on the location of the network links is available, these methods become totally unfeasible. This paper presents a method to identify the network topology using only nodal measures obtained by means of phasor measurement units.

Keywords: accuracy; distribution network; measurement errors; numerical methods; phasor measurement unit; power system state estimation; system of equations; topology identification

1. Introduction

In transmission networks, state estimation (SE) is a key function, since its results are used in fundamental network applications, such as optimal power flow, evaluation of available transfer capability, estimation of voltage and transient stability [1]. SE processes and filters real-time data (in terms of analogic measures, i.e., line flow, bus injection, voltage magnitude, or digital measures, i.e., switching devices status) available at the control centre of the transmission system to best estimate the current state of the network in order to provide dependable results to other network applications [2]. Typically, the aforementioned real-time applications are based on the knowledge of network topology, which is determined from the switching devices status [3]. Moreover, SE also uses topology to

estimate the state and to identify bad data [1]. For these reasons, a great effort has been lavished to detect topology error in transmission systems [2–7].

Traditionally, power distribution networks are either not observable or only partially observable [8], since they were excluded, in the past, from the deployment of communication infrastructures, automation, monitoring, and control systems [9]. Recently, the distribution system is facing a deeply transformation due, on the one hand, to the load growth and the proliferation of new "actors" (distributed generators, prosumers, electric vehicles, etc.) and, on the other hand, to the increasing demand for high quality, efficiency, security, reliability of energy supply, leading towards Smart Grid (SG) implementation [10]. In this complex scenario, SG implementation asks for the introduction of innovative paradigms, network control and management techniques, as well as planning and maintenance strategies that, in turn, ask for the integration of additional infrastructures, computational resources and data storage, for information exchange and processing, electrical vehicles, intelligent devices, network optimization procedures, remote control, automation and so on [11–17]. Many benefits are expected from this distribution network upgrade [18]; on the other hand, it also involves new problems related to distribution system vulnerability to cyber-physical attacks [19–21].

Many countermeasures against these attacks [22] require the knowledge of the network topology [23]. Nevertheless, many of the aforementioned applications are based on network topology too [9,24,25]. As said before, topology identification topic has been well investigated for transmission networks, differently from the distribution system, for which a traditional topology identification technique is not available at present. Of course, the distribution system upgrade could enable the Distribution System Operator (DSO) to adopt techniques similar to the ones used in transmission networks or even to develop new ones [25,26]. It is worth to highlight that these techniques are based on the assumption that the location of each link in the network is known, then "topology identification" means to determine if a link is open or closed [2–7,25,26]. A well-suited cyber attack can destroy such information and make network topology identification impossible, as well as make countermeasures and network applications impracticable. New techniques based on phasor measurement units (PMUs), specifically devised for distribution networks, could cope with this issue since they can provide information about power injection and voltage at the network nodes (therefore, the current absorbed from/supplied to the nodes is also known). In [9] the problem of reconstructing the topology of a portion of the distribution network using a dataset of voltage measurements is investigated, but it assumes that all power lines in the grid have the same inductance/resistance ratio and it neglects the information about power injections.

On the other hand, [23] performs a blind topology identification only by using power injection data at each bus, neglecting information about voltages, and it considers a simplified power flow, that is the linear DC power flow. In [8] both nodal measurement types are considered, but the algorithm is based on a linear coupled approximation for lossless AC power flow and the results are conditioned on a given assumption regarding correlations in power injections at the non-substation buses. In [27] network connectivity verification is proposed. It is based on an algorithm that neglects the presence of dispersed generation along the distribution network, since it assumes that voltage magnitude always decreases downstream along the feeder as it usually occurs in traditional passive distribution networks only.

The method proposed in this paper is able to identify the network topology after a critical event such as a cyber-physical attack or, more in general, when there is no information about the

connections among the network nodes. It is able to exploit the knowledge of both nodal current and voltage measured by means of PMUs, and it is based on a full AC power flow avoiding all the limitations of the aforementioned methods.

Section II describes the method proposed for identifying the network admittance matrix. Section III describes the technique adopted to analyze the real part of the matrix in order to identify the network topology accounting for PMUs accuracy. Finally, section IV reports some numerical results.

2. Nodal admittance matrix identification by means of synchrophasors

The one-line diagram of a feeder line of a three-phase, symmetrical, radial MV distribution network is considered. A graphic representation of the network is reported in Figure 1, where a node (filled point) is a network bus where customers and/or generators are connected, or a switching substation.

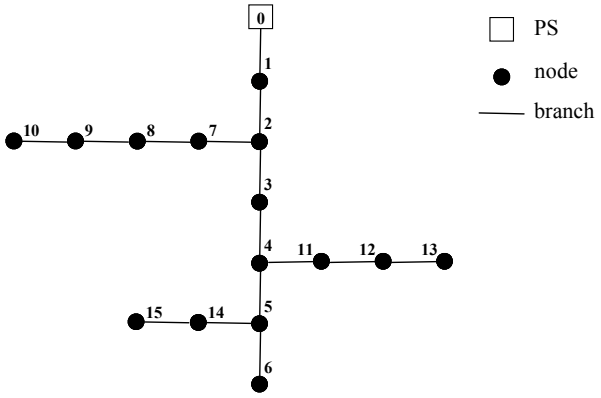

Figure 1. One-line diagram of a feeder line of a three-phase, symmetrical, radial distribution network.

A branch (solid line) is a link that represents the electrical equipment connecting two nodes. The primary substation (PS) bus takes the labeled 0, the other nodes are sequentially numbered such that the ones from the PS to node i take a number lower than i. These nodes are said "upstream" from node i while node i is said "downstream" from them. From the graph theory a tree, as the radial network shown in Figure 1, has n branches and $n+1$ nodes. So, each branch can be numbered as the downstream node which it is connected to. The Kirchhoff's current law can be written for each node and the following equation in matrix form can be written [28]:

$$[\mathbf{Y}][\mathbf{V}] = [\mathbf{J}] \tag{1}$$

where:

[Y] size (n+1) × (n+1)—is the nodal admittance matrix, whose elements can be directly obtained by topological inspection according to the following rule:

(1) y_{ii} (self-admittance) is the sum of the admittances related to the branches that are connected to node i;

(2) y_{ij} (mutual admittance) is the negative of the sum of the admittances related to the common branches between nodes *i* and *j*;

[V] size (n+1) × 1—node voltages array;

[J] size (n+1) × 1—load/generator currents array; the real part of the current is considered positive when it is injected into the node ("absorbed" load current).

The equations in (1) are linear and complex. In the considered scenario, the elements in **[Y]** are unknown and do not change; the elements in **[V]** and **[J]** are the synchrophasors (Figure 2).

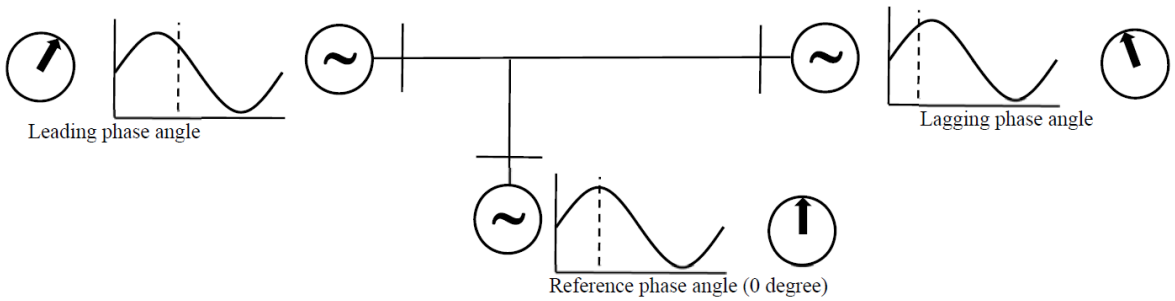

Figure 2. Synchrophasors [29].

After a cyber attack that destroyed the stored information about the network structure, firstly, **[Y]** is found by exploiting a set of equation systems like the one in (1) and the PMUs information about nodal voltages and currents. Subsequently, the network admittance matrix is used to identify the network topology, thus restoring the missing information. In fact, the network admittance matrix encloses information about topology by construction, since y_{ij} equal to zero implies that there is not a link between node *i* and *j*, otherwise the nodes are connected to each other. On the other hand, in a real world application, such a term could differ from zero even when there is not a link between node *i* and *j* due to voltage and current data accuracy and numerical issues.

At the first step, the nodes belonging to a given feeder line, *fl*, have to be found to identify the nodal admittance related to *fl* itself. To this aim, the sending-end switch of the feeder line is set to "closed status", while the sending-end switches of the other feeder lines are set to "open status", to search for the nodes belonging to *fl*. More specifically, the nodes whose voltage syncrophasor differs from zero are the ones belonging to *fl*. Then another switch is set to "closed status", and the new nodes whose voltage syncrophasor differs from zero to another feeder line. For a given PS, this procedure stops when the nodes of its feeders are identified.

After that, the following method could be adopted to identify matrix **[Y]** of a feeder. Note that, to convert **[Y]** of Eq1 into an equivalent vector, in order to explicit the unknowns, let us define the row string vector operator, *rs* with size 1 × $(n+1)^2$, the column string vector operator, *cs* with size $(n+1)^2 \times 1$, and the vectorization operator, ν (transposed row string vector) with size $(n+1)^2 \times 1$, as follows [30]:

$$[\mathbf{rs}(\mathbf{Y})] \equiv \begin{bmatrix} y_{0,0} & y_{0,1} & \cdots & y_{0,n} & y_{1,0} & y_{1,1} & \cdots & y_{1,n} & \cdots & y_{n,0} & y_{n,1} & y_{n,n} \end{bmatrix} \quad (2)$$

$$[\mathbf{cs}(\mathbf{Y})]\equiv[y_{0,0}\quad y_{1,0}\quad \cdots\quad y_{n,0}\quad y_{0,1}\quad y_{1,1}\quad \cdots\quad y_{n,1}\quad \cdots\quad y_{0,n}\quad y_{1,n}\qquad y_{n,n}]^T \tag{3}$$

$$[\mathbf{v}(\mathbf{Y})]\equiv[\mathbf{rs}(\mathbf{Y})]^T=[\mathbf{cs}(\mathbf{Y}^T)] \tag{4}$$

Therefore, the following relations hold:

$$[\mathbf{rs}(\mathbf{J})]=[\mathbf{J}]^T \tag{5}$$

$$[\mathbf{cs}(\mathbf{J})]=[\mathbf{J}] \tag{6}$$

$$[\mathbf{v}(\mathbf{J})]=[\mathbf{J}] \tag{7}$$

Moreover, the following relation is obtained by considering the standard multiplication of three matrices M_1, M_2, M_3 [30]:

$$[\mathbf{v}(\mathbf{M_1M_2M_3})]=[\mathbf{M_1}]\otimes[\mathbf{M}_3]^T[v(\mathbf{M}_2)] \tag{8}$$

in which symbol ⊗ denotes the Kronecker product.

Note that, Eq 1 does not change by pre-multiplying it by the identity matrix **[I]** (with size n+1):

$$[\mathbf{I}][\mathbf{Y}][\mathbf{V}]=[\mathbf{J}] \tag{9}$$

Considering relations (7) and (8), this system of equations can be rewritten as follows:

$$\overbrace{[\mathbf{I}]\otimes[\mathbf{V}]^T}^{(n+1)\,x\,(n+1)^2}\ [\mathbf{v}(\mathbf{Y})]=[\mathbf{J}] \tag{10}$$

For example, when $n = 2$:

$$\begin{bmatrix} V_0 & V_1 & V_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & V_0 & V_1 & V_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & V_0 & V_1 & V_2 \end{bmatrix} \begin{bmatrix} y_{0,0} \\ y_{0,1} \\ y_{0,2} \\ y_{1,0} \\ y_{1,1} \\ y_{1,2} \\ y_{2,0} \\ y_{2,1} \\ y_{2,2} \end{bmatrix} = \begin{bmatrix} J_0 \\ J_1 \\ J_2 \end{bmatrix} \tag{11}$$

It is apparent that the number of equations is lower than the number of unknowns and, consequently, acquiring one set only of concurrent measures by means of the PMUs is not enough to calculate the elements of **[Y]**. Therefore, more equations are necessary, in other words many concurrent acquisitions at different instants must be performed. The first set of $(n+1)$ equations is obtained by using in (10) the nodal measures acquired at time t_1 by means of the PMUs. The second set of equations is obtained thanks to the subsequent concurrent measures, acquired at time t_2, and so on.

Therefore, the following system of equations is obtained by considering $(n+1)$ concurrent measures in $(n+1)$ different instants:

$$\underbrace{\begin{bmatrix} [\mathbf{I}]\otimes[\mathbf{V}(t_1)]^T \\ [\mathbf{I}]\otimes[\mathbf{V}(t_2)]^T \\ \cdots \\ [\mathbf{I}]\otimes[\mathbf{V}(t_{n+1})]^T \end{bmatrix}}_{[\mathbf{V'}]}[\boldsymbol{\nu}(\mathbf{Y})]=\underbrace{\begin{bmatrix} [\mathbf{J}(t_1)] \\ [\mathbf{J}(t_2)] \\ \cdots \\ [\mathbf{J}(t_{n+1})] \end{bmatrix}}_{[\mathbf{J'}]} \tag{12}$$

where:

$$\begin{aligned} &[\mathbf{V}(\tau)]=[V_0(\tau)\ \ V_1(\tau)\ \ \ldots\ \ V_k(\tau)\ \ \ldots\ \ V_n(\tau)]^T \\ &[\mathbf{J}(\tau)]=[J_0(\tau)\ \ J_1(\tau)\ \ \ldots\ \ J_k(\tau)\ \ \ldots\ \ J_n(\tau)]^T \\ &\textit{with}\quad \tau=t_1,t_2,\ldots,t_{n+1} \end{aligned} \tag{13}$$

[V'], with size $(n+1)^2 \times (n+1)^2$, is the coefficient matrix and **[J']**, with size $(n+1)^2 \times 1$, is the known term array. System (12) is resolvable provided that the measures are not correlated among themselves, that is for each set of values measured at a given instant there must not exist a combination of the other sets that led to the same values in the considered set.

Note that, even though there are $(n+1)^2$ elements in **[Y]**, the number of admittances to be found is $(n+1)(n+2)/2$ for symmetry reasons. Let us apply the reduced form of operator ν to account for this symmetric matrix [30]:

$$\begin{aligned} [\bar{\boldsymbol{\nu}}(\mathbf{Y})]=&\Big[\overbrace{y_{0,0}\ \ y_{0,1}\ \ \cdots\ \ y_{0,n}}^{1\times(n+1)}\ \ \overbrace{y_{1,1}\ \ y_{1,2}\ \ \cdots\ \ y_{1,n}}^{1\times n}\ \ \cdots\ \ \overbrace{\cdots\ \ \cdots}^{1\times(n+1-k)}\ \ \cdots\ \ \overbrace{y_{n,n}}^{1\times 1}\Big]^T= \\ &=[\ldots\ y_{rc}\ \ldots]^T \quad \text{where } r=0,1,2,\ldots n \text{ and } c=r,r+1,\ldots n \end{aligned} \tag{14}$$

The reduced operator in (14) can be obtained from the one in (4) by applying the row deletion matrix, **[D]** with size $0.5(n+1)(n+2) \times (n+1)^2$:

$$[\bar{\boldsymbol{\nu}}(\mathbf{Y})] = \mathbf{[D]}\ [\boldsymbol{\nu}(\mathbf{Y})] \tag{15}$$

where **[D]** is obtained from a $(n+1)^2$-dimensioned identity matrix by deleting some rows. More specifically, the i-th row of this matrix is deleted if the i-th element of $[\boldsymbol{\nu}(\mathbf{Y})]$ is y_{rc}, with r greater than c. For example, when $n = 2$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \leftrightarrow \\ \leftrightarrow \\ \\ \end{matrix} \begin{bmatrix} y_{0,0} \\ y_{0,1} \\ y_{0,2} \\ y_{1,0} \\ y_{1,1} \\ y_{1,2} \\ y_{2,0} \\ y_{2,1} \\ y_{2,2} \end{bmatrix}$$

$$\Rightarrow \quad [\mathbf{D}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{16}$$

On the other hand, the operator ν can be obtained from the reduced one by applying the row addition matrix, **[A]**:

$$[\mathbf{v}(\mathbf{Y})] = [\mathbf{A}]\left[\bar{\mathbf{v}}(\mathbf{Y})\right]$$
$$\text{size of } [\mathbf{A}] \quad (\mathrm{n}+1)^2 \times \frac{(\mathrm{n}+1)(\mathrm{n}+2)}{2} \tag{17}$$

where **[A]**—size $(n+1)^2 \times 0.5(n+1)(n+2)$ —is obtained from a $(n+1)^2$-dimensioned identity matrix by deleting some columns and adding them to the others. More specifically, the i-th column of this matrix is deleted if the i-th element of $[\mathbf{v}(\mathbf{Y})]$ is y_{rc}, with r greater than c, and the column deleted is added to the j-th column provided that the j-th element of $[\mathbf{v}(\mathbf{Y})]$ is y_{cr}. For example, when $n = 2$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow [\mathbf{A}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{18}$$

By applying to ν the row deletion and addition matrix in turn, the following relation holds:

$$[\mathbf{v}(\mathbf{Y})] = [\mathbf{A}]\left[\bar{\mathbf{v}}(\mathbf{Y})\right] = [\mathbf{A}][\mathbf{D}][\mathbf{v}(\mathbf{Y})] \tag{19}$$

Therefore, the equation system in (12) can be rewritten as follows to account for symmetries:

$$[\mathbf{V}'][\mathbf{A}]\left[\bar{\mathbf{v}}(\mathbf{Y})\right] = [\mathbf{J}'] \tag{20}$$

and the sub-set of independent equations in the system is obtained by applying the row deletion

matrix, provided that the measures are not correlated among themselves. Pre-multiplying the terms by the deletion matrix, the relations in (20) can be written as:

$$[\mathbf{D}][\mathbf{V'}][\mathbf{A}]\left[\overline{\mathbf{v}}(\mathbf{Y})\right]=[\mathbf{D}][\mathbf{J'}] \tag{21}$$

and the following reduced system of equations is obtained:

$$\begin{aligned}&\left[\overline{\mathbf{V'}}\right]\left[\overline{\mathbf{v}}(\mathbf{Y})\right]=\left[\overline{\mathbf{\Delta J'}}\right]\\ &with \quad \left[\overline{\mathbf{V'}}\right]=[\mathbf{D}][\mathbf{V'}][\mathbf{A}]\\ &\qquad\quad \left[\overline{\mathbf{\Delta J'}}\right]=[\mathbf{D}][\mathbf{\Delta J'}]\end{aligned} \tag{22}$$

The reduced system of equations accounts for symmetries. On the one hand, it includes $(n+1)(n+2)$ unknowns, that are the real and imaginary parts of the admittances; on the other hand, there are $(n+1)(n+2)$ equations because each complex linear equation can be treated as two real linear equations.

3. Topology identification by using the nodal admittance matrix

The topology identification method proposed in this section is based on the use of the real part of **[Y]**, i.e., the network conductance matrix **[G]**. The resulting information about the links in the network are stored in an array, **[L]**, where a link between node i and j is represented by a number, k, which is computed as follows:

$$\begin{aligned}&k = i(n+1)+ j\\ &with \;\; i \in [0, n],\, j \in [i+1, n]\end{aligned} \tag{23}$$

For example, when the network depicted in Figure 1 is considered, the link array **[L]** is:

$$[\mathbf{L}]=\begin{Bmatrix}1,\,18,\,35,\,39,\,52,\,69,\,75,\,86,\,94,\\ 120,\,137,\,154,\,188,\,205,\,239\end{Bmatrix} \tag{24}$$

Information about the topology can be easily derived from **[G]** since the element at position (i,j) is equal to zero when there is not a link between nodes i and j, otherwise the nodes are connected to each other. On the other hand, in practical applications, such an element could differ from zero even when nodes i and j are not directly connected to each other due to PMUs accuracy and numerical issues. Therefore, a method to properly analyse the conductance matrix is proposed in the following. More specifically, such a method checks the sign of the mutual conductance values and compares some of them with the self-conductance. Moreover, it imposes radiality and connectivity constraints. The main steps of the proposed method are reported in Figure 3, where:

a. *ascend_order_position* is a function that returns an array P such that:

$$G\left[i,P\left[p-1\right]\right]<G\left[i,P\left[p\right]\right] \quad with \quad p\in\left[1,n\right] \tag{25}$$

when it is applied to (i, **G**);

b. α is a negative number with magnitude lower than one;

c. *apply_eq23* is a function that applies equation 23 (even when j is lower than $i+1$);

d. *remove_duplicated_links* is a function that removes from L each link k obtained for i higher than j;

e. *force_radiality* is a function that removes as many links as necessary to avoid loops;

f. *force_connectivity* is a function that adds as many links as it is necessary to avoid isolated set of nodes while keeping the radiality constraint satisfied.

The first link stored in **[L]** is the one connecting node 0 to the downstream node. Note that only one node is connected to node 0 when a specific feeder line is considered. Therefore, looking at the first row of **[G]**, the column with the lowest value (that is the negative number with higher magnitude) indicates the node connected to 0 and, consequently, the related link is stored in **[L]** (see rows 3–4 in Figure 3).

When a generic node i is considered, the previous mechanism is firstly adopted. In other words, looking at the i-th row of **[G]**, column j with the lowest value, i.e., **G**[i,**P**[1]], is found and the link between i and j is stored (see rows 7–8 in Figure 3). After that, the lowest one among the other elements in row i, i.e., **G**[i,**P**[2]], indicates another node that could be connected to i. The related link is stored in L provided that:

(1) the sum of the mutual-conductance magnitudes considered until that step does not exceed the self-conductance;
(2) the aforementioned lowest element is negative; and
(3) it has a magnitude comparable with the self-conductance (row 11 in Figure 3).

With reference to point 3, a coefficient α is used to establish if the element **G**[i,**P**[p]] can be considered exactly as "zero" although it actually differs from "zero". Such a difference is due to data accuracy and numerical issues.

After that, the next lowest value is considered and so on, as long as the three aforementioned conditions are satisfied (rows 11–15 in Figure 3). When the p-th lowest value is negative and it is comparable with the self-conductance (i.e., it satisfies conditions 2 and 3), but it hasn't been stored in **[L]** since the sum exceeds the self-conductance (i.e., it does not satisfy condition 1), the link is stored anyway if the addition of the mutual-conductance leads to a better approximation of the self-conductance (rows 16–19 in Figure 3).

A link between β and γ could be stored in **[L]** when i is equal to β, and the same link could be stored again in **[L]** when i is equal to γ. In other words, the same link is stored twice.

Conversely, a link between β and γ could be stored in **[L]** when i is equal to β, even though the link is supposed not to exist when i is equal to γ.

In both cases the link is added although in the first case it is more probable that the link actually exists. The function *remove_duplicated_links* erases the duplicates in the occurrence of the first case to avoid storing the same link twice (row 22 in Figure 3). Subsequently, two functions that, respectively, remove and add links to account for the topology constraints are executed (rows 23–24 in Figure 3).

```
1   L := ∅
2   i := 0
3   P := ascend_order_position (i, G)
4   L := L ∪ {apply_eq23(i,P[0])}
5   i := 1
6   while (i ≤ n)
7   P := ascend_order_position (i, G)
8   L := L ∪ {apply_ eq23(i,P[0])}
9   sum := |G[i,P[0]]|
10  p := 1
11  while ( p < n & G[i,P[p]] < αG[i,i] & |G[i,P[p]]|+sum ≤ G[i,i] )
12  L := L ∪ {apply_ eq23(i,P[p])}
13  sum := sum+|G[i,P[p]]|
14  p := p+1
15  endwhile
16  if ( p < n & G[i,P[p]] < αG[i,i] & |G[i,P[p]]| < 2(G[i,i]-sum) )
17  L := L ∪ {apply_ eq23(i,P[p])}
18  sum := sum+|G[i,P[p]]|
19  endif
20  i:=i+1
21  endwhile
22  L := remove_duplicated_links (L)
23  L := force_radiality (L)
24  L := force_connectivity (L)
```

Figure 3. Main steps of the method adopted to analyze the network conductance matrix in order to figure out the topology.

The ability of the method to correctly identify each link is strongly dependent on the accuracy of **[G]**, which in turn depends on the accuracy of the adopted PMUs, on the method adopted to solve the reduced system of Eq 22, and on the precision data type. Assuming a steady-state condition, a simple method to limit the impact of the PMUs' accuracy on the network conductance matrix is to repeat the acquisition of the set of nodal measures. More specifically, the values of V_i and I_i at time τ in Eq 13 can be obtained by averaging their measured values in the neighbourhood of τ.

For example, when R consecutive acquisitions are performed in the neighbourhood of a given time $\tau = t_m$, the voltage at the i-th node is computed as:

$$V_i(t_m) = \frac{\sum_{s=1}^{R} V_i\left(t_m + \left(s - \frac{R+1}{2}\right)\Delta\right)}{r}$$
$$with \qquad t_{m-1} < t_m - \left(\frac{r-1}{2}\right)\Delta \qquad (26)$$
$$t_m + \left(\frac{r-1}{2}\right)\Delta < t_{m+1}$$

where Δ has to be a time interval small enough to ensure that the nodal voltages and currents do not change (accounting for R). The voltage (and current) values computed by using (26) are used in (22) to improve admittance matrix accuracy.

4. Case study

The method proposed to estimate the nodal admittance matrix and to identify the topology by using its real part has been implemented in a Java program. The program has been applied to the distribution network represented in Figure 1. The frequency has been assumed always equal to the nominal one and the coefficient α has been set to –0.05.

Firstly, it has been tested in case of no measurement error. To this aim, the line voltage magnitude at node 0 has been assumed always equal to 20 kV with phase 0, while 16 sets of simultaneous nodal voltages are randomly created for the other 15 nodes imposing that nodal voltage constraints are satisfied (±10% of nominal voltage as established by the standard EN 50160). After that, the simultaneous values of nodal currents are obtained by applying Eq 1 for each voltage set, so that each node could appear as a load point, a generation point or both. These sets of voltage and current values are provided to the program implementing the proposed method in order to emulate the concurrent measures of both voltage and current acquired in 16 different time instants.

The program has correctly evaluated the values of both real and imaginary parts of the nodal admittance matrix, as well as it has correctly identified the network topology (i.e., the identified link array is equal to the one in 24).

Subsequently, the program has been tested considering different PMUs accuracies. ANSI C 12.20 is the American national standard for electricity meter. The meters should satisfy Accuracy Class 5 (i.e., ±0.5%) at least, the best ones comply with Accuracy Class 2 (i.e., ±0.2%). Usually, PMU are very precise meters because their accuracy is less or equal to 0.1% [31].

For a given accuracy, 100 tests have been performed and, for each test, the estimated link array has been compared with the correct one in order to compute the number of elements correctly identified. The average percentage of links correctly identified is reported in Figure 4 for different number of consecutive measures, with $R \in [1,10]$. Therefore, 50,000 tests have been performed to evaluate the performance of the proposed approach on the 16-node test network of Figure 1.

In Figure 4, for a given PMUs' accuracy and R, the height of the related rectangle is equal to the average number of links correctly identified expressed in percentage.

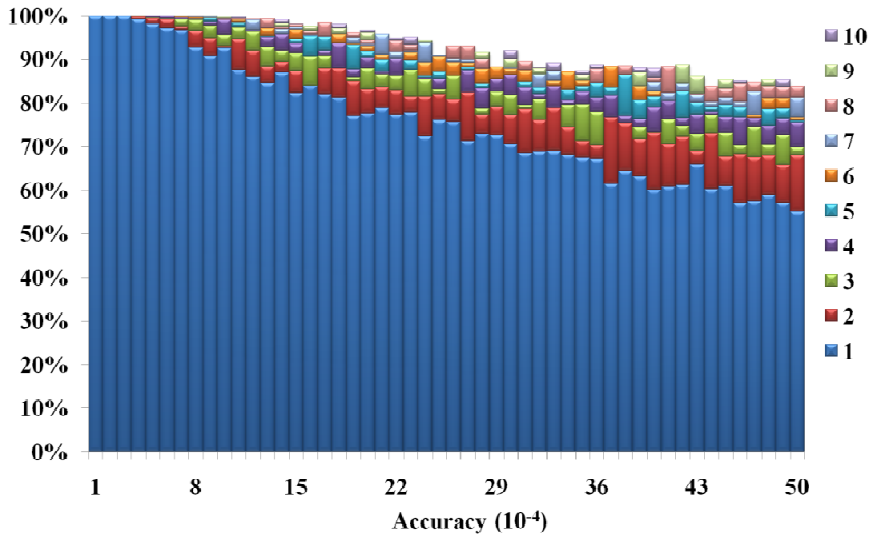

Figure 4. Average percentage of links correctly identified for the 16-node test system (Figure 1).

Moreover, for a given PMUs' accuracy the rectangles are stacked one behind the other for increasing values of R. More specifically, the first one, i.e., the one in front, refers to $R = 1$, the second refers to $R = 2$ and so on. This implies that, for a given accuracy, the rectangle related to $R = 2$ is visible only when its height is greater than the one related to $R = 1$. Similarly, for example, a rectangle related to $R = 5$ is visible only if its height is greater than the ones related to r equal to 1, 2, 3 and 4.

Figure 4 shows that the proposed topology identification method always enable to identify correctly the network without any new acquisition (i.e., $R = 1$) when the PMUs' accuracy is less than 0.04%. Moreover, more than the 90% of the links are correctly identified when only one acquisition is performed (i.e., $R = 1$) and the PMUs' accuracy is less or equal to 0.1%. When r is greater than 8, the method always enables to identify correctly the network if the PMUs' accuracy is less or equal to 0.1%. Moreover, even when poor quality PMUs are considered, for example with accuracy up to 0.5% (i.e., Accuracy Class 5), at least the 90% of the links are correctly identified if many re-acquisitions are executed.

The main fact arising from the results is that few re-acquisitions are necessary to obtain correct results with typical PMU's accuracy. Finally, the worse the PMUs' accuracy the greater the advantage of performing re-acquisitions.

The program has also been applied to the IEEE test feeder with 33 nodes [32] represented in Figure 5. The link array **[L]** related to the network is the following:

$$[\mathbf{L}] = \begin{Bmatrix} 1, 35, 51, 69, 88, 103, 137, 171, 190, 205, \\ 239, 273, 307, 341, 375, 409, 443, 477, \\ 511, 545, 613, 647, 681, 749, 783, 851, \\ 885, 919, 953, 987, 1021, 1055 \end{Bmatrix} \tag{27}$$

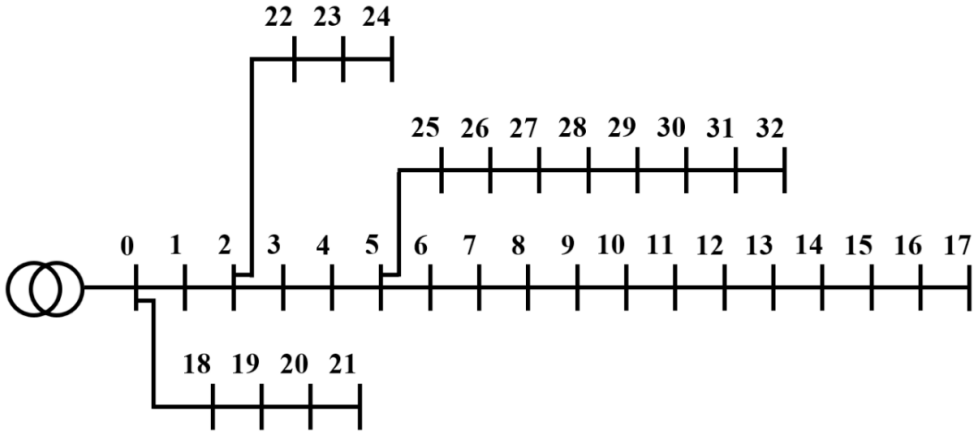

Figure 5. 33-node test system.

Similarly, to the previous test system, 33 sets of simultaneous nodal voltages are randomly created for node 1–32 then the related nodal currents are computed by applying Eq 1. These sets emulate the concurrent measures of both voltage and current acquired in 33 different time instants. When no measurement error is considered, that is the previous values are used without any modification, the proposed method has been able to find the correct value of all elements in the nodal admittance matrix. Moreover, the identified link array is equal to the one in 27 and hence the network topology has been correctly identified too. After that, 100 tests have been performed for each PMUs' accuracy as performed for the 16-node test system. More specifically, a slightly modification of the sets of concurrent measures has been done accordingly the given accuracy. The average percentage of links correctly identified is reported in Figure 6.

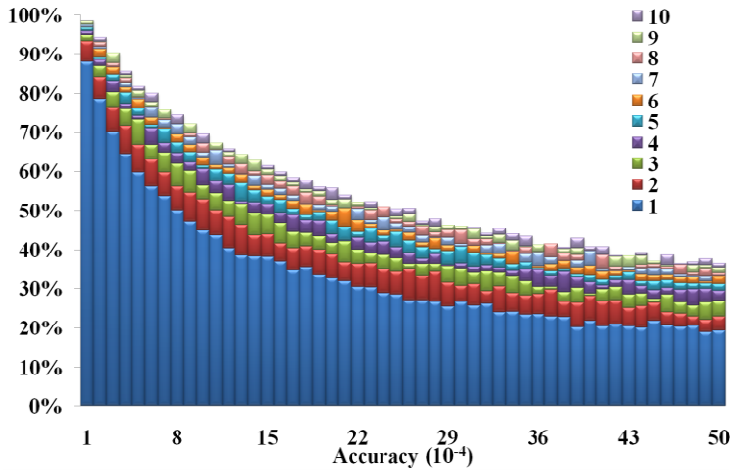

Figure 6. Average percentage of links correctly identified for the 33-node test system (Figure 5).

Comparing with the results obtained for the previous networks it is evident that the performance of the method degrades when the size of the network is doubled. On the other hand, the positive effect of performing more than one measure acquisition is very considerable even if the PMUs' accuracy is small, differently from the 16-node test system.

Therefore, the main fact arising from the tests performed on the 33-node test system is that the method proposed to estimate the nodal admittance matrix and to identify the topology still works but a large number of re-acquisitions are necessary to obtain sufficient or good results with typical PMU's accuracy.

5. Conclusion and future work

The basic theoretical aspects of the proposed method to identify the network topology using only voltage and current measures have been described. The method firstly uses these data to evaluate the elements of the nodal admittance matrix. It has been mathematically shown that to evaluate the value of all elements is necessary to concurrently measure both voltage and current at all nodes. Moreover, to this aim is also necessary to perform concurrent measures in different instants and the number of instants has to be equal to the number of nodes.

The second part of the method uses the real part of the nodal admittance matrix to identify the network topology. It checks the sign of the mutual conductance values and compares some of them with the self-conductance, and then it imposes radiality and connectivity constraints. The performance of the method to correctly identify each link strongly depends on the matrix accuracy and, in turn, on the accuracy of the adopted PMUs, on the numerical method adopted to solve the system of equations as well as on the precision of data storage.

The ability of the proposed method to identify the topology has been tested on two networks considering different PMUs' accuracy. When typical PMUs' accuracy is considered, the results show that the topology is always correctly identified in small networks. As the number of nodes increases, sometimes the topology is not correctly identified. However, a very great number of links in the network is correctly identified, and almost all if the measurements are very accurate.

Therefore, the proposed method is a promising starting point in the field of topology identification in distribution network subject to cyber attacks that have destroyed all information about network structure, historical data and so on. Further works will regard, by one hand, the development of very accurate PMUs, the research of the best method to solve the system of equations used for identify the nodal admittance matrix as well as the investigation of data storage precision on the performance. On the other hand, new features and improvement are necessary in the view of fully dependable topology identification in large networks.

Conflict of interest

The authors declare no conflict of interest in this paper.

References

1. Monticelli A (2004) Electric power system state estimation. *P IEEE* 88: 262–282.
2. Kumar DMV, Srivastava SC, Shah S, et al. (1996) Topology processing and static state estimation using artificial neural networks. *IEE P-Gener Transm D* 143: 99–105.
3. Clements KA, Costa AS (2002) Topology error identification using normalized Lagrange multipliers. *IEEE T Power Syst* 13: 347–353.

4. Singh N, Glavitsch H (1991) Detection and identification of topological errors in on line power system analysis. *IEEE T Power Syst* 6: 324–331.
5. Costa IS, Leao JA (1993) Identification of topology errors in power system state estimation. *IEEE T Power Syst* 8: 1531–1538.
6. Silva APAD, Quintana VH, Silva APAD, et al. (1995) Pattern analysis in power system state estimation. *Int J Elec Power* 17: 51–60.
7. Singh D, Pandey JP, Chauhan DS (2005) Topology identification, bad data processing, and state estimation using fuzzy pattern matching. *IEEE T Power Syst* 20: 1570–1579.
8. Deka D, Backhaus S, Chertkov M (2015) Structure learning and statistical estimation in distribution networks-Part II. *Eprint Arxiv*, 1–1.
9. Bolognani S, Bof N, Michelotti D, et al. (2013) Identification of power distribution network topology via voltage correlation analysis. IEEE Conference on Decision and Control. *IEEE*, 1659–1664.
10. Moslehi K, Kumar R (2010) A reliability perspective of the smart grid. *IEEE T Smart Grid* 1: 57–64.
11. Conti S, Rizzo SA, El-Saadany EF, et al. (2014) Reliability assessment of distribution systems considering telecontrolled switches and microgrids. *IEEE T Power Syst* 29: 598–607.
12. Rueda JL, Guaman WH, Cepeda JC, et al. (2013) Hybrid approach for power system operational planning with smart grid and small-signal stability enhancement considerations. *IEEE T Smart Grid* 4: 530–539.
13. Moeini-Aghtaie M, Farzin H, Fotuhi-Firuzabad M, et al. (2016) Generalized analytical approach to assess reliability of renewable-based energy hubs. *IEEE T Power Syst* 32: 368–377.
14. Minciardi R, Sacile R (2012) Optimal control in a cooperative network of smart power grids. *IEEE Syst J* 6: 126–133.
15. Zubo RHA, Mokryani G, Rajamani HS, et al. (2016) Operation and planning of distribution networks with integration of renewable distributed generators considering uncertainties: A review. *Renew Sust Energ Rev* 72: 1177–1198.
16. Soltani NY, Kim SJ, Giannakis GB (2015) Real-time load elasticity tracking and pricing for electric vehicle charging. *IEEE T Smart Grid* 6: 1303–1313.
17. Gungor VC, Sahin D, Kocak T, et al. (2011) Smart grid technologies: communication technologies and standards. *IEEE T Ind Inform* 7: 529–539.
18. Hamilton B, Summy M (2011) Benefits of the smart grid [In My View]. *IEEE Power Energy M* 9: 104–112.
19. Ansari N (2013) CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid. *IEEE T Emerg Top Com* 1: 33–44.
20. Conti S, Corte AL, Nicolosi R, et al. (2016) Impact of cyber-physical system vulnerability, telecontrol system availability and islanding on distribution network reliability. *Sust Energ Grid Netw* 6: 143–151.
21. Davis KR, Davis CM, Zonouz SA, et al. (2015) A cyber-physical modeling and assessment framework for power grid infrastructures. *IEEE T Smart Grid* 6: 2464–2475.
22. Hug G, Giampapa JA (2012) Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks. *IEEE T Smart Grid* 3: 1362–1370.
23. Li X, Poor HV, Scaglione A (2016) Blind topology identification for power systems. IEEE International Conference on Smart Grid Communications. *IEEE*, 91–96.

24. Erseghe T, Tomasin S, Vigato A (2013) Topology estimation for smart micro grids via powerline communications. *IEEE T Signal Proces* 61: 3368–3377.
25. Zhao L, Song WZ, Tong L, et al. (2014) Topology identification in smart grid with limited measurements via convex optimization. Innovative Smart Grid Technologies-Asia. *IEEE*, 803–808.
26. Singh R, Manitsas E, Pal BC, et al. (2010) A recursive bayesian approach for identification of network configuration changes in distribution system state estimation. *IEEE T Power Syst* 25: 1329–1336.
27. Luan W, Peng J, Maras M, et al. (2015) Smart meter data analytics for distribution network connectivity verification. *IEEE T Smart Grid* 6: 1964–1971.
28. Chatzarakis GE (2009) Nodal analysis optimization based on the use of virtual current sources: A powerful new pedagogical method. *IEEE T Educ* 52: 144–150.
29. Phasor Advanced FAQ-Phasor-RTDMS: Available from: http://www.phasor-rtdms.com/phaserconcepts/phasor_adv_faq.html.
30. Vetter WJ (1975) Vector structures and solutions of linear matrix equations. *Linear Algebra & Its Applications* 10: 181–188.
31. Depablos J, Centeno V, Phadke AG, et al. (2004) Comparative testing of synchronized phasor measurement units. Power Engineering Society General Meeting. *IEEE* 1: 948–954.
32. Baran ME, Wu FF (1989) Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE T Power Deliver* 4: 1401–1407.

AIMS Press