Title: **Short Circuit Model Data Requirements and Improvements for an Automated Fault Location Program**

Authors: Xinyang Dong, Juergen Holbach, Robbie James

Keywords: Automated Fault Location, Short Circuit Modeling, Automation, Naming Convention, Data Management, Data Cleaning, GIS data

I. Introduction

As capital projects continue to increase, new technology advances, and additional compliance requirements take effect, utilities and transmission owners are faced with challenges related to data management and data cleaning. One challenge is related to fault location: after a fault happens, generating an accurate fault location estimate and promptly providing that information to the field is crucial for protection engineers and utilities. To meet this need, an automated advanced fault location (AFL) system has been developed to provide a more streamlined fault location process and a more accurate fault location calculation result [1]. The AFL program automatically loads fault records, connects to a short-circuit model, calculates the fault location using different algorithms, and utilizes an expert system to determine the best fault location estimate. The fault location calculation requires an up-to-date and accurate short-circuit model to have the desired accuracy. Having a high-quality short-circuit model is crucial for the fault location program, as well as other automated processes and studies.

This paper discusses the data requirements for a transmission network short-circuit model and how the model is used in the AFL system. In addition, methods of cleaning the data to improve the data quality in an automated way are also presented. We also discuss the benefits of having a short-circuit model with high data quality, including how it benefits other automated processes and studies, such as wide-area coordination studies and PRC-027. Later in the paper, we share some common difficulties that utilities experience and how they are addressed. Finally, we summarize our experience and discuss future steps to enable advanced uses for the AFL program.

II. Realistic Example of Issues in the Short Circuit Model for the AFL Program

The short circuit model of a transmission system is used in studies by engineers in protection, as well as in other departments such as planning. These studies are generally performed manually in many utilities, and the short circuit database is often maintained manually as well. When adding data into the short circuit model, the engineer focuses on the impedance values that describe a power system object (line, transformer, etc.) as they are essential for the short circuit study. Other data objects, like bus name, bus type, line name, line length, etc., are often not entered following a standard convention.

For the automated AFL program to work, a couple of data attributes are essential to enable automated processing. The base waveform information used in the fault location will typically be provided by a fault record from a protection relay or digital fault recorder. Utilities often use an identifier of the affected line in the file name of the fault record. It is essential that the AFL program be able to identify the associated line in the short circuit model automatically. After the line is identified in the short circuit model, the AFL program will initiate the optimal fault location estimate based on the line's detailed impedance model and then report the fault location in terms of impedance and mileage. In this process, the following information from the short circuit model forms the basis of the study:

- Line topology
- Line and substation identifier
- Line impedances and length
- Fault current level for the surrounding network

To better understand the details mentioned above, we will go through an example of a fault on the BRIGHTVILLE–PAGERVILLE 60-kV line.

A. Case Study Illustrating the Data Requirements

In this section, we discuss an example of finding the fault location on the BRIGHTVILLE–PAGERVILLE 60-kV line using the network model stored in a short circuit program and fault records extracted from the field relays.

The fault records provided from the field have the following names (Figure 1), which identify the fault line and the line terminal from which the record was taken.

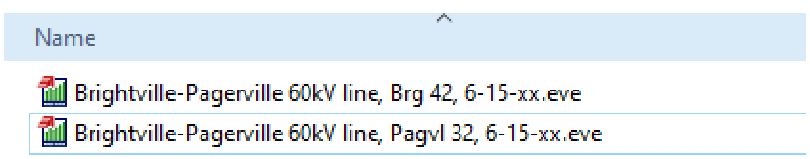

Figure 1. Fault Records

The naming convention, in this case, shows good practice for manual usage, but such a naming convention needs to be applied consistently across the utility. As a minimum requirement for an automatic fault location tool to be able to process the record automatically, the name should identify the following:

- Line name
- Line terminal
- Circuit number
- Date and time of the record
- Voltage level

IEEE proposed the use of a naming convention for fault records in IEEE Std C37-232: "IEEE Standard for Common Format for Naming Time Sequence Data Files (COMNAME)." [2] This standard is an excellent resource for utilities that are in the process of developing a naming convention for fault records.

The AFL program automatically loads the data from the short circuit model and identifies the line for which the fault record was provided. For the first example, let's examine problems that can arise when the entered data in the short circuit model did not follow a strict convention. The complete topology of the line in the short circuit model is shown in the screenshot in Figure 2. The line lies between the BRGHTVILLE and PAGVILLE buses and consists of three sections and two taps branching out along the line. None of the buses were marked as tap buses (otherwise, there would be a “T” showing under the bus).

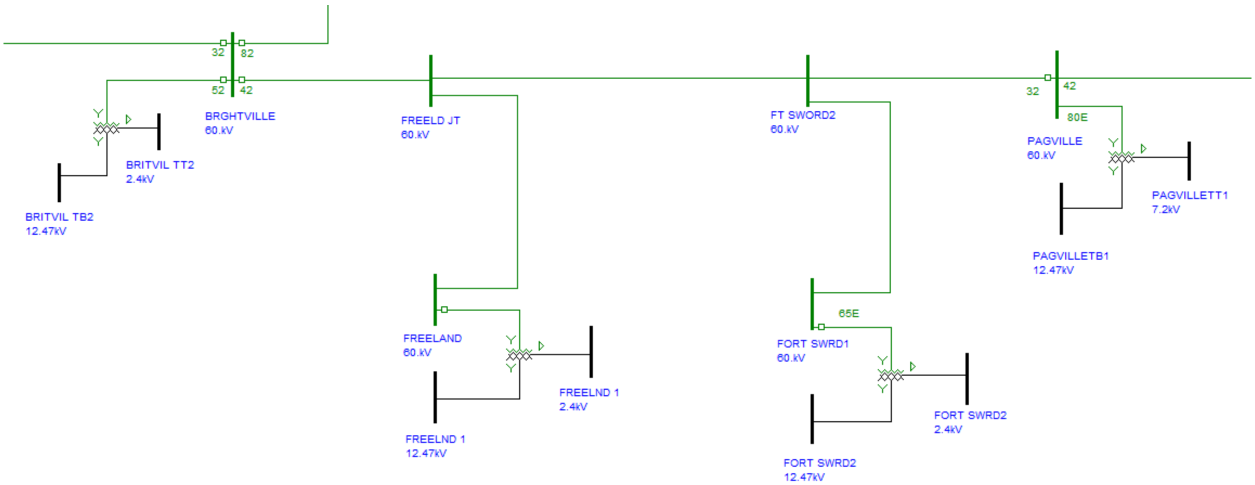

Figure 2. BRIGHTVILLE–PAGERVILLE 60-kV Line

When the AFL program is parsing the faulted line, it will not be able to correctly identify the topology. In this case, the user will have to assist and manually select the line that should be used for further processing. As shown in Figure 3, the program displays the line as five separate lines rather than a single “BRIGHTVILLE–PAGERVILLE” line. Because no line name is entered in the model, the fault location program uses the from-bus and to-bus to generate a line name.

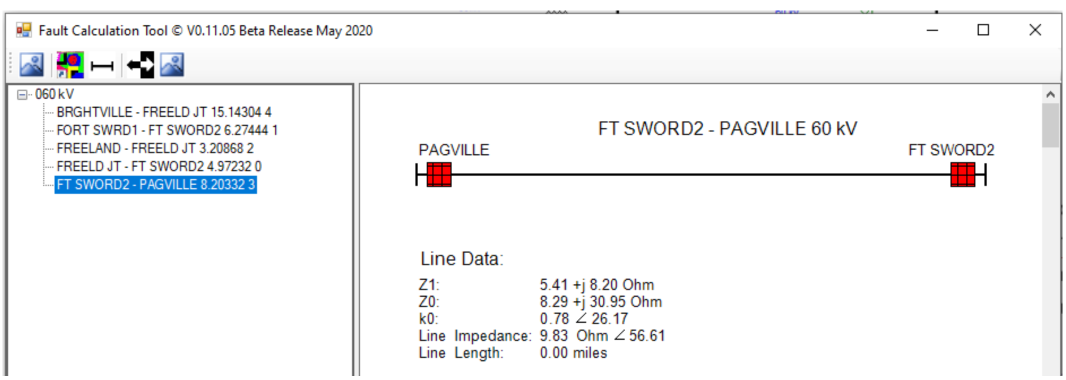

Figure 3. Load Network in Fault Location Program

If the line is incorrectly identified as “FT SWORD2 - PAGVILLE”, the fault location results may not be useful. If we load the fault records and calculate the fault location using various techniques, we see the fault location estimates range from ~200–600% of the line impedance (Figure 4).

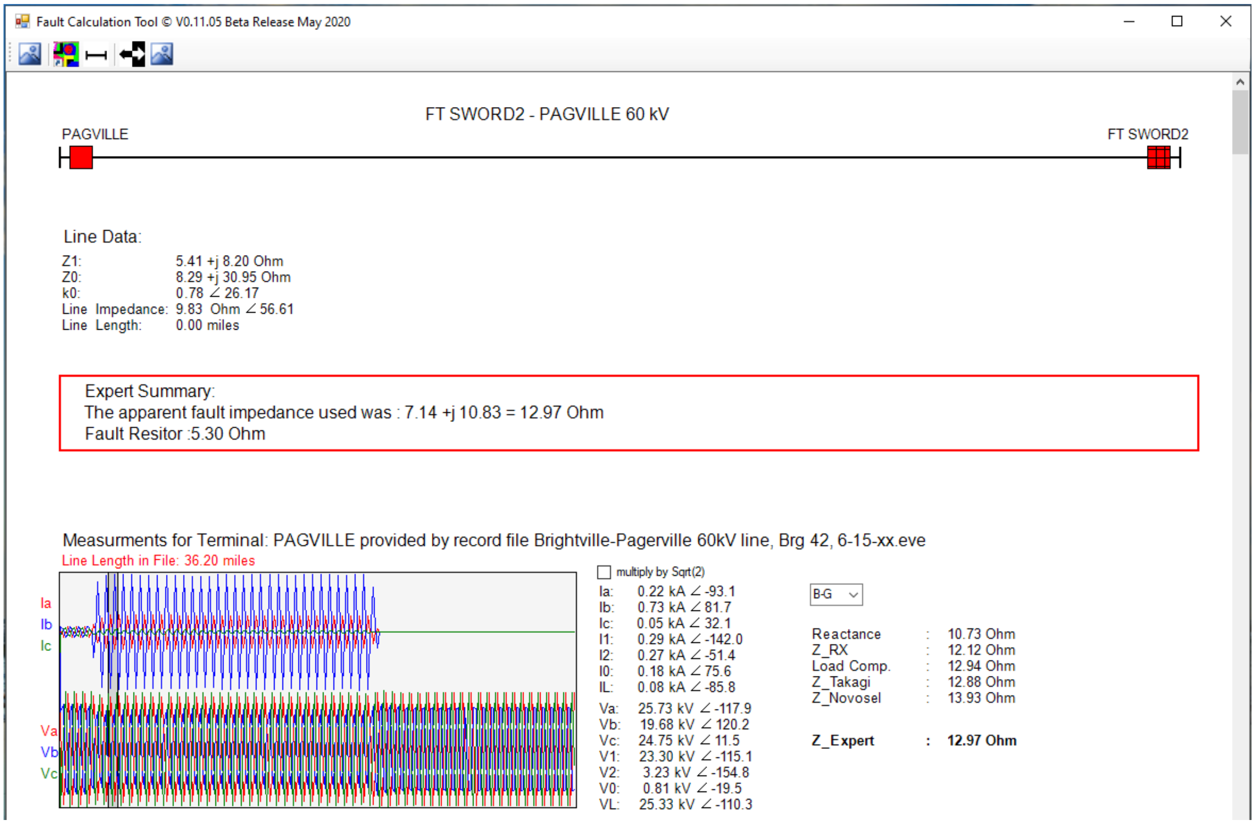

Figure 4. Load Fault Records in Fault Location Program

Obviously, the result is not very useful without the location calculated along the correct line; in this case, we are missing the following information in the short circuit model for the fault location program to function as expected:

- Line name and line length
- Which intermediate buses are tap buses

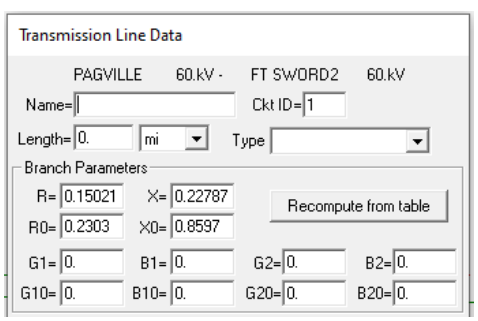

Figure 5. Missing Line Information

The protection engineer can manually fix the bus types for the mid-point buses assuming they have an operating diagram (a system map drawing) and GIS database with the line lengths. While it is easy to manually add the missing information to the short circuit model for a single case, consider the effort of adding missing data for numerous lines across the entire transmission short circuit model. Depending on the size of the network, this could multiply the workload by hundreds or even thousands.

This lack of consistency in the short circuit model reveals a major issue: an insufficient or non-existent naming convention for the network bus and transmission lines. The next section details how the network model was automatically updated to correct several problems.

III. Update the Short Circuit Model Automatically

The AFL system has the potential to automatically load the short circuit model and fault records, calculate and identify faults, and generate an automated report. However, to realize the tool's full potential, a short circuit model with clean data is required. To address existing issues in the short circuit network model, the main improvements implemented are as follows:

- A: Create naming convention
- B: Correct bus type
- C: Line length

These three issues are covered in the sections below.

A. Create Naming Convention

A naming convention for power system elements in the short circuit model must be utilized for an automated AFL program. This naming convention should be used in all utility databases that store information about the power system elements so that the transmission lines and other elements can be identified across different platforms, including the short circuit model, GIS database, etc. The naming convention(s) will form the basis to enable the automated process for bringing in data from another platform (e.g., line length, relay settings, etc.). In this case, the naming convention should have a clear and simple format that is unique for every power system element and, of course, convey information about the power system element. In implementing this project, we discovered that such a naming convention already existed for transmission lines in the GIS database and asset management database of this utility; however, it had not been utilized for short circuit modeling.

The existing naming convention in the GIS database consists of a 5-digit line number for each transmission line. The first digit of the 5-digit number stands for the voltage level: 6 means 60 kV, 1 means 115 kV, 2 means 230 kV, and so on. Digits 2 through 5 comprise a unique identifier assigned to the line (e.g., 0001). For tapped lines, the line number will have a letter at the end. For example, the line number for the BRIGHTVILLE–PAGERVILLE transmission line is 60001. The two tapped lines on this line are FREELAND TAP and FORT SWORD TAP, with line numbers 60001A and 60001B. The line number is also unique for each transmission line, meaning it fits the requirements of simplicity and usefulness. Therefore, it makes sense to use this existing convention to bring data into the short circuit model.

B. Update Bus Type

The first improvement is to have a cleaner line topology so that the AFL or other automated program has a way of identifying where the line starts and where it ends. This can be achieved by cleaning up the bus types in the short circuit model. Once all the mid-point buses are marked as tap buses, it is easy to get the correct line topology by defining line boundaries as real buses.

Figure 6 shows how the fault location program correctly identified the line topology once the buses were correctly marked. If we load the fault location at "PAGVILLE" in this topology, we will get a much more meaningful fault calculation result.

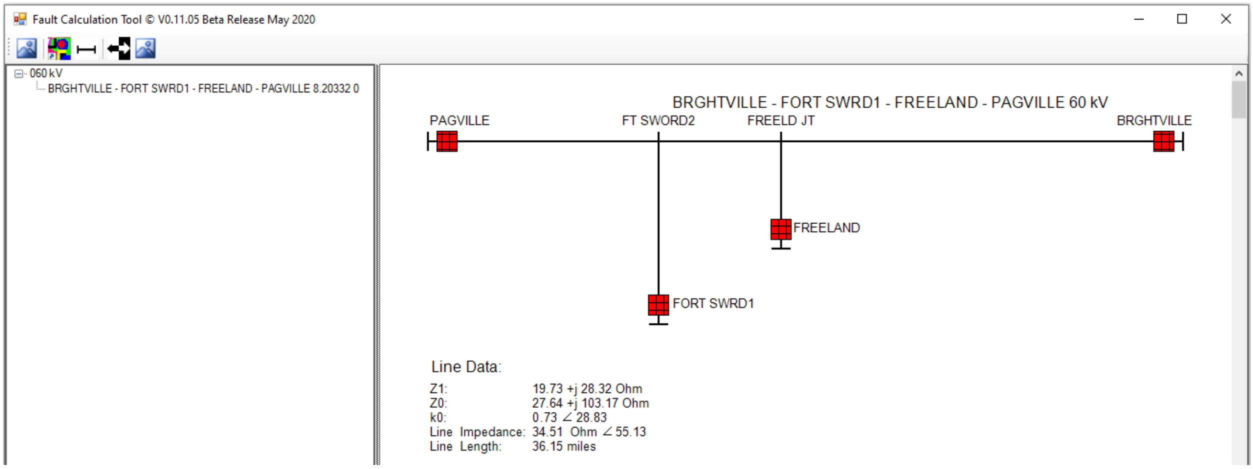

Figure 6. Load Updated Network in Fault Location Program

How to decide if a bus is a mid-point bus that needs to be updated to tap bus? One way is to identify line topology visually and correct mid-point buses along the line manually using the operating diagram (a system map drawing) as a reference, one line at a time. Another way is to find the all buses and line sections for any given number of lines automatically, mark only the mid-point buses as tap bus, in an automated fashion. In this case, the existing naming convention of bus names made the second option—systematically update the bus type—possible. The naming convention for a transmission line is "Substation A – Substation B #n, xxkV", where end-point buses are named after Substation A and B, n is the circuit number, and xx is the voltage level(e.g., BRIGHTVILLE –PAGERVILLE #n, 60kV). If we can automatically find buses "Substation A" and "Substation B", and the buses in between, we can resolve this issue. This process is broken down in two steps described as the following.

<u>APPROXIMATE STRING MATCHING</u>: In this project, the utility did not employ a strict bus naming convention across the entire utility. As a result, the bus names were different depending on what data source was used. In this example, the field crew named the fault records using the full name of the bus terminals, whereas, in the short circuit model, abbreviations were often used.

To enable automated bus marking, the first step is to find the bus representing Substation A and Substation B, which is often not a one-to-one match (i.e., full name vs. abbreviation). Using the BRIGHTVILLE–PAGERVILLE line as an example, Table 1 summarizes the substation names versus the bus names representing the substations in the short circuit program (differences are highlighted in red).

Table 1. Substation Name versus Bus Name

Substation Name	Bus Name
BRIGHTVILLE	BRGHTVILLE
PAGERVILLE	PAGVILLE
FREELAND	FREELD JT
	FREELAND
FORT SWORD	FT SWORD2
	FORT SWRD1

In this example, the bus name is rarely the same as the substation name (e.g., only one of the FREELAND buses); also, there are variations of bus names representing the same substations (e.g., FREELAND and FORT SWORD).

Approximate string matching (also referred to as fuzzy string searching) is a technique to provide solutions for such issues. The algorithm used in the process is the Levenshtein distance algorithm. It calculates a "distance" between two strings, which represents how close the two strings are. Given a threshold of this distance, the automated process can decide if a string is similar enough to the target string. The algorithm helps find the actual bus names in a short circuit model given an approximate string (the substation name). For a substation with multiple buses modeled at the same voltage level, this automated process will return multiple possible matches for further processing.

Having found the corresponding bus(es) that represent Substation A and Substation B, the next step is to find the correct path containing one or multiple line sections in the short circuit model.

FIND LINE BETWEEN TWO BUSES: A common problem is finding the shortest path between two nodes in a network (significant research has been performed, and many algorithms have been created to address this issue). However, the challenge here is that 1) the shortest path is often not the line we are looking for, and 2) the two nodes are not defined (rather, there is a list of possible two-node combinations).

An algorithm was developed to do the following: 1) search all possible paths from both nodes and find multiple paths where the paths from both nodes meet, 2) repeat for all possible node combinations, and 3) return all possible paths. In this paper, we will refer to this as the Flooding Algorithm.

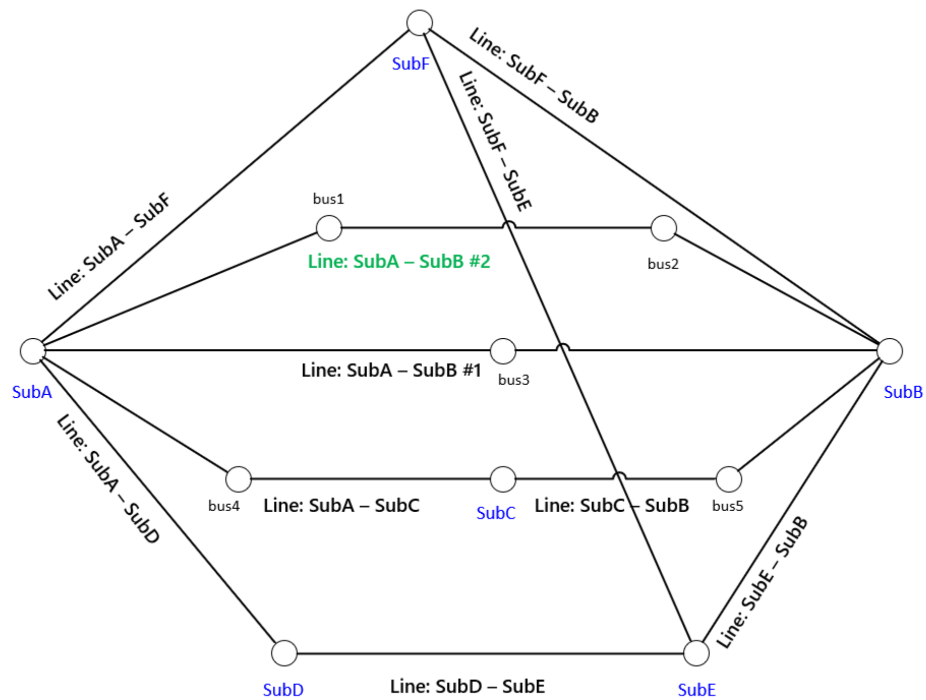

Figure 7. Network Example

Suppose we are trying to find the path of line SubA – SubB #2 from the network above (shown in green). All the nodes are buses in the short circuit model. Six of them with the name “SubX” represent substations that are end terminals of transmission lines, the rest of buses are mid-point buses. All transmission line names are marked on the graphic only, and all buses are real. Assuming that we have correctly identified the buses that match to substations SubA and SubB in the approximate string matching. As the figure 7 shows, there are several possible line paths and the task is to identify the path for line SubA – SubB #2, mark any mid-point buses as tap buses, and update the line sections with the agreed-upon naming convention.

One way to identify the line, would be to select the shortest path between SubA and SubB , what would return two paths in this example:

- SubA – SubB #1
- SubA – SubF and SubF – SubB.

However, finding the shortest path is not a reliable criterium, as neither of these two paths are the line we are looking for. So, we need to expand the selection of paths from using the minimum number of nodes to a larger group to find the target line. The Flooding Algorithm returns all the shortest paths and all the paths n nodes longer than the shortest path (n is a user definable parameter that limits the selection). Figure 8 presents what will be returned in the network example, given n = 1. In this case, the correct path of the target line is displayed in the fourth row. User can select the correct path for the automated process to tell the program which path is the target line.

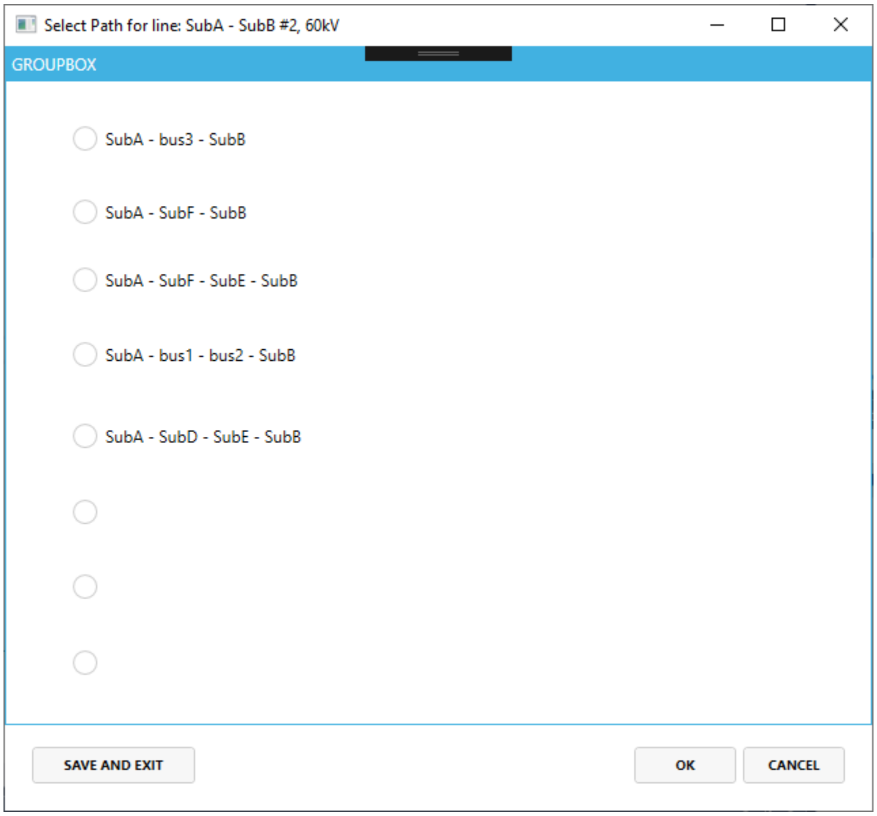

Figure 8. Select Path for Line

In our approach, another program was developed to automatically update and correct the data issues.

For quality assurance purposes, all findings and proposed updates were stored in a mapping file (generated in the steps mentioned in II. B). The file contains two lists: 1) a list of buses that need to be updated to tap buses and 2) a list of line sections with the correct line name and circuit number. An auto-update program can read this mapping (Excel) file and update the short circuit program automatically after review and approval of the changes (in the example of the ASPEN OneLiner OLR file, this happens via an Application Programming Interface, or API functions). The auto-update program automates the process of updating the model, which can save days or weeks of effort updating thousands of buses and line sections without creating unintentional results.

<u>RESULTS:</u> The automated process mentioned above automates 80% of the effort in improving the short circuit model data quality without the need of checking the operating diagrams (system map drawings). It minimizes the manual work of identifying each line and bus to be updated, automatically generates an Excel file summarizing all proposed changes, and can update the short circuit model automatically once the file is reviewed and approved. So, while this workflow and algorithm does not yet work 100% for the more complicated (approximately 20% of) lines (e.g., multiple-terminal lines, or tap lines that is tapped on another tap line); the intermediate Excel file does allow engineers to manually prepare some of the more complicated lines for auto-update.

C. Calculate and Update Line Length

Line length is a very important piece of required data for the fault location program to generate a desirable result. With all the line topology "cleansed" in the network model, the next step is to map the short circuit model of the transmission line with the line length data typically stored in a different platform (e.g., in a GIS database or Lidar data files/databases). The GIS database provides a holistic view of the network in terms of geocoordinates, maps, and calculated lengths of conductors, while Lidar data provide a more detailed line conductor data to generate impedances and lengths and may be more accurate because it measures the actual conductor length (rather than providing a calculated estimate). In this paper, we focus on the GIS database and discuss where in the database the line length data is stored.

There are primarily two types of GIS data that can be used to import the line length into the short circuit model. One is the GIS section length, where transmission lines are broken down into sections, and the length of sections are stored in the GIS database; the other is transmission tower and their geo-coordinates. The engineers at utilities are generally more familiar with the GIS section length, so it is often preferred when updating the line length in the model. However, the section breakdown is not always consistent with what is in the model, so the geo-coordinates are also used in calculating the length when the GIS section length is not suitable for use. Figure 9 presents an example of using geo-coordinates to calculate the length of the BRIGHTVILLE–PAGERVILLE line.

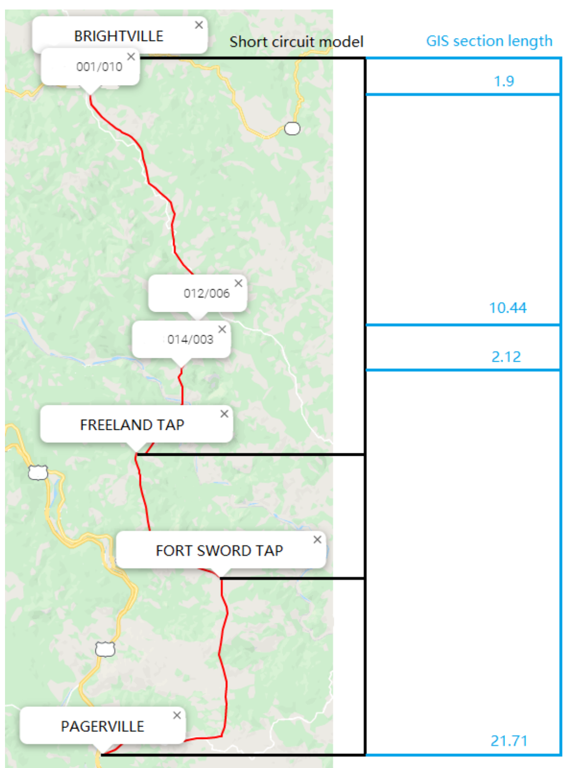

Figure 9. Line Section Breakdown in Short Circuit Model and GIS Database

As shown in Figure 9, the BRIGHTVILLE–PAGERVILLE line is modeled in the short circuit program as three separate line sections with the mid-points being FREELAND TAP and FORT WORD TAP. However, in the GIS section lengths (in blue), the sections are represented differently than in the breakdown from the short circuit program (making the section lengths not suitable for update). In this scenario, a list of tower coordinates is needed to calculate the length based on the model in ASPEN OneLiner. The short circuit model breaks down the line at FREELAND TAP and FORT SWORD TAP. If we attain the geo-coordinates of towers for line 60001A FREELAND TAP and 60001B FORT SWORD TAP, and find the tower on those lines that is closest to line 60001 BRIGHTVILLE–PAGERVILLE towers, then we can find where to break down the total length of the line and update the short circuit model accordingly. This is another process that is automated and proved to save engineers huge amount of time with high data accuracy.

IV. Additional Benefits for an Updated Short Circuit Model

With the line name, topology, and line length information standardized and entered into the short circuit model via the agreed-upon naming convention(s), the fault location program can be fully utilized. Additionally, this higher-quality data can be used as primary keys (unique identifiers) to link to data in other data sources. Figure 10 presents an overview of other potential applications.

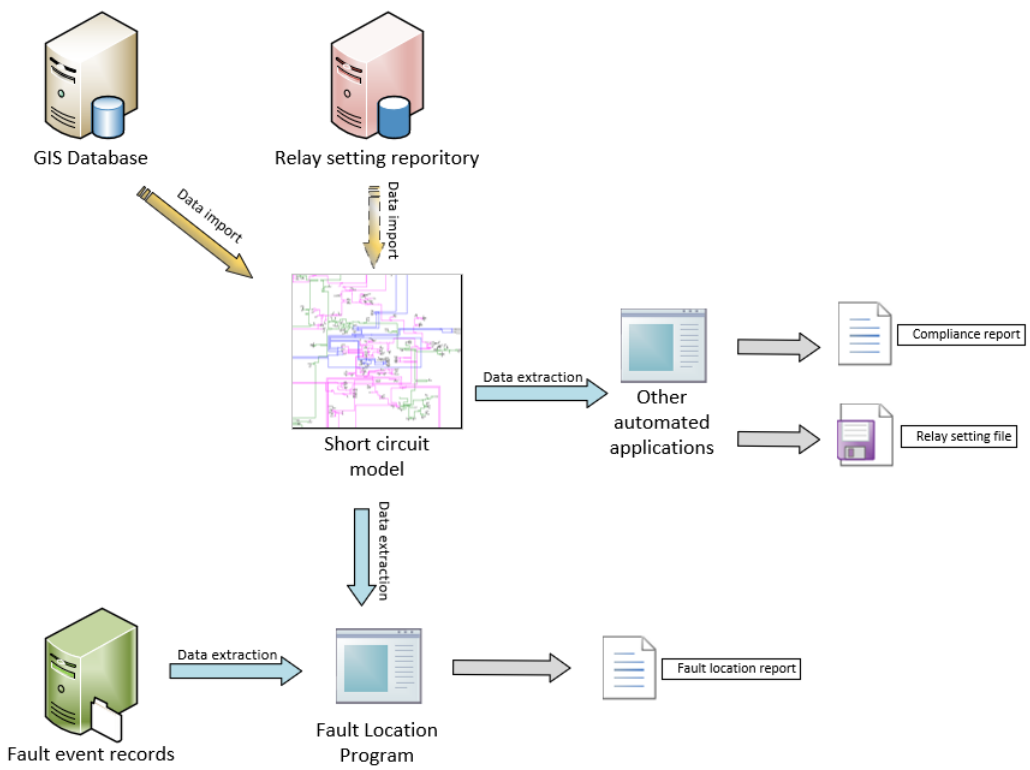

Figure 10. Potential Application and Usage

As shown in Figure 10, the fault location program can extract data automatically from the short circuit model and fault event records and generate a fault location report. This is made possible with the improvements mentioned in the previous section (line topology cleaned up and naming convention established). Since the additional information is added to the short circuit model, other databases such as the relay settings repository can also be linked, which would enable automatic relay setting import.

Other automated applications can also extract and use data from the short circuit model to further automate other studies, for example, to help create relay settings.

V. Project Collaboration and On-Going Efforts

The implementation of this project required collaboration among various team members. Some key points are highlighted below.

A. Protection Engineer Feedback

Protection engineers at the utility generally spend a lot of time using or maintaining the short circuit model. Cooperation is essential in addressing the challenges in improving the short circuit model, and it is important to keep the protection engineers involved, get their feedback, and work together to achieve the best result.

According to the protection engineers' feedback, the improvements mentioned in this paper are 80% complete, and the efforts are benefiting their day-to-day work. These improvements not only benefit the fault location program and other automated studies, but they also help manual fault location studies by making high-quality line length data readily available within the short circuit model. Creating a standard bus-type naming convention also helps create consistency across the protection group for use in fault studies. Such an improvement initiative takes extra time and effort in the short-term, but the long-term efficiency gains are worth the effort.

B. Cross-Department Collaboration

GIS or asset management databases are generally not managed by the protection department and do not always use the same platform. Further, the data often use different naming conventions or identifiers than those used for power system protection.

With the help of the database administrator(s) or engineers from other departments, we can better understand and explore data that may be useful. For example, the line number convention from Section II was not previously used by (or familiar to) most protection engineers, and the GIS database is also a great example of the data complexity inherent in databases designed for different uses. The GIS database administrator was instrumental in identifying the data that best suited our interest. For the efficient use of data managed by other departments, the importance of cross-department collaboration cannot be stressed enough.

C. Maintenance Plan

The data cleaning and overall improvements made to the short circuit model require a significant up-front investment, but it is also important to recognize that a long-term maintenance plan is also needed. Protection engineers are the main stakeholders of the short circuit model when an update is required. Once the improvement efforts are completed, training is recommended to bring everyone up to speed with what happened and why. An updated guideline for the short circuit model also needs to be created and distributed so that a new process in the protection department is established going forward.

VI. Conclusion

As technology advances, more and more data are being generated, stored, and distributed at utilities. The value of the data available at the utility is often underestimated and underutilized. This paper shows an example of how data from other departments can be used to benefit the protection department in the execution of fault analyses and the potential and advantages of initiating such efforts. At the same time, it reveals some of the reasons why the data is not currently being fully utilized. The main reasons for this are:

- Lack of naming convention(s): Strict and consistent naming conventions in the short circuit model require extra maintenance efforts, which adds a burden to the daily work of protection engineers.
- As various data entities are created and maintained by different departments, it is difficult to coordinate and implement universal naming conventions in advance, particularly if the overall goal is vague.
- Automation is often the main driver behind data quality improvements. When automation is not used to its full extent, the issue is not a top priority.

This paper discussed the short circuit model data requirements and improvements needed to realize the AFL program's full potential. It is an example of how automation can be the driver behind data cleaning and how automation can be used to clean the network model data.

The AFL system is a roadmap for automating the entire fault identification, calculation, and reporting process. We have removed the roadblocks in the short circuit model and would like to apply the experience to aligning data in other steps of the process. For example, currently, fault records are primarily pulled and saved manually. This creates a similar challenge: the event records do not have a consistent naming convention for automation usage. Which terminal of the line does the fault record belong to? Which relay generated the record? With the protection elements of all relays modeled in the short circuit model, the fault location program can be used to simulate faults and check relay operations to validate the field operation of actual relay events and potentially identify and report a slow circuit breaker. But to achieve this goal, an investigation is needed into the requirements and potential improvements of fault record extraction and renaming.

BIBLIOGRAPHY

[1] M. Anderson, J. Sykes, J. Holbach, E. Udren, D. Novosel, X Dong, "Addressing the Challenges of Building a Fully Automated Fault Location System," Georgia Tech Relaying Conference 2019

[2] IEEE C37.232-2011 "IEEE Standard for Common Format for Naming Time Sequence Data Files (COMNAME)"

[3] IEEE Guide C37.114 "Guide for Determining Fault Location on AC Transmission and Distribution Lines".

[4] Takagi, et al.," Development of a New Fault Locator Using the One-Terminal Voltage and Current Data" IEEE Transaction on PAS, Vol PAS 101, No8, August 1082, pp.2892-2898

[5] W. Peterson, D. Novosel, D. Hard, T.W. Cease, J. Schneider, "Tapping IED Data to find Transmission Faults", IEEE Computer Applications in Power ,Volume: 12, Issue: 2, Apr 1999.

[7] C. Black, M. Behe, G Wilson, P Myrda, "Automated Fault Location Analysis", EPRI 2016.

BIOGRAPHIES

Xinyang Dong received her MASc degree from North Carolina State University and has been with Quanta Technology since 2016. During her time at Quanta Technology, she has contributed to wide-area coordination studies, line and transformer protection, feeder studies and relay coordination, fault location algorithms, and short circuit analysis using industry-standard short-circuit packages. She has also automated protection studies, and process of data validation, cleanup, and management in various programming languages.

Juergen Holbach, PhD, *Senior Director of Automation & Testing,* has more than 30 years of experience in the design and application of protective relaying. An IEEE member and chairman, he has published over a dozen papers and holds three patents. In 2009, Juergen received the Walter A. Elmore Best Paper Award from the Georgia Tech Relay Conference. Juergen's areas of expertise include Protection & Automation, Automated Fault Location, Impact of inverter-based resources onto protection, RTDS Testing, and IEC 61850 compliance.

Robbie James received his BASc from University of California, Berkeley and MASc from California State University, Fresno. He started his career at PG&E in 2008 in distribution engineering and has held his current position as a transmission system protection engineer since 2010. He is a registered professional engineer in California.