EX PARTE OR LATE FILED

AT&T

Frank S. Simone
Government Affairs Director

Suite 1000
1120 20th Street, N.W.
Washington, DC 20036
202 457-2321
FAX 202 457-2165
fsimone@lgamgw.attmail.com

February 3, 1998

RECEIVED
FEB - 3 1998
FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

Ms. Magalie Roman Salas
Secretary
Federal Communications Commission
1919 M Street, N. W., Room 222
Washington, D. C. 20554

Re: *Ex Parte*, CC Docket No. 96-98; RM 9101 - Implementation of the Local Competition Provisions of the Telecommunications Act of 1996

Dear Ms. Roman Salas:

A copy of the enclosed analysis, Performance Measurements - Implementation Considerations, was delivered today to Jake Jennings, Radhika Karmarkar, Wendy Lader, Michael Kende, Don Stockdale and Carol Mattey of the Common Carrier Bureau for inclusion in the record of the above-mentioned proceeding.

Two copies of this Notice are being submitted to the Secretary of the FCC in accordance with the Section 1.1206(a)(1) of the Commission's rules.

Sincerely,

Enclosures

cc: C. Mattey
D. Stockdale
J. Jennings
R. Karmarkar
W. Lader
M. Kende

No. of Copies rec'd 0+2
List ABCDE

Performance Measurements - Implementation Considerations

State and/or Federal orders imposing performance measurements obligations on ILECs should be prescriptive and detailed if the desired result of determining the ILEC's compliance with its nondiscrimination obligations is to be achieved. A commission order requiring ILECs to file performance measurements should include a proposed plan and a schedule for implementation in order to achieve this objective.

This paper describes a set of steps, which if incorporated in an order or post-order activity, will increase the usability of the resultant performance measurements.

A. The Performance Measurement Plan

An appropriate performance measurement plan is needed. The plan should include clear and precise definitions of: the measurements to be made; the measurement data to be collected; the conditions for which data are excluded from the measurements and the reporting dimensions for which measurements may be disaggregated for purpose of comparisons. The plan should provide for a start up period that is of sufficient length during which measurements demonstrate stability across a sufficiently broad range of resold services and unbundled network elements. It should include generally acceptable statistical tests of differences to be used in comparing the CLEC performance measurement with the ILEC performance data. It should also specify the report content, reporting schedule and compliance criteria which are to be used to determine the ILEC's compliance with its nondiscriminatory obligations. Finally, it should provide for appropriate audit procedures so that all parties can rely with confidence on the data reported by the ILEC.

The ILEC should demonstrate that the measurement plan will gather and retain data in a manner that permits meaningful tests for statistically significant differences in performance. The measurement plan should permit each measurement to be tested and a determination made, at a generally acceptable level of statistical significance, that the CLEC results are no worse than those experienced by the ILEC's own retail local service operations or those of any of its affiliates. The 95% confidence level is the preferred level of statistical significance.

B. Statistical Testing Considerations

<u>Minimum Sample Size And Its Significance</u>

The performance being measured is based on a data set which represents a specific outcome, or realization, of a process which contains randomness. This

random process is one which theoretically repeats itself an indefinite number of times. This data set is therefore a sample in that it represents only one of a very large number of possible repetitions of the random process. Since time cannot be repeated, it is also the only outcome of the random process that will be measured for a given time or time interval (e.g., the month of June, 1998).

One statistical test compares the difference between the sample means for two samples, usually different sample sizes, which are taken over the same time period. In this context, one sample is from the ILEC's process (e.g. service order intervals for a set of N_1 orders) and the other is from the CLEC process (e.g. service order intervals for a set of N_2 orders). N_1 and N_2 are the sample sizes.

For purposes of this discussion, assume that the individual elements of the data set are service order intervals. The data set itself contains variation from data element to data element being measured (i.e., the order interval for each order will usually be somewhat different). This is reflected in the variance of the sample. As the sample size increases, the probability distribution of the sample mean becomes approximately normal (the assumption associated with the test), the variance of the sample mean approaches zero, the sample variance approaches the population variance and the sample mean approaches the process mean for the specific time period for which the measurements are made (e.g., the month of June, 1998). The size of the sample for which the statistical testing assumption is valid is approximately 30. However, the size of the sample for the variability of the sample mean to be insignificantly small may be two or more orders of magnitude greater than 30.

This gives rise to two possible criterion for selecting the minimum sample size. The first is to define the minimum sample size so that the underlying test assumptions are valid. The other is to select the sample size so that statistical variation associated with the test statistic (e.g., the z statistic) is essentially zero and conclusions can be drawn based on direct comparisons without the need for the statistical test.

As a result, the criterion for selecting the sample size has a considerable consequence on the degree of disaggregation that may be tested and reported. A sample size which is based on valid test methodology will be much smaller than that for a sample size which is based on minimal variance. The smaller sample size will exhibit more variability, thus requiring larger differences to exist in the means being compared before an "out of parity" determination could be made. A sample size which is based on minimal variance, and thus allows a more direct comparison, will require large sample sizes and consequently rule out the possibility for levels of disaggregation which may be needed in order to prevent masking of discrimination due to extensive aggregation.

The primary purpose for the application of comparative performance measurements is to determine, at the appropriate level of disaggregation, whether the ILEC is in compliance with its nondiscrimination obligation. Therefore, the desired criterion for sample size is one that permits greater

disaggregation. This in turn recognizes the need for larger differences to reject parity than would be required for sample sizes based on the minimal variance criterion.

C. Compliance Criteria

The following criteria should be used to determine an ILEC's compliance with its nondiscrimination obligations. The criteria cover three dimensions. Being in compliance means meeting the compliance criteria along all three dimensions.

a) Maximum number of comparisons failing the test: It is necessary to determine the degree to which individual tests are allowed to "fail" before an ILEC may be found to be discriminatory. For example, at the 95% confidence level, there is a 5% chance that an individual comparison will fail the test when, in fact, the results are not different. One would expect that, on average, five out of 100 comparisons of measurements that are not different would fail the test at the 95% confidence level. Therefore, using this confidence level, a reasonable reporting conclusion would be that nondiscrimination is achieved if no more than 5% of the individual comparisons "failed" the test. That is, this test would be subject to a 5% threshold.

b) Maximum repeating measurements failing the test: Using the preceding example, one would expect that, statistically speaking, no more than .25% (.05 x .05 = .0025) of the performance measurements would fail the test in two consecutive periods (i.e., this month and the previous month). Therefore, using this confidence level, a reasonable reporting conclusion would be that nondiscrimination is achieved if no more than .25% of the individual comparisons "failed" the test in two consecutive months.

c) Measurements exhibiting extreme differences: Depending on the market entry phase, a relatively small number of performance measurements may "make or break" the CLEC's ability to effectively compete. For example, at initial market entry, very large delays in obtaining Pre-Ordering information or very large differences in CLEC vs ILEC service order intervals would prevent the CLEC from effectively gaining market entry. The downstream operational performance categories such as Maintenance and Repair, Billing, Network Performance, etc. become important after a significant customer base is established. An extreme difference is defined here as a difference between the CLEC and ILEC performance measurements of at least three standard deviation ("three sigma").

Therefore, in order to ensure that an ILEC complies with its nondiscrimination obligations, compliance criteria that demonstrate multiple dimensions along which the ILEC can deter meaningful market entry should be used. That is, an ILEC may meet compliance criteria along one dimension and not along other

dimensions. In this context, being in compliance means meeting the compliance criteria along all dimensions.

See Appendix A for additional discussion on compliance criteria.

D. Process Used To Determine Parity

The following is a description of a stepwise process which can be used to develop the comparative performance reporting.

1. Collect and maintain performance data on a monthly basis. Include reporting dimensions with each measurement.
2. Calculate the measurements based on the measurement formulas, data definitions and data exclusions.
3. Make comparisons, disaggregated by reporting dimensions, based on the same measurements for the CLEC performance and the ILEC performance using the appropriate statistical tests;
    - When identical ILEC performance measurements are not obtainable, make comparisons based on the CLEC performance and analogous ILEC performance.
    - However, when a particular ILEC performance measurement, or an appropriate analog, is not available for comparison, the CLEC performance measurement should be compared to a benchmark performance level. A special study should be utilized by the ILEC to establish the benchmark performance level whenever a reasonable ILEC analog does not exist. The study method should be disclosed in advance and an audit of the results should be permitted.
    - Each measurement has its associated reporting dimensions. Reporting dimensions are pre-established markings of the data collected for each performance measurement and are required to permit meaningful sorting and direct comparison of results. Each pair of measurements (CLEC performance measurement and either ILEC performance measurement or ILEC benchmark) may contain values based on one or more reporting dimensions. For example, the markings associated with the data collected for an ordering and provisioning measurement can provide the ability to report results based on service type (e.g., Business POTS and Residence POTS).
    - Parity is a showing that measurements of CLEC performance and ILEC performance, based on identical or analogous functionality, meet specific comparison criteria using the appropriate statistical tests. Statistical comparison tests are based on the maximum statistical

variation allowed for the data points[1] being compared. If the two data points are different by more than this maximum, they are considered to be statistically "not the same". For example, a statistical test of means would compare the z statistic, which is a function of the difference between the ILEC and CLEC performance measurement means, with a 95% confidence level value of 1.65. If the value of z was less than or equal to 1.65, the test of parity passed. If z was greater than 1.65, the test of parity failed.

When an analog is not available, nondiscrimination may be demonstrated by measuring CLEC performance against an ILEC benchmark that provides CLECs a meaningful opportunity to compete.

4. Determine compliance based on the above compliance criteria:
5. Prepare compliance reports

---

1 A data point pair is either 1) a CLEC performance measurement result and an ILEC performance measurement result; or 2)a CLEC performance measurement result and an ILEC (or LCUG default) benchmark.

Appendix
Performance Reporting

Consequence of Statistical Error On Compliance Determination

The difference between two test statistics (e.g., sample means) is used to determine if they are or are not alike. In making this determination, it is also necessary to establish a confidence level associated with this determination. This is accomplished by choosing the critical value for the test statistic (e.g., z statistic when testing the differences in sample means). A larger critical value means that there is a smaller likelihood that a comparison will erroneously reject the hypothesis that the CLEC and ILEC performance measurements are the same. However, a larger critical value will also increase the probability of accepting this hypothesis when it is not true. This situation creates competing objectives: choosing the critical value in order to reduce the probability of rejecting the hypothesis when it is true also increases the probability of accepting the hypothesis when it is not true. A confidence level must be chosen so as to strike a balance between these competing objectives.

However, a consequence of choosing a confidence level of less than 100% in the determination of compliance, is that there is a non-zero probability (e.g., 5%) that the statistical tests will produce the erroneous conclusions that CLEC and ILEC performance measurements are not at parity. Therefore, a compliance determination must recognize that there is a threshold of "failed tests" that must be exceeded before an ILEC can be determined as being out of compliance with its nondiscriminatory obligations.

Support For The Use of The 95% Confidence Level

Extreme limits on the confidence level are not appropriate in cases where the importance of making an error in rejecting the hypothesis (i.e., parity) is approximately the same as making an error in accepting it[2]. Furthermore, the California Public Utilities Commission recently approved a "parity" provision developed by the staff of the Commission, which has been incorporated into the AT&T-Pacific Bell Interconnection Agreement. That provision states that, for purposes of defining Pacific's contractual service quality obligations to AT&T, "parity is achieved when Pacific's service performance [for AT&T], as defined by the designated comparable measures, is within 1.65 standard deviations (90% confidence level) of [Pacific's] average retail performance for the equivalent retail product or service."[3] In other cases, a difference of one standard deviation

---

2 *See Kendon v. AT&T Technologies*, 883 F.2d 388, 397-98 (5th Cir. 1989) (rejecting claim that a disparity of three standard deviations is an appropriate minimum requirement for statistical significance in discrimination cases).

3 *See* AT&T-Pacific Interconnection Agreement, Att. 17, p. 2. While the use of 1.65 standard deviations produces a 90% probability (90% confidence level) that a result outside of the parity range (that is, either significantly better or significantly worse than the mean) is not a random occurrence, where -- as in this case -- the only relevant issue to be tested is whether performance for one group was significantly ***worse*** (which may be a higher or lower number depending on the measurement), the only results of interest are those that are both outside the parity range and worse (a "one-tailed" test). If only

was judged as appropriate in determining whether there were significant differences in the quality of water when water samples were compared to a standard[4].

There are two types of risks associated with making a false conclusion. Type I: Risk defined as the probability that parity will be falsely rejected. That is, a comparison of ILEC and CLEC measurements falsely yielded a conclusion that the ILEC was not at parity. Type II: Risk defined as the probability that parity will be falsely accepted. That is, a comparison of ILEC and CLEC measurements falsely yielded a conclusion that the ILEC was at parity. While the later risk type cannot be quantified in the statistical testing methodology (i.e., test of means), it can be qualitatively identified relative to the first risk type The three cases cited can be placed on a risk scale as shown in the following table:

	Degree Of Risk In Rejecting Parity		
Type I Risk	High	Medium	Low
Type II Risk	Low	Medium	High
Confidence Level	84%	95%	99.99%
Standard deviations	1.00	1.65	3.00

In the above table, 84% represents a case in which it was concluded that the importance of water quality favored a higher risk of inappropriately rejecting a conclusion that water quality was at parity with the standard. The 99.99% represents a case in which it was concluded that the weight given to falsely rejecting nondiscrimination (parity) should be more in balance with the weight associated with falsely accepting it. Thus, limiting the probability of falsely rejecting parity to .01% was considered as being extreme.

Translating the High and Low Risk levels to consequences in the ILEC vs CLEC comparisons, a difference representing one standard deviation would create a condition which probably unduly favors the CLEC while a difference representing three standard deviations unduly favors the ILEC. Therefore, choosing 95% (i.e., a 5% probability of rejecting parity when it is true) strikes a more reasonable balance, based on the cases cited, between risk to ILEC vs CLEC.

---

worse performance is pertinent, the use of 1.65 standard deviations corresponds to a ***95%*** confidence level that a ***worse*** result outside the parity range is not a random occurrence. *See Palmer v. Shultz*, 815 F.2d 84, 92, 95-96 (D.C. Cir. 1989).

4 See AMAX, Inc. vs Colorado Water Quality Control Comm. 790 P.2d 879, Dec 21, 1989, Colorado Court of Appeals, Division Two.