Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Generalized persistent fault detection in distribution systems using network flow theory

Márton Greber * **Attila Fodor** * **Attila Magyar** *

* *Department of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem u. 10., Veszprém, H-8200, Hungary, ({greber.marton;fodor.attila;magyar.attila}@virt.uni-pannon.hu)*

Abstract: Persistent faults are steady state anomalies with a magnitude which does not necessary trigger general protective gear. It is present in various types of distribution networks, as leak in pipe networks or as high-impedance fault in electric systems. As smart meters come into general use, distribution systems are upgraded to have advanced metering infrastructure which can be used for diagnostic purposes. Different kind of detection methods are presented in different physical domains therefore comparison is cumbersome. The main achievement of the presented work is the formulation of an abstract system description, in order to tackle problems from various domains on a common ground. The notions of the well established general network theory are being used as a solid foundation. In this framework a general extension of flow networks is presented for distribution systems with measurement data available. Detecting faults at metered points is tackled in the literature, this problem is translated into the proposed representation. On the other hand a novel problem is described, the fault at unknown location between two metered points. The applicability of the abstract description to specialized distribution systems is presented through a simple case study.

Keywords: Distribution networks; Smart power applications; Steady-state errors ; Losses; Networks

1. INTRODUCTION

Distribution networks have the soul purpose to deliver resources from some supply entity to end consumers. According to the type of the delivered resource, the networks take different appearances in the real world. Water and gas are transported through pipes, electricity is delivered through conductors. Effects from various sources like weather, geographical location or human activity cause distribution lines to be vulnerable to faults. Persistent faults are steady state type faults. They can last relatively long because the magnitude of the faults is not big enough to trigger general protective gear.

Persistent faults can take many shapes according to the physical system. In electrical power distribution systems high impedance faults (HIF) can happen when at some point the distribution conductor is grounded through a high impedance, (Ghaderi et al. (2017)). Another example is the theft of electricity, also called non-technical losses (NTL). In this case live wires are manipulated to tap electricity or the measurement equipment is tampered, in order to hide consumption (Viegas et al. (2017)). Since more current needs to be supplied in order to keep existing demand satisfied, both cases lead to excessive load of the line and heat build up, in extreme case fire and blackout. In case of pipe network there are two main definitions called unaccounted for gas (UFG) and non-revenue water (NRW), both of these cover the same topics, just in different technological context. In case of both, gas and water, resources are lost before reaching end customers. This is influenced by many factors like leakages, measurement variations and meter tampering, in case of gases there are also emissions . Real world case studies show, that leakage plays a considerable role both in UFG (Shafiq et al. (2018)) and NRW (Kanakoudis and Muhammetoglu (2013)).

Distribution systems nowadays are getting more and more support in terms of measurement equipment. Advanced metering infrastructure (AMI) relies on two way communication between the utility and the consumers (Jha et al. (2014)). This is enabled through smart meters, which are measurement devices capable of communication, to provide consumption data periodically. According to the type of distribution network, these are prominent in both electrical (Bahmanyar et al. (2016)), and water distribution networks (Singapore (2016)). This data can also be used to perform various diagnostics on the network.

In the case of NTL the general power balance equations need to be modified in order to account for wasted load. Using the extended expressions it was noted that observing and comparing the voltage drop on line segments is an indicator, and can be used for NTL detection (Bula et al. (2016)). Since distribution networks can have large sizes, the question of structural decomposition and splitting a

complex search into smaller units is an important task. By incorporating parameter and measurement uncertainty into the model, the real world applicability of the developed greatly increased (Pózna et al. (2019)).
In case of water distribution networks, leakage detection can be formulated as an optimization problem where the difference between measured and estimated pressures is minimized. By incorporating a hydraulic simulator, leakage scenarios are generated to locate leaks (Sousa et al. (2015)). Until now, deficit was only assumed on metered points of the network. Another question is how to detect leaks between two metered points of the network, lying on a given pipe section. A steady state algorithm was developed using pipeline dynamics, in order to test arbitrary points of a pipe section for leaks. This is achieved by a developed finite-difference approximation of the pipeline dynamics (Lizarraga-Raygoza et al. (2018)).
It can be seen that solutions for different aspects of the problem are already well established. However they are scattered in the different physical domains. On the other hand all of these distribution systems are governed by the same fundamental principles, for example the conservation of energy or the general Kirchoff's laws. Flow networks are mathematical constructs involving graphs, and associating special properties to them (Iri (1969)). It is an already well established part of applied graph theory, capable of serving as a foundation to this task. The main focus of this work is to describe general distribution system concepts in this abstract representation. Main emphasis is on the formal extension of flow networks, such that faults from various technological systems can be translated into it.

2. NETWORK FLOW

Graphs are mathematical constructs, which allow problems to be translated into a structure consisting of vertices and edges. However in order to represent real world transportation networks in this context, the notion of graphs can be extended. By associating general attributes to vertices and edges, the resulting structure is called a flow network (Rockafellar (1998)).

2.1 General notions

In general a graph(G) is defined by three main components, vertices, edges and the connections between them:

$$G = (V, E, \partial^+, \partial^-), \tag{1}$$

where V denotes the set of vertices $V = \{v_1, v_2, \ldots, v_m\}$, E represents the set of edges $E = \{e_1, e_2, \ldots, e_n\}$ and the connections are described as incidence relations: $\partial^+ : E \to V$, $\partial^- : E \to V$. Throughout this work, directed graphs are used, which means that all the edges have directions. A given edge e, is pointing from vertex $\partial^+(e)$ to vertex $\partial^-(e)$ (see Fig. 1.).

A flow network (N) consists of a graph, which is coupled with special attributes (Iri (1996)):

$$N = (G, \xi, \eta, \zeta, f_1, \ldots, f_n), \tag{2}$$

where ξ is the edge flow $\xi : E \to \mathbb{R}$, η defines the edge tension $\eta : E \to \mathbb{R}$ and the relation between those two quantities is described by some branch characteristic function: $\eta_i = f_i(\xi_i)$ (see Fig. 1.). One vertex can be chosen as a reference point (v_{ref}) for measurements. This way we can associate potential values to vertices: $\zeta : V \to \mathbb{R}$, which represent the tension between some arbitrary vertex and the reference node. For a given edge, if one of the vertices is v_{ref} the branch characteristic description can be written as: $\zeta_i = f_i(\xi_i)$.

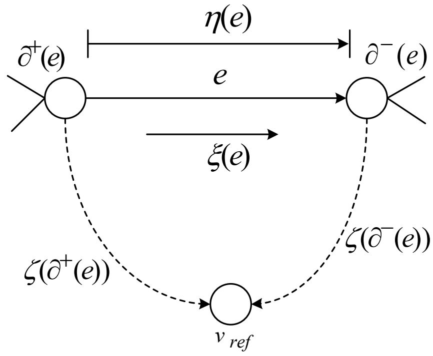

Fig. 1. Flow network notations

2.2 Distribution network flow

According to the function of a given edge, in distribution networks three main component categories can be distinguished. Resources are being fed into the network through supply participants. These are transported through distribution line elements, to reach end consumer vertices. In order to give bounds to the topologies, the direction of these edges are fixed according to the rules below. Consumer edges point from arbitrary non-reference vertices to the reference vertex:

$$E_c = \{e \mid \partial^+(e) \in V \setminus v_{ref},\ \partial^-(e) = v_{ref}\}. \tag{3}$$

Supply edges point from the reference node to some arbitrary non-reference vertices:

$$E_s = \{e \mid \partial^+(e) = v_{ref},\ \partial^-(e) \in V \setminus v_{ref}\}. \tag{4}$$

Distribution line edges point from some non-reference vertices to some arbitrary non-reference vertices:

$$E_l = \{e \mid \partial^+(e) \in V \setminus v_{ref},\ \partial^-(e) \in V \setminus v_{ref}\}. \tag{5}$$

The edge set for a distribution system flow network graph is defined as the union of the above mentioned sets:

$$E = E_c \cup E_s \cup E_l. \tag{6}$$

Taking a distribution flow network, the general problem statement is the following, given

- Customer flow desires: $\xi(e),\ \forall e \in E_c$
- Supply at a given potential level: $\zeta(\partial^-(e)),\ \forall e \in E_s$
- Distribution system parameters: $f_1, \ldots, f_n$

solve for:

- Changes in potential levels: $\zeta(v),\ \forall v \in V \setminus v_{ref}$
- Line flows: $\xi(e),\ \forall e \in E_l$.

For later use, the solution to this problem statement is described by the function $\mathcal{F}$.

$$\mathcal{F} : (N, v) \to \mathbb{R} \tag{7}$$

This takes a flow network N, solves the above described distribution network flow problem, and for a given vertex v it returns the calculated $\zeta(v)$ potential.

3. MEASURED FLOW NETWORK

The distribution flow network can be a projection of a real system. In order to incorporate measurement data for diagnostic purposes a similar but new concept is defined as measured flow network:

$$\tilde{N} = (G, \tilde{\xi}, \tilde{\zeta}, f_1, \ldots, f_n), \tag{8}$$

where $\tilde{\zeta}$ is the measured potential $\tilde{\zeta} : V \rightarrow \mathbb{R}$ and $\tilde{\xi}$ is the measured flow $\tilde{\xi} : E \rightarrow \mathbb{R}$. This measured quantities represent different categories of real world measurement sources. Potential is measured at the non-reference vertices of consumers and supply points:

$$\tilde{\zeta}(v),\ v \in \{\partial^+(e) \mid e \in\ E_c\} \cup \{\partial^-(e) \mid e \in E_s\} \tag{9}$$

Flow is measured on consumer and supply edges of the network:

$$\tilde{\xi}(e),\ e \in E_c \cup E_s \tag{10}$$

4. PERSISTENT FAULT AT REGISTERED POINTS

4.1 Description

In case of fault situations regarding only registered points, the graph of the flow network and the graph of the measured flow network are identical. This means that the feeder and all the customers are measured points. If there is persistent fault, there is flow deficit, meaning that the registered amount of consumption is less than the supplied flow quantity.

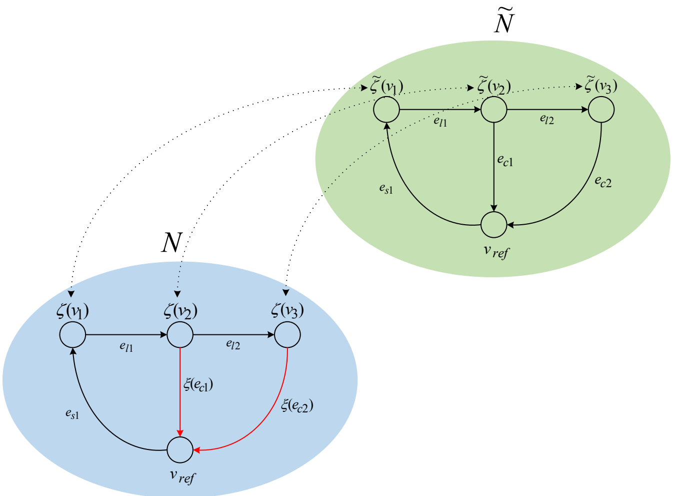

Fig. 2. Example flow network compared to measured configuration

In Fig. 2. a simple flow network is depicted. There is one supply and two consumer elements. In case of a fault situation, the task is to determine the consumed flows to such an extent, that the potential values of the measured network flow structure are reproduced. The flow values in N are the variables which need to be determined (see red edges in Fig. 2.).

4.2 Detection

We have a flow network (N) and an associated measured network configuration ($\tilde{N}$). In order to verify the measurement data of $\tilde{N}$, take the measured flows $\tilde{\xi}$, plug them into the flow network N, and solve the network flow problem. This will provide the potential levels, according to the flow measurements. If the calculated potential values doesn't match the measured potential levels:

$$\exists v \in V \setminus v_{ref} \rightarrow \zeta(v) \neq \tilde{\zeta}(v), \tag{11}$$

the network has deficit. If the supplied amount of flow for a network doesn't equals the consumed amount of flow, it is another indicator of a faulty network:

$$\sum\{\tilde{\xi}(e) \mid e \in E_s\} \neq \sum\{\tilde{\xi}(e) \mid e \in E_c\}. \tag{12}$$

In order to determine the valid consumer flow values the problem can be formulated as follows: minimize the difference between measured and calculated potential levels, by adjusting the flow network's customer flows ξ_{E_c}.

$$\arg\min_{\xi_{E_c}} \frac{1}{2} \sum \left\{ \left(\mathcal{F}(N(\xi_{E_c}), v) - \tilde{\zeta}(v) \right)^2 \mid \forall v \in V \right\} \tag{13}$$

5. PERSISTENT FAULT AT NON-METERED POINTS

5.1 Description

Faults can happen between metered points, which are new faulty nodes and are not present in the previous configuration. The goal is still the same, reproducing the potential measurement values from $\tilde{N}$, by adjusting the consumer flow values in N. In order to represent these possible fault locations, the graph of N needs to be extended. The inserted node and edge properties are the unknowns, which must be adjusted to reproduce $\tilde{\zeta}$.

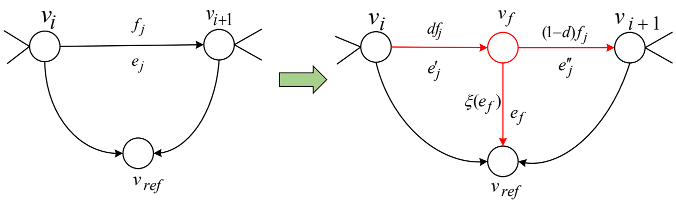

Fig. 3. Diagnostic node insertion

5.2 Detection

In order to be able to formulate the fault detection problem, the creation of the extended network definition is necessary. For every $e \in E_l$ line element, a new hypothetical fault node is inserted, by cutting in half the given edge. To represent the distance of the fault, from adjacent nodes the dividing point $d \in \{x \mid 0 < x < 1\}$ is introduced. Spatial relation from the two nodes is represented by dividing the branch characteristic using d. The faulty flow values are modelled as a new vertex between the fault node v_f and v_{ref}, having $\xi(e_f)$ flow rate.

The diagnostic model needs are extended, in order to be able to calculate with non-metered points. Therefore all the distribution line elements must be changed, resulting in a set of new parameters. Let V_f denote the vertex set of all the inserted fault nodes $V_f = \{v_{f1}, v_{f2}, \ldots, v_{fk}\}$, E_f the appropriate fault flow edges $E_f = \{e_{f1}, e_{f2}, \ldots, e_{fk}\}$. Existing line elements are cut into two pieces, for every

$e_l \in E_l$, two vertices are created $e_l^{'}$, $e_l^{''}$. The extended line segment set $\hat{E}_l$ is defined as:

$$\hat{E}_l = \{e_1^{'}, e_1^{''}, e_2^{'}, e_2^{''}, \ldots, e_n^{'}, e_n^{''}\}. \tag{14}$$

From a flow network perspective the fault flow edges in E_f can be treated as consumer edges. The extended consumer edge set $\hat{E}_c$, is the union of E_c and the fault edges: $\hat{E}_c = E_c \cup E_f$. In order to extend an existing G graph, the following steps can be used:

(1) given an existing $e \in E_l$ line element with $\partial^+(e)$ and $\partial^-(e)$ end nodes
(2) do the cutting, e will become e' and e'', and insert the hypothetical fault node v_f:
   - $\partial^+(e') = \partial^+(e)$ and $\partial^-(e') = v_f$
   - $\partial^+(e'') = v_f$ and $\partial^-(e'') = \partial^-(e)$
(3) insert the edge e_f representing the deficit flow: $\partial^+(e_f) = v_f$ and $\partial^-(e_f) = v_{ref}$

Using this preliminary statements the extended diagnostic graph can be defined:

$$\hat{G} = (\hat{V}, \hat{E}, \partial^+, \partial^-), \tag{15}$$

where $\hat{V}$ is the extended vertex set $\hat{V} = V \cup V_f$ and $\hat{E}$ is the extended edge set $\hat{E} = E_s \cup \hat{E}_c \cup \hat{E}_l$.

Using this extended graph notation, the extension for the network flow can be constructed:

$$\hat{N} = (\hat{G}, \hat{\xi}, \hat{\zeta}, D, f_1, \ldots, f_n), \tag{16}$$

where D is the set of division points, $D = \{d_1, d_2, \ldots, d_k\}$. Flow network calculation on the extended network is defined in the same way as it was defined before for a general N.

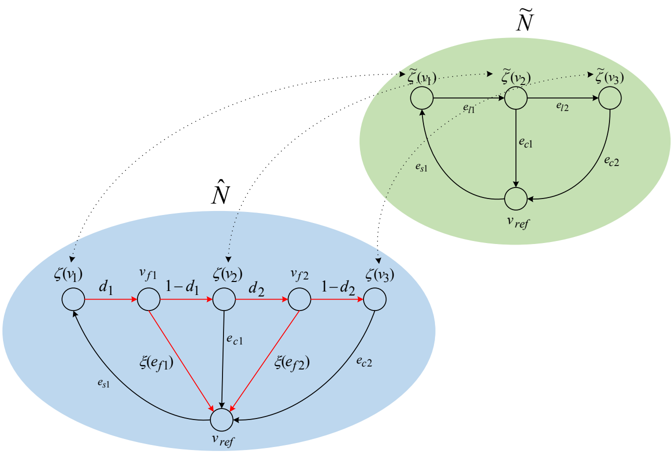

Fig. 4. Extended network compared to measured flow configuration

The optimization problem which lies under this fault detection task is similar to the registered point case. However now the graphs of the calculated and the measured networks are different. The values of the division points and the theft flows needs to be adjusted in such a way that on the metered points $v \in V$ the difference between measured and calculated potential values is minimized. By denoting the set of flows associated to fault edges by $\hat{\xi}_{E_f} = \{\hat{\xi}(e) \mid e \in E_f\}$, the optimization problem can be formulated as follows:

$$\arg\min_{\hat{\xi}_{E_f}, D} \frac{1}{2} \sum \left\{ \left(\mathcal{F}(\hat{N}(\hat{\xi}_{E_f}, D), v) - \tilde{\zeta}(v) \right)^2 \mid, \forall v \in V \right\}$$
$$\text{s.t.: } \sum\{\hat{\xi}(e) \mid e \in E_s\} = \sum\{\hat{\xi}(e) \mid e \in \hat{E}_c\} \tag{17}$$

5.3 Genetic algorithm

Throughout network flow calculations the solution method consists of using Kirchoff's laws and the branch characteristics. As stated before, consumption flows are an input parameters for the network flow, these are incorporated in the Kirchoff's laws. The second components are branch characteristics, for obtaining the results in potentials. In the optimization formulation above, both the branch characteristics, and part of the network consumptions are variables. Therefore the problem is of non-linear nature, with high number of unknowns. However the structure of the problem offers some opportunities to tailor the optimization method for this specific problem.

Genetic algorithms (GA) are optimization procedures inspired by the evolution theory of Darwin. For a given problem some solution candidates are generated, and the goodness of them is evaluated. This is a generation, and members are transferred into an upcoming generation according to some rules (Sivanandam and Deepa (2008)):

- elitism: some percent of the top ranked members are automatically transferred,
- recombination: new solutions are created by combining the properties of existing candidates,
- mutation: the solutions are tweaked in a random fashion .

In the new generation the goodness is evaluated again, the the procedure continues until some threshold value is reached or the maximum generation limit is exceeded. Genetic algorithms offer a robust method even for hard optimization problems, since the solution algorithm is almost the same in every case. However as with all soft computing techniques it doesn't necessary give exact optima, most of the time close to optimal solutions are reached.

In case of evolutionary optimization the two main tasks are to determine the solution(chromosome) representation, and the measure of goodness, the fitness function. By distributing the contents of the unknown sets into one vector a simple chromosome representation is obtained:

$$p_i = [\xi(e_1),\ d_1, \xi(e_2),\ d_2, \ldots, \xi(e_k),\ d_k]. \tag{18}$$

As the measure of the fitness the optimization expression of equation 17. is used.

5.4 Two-stage evolutionary optimization

As mentioned before, the fault detection optimization problem is highly non-linear and since all the distribution edges are cut into half, the search space is enormous. Therefore a single run genetic algorithm would require a high population size, with a high number of generations. Experimental runs showed that it is still not enough to pinpoint a fault location. What the results showed, so called hot-spots of the fault were identified. Meaning that fault flows were inserted in some neighbourhood of the actual fault location. The idea is to run two consequent optimization runs, where the first is to approximate the fault hot-spots, and the second to pinpoint accurate location and flow values. The effect of this two-stage GA is the implicit restriction of the search space during the whole process.

In the first stage the initial population $P = [p_1,\ p_2, \dots]$ is created in a totally random manner. Meaning that the entries of each p_i are randomized, but in such a way that the constraint of equation 17. is taken into account. By taking the supply flows, and subtracting the metered consumption, the amount of deficit flow can be determined:

$$\xi_\Delta = \sum\{\tilde{\xi}(e) \mid e \in E_s\} - \sum\{\tilde{\xi}(e) \mid e \in E_c\}. \qquad (19)$$

The only thing to take care is that the sum of the fault flow values in a given chromosome are equal to ξ_Δ. After executing the GA, using the incidence relations and the best member of the final population the number of hot-spots, and the cumulative fault flow in that region can be determined. We omit accurate location approximation and fault flow detection, by using a lower population and generation count, but the process is accelerated.

In the second stage, the initial population is generated using the hot-spot informations from the first run. Let us have h number of regions of interest, and the local cumulative fault flows are denoted by $\xi_i^H,\ i \in \{1, \dots, h\}$. The initialization is done, by assigning random flow indexes to each ξ_i^H in each chromosome. In this step the search space is heavily reduced in the flow dimensions, by using the cumulative flow values the real task is to determine the correct set of division points.

6. CASE STUDY

A simple case study example is created in order to illustrate the methodology. The topology of the network is presented in Fig. 5. This is an electrical network with radial layout. The feeder is at node 1, the reference vertex is v_{22}.

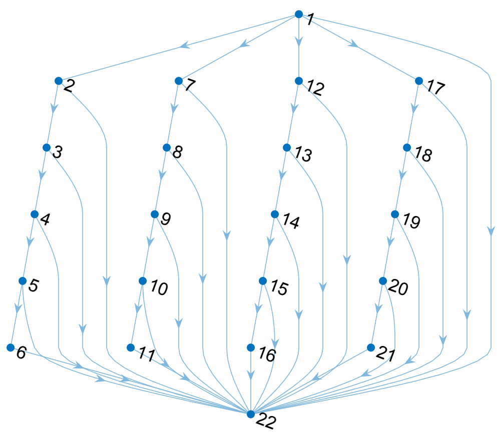

Fig. 5. Case study network

The feeder point is modelled as a constant voltage source, consumer edges are constant current sources and connection line elements are resistive components. This is a homogeneous test case, meaning that all the line elements and all the consumer elements are of equal value. The line elements have a resistance of 0.1Ω and all the consumers are $1A$. The feeder point supplies the network as a $230V$ voltage source.

In order to use the methodology presented in the previous sections, the electrical network must be translated into that framework. Network currents are equal to edge flows: $I \rightsquigarrow \xi$, and electrical potential is equal to network flow potential: $\phi \rightsquigarrow \zeta$. The translated distribution system properties are:

- Customer flow desires: $\xi(e) = 1,\ \forall e \in E_c$.
- Supply at a given potential level: $\zeta(\partial^-(e)) = 230,\ \forall e \in E_s$.
- Distribution system line parameters: $\eta_i = 0.1\xi_i$

Resistive networks are basic electrical systems, which are covered by Ohm's law:

$$U = RI, \qquad (20)$$

where U is voltage, R is resistance and I is current. This is a linear network, by using node quantities the flow problem can be written as $\boldsymbol{\phi} = \boldsymbol{RI}$, where $\boldsymbol{\phi} \in \mathbb{R}^m$ is the potential vector, $\boldsymbol{R} \in \mathbb{R}^{m\times m}$ is the nodal resistance matrix and $\boldsymbol{I} \in \mathbb{R}^m$ is the nodal current injection vector. The network flow calculation $\mathcal{F}$ is thereby a linear problem. The following simulations were performed using MATLAB where the methodology was implemented.

6.1 Fault at registered point

Let us suppose that we have the above introduced test case, and persistent faults are only allowed at registered points. In order to generate measurement data, a simulation was run $(\tilde{N})$, where the consumption connected to nodes 9 and 20 is changed from 1 to 2 and 3.5. The resulting potential values are stored for $\tilde{\zeta}$. The task is to find the faulty nodes, utilizing the measurement potential values.

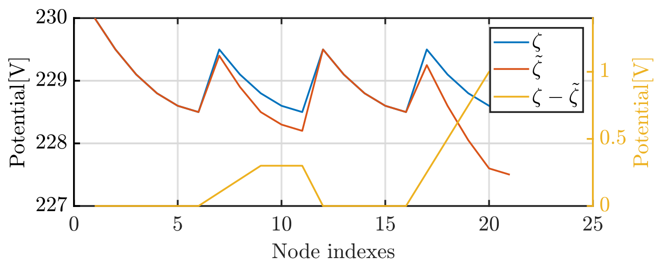

Fig. 6. Potential values

We run the network flow calculations and plot every potential value (Fig. 6. blue line). Next we observe the available measurement data, which is visualized with red in Fig. 6. It can be seen that there is a substantial difference between them. This satisfies the deficit criteria state in equation 11. The faults can be detected using linear least-squares formulation, since the network flow calculation of the underlying DC network is a linear problem. The inbuilt function lsqlin() in MATLAB was used in order to solve this problem, since it can solve tasks in the form of:

$$\min_x \frac{1}{2}||Cx - d||_2^2, \qquad (21)$$

where Cx represents the linear flow calculation and d the measured quantity. Running the optimization function, and plotting the results, the faulty nodes can be observed, alongside with the respective fault flows(see Fig. 7.)

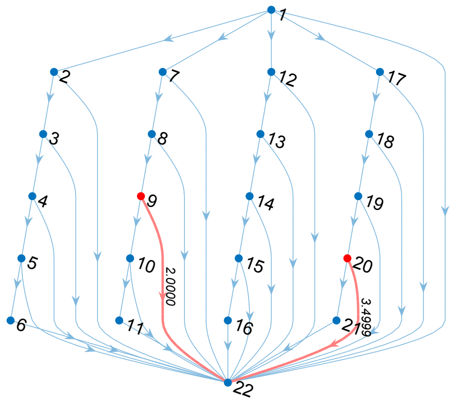

Fig. 7. The network with the detected fault nodes

6.2 Fault at non-metered points

Two faults are inserted also in this test case. However, now they are inserted between two existing vertices. One fault was inserted between nodes 9 and 10, with a dividing point of 0.331 and a fault flow of 3.3. The second fault lies between nodes 20 and 21, with a dividing point of 0.218 and a fault flow of 1.7. in order to generate measurement data, these faulty points are inserted into the graph. Potential measurements are taken from vertices $2, \ldots, 21$, and stored in a measurement flow network $\tilde{N}$. Using the nominal consumption the expected potential profile is calculated by solving the linear network flow problem. The two sets are compared, and since there is a difference, one can begin with the detection algorithm.

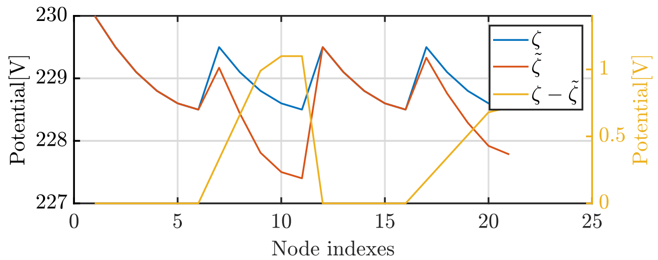

Fig. 8. Potential values

The solution is reached using the two-stage evolutionary algorithm. In both stages a population size of 5000, a maximum generation number of 50 is used. The rate of elite selection is 20%, whereas the mutation is set to 2%. For reproduction, single-point crossover is utilized. In order to ensure the optimization constraint of equation 17. a penalty function was introduced. The fitness of the chromosomes is calculated through the optimization expression. If the sum of the overall incorporated hypothetical flows is different from the deficit flow ξ_Δ, the fitness function is multiplied by some tunable input factor.

Table 1. illustrates simulation results for a two-stage evolutionary optimization run. s and t represent source and target vertex indexes, which serve as an identifies to the parameters between two registered nodes. ξ_i and d_i are the arguments of the optimization which are represented in a chromosome of a population.

Table 1. Simulation results

		I. stage		II. stage	
s	t	d_i	ξ_i	d_i	ξ_i
1	2	0.08827	0.5	0.83127	0
2	3	0.49962	0	0.85381	0
3	4	0.34439	0.1	0.91921	0
4	5	0.88343	0	0.97254	0
5	6	0.67107	0.2	0.20960	0
1	7	0.97987	0	0.34053	0
7	8	0.87945	0.6	0.24550	0
8	9	0.81779	0.4	0.18217	0
9	10	0.56802	0.7	0.23536	3.24880
10	11	0.56116	0.3	0.77113	0
1	12	0.02425	0.3	0.26139	0
12	13	0.96596	0.1	0.89373	0
13	14	0.14662	0	0.25992	0
14	15	0.85332	0	0.42411	0
15	16	0.48808	0.1	0.19582	0
1	17	0.12770	0.2	0.03282	0
17	18	0.59779	0.5	0.06854	0
18	19	0.92438	0.3	0.68289	0
19	20	0.67647	0.3	0.34518	1.75120
20	21	0.61723	0.4	0.43659	0

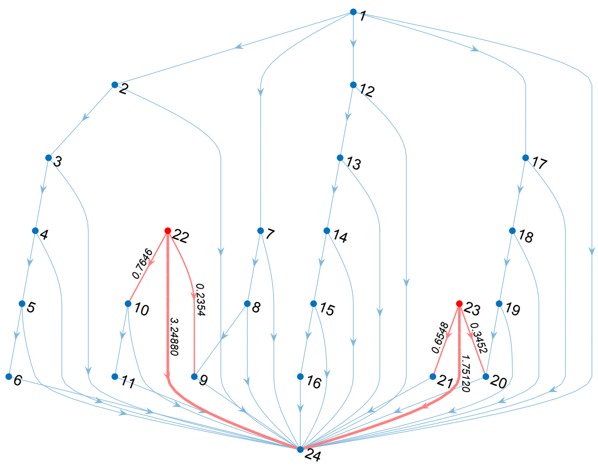

Fig. 9. The network with the detected fault nodes

The variable ξ_Δ was found to be 5, using equation 19. In the first stage this amount was distributed randomly in the initial population generation, as well as the dividing points. The results of this part can be found in the I. stage column. Taking a look at the network graph vertex numbering, it can be seen that the first five rows of the table are for the first radial branch, as seen from the left (Fig. 5.), the second five entries represent the second branch from the left and so on. It can be seen that there is a hot-spot between vertices 9 and 10, since there is a local fault flow maxima. Using the same principle, the edge, having end nodes 17 and 18 can also be identified as a local hot-spot area. Taking the sum of the fault currents in each local neighbourhood we can find the local cumulative flows: $\xi^H = \{2,\ 1.7\}$. In the second stage the initialization is done, by spreading these values randomly. The results of this run are shown in the II. Stage column of Table. 1. The average absolute flow error is 0.0512, and the average absolute division point error is 0.11125. Finally the faults are visualized on a partially extended graph (see Fig. 9.) containing only the detected non-registered faulty vertices.

7. CONCLUSION

The extensions of the general network flow theory has been accomplished in the article, in order to handle measurement data, and extensive diagnostic methods. Since the method uses first engineering principles, it can describe pipe networks, like water, gas and also electric power networks. It can serve as a common language for fault detection studies ranging through different physical systems.

In the presented paper the main emphasis has been the detection of persistent faults. It has been noted that persistent faults are present in different kind of distribution systems. In order to have a common language for detection research the theory of network flow has been used. Since distribution systems follow certain rules with respect to the topology, it has been incorporated into the formal definitions. The notion of measured flow network has been introduced, in order to incorporate smart meter measurement data into the diagnostic methods. The fault phenomena have been divided into two sections: persistent fault at registered points, and non-metered points. Both problems have been translated into least squares optimization problems. In the first case according to the system equations, linear or non-linear least squares optimization solves the problem. However, in the case of non-metered points the flow networks need to be extended. This novel flow network extension has been used to formulate this problem. This resulted in a hard optimization problem, where solution was achieved through a two-stage evolutionary algorithm. The methodology has been illustrated through a simple case study. It has been shown, that detection methods formulated in the flow network reference system can be translated to specialized physical systems. The research regarding those system archetypes can be brought closer using the proposed methodology, since it offers a common language.

ACKNOWLEDGEMENTS

Márton Greber was supported by the ÚNKP-19-2 new national excellence program of the Ministry for Innovation and Technology. A. Magyar was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Attila Magyar was supported by the ÚNKP-19-4 new national excellence program of the Ministry for Innovation and Technology.

REFERENCES

Bahmanyar, A., Jamali, S., Estebsari, A., Pons, E., Bompard, E., Patti, E., and Acquaviva, A. (2016). Emerging smart meters in electrical distribution systems: Opportunities and challenges. In *2016 24th Iranian Conference on Electrical Engineering (ICEE)*, 1082–1087. doi:10.1109/IranianCEE.2016.7585682.

Bula, I., Hoxha, V., Shala, M., and Hajrizi, E. (2016). Minimizing non-technical losses with point-to-point measurement of voltage drop between "smart" meters. *IFAC-PapersOnLine*, 49(29), 206 – 211. doi: https://doi.org/10.1016/j.ifacol.2016.11.103. 17th IFAC Conference on International Stability, Technology and Culture TECIS 2016.

Ghaderi, A., Ginn, H.L., and Mohammadpour, H.A. (2017). High impedance fault detection: A review. *Electric Power Systems Research*, 143, 376 – 388. doi: https://doi.org/10.1016/j.epsr.2016.10.021.

Iri, M. (1969). Network flow, transportation, and scheduling; theory and algorithms. *Mathematics in Science and Engineering*, 57.

Iri, M. (1996). Network flow — theory and applications with practical impact. In *System Modelling and Optimization*, 24–36. Springer US. doi:10.1007/978-0-387-34897-1_3.

Jha, I.S., Sen, S., and Agarwal, V. (2014). Advanced metering infrastructure analytics — a case study. In *2014 Eighteenth National Power Systems Conference (NPSC)*, 1–6. doi:10.1109/NPSC.2014.7103882.

Kanakoudis, V. and Muhammetoglu, H. (2013). Urban water pipe networks management towards non-revenue water reduction: Two case studies from greece and turkey. *CLEAN - Soil, Air, Water*, 42(7), 880–892. doi: 10.1002/clen.201300138.

Lizarraga-Raygoza, A., Delgado-Aguinaga, J., and Begovich, O. (2018). Steady state algorithm for leak diagnosis in water pipeline systems. *IFAC-PapersOnLine*, 51(13), 402 – 407. doi: https://doi.org/10.1016/j.ifacol.2018.07.312. 2nd IFAC Conference on Modelling, Identification and Control of Nonlinear Systems MICNON 2018.

Pózna, A.I., Fodor, A., and Hangos, K.M. (2019). Model-based fault detection and isolation of non-technical losses in electrical networks. *Mathematical and Computer Modelling of Dynamical Systems*, 25(4), 397–428. doi:10.1080/13873954.2019.1655066.

Rockafellar, R.T. (1998). *Network Flows and Monotropic Optimization*. Athena Scientific.

Shafiq, M., Nisar, W.B., Savino, M.M., Rashid, Z., and Ahmad, Z. (2018). Monitoring and controlling of unaccounted for gas (ufg) in distribution networks: A case study of sui northern gas pipelines limited pakistan. *IFAC-PapersOnLine*, 51(11), 253 – 258. doi: https://doi.org/10.1016/j.ifacol.2018.08.284. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

Singapore, P.U.B. (2016). Managing the water distribution network with a smart water grid. 1(1). doi: 10.1186/s40713-016-0004-4.

Sivanandam, S. and Deepa, S.N. (2008). *Introduction to Genetic Algorithms*. Springer Berlin Heidelberg. doi: 10.1007/978-3-540-73190-0.

Sousa, J., Ribeiro, L., Muranho, J., and Marques, A.S. (2015). Locating leaks in water distribution networks with simulated annealing and graph theory. *Procedia Engineering*, 119, 63 – 71. doi: https://doi.org/10.1016/j.proeng.2015.08.854. Computing and Control for the Water Industry (CCWI2015) Sharing the best practice in water management.

Viegas, J.L., Esteves, P.R., Melício, R., Mendes, V., and Vieira, S.M. (2017). Solutions for detection of non-technical losses in the electricity grid: A review. *Renewable and Sustainable Energy Reviews*, 80, 1256 – 1268. doi:https://doi.org/10.1016/j.rser.2017.05.193.