CONF-970456--1
ANL/DIS/CP--92368

SIMPLIFYING GENERATOR SITING AND CONTRACT OPTIONS

James A. Kavicky
Energy Systems Engineer
Argonne National Laboratory
Argonne, IL

and

S.M. Shahidehpour
Dean of the Graduate College
Illinois Institute of Technology
Chicago, IL

ABSTRACT

This paper demonstrates the application of a clustering method designed to aggregate the effects of individual generation shift factors, while preserving acceptable parallel path characteristics. The paper illustrates how several statistical metrics and modeling heuristics are employed to determine criteria for clustering acceptance. Tradeoffs between cluster precision and effective cluster aggregation are illustrated using a four-area system. These tradeoffs are shown to affect final cluster acceptability. The paper applies the technique to generator siting and contract option applications to reduce transmission modeling complexity and suggests it can be used as a screening method to identify the need for detailed system studies.

1. INTRODUCTION

The historical aspects of parallel path influences on interconnected system behavior have been previously reviewed in [1]. Further operational complications can be expected as the number of interutility transactions increases in a deregulated environment, thereby compounding the impacts of parallel path effects. Current methods for modeling parallel flows rely heavily on power flow simulations to represent their impacts on network behavior and transaction simulation. Since individual generators and delivery points impact interutility tie lines in different ways, the dependence between generation dispatch and interutility tie line flows is easily observed through the use of generator shift factors (GSFs).

The approach introduced in [2] demonstrates the application of a clustering method to aggregate the effects of individual GSFs, while preserving acceptable parallel path characteristics. The paper applies the technique to generator siting [3] and contract option applications to reduce transmission modeling complexity and suggests it can be used as a screening method to identify the need for detailed system studies. This paper further illustrates the method by examining its performance on a 31-bus, four-area system. Several statistical metrics [4] and modeling heuristics are employed to determine criteria for clustering acceptance. Tradeoffs between cluster precision and effective cluster aggregation are illustrated, as these issues affect final cluster acceptability.

The process is easily summarized as follows. GSFs are determined for all interutility tie lines that interconnect adjacent systems assumed to be utilities. By observing interutility tie line sensitivities associated with various cases, buses having similar GSF responses are grouped together. The similarities of various GSFs provide the criteria used to cluster the buses. The end result is a reduced number of representative cases (or subareas) based on the GSF performance derived from the typical system topology used in the study. The method used in this paper is aimed at avoiding ad hoc methods of determining representative areas by using methods to naturally cluster system buses into areas that electrically respond to perturbations in a very similar manner (within some tolerable error margin).

2. METHOD DESCRIPTION

The following steps sequentially describe the adopted clustering methodology.

1. Determine base-case topology for system. Figure 1 shows a view of a 31-bus, four-area system, and Figure 2 shows the corresponding system one-line diagram.

2. Determine GSF pairs. Figure 3 shows the system after all radial lines have been removed by the Distribution Factor (DFAX) program. These remaining 23 buses represent the set of all possible buses available for generator shift pairs. The 10 interutility tie lines shown in Figure 1 represent the monitored line set used for all GSFs in this study. The total sample space defined by this topology is: $23*(23-1) = 506$ cases, where each case represents 10 variables.

3. Format cluster algorithm input data. DFAX output reports are parsed to generate a formatted input file suitable for the clustering program. Input data is sorted and modified to avoid including bidirectional GSFs (e.g., Bus 23 to Bus 1 and Bus 1 to Bus 23), since the corresponding

DISCLAIMER

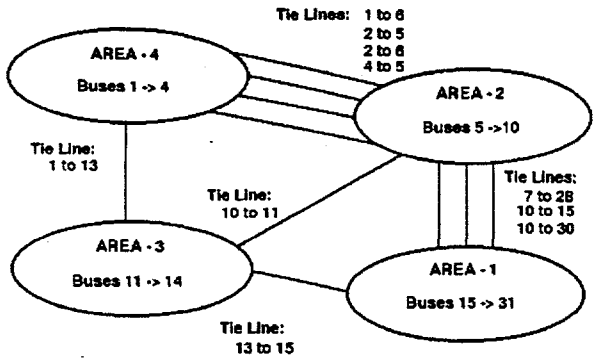

Figure 1. Four-Area System

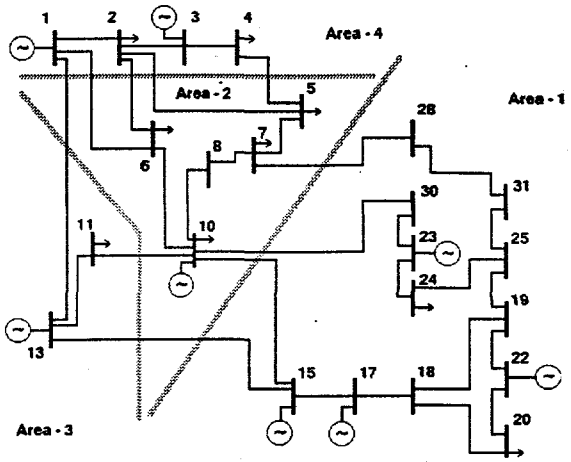

Figure 3. One-Line Diagram with Radial Lines Removed

GSFs are identical except for a sign change. In this study, 506 cases are reduced to 506/2 = 253 cases.

4. Run cluster algorithm. The cluster algorithm is Euclidean-based, so an initial Euclidean distance, called epsilon, is chosen to guide clustering tightness. A second input parameter is also provided to regulate transaction participation. Both parameters are discussed more in the next section.

5. Briefly review statistical metrics. If various statistics generally satisfy modeling requirements, proceed to Step 6. Otherwise, repeat Step 4 using a different maximum Euclidean tolerance. For example, if epsilon is large, fewer subareas are formed (i.e., less clusters having more cases in each cluster). If epsilon is small, more subareas are defined to keep the closeness among cases extremely tight.

6. Parse results to construct subarea visualization. Generate a view of partitioned subareas within areas according to cluster results. This two-dimensional view summarizes the aggregated multi-dimensional GSF cases.

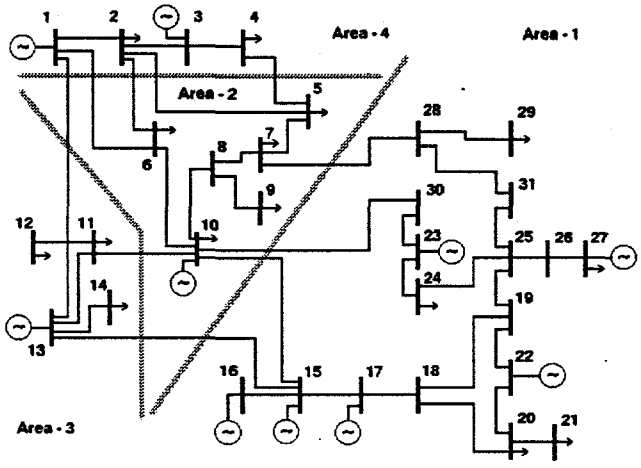

Figure 2. System One-Line Diagram

7. Analyze and evaluate results. Compare statistical results and subarea partitions. Evaluate alternatives between numerical accuracy and aggregation level. Repeat Step 4, if necessary, using an appropriate epsilon variant.

8. Re-evaluate clusters if necessary. Once a final level of aggregation is achieved, some clusters may need to be re-evaluated to obtain improved statistical information.

9. Apply clustered results. Subject to the intended modeling requirements, the cluster centroid or a representative case nearest to the cluster centroid can be used to characterize all cases within each validated cluster.

3. CLUSTER ACCEPTANCE CRITERIA

Several factors influence cluster acceptance. The methodology outlined in this paper makes use of both statistical and heuristic methods. Clustering applications, by definition, are regarded as heuristic in nature. The approach presented here also permits several heuristics that are particular to the application of distribution factors. The following paragraphs provide some definitions and describe the acceptance criteria available when using the proposed methodology.

3.1 Euclidean Distance — Epsilon

The most significant clustering parameter is the threshold that determines the maximum tolerable pair-wise Euclidean distance. The combination of two GSFs defines a multi-dimensional vector represented as a single Euclidean distance. Resulting distances below the specified value of epsilon can be used to further expand the number of cases in a cluster, if all of the elements within the cluster are mutually inclusive. As a result, larger values of epsilon tend to produce fewer clusters, while smaller values of epsilon tend to produce more clusters. Epsilon is the most critical metric governing cluster outcome, because it regulates the level of compromise among clustered cases.

3.2 Cluster Mean or Cluster Centroid

The cluster centroid, or mean, is a single vector used to characterize the GSFs resulting from the various cases represented in the cluster. A proper cluster centroid preserves the interarea energy balance found in the represented GSFs. Assurance of energy balance is provided in the method.

3.3 Minimum and Maximum Values and Skewness

Without exception, the clustering process invites compromise. The minimum and maximum values that deviate from the mean are stored for each interutility tie line comprising the clustered cases. These values are most likely noncoincident among all clustered cases. However, they can still be used to convey important information regarding centroid positioning relative to the entire cluster. This characteristic is referred to as cluster skewness. The minimum and maximum values are used to construct minimum and maximum Euclidean distances with respect to the cluster mean. A cluster with minimum and maximum vectors having the same value indicates a cluster centroid positioned through the center of the represented cases. Minimal skewness is a good cluster property.

3.4 Standard Deviation and Coefficient of Variation

The standard deviation is determined for each cluster. Small values of standard deviation indicate close agreement among all cases included in the cluster. The coefficient of variation provides an improved acceptance criterion. Large standard deviations that correspond to large means are less significant than large standard deviations that correspond to small means. The coefficient of variation numerically conveys this information. However, cases fulfilling the latter situation can represent a less significant event, as illustrated later in the paper, if certain modeling assumptions are applied.

3.5 Maximum Percent Error

The largest percent error referenced to the mean is recorded for each cluster. Individual minimum and maximum values are used to calculate this metric. Although the coefficient of variation is similar to the percent error, the percent error is very intuitive, because it is a familiar numeric statistic. Larger percent errors imply larger deviations from the mean, which indicates that a cluster can be too large if modeling requirements are strict.

3.6 Transaction Participation Threshold

The second parameter available on the cluster command line is the transaction participation threshold (TPT). Interactions among areas below a threshold value can be regarded as insignificant in some applications. For example, a transfer from Area A to B may only produce a 5% transfer between Areas A and C. The transaction through Area C can be ignored if modeling requirements justify this relaxed condition. The TPT is compared against the cluster means.

Applying a moderate TPT value can greatly impact the cluster maximum percent error, since the algorithm exempts those errors from being considered. In this situation, an asterisk is printed adjacent to the coefficient of variation to indicate that the corresponding line meets the TPT criterion. The next significant maximum error is then recorded for the cluster.

3.7 Normalized Percent Error Frequency Distribution

Derived from the maximum percent error metric, a normalized frequency distribution of percent error is provided as a summary of overall cluster performance. The distribution is expressed as a percentage that indicates the number of clusters within a particular percent error range with respect to the total number of clusters. This summary is used in this study to illustrate the performance of various epsilon and TPT values.

3.8 Violations

3.8.1 Internal Violation. Clusters formed having GSF bus pairs belonging to the same subarea are prohibited.

3.8.2 Multi-Area Violation. In some situations, clusters can be formed having cases belonging to different areas. For example, generators at Buses 2 (in Area 4) and 5 (in Area 3) can simultaneously serve a load at Bus 11 (in Area 3), if the independent cases are clustered together. The resulting generation is spilt between generators (say, 60% at Bus 2, and 40% at Bus 5). Although the cluster mean preserves overall energy balance among areas (verified against a load flow simulation), these simultaneous contracts may not be desirable in a particular modeling application. A similar situation can produce split load buses. Consequently, formation of clusters having either of these properties is not permitted in this algorithm. A large percent error and coefficient of variance can indicate a multi-area violation.

3.8.3 Subarea Violation. Similar to the multi-area violation, the subarea violation is detected to prevent cluster formations that violate subarea boundaries. In this situation, interarea energy balances behave properly. However, either generator or load buses are split among multiple subareas within the respective generating or demand areas. A large percent error and coefficient of variance can indicate a subarea violation. The formation of clusters violating subarea boundaries is not permitted, because other firm cluster definitions would be jeopardized.

3.8.4 Direction Violation. If a particular mean of a cluster has an opposite sign than either the maximum or minimum value signs, a direction violation is recorded. This

violation is flagged as a warning to the user, and it does not impact cluster formation in any way. However, means below the specified TPT are exempt from this violation under the assumption that these means may be considered zero in the modeling environment.

4. GENERAL EVALUATION

The reduction process typically begins by evaluating the impact that different values of epsilon have on cluster formation. As alternative values are attempted, the order of cluster formation indicates the relative similarities among GSF cases included in each cluster. Patterns begin to emerge, which help to guide the selection of appropriate epsilon values.

The four-area example is used to demonstrate the cluster formation associated with three different epsilon values: Figure 4 for epsilon = 0.500; Figure 5 for epsilon = 0.750; and, Figure 6 for epsilon = 1.000. Examination of these three figures illustrates a definite pattern as clusters tend to expand and absorb electrically similar buses.

The subareas shown in Figures 4, 5, and 6 are derived from the clusters generated by the clustering algorithm. However, individual clusters cannot be used to suggest subareas unless all clusters support the suggestion. There must be complete agreement among all subareas, as they are combined pair-wise with adjacent subareas.

The number of interarea contracts and reduced contracts resulting after cluster formation is tabulated in Table 1 for each value of epsilon. Of the 506 bidirectional contracts mentioned earlier, only 340 contracts are interutility bidirectional contracts. This number can be confirmed by using Figure 3 to count the number of interarea pair-wise bus combinations. Table 1 also summarizes the composite percent reduction in contracts resulting from the clustering process.

During this phase of analysis, the primary metrics used to determine cluster acceptance are the percent contract reduction, the normalized frequency distribution for varying levels of epsilon, and the selected TPT value. Table 2 shows the frequency distribution for each value of epsilon using a TPT = 0.0. A reasonable TPT = .080 (representing an 8% transaction participation threshold) is also used to show the percent error impact of an active TPT (see Table 3). Comparison of Tables 2 and 3 illustrates the sensitivity of very large percent errors to small (< .080) means. A TPT between 5% to 15% may be appropriate considering the inherent aggregation within clusters. Note that Tables 2 and 3 depict the frequency distribution of all generated clusters prior to enforcing internal, multi-area, or subarea violations. As a result, these numbers are further evaluated in the next section.

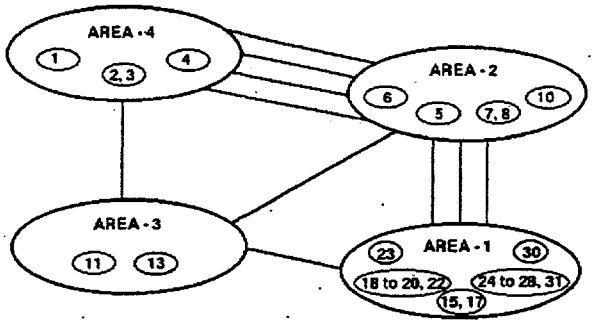

Figure 4. 0.500 Cluster Results

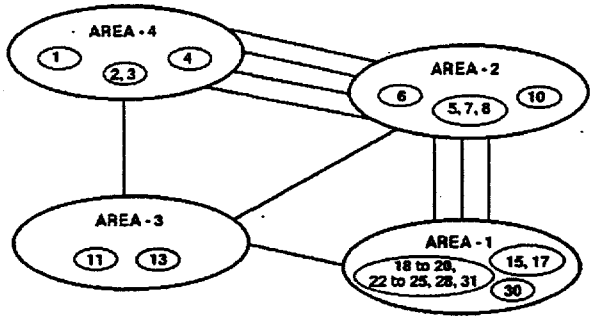

Figure 5. 0.750 Cluster Results

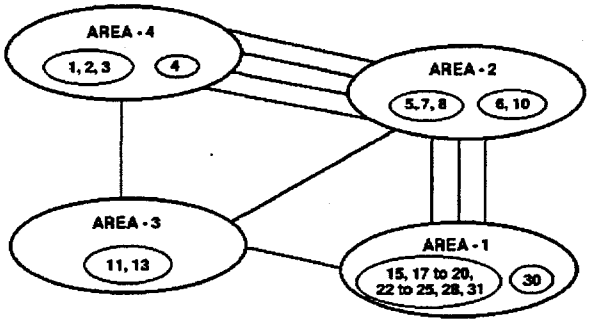

Figure 6. 1.000 Cluster Results

Table 1. Interarea Contract Comparison

Epsilon	Total Contracts	Reduced Contracts	Percent Reduction
0.500	340	142	58.2
0.750	340	90	73.5
1.000	340	36	89.4

5. DETAILED RE-EVALUATION

The initial examination of Tables 1 and 3 permits the selection of an appropriate level of epsilon for the system under consideration. The 8% TPT value is appropriate for the intended modeling requirements, and a 73.5% reduction in contracts is reasonable. These tradeoffs between numerical accuracy and contract reduction are acceptable for the requirements set forth in this study, and can be achieved with a corresponding value of epsilon set to 0.750.

The detailed re-evaluation described in this section uses the 0.750 value of epsilon and the subarea representation in Figure 5 as the basis for improved statistical analysis. This process determines the final contract definitions and their associated statistics. These final numerical values can be used to assess resulting cluster performance.

By direct examination of Figure 5, the number of interarea, bidirectional contracts for Area 1, Area 2, Area 3, and Area 4 is 3*8=24, 3*8=24, 2*9=18, 3*8=24, respectively. The total of 90 bidirectional contracts agrees with the data tabulated in Table 1. Therefore, a total of 45uni-directional contracts results from the 90 bidirectional contracts defined by this analysis.

Table 2. Percent Frequency Distribution for Varying Values of Epsilon Using TPT = 0%

Maximum Percent Error	0.500	0.750	1.000
$i \le 10.0$	12.0	12.2	15.8
$10.0 < i \le 25.0$	0.0	0.0	0.0
$25.0 < i \le 50.0$	10.0	7.3	10.5
$50.0 < i \le 75.0$	6.0	7.3	5.3
$75.0 < i \le 100.0$	10.0	7.3	0.0
$100.0 < i \le 150.0$	14.0	12.2	0.0
$150.0 < i \le 200.0$	12.0	12.2	10.5
$200.0 < i \le 300.0$	12.0	14.6	21.1
$300.0 < i \le 400.0$	8.0	4.9	0.0
$400.0 < i \le 800.0$	4.0	4.9	0.0
$800.0 < i \le 1200.0$	2.0	2.4	10.5
$1200.0 < i \le 1800.0$	0.0	2.4	10.5
$1800.0 < i$	10.0	12.2	15.8

Table 3. Percent Frequency Distribution for Varying Values of Epsilon Using TPT = 8%

Maximum Percent Error	0.500	0.750	1.000
$i \le 10.0$	16.0	14.6	21.1
$10.0 < i \le 25.0$	12.0	9.8	0.0
$25.0 < i \le 50.0$	28.0	24.4	15.8
$50.0 < i \le 75.0$	14.0	12.2	0.0
$75.0 < i \le 100.0$	8.0	9.8	0.0
$100.0 < i \le 150.0$	20.0	24.4	21.1
$150.0 < i \le 200.0$	2.0	0.0	26.3
$200.0 < i \le 300.0$	0.0	4.9	10.5
$300.0 < i \le 400.0$	0.0	0.0	5.3

The value of epsilon = 0.750 produces 41 clusters. Of these 41 clusters, 18 clusters directly correspond to valid clustered contracts. The remaining clusters produce violations that inhibited their use in the analysis at the 0.750 level of epsilon. Contracts represented by the 18 clusters are all eight contracts between Areas 2, 3, 4 and Area 1, Buses (18 to 20, 22 to 25, 28, 31), all eight contracts between Areas 2, 3, 4 and Area 1, Buses (15, 17), one contract between Area 4, Bus 4 and Area 2, Buses (5, 7, 8), and one contract between Area 2, Buses (5, 7, 8) and Area 3, Bus 13.

Given the 45 interarea contracts and the 18 contracts represented using the clusters formed at epsilon = 0.750, the remaining 45-18 = 27 contracts need to be re-evaluated by using smaller epsilon values. A smaller epsilon is necessary to repartition clusters that were aggregated beyond permissible limits and violated subarea boundaries when epsilon = 0.750. For example, epsilon = 0.500 separated the Bus (5, 7, 8) contracts that had combined destinations (i.e., (Buses (11, 13)). Epsilon = 0.250 broke down the Buses (1, 2, 3) subarea into distinct Bus 1 and Bus (2, 3) clusters as suggested in Figure 5. The following nine contracts are determined using these two values of epsilon and a TPT of 0.080 to remain consistent with established modeling requirements:

- Area 2, Buses (5, 7, 8) to Area 3, Bus 11
- Area 2, Buses (5, 7, 8) to Area 1, Bus 30
- Area 2, Buses (5, 7, 8) to Area 4, Bus 1
- Area 2, Buses (5, 7, 8) to Area 4, Buses (2, 3)
- Area 4, Buses (2, 3) to Area 2, Bus 6
- Area 4, Buses (2, 3) to Area 2, Bus 10
- Area 4, Buses (2, 3) to Area 3, Bus 11
- Area 4, Buses (2, 3) to Area 3, Bus 13
- Area 4, Buses (2, 3) to Area 1, Bus 30.

Choosing a smaller value of epsilon to achieve repartitioning also improves the statistical behavior of the new clusters.

The nine contracts above are subtracted from the 27 contracts prior to the re-evaluation to give 18 remaining contracts. These 18 contracts are represented as individual GSFs, since they represent singleton clusters (i.e., clusters that contain only one element). The contract between Area 4, Bus 1 and Area 2, Bus 6 is representative of a singleton cluster.

In summary, the total number of uni-directional contracts can be determined by adding the 18 singleton clusters, 9 re-evaluated clusters, and 18 clusters resulting from epsilon = 0.750. The resulting frequency distribution of the 45 contracts are shown in Table 4. The corresponding reduction in generator siting alternatives is (1-11/23)*100= 52%. Since there is no statistical error in singleton clusters, a new row indicating a 0% error is added to the distribution. In addition, more than half of the clusters in the 100% to 150% range are actually very close to 100%.

6. SUMMARY

The application of clustering techniques provides a data reduction benefit as the GSFs in interutility tie lines are examined. GSFs are chosen as the clustering criteria, because they support parallel path analysis and provide an acceptable method for transmission modeling. The relationship between contract impacts on these tie lines and the choice of generation siting alternatives is also facilitated by the use of GSFs. However, detailed network studies are needed to fully understand transmission system impacts.

Table 4. Final Percent Frequency Distribution for Epsilon = 0.750 Using TPT = 8%

Maximum Percent Error	0.750
i = 0.0	40
0.0 < i ≤ 10.0	2.2
10.0 < i ≤ 25.0	8.9
25.0 < i ≤ 50.0	20.0
50.0 < i ≤ 75.0	11.1
75.0 < i ≤ 100.0	0.0
100.0 < i ≤ 150.0	17.8

A four-area system illustrates the method's applicability to a 10 interutility tie example. Several statistical metrics and modeling heuristics are employed to determine criteria for clustering acceptance. Tradeoffs between cluster precision and effective cluster aggregation are described and illustrated using the example system.

An acceptable error margin produces a reduction in the number of representative cases based on the GSF performance. The number of contracts needed to characterize the system is reduced by nearly 75%, while the number of generator siting options is reduced by 52%. This method can be applied to generator siting and contract option applications to reduce transmission modeling complexity or may be employed as a screening method to identify the need for detailed system studies.

7. ACKNOWLEDGEMENT

This manuscript has been authored by a contractor of the U.S. Government under contract no. W-31-109-ENG-38. Parts of this work have been directly funded by the U.S. Department of Energy, Office of Fossil Energy.

8. REFERENCES

[1] J.A. Kavicky and S.M. Shahidehpour, "Parallel Path Aspects of Transmission Modeling," *IEEE Transactions on Power Systems*, Vol. 11, No. 3, pp. 1180-1190, 1996.

[2] J.A. Kavicky and S.M. Shahidehpour, "Determination of Generator Siting and Contract Options Based on Interutility Tie Line Flow Impacts," *IEEE Transactions on Power Systems*, 97WM-518, 1997.

[3] R.R. Austria and M.R. Pangilinan, "A Closer Look at Long-Term Transmission Planning in a Competitive Environment," *Power Technology*, No. 82, Power Technologies, Inc., July 1995.

[4] G.C. Canavos, *Applied Probability and Statistical Methods*, Little, Brown & Company, New York, NY, 1984.