# Data-Driven Event Diagnosis in Transmission Systems With Incomplete and Conflicting Alarms Given Sensor Malfunctions

Yazhou Jiang, *Member, IEEE*, and Anurag K. Srivastava, *Senior Member, IEEE*

***Abstract*—Accurate fault event diagnosis with incomplete and conflicting alarms given sensor malfunctions is a challenging problem for power system operators. To solve this problem, this study proposes a data-driven approach based on Mixed Integer Linear Programming (MILP) for fast determination of fault event scenarios with uncertainties. The uncertainties include failures and malfunction of relays and circuit breakers (CBs) as well as incomplete/incorrect sensor alarms at the control center. To improve the accuracy for fault event diagnosis, redundant alarms from multiple sources, i.e., Phasor Measurement Units (PMUs), Supervisory Control and Data Acquisition (SCADA), and Sequence of Events Recorders (SERs) are jointly used in this study. The temporal correlation of sensor alarms is incorporated in the constraints of the MILP model. The resulting data-driven algorithm determines the most credible fault scenario that is well supported by the available sensor alarms at the control center. Simulation results of the IEEE 14-bus system, the synthetic South Carolina 500-bus system, and a real-world complex event scenario demonstrate the effectiveness of the proposed approach for accurate and efficient fault event diagnosis.**

***Index Terms*—Alarm message, analytical model, data-driven, fault diagnosis, mixed integer linear programming, outage management, power system protection.**

## NOMENCLATURE

| | |
|---|---|
| MP | Main protection |
| PBP | Primary backup protection |
| SBP | Secondary backup protection |
| BFP | Breaker failure protection |
| M | Large number |
| $r_i^{MP}$, $r_i^{PBP}$, $r_i^{SBP}$, $r_i^{BFP}$ | MP, PBP, SBP, BFP of relay $i$ |
| $R_i^{MP}, R_i^{PBP}, R_i^{SBP}, R_i^{BFP}$ | MP, PBP, SBP, BFP alarm from relay $i$ |
| $f_{r_i^{MP}}, f_{r_i^{PBP}}, f_{r_i^{SBP}}, f_{r_i^{BFP}}, f_{CB_i}, f_{CB_i}^{PMU}$ | Decision variables to indicate the failure of MP, PBP, SBP, BFP of relay $i$, $CB_i$ and PMU |
| $m_{r_i^{MP}}, m_{r_i^{PBP}}, m_{r_i^{SBP}}, m_{r_i^{BFP}}, m_{CB_i}, m_{CB_i}^{PMU}$ | Decision variables to indicate malfunction of MP, PBP, SBP, BFP of relay $i$, $CB_i$ and PMU |
| $t_{R_i}^{MP}, t_{R_i}^{PBP}, t_{R_i}^{SBP}, t_{R_i}^{BFP}, t_{CB_i}$ | Time tags of MP, PBP, SBP, BFP and $CB_i$ status alarms |
| $x_i^{MP}, x_i^{PBP}, x_i^{SBP}, x_i^{BFP}, x_{CB_i}$ | Decision variables to denote incorrect time tags of MP, PBP, SBP, BFP of relay $i$ and $CB_i$ alarm |
| $\varepsilon_{MP}, \varepsilon_{PBP}, \varepsilon_{SBP}, \varepsilon_{BFP}, \varepsilon_{CB}$, | Timing tolerance for MP, PBP, SBP, BFP and CB alarms |
| $N_{Fail.}$, $N_{Malf.}$, $N_{Incor.}$ | No. of alarm failures, malfunctions, and incorrect time tags |
| $\omega_1$, $\omega_2, \omega_3$ | Weighting factors for alarm failures, malfunctions, and incorrect time tags |
| $t_k^{ftl}$ | Inception time of event $k$ |
| $r_i^{pilot}$ | Pilot protection signal of $L_j$ |
| $\Delta T_{PBP}, \Delta T_{SBP}, \Delta t_{BFP}$ | Protection coordination time of PBP and SBP, and breaker failure detection time of BFP |
| $R_i^{CB}$ | Decision variable to denote the alarm that trips $CB_i$ |
| $CB_i^{PMU}$ | $CB_i$ status alarm from PMU |
| $\Omega_s, \Omega_{Line}$ | Set of sensors and transmission lines |
| $y_{CB_i}^{MP}, y_{CB_i}^{PBP}, y_{CB_i}$ | Statuses of $CB_i$ after MP and/or PBP activates, and the final status of $CB_i$ |

## I. INTRODUCTION

FAULT event diagnosis has been widely recognized to be critical for reliable operation of power systems [1], [2]. A fault and the resulting actions of protective relays and CBs generally lead to a power outage. Over the last decades, significant efforts have been devoted to development of technologies and tools for advanced fault event diagnosis for applications in Energy Management Systems (EMS) at the control center. As a result, considerate progresses have been made and the engineering practice for fault diagnosis has gradually evolved from the domain knowledge-based analysis to the most recent analytical model-based techniques [3], [4].

Manuscript received December 1, 2018; revised May 8, 2019 and September 8, 2019; accepted October 12, 2019. Date of publication October 16, 2019; date of current version January 22, 2020. This work was supported by U.S. National Science Foundation under Awards CNS-1329666 and CNS-1840192. Paper no. TPWRD-01472-2018. *(Corresponding author: Yazhou Jiang.)*

Y. Jiang was with Washington State University, Pullman, WA 99164 USA. He is now with the GE Global Research Center, Niskayuna, NY 12309 USA (e-mail: yazhou.jiang@ge.com).

A. K. Srivastava is with Washington State University, Pullman, WA 99164 USA (e-mail: anurag.k.srivastava@wsu.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2019.2947671

Fault diagnosis is a primary task for system operators and protection engineers when a power outage occurs. System operators need to infer the faulted component and sequential actions of protective relays from the available sensor alarms at the control center. In an outage, failures or malfunctions of relays and CBs may occur, and sensor alarms may be delayed or missing due to communication issues. Accurate identification of these complex fault scenarios with sensor malfunctions requires extensive domain knowledge of power systems and its protection configuration. To assist system operators with decision-making of fault event diagnosis, various classifications of technologies, i.e., Knowledge Based Systems (KBSs), Model Based Systems (MBSs), and Artificial Intelligent Systems (AISs) have been developed. KBSs have been proposed in [5]–[7] for determination of the fault scenario by comparing actual event messages with simulated messages. MBSs have been elaborated in [8] to represent correlation between physical power system components and event messages. However, a KBS or an MBS relies heavily on the knowledge modules and the expertise of system models for an accurate fault event diagnosis. AISs such as Petri Nets [9], [10], Neural Network [11], and Bayesian networks [12] have also been proposed for fault diagnosis. The recent study is focused on analytical methodologies to diagnose the fault event of power systems [3], [4]. The causality of a faulty component and the resulting alarms is modeled in an analytical way to determine the most credible fault scenario that is well supported by the available sensor alarms at the control center. Tools based on analytical methodologies have been developed to assist system operators with decision-making. A tool named Generalized Alarm Analysis Module (GAAM) has been integrated into EMS at the control center in Italy [13], [14]. Note that Ref. [13] is based on multiple hypothesis analysis by hypothesizing the fault scenario and calculating its credibility. The hypothesis with the highest credibility is deemed to be faulty. The drawback of these methodologies is a large number of hypotheses and the given hypotheses may not capture the true fault scenario. Moreover, the temporal correlation of sensor alarms and the multiple sources of alarms have not been considered. Other tools have been developed and tested in the regional system of China [4]. These tools use Sequential Event Recorders (SERs) or SCADA alarms including CB statuses for fault diagnosis. To improve the accuracy, methods incorporating the time tags of alarms are developed in [4]. A five-digit algorithm based on high fidelity data from newly installed PMUs is introduced in [15] to determine the faulted transmission line.

While significant progress has been made in development of advanced methodologies and tools for fault event diagnosis in transmission systems, it is worth noting that power system is highly nonlinear and fault diagnosis with uncertainty is a complex problem. Indeed, the state-of-the-art analytical techniques and tools are primarily based on multiple-hypotheses analysis [4], [14], [16]. Each component in the outage area is hypothesized to be faulty and the causality of the hypothesized faulty component and activated relays is modeled for determination of the most credible fault event. For example, a methodology based on multiple-hypothesis analysis is proposed in [16] for outage management of distribution systems incorporating information from smart meters. Ref. [16] is used to identify the outage area and the faulted line sections based on a radial structure of distribution feeder circuits. The number of smart fault indicators with a failure or malfunction and the pair number of a missed protection coordination of recloser-fuse or fuse-fuse in distribution systems are first hypothesized and the credibility for each hypothesis is quantified by how well it is supported by data from smart meters and smart fault indicators. The detailed protection coordination has not been modeled and the temporal correlations of sensor alarms have not been addressed. In the analytical models for event diagnosis of transmission systems, the constraints are nonlinear as reported in [4]. The nonlinearity and the resulting complexity together with numerous hypotheses make it hard to be solved in an efficient manner. Heuristic such as Tabu Search [4], particle swarm optimization [17] or Generic algorithm (GA) [18], has to be deployed as solution methodologies for nonlinear analytical models. To the best of the authors' knowledge, the state-of-the-art research on analytical models for fault event diagnosis, especially with incomplete and conflicting alarms, is inadequate in the following aspects:

- A large number of hypotheses have to be generated to capture the possible faulted component and the resulting actions of protective relays and CBs. The number of hypotheses increases exponentially with regard to the number of components involved in the outage area.
- For each hypothesis, nonlinear analytical techniques are used to model the causality of the faulted component and the resulting actions of protective relays and CBs with uncertainties. Nonlinearity leads to additional complexity and makes it hard for on-line applications.
- Heuristic is employed to solve the nonlinear analytical models for determination of the most credible fault event scenario that is well supported by the available sensor alarms at the control center. The global optimality of the optimization is not guaranteed, and the true fault scenario may be missed due to suboptimality.
- Existing studies use alarms from SERs, SCADA alarms, or PMUs to diagnose the fault. Technologies to incorporate redundant alarms from multiple sources in an automatic manner are not available.
- The available technology has not fully modeled the temporal correlation of sensor alarms for an accurate fault event diagnosis.

Transmission utilities are faced with the challenge of how to leverage sensor data available for system operators to better support the grid operation with a better accuracy. To bridge these gaps of the existing solutions, this study proposes a new data-driven approach to leverage alarms from multiple sources such as SCADA, PMU, SERs for event diagnosis considering incorrect and incomplete sensor alarms as well as temporal correlations of alarms given sensor malfunctions. Key contributions of the proposed approach are: 1) data-driven approach by leveraging sensor alarms from multiple sources including SERs, PMUs and SCADA data for accurate fault event diagnosis with uncertainties; 2) integrating the data driven approach with an analytical MILP model that considers the failures or malfunctions of protective relays and CBs as well as incomplete/incorrect alarms. The most credible fault event scenario is determined by minimizing the discrepancy of the determined scenario with the available sensor alarms while respecting the principle of protection systems. The temporal correlation of sensor alarms is explicitly incorporated in the analytical model to capture the temporal abnormality of actions of protective relays and/or CBs; 3) testing and validation of the proposed approach using the IEEE-14 bus system, the synthetic South Carolina 500-bus system, and a real-world complex event scenario.

## II. Fault Diagnosis With Uncertainties

When a permanent fault occurs in power systems, relays at substations are configured to detect the fault in a timely manner and trip open the CBs for fast fault isolation. In this process, relays may be incorrectly configured, and CBs may fail to operate upon receiving a tripping signal. These uncertainties together with missing sensor alarms and uncertainties of transmission line parameters significantly complicate the process of fault event diagnosis. To ensure a fault to be isolated in a timely manner, redundant protection configurations such as pilot protection, distance relays, and breaker failure relays, are widely deployed in transmission systems. For instance, many utilities select pilot protection as MP for transmission lines with a voltage level of 345 kV or above due to its fast response to isolate the fault while distance relays are used as a backup. When the MP fails to isolate the fault, the backup protection, i.e., PBP and SBP, is expected to operate to open the CB. The statuses of CBs are sent back to the control center through SCADA for system operators to diagnose the fault event scenario. In the meanwhile, sequential actions of relays and CBs with timing tags are recorded by SERs at the substation, which can be acquired by EMS in an online manner. Some newly installed PMUs at substations also monitor and send CB statuses and estimated phasors to the control center through Phasor Data Concentrators (PDCs). It is worth noting that ‘M’ type measurement class PMU typically uses a longer window to estimate phasors compared to ‘P’ type protection class PMUs as part of relays or standalone PMUs following IEEE C37.118 and IEEE Test Suite Specification (TSS). The additional delay in ‘M’ type will not impact the performance of the proposed algorithm as phasor will be still timestamped in middle of the window and time synchronized in PDC. Moreover, delay is typically in fraction of seconds for ‘M’ type of PMUs, which will not impact the proposed applications with several seconds/minutes timeline. PMU data is used after the event to compute the fault status and information while transient response is not required for this. Additionally, as discussed with engineers from utility companies such as New York Independent System Operator (NYISO) and American Electric Power (AEP), PMUs can also be customized to monitor the status of circuit breakers using user defined bits. Different from SERs which use the local substation clock, PMUs data at different substations are synchronized by using the Global Positioning System (GPS) with an error less than one millisecond. Alarms from multiple resources, i.e., SERs, SCADA, and PMUs, are jointly used in this study for accurate fault event diagnosis with uncertainties.

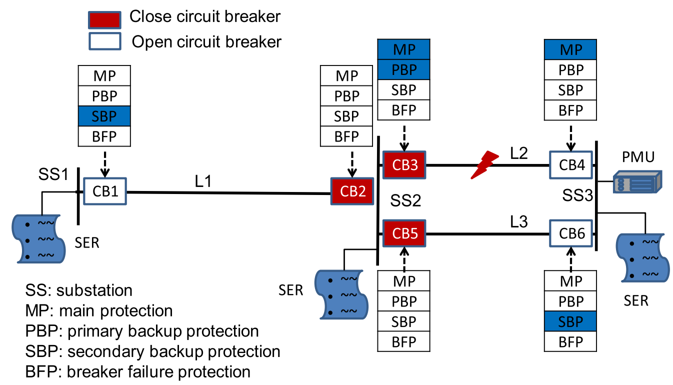


Fig. 1. Simplified power grid with protection systems.

TABLE I
Alarms at the Control Center for Fault Diagnosis

| Alarm Source | Time tags | Received alarms |
|---|---|---|
| SERs | 14:2:21:100 | MP of CB4 operates |
| | 14:2:21:134 | CB4 trips |
| | 14:2:21:601 | SBP of CB1 operates |
| | 14:2:21:602 | SBP of CB6 operates |
| | 14:2:21:635 | CB1 and CB6 trip |
| PMU | 14:2:21:134 | CB4 trips |
| | 14:2:21:635 | CB6 trips |
| SCADA | | CB1, CB4, and CB6 open |

A simplified power system as shown in Fig. 1 is used for illustration of the protection system. Suppose a fault occurs on line L2 and MPs of CB4 and CB3 are expected to trip the CBs instantaneously without any intentional delays. If CB4 opens successfully while MP fails to trip open CB3 due to the incorrect configuration, PBP of CB3 is configured to trip open it with a given delay. If the tripping signal does not transmit successfully to the CB due to an incorrect configuration or others, SBPs of CB1 and CB6 are designed to trip CB1 and CB6 respectively to isolate the fault. The sequential actions of relays and CBs are recorded by SERs. In Fig. 1, the PMU installed at Sub3 monitors the status of CB4 and CB6. For the given fault scenario, the alarms received from SERs, SCADA, and PMU(s) are summarized in Table I for fault event diagnosis. Note that the time tags of SCADA alarms are not considered since the time tags are usually added on using the computer clock at the control center when the alarms arrive.

## III. Fault Diagnosis Process

The framework of the proposed methodology for fault event diagnosis is shown in Fig. 2. Collected sensor alarms from SCADA, PMUs and SERs are first used to determine the outage area by the proposed methodologies in [13] or [19]. Ref. [19] is focused on the system topology determination for applications of state estimation. The causality of a fault and the consequential actions from relays and circuit breakers for event diagnosis has not been addressed in [19]. The sensor data together with system knowledge such as power network topology and protection configuration serve as the input to the proposed MILP optimization

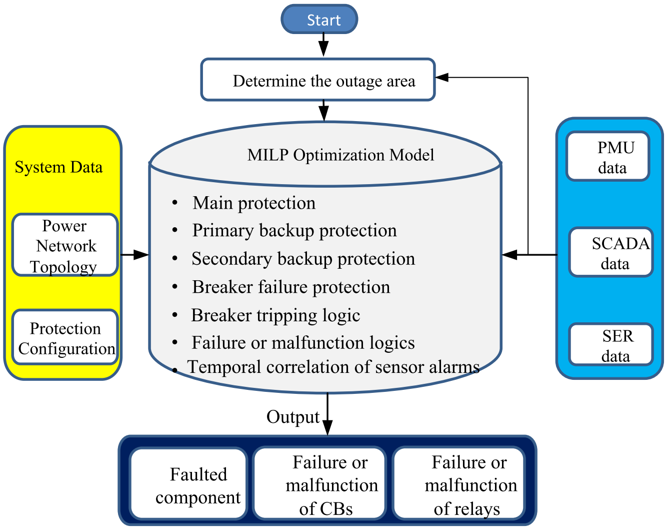


Fig. 2 Fault diagnosis framework.

for fault event diagnosis. Included in the proposed approach is also the temporal correlation of alarms while respecting protection operating logics for determination of the most credible fault event scenario that is well supported by the available alarms. The output is the event diagnosis report including faulted component, failure or malfunction of CBs & relays.

## IV. MILP Model for Fault Event Diagnosis

### A. Assumptions

In North America, transmission networks are usually equipped with distance relays with 3-zone protection configuration, pilot protection, and breaker failure relays. Pilot protection and Zone1 protection of distance relays are regarded as MP of a transmission line while Zone2 and Zone3 protections are regarded as the PBP and SBP, respectively. In this study, the assumption is made as follows for the purpose of illustration:

- Faults are isolated by the redundant protection systems, i.e., MP, PBP, SBP and BFP.

The redundant protection system is configured to isolate faults while accommodating potential failures from relays or breakers. In addition, the proposed models can be applied to diagnose any electric assets such as transmission lines, bus bars, and transformers with different protection configurations.

### B. Objective Function

In this study, a sensor indicates a relay or a CB. If the alarm from a sensor is not available but expected at the control center, it will be defined as a failure of the sensor. As such, the sensor failure includes physical device failures, missing or delayed sensor alarms. On the other hand, if a sensor alarm is not expected while it is available, it is defined as a malfunction of the sensor. This study considers multiple failures and anomalies for event diagnosis. The objective function is to minimize the discrepancy between the estimated sensor alarms and available ones at the control center for fault diagnosis. The discrepancy is quantified by weighted summation of the total no. of sensor failures, malfunction, and incorrect timing tags. That is

$$\text{Min} \quad \omega_1 * N_{Fail.} + \omega_2 * N_{Malf.} + \omega_3 * N_{Incor.} \tag{1}$$

For transmission utilities, the historical data can be used to determine the weighting factors. For example, if n1 out of N1 sensor failures occur in history, $\omega_1$ is determined to be (N1-n1)/N1. The same procedure can be applied to $\omega_2$, and $\omega_3$.

### C. Constraints

The constraints are meant to model the causality of the faulty component and the resulting actions of relays & CBs with uncertainties while respecting the principles of configured protection systems. The temporal correlation of time tags of redundant sensor alarms is also incorporated in the constraints of the proposed optimization.

*1) Constraints of MP:* Suppose that $r_i^{MP}$ is the MP of line $L_j$. $r_i^{MP}$ is expected to operate when a fault occurs at Zone1 protection distance of $L_j$ or at the Zone2 protection region and a pilot protection signal $r_i^{pilot}$ from the remote end of $L_j$ is also received. The alarm $R_i^{MP}$ should be received when $r_i^{MP}$ operates properly without failures or $r_i^{MP}$ should not operate while it malfunctions to trip. This logical relationship is modeled in Eq. (2) as

$$\begin{cases} r_i^{MP} = L_j^{Z1} \vee \left(L_j^{Z2} \wedge r_i^{pilot}\right) \\ R_i^{MP} = \left(r_i^{MP} \wedge \sim f_{r_i^{MP}}\right) \vee \left(\sim r_i^{MP} \wedge m_{r_i^{MP}}\right) \end{cases} \tag{2}$$

Since MP protection is designed to operate instantaneously without intentional delays upon occurrence of a fault, time tag $t_{r_i}^{MP}$ is expected to be within the time interval given in (3) as

$$\begin{cases} -x_i^{MP} * M - \varepsilon_{MP} \le R_i^{MP} * \left(t_{R_i}^{MP} - t_k^{ftl}\right) \\ R_i^{MP} * \left(t_{r_i}^{MP} - t_k^{ftl}\right) \le \varepsilon_{MP} + x_i^{MP} * M \end{cases} \tag{3}$$

*2) Constraints of PBP:* Suppose that $r_i^{PBP}$ is the PBP of line $L_j$ and is associated with $CB_i$. PBP is expected to trip when *i)* $r_i^{PBP}$ detects a fault on $L_j$; *ii)* its setting time for coordination with MP is released; *iii)* $CB_i$ is not open by MP; *iv)* there is no breaker failure. Alarm $R_i^{PBP}$ is available when $r_i^{PBP}$ operates properly without failures or it should not operate while it malfunctions to trip. That is

$$\begin{cases} r_i^{PBP} = L_j \wedge \sim y_{CB_i}^{MP} \wedge \sim f_{CB_i}^{MP} \\ f_{CB_i}^{MP} = \sim y_{CB_i}^{MP} \wedge R_i^{MP} \\ R_i^{PBP} = \left(r_i^{PBP} \wedge \sim f_{r_i^{PBP}}\right) \vee \left(\sim r_i^{PBP} \wedge m_{r_i^{PBP}}\right) \end{cases} \tag{4}$$

The time tag of alarm $R_i^{PBP}$ is expected to be within the range from $t_0 + \Delta T_{PBP} - \varepsilon_{PBP}$ to $t_0 + \Delta T_{PBP} + \varepsilon_{PBP}$ as

$$\begin{cases} -x_i^{PBP} * M - \varepsilon_{PBP} \le R_i^{PBP} * \left(t_{R_i}^{PBP} - t_k^{ftl} - \Delta T_{PBP}\right) \\ R_i^{PBP} * \left(t_{R_i}^{PBP} - t_k^{ftl} - \Delta T_{PBP}\right) \le \varepsilon_{PBP} + x_i^{PBP} * M \end{cases} \tag{5}$$

*3) Constraints of SBP:* Use $r_i^{SBP}$ to denote the SBP of line $L_j$. $CB_k$ is associated with MP and PBP designed to isolate $L_j$ if it is faulted. SBP is expected to issue a tripping signal when $r_i^{SBP}$ detects a fault and the time for coordination with PBP is released. The underlying logic is that neither MP or PBP has successfully tripped open $CB_k$ and there is no breaker failure of $CB_k$. In a similar fashion, the alarm $R_i^{SBP}$ from $r_i^{SBP}$ is available under these two scenarios: a) $r_i^{SBP}$ operates properly without failures; b) $r_i^{SBP}$ malfunctions to report a fault. That is

$$\begin{cases} r_i^{SBP} = L_j \wedge \sim y_{CB_k}^{MP} \wedge \sim y_{CB_k}^{PBP} \wedge \sim f_{CB_k}^{MP} \wedge \sim f_{CB_k}^{PBP} \\ f_{CB_k}^{PBP} = \sim y_{CB_k}^{PBP} \wedge R_k^{PBP} \\ R_i^{SBP} = \left(r_i^{SBP} \wedge \sim f_{r_i^{SBP}}\right) \vee \left(\sim r_i^{SBP} \wedge m_{r_i^{SBP}}\right) \end{cases} \tag{6}$$

The time of the tripping signal from SBP is expected to be in the interval $[t_0 + \Delta T_{SBP} - \varepsilon_{SBP}, t_0 + \Delta T_{SBP} + \varepsilon_{SBP}]$ as

$$\begin{cases} -x_i^{SBP} * M - \varepsilon_{SBP} \leq R_i^{SBP} * \left(t_{R_i}^{SBP} - \Delta T_{SBP} - t_k^{ftl}\right) \\ R_i^{SBP} * \left(t_{R_i}^{SBP} - \Delta T_{SBP} - t_k^{ftl}\right) \leq \varepsilon_{SBP} + x_i^{SBP} * M \end{cases} \tag{7}$$

*4) Constraints of BFP:* Let $r_i^{BFP}$ be the BFP of $CB_i$. BFP is expected to operate when $CB_i$ receives a tripping signal from its associated relays such as MP, PBP, and SBP while it fails to open its mechanical contact. If $r_i^{BFP}$ operates properly without failures or it malfunctions to report a breaker failure, the alarm $R_i^{BFP}$ will be available for fault diagnosis. The underlying logic is

$$\begin{cases} R_i = R_i^{MP} \vee R_i^{PBP} \vee R_i^{SBP} \\ r_i^{BFP} = R_i \wedge \sim y_{CB_i} \\ R_i^{BFP} = \left(r_i^{BFP} \wedge \sim f_{r_i^{BFP}}\right) \vee \left(\sim r_i^{BFP} \wedge m_{r_i^{BFP}}\right) \end{cases} \tag{8}$$

The tripping time of the BFR falls in the given range as

$$\begin{cases} -x_i^{BFP} * M - \varepsilon_{BFP} \leq R_i^{BFP} * \left(t_{R_i}^{BFP} - \Delta T_i^{BFP} - t_k^{ftl}\right) \\ R_i^{BFP} * \left(t_{R_i}^{BFP} - \Delta T_i^{BFP} - t_k^{ftl}\right) \leq \varepsilon_{BFP} + x_i^{SBP} * M \end{cases} \tag{9}$$

where $\Delta T_i^{BFP} = \Delta t_{BFP} + R_i^{MP} * (t_{R_i}^{MP} - t_k^{ftl}) + R_i^{PBP} * (t_{R_i}^{PBP} - t_k^{ftl}) + R_i^{SBP} * (t_{R_i}^{SBP} - t_k^{ftl})$.

*5) Constraints of CB:* A CB is expected to open the mechanical contact when a tripping signal is received from its relays, i.e., MP, PBP, SBP, and BFP. The underlying logic is that a tripping signal is received, and the CB operates without failures; or no tripping signal is received but the breaker malfunctions. Suppose $y_{CB_i}$ indicates the availability of alarm from $CB_i$. The logic is

$$\begin{cases} R_i^{CB} = R_i \vee \left(\sum_{j\in\Omega_{CB_i}}^{\vee} R_j^{BFP}\right) \\ y_{CB_i} = \left(R_i^{CB} \wedge \sim f_{CB_i}\right) \vee \left(\sim R_i^{CB} \wedge m_{CB_i}\right) \\ y_{CB_i} \geq y_{CB_i}^{MP} \\ y_{CB_i} \geq y_{CB_i}^{PBP} \end{cases} \tag{10}$$

The constraint of the opening time $t_{CB_i}$ of $CB_i$ is given as

$$\begin{cases} -x_{CB_i} * M - \varepsilon_{CB} \leq y_{CB_i} * \left(t_{CB_i} - \Delta T_{CB_i} - t_k^{ftl}\right) \\ y_{CB_i} * \left(t_{CB_i} - \Delta T_{CB_i} - t_k^{ftl}\right) \leq \varepsilon_{CB} + x_{CB_i} * M \end{cases} \tag{11}$$

where $\Delta T_{CB_i} = \sum_{j\in\Omega_{CB_i}} (R_j^{BFP} * \Delta t_{BFP}) + R_i^{MP} * (t_{R_i}^{MP} - t_k^{ftl}) + R_i^{PBP} * (t_{R_i}^{PBP} - t_k^{ftl}) + R_i^{SBP} * (t_{R_i}^{SBP} - t_k^{ftl}) + \Delta t_{CB}$.

*6) Constraint of Event Inception Time:* The fault inception time $t_k^{ftl}$ of a component is correlated with the decision variable of its status, which is modeled as:

$$t_k^{ftl} = L_j * t_{j0}^{ftl} \tag{12}$$

Since both $L_j$ and $t_{j0}^{ftl}$ are decision variables, (12) will be nonlinear. Using Big-M theory, (12) is converted into a linear constraint as

$$-1000 * (1 - L_j) \leq t_k^{ftl} - t_{j0}^{ftl} \leq 1000 * (1 - L_j) \tag{13}$$

*7) Constraint to Incorporate PMU Data:* When a PMU is installed in the substation, the statuses of CBs are monitored by the PMU. The logic is

$$\begin{cases} y_{CB_i} - CB_i^{PMU} \leq f_{CB_i}^{PMU} \\ CB_i^{PMU} - y_{CB_i} \leq m_{CB_i}^{PMU} \end{cases} \tag{14}$$

*8) Constraints to Count Alarm Discrepancy:* Sensor failures, sensor malfunctions, and incorrect alarm timing tags are counted as

$$\begin{cases} N_{fail.} = \sum_{i\in\Omega_s} (f_{r_i^{MP}} + f_{r_i^{PBP}} + f_{r_i^{SBP}} + f_{r_i^{BFP}} \\ \qquad + N_{CB_i} * f_{CB_i} + f_{CB_i}^{PMU}) \\ N_{Malf.} = \sum_{i\in\Omega_s} (m_{r_i^{MP}} + m_{r_i^{PBP}} + m_{r_i^{SBP}} + m_{r_i^{BFP}} \\ \qquad + N_{CB_i} * m_{CB_i} + m_{CB_i}^{PMU}) \\ N_{Incor.} = \sum\limits_{i\in\Omega_s} \left(x_i^{MP} + x_i^{PBP} + x_i^{SBP} + x_i^{BFP} + x_{CB_i}\right) \end{cases} \tag{15}$$

where $N_{CB_i}$ is used to denote the number of available alarms for each CB. For instance, if the open status alarm of $CB_i$ is available from SERs as well as the SCADA, $N_{CB_i} = 2$.

*9) Constraints of the Decision Variables:*

$$L_i = L_i^{Z1} + L_i^{Z2}, \forall i \in \Omega_{Line} \tag{16}$$

In this study, M is selected as 750 to impose constraints (3), (5), (7), (9), and (11) in the optimization to determine if the timing tag of an alarm is incorrect when the time tags use millisecond as the unit. To strategically select M, the worst scenario can be used. The worst fault scenario is that SBP acts to trip open a CB while the CB fails and the BFP trips open the CB instead. Given the protection coordination time of SBP, say 0.5 second, and BFP to detect a CB failure as 0.2 second, CB breaker opening time as two cycles, M can be any number larger than 0.734 second, which is equal to {t for SBP to operate} +{t to open CB contact}+{t for BFP to operate}+{t to open CB contact}. $\varepsilon_{\mathrm{MP}}$, $\varepsilon_{\mathrm{PBP}}$, and $\varepsilon_{\mathrm{SBP}}$, are selected to be 5 ms in this study and $\varepsilon_{\mathrm{BFP}}$ & $\varepsilon_{\mathrm{CB}}$ are selected as 34 ms. Essentially the error tolerant parameters are selected to account for clock inaccuracy or any timing issues for MP, PBP, SBP, BFP, CB to act when a fault occurs. A GPS can maintain a clock error to be within 1 ms. A digital relay may have 16, 32, or 64 samples each cycle depending on its configuration. To accommodate these uncertainties, $\varepsilon_{\mathrm{MP}}$, $\varepsilon_{\mathrm{PBP}}$, and $\varepsilon_{\mathrm{SBP}}$ are selected to be 5 ms, which is slightly larger than 2.04 ms, a summation of 16.7/16 ms and 1 ms. $\varepsilon_{\mathrm{BFP}}$ & $\varepsilon_{\mathrm{CB}}$ need to be selected to further accommodate the inaccuracy of breaker opening time. Since activation of a breaker is opening the physical contact, which may take one to two cycles. Therefore, the thresholds can be selected as 34 ms. In the meanwhile, system operators are also provided with the flexibility of relaxing the temporal correlations of alarms by setting $\varepsilon_{\mathrm{MP}}$, $\varepsilon_{\mathrm{PBP}}$, $\varepsilon_{\mathrm{SBP}}$, $\varepsilon_{\mathrm{BFP}}$ and $\varepsilon_{\mathrm{CB}}$ to a large number, such as 1000 second if desired.

### *D. Convert Nonlinear Logical Constraints Into a Linear Combination of Decision Variables*

Logical operation constraints can be converted into a linear combination of decision variables with the principles given as

*1) Conversion of logical "Or"constraints:*

$$w = g_1 \vee g_2 \vee \cdots \vee g_{n-1} \vee g_n$$

$$\Leftrightarrow \begin{cases} w \geq g_i, \forall i \in \{1,2,3,\ldots,n\} \\ w \leq g_1 + g_2 + \cdots + g_{n-1} + g_n \end{cases}$$

*2) Conversion of logical "And"constraints:*

$$v = g_1 \wedge g_2 \wedge \cdots \wedge g_{m-1} \wedge g_m$$

$$\Leftrightarrow \begin{cases} v \leq g_i, \forall i \in \{1,2,3,\ldots,m\} \\ v \geq g_1 + g_2 + \cdots + g_{m-1} + g_m - (m-1) \end{cases}$$

where *w*, $g_1, \ldots, g_n$, and *v* are binary decision variables.

Using these principles, the logical operation constraints (2), (4), (6), (8), and (10) are converted into a linear combination of decision variables. The details are given in the Appendix. It is worth noting that decision variables $h_{i1}, \ldots, h_{i11}$ add complexity into the optimization with an increased number of decision variables while assisting with the conversion into a MILP formulation. The simulation results in Section V demonstrate that the additional decision variables will not compromise the potential of the proposed approach for online applications given its highly computational performance and global optimality resulting from the intrinsic linearity. It is also worth noting that even though the principles have been proposed in [20], there are multiple novel contributions of the proposed work. This study developed a new algorithm by modeling the causality of the fault and the observed alarms as well as considering temporal correlations of alarms given sensor malfunctions for event diagnosis in transmission systems. In contrast, Ref. [20] is dedicated to modeling the spatial locations of smart fault indicators and line sections and using received data from fault indicators to estimate the faulty line sections without consideration of the protection system in distribution systems. The problems addressed in this study is different than the one addressed in authors' previous study [20]. To the best knowledge of the authors, this study is the first to propose an analytical model based on MILP for event diagnosis with incomplete and conflicting alarms given sensor malfunctions. The causality and temporal connections of observed alarms, while respecting the redundant protection coordination and possible failure mechanism are first modeled in the proposed optimization, which is novel compared to the state-of-the-art. And problem formulation and preparing set of equations to be solved by MILP is novel.

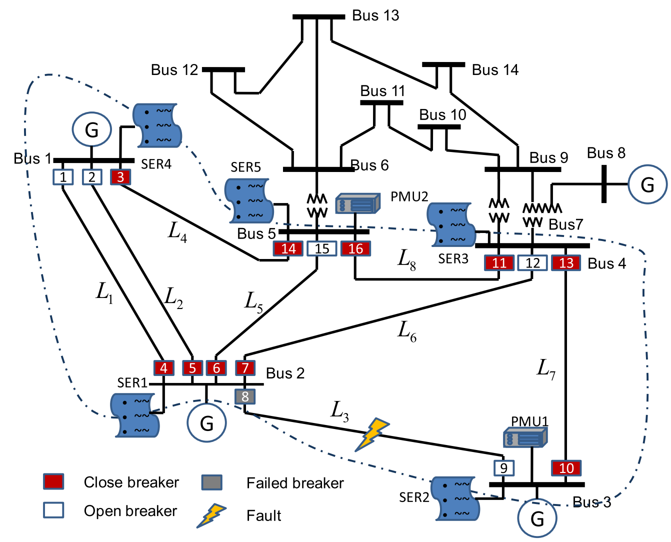


Fig. 3. Fault scenario for IEEE 14-bus system.

## V. Simulation Results

The proposed approach for fault event diagnosis is tested with different cases, i.e., IEEE 14-bus system [21], synthetic South Carolina 500-bus system [22], and real-world event scenario in [4]. Fault events including relay failures, breaker failures, missing/incorrect alarms and multiple faults, are used to demonstrate the effectiveness of the proposed approach for on-line applications. The IEEE 14-bus system and the synthetic South Carolina 500 bus system are modeled in Real-Time Digital Simulator (RTDS) to simulate the fault scenario. The optimization model is implemented in CPLEX 12.7.1 on a computer with i5-3340M CPU and 4 GB memory.

### *A. IEEE 14-Bus System*

Fault scenario:

- A fault occurs at $L_3$ at time 11:12:1:189;
- The MP of CB9 at the substation of Bus3 operated correctly and the CB tripped open instantaneously;
- The MP and PBP of CB8 at the substation of Bus2 did not operate due to a wrong configuration;
- The SBP of CB1, CB2, CB12, and CB15 operated and tripping signals were issued to open the CBs.
- CB1, CB2, CB12, and CB15 tripped open correctly while the timing tag for CB12 status was incorrectly recorded.

The outage area is determined as shown in the dashed line in Fig. 3. Eight transmission lines, sixteen CBs, and sixteen sets of relays including MP, PBP, SBP, and BFP are involved in the outage as given in Table II. Statuses of CBs 9, 10, 14, 15, and 16 are monitored by PMU1 and PMU2. SERs 1-5 at substations record the sequential actions of relays and CBs. The statuses of CBs are available at SCADA in the control center. The alarms from PMUs, SERs, and SCADA are shown in Table III. Note that the system operator determines the outage area based on alarms and measurements as reported in [13]. The SCADA

TABLE II
COMPONENTS IN OUTAGE AREA

| | |
|---|---|
| 8 Transmission Lines | $L_1, \ldots, L_8$ |
| 16 CBs | $y_{CB_1}, y_{CB_2}, \ldots, y_{CB_{15}}, y_{CB_{16}}$ |
| MP of CBs | $r_1^{MP}, r_2^{MP}, \ldots, r_{16}^{MP}$ |
| PBP of CBs | $r_1^{PBP}, r_2^{PBP}, \ldots, r_{16}^{PBP}$ |
| SBP of CBs | $r_1^{SBP}, r_2^{SBP}, \ldots, r_{16}^{SBP}$ |
| BFP of CBs | $r_1^{BFP}, r_2^{BFP}, \ldots, r_{16}^{BFP}$ |

TABLE III
ALARMS FROM MULTIPLE SOURCES

| Alarm Source | Time Tags | Alarms at Control Center |
|---|---|---|
| SERs | 11:12:1:200 | MP of CB9 operated |
| | 11:12:1:233 | CB9 operated to open |
| | 11:12:1:700 | SBP of CB12 operated |
| | 11:12:1:701 | SBP of CB1 and CB2 operated |
| | 11:12:1:702 | SBP of CB15 operated |
| | 11:12:1:734 | CB1 and CB2 open |
| | 11:12:1:735 | CB15 operated to open |
| | 11:12:1:835 | CB12 open |
| PMU | 11:12:1:233 | CB9 tripped open from PMU1 |
| | 11:12:1:735 | CB15 tripped open from PMU2 |
| SCADA | | CB1, CB2, CB9, CB12, CB15 open |

alarms indicate that CB1, CB2, CB9, CB12, and CB15 are open; and the MP of CB9 as well as the SBPs of CB1, CB2, CB12, and CB15 operated to open the CBs; From PMU1 and PMU2, CB9 and CB15 are open. The SERs recorded the sequential actions of CBs and relays. In the meanwhile, the timing tags from SER logs are leveraged in the optimization model. Use 11:12:1:200 as the reference time and the timing tags of alarms is given as

$$t_R^{MP} = [\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, 0, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset]$$
$$t_R^{SBP} = [0.501, 0.501, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, 0.5, \emptyset, \emptyset, 0.502, \emptyset]$$
$$t_{CB} = [0.534, 0.534, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, 0.033, \emptyset, \emptyset, 0.635, \emptyset, \emptyset, 0.535, \emptyset]$$

The received alarms are input into optimization as

$$R^{MP} = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]$$
$$y_{CB} = [1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0]$$
$$R^{SBP} = [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0]$$
$$N_{CB} = [2, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 2, 0]$$

$\emptyset$ is used to denote the unavailability of information. The weighting factors of $\omega_1$, $\omega_2$, and $\omega_3$ are set to be 1. Put the alarms and their timing tags together with the parameters as the input to the model. The optimization is executed, and it takes around 5 ms to solve the optimization and the decision variables are determined to be

$$L = [0, 0, 1, 0, 0, 0, 0, 0]$$
$$r^{MP} = [0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]$$
$$r^{SBP} = [1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0]$$

TABLE IV
FAULT DIAGNOSIS REPORT

| | | | |
|---|---|---|---|
| Fault occurrence t | 11:12:1:200 | Failed CB | None |
| Faulty Component | L3 | Malfunctioned CB | None |
| Failed Relay | $r_8^{MP}\&r_8^{PBP}$ | Missing alarm | None |
| Malfunctioned relay | None | Incorrect time tag | CB12 |

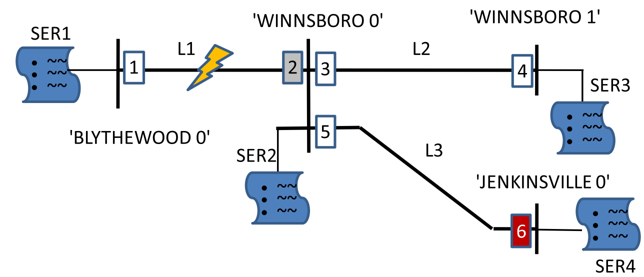


Fig. 4. South Carolina 500-bus system.

$$f_{r^{MP}} = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]$$
$$f_{r^{PBP}} = [0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]$$
$$x_{CB} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]$$

From the simulation results, the 3rd item of $L$ is "1", which indicates that L3 is the faulty transmission line and there is an MP failure and PBP failure of CB8 since the 8th items of $f_{r^{MP}}$ and $f_{r^{PBP}}$ are equal to "1". The timing tag of $CB_{12}$ from SERs is incorrect and CB1, CB2, CB9, CB12, and CB15 operated correctly to open. The fault diagnosis report is given in Table IV. It is demonstrated that the proposed algorithm can handle relay failures and incorrect time tags from the alarms. The linear optimization model is solved in 5 ms. The high computational performance and optimality of the proposed approach makes it potential for online applications for fault event diagnosis. Note that this study does not take the time for MP to trigger open CB into account. The fault inception time is determined to be 11:12:1:200 instead of the true fault inception time 11:12:1:189.

## B. Synthetic South Carolina 500-Bus System

The synthetic South Carolina 500-bus system is a representative power grid model derived from the public information with no confidential critical energy infrastructure information as described in [23], [24]. To validate the proposed approach in this study, a fault occurred on the transmission line L1 between the substations of "BLYTHEWOOD 0" and "WINNSBORO 0" at 8:32:11:348 and CBs 1, 3, 4, 5 opened to isolate the fault.

No substations involved in the outage are installed with PMUs. The alarms from SERs and SCADA are given in Table V. Using these alarms as input, the optimization model is solved in 11 ms and some key decision variables are given as

$$L = [1, 0, 0]$$
$$R = [1, 1, 0, 0, 0, 0]$$
$$R^{CB} = [1, 1, 1, 0, 1, 0]$$

TABLE V
ALARMS FROM MULTIPLE SOURCES

| Source | Time Tags | Received Alarms |
|---|---|---|
| SERs | 8:32:11:351 | MP of CB1 operated |
| | 8:32:11:352 | MP of CB2 operated |
| | 8:32:11:383 | CB1 tripped to open |
| | 8:32:11:551 | BFP of CB2 tripped |
| | 8:32:11:583 | CB3 and CB5 tripped open |
| SCADA | | Open CB1, CB3, CB4 and CB5; |

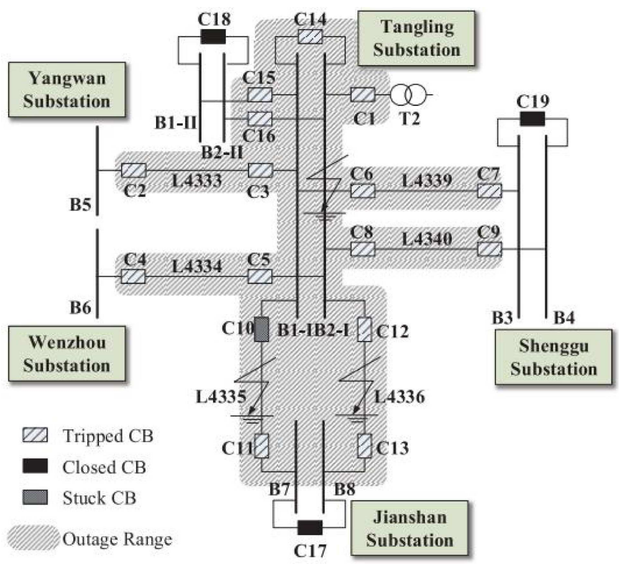


Fig. 5. Real-world fault scenario [4].

$$f_{CB} = [0, 1, 0, 0, 0, 0]$$
$$m_{CB} = [0, 0, 0, 1, 0, 0]$$
$$t^{ftl} = [0.5,\ 0,\ 0]$$

Since the reference time is 8:32:11:351, the fault occurrence time is inferred to be 8:32:11:351:5 from $t^{ftl}$ and the faulty line is L1 as inferred from the decision variable L. From the decision variable $f_{CB}$, there is a breaker failure of CB2. Since there is no evidence to indicate the action of the protection of CB4 while CB4 is reported to be open from SCADA, it is determined that CB4 malfunctioned to report the open status. From this scenario, the approach is demonstrated to handle breaker failures and incorrect alarms.

## C. Performance of the Algorithm Using Real-World Scenario

The real-world scenario from [4] is tested using the proposed algorithm. The true fault scenario is shown in Fig. 5. The fault scenario is briefly summarized as: faults occurred on L4335 first, L4336 30 ms later, and B2-I 340 ms later; the C10 was stuck to open and MP of L4336 at Tangling Substation was missing. The sequential actions of fault scenario have been discussed in detail in [4]. The components in the outage area is given in Table VI and the received alarms are shown in Table IX.

TABLE VI
COMPONENTS IN OUTAGE AREA

| | |
|---|---|
| L4333, L4334, L4335, L4336, L4339, L4340, B1-I, B2-I | $L_1, \ldots, L_6, L_7, L_8$ |
| C1, C2, …, C16 | $y_{CB_1}, \ldots, y_{CB_{15}}, y_{CB_{16}}$ |
| MP of CBs | $r_1^{MP}, r_2^{MP}, \ldots, r_{16}^{MP}$ |
| PBP of CBs | $r_1^{PBP}, r_2^{PBP}, \ldots, r_{16}^{PBP}$ |
| SBP of CBs | $r_1^{SBP}, r_2^{SBP}, \ldots, r_{16}^{SBP}$ |
| BFP of CBs | $r_1^{BFP}, r_2^{BFP}, \ldots, r_{16}^{BFP}$ |

To accommodate the complex scenarios, up to 4 faults are hypothesized to be in the outage area and the optimization is solved for each assumed scenario. The computational times for one-fault scenario, two-fault scenario, three-fault scenario, and four-fault scenario are 570 ms, 740 ms, 510 ms, 590 ms, respectively. The results are shown in Table X. From the simulation results, for the hypothetical scenario with one fault, the no. of discrepancy $N_{Disc.}$, which is the summation of $N_{Fail.}$, $N_{Malf.}$ and $N_{Incor.}$, is calculated to be 32. $N_{Disc.}$ for the two-fault scenario, three-fault scenario, and four-fault scenario is calculated to be 10, 6, 6, respectively. From the simulation results, it is inferred that when three sequential faults, which are L4335@3.25 ms, L4336@30 ms, B2-I@340.5 ms, occur, the discrepancy between the expected and received alarms is at the minimum value. Therefore, the three-fault scenario is the most credible to be the true scenario. It is also deduced that MP of L4336@Tangling failed since the alarm is determined to be missing, and there is a breaker failure of C10. Confidence on the result can be further computed by fit between hypothesis and converged optimization scenarios by following equations:

$$\text{Trust Score} = 1 - N_{Discp.}/N_{Alarm}$$

where $N_{Discp.}$ is the number of discrepancies in each hypothesized scenario from the converged optimization and $N_{Alarm}$ is the number of total alarms available for fault diagnosis.

For a scenario with a complete alarm set without failures or malfunction, Trust Score will be 1. The score represents the confidence level of the determined scenario to the ground truth. By applying the confidence level, the event diagnosis results for the real-world scenario are updated for each hypothesized number of faults in the scenario. It clearly shows that for the hypothesis with a single fault in the system, the score value is 0.47, which indicates a low credibility of the determined scenario to be the ground truth. In the meanwhile, the credibility for the scenarios with three and four faults is the same while it is less likely for more faults occurred in the scenario what led to the same observed alarms and measurement sets. It can be concluded that the trust score value can effectively reflect the confidence of the fault diagnosis results by reflecting the incorrect and incomplete alarms and measurement in the collected data set. As a result, the proposed approach provides the decision-making of event diagnosis for transmission utilities.

### *D. Comparison of the Analytical Model-Based Event Diagnosis Methodologies*

Analytical model-based event diagnosis methodologies in [4] and [14] have been implemented at EMS. From the simulation results claimed in the literature, both algorithms can support the system operators and protection engineers to diagnose the event. The pros and cons of the proposed approach compared to the two algorithms are summarized in Table XI. Assume there are $n$ components, $m$ relays, $k$ CBs in the outage region. In the meanwhile, the maximum no. of faults in the outage region is assumed to be $f$. $m$, $n$ and $f$ are not directly correlated while their values are dependent upon the outage area and the system protection configurations. In this study, $f$ is assumed to be limited to 4 for the real-world scenario as indicated in Table X. The proposed approach outperforms the state-of-the-art by significantly reducing the hypotheses, handling complex fault scenarios, the linearity of the proposed model for event diagnosis, global optimality and incorporation of the temporal correlation of alarms. The result is a highly computational methodology for accurate fault event diagnosis in an online application. It is worthy to further clarify the comparison of the proposed approach to References [4] and [14] as:

- the proposed methodology is essentially a MILP model and the linearity of the approach ensures the "Global optimality". Indeed, the proposed technology is a data driven approach and the effectiveness and accuracy of the proposed approach is based on the completeness and quality of the data. As demonstrated in the study, the proposed methodology already has the capability to handle incomplete or incorrect alarms to some extent which indicates that not all components are required to be measured by devices such as PMUs, SERs, or SCADA. If too many alarms and measurements data are missing, results will not converge or will converge with less confidence as indicated by trust score.
- the fewer hypotheses will enhance the computational performance of the proposed approach targeted for online applications and reduce the efforts from system operators in decision-making. Ideal will be considering all possible hypotheses and intelligent pruning for computationally efficient performance for online applications. If intelligent pruning or analysis is not done, more hypotheses will result in longer computational time.
- the proposed methodology is a MILP model while Reference [4] is based on nonlinear analytical models and heuristic techniques such as Tabu searching have to be employed as solution methodologies. Essentially, all the constraints from Reference [4] have been or can be included in the proposed MILP methodology while developing a linear formulation and the optimization can be directly solved by commercial MILP solvers. Therefore, the proposed MILP approach offers clear differences for better computational performance and higher chances of convergence and robustness even with increasing problem complexity while not compromising the accuracy than that in Reference [4].

TABLE VII
ALARMS RECEIVED FOR FAULT DIAGNOSIS

| Source | Time Tags | Received Alarms |
|---|---|---|
| SERs | 10:12:21:121 | MP of CB1 operated |
| | 10:12:21:153 | CB1 tripped open |
| | 10:12:21:624 | SBP of CB4 operated |
| | 10:12:21:657 | CB4 tripped open |
| SCADA | | CB1, CB2, CB4 are open |

TABLE VIII
IMPACT OF WEIGHTING FACTORS ON FAULT DIAGNOSIS RESULTS

| Weighting Factors | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|---|---|---|---|---|---|---|
| | 1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 |
| Fault occurrence t | 10:12:21:121 | | | 10:12:21:121 | | |
| Faulty Component | L1 | | | L1 | | |
| Failed Relay | $r_2^{MP}$ | | | $r_2^{MP}\&r_2^{PBP}\&r_2^{BFP}\&r_6^{SBP}$ | | |
| Malfunctioned relay | $r_4^{SBP}$ | | | Ø | | |
| Failed CB | Ø | | | Ø | | |
| Malfunctioned CB | CB4 | | | Ø | | |
| Incorrect time tag | Ø | | | Ø | | |

### *E. Impact of Weighting Factors in the Objective Function on the Fault Diagnosis Results*

The selection of weighting factors of the objective function in (1) will impact the fault diagnosis results from the proposed optimization. When $\omega_1$ is selected to be high while $\omega_2$ & $\omega_3$ are selected as a low value, the proposed optimization will identify the fault scenario with the minimum sensor failures as the diagnosis results. In contrast, if $\omega_3$ is selected high while $\omega_1$ & $\omega_2$ are low, the proposed optimization will identify the fault scenario with the fewest incorrect timing tags. To further evaluate the effect of weighting factors on the result of event diagnosis, a combination of $\omega_1$, $\omega_2$ and $\omega_3$ is selected and the optimization is solved for a fault scenario using Synthetic South Carolina 500-Bus System in Fig. 4.

The fault scenario is that a fault occurred at L1 at 10:12:21:120 and MP tripped open CB1 and CB2 while the protection alarm MP for CB2 was missing. And SBP of CB4 triggered open CB4 incorrectly. The alarms received at the control center is summarized in Table VII. For a combination of the weighting factors in (1), the corresponding results are summarized in Table VIII. It is shown that when $\omega_2$ is high, the proposed optimization is aimed to minimize the number of sensor malfunctions, which leads to the incorrect determination of the fault scenario. While $\omega_1$ is high, the true fault scenario is identified. The results show that the weighting factors may have an impact on the fault diagnosis result from the proposed optimization. One potential technology to eliminate the effect of weighting factors on the result of event diagnosis is to hypothesize the number of incomplete alarms, incorrect alarms, and malfunctioning alarms directly in each scenario and carry out the proposed analytical model for each hypothesis. This approach based on multiple-hypothesis analysis can explicitly consider all combinations of issues in event diagnosis while the computational complexity will be exponentially increasing with regard to the potential abnormalities in the event. While it is worth to investigate and to implement this approach, given the scope of work, the authors will elaborate it in the future

TABLE IX
ALARMS FROM THE EVENT

| Time (ms) | Substation | Alarm | Time (ms) | Substation | Alarm |
|---|---|---|---|---|---|
| 0 | Tangling | MP of L4335 in relay set I operates | 250 | Tangling | C6 is tripped |
| 2 | Tangling | MP of L4335 in relay set II operates | 253 | Tangling | C15 is tripped |
| 5 | Jianshan | MP of L4335 in relay set I operates | 261 | Yangwan | C2 is tripped |
| 6 | Jianshan | MP of L4335 in relay set II operates | 265 | Shenggu | C7 is tripped |
| 30 | Jianshan | MP of L4336 in relay set I operates | 340 | Tangling | MP of B2-I in relay set I operates |
| 30 | Jianshan | MP of L4336 in relay set II operates | 341 | Tangling | MP of B2-I in relay set II operates |
| 52 | Jianshan | C11 is tripped | 350 | Wenzhou | MP of L4334 in relay set I operates |
| 78 | Tangling | C12 is tripped | 350 | Wenzhou | MP of L4334 in relay set II operates |
| 81 | Jianshan | C13 is tripped | 351 | Shenggu | MP of L4340 in relay set I operates |
| 160 | Tangling | C10 is tripped but failed to open | 389 | Tangling | C1 is tripped |
| 203 | Tangling | BFP of C10 operates | 390 | Tangling | C5 is tripped |
| 211 | Yangwan | MP of L4333 in relay set I operates | 390 | Tangling | C8 is tripped |
| 212 | Shenggu | MP of L4339 in relay set I operates | 390 | Tangling | C16 is tripped |
| 249 | Tangling | C3 is tripped | 395 | Shenggu | C9 is tripped |
| 249 | Tangling | C14 is tripped | 399 | Wenzhou | C4 is tripped |

TABLE X
EVENT DIAGNOSIS RESULTS

| No. of Fault | Fault Loc. & Time | Fail. Relay | Mal. Relay | Fail. CB | Mal. CB | Discrep. No. | Trust score |
|---|---|---|---|---|---|---|---|
| 1 | L4335@3.25ms | ∅ | MP of L4336@Jianshan, MP of B2-I@Tangling, MP of L4334@Wenzhou, MP of L4340@Shenggu | C10 | C1, C4, C5, C8, C9, C12, C13, C16 | 32 | 0.47 |
| 2 | L4335@3.25ms, B2-I@340.5ms | ∅ | MP of L4336@Tangling | C10 | C12, C13 | 10 | 0.83 |
| 3 | L4335@3.25ms, L4336@30ms, B2-I@340.5ms | MP of L4336@Tangling | ∅ | C10 | ∅ | 6 | 0.9 |
| 4 | L4335@3.25ms, L4336@30ms, B2-I@340.5ms, L4334@350ms | MP of L4336@Tangling | ∅ | C10 | ∅ | 6 | 0.9 |

TABLE XI
COMPARISON OF ALGORITHMS FOR EVENT DIAGNOSIS

| Methodology | No. of Hypotheses | Temporal Correlation | Linearity of Methodology | Handling Multiple Faults | Global Optimality | Solving Methodology |
|---|---|---|---|---|---|---|
| Algorithms in [4] | Up to $2^{m+n+k}$ | √ | × | √ | × | Heuristic |
| Algorithm in [14] | $n$ | × | √ | × | ∅ | Logic Reasoning |
| MILP algorithm in this study | $f$ | √ | √ | √ | √ | MILP solver |

study. Nevertheless, presented approach study will have same formulation and framework but a greater number of hypotheses. Note that, our assumptions in this study is utilizing best possible way to judge weighting factors similar to the one already existing in multiple power system operational practices such as weighting factor in state estimation for sensor data and failure rate assumptions in reliability analysis.

### *F. Discussion of the Proposed Methodology*

The proposed event diagnosis technology is aimed for application in a wide area system. The sampling, filtering and estimation algorithms impact the phasor estimation algorithm. But not all applications will be impacted by these differences assuming that PMUs meet the performance requirement of IEEE C37.118 and tested following IEEE TSS. As demonstrated in study using PMU Application Requirements Test Framework and related papers, some applications are sensitive, and some applications are not sensitive to the observed error caused by different types of PMUs. Applications addressed in this study will be least sensitive to PMU errors as it is based on a large threshold for fault currents and circuit breaker status. Hence data synchronization and sampling rates even being important for some other applications will be less important for the one addressed here. In the meanwhile, the proposed algorithm is able to determine the event inception time and handle missing time tags for event diagnosis. If GPS signal is not available, the proposed approach can still diagnose the event while the accuracy for event diagnosis may be affected since the temporal correlation of alarms is not fully utilized. As demonstrated in the real-world case in Section V-C, the proposed approach shows a good performance to handle real-world complex event scenario with multiple faults, missing alarm, and breaker failure. Note that missing alarms can be due to communication delay, device failure, or other reasons. However, the proposed approach needs to be further explored to handle problems such as cascading line trips. This point is discussed for potential applications of fault event diagnosis

## VI. Conclusion and Future work

This study presents a data driven approach based on MILP for fault event diagnosis of transmission systems considering the incomplete and incorrect alarms with sensor malfunctions. The proposed approach is demonstrated to be capable to handle complex fault events including multiple faults, missing fault inception time, failures and malfunctions of protective relays or CBs as well as incorrect alarm time tags. Alarms from multiple sensor sources such as PMUs, SERs, and SCADA data are jointly

$$(2) \Leftrightarrow \begin{cases} h_{i1} \le L_j^{Z2} \\ h_{i1} \le r_i^{pilot} \\ h_{i1} \ge L_j^{Z2} + r_i^{pilot} - 1 \\ r_i^{MP} \ge L_j^{Z1} \\ r_i^{MP} \ge h_{i1} \\ r_i^{MP} \le h_{i1} + L_j^{Z1} \\ h_{i2} \le r_i^{MP} \\ h_{i2} \le 1 - f_{r_i^{MP}} \\ h_{i2} \ge r_i^{MP} - f_{r_i^{MP}} \\ h_{i3} \le 1 - r_i^{MP} \\ h_{i3} \le m_{r_i^{MP}} \\ h_{i3} \ge m_{r_i^{MP}} - r_i^{MP} \\ R_i^{MP} \ge h_{i2} \\ R_i^{MP} \ge h_{i3} \\ R_i^{MP} \le h_{i2} + h_{i3} \end{cases}$$

$$(4) \Leftrightarrow \begin{cases} r_i^{PBP} \le L_j \\ r_i^{PBP} \le 1 - y_{CB_i}^{MP} \\ r_i^{PBP} \le 1 - f_{CB_i}^{MP} \\ r_i^{PBP} \ge L_j - y_{CB_i}^{MP} - f_{CB_i}^{MP} \\ f_{CB_i}^{MP} \le 1 - y_{CB_i}^{MP} \\ f_{CB_i}^{MP} \le R_i^{MP} \\ f_{CB_i}^{MP} \ge R_i^{MP} - y_{CB_i}^{MP} \\ h_{i4} \le r_i^{PBP} \\ h_{i4} \le 1 - f_{r_i^{PBP}} \\ h_{i4} \ge r_i^{PBP} - f_{r_i^{PBP}} \\ h_{i5} \le 1 - r_i^{PBP} \\ h_{i5} \le m_{r_i^{PBP}} \\ h_{i5} \ge m_{r_i^{PBP}} - f_{r_i^{PBP}} \\ R_i^{PBP} \ge h_{i4} \\ R_i^{PBP} \ge h_{i5} \\ R_i^{PBP} \le h_{i4} + h_{i5} \end{cases}$$

$$(6) \Leftrightarrow \begin{cases} r_i^{SBP} \le L_j \\ r_i^{SBP} \le 1 - y_{CB_k}^{MP} \\ r_i^{SBP} \le 1 - y_{CB_k}^{PBP} \\ r_i^{SBP} \le 1 - f_{CB_k}^{MP} \\ r_i^{SBP} \le 1 - f_{CB_k}^{PBP} \\ r_i^{SBP} \ge L_j - y_{CB_k}^{MP} - y_{CB_k}^{PBP} \\ \quad - f_{CB_k}^{MP} - f_{CB_k}^{PBP} \\ f_{CB_k}^{PBP} \le 1 - y_{CB_k}^{PBP} \\ f_{CB_k}^{PBP} \le R_k^{PBP} \\ f_{CB_k}^{PBP} \ge R_k^{PBP} - y_{CB_k}^{PBP} \\ h_{i6} \le r_i^{SBP} \\ h_{i6} \le 1 - f_{r_i^{SBP}} \\ h_{i6} \ge r_i^{SBP} - f_{r_i^{SBP}} \\ h_{i7} \le 1 - r_i^{SBP} \\ h_{i7} \le m_{r_i^{SBP}} \\ h_{i7} \ge m_{r_i^{SBP}} - f_{r_i^{SBP}} \\ R_i^{SBP} \ge h_{i6} \\ R_i^{SBP} \ge h_{i7} \\ R_i^{SBP} \le h_{i6} + h_{i7} \end{cases}$$

$$(8) \Leftrightarrow \begin{cases} R_i \ge R_i^{MP} \\ R_i \ge R_i^{PBP} \\ R_i \ge R_i^{SBP} \\ R_i \le R_i^{MP} + R_i^{PBP} + R_i^{SBP} \\ r_i^{BFP} \le 1 - y_{CB_i} \\ r_i^{BFP} \le R_i \\ r_i^{BFP} \ge R_i - y_{CB_i} \\ h_{i8} \le r_i^{BFP} \\ h_{i8} \le 1 - f_{r_i^{BFP}} \\ h_{i8} \ge r_i^{BFP} - f_{r_i^{BFP}} \\ h_{i9} \le 1 - r_i^{BFP} \\ h_{i9} \le m_{r_i^{BFP}} \\ h_{i9} \ge m_{r_i^{BFP}} - r_i^{BFP} \\ R_i^{BFP} \ge h_{i8} \\ R_i^{BFP} \ge h_{i9} \\ R_i^{BFP} \le h_{i8} + h_{i9} \end{cases}$$

$$(10) \Leftrightarrow \begin{cases} R_i^{CB} \ge R_i \\ R_i^{CB} \ge \sum_{j \in \Omega_{CB_i}} R_j^{BFP} \\ R_i^{CB} \le R_i + \sum_{j \in \Omega_{CB_i}} R_j^{BFP} \\ h_{i10} \le R_i^{CB} \\ h_{i10} \le 1 - f_{CB_i} \\ h_{i10} \ge R_i^{CB} - f_{CB_i} \\ h_{i11} \le 1 - R_i^{CB} \\ h_{i11} \le m_{CB_i} \\ h_{i11} \ge m_{CB_i} - R_i^{CB} \\ y_{CB_i} \ge h_{i10} \\ y_{CB_i} \ge h_{i11} \\ y_{CB_i} \le h_{i10} + h_{i11} \\ y_{CB_i} \ge y_{CB_i}^{MP} \\ y_{CB_i} \ge y_{CB_i}^{PBP} \end{cases}$$

used for accurate fault event diagnosis in an automatic manner. The temporal correlation of alarms is explicitly modeled in the constraints for improved fault diagnosis. The high computational performance of the proposed algorithm is demonstrated by the testing cases and the proposed methodology is potential for on-line applications for fault diagnosis at the control center. Future study needs to further validate the algorithm with more scenarios and hypothesize explicitly all combinations of issues in the outage for fault event diagnosis.

## Appendix

See the equation shown on the previous page, where $h_{i1}$, $h_{i2}, \ldots, h_{i10}$ and $h_{i11}$ are additional decision variables to assist the conversion of logical nonlinear constraints into a linear combination of decision variables.

## References

[1] Y. Sekine, Y. Akimoto, M. Kunugi, C. Fukui, and S. Fukui, "Fault diagnosis of power systems," *Proc. IEEE*, vol. 80, no. 5, pp. 673–683, May 1992.

[2] M. Kezunovic, "Smart fault location for smart grids," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 11–22, Mar. 2011.

[3] W. Guo, F. Wen, G. Ledwich, Z. Liao, X. He, and J. Liang, "An analytic model for fault diagnosis in power systems considering malfunctions of protective relays and circuit breakers," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1393–1401, Jul. 2010.

[4] Y. Zhang, C. Y. Chung, F. Wen, and J. Zhong, "An analytic model for fault diagnosis in power systems utilizing redundancy and temporal information of alarm messages," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4877–4886, Nov. 2016.

[5] M. Y. Park, M. Lefley, B. Ramsay, and I. Moyes, "An abductive fuzzy knowledge based system for fault diagnosis in a power system," *Proc. Int. Conf. Power Syst. Tech.*, pp. 345–350, Aug. 2000.

[6] Z. A. Vale and A. M. e Moura, "An expert system with temporal reasoning for alarm processing in power system control centers," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 1307–1314, Aug. 1993.

[7] C. Y. Teo, "Conventional and knowledge based approaches in fault diagnosis and supply restoration for power network," *IEEE Trans. Power Syst.*, vol. 13, no. 1, pp. 8–14, Feb. 1998.

[8] S. D. J. McArthur, R. Dysko, J. R. McDonald, S.C. Bell, R. Mather, and S. M. Burt, "The application of model based reasoning within a decision support system for protection engineers," *IEEE Trans. Power Del.*, vol. 11, no. 4, pp. 1748–1754, Oct. 1996.

[9] L. Xu and M. Kezunovic, "Implementing fuzzy reasoning Petri-Nets for fault section estimation," *IEEE Trans. Power Del.*, vol. 23, no. 2, pp. 676–685, Apr. 2008.

[10] J. Sun, S. Y. Qin, and Y. H. Song, "Fault diagnosis of electric power systems based on fuzzy Petri nets," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 2053–2059, Nov. 2004.

[11] G. J. Cardoso, J. G. Rolim, and H. H. Zurn, "Application of neural network modules to electric power system fault section estimation," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1034–1041, Jul. 2004.

[12] Z. Yongli, H. Limin, and L. Jinling, "Bayesian networks-based approach for power systems fault diagnosis," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 634–639, Apr. 2006.

[13] H. Miao, M. Sforna, and C. C. Liu, "A new logic-based alarm analyzer for on-line operational environment," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1600–1606, Aug. 1996.

[14] J. W. Jung, C. C. Liu, M. G. Hong, M. Gallanti, and G. Tornielli, "Multiple hypotheses and their credibility in on-line fault diagnosis," *IEEE Trans. Power Del.*, vol. 16, no. 2, pp. 225–230, Apr. 2001.

[15] B. Cui, A. Srivastava, and P. Banerjee, "Automated failure diagnosis in transmission network protection system using synchrophasors," *IEEE Trans. Power Del.*, vol. 33, no. 5, pp. 2207–2216, Oct. 2018.

[16] Y. Jiang, C. C. Liu, M. Diedesch, E. Lee, and A. K. Srivastava, "Outage management of distribution systems incorporating information from smart meters," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 4144–4154, Dec. 2016.

[17] Z. Y. He, H. D. Chiang, C. W. Li, and Q. F. Zeng, "Fault-section estimation in power systems based on improved optimization model and binary particle swarm optimization," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2009, pp. 1–8.

[18] L. He, K. Jia, and Z. Fan, "The immune genetic algorithm in fault diagnosis of modern power system," in *Proc. Int. Conf. Edu. Technol. Comput.*, 2010, pp. V4-26–V4-29.

[19] M. Prais and A. Bose, "A topology processor that tracks network modifications over time," *IEEE Trans. Power Del.*, vol. 3, no. 3, pp. 992–998, Aug. 1988.

[20] Y. Jiang, "Towards detection of distribution system faulted line sections in real-time: A mixed integer linear programming approach," *IEEE Trans. Power Del.*, vol. 34, no. 3, pp. 1039–1048, Jun. 2019.

[21] IEEE Std 14-Bus System. [Online]. Available: http://icseg.iti.illinois.edu/ieee-14-bus-system/. Accessed on: 2018.

[22] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, Jul. 2017.

[23] Electric Grid Test Case. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg500/. Accessed on: 2018.

**Yazhou Jiang** (S'12–M'17) received the Ph.D. degree from Washington State University, Pullman, WA, USA, in 2016. He is currently a Power Systems Engineer with GE Global Research Center, Niskayuna, NY, USA. At GE, he is leading or contributing to several projects such as pumped storage hydro for renewable integration, distribution automation, grid impact from cyber-attacks, and flexible resources at demand side for grid operation. His research interests include distribution automation, fault diagnosis of power system, and optimization methods for power system applications.

**Anurag K. Srivastava** (S'97–M'05–SM'08) received the Ph.D. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2005. He is currently an Associate Professor of electric power engineering with Washington State University, Pullman, WA, USA, and the Director of the Smart Grid Demonstration and Research Investigation Lab, Energy System Innovation Center. He is an Editor for the IEEE Transactions on Smart Grid, IEEE Transactions on Power Systems, *IET Generation, Transmission and Distribution*, and Elsevier *Sustainable Computing*. He is an IEEE Distinguished Lecturer and Author of more than 300 technical publications including a book on power system security and three patents. His research interest includes data-driven algorithms for power system operation and control including resiliency analysis.