Optimal Power Flow Contingency Planning for Electrical Grid Systems under N-K Failure Scenarios

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the requirements for the Degree Master of Science

(Systems Engineering)

Sarah Woehleke

May 2015

TEXT and rectangles in blue will NOT show on printed copy

Type final title of thesis or dissertation (M.S. and Ph.D.) below. If your title has changed since your submitted an Application for Graduate Degree, notify Graduate Office.

Optimal Power Flow Contingency Planning for Electrical Grid Systems under N-K Failure Scenarios

A Thesis

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment

of the requirements for the degree

Master of Science

by

Name

Sarah Woehleke

Month degree is awarded August

Year 2015

APPROVAL SHEET

The thesis

is submitted in partial fulfillment of the requirements

for the degree of

Master of Science

S. Wehlke

AUTHOR *signature*

The thesis has been read and approved by the examining committee:

Please insert committee member names below:

Alfredo Garcia

Advisor

James Lambert

Michael Smith

Accepted for the School of Engineering and Applied Science:

James H. Ay

Dean, School of Engineering and Applied Science

Month degree is awarded August

Year 2015

Print Form

Abstract. The current N-1 requirements for electrical grid contingency planning are no longer adequate to address the increasing complexity of a grid system containing growing numbers of automated components and renewable energy generation sources. The often unpredictable nature of renewable generation resources coupled with the vulnerability of the automated components to cyber interventions opens up grid systems to the possibility of coordinated cyber-physical attacks on multiple grid components as well as simultaneous multiple component failures. This research will address methods for grid security contingency planning on N-K failure and attack scenarios by testing different parameters within the Cross-Entropy Method framework for combinatorial optimization.

Contents

1. Acknowledgments

I would like to acknowledge the Commonwealth Coalition for Advanced Logistics Systems (CCALS) and the support of the LMI Research Institutes Academic Partnership Program along the significant guidance and expertise provided by members of LMIs Energy and Environment Group including Margaret Brennan Hogan, Stuart Funk, and Karen Munson. I would like to personally thank everyone who supported me within the CCALS team, especially Thomas Polmateer, James Lambert, and Mark Manasco.

I would like to extend a special thanks to Michael Smith for always asking the best questions and prompting thoughtful discussions on any topic. Finally, I owe so much to my advisor, Alfredo Garcia, for the countless hours of question-answering and guidance, without which this research would not have been possible.

2. Problem Statement

Utility and micro-grid operators, as well as outside evaluators, must choose a method for evaluating both cyber and physical system security under dynamic conditions in a system that is vulnerable to both malicious attacks and equipment malfunctions. Current grids are set up to maintain full operation should one component fail, but the emergence of new threats and technologies has highlighted the need for security planning under N-K failure scenarios. The problem inherent to N-K contingency planning is the combinatorial complexity that arises when calculating the total number of N-K attacks on a system with a large number of individual components. In such a case the individual likelihood of any N-K failure occurring without prior knowledge of the system is close to zero.

Despite the low probability of any particular attack, there is still a need to protect the system against unlikely attacks that could be extremely costly or cause a system failure. In risk analysis terms these failures are known as a low probability/high consequence events. Many industries, like transportation/aviation, attempt to model these events in order to increase the secure operation of their systems. In these cases, because the likelihoods of any failure happening are small and undifferentiated, the severity/cost as a consequence of the failure becomes the essential variable in differentiating the possible failures.

Many problems faced by power system engineers (or network engineers in general) can be categorized as combinatorial optimization problems; for example: transmission network expansion planning problems, unit commitment problems, optimal placement of phasor measurements units for state estimation, or the design of special protection schemes. A generic combinatorial optimization problem can be written as:

$$u* = arg\max_{u \in U} S(u) \tag{1}$$

where $S(\cdot)$ signifies the function of performance and U is the discrete search space of the optimization problem. Because U is generally of a large dimension - i.e. the search space is described by a large number of components. Here is where the term "combinatorial explosion" becomes relevant; solving a combinatorial problem through simple enumeration creates a complexity that grows exponentially as each subsequent dimension is considered within U. Therefore, simple enumeration cannot reasonably be used without the assistance of extreme computational power where U is large in dimension.

In this example, U is the n-dimensional set of operating costs per hour associated with possible combinations of n-k failure scenarios of any given grid system. Even if k is limited to a fraction of the total n components, enumerating the total set of failures creates the exact combinatorial explosion described above. So then the problem at hand is how to evaluate and estimate a probability distribution of the set of failures in U in a way that maximizes the available computing power while minimizing processing time and variance of results.

Specifically, the goal is to develop an algorithm for solving the combinatorial optimization problem described by Equation (1). If we let X be a random variable taking its value in some discrete space X with a probability mass function (pmf) $f(\cdot)$, $S'(\cdot)$ be a real-value function defined on X and γ be a real number. In the rare-event simulation context, we are estimating the probability of occurrence l of an event $\{S'(X_j) \leq \gamma\}$. Here, we must take to considerations into account:

(1) The event $\{S(u) \leq \gamma = S(u*)\}$, where u is a random variable taking its values in U with a pmf $f_u(\cdot)$, tends to be a rare event. For example, if $f_u(\cdot)$ is a uniform pmf possessing a unique maximum, the probability of this event is $\frac{1}{\#U}$;

(2) When solving this rare-event problem, the goal of the method is to obtain, as a result, a pmf which is close to the real one, and, therefore, likely to generate samples u for which the difference between the value of $S(u)$ and the 'real' probability of u is minimized.

Based on these two considerations, several examples of network optimization, both in and out of the power industry, are relevant to solving the problem at hand. These will be explored in the following section in depth, and include minimal cut problems with network applications and cross-entropy optimization (both in and out of the power context).

In addition to developing a combinatorial optimization algorithm that minimizes the variance of our estimated pmf, there still exists the problem of how best of present the resulting pmf to a utility operator to inform their decision-making process. While developing an interactive dashboard is the ultimate goal, it still exists well outside the scope of this research. The literature review will include a few example of applications of decision-making tools, and we will ultimately provide recommendations as to the type of interface that can be constructed to maximize the usability of this information for stakeholders of varying degrees of technical proficiency.

3. Literature Review

3.1. Background

Large amounts of time and effort have been devoted to securing the nations electrical grid system against equipment failure from malfunction or malicious interference in order to prevent significant power outages. Due to the large-scale and complex nature of the system, the process of securing vital grid components is fraught with complication. As electrical grids grow and change, the security process grows and changes as well. Consequently, there is a need for a dynamic decision making tool for utility planners and managers of micro-grids, which will enable them to target investments in the grid for maximizing the overall security of the electrical system.

In addition to the standard concerns about security and risk management, the growing significance of computer systems and renewable energy electrical generation sources in the operation of the power grid brings about the need to secure these components against previously unforeseen vulnerabilities. Consequently, while previous grid configurations were not especially vulnerable to cyber attacks, these are very real threats to the new systems. Much work has been done in securing the electrical grid from cyber attacks and failures of computer systems, but the addition of unpredictable electrical generation from renewable sources (wind and solar specifically) opens up the system to an entirely new set of possible cyber and physical failures.

Current reliability standards for electrical grids focus on the systems being able to withstand an attack or failure of a single component of the power system. This means that should an essential generating facility fail, there must be another facility that can produce the required amount of electricity to make up the difference in production to meet demand in order to minimize load shedding or loss of service.

Consequently, certain facilities are redundant and must be kept up and running for security purposes, even if they might not be needed at a given time; this extra generating capacity is known as spinning reserves. In the case of a substation or a transmission line going down, there must be an alternate path over which electricity may be routed so that demand can still be met. These requirements are known as N-1 contingency plans, although more recently some standards have been requiring this ability for two-component coordinated attacks or failures (N-2). There are no current requirements for more than two coordinated component failures.

Despite this deficiency in current requirements, a coordinated attack of multiple components or simultaneous (or cascading) failures of multiple components is not unlikely, and may be beyond the scope of existing contingency plans. Therefore, there is a distinct need to address the problem of multiple failures with an additional method of contingency planning. Because of the unknown and varying nature of this type of attack, the process of security planning is known as N-K contingency planning (K being the number of possible components taken down in a coordinated attack).

This problem will certainly play a future role in planning of power system operation for both large-scale power systems and micro-grid installations, and current security software systems (Energy Management Systems or EMS) will need to be updated to take this into account. To this end, there is a need for the development of a methodology for evaluating the N-K system failures and informing relevant stakeholders and systems planners about effectively targeting security investments and optimal grid design.

3.2. Prior Work

While the concept of N-K contingency planning is fairly new, a good amount of academic work has been devoted to solving different aspects of the problem. The primary concern of much of the work is developing an optimization method for determining the most efficient flow of power under N-K failure conditions. The process of determining the optimal flow of power, which minimizes generation cost while still meeting demand given the constraints of the system, is known as economic dispatch or unit commitment.

This process is usually conducted with a linear programming model that determines: (i) the amount of power each generation facility should produce (favoring low marginal cost facilities given no additional constraints) and (ii) how the power should be routed given the capacity and configuration of the transmission lines and substations. The power flow over a given line cannot exceed the maximum flow or voltage on that line, and a power balance must be maintained at the buses or substations. While this flow can usually be maintained given any single component failure under current security requirements, the emergence of cyber security threats and the possibility of coordinated cyber-physical attacks mean that K-1 contingency planning may no longer be sufficient.

The ultimate goal of any N-K contingency planning is to maintain delivery of electricity in order to meet demand, no matter what externalities may occur. With this goal in mind, current research has opted to approach the problem from a few different directions.

3.2.1. Optimization Methods

Optimization methods[2] including Benders Decomposition and Cutting Plane Algorithms have been used to run tests on sample power systems to explore the possibility of identifying optimal power dispatch under different N-K scenarios. This approach is targeted at situations where a specific N-K scenario has already been identified, and a new way to route power needs to be established in order to meet demand. These methods do not take cost into account and will have limited applicability should there be more than one or two possible N-K scenarios that need to be evaluated.

3.2.2. Security Methods

Other researchers have taken the approach of designing or retrofitting a system to withstand N-K scenarios by creating islands that are able to continue running despite failures in other parts of the grid. This method also achieves the goal of limiting cascading failures in a system in which failures occur not because of a specific attack or malfunction but because of reliance on other separate grid components that have failed or been attacked. In fact, several recent papers on N-K security have been focused exclusively on treating the K failures as cascading failures, rather than as attacks.[3]

However, because N-K security contingency planning is a fairly recent consideration in the power industry and has not yet been mandated in federal or state requirements, the existing literature is relatively sparse. In order to find relevant prior work, it is necessary to delve into other optimization problems that have similar challenges to solving the N-K problem. Since the issue is by nature a combinatorial problem, the closest parallels can be found in optimization algorithms that attempt to address combinatorial problems.

3.2.3. Combinatorial Optimization

Combinatorial optimization consists of finding an optimal solution from a finite set of solutions, especially in cases where enumerative methods are not feasible due to the issue of combinatorial explosion. Combinatorial optimization methods are often used in network planning and problem solving for issues like optimal investing in security, quickest path problems, and optimal dispatch of resources. Consequently, the method is often used within the power industry, as well as transportation, manufacturing, and computing.

Some of the combinatorial optimization problems faced by power system engineers (or network engineers in general) include: transmission network expansion planning problems [28], unit commitment problems [25], optimal placement of phasor measurements units for state estimation [26], or the design of special protection schemes [27].

Within combinatorial optimization, there are many different commonly used algorithms. Because of the rare nature of each of the various N-K scenarios, the Cross-Entropy (CE) sampling method within the combinatorial optimization framework is ideal.

3.2.4. Cross-Entropy Optimization

The CE method is intended for as an approach for combinatorial and rare event simulation. The method was pioneered in 1997 by Reuven Rubinstein, deriving its name from cross-entropy (Kullback-Leibler) distance which is way of measuring the value of information from an experiment (what are the chances of a surprise result). The distinction of the CE method from other combinatorial optimization methods is that it provides a precise framework for setting updating and learning rules optimally. [11] The method is similar to determining importance measures when developing fault trees.

It has been successfully applied to grid optimization problems involving economic dispatch as well as other difficult combinatorial optimization problems including scheduling problems and combinatorial optimization problems with the genome as well as several well-known problems [11] :

- The maximal/minimal cut problem - Partitioning the points in system into two subsets so that the sum of the weights of the edges going from one set to the other is maximized/minimized;
- The traveling salesman problem (TSP) - Given a set of cities and relative distances, what is the shortest possible route through the set which returns the salesman to the original city having visited each city exactly once;
- The quadratic assignment problem - (factory placement problem) Given a set of n facilities and a set of n locations, a distance is specified for each pair of locations, and for each pair of facilities a weight or flow is specified (e.g., the amount of supplies transported between the two facilities). The problem is to assign all facilities to different locations with the goal of minimizing the sum of the distances multiplied by the corresponding flows;
- The clique problem - Any problem related to finding the complete set of subsets or subgraphs "cliques" in a set or graph given some relation between the relevant elements;
- The buffer allocation problem (BAP) for production lines - Given a number of stations of a line, K, and a fixed number of servers assigned to each station with work allocation, w, how can one best allocated a certain fixed number of buffer slots,

N, among the $K-1$ intermediate buffer locations of a production line in order to meet the specified objective?

These examples can be either deterministic (TSP) or noisy/simulation-based (BAP) and stochastic problems. In this case, the minimal cut problem is very similar to the ultimate goal of our analysis, specifically the case where a network operator is evaluating the smallest possible subset of nodes that can be removed from the system and cause a failure in network communication. This method has been used within the power systems context to assist in extreme event detection [29] [30] [31]

Some more specific applications of the CE method include: continuous multi-extremal optimization [17]; continuous optimal control problems [15]; DNA sequence alignment [20]; mixed integer nonlinear programming [14]; multidimensional independent component analysis [18]; network reliability optimization [22]; neural and reinforcement learning [24]; resource allocation in stochastic systems [21]; and vehicle routing optimization with stochastic demands [23].

Within power system optimization, the CE method has been used to solve unit commitment problems in Damien Ernst (et al)'s 2007 paper entitled "The Cross-Entropy Method for Power System Combinatorial Optimization Problems" [10]. While not specifically addressing N-K contingency planning, the algorithm proposed by Ernst (et al) is designed to solve optimization problems with unit commitment that directly parallel those in N-K contingency planning. Both problems require the enumeration of large numbers of possible combinations of grid components; in the case of unit commitment it is to find the lowest cost combination of components for generation and transmission while in the case of contingency planning it is to find the highest consequence combination of failed grid components.

While these methods all attempt to solve essential parts of the N-K conundrum and similar problems, there is currently no coherent method for informing decision-making in the face of potential N-K failures when it comes to investing in the prevention of potential cyber and physical grid attacks. Such a method would need to take into account the associated costs and potential power flow disruption of specific N-K attacks or failures. While optimal power flow models are essential to solving this problem, they are not sufficient in informing decision-makers about the value of changing or upgrading the existing system in order to prevent the incurred costs and power outages from N-K scenarios. Therefore, the method to be pursued in this research represents an adaptation of CE methods from other arenas to best fit the N-K problem.

3.3. Relevant Legislation

While N-1 contingency planning is the current federal requirement, the need for greater physical infrastructure system security has gained more attention in recent years. The US government has approached this problem with the National Infrastructure Protection Plan which was issued by the Department of Homeland Security in 2006 (this was later revised in 2009 and again in 2013). The plan is intended to set up a framework for interdepartmental cooperation between the government and private sector to ensure the security and resilience of the physical infrastructure system. Figure 1 outlines this plan [8]:

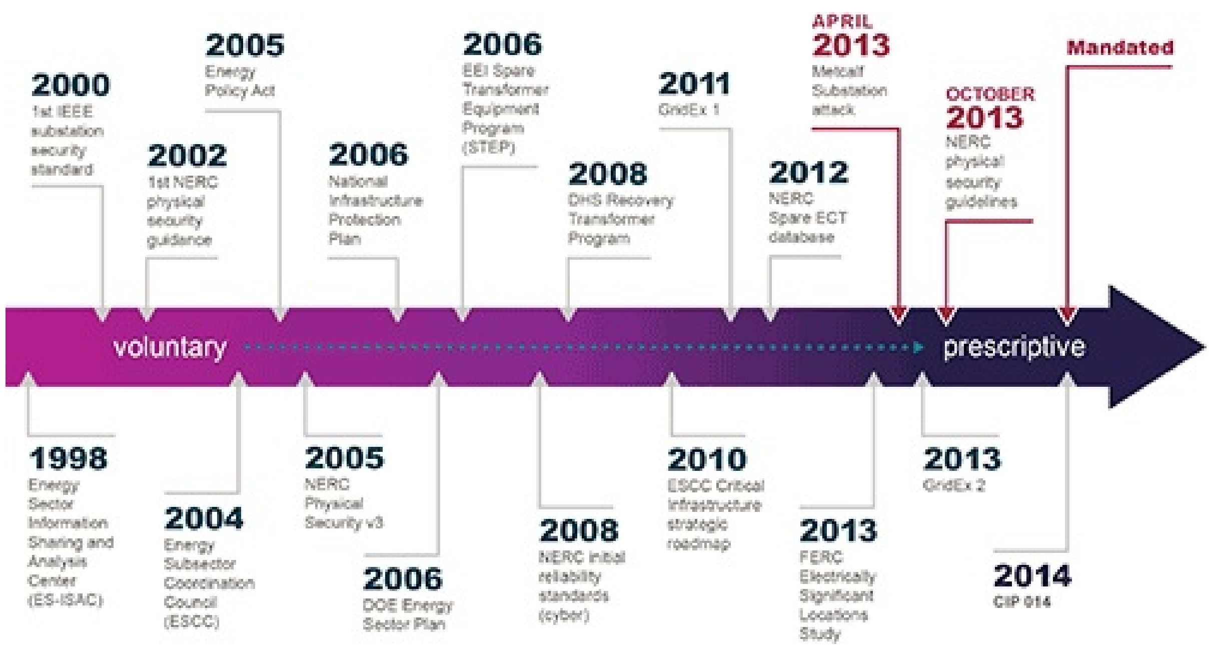

FIGURE 1. Physical Grid Security Timeline

On March 7, 2014 the North American Electric Reliability Corporation (NERC) issued an order for a new set critical infrastructure protection (CIP) standards for physical power system security which were approved by the Federal Energy Regulatory Commission (FERC) and became effective on January 26, 2015 [8]:

"The proposed Reliability Standards should require owners or operators of the Bulk-Power System, as appropriate, to identify facilities on the Bulk-Power System that are critical to the reliable operation of the Bulk-Power System. Then, owners or operators of those identified critical facilities should develop, validate and implement plans to protect against physical attacks that may compromise the operability or recovery of such facilities." [9]

In the order, NERC calls for the following six-step regulatory standard [9]:

(1) Perform a risk assessment based on "objective analysis, technical expertise, and experienced judgment to identify critical facilities
(2) Outside verification of risk assessment by an unaffiliated third party
(3) Communication between transmission system owners and operators
(4) Evaluate potential threats and vulnerabilities of critical facilities
(5) Develop and implement a security plan to protect against attacks on critical facilities
(6) Outside verification of threat and vulnerability assessment and security plan by an unaffiliated third party

There are several implications of the new regulatory standard. First of all, there is a recognized need for additional security to physical power systems. Second, risk analysis methods will need to be able to identify critical facilities within the network at which to target their security plans. Additionally, the requirement for an outside evaluator means that there is an opportunity to market a flexible method for identifying at risk components within a power network without the hands-on expertise that utility planners possess. [8]

4. Methods

4.1. Data Source

The inherent problem with evaluating the security of electrical grid systems as an unaffiliated third party, is that grid data is not made publicly available for security reasons (catch

22). Luckily, IEEE and several academic sources have created test power flow models based on real systems. The two systems used in this research are listed below:

(1) The IEEE 57 Bus Test Case represents a portion of the American Electric Power System (in the Midwestern US) as it was in the early 1960's. The data was kindly provided by Iraj Dabbagchi of AEP and entered IEEE Common Data Format by Rich Christie at the University of Washington in August 1993. [7]
(2) The six-Bus System with 11 transmission lines and three generators is courtesy of Professor Bruce Wollenberg of the University of Minnesota who has kindly made available the programs distributed with his well known textbook, Wood and Wollenberg, Power System Operation and Control 2nd Ed. [6]

The first example (a relatively simple test grid) in Figure 2 provides a good illustration of how combinatorial explosion from the data can limit the analytical methods used:

As shown in Figure 2, the system is composed of 57 buses, 80 lines, and 7 generators. The total number of: N-2 scenarios would therefore be 10,296; N-3 scenarios would be 487,344; N-4 scenarios would be 17,178,87, and so on. As the systems become larger and more complex, the combinatorial problem becomes even more profound. Using traditional statistical sampling methods to pare down the number of N-K scenarios to run would still require a prohibitively large sample for the typical decision-maker.

4.2. Optimal Power Flow Modeling

While the CE method can be applied to optimal power flow (OPF) modeling in addition to standard optimization methods, for this research, the standard optimization models used by the MATPOWER program in Matlab (and by extension, the Python program PYPOWER) will be used. While not the main focus of this work, the OPF solution is the main input for the CE algorithms tested. The OPF solver used to generate the cost of operating the power system under the varying set of failures is the PYPOWER package in the open-source Python programming language. PYPOWER is a port of MATPOWER to Python

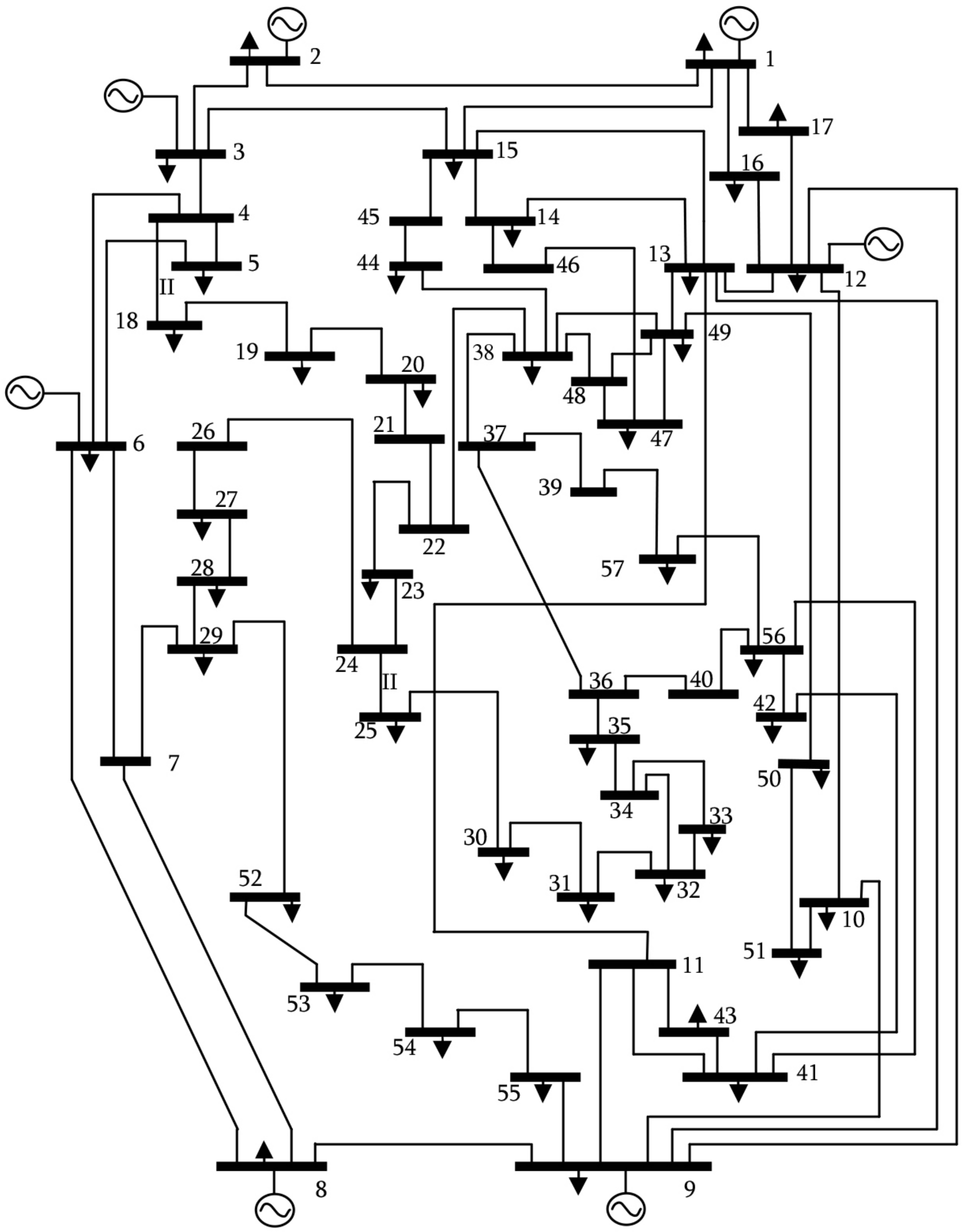

Figure 2. IEEE 57-Bus Test Grid, December 1961 [7]

that solves power flow and OPF problems for AC and/or DC power. For the purposes of estimating costs of N-K failures, the AC OPF module is sufficient.

[From MATPOWER Manual] [13]:

4.2.1. Branches

All transmission lines, transformers and phase shifters are modeled with a common branch model, consisting of a standard π transmission line model, with series impedance $z_s = r_s + jx_s$ and total charging susceptance b_c, in series with an ideal phase shifting transformer. The transformer, whose tap ratio has magnitude τ and phase shift angle θ_{shift}, is located at the $from$ end of the branch, as shown in Figure 3. The parameters r_s, x_s, b_c, τ and θ_{shift} are specified directly in columns BR_R (3), BR_X (4), BR_B (5), TAP (9) and $SHIFT$ (10), respectively, of the corresponding row of the branch matrix.

The complex current injections i_f and i_t at the $from$ and to ends of the branch, respectively, can be expressed in terms of the 22 branch admittance matrix Y_{br} and the respective terminal voltages v_f and v_t

$$\begin{bmatrix} i_f \\ i_t \end{bmatrix} = Y_{br} \begin{bmatrix} v_f \\ vt \end{bmatrix} \tag{2}$$

With the series admittance element in the τ model denoted by $y_s = 1/z_s$, the branch admittance matrix can be written

$$Y_{br} = \begin{bmatrix} y_s + j\frac{b_c}{2}\frac{1}{\tau^2} & -y_s\frac{1}{\tau e^{j\,shift}} \\ y_s\frac{1}{\tau e^{j\,shift}} & y_s + j\frac{b_c}{2} \end{bmatrix} \tag{3}$$

If the four elements of this matrix for branch i are labeled as follows:

$$Y_{br}^i = \begin{bmatrix} y_{ff}^i & y_{ft}^i \\ y_{tf}^i & y_{tt}^i \end{bmatrix} \tag{4}$$

then four $n_l 1$ vectors Y_{ff} , Y_{ft}, Y_{tf} and Y_{tt} can be constructed, where the i-th element of each comes from the corresponding element of Y_{br}^i. Furthermore, the $n_l n_b$ sparse connection matrices C_f and C_t used in building the system admittance matrices can be defined as follows. The $(i, j)^{th}$ element of C_f and the $(i, k)^{th}$ element of C_t are equal to 1 for each branch i, where branch i connects from bus j to bus k. All other elements of C_f and C_t are zero.

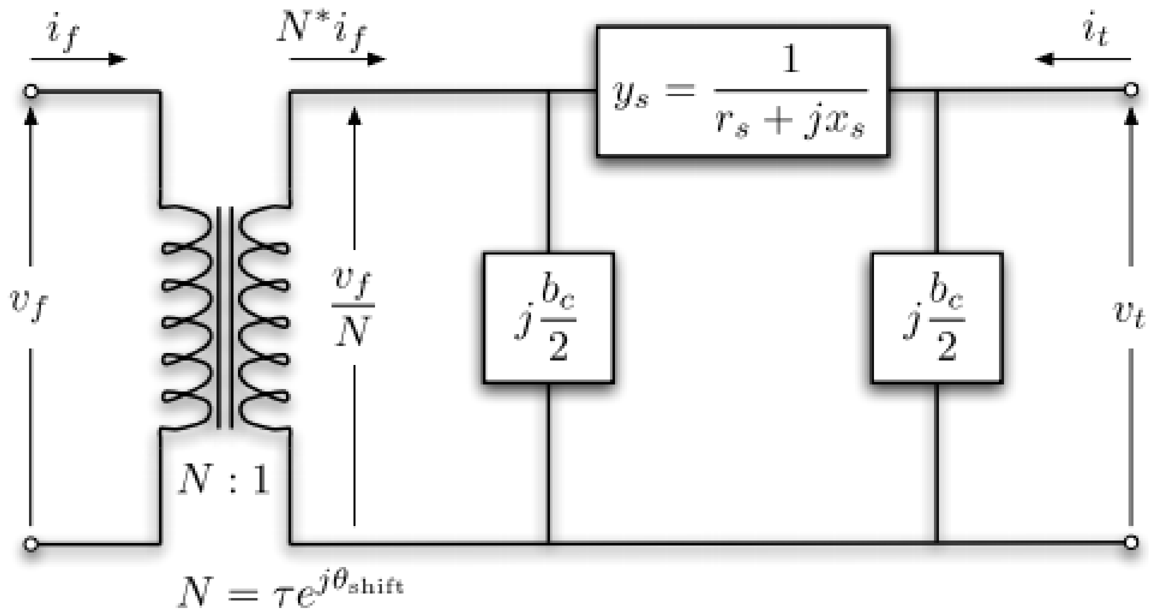

FIGURE 3. Branch Model

4.2.2. Generators

A generator is modeled as a complex power injection at a specific bus. For generator i, the injection is

$$s_g^i = p_g^i + jq_g^i \tag{5}$$

Let $S_g = P_g + jQ_g$ be the $n_g \times 1$ vector of these generator injections. The MW and MVAr equivalents (before conversion to p.u.) of p_g^i and q_g^i are specified in columns PG (2) and QG (3), respectively of row i of the *gen* matrix. A sparse $n_b n_g$ generator connection matrix C_g can be defined such that its $(i,j)^{th}$ element is 1 if generator j is located at bus i and 0 otherwise. The $n_b \times 1$ vector of all bus injections from generators can then be expressed as

$$S_{g,bus} = C_g S_g \tag{6}$$

4.2.3. Loads

Constant power loads are modeled as a specified quantity of real and reactive power consumed at a bus. For bus i, the load is

$$s_d^i = p_d^i + jq_d^i \tag{7}$$

and $S_d = P_d + jQ_d$ denotes the $n_b \times 1$ vector of complex loads at all buses. The MW and MVAr equivalents (before conversion to p.u.) of p_d^i and q_d^i are specified in columns PD (3) and QD (4), respectively of row i of the bus matrix.

Constant impedance and constant current loads are not implemented directly, but the constant impedance portions can be modeled as a shunt element described below. Dispatchable loads are modeled as negative generators and appear as negative values in S_g.

4.2.4. Shunt Elements

A shunt connected element such as a capacitor or inductor is modeled as a fixed impedance to ground at a bus. The admittance of the shunt element at bus i is given as $y_{sh}^i = g_{sh}^i + jb_{sh}^i$ and $Y_{sh} = G_{sh} + jB_{sh}$ denotes the $n_b \times 1$ vector of shunt admittances at all buses. The parameters g_{sh}^i and b_{sh}^i are specified in columns GS (5) and BS (6), respectively, of row i of the bus matrix as equivalent MW (consumed) and MVAr (injected) at a nominal voltage magnitude of 1.0 p.u and angle of zero.

4.2.5. Network Equations

For a network with nb buses, all constant impedance elements of the model are incorporated into a complex $n_b n_b$ bus admittance matrix Y_{bus} that relates the complex nodal current injections I_{bus} to the complex node voltages V :

$$I_{bus} = Y_{bus}V \tag{8}$$

Similarly, for a network with n_l branches, the $n_l n_b$ system branch admittance matrices Y_f and Y_t relate the bus voltages to the $n_l 1$ vectors I_f and I_t of branch currents at the from and to ends of all branches, respectively:

$$I_f = Y_f V \tag{9}$$

$$I_t = Y_t V \tag{10}$$

If $[\cdot]$ is used to denote an operator that takes an $n1$ vector and creates the corresponding nn diagonal matrix with the vector elements on the diagonal, these system admittance matrices can be formed as follows:

$$Y_f = [Y_{ff}]C_f + [Y_{ft}]C_t \tag{11}$$

$$Y_t = [Y_{tf}]C_f + [Y_{tt}]C_t \tag{12}$$

$$Y_{bus} = C_f^T Y_f + C_t^T Y_t + [Y_{sh}] \tag{13}$$

The current injections can be used to compute the corresponding complex power injections as functions of the complex bus voltages V :

$$S_{bus}(V) = [V]I_{bus}^* = [V]Y_{bus}^* V^* \tag{14}$$

$$S_f(V) = [C_f V]I_f^* = [C_f V]Y_f^* V^* \tag{15}$$

$$S_t(V) = [C_t V]I_t^* = [C_t V]Y_t^* V^* \tag{16}$$

The nodal bus injections are then matched to the injections from loads and generators to form the AC nodal power balance equations, expressed as a function of the complex bus voltages and generator injections in complex matrix form as

$$g_S(V, S_g) = S_{bus}(V) + S_d C_g S_g = 0 \tag{17}$$

4.2.6. Power Flow

The standard power flow or load flow problem involves solving for the set of voltages and flows in a network corresponding to a specified pattern of load and generation. [Pypower] includes solvers for both AC and DC power flow problems, but only AC power is used for these purposes both of which involve solving a set of equations of the form

$$g(x) = 0 \tag{18}$$

constructed by expressing a subset of the nodal power balance equations as functions of unknown voltage quantities.

4.2.7. AC Power Flow

By convention, a single generator bus is typically chosen as a reference bus to serve the roles of both a voltage angle reference and a real power slack. The voltage angle at the reference bus has a known value, but the real power generation at the slack bus is taken as unknown to avoid overspecifying the problem. The remaining generator buses are typically classified as PV buses, with the values of voltage magnitude and generator real power injection given. These are specified in the VG (6) and PG (3) columns of the *gen* matrix, respectively. Since the loads P_d and Q_d are also given, all non-generator buses are classified as PQ buses, with real and reactive injections fully specified, taken from the PD (3) and QD (4) columns of the bus matrix. Let I_{ref}, I_{PV} and I_{PQ} denote the sets of bus indices of the reference bus, PV buses and PQ buses, respectively. The bus type classification is specified in the Matpower case file in the BUS_TYPE column (2) of the bus matrix. Any isolated buses must be identified as such in this column as well.

In the traditional formulation of the AC power flow problem, the power balance equation is split into its real and reactive components, expressed as functions of the voltage angles Θ and magnitudes V_m and generator injections P_g and Q_g, where the load injections are assumed constant and given:

$$g_P(\Theta, V_m, P_g) = P_{bus}(\Theta, V_m) + P_d C_g P_g = 0 \tag{19}$$

$$g_Q(\Theta, V_m, Q_g) = Q_{bus}(\Theta, V_m) + Q_d C_g Q_g = 0 \tag{20}$$

For the AC power flow problem, the function g(x) is formed by taking the left-hand side of the real power balance equations for all non-slack buses and the reactive power balance equations for all PQ buses and plugging in the reference angle, the loads and the known generator injections and voltage magnitudes:

$$g(x) = \begin{bmatrix} g_P^{\{i\}}(\Theta, V_m, P_g) \\ g_Q^{\{j\}}(\Theta, V_m, Q_g) \end{bmatrix} \quad \begin{matrix} \forall i \in I_{PV} \cup I_{PQ} \\ \forall j \in I_{PQ} \end{matrix} \tag{21}$$

The vector x consists of the remaining unknown voltage quantities, namely the voltage angles at all non-reference buses and the voltage magnitudes at PQ buses:

$$x = \begin{bmatrix} \theta_{\{i\}} \\ {}^{\{j\}} \\ v_m \end{bmatrix} \quad \begin{matrix} \forall i \notin I_{ref} \\ \\ \forall j \in I_{PQ}. \end{matrix} \tag{22}$$

This yields a system of nonlinear equations with $n_{pv} + 2n_{pq}$ equations and unknowns, where n_{pv} and n_{pq} are the number of PV and PQ buses, respectively. After solving for x, the remaining real power balance equation can be used to compute the generator real power injection at the slack bus. Similarly, the remaining $n_{pv} + 1$ reactive power balance equations yield the generator reactive power injections.

[PYPOWER] includes four different algorithms for solving the AC power flow problem. The default solver is used for this research and is based on a standard Newtons method using a polar form and a full Jacobian update at each iteration. Each Newton step involves computing the mismatch $g(x)$, forming the Jacobian based on the sensitivities of these mismatches to changes in x and solving for an updated value of x by factorizing this Jacobian. This method is very standard described in detail in many textbooks.

4.2.8. Standard AC OPF

The standard optimization formula for the OPF takes the form:

$$
\begin{aligned}
\min_{X} \quad & f(x) \\
\text{subject to} \quad & g(x) = 0 \\
& h(x) \leq 0 \\
& x_{min} \leq x \leq x_{max}
\end{aligned}
$$

The optimization vector x for the standard AC OPF problem consists of the $n_b \times 1$ vectors of voltage angles Θ and magnitudes V_m and the $n_g \times 1$ vectors of generator real and reactive power injections P_g and Q_g.

$$
x = \begin{bmatrix} \Theta \\ V_m \\ P_g \\ Q_g \end{bmatrix} \tag{23}
$$

The objective function is simply a summation of individual polynomial cost functions f_P^i and f_Q^i of real and reactive power injections, respectively, for each generator:

$$
\min_{\Theta, V_m, P_g, Q_g} \sum_{i=1}^{n_g} f_P^i(p_g^i) + f_Q^i(q_g^i) \tag{24}
$$

In the traditional formulation of the AC power flow problem, the power balance equation in is split into its real and reactive components, expressed as functions of the voltage angles Θ and magnitudes V_m and generator injections P_g and Q_g, where the load injections are assumed constant and given:

$$
g_P(\Theta, V_m, P_g) = P_{bus}(\Theta, V_m) + P_d - C_g P_g = 0 \tag{25}
$$

$$
g_Q(\Theta, V_m, Q_g) = Q_{bus}(\Theta, V_m) + Q_d - C_g Q_g = 0 \tag{26}
$$

The equality constraints are simply the full set of $2 \cdot n_b$ nonlinear real and reactive power balance equations above. The inequality constraints consist of two sets of n_l branch flow limits as nonlinear functions of the bus voltage angles and magnitudes, one for the from end and one for the to end of each branch:

$$h_f(\Theta, V_m) = |F_f(\Theta, V_m)| - F_{max} \leq 0 \tag{27}$$

$$h_t(\Theta, V_m) = |F_t(\Theta, V_m)| - F_{max} \leq 0 \tag{28}$$

The flows are typically apparent power flows expressed in MVA, but can be real power or current flows, yielding the following three possible forms for the flow constraints:

$$F_f(\Theta, V_m) = \begin{cases} S_f(\Theta, V_m), & \text{apparent power} \\ P_f(\Theta, V_m), & \text{real power} \\ I_f(\Theta, V_m), & \text{current} \end{cases} \tag{29}$$

where $P_f \leq \{S_f\}$ and the vector of flow limits F_{max} has the appropriate units for the type of constraint. It is likewise for $F_t(\Theta, V_m)$. The variable limits include an equality constraint on any reference bus angle and upper and lower limits on all bus voltage magnitudes and real and reactive generator injections:

$$\begin{aligned} &\theta_i^{ref} \leq \theta_i \leq \theta_i^{ref}, && i \in I_{ref} \\ &v_m^{i,min} \leq v_m^i \leq v_m^{i,max}, && i = 1...n_b \\ &p_g^{i,min} \leq p_g^i \leq p_g^{i,max}, && i = 1...n_g \\ &q_g^{i,min} \leq q_g^i \leq q_g^{i,max}, && i = 1...n_g. \end{aligned}$$

4.2.9. Adapting PYPOWER for N-K Contingency Planning

Because the PYPOWER extension of the MATPOWER program is not set up to be used as a risk analysis tool, some modifications to the program are necessary in order to evaluate N-K failure scenarios. First of all, in order to simulate "failures" certain components must be "switched off" or "removed" from the analysis. In the case of power generation facilities, this is a simple matter of designating that the plant is not operational in the OPF function by simply changing a 0 to a 1 in the code. The same method can be used to switch off power transmission lines. Because this functionality was not built into the program for the buses in the system, a work-around was used wherein all of the lines solely terminating at a specific bus are switched to “off” to simulate the effects of that bus becoming non-operational. This mimics the real world implications of a blown transformer (or other bus malfunction) as it would render all of its dependent lines unable to deliver power in that direction, forcing the OPF solution to a different route.

4.3. Solution Methodologies; Combinatorial Optimization

4.3.1. The CE Method

Because of the rare and combinatorial nature of each of the various N-K scenarios, the cross-entropy (CE) sampling method is ideal for determining the highest-cost failure scenarios. The CE method has the advantage of using a simple adaptive procedure for estimating the optimal reference parameters. In this exercise, the updating parameters are set so that the CE method becomes a greedy sampling methodology. This means that scenarios that have a higher initial weights (based on relative operating cost of each failure) are more likely to be sampled. This method reduces the variance of the final sample by targeting the more likely scenarios from the perspective of an intelligent attacker.

5. Results

The framework for setting up the different algorithms for estimating the pmf of $S(u)$ is as follows:

(1) Run existing optimal power flow optimization model using a sample of N-K scenarios
(2) Update the relative weights of the most damaging/costly scenarios using the results of the OPF model on the initial sample to develop a parameter for updating subsequent samples
(3) Resample using the updated parameters from Step 2

The relevant variable in developing an updating parameter for combinatorial optimization is purely operating cost in this exercise. In other words, it is assumed that the likelihood of a specific scenario occurring is tied to the cost of delivering power should that failure occur. Because of the variable cost of running different power plants and the losses from routing power to the delivery nodes, different OPF solutions have different operating costs. Obviously the favored solution sticks to the lowest variable cost power generating facilities and the lowest loss route through the system. However, an intelligent attacker will attempt to discern multiple component failures that will force the power delivery to be routed through high variable cost facilities and inefficient power lines and buses.

While the ultimate goal for a malicious party may be a full system failure, our analysis is restricted to examples where power delivery is able to be maintained. It would be fairly simple to update the framework to include a proxy for operating cost associated with lost power delivery based on estimated value of service to the power operator. In general this should be assumed to be much more costly than scenarios where power flow is able to be maintained in the face of N-K failures. Should this algorithm be adapted to smaller systems, it will be necessary to develop a value of service estimation because the minimal cut solution will be reached fairly quickly and will limit the ability to evaluate the system if this is not considered.

The determination of an informed sample of N-K scenarios requires anticipating the most damaging set of component failures, either from malicious interference or simple malfunctions. From the perspective of a utility planner, the most damaging scenarios are going to involve the smallest number of failed components possible (in this case: K¡4) because it is generally unlikely that a large number of components will independently fail simultaneously. Similarly, a logical attacker of the system would be unlikely to attack more components than is necessary to cause a large amount of damage, so smaller values of K are deemed to be more likely than larger values. Therefore, in most scenarios it would be beneficial to start with sampling and running N-2 scenarios, and iteratively increasing K.

In the following examples, the default value for K is 2. While $k \geq 2$ scenarios are considered, they are only tested on the model that produced the lowest variance in results for N-2 testing.

5.1. A Note on Weights

As different scenarios are run, it is likely to become apparent that certain nodes (and combinations thereof) within the system hold a higher damage potential. Using this information, it would be possible to assume that these components and combinations of components are more likely to show up in other N-K scenarios, informing the weights of component failures in $K + n$ testing.

From the perspective of the researcher, the weights for individual component failures can be estimated based on how essential those nodes are in lower operating costs OPF solutions and/or their likelihood of being present in high operating cost OPF solutions. Without the system knowledge of an operator, it is difficult to determine the exact relationship between individual component weights and N-K scenario weights. In order to inform the sampling process for initial N-K scenarios in each of the following methods, the relationship between individual component failure and its probability of failure in conjunction of components must be estimated based on the task at hand. This idea will be expanded on in the following sections.

5.2. Simple Random Sampling

The first method tested features no assumed initial weights and no updating the weights of N-2 scenarios based on the results from previous tests (i.e. simple random sampling). This method is outlined as follows:

(1) Run existing optimal power flow optimization model using a random (uninformed) sample of N-K scenarios

By design, this method assumes equal weighting of all scenarios within U $(\frac{1}{\#u})$. As expected this method produces results that vary highly depending on how the randomization of the algorithm is set initially. This variance value is assumed to be the "baseline" value on which subsequent algorithms are trying to improve. Obviously this can hardly be deemed to be a "model" at all, but it serves the purpose of illustrating how difficult the task at hand is without any improvements to the sampling method. Because of the large number of possible combinations at hand, any sampling strategy must have some sort of "intelligent" quality in order to produce results with any kind of robustness.

The various improvements to the sampling method are outlined in the following sections.

5.3. Cross Entropy Analysis

For these purposes a general CE method is a simple improvement over simple random sampling because it uses parameters to update the initial weights with each iteration of the algorithm. This method has been adapted to suit N-K analysis as follows:

(1) Generate a random data sample (trajectories, vectors, etc.) from U;
(2) Run sample points through OPF model;
(3) Update the parameters of the random sample based on the OPF results to produce a better sample in the next iteration;
(4) Repeat steps 1-3 until designated stopping point

The first issue when running this method is how to best set the parameters. Based on the previous discussion, these parameters are informed by the operating costs resulting from running the various N-K sample points through the OPF model. A whole paper could easily be written solely on how best to set the parameterization. The goals for setting the parameter in this case are so that the sampling strategy will have a "greedy" quality. This means that the higher cost OPF solutions are of much more interest to the researcher, so any sampling method should favor those results. This is because the goal is to identify the most damaging scenarios, so while the relative probability of low cost N-K scenarios are interesting, they are not the primary focus of this research.

The parameters can then be set from the perspective of the grid manager or security expert. Because there are 10,296 N-2 scenarios in the test grid, it can be assumed that the operator could reasonably evaluate only some small subset of those scenarios. To make it simple, it is assumed that this number is 100 scenarios for N-2. This means that the top 5% of the results are the most relevant to the decision maker. Therefore the parameters are set so that with each iteration of the sampling process, the top 5% costly OPF solutions are updated to reflect 90% of the total density function while the remaining scenarios are updated to reflect the remaining 10%. Because there tends to be little variation in the final OPF solutions of the top 5% of scenarios, this weight is distributed evenly among the set of N-K scenarios. However, this may not be true for all grids, so parameters should be tested and evaluated carefully before deciding on a final methodology.

5.4. CE Method with Informed Initial Weights from Optimal Power Flow

In order to come up with a set of weights to inform initial sampling efforts, it is necessary to evaluate the system configuration. In the actual application of this process, experienced utility planners will be able to use their knowledge of the system to identify specific components or combinations of components that are particularly necessary to the operation of the system. This existing knowledge of the system can be used to update the initial weights so that all of the initial N-K scenarios are not equally likely.

In the case where there is no insider information on system operation, deductions about the initial weights of certain failures can be determined from the results of the Optimal Power Flow (OPF) model. In this case, components that are operating at or near capacity under base conditions (no failures) are deemed to be particularly vulnerable to attack or failure, and the initial weights can be adjusted accordingly. This method is outlined as follows:

(1) Generate a weighted data sample using results from OPF results on N-0 and N-1 scenarios from U;
(2) Run sample points through OPF model;
(3) Update the parameters of the random sample based on the OPF results to produce a better sample in the next iteration;
(4) Repeat steps 1-3 until designated stopping point

Because the individual component initial weights from step 1 cannot be assumed to have an independent and mutually exclusive relationship to one another when assessing the relative weights of N-2 scenarios, another method must be used to develop initial weights for the N-2 scenarios from this information. In this case, the update is fairly simple - because the number of components that are deemed essential to the lowest cost OPF solutions are about 5% of the total number of components, the N-2 scenarios in which they are present can be updated similarly to the parameterization process explained in the previous section. Then CE sampling can be performed using the same method explained in the previous section.

5.5. Importance Sampling with Informed Initial Weights from Sub-Optimal Power Flow

This method attempts to take advantage of the fact that an informed attacker is going to be thinking in terms of maximizing operating cost in a given OPF model instead of minimizing it like an operator would. So instead of developing a set of initial weights based on OPF minimum cost models for N-0 and N-1 scenarios, this method develops the set of initial weights using OPF maximum cost models on those two scenarios. While this provides a marked improvement over the minimum cost model, it is not initially clear why this should be the case. Essentially this process is just approaching the updating from the other side, creating a mirror effect. However, since the algorithm is based on a "greedy" sampling strategy that favors high cost scenarios, using a set of informed priors based on the maximum operating cost solutions guides the search to these scenarios much more effectively than the inverse case. The steps for this methods are outlined as follows:

(1) Generate a weighted data sample using results from OPF maximum cost results on N-0 and N-1 scenarios from U;
(2) Run sample points through OPF model;
(3) Update the parameters of the random sample based on the OPF results to produce a better sample in the next iteration;
(4) Repeat steps 1-3 until designated stopping point

This method can be further improved upon by finding the next costliest scenarios and so on until a designated stopping point. Using this approach, it becomes clear quite quickly which components are most likely to remain untouched by an intelligent attacker (high variable cost generators and low efficiency transmission lines).

5.6. Summary of Results

While the output from each of the individual iterations of the model is interesting and informative, the ultimate goal is to produce a methodology for flexibly evaluating electrical grid risk. Because of the variable nature of using a sampling strategy to evaluate the cost of N-K scenarios, results from different runs can vary slightly, but ultimately provide important insight into the security risks of the system. Table 1 illustrates the results of running a simpler version of the model on a small system (Figure 4); orange areas show severely costly N-2 combinations while yellow areas show moderately costly combinations. Bus is abbreviated to B, Generator to G, and Line to L. The top row lists the first failing component, while the first column lists the second failing component. Each cell then represents the combination of those two components (because one component cannot along with itself, only half of the cells are used).

	B1	B2	B3	B4	B5	B6	G1	G2	G3	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
B1	0	0.044	0.044	0.015	0.015	0.015	0.015	0.015	0.015	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
B2	0	0	0.044	0.015	0.015	0.015	0.015	0.015	0.015	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
B3	0	0	0	0.015	0.015	0.015	0.015	0.015	0.015	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
B4	0	0	0	0	0.044	0.044	0.015	0.015	0.015	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
B5	0	0	0	0	0	0.044	0.015	0.015	0.015	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
B6	0	0	0	0	0	0	0.015	0.015	0.015	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
G1	0	0	0	0	0	0	0	0.044	0.044	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
G2	0	0	0	0	0	0	0	0	0.044	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
G3	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
L1	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
L2	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
L3	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
L4	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001
L5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001
L6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001	0.001
L7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001	0.001
L8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001	0.001
L9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.001
L10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.001
L11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 1. Results of Model on 6-Bus System

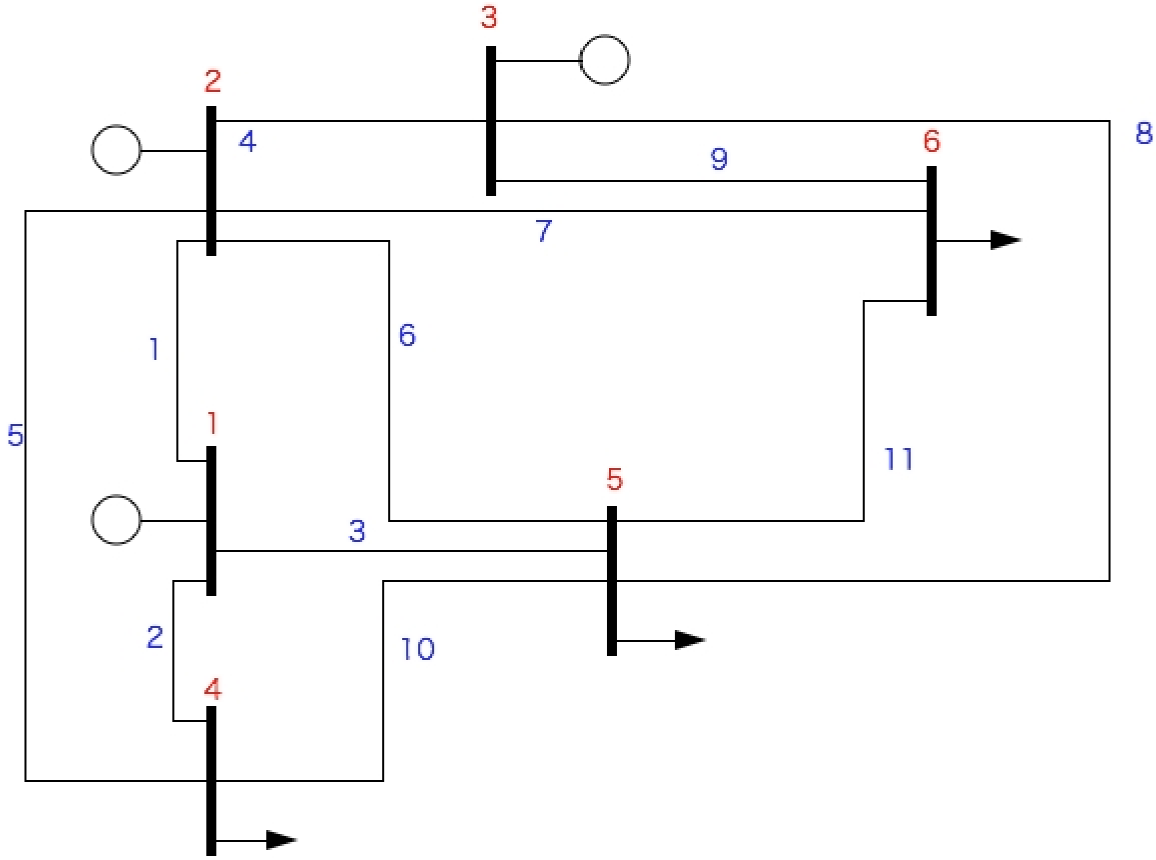

FIGURE 4. Six-Bus System with 11 Transmission Lines and Three Generators [6]

The purpose of presenting the results of the model on a smaller system is because it allows us to present the failures and associated risk of failures in a viewable format. Each of the various cells represents a potential N-2 failure, and each of the numbers in the cells represents the weight associated with the risk and cost of the given failure. Once systems and N-K failures get larger than the illustrated example, visualizing the results becomes a problem all on its own.

While the results of the 57-bus example are more difficult to present in a friendly visual format, it is possible to draw many conclusions about the model from the results. First of all, while the ranking of most vulnerable N-K combinations may vary, the nodes of the system that are included in the list of most vulnerable N-K combinations is fairly consistent. This means that at least in the 57-bus system, particularly valuable components tend to show up repeatedly in the list of most damaging N-K scenarios. Specifically the nodes containing generators or output buses are most likely to show up in the highest 25% of damaging scenarios. This is extremely useful for decision-makers because it will allow them to identify points of the system that are especially crucial for continuous operation.

The scale of the failures is very important to the results of the model. As mentioned previously, it is important to quantify how valuable it is to maintain complete service at each of the demand nodes. The total generation capacity of the system is 1975.9 MW per hour, with a typical hourly generation of 1270 MW. Under optimal conditions generation costs are about $42,000 per hour, and only about 1.3% of the electricity generated is lost during transmission (about $540 per hour). However, because optimal power flow favors the cheapest per unit generator, if prices change or that generator goes out of service, the prices can increase by nearly 50%. Similarly, if the most efficient power lines go out of service, transmission loss can swell to nearly 10% of total production.

There are several options for presenting the final results of this type of risk analysis to a decision maker. The most intuitive approach would be an interactive map of the grid that highlights specific components that are show up repeatedly in highly costly N-K scenarios. While this approach is extremely useful when looking for single components in which to invest security dollars, it does not highlight combinations of component failures that might be especially problematic. See Figure 5 for an example of this output; red highlighted areas show highly vulnerable components while yellow areas show moderately vulnerable components.

Ideally this approach would allow the user to select a component and view the list of damaging N-K scenarios it is a part of along with the associated weights for each of those scenarios. Using generator 1 as an example, the list would look something like: Generator 1 with Generator 9 - .016%; Generator 1 with Bus 22 0.011%, etc. In a computing environment with large amounts of processing power, this list would include total associated cost in order to further inform the decision maker.

The alternative to a visual representation of the grids most costly component failures would be a simple list of the most damaging N-K scenarios under this methodology. While

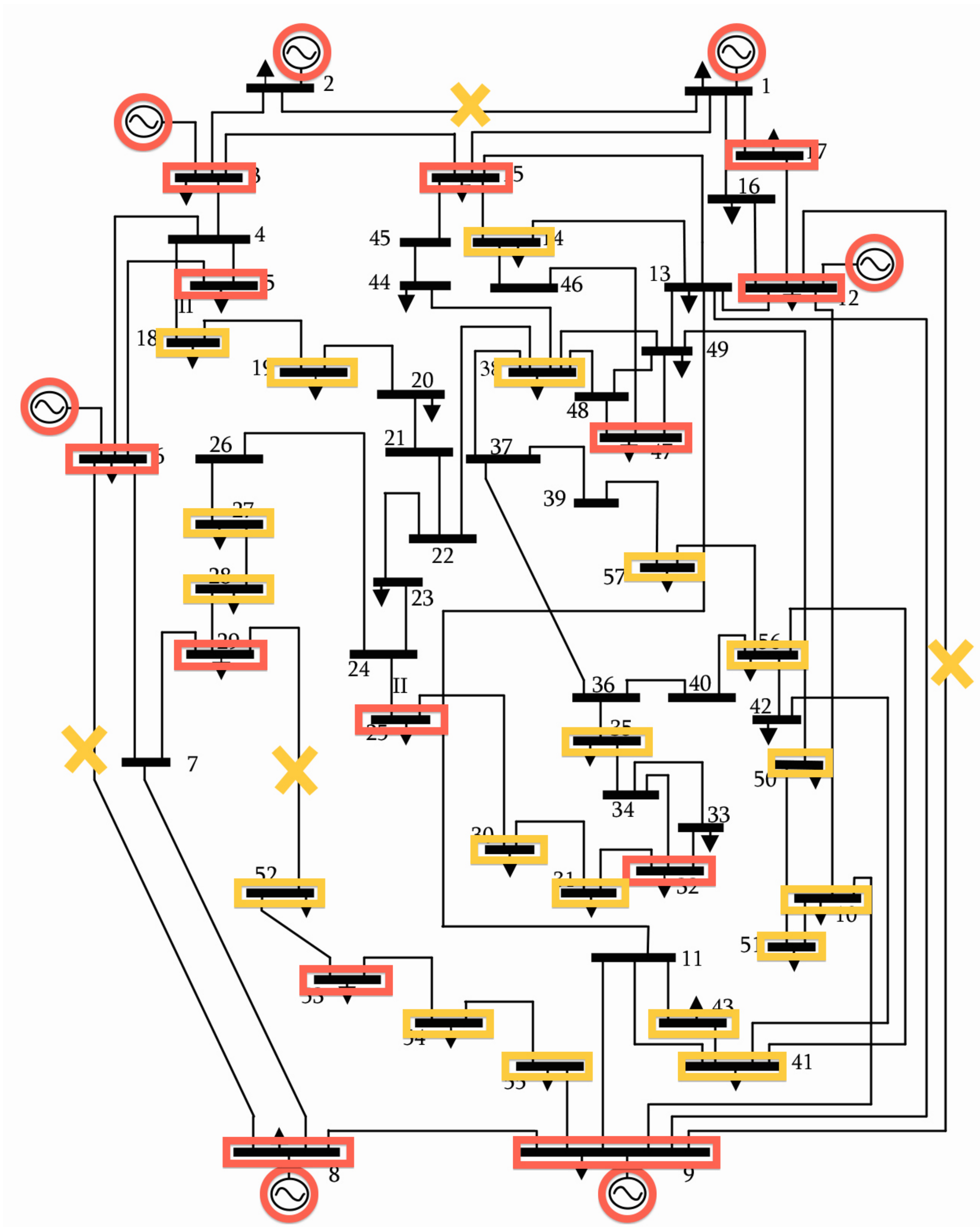

FIGURE 5. IEEE 57-Bus Test Grid, December 1961 [7]

this is the most exhaustive way of presenting the data, it gives less actionable information to the user, especially if they are not an expert in the system.

6. Validation Methods and Discussion of Potential Future Work

6.1. Validation Methods

While it is difficult to propose an independent validation method for this solution when employed on larger systems, it is possible to test the sampling method on a smaller system and compare it to the same process run with the full number of N-K scenarios. However, there are limitations to this type of test. Because of the inherent differences in system configuration, a test on a smaller network may provide misleading results. This can be due to the fact that the costs of failures on smaller systems tend to be clustered around a small set of points, rather than distributed evenly. The ideal way to test this method is to use a system with a large amount of computing power to run the OPF model on an entire set of N-K scenarios to compare the results with the CE sampling method.

In order to compare the efficacy of different combinatorial optimization algorithms on the N-K problem, the relevant variable is the variance of the results. Because of the combinatorial explosion associated with the n-dimensional experiment space U, enumeration methods are not feasible without extreme computing power. Therefore all algorithms must employ some sort of importance sampling method. In order to rank the results of these methods, it is useful to compare them to an example where there is no update based on parameterization (i.e. simple random sampling).

In order to test the effectiveness of using informed initial weights, it is possible to run a simulation where all initial weights are equal in the set of N-K scenarios and compare to the simulation where initial weights are informed by component capacity and the results of previous N-K scenarios. The ultimate goal of the model is to identify most damaging N-K scenarios, so a simple comparison of the relative costs of the most likely scenarios from each model can inform the effectiveness of using this information. In summary, even on relatively smaller samples, it is apparent that using initial weights more reliably identifies higher cost scenarios than a simple random method.

6.2. Potential Future Work

With increased automation and additional computing power, it will be possible to further test this methodology to ascertain additional improvement to the model and to develop a marketable risk assessment tool for utility operators and outside parties. The ability to customize the initial weights and system configuration based on the users knowledge makes this method extremely flexible and adaptable to almost any power system, whether a micro-grid or municipal utility. With legislation on the way to bolster risk assessment requirements for physical system security, this type of work is especially relevant.

Additionally, users have been instituting similar methods along with machine learning models to ascertain the structure of different papers using probabilistic topic models to analyze text. Topic models are algorithms used to discover the major themes within an unstructured group of documents and organize the collection by these uncovered themes. The need for an algorithm to do this analysis comes from the fact that as more texts become available online, there is not enough human power to analyze each document individually. The different models used include Latent Dirichlet Allocation (LDA), which assumes a topic to be a distribution over a fixed vocabulary. This method assumes that the hidden structure of the document can be approximated with a posterior probability distribution, which can be determined from the observed variables the words in the given document.

While on the surface, this topic has very little to do with analyzing risk in electrical grids, they share the similar problem of having too much data to analyze using traditional methods. While text analysis uses assumptions about the frequency and combination of certain key words, N-K risk analysis uses the frequency and combination of different types of grid components (generators, lines, substations, etc.). Furthermore, both problems are dealing with a hidden structure (document topic vs. cost of power failure/type of attack), which may be derived from observable components (words or parts of the grid). Other fields have taken advantage of the existing text-based algorithms by adapting them to their own area of concern. For example, the LDA model has been used by population geneticists to determine ancestral heritage within a sample of individuals, even when dealing with incomplete genetic data.

A potential methodology for adapting an LDA model to be used to determine risk within an electrical grid is to identify a set of topics or categories of failure or attack scenarios. Just like certain words are more likely to show up in texts of a certain topic structure, certain components may be more likely to show up in different categories of N-K failures. For example, a failure caused by natural elements may be more likely to be comprised of a set of power lines that have been taken down than any other combination of components.

On the other hand, a premeditated cyber-physical attack may be more likely to be made up of a variety of different components chosen to maximize damage to the grid. This would allow for additional input from grid operators and other insiders with knowledge into the system to help adjust the initial weights of different scenarios based on their experience. Additional adjustments to the initial weights of N-3 and greater scenarios can be determined from inferences derived from N-1 and N-2 scenarios. The additional information about the prior distribution can inform the structure of the posterior distribution and cut down significantly on the computing power required.

Another benefit of this type of method is that by automatically categorizing the different scenarios into topics, the program is already making it easier to present the data to the user. Because the underlying goal is to find hidden connections between the different variables and topics, the process of aggregating the findings into a coherent user interface is much more simple. This type of approach promises to be a more elegant method of probabilistic modeling, both at the analysis and presentation levels.

References

[1] N.J. Gordon, D.J. Salmond, A.F.M. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation” IEE proceedings-F, Vol. 140, No. 2, April 1993

[2] R. Chen, A. Cohn, N. Fan, and A. Pinar, “N-k-ϵ Survivable Power System Design to appear in 12th International Conference on Probabilistic Methods Applied to Power Systems”, (PMAPS12).

[3] J. Yan, Y. Zhu, H. He, and Y. Sun, “Multi-Contingency Cascading Analysis of Smart Grid Based on Self-Organizing Map,” IEEE Trans. Information Forensics and Security, vol. 8, no. 4, pp. 646-656, 2013.

[4] B. Xie, R. Passonneau, H. Dutta, J. Miaw, A. Radeva, A. Tomar, and C. Rudin. “Progressive Clustering with Learned Seeds: An Event Categorization System for Power Grid Proceedings of the 24th International Conference on Software Engineering & Knowledge Engineering (SEKE)”, pp. 100-105, 2012.

[5] Pearl, J. “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”, San Mateo, CA, USA: Morgan Kaufmann Publishers, (1988).

[6] Wood and Wollenberg, Power System Operation and Control 2nd Ed.

[7] I. Dabbagchi, R. Christie, AEP and University of Washington, IEEE Common Data Format, 1993.

[8] Shumard, R. “Utilities Look to Get Started with NERC CIP-014-1 Physical Security Standard”, http://www.energybiz.com/article/15/02/utilities-look-get-started-nerc-cip-014-1-physical-security-standard, 2015.

[9] NERC, “Physical Security Standard Implementation”, http://www.nerc.com/pa/CI/Pages/Physical-Security-Standard-Implementation.aspx.

[10] D. Ernst, M. Glavic, G.-B. Stan, S. Mannor, L. Wehenkel, and Gif sur Yvette Supelec. “The cross-entropy method for power system combinatorial optimization problems”. In Power Tech, 2007 IEEE Lausanne, pages 1290 1295, Lausanne, 2007, http://www.montefiore.ulg.ac.be/services/stochastic/pubs/2007/EGSMW07/EGSMW07.pdf.

[11] P. De Boer, D. Kroese, S. Mannor, R. Rubinstein, “A Tutorial on the Cross-Entropy Method”, Annals of Operations Research 134, 1967, 2005, http://www.maths.uq.edu.au/ kroese/ps/CEopt.pdf.

[12] Z. Botev, D. Kroese, R. Rubinstein, P. l’Ecuyer, “The Cross-Entropy Method for Optimization”, http://www.maths.uq.edu.au/ kroese/ps/aortut.pdf.

[13] R. Zimmerman, C. Murillo-Sanchez, Matpower 5.1 User's Manual, 2015, http://www.pserc.cornell.edu/matpower/manual.pdf.

[14] R. P. Kothari and D. P. Kroese. Optimal generation expansion planning via the cross-entropy method. In M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, editors, Proceedings of the 2009 Winter Simulation Conference., pages 14821491, 2009.

[15] A. Sani. The Spread of HIV/AIDS in Mobile Populations: Modelling, Analysis, Simulation. LAP LAMBERT Academic Publishing, Berlin, 2010.

[16] A. Sani and D. P. Kroese. Controlling the number of HIV infectives in a mobile population. Mathematical Biosciences, 213(2):103112, 2008.

[17] D. P. Kroese, S. Porotsky, and R. Y. Rubinstein. The cross-entropy method for continuous multi-extremal optimization. Methodology and Computing in Applied Probability, 8(3):383407, 2006.

[18] Z. Szabó, B. Póczos, and A. Lörincz. Cross-entropy optimization for independent process analysis. In Independent Component Analysis and Blind Signal Separation, volume 3889, pages 909916. Springer-Verlag, Heidelberg, 2006.

[19] Z. I. Botev and D. P. Kroese. Global likelihood optimization via the cross-entropy method with an application to mixture models. In Proceedings of the 36th Winter Simulation Conference, pages 529535, Washington, D.C., 2004.

[20] J. Keith and D. P. Kroese. Sequence alignment by rare event simulation. In Proceedings of the 2002 Winter Simulation Conference, pages 320327, San Diego, 2002.

[21] I. Cohen, B. Golany, and A. Shtub. Resource allocation in stochastic, finite capacity, multi-project systems through the cross entropy methodology. Journal of Scheduling, 10(1):181193, 2007.

[22] D. P. Kroese, K.-P. Hui, and S. Nariai. Network reliability optimization via the cross-entropy method. IEEE Transactions on Reliability, 56(2):275287, 2007a

[23] K. Chepuri and T. Homem-de-Mello. Solving the vehicle routing problem with stochastic demands using the cross entropy method. Annals of Operations Research, 134(1):153181, 2005.

[24] Y. Wu and C. Fyfe. Topology perserving mappings using cross entropy adaptation. In 7th WSEAS international conference on artificial intelligence, knowledge engineering and data bases, Cambridge, 2008.

[25] S.A. Kazarlis, A.G. Bakirtzis, and V. Petridis. A genetic algorithm solution to the unit commitment problem. IEEE Transactions on Power Systems, 11(1):8392, February 1996.

[26] F.J. Marin, F. Garcia-Lagos, G. Joya, and F. Sandoval. Genetic algorithms for optimal placement of phasor measurement units in electrical networks. Electronics Letters, 39(19):14031405, September 2003.

[27] C. Moors, D. Lefebvre, and T. Van Cutsem. Combinatorial optimization approaches to the design of load shedding schemes against voltage instability. In Proceedings of the 32nd North American Power Symposium (NAPS), pages 10141021, 2000.

[28] [16] R. Romero, R.A. Gallego, and A. Monticelli. Transmission system expansion planning by simulated annealing. IEEE Transactions on Power Systems, 11(1):364369, February 1996.

[29] Kosut, O., "Max-flow min-cut for power system security index computation," Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th , vol., no., pp.61,64, 22-25 June 2014

[30] B. Lesieutre, A. Pinar, S. Roy, "Power System Extreme Event Detection: The Vulnerability Frontier", 2006.

[31] B. Lesieutre, S. Roy, V. Donde, A. Pinar, "Power System Extreme Event Screening using Graph Partitioning", 2008.