Impact Quantification of Hypothesized Attack Scenarios on Bus Differential Relays

†Rashiduzzaman Bulbul, *Yuan Gong, ‡Chee-Wooi Ten, §Andrew Ginter, and *Shengwei Mei
†Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI 49931.
*Department of Electrical Engineering, Tsinghua University, Beijing, China.
§Waterfall Security Solutions, 21 Hamelacha St., Afek Industrial Park, Rosh Ha‘ayin, 48091, Israel.
Emails: {†mbulbul, ‡ten}@mtu.edu, {*y-gong09, meishengwei}@tsinghua.edu.cn, §andrew.ginter@waterfall-security.com.

Abstract—The evolution of protective relays, from electromechanical to digital devices, has enhanced the system-wide monitoring and control. In most substation configuration, intelligent electronic devices (IEDs) of bus differential relays remain one of the most vital protection schemes to prevent equipment from overloading damage. The proposed study is to systematically identify the critical substation/IED hypothesized attack scenarios by eliminating large number of insignificant cases in an online environment. A statistical model combined with real-time simulation is proposed to identify the critical combinations of substation outages. The metrics include (1) substation risk index, (2) dependency of substation combinations, and (3) criticality of time windows which are validated using IEEE 30-bus system in this study. The results of combinatorial search based on a substation list are compared using depth first search (DFS), breadth first search (BFS), and random selection.

Index Terms—Cyber-contingency analysis, microprocessor-based relays, substation cybersecurity.

I. INTRODUCTION

The advanced information and communication technology integrated with sophisticated control mechanism can improve system-wide coordination on efficiency and reliability of electrical grids [1]. The enhanced cyberinfrastructure and technology, if not carefully audited, may potentially be prone to electronic intrusions due to the vulnerabilities of access points in substation networks [2]. Physical access control of communication network and lack of connectivity of power control networks to external networks might be considered as obstacles against the threats. The sharing of network database between corporate and enterprise networks increases the complexity of system management as the weak access points of system security can be compromised [3]. One example of advanced persistent threat, Stuxnet, reveals themselves as potential attack agents launched by a group of attackers or organizations who can systematically penetrate the network of a specific target or entity for a prolonged period [4].

Misconceptions of cybersecurity on power control systems has been reported with generic recommendation based on supporting infrastructure [5]. The digital protective relays of substation communication have been identified with the possible remote access points of cyber-vulnerabilities between substations and control center [6]. Coordinated cyberattack to multiple substations has been recognized as a triggering potential that might lead to a system blackout [7]. In the recent years, there are standardized schemes and objects for substation automation have been implemented for protective IED integration [8]–[10]. The design of standard function blocks of substation automation and its database enhancement have been proposed by exploring implemented concepts and design with available intelligent systems [11], [12]. Although substation automation network is not publicly accessible, trustworthy computing paradigms with distributed intelligence are crucial to the development of security related framework [13].

According to North American Electric Reliability Corporation (NERC) standard, electric power grids are designed to withstand a single contingency (N − 1 contingency) without violating any system security constraints [14]. NERC security analysis ranks the most critical reasons for system contingencies such as protection system misoperation, AC substation equipment failure, or automatic transmission events at a given operating condition [15]. A full redundancy of defense systems may not be immune from an intelligent coordinated attack [16]. A coordinated attack can trip multiple generators, transmission lines, loads, or transformers from a power grid that can aggravate the system operating conditions. Violation of predefined threshold values of bus voltages, system frequency, and branch flows may lead to cascading failure of a system blackout [17].

Bus differential protection is one of the most vital schemes as misoperation of the bus protection can disconnect large number of generators, loads, and lines in a substation. The nodal disconnection can overload other in-service transmission lines that can cause system instability issues [18]. There have been several methods introduced to enumerate critical substation combinations or cascading sequence but the challenges of combinatorial explosion remain [19]. Cyber-contingency analysis includes steady-state evaluation using AC power flow techniques that is evaluated based on the hypothesized substation outages [20]. A preliminary investigation on buses associated with the IEDs under one substation is reported [21].

As system loading level varies over time, the criticality of each substation (node) can be different [22], [23]. Ideally, a tool that systematically enumerates the worst case combinations based on the operating status and the relationship between the IED and breakers would assist system dispatchers to understand the pivotal nodes of a power system and the potential mitigation strategies can be reconfigured. This will
improve the cyber-situation awareness in a control room and will help to plan appropriate strategies against coordinated cyberattack [24]. This paper is concerned about identification of critical substation combinations that are connected with the bus differential relays corresponding to each substation.

II. CONNECTION BETWEEN IEDs AND BREAKERS

Fig. 1 depicts an example of bus differential protection scheme for a substation with single-bus-single-breaker configuration. Bus 1 is connected to a generator, a transmission line, a transformer and bus 2 with a tie line breaker. Buses 3 and 4 are connected with transformer and other feeders. Intuitively, operation of a bus differential protection in a substation can disconnect multiple components from the system. Each of the two differential relays are configured to protect two buses where each busbar is defined as a zonal protection. Each zone can be set up differently based on applications, current transformer ratio, and bus topology [25]. Attacks on the differential relays can be executed by an intruder in the substation network; (s)he can manipulate the zonal settings on each IED.

The status of a zone and differential relay is defined by a binary variable of 1 or 0 that indicates either these are compromised or otherwise. The zone setting of a microprocessor-based differential relay is configured with a number of zone configuration files. The cyberintruder would have to compromise the relay with administrative privilege in order to modify those files. If \( z \) is the number of zones and \( r \) number of differential relays are associated with a single busbar in a substation, then complete busbar outage \( (\hat{B}) \) is denoted by

\[
\hat{B} = \prod_{i=1}^{r} \prod_{j=1}^{z_i} R_{i,j}
\]

where \( R_i \) is the status of \( i \)-th of differential relay, \( z_i \) is the total number of zones protected by a \( i \)-th relay, and \( Z_{i,j} \) is the status of \( j \)-th zone protected by a \( i \)-th relay. Depending on the substation configuration, the connection among the number of IEDs mapping to each breaker can result in the total combinations of the zonal combinations.

In this study, hypothesized substation outages based on the worst bus outage combination are analyzed. This is a combinatorial problem with differential zones where each hypothesized bus outage in a substation has distinct impact on the system. The worst case scenario is a complete substation outage which ensures de-energization of multiple transmission lines, transformers, loads, generators that affects the flow of power to other part of the network due to the nodal disconnection. The number of combination of bus outages depends on substation size, topology configuration, and the number of protection zones. If \( z_s \) is the set of all differential protection zones in a substation, then the total enumeration of combinations is:

\[
C_{\text{max}} = \sum_{k=1}^{z_s} C_{k}^{z_s} = \sum_{k=1}^{z_s} \frac{z_s!}{k!(z_s-k)!}
\]

where \( C_{\text{max}} \) is the set of all possible zonal combinations. Based on substation size and busbar topology, only a small number of combinations results in substation outage, i.e., \( C_{\text{worst}} \subseteq C_{\text{max}} \).

In Fig. 1, there are 4 zones of bus differential protection which has 15 combinations of differential zones. Among them, only 2 combinations can electrically disconnect the entire substation. De-energization of zone 1 and zone 2 disconnects all incoming power to transformers which results an entire substation outage. It should be noted that the combination consists of entire set of protection zone will always in the \( C_{\text{worst}} \) set. But the attacker’s interest would minimize the number of zones to manipulate for a complete substation disconnection. According to Fig. 1, \( C_{\text{worst}} = \{(z_1 \cap z_2) \cup (z_1 \cap z_2 \cap z_3) \cup (z_1 \cap z_2 \cap z_4) \cup (z_1 \cap z_2 \cap z_3 \cap z_4)\} \). The combination of \( C_{\text{worst}} \) which consists of minimum number of zones is considered as the most significant combination and is denoted by \( C_{\text{worst}} \). In this case, \( C_{\text{worst}} = (z_1 \cap z_2) \). The set of relays for protection of \( C_{\text{worst}} \) is considered as the most critical relays of substation and is denoted by \( R \) where in this
III. EXTENDED REVERSE PYRAMID MODEL (RPM)

The RPM model proposed in [21] is an elimination approach to effectively determine nonconvergent cases from the large pool of combination list. At the first level of RPM, a substation list is split into several segments exploring the network topology of a power system using breadth-first-search (BFS) algorithm. The remainder levels of RPM are constructed based on pairwise merge criterion described in [21] which ensures the maximum interaction of substations among all segments. Nonconvergent combinations for each segment are identified through power flow simulation. In this paper, RPM model is validated by extending additional 2 segmentation methods to verify if the risk index of substations and the results are consistent and also be checked the difference among these algorithms. Risk index is the measure of individual substation criticality for the loading level of a given system that is defined in section IV.

1) Random Selection: Using this method the substations for each segment at the first level are randomly selected. The number of segments in first level of RPM is \( N_{seg} = \frac{S}{2} \), where \( S \) and \( s \) are the number of substations in a power system and each segment, respectively. The subset of substation for each segment is denoted by \( ss(i, j) \) where \( i \) indicates level and \( j \) indicates segment. For the first level of RPM, each \( ss(i = 1, j \leq N_{seg}) \) is a random set, consisting of \( s \) number of substation and \( ss(\cdot) \subset S_{sub} \) where \( S_{sub} \) is a complete set of all substations in a power system. Since the number of substations for each segment is limited to the number of substation in each segment, all segments are uniformly divided.

2) Depth First Search (DFS): It is a graph traversal algorithm that navigates the vertices of graph first in deep then wide. This algorithm traverses from a root node in depth until a visited vertex is found [26]. When the current vertex has only visited neighbor left, it is then marked as visited. The algorithm then backtracks to the first vertex which is not traversed. The DFS constructs a spanning tree and establishes an order to each vertex according to when it is first visited. Based on the order, \( S_{sub} \) is split into several segments where the substations of a particular segment are connected in depth. The formulation of hypothesized outage events using DFS algorithm is to ensure the de-energization of those substations that are close to each other. To start the search process, a root substation is randomly selected from \( S_{credible} \subset S_{sub} \). A \( S_{credible} \) is the set of substations having higher degree of connectivity with other substations. The order of \( S \)-th substation from root is defined as:

\[
\text{Ord}(R) = \begin{cases} 
0, & \text{if busbar is root node} \\
v, & \text{otherwise}
\end{cases}
\]

where \( v \) is the count of the discovery of \( S \)-th substation. The subset of substation for the first level of RPM is \( ss(i = 1, j \leq N_{seg}) \), which is a set of \( s \) number of substation from the order \( \text{Ord} = \{s(j - 1)\} \) to \( \text{Ord} = \{(j \cdot s) - 1\} \). As a result of the first \( s \) substation are assigned to first segment, the next \( s \) substation will follow in the second segment, and so on.

3) Breadth First Search (BFS) [21]: This is the opposite traversal algorithm to find vertices in breadth first then depth. The algorithm searches for all vertices for \( N_{seg} \) times with different order of worst case substation outage. The subset of substations for all segments at the first level of RPM is denoted by \( ss(i = 1, j \leq N_{seg}) \), which is a set of \( s \) number of substation with minimum order to \( \text{root}(j) \). The BFS algorithm ensures that substations of a segment are directly connected to root node. Each segment at the first level of RPM represents an area of power system network.

Fig. 2 shows an enumeration example of the IEEE 14-bus system using the aforementioned 3 approaches for identification of nonconvergent substation combinations. The number of segments is initiated with \( N_{seg} = 2 \). At the level 1, the nonconvergent substation combinations vary and list is merged between the two segments to form level-2 segment. Among those substations, new nonconvergent combinations are determined. This determines a complete nonconvergent combinations for each method. The unique list of nonconvergent combinations shows in green color box are the overlapping combinations with other methods. The orange boxes indicate
at least two methods have the same combination while the white boxes are the unique combinations on those methods that have no overlapping. The total maximum combinations of green, orange, and white boxes are 9, 2, and 3, respectively. The summation of a complete enumeration for the IEEE 14-bus and 30-bus systems would have been for each with total combinations of 16,383 and 1,073,741,823.

IV. IDENTIFICATION OF BOTTLENECK LIST

Fig. 3 presents the proposed algorithm to identify bottleneck list of substation combinations, risk index of individual substation, and dependency index of substation combinations. As the criticality of substation depends on system loading level of time slots, the proposed algorithm starts with identifying a specific time slot. Input data for each time slot contains the power flow model, individual line status, and amount of load connected to each substation.

A. Bottleneck List of Combinations

Each simulation of RPM provides a list of nonconvergent cases that is from all segments at the first level. The limitation of RPM is that it does not enumerate all possible nonconvergent combinations from the set of substation using single simulation. However, this can be validated with multiple simulations to ensure the consistency of the impact results for each substation. If \( N_{sim}(t) \) is considered to find the list of nonconvergent combinations at a given time, there might some repetitive combinations on the list. Probability function is used to sort out the most frequent combinations from the list. The sorted list of substation combinations is the bottleneck list of combinations. The number of nonconvergent combination depends on the system loading level. Historical simulation results along with real-time results of RPM can be used to find the list of nonconvergent combination, which will incorporate the trend of loading level change of the system. Probability of specific combination (of bottleneck list) is as follow:

\[
P_r(c_t) = \frac{N_{occ}(t)}{N_{sim}(t)} \tag{4}
\]

where \( c_t \) is a combination of a bottleneck at \( t \)-th time slot, \( N_{sim}(t) \) is the number of simulation for \( t \)-th time slot, and \( N_{occ}(t) \) is the number of occurrence for that particular combination in \( N_{sim}(t) \) simulation result at \( t \)-th time slot. A threshold value of \( P^*_r(c_t) \) is used to find the bottleneck list of nonconvergent combinations.

B. Risk Index of Individual Substation on Bottleneck list

For the purpose of substation risk analysis, bottleneck list is further examined. Since the criticality of substation depends on loading level, it is necessary to find an index which determines the individual substation risk. Individual substation risk index will help control center operator to rank the substation and decide which substation should be under close observation. In this study, substation risk index is defined as:

\[
R(s_t) = \begin{cases} 
1.0 & \text{if } k = 1 \\
\frac{N_s}{N_b} & \text{if } k > 1 
\end{cases} \tag{5}
\]

where \( k \) is the number of substation in a combination, \( s_t \) denotes particular substation in a bottleneck of \( t \)-th time slot, \( N_s \) is the number of combination in bottleneck list, contain the specific substation(s), and \( N_b \) is the total number of combinations on bottleneck list for \( k > 1 \).

The equation shows that if there is only one substation in a combination of bottleneck list, highest risk index, i.e., 1.0, will be given to that particular substation. This implies that de-energization of that substation might lead to systemic risk of potential cascading failure. Since any combination on bottleneck list might be problematic, it is important to find the number of combinations in which a particular substation that exists. The risk index for those substations with more than one combination is given by the ratio of \( N_s \) and \( N_b \). Higher value
of risk index indicates that higher number of combinations are associated with particular substation(s).

C. Combination Dependency on Bottleneck List

The combinations on the bottleneck list might have common substation among them. If so, the combinations have interdependency. Based on the substations of a combination, each combination represents a group of combinations where other combinations of that group has at least one common substation at a particular combination. The dependency increases with the number of combinations. The combinations are then prioritized based on their dependency. The dependency of a combination is defined as:

$$D(c_t) = m \sqrt{\frac{\sum_{i=1}^{k} N_{comm}(i)}{\sum_{i=1}^{m} N_{sub}(i)}} \quad \forall \ k > 1$$

(6)

where $c_t$ is a bottleneck combination at t-th time slot, $m$ is the number of combination in a group that associated with $c_t$, $N_{sub}$ is the number of substations for each combination, $k$ is the number of substation in $c_t$, and $N_{comm}$ is the number of combination in which t-th substation is common.

Since $\sum N_{comm}$ is less than $\sum N_{sub}$, the right term of this equation would be less than 1.0. The ratio of $\sum N_{comm}$ and $\sum N_{sub}$ is given $m$-th root which will increase the value of $D(c_t)$ with increased combinations in a group. The $\sum N_{comm}$ and $\sum N_{sub}$ are equal only when all the combinations in a group are same which is unlikely. If any combination doesn’t have any common substation with other combinations it has no dependency.

V. Simulation Study

To validate the proposed methodology of finding the critical buses/substations and their combinations at specific time period, the IEEE 30-bus system is used for the simulation study. There are 100 simulation cases are set up representing 100 time windows to ensure statistically consistency on the results. In this simulation, all bus differential protection schemes are linked to command-to-physical devices that interact with the physical devices of circuit breakers for each substation. Assumptions of the worst case scenario is that there are coordinated attacks among multiple substations to de-energize the buses by modifying the breaker trip settings of bus differential relay. Dispatchers at the energy control center get the statuses of transmission lines, loads, and generators from remote terminal units (RTUs) for each time slot. One or two transmission circuits are randomly chosen to be de-energized in the simulations which represents maintenance activities or load shedding. Load profile of each substation is modified over 12 time slots for IEEE 30-bus system to represent the load variation for entire system over time. The most frequent nonconvergent combinations from 100 simulation cases are sorted as bottleneck list based on the threshold value of $P_r(c_t)$ which is 0.2 using the IEEE 30-bus system.

The system consists of $N_{seg} = 5$ with each of those segments with 6 substations. To begin with the navigation, the highest degree of connectivity, i.e., at least 4 lines connected in this case, is selected from the test case. The number of substations in each segment $N_s = 6$ is randomly chosen from the list $S_{credible} = \{1, 2, 4, 5, 6, 10, 12\}$. Fig. 4 shows the non-zero dependency index for the combinations of bottleneck list. The insight of the dependency index is that it quantifies overlapping bus combination with others. In a particular time slot, all nonconvergent combinations is ranked based on the dependency index. In Fig. 4, combination (22, 6, 15) is an important one as it shows the highest interaction with other combinations because buses 6, 15, and 22 are common with other 9, 7, and 4 combinations, respectively. As dependency index is obtained from the number of common substations, it might vary with the size of bottleneck list. The proposed methodology analyzes the combinations on the bottleneck list which is useful to categorize individual combination.

The proposed method first determines individual busbar in each substation based on its criticality and nonconvergent combinations at a specific time slot. Fig. 5 shows the risk index for individual busbar over 12 different time slots using DFS algorithm. The value of $R(s_t)$ is between 0 and 1. Buses 9, 12, 25, and 27 are found as the most critical nodes of IEEE 30-bus system since removal of any busbar results in a nonconvergent power flow solution. For this reason, all of these busbar are showing risk index 1.0 over 12 time slots. No combination on bottleneck list contains busbars 8, 11, 14, 21, 26, 29, and 30. This shows all of these buses are with zero risk index.

Nonconvergent list of substation combinations depends on system and individual substation loading levels as well as the size of bottleneck list might vary over time. As a result, simulations show that the risk index for each busbar vary slightly over the 12 time slots. Busbar 6 is showing risk index 1.0 for only time slot-7 and 12 which indicates that system loading level for these two time slots are critical for the busbar. The remainders for busbar 6 of its risk index are around $\approx 0.31$. In most time slots, busbars 4, 6, 10, and 15 shows risk index more than 0.2, which indicate approximately 20% combinations of bottleneck list contain any of these busbars. Buses 1, 2, 22, 24, 28, and 16 have risk index in between 0.1 to 0.2 over the time slots. Risk index of substations 3, 5, and 7 are less than 0.038. It is noted that busbars having risk index more than 0.1 are showing consistent risk index over 12 time slots. The reason is that risk index for individual busbar depends on the size of bottleneck list and bottleneck list depends on threshold value of $P_r(c_t)$. The threshold value of $P_r(c_t)$ ensures the most frequent combinations to be listed on bottleneck list. The combinations with higher $P_r(c_t)$ mostly consists of those busbars having risk index more than 0.1. This observation demonstrates the effectiveness of proposed model to identify the most critical substations in the system.

Simulation results of a particular time slot for three segmentation methods show similar pattern of risk index of individual busbars. There might be an insignificant deviation among risk indices of a particular busbar obtained from different segmentation methods. However, there is no abrupt deviation of risk index on any busbars. This can be discerned from the observation that the consistency of individual risk indices over three different segmentation methods. Average deviation of
busbar risk index is determined using Eq. (7):

$$AD_s = \frac{1}{T} \sqrt{\frac{1}{T} \sum_{t=1; i \neq j}^{T} \left( R_{s,i}(t) - R_{s,j}(t) \right)^2} \tag{7}$$

where $AD_s$ is the risk index average deviation of $s$-th substation, $T$ is the total number of time slot, $i$ and $j$ indicate different segmentation method, and $R_{s,i}(t)$ is the risk index of $s$-th substation at $t$-th time slot.

The numbers obtained from Table I is based on Eq. (7) where $R_{s,i}$ and $R_{s,j}$ both are selected from either of the two of Random Selection, BFS, and DFS. This compares the deviation of risk indices over past 12 time slots. Table I shows all buses have very low risk index deviation except substations 16 and 22 (those two rows are bold) because the average risk index of these substations using DFS algorithm, which is higher than the risk index obtained using random selection and BFS algorithms. Lower risk index deviation of busbar demonstrates that the proposed RPM method effectively identifies critical buses and their combinations regardless of segmentation methods.

VI. CONCLUSION

The proposed study establishes impact quantification metrics to the security framework on substation outages. The outages are mapped to IEDs of bus differential relays that can be hypothetically compromised and manipulated by electronic intruders. A systematic approach introduced in this paper using RPM method is to avoid combinatorial explosion. Three segmentation methods are used in the simulation studies to ensure that the proposed 3 metrics demonstrate consistent simulation results among them. Future work includes validating the methods with larger simulation test cases.

The proposed cyber-contingency analysis systematically enumerates the “worst case scenarios” by eliminating insignificant combinations from the convergent cases. The conventional way of N-1 or higher order contingency analysis does not capture the worst outcome of hypothesized substation cyberattacks. This would help improving situation awareness in a control room to identify critical combinations from nonconvergent power flow solutions. Future work includes comparing the other combinatorial methods to verify if potential performance time and accuracy of worst case combinations can be improved.

### REFERENCES


### Fig. 5. Risk Index of Individual Substation Using DFS on IEEE 30-Bus System

<table>
<thead>
<tr>
<th>Substation ID</th>
<th>Time slot-1</th>
<th>Risk index</th>
<th>Substation ID</th>
<th>Time slot-2</th>
<th>Risk index</th>
<th>Substation ID</th>
<th>Time slot-3</th>
<th>Risk index</th>
<th>Substation ID</th>
<th>Time slot-4</th>
<th>Risk index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
</tr>
<tr>
<td></td>
<td>0 0.5 1</td>
<td>Substation ID</td>
<td>Time slot-5</td>
<td>Risk index</td>
<td>Substation ID</td>
<td>Time slot-6</td>
<td>Risk index</td>
<td>Substation ID</td>
<td>Time slot-7</td>
<td>Risk index</td>
<td>Substation ID</td>
</tr>
<tr>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
</tr>
<tr>
<td></td>
<td>0 0.5 1</td>
<td>Substation ID</td>
<td>Time slot-9</td>
<td>Risk index</td>
<td>Substation ID</td>
<td>Time slot-10</td>
<td>Risk index</td>
<td>Substation ID</td>
<td>Time slot-11</td>
<td>Risk index</td>
<td>Substation ID</td>
</tr>
<tr>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
<td></td>
<td>1 5 10 15 20 25 30</td>
<td>0 0.5 1</td>
</tr>
</tbody>
</table>


