July 31, 2001

Via Email

Ms. Kathleen Cummings
State Corporation Commission
Division of Communications
1300 East Main Street
Richmond, VA 23219

Re: Ex. Rel. State Corporation Commission, Ex Parte: In the Matter of Establishment of a Collaborative Committee to Investigate Market Opening Measures; Case No. PUC 00026

Dear Ms. Cummings:

WorldCom submits the following letter in lieu of comments on the June 22, 2001 Verizon Virginia, Inc. and Verizon South Inc. ("Verizon" or "VZ") proposed performance plan for the Commonwealth of Virginia. To encourage the development of effective and sustainable competition in the local exchange market, the State Corporation Commission ("Commission") must ultimately adopt stringent, self-effectuating remedies for violations of performance standards by Verizon. An effective remedies proposal must have consequences that are severe enough to deter misconduct rather than merely being a cost of doing business. Only through such stringent remedies will VZ be incented to modify its behavior and strive toward compliance. VZ's plan, as proposed, will do nothing to incent VZ to provide quality service to its competitors.

I. Structure and Incentive Levels Proposed by VZ Will Not Incent VZ to Improve Performance

WorldCom is concerned with the structure and incentive levels proposed by VZ in it's performance plan. VZ proposes per occurrence remedies that are extremely low -- $9 to $42 for three levels of misses. These per occurrence incentive levels proposed by VZ could easily become simply a cost of doing business and provide no incentive to VZ to expend human and capital resources to actually improve performance. In Michigan, the Commission recently compresses the severity categories into one and adopted significantly higher remedy amounts as reflected in the below chart.

<u>Michigan</u>

LIQUIDATED DAMAGES TABLE FOR TIER-1 MEASURES

Per occurrence						
Measurement Group	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6 and each following month
	$150	$300	$600	$800	$1000	$1200

Per Measure						
Measurement Group	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6 and each following month
	$20,000	$40,000	$60,000	$80,000	$100,000	$120,000

In addition, in Texas, Kansas, Oklahoma and Missouri, higher remedy amounts have also been adopted, as reflected below:

Texas, Kansas, Oklahoma, Missouri
LIQUIDATED DAMAGES TABLE FOR TIER-1 MEASURES

Per occurrence						
Measurement Group	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6 and each following month
High	$150	$250	$500	$600	$700	$800
Medium	$75	$150	$300	$400	$500	$600
Low	$25	$50	$100	$200	$300	$400

Per Measure/Cap*						
Measurement Group	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6 and each following month
High	$25,000	$50,000	$75,000	$100,000	$125,000	$150,000
Medium	$10,000	$20,000	$30,000	$40,000	$50,000	$60,000
Low	$5,000	$10,000	$15,000	$20,000	$25,000	$30,000

***** **For per occurrence with cap measures, the occurrence value is taken from the per occurrence table, subject to the per measure with cap amount.**

These higher amounts provide more of an incentive to the incumbent to provide quality service to its competitors. WorldCom requests that Staff, and in turn the Commission, adopt per occurrence remedy levels that are significantly higher than those proposed by VZ to effectively eliminate any financial incentives that VZ may have to provide competitors with poor quality service.

WorldCom also requests that the Staff propose to the Commission that per measure incentives be added in addition to per occurrence incentives in situations of sustained and/or severe poor performance by VZ. Even if set at levels significantly higher then the VZ proposal for Virginia, per occurrence remedy plans alone, are inadequate to deter poor performance by the incumbent. Per Occurrence plans may work when robust competition has already developed and few new products are coming to market. However, for the current Virginia market, where competition is still struggling for a foothold, a per occurrence incentive plan can easily become a cost VZ will readily pay to stifle competition. Per occurrence remedy plans fail to protect competitors when they need it most, by keeping remedies the lowest when competitors are just beginning to ramp up in a market or launching new services in competition with the incumbent.

A combined per occurrence and per measure approach is best for opening new markets to competition and ensuring that CLEC's new market entry activity or launches of new service offerings are not crushed at introduction with no substantial financial risk to VZ. Per occurrence remedies may be reasonable for first time, low level misses but once misses extend into the second, third or fourth month, then per measure remedies should be invoked in addition to per occurrence incentives.

WorldCom is also concerned with how VZ proposes to address the frequency of a miss. VZ suggests handling the problem of consecutive failures in a manner that only increases the remedy by 50% for 2 or more consecutive months and then remains constant after three months. This will bring little comfort to a CLEC experiencing poor performance for more than three months at a time. The plan proposed by VZ does not emphasize enough the detrimental effect of ongoing performance failures. In addition, the plan as proposed does not provide a great enough incentive for VZ to change its behavior. Each month that goes by without quality service provided by VZ affects WorldCom's customers and reflects poorly on WorldCom -- this impression degrades further the longer the poor performance continues. Given that some of these consecutive misses/failures could be ongoing for the same customer orders, the remedy should increase by at least 100% each month, not a mere 50%. WorldCom recommends adding one month's remedy amount for each consecutive month; that is, starting with a base (x) remedy amount for the first month missed. If the second month is missed the remedy amount becomes 2x; the third month is 3x and so on up to 6x for six consecutive months of poor performance. This approach is currently used for per measure remedies in the Texas, Michigan, Kansas, Oklahoma, and Missouri plans.

WorldCom further requests that the Staff consider requiring VZ, once it has missed a standard and is paying a higher incentive due to the miss, to show compliance with the standard for three consecutive months before it is allowed to return to normal incentive levels. Further, if non-compliance with the same standard occurs again, WorldCom requests that the Staff require VZ to illustrate compliance with the standard for six consecutive months before a return to first month incentive levels can occur. These requirements will provide VZ with the incentive to remedy existing performance problems.

Finally, VZ notes in its proposal that not all measures in the guidelines are included in its incentive plan. Specifically, VZ notes that guidelines that have no performance standard or are redundant with other measures that are eligible for incentive credits have not been included in credit calculations under its proposal. VZ should not be allowed to unilaterally determine which measures should be included in the incentive plan. It is clear from collaboratives in other jurisdictions that the CLECs do not necessarily agree with what VZ deems to be redundant. The Staff and the parties should be given an opportunity to sit down with VZ and discuss the selection of the metrics to be included in the incentive plan once the measures themselves are finalized for Virginia.

II. Use of the K-factor As Proposed by VZ is Not Appropriate

VZ proposes the use of the K-factor as a means of offsetting the 5% probability that VZ will be found out of parity when performance is actually in parity. WorldCom does not agree with the use of the K-factor in this proposal. WorldCom believes that no forgiveness should be allowed with a statistical test that has a 95% confidence level, i.e. uses a -1.645 critical value. Any modified Z score that is negative indicates that the CLEC received poorer performance than the ILEC. Using a –1.645 score before even low-level remedies are triggered provides adequate protection against random variation. Most notable among ILEC proposals for forgiveness of performance failures resulting from random variation is the K-table used to exclude failed metrics from remedies in Texas. This mechanism stemmed from an AT&T statistical expert's proposal for determining compliance with the Act if larger numbers of tests for CLECs are aggregated into one report.[1] The mechanism was never meant to offset Type I errors (ILEC wrongly found guilty—a chance of 5% with a –1.645 critical value) as proposed here by VZ. The Texas K-Table is calculated improperly for its current use, and does not take into account Type II errors (ILEC wrongly found not guilty—usually a greater chance than Type I errors at confidence levels at or above 95%). If any forgiveness is due statistically, WorldCom's statistical consultant proposes that it should only be three forgivenesses per five years. As the Texas K-Table is designed, (1) failures that are repeated consecutively (thus diminishing the chance that the failure was random) and (2) failures at large negative z scores and even 99.9% confidence levels with large means differences could be

[1] WorldCom at one point proposed using the K-Table only for a three-tiered remedy plan where the second tier went to CLECs as an added remedy when a large percent of metrics were missed.

forgiven. A more detailed paper on this subject from WorldCom's statistical consultant Dr. John Jackson of Auburn University is attached to these comments.

III. VZ's Plan Omits Several Integral Provisions of the New York Plan

WorldCom also proposes that since the reliable operation of CLEC interfaces with VZ's Operation Support Systems is crucial to the development of competition, the Staff should include in any proposal to the Commission the NY Change Control Assurance Plan. VZ must be subject to remedies for failure to follow proper change control notice, documentation, software certification and error correction processes. Discrimination in this area can have an especially negative effect on local growth opportunities.[2]

In addition, the New York Performance Assurance Plan created, what are referred to as Special Measures, a super measurement-based remedy, focused on past performance weaknesses of Verizon. In New York, the Special Measures divides large remedy amounts ($2.5 million quarterly for flow through, $2 million monthly for hot cuts, $2 million monthly for missing notices, for example) among the competitors when benchmarks are missed for certain metric groups. In New York the Special Measures addressed the following 5 areas of service quality, which are particularly critical: UNE flow through, UNE ordering, Hot Cuts, Local Service Request Confirmations and Reject Notices. The New York Commission believed that these measures subject to the Special Measures provision were persistent problems that needed a significant incentive to outweigh the costs and competitive advantages of not fixing the underlying operational problems. The Staff should propose that the Special Measures provision created in New York be adopted in Virginia. Finally, the Staff should also propose to the Commission that 5% of the proposed annual cap (if a cap is adopted) be proposed for additional special measures incentives to address weak areas in VZ performance highlighted by the Third Party OSS test, which is ongoing in Virginia. All of these adjustments to the VZ proposal would render the incentive plan more effective and help to ensure reliable OSS, which is critical to the development of effective and sustainable competition in the local exchange market.

IV. The Performance Plan Should be Implemented Immediately and Should Not Supercede Contract Remedies

The Staff should propose that the performance incentive plan take effect as soon as the Commission adopts it, in order to encourage compliance with the market opening requirements of section 251 of the 1996 Telecommunications Act. Current contract remedies do not sufficiently deter discrimination. It is essential to impose sufficient performance plan remedies in addition to the existing contract remedies available to individual CLECs in order to motivate VZ to open its markets. In recognition of this need, the state commissions in Texas and Pennsylvania chose not to wait for 271 approval to implement remedies that improve on existing contractual remedies.[3] Moreover, even in New York, where substantial individual contract remedies were in

---

[2] Further, once the Metrics Change Control Process that is being developed in New Jersey is finalized, the Staff should include it in Virginia's performance incentive plan.

[3] See Opinion and Order, Joint Petition of NEXTLINK Pennsylvania, Inc., et al., for an Order Establishing a Formal Investigation of Performance Standards, Remedies, and Operations Support Systems Testing for Verizon-Pennsylvania, Inc., Pennsylvania PUC Case No. P-00991643 (December 31, 1999), at 178-180; Order Adopting the Texas 271 Agreement, Texas PUC Project No. 16251 (October 13, 1999).

place, the “special measures” provisions of the PAP, dealing with certain particularly sensitive service areas, as outlined above, were implemented in the last quarter of 1999, prior to VZ-NY’s receipt of federal 271 approval. The Staff should propose to the Commission that the remedies plan be implemented as soon as possible, prior to 271 approval, so that the plan can be an active tool to open the local market as well as a safeguard against backsliding after VZ’s 271 approval.

Also, the Staff proposal to the Commission should reflect that the incentive plan would not supersede remedies available in the contract. In considering the deterrent value of remedies, Verizon’s proposed remedies should not supersede and replace existing contract remedies available to individual CLECs under interconnection agreements or legal settlements, as well as obviate any need for such remedies in future agreements. WorldCom does not ask for duplicate remedies, but where the same metric exists on two plans, VZ should pay the higher of the two amounts. VZ should be subject to remedies for all metrics covered by either plan. In New York, payments under the Performance Assurance Plan are cumulative to the payments available to individual CLECs under their individual interconnection agreements making total monetary incentives for good performance higher than those expressed on the face of the New York Plan. In fact, the FCC's order approving VZ-NY's 271 application specifically pointed to the liquidated damages in VZ's interconnection agreements as other remedies available to CLECs that would counter VZ's incentives to discriminate and thus partially justified the monetary cap contained in the New York Plan on this ground. The New York 271 Order at ¶ 430 explains that it may permit 271 entry even though a state Performance Assurance Plan alone provides less than full protection against anticompetitive behavior because of additional incentives for ILEC compliance including "payment of liquidated damages through many of its individual interconnection agreements." Thus, eliminating this additional remedy, as proposed by VZ in Virginia, undermines overall deterrence.

Finally, parties should be allowed to negotiate additional measures and remedies in their contracts, but contract inclusion should not be necessary to gain coverage and receive performance reporting under the incentive plan. The Commission should retain the authority to review the incentive plan and the measures in a formal proceeding annually, and it should retain control in order to increase remedies, if necessary, to gain compliance in the interim in problematic performance areas.

V. Bill Credits Proposed by VZ Should be Replaced with Direct Payments

To ensure that remedies are not constrained by the amount of future business given by a harmed CLEC to Verizon, direct payments are preferable to bill credits. Credits create the perverse incentive of requiring a customer to buy more service to gain remedies for past poor service. Direct payments ensure that CLECs immediately receive the full amount of remedies rather than awaiting subsequent bills. This is especially necessary in VZ's Virginia plan where incentive credits for a given month might not appear until 3 or 4 billing cycles have passed. This time lapse could severely affect harmed CLECs who are trying to pay credits to customers, add resources to handle the inefficiencies, settle lawsuits and escalate problems caused by their major competitor and supplier. Consequently, VZ should pay remedies monthly and not gain from the float of monies due CLECs, as proposed.

Further, CLECs often must resort to withholding bill payments to gain VZ's attention to errors in billing when requests for adjustments are ignored. Bill credits may diminish the attention getting effect of this action. VZ may use the bill credits as offsets of the amount withheld by CLECs rather than adjusting the existing billing errors and then applying the plan. In addition, it is important for CLECs to know which metrics the credits apply to. This is impossible if a bill credit is received, as today in other states, with no explanation. A miscellaneous bill credit could be taken for many other things besides a remedy. Direct payments, if accompanied by an explanatory invoice, also make the amounts paid easier for CLECs to audit than bill credits. This may facilitate the self-policing aspect of the plan and reduce the Commission oversight.

At a minimum, if the Commission decides to support credits versus direct payments in full to CLECs, the VZ plan should be clarified. It should state that bill credits may be applied to any future bills, including access bills, and that if VZ no longer bills a CLEC, VZ will pay the CLEC by check. Without this modification, CLECs who have been forced out of business altogether (or out of a particular line of business), by VZ discrimination will not be able to recover under the VZ plan for that discrimination.

VI. VZ's Plan Should Not Have a Cap on the Total Dollar Amount at Risk

WorldCom is opposed to absolute monetary caps on remedy plans. If Staff determines that a cap is necessary, WorldCom supports a "procedural cap" that, when reached, allows VZ to seek regulatory review of the remedy payments that are due. VZ would continue to make payments into a designated account until the Commission decides if VZ has presented sufficient justification for not paying remedies in excess of the procedural cap. VZ would have the burden of showing, by clear and convincing evidence, that the remedies due in excess of the procedural cap are unwarranted. The Commission would then decide whether and to what extent the amount in excess of the procedural cap should be paid out.

If the Staff determines that an absolute cap is appropriate, it must do so at a level that makes discrimination the least rational choice for VZ. VZ's proposed cap of $36.3 million does not satisfy this criterion. It is not explained anywhere in VZ's proposal how the cap that it proposes for Virginia was derived. The FCC, in its analysis of the New York Plan monetary cap, determined that the original amount – prior to the addition of $24 million to cover missing metrics issues – was 36% of net local return. It is unclear from VZ's proposal what percent of net local return the $36.3 million annual cap represents. In addition, with the subsequent additions to the New York Plan that the NY PSC made subsequent to VZ-NY's 271 approval, the New York Plan now places 44% of net local return at risk.

Using publicly available ARMIS data, 36% of VZ's net local return in Virginia is $189.3 million and 44% of VZ's net local return in Virginia is $231.3 million.

Data for Verizon & GTE Virginia from ARMIS 43-01 (2000)						
(Downloaded from FCC Web Site: http://www.fcc.gov/ccb/armis/)						
Year	Company Name	Row_#	Row_Title	Total_b	State_g	Interstate_h
2000	Verizon & GTE	1090	Total Operating Revenues	2,471,687	1,611,801	694,849
2000	Verizon & GTE	1190	Total Operating Expenses	1,619,910	1,050,468	413,717
2000	Verizon & GTE	1290	Other Operating Income/Losses	-7,351	58	21
2000	Verizon & GTE	1390	Total Non-operating Items (Exp)	-16,926	-24,404	-1,895
2000	Verizon & GTE	1490	Total Other Taxes	116,508	80,897	32,795
2000	Verizon & GTE	1590	Federal Income Taxes (Exp)	222,671	151,877	77,489
2000	Verizon & GTE	1915	Net Return			172,763

FCC's Net Return Calculation*				
		Net Return	36% Net Return	44% Net Return
#REF!	"Net Return"	525,784	189,282	231,345

*Calculations in testimony based on FCC NY 271 Order at ft. 1332: "To arrive at a total "Net Return" figure that reflects both interstate and intrastate portions of revenue derived from local exchange service, we combined line 1915 (the interstate "Net Return" line) with a computed net intrastate return number (total intrastate operating revenues and other operating income, less operating expenses, non-operating items and all taxes)." Following the FCC's guidelines, the 'Net Return' is [172763+1611801+58 - (1050468+-24404+80897+151877)]= $525784.

Both amounts are a far cry from the $36.3 million that VZ has proposed. The Staff must ensure that remedy levels are set at a minimum level that will incent VZ to expend the resources and capital to correct disparities that add to its bottom line by stalling competition in the local market in Virginia. The amount proposed by VZ will not provide such an incentive. WorldCom proposes that the cap, if adopted, be set at 44% of VZ's net local return as adopted in New York and Georgia. In Virginia that would be $231.3 million.

Finally, WorldCom objects to VZ's inclusion of the following as Force Majeure events: "labor slowdowns, picketing, or boycotts, unavailability of equipment, parts or repairs thereof." VZ is capable of anticipating and planning for such events. If they choose not to act to resolve them, and such lack of action causes a missed metric, that miss should generate a remedy. With respect to "unavailability of equipment, parts or repairs thereof," how can CLECs be assured that available equipment is not being reserved for VZ's customers to the detriment of CLEC customers? On non-parity measures, it would be impossible to know.

By definition "force majeure" means: 1) superior or irresistible force; 2) an event or effect that cannot be reasonably anticipated or controlled.[4] While this definition clearly applies to the rest of the items in VZ's force majeure category ("unusually severe weather conditions, earthquake, fire, explosion, flood, epidemic, war, revolution, civil disturbances, acts of public enemies, any law, order, regulation, ordinance or requirement of any governmental or legal body…and any acts of God), it does not apply to the events at issue. Labor slowdowns, picketing, and boycotts are not irresistible forces because they can be avoided or controlled through negotiations; they can be reasonably anticipated there as well. Equipment issues can be controlled and anticipated by no more superior a force than an up-to-date inventory. These events simply do not fit into the same category as those qualified to be force majeure events. VZ's definition of force majeure should be modified as detailed above.

****

In conclusion, WorldCom requests that the Staff adjust VZ's proposed incentive plan as discussed above before submitting the plan to the Commission for approval. Only through the implementation of these changes can the Staff and the Commission ensure that VNJ will be incented to modify its behavior and strive toward compliance

[4] Merriam-Webster's Collegiate Dictionary, Tenth Edition copyright © 2001 by Merriam-Webster, Incorporated.

with performance standards and in turn provide competitors with the quality of service to which they are entitled.

Respectfully submitted,

Kimberly A. Wild

Cc: Service List (via email)

RANDOM VARIATION, "FORGIVENESSES", AND "K-TABLES": A CLEC PERSPECTIVE

By

John D. Jackson
Professor of Economics
Auburn University, AL 36849

I. Introduction

The Telecommunication Act of 1996 provided for ILEC entry into the long distance telephone service market after CLECs were allowed to enter the various local telephone service markets. This CLEC entry, in turn, is predicated upon their ability to purchase from the ILEC various services crucial to their ability to compete in the local market. Consequently, the Act further requires that the ILEC provide these services to the CLECs at a quality level at least equal to that they provide to their own customers. Thus, the evaluation of parity in local service provision has become a central issue in all proceedings concerning ILECs' (1) obligation to open their local markets under the Act's section 251 and (2) opportunity to enter the in-region long distance market after satisfying the conditions set for in the Act's section 271. As a result, statistical means difference tests, typically based on (some version of) the Local Competition Users Group (LCUG) Modified Z statistic, have become the cornerstone in the evaluation of service quality provision. Indeed, test results are not only used to determine whether the ILEC has discriminated against the CLEC in service quality provision, they also enter into the determination of the magnitude of the penalty involved according to several performance assurance plans (such as those proposed by SBT, BST, and AT&T).

When one makes a decision concerning the presence or absence of parity in service provision based on a statistical test, he or she can err in one of two possible ways. One could conclude that discrimination in service provision exists when in fact it does not, or one could conclude that discrimination does not exist when in fact it does. Because the null hypothesis of the test assumes "no discrimination," the former error involves the rejection of a true null; it is called a type I error. The latter error involves the acceptance of a false null; it is called a type II error. Proposals made by some ILECs that use the notion of "random variation" as a basis for suggesting that some of their discriminatory acts (as determined by failed parity tests) should be "forgiven" (i.e., not penalized), where the number of violations to be forgiven is sometimes determined by a "K-Table" (see, e.g., the SBT plan), are founded exclusively on the existence of type I error. The purpose of this paper is to examine the underpinnings of such proposals and to evaluate their appropriateness from a CLEC perspective.

II. FORGIVING FAILED TESTS: THE BASIC RATIONALE AND A CLEC REACTION

The fundamental statistical test of parity service provision employed in almost all of the proposed performance assurance plans (PAPs) is a simple one-tailed means difference test conducted at the $\alpha=0.05$ level of significance. Since the probability of committing a type I error is equal to the level of significance of the test, each parity test incurs a five percent chance of concluding discrimination in service provision when parity in fact exists. ILECs describe such a decision as the result of "random variation" in the test statistic and not the result of actual discrimination on their part. They use this idea as the basis for the following argument:

> Suppose we supply the CLECs with 100 submeasures per month that are subject to parity testing. Each submeasure stands a 5% chance of failing its test each month due solely to random variation. Thus, *even if we supply every service in parity every month*, over the course of a year, each submeasure can be expected to fail 0.6 (12 mo. x .05) tests. (Since it is hard to think about failing a fraction of a test, aggregating further over time is helpful: Failing 0.6 tests in one year is equivalent to failing 3 tests in 5 years.) This means that, *even though we always are in parity*, testing 100 submeasures per month

> implies that 60 (0.6 x 100) tests will be failed over the course of a year (300 tests in 5 years) due strictly to random variation. (None could be failed due to discrimination, since it is explicitly assumed away). This result, in turn, implies that we should be "forgiven" (i.e., not penalized for) five test failures per month (60 per yr. / 12 mo.), since this is the number of tests (out of 100) that would be expected to fail due solely to random variation (*even if we are always in parity*)."

Honesty compels me to admit that the above is not really what the ILECs typically argue -- although it is certainly what they should argue. Usually, ILECs unabashedly ignore the statistical underpinnings that determine the "appropriate" number of forgivenesses, and they inflate the number of forgivenesses they demand with no obvious basis whatsoever. A personal anecdote will illustrate: In February 1999, I was involved (as a statistical consultant for MCI Telecommunications) in a joint workshop (CLECs, Pacific Bell, and the Public Utilities Commission's staff and Administrative Law Judge), which constituted the first attempt to produce a unified remedy plan for ILECs in California. At that time, the CLECs were proposing an "equal risk" approach to parity testing. Without going into detail, equal risk is an alternative to forgiveness for dealing with random variation. It involves the selection of a critical value of the test statistic that equates the probability of type I and type II errors so that the expected value of inappropriate penalty payments is zero. In any event, some exploratory work using CA data by Dr. Clark Mount-Campbell had suggested that a Z value of 1.04 would equalize the probabilities of type I and type II error at 0.15 (i.e., $\alpha = \beta = 0.15$). Thus the CLECs were proposing that all parity tests be conducted at an $\alpha = 0.15$ level of significance. PacBell, ignoring the equal risk aspects of the testing procedure, insisted that each submeasure would fail about two tests each year due to random variation. (Presumably, PacBell arrived at this figure by noting that 12 months x 0.15 probability of a type I error

= 1.8, or approximately 2, tests expected to fail each year due to random variation.) Thus PacBell demanded one forgiveness per sub measure every six months to compensate them for random variation. At the same time, PacBell argued that the appropriate significance level should be α=0.05 (or Z_{Crit}=1.645 rather than 1.04), implying as shown above, about one forgiveness per submeasure every 18 months. (As an interesting aside, the CLECs, mistakenly viewing forgivenesses as a bargaining chip and also ignoring the equal risk aspects of the testing procedure, had pretty much agreed to grant PacBell one forgiveness per submeasure every six months if PacBell would agree to test at the α = 0.15 level.) To make a long story short, no unified plan (at least in terms of critical values and remedy levels) came out of that workshop. And remedy plan issues remain in litigation before the PUC. Subsequent to the initial CA workshop discusions, Bell Atlantic-New York was granted 271 approval by the FCC. In approving the BANY PAP, the FCC noted the appropriatenesss of a one-tailed parity test undertaken at the α = 0.05 level of significance (Z_{Crit}=1.645). As result, most subsequent PAPs (Pennsylvania and Texas) have adopted a 1.645 critical value for judging parity. Massachusetts copied New York and is using in addition to a 1.645 critical value a repeated 0.8225 critical value as a component in scoring whether parity performance has been achieved.

While the above anecdote is only one instance of an ILEC's tendency to inflate the number of forgivenesses, it is symptomatic of a general propensity. A number of states served by Southwestern Bell Telephone Company (SWBT) are currently considering a PAP modeled after their Texas plan. The Texas plan determines the number of forgivenesses from a "K-Table," which consists of a set of test numbers and corresponding forgiveness (and critical Z) values. The table basically says to the reader, "You tell me how many tests you are going to conduct, and I will tell you how many parity violations must be forgiven to correct for random variation (and the appropriate Z_{Crit} value to use in the tests)." The number of forgivenesses is called "K" in the table, hence the name. As will be shown later, this table overstates the statistically appropriate number of forgivenesses justified to correct for random variation by a factor of twenty to one hundred percent, depending on the number of tests undertaken. Thus, when forgivenesses are used to correct for potential problems arising from random variation, there is a clear tendency for ILECs to overstate the justified number.

In concluding this overview, it is important to note that many view forgivenesses, whether justified by random variation or not, as THEFT! While this is a harsh view, it is, to many CLECs, appropriate. In their view, forgivenesses allow ILECs to violate the law, by providing CLECs with discriminatory service levels, without being penalized. Three tenets form the basis for this view.

(i). Computing the extent of random variation and the appropriate number of forgivenesses according to the ILEC approach outlined above requires the assumption that the ILEC always provides parity service. Many CLECs find this assumption ludicrous. They point out that if it were true, there would be no need for parity testing, and with no statistical testing, there would be no random variation in the test statistic, and hence no need for forgivenesses. The most fundamental rationale for performance appraisal and parity testing is that the ILEC has an incentive to maintain its monopolistic position in the local market and will do so by providing inferior service levels to competing CLECs unless its service provision performance is carefully monitored. Thus the mere fact that we are trying to put together a PAP gives lie to the assumption that the

ILEC always provides parity service

It can also be argued that the number of forgivenesses justified if this assumption were true would be an overstatement of the appropriate number of forgivenesses, given that is not true. Thus a corrected number of forgivenesses could be obtained by weighting the original number of forgivenesses by the probability that the ILEC provided parity in its supply of every submeasure. But even in this case, many CLECs would argue that a false sense of propriety has been given to an essentially worthless idea -- nothing is to be gained by placing any credence in a procedure based on such an unrealistic hypothetical.

(ii) Random variation and its associated forgivenesses ignore the possibility of type II error. Recall that when someone bases their conclusions on a statistical test, they can make two types of errors. They could conclude parity is not present when in truth it is, a type I error; or they could conclude parity is present when in fact it is not, a type II error. As explained above, ILEC random variation arguments exploit the former type of error but ignore the latter. Clearly, when a type II error occurs -- the ILEC is judged in parity when in fact it is discriminating against the CLEC -- the ILEC avoids paying a penalty it should pay. *In fairness, if the CLEC owes the ILEC a forgiveness when the ILEC is asked to pay a penalty it should not have to pay due to type I error, then the ILEC owes the CLEC a "forgiveness" if it avoids paying a penalty it should pay due to a type II error.* The problem is that determining how many forgivenesses of the second type the ILEC owes the CLEC requires the computation of the probability of a type II error. This computation requires, in turn, knowledge of the extent to which parity was violated (so as to locate the distribution of sample means differences under the alternative hypothesis). Since this information is not generally available to the analyst, this latter computation, and the implied forgivenesses associated with it, is typically ignored.

There are, however, several ways to take type II errors, as well as random variation, into account in performance appraisal questions. One method is an "equal risk" approach, as developed in current PAPs of AT&T and BST. As this approach has already been outlined, an example will serve to illustrate the point. It turns out that a delta value of 0.1 and a CLEC sample size of about 400 will produce a balancing critical value of $Z_{Crit}=1.04$ which equates the probability of making a type I error (α) with the probability of making a type II error (β) at a value of 0.15. Now suppose we conduct 100 tests this month. Under these conditions, the ILEC would be judged to owe penalties on 15 submeasures that it should not have to pay (due to type I error), but it would also avoid paying penalties on 15 submeasures that it should have to pay (due to type II error). In the end, fifteen penalties, plus those for any other submeasures found out of parity, are owed, and fifteen penalties, plus those for any other submeasures found out of parity, are paid. The errors cancel each other out and there is no mistake in penalty assessment.

There is no doubt that such an equal risk approach has a certain appeal for parity testing and performance appraisal. An obvious advantage is that it obviates the need to treat forgivenesses and K-Tables. Unfortunately, operationalizing the approach encounters some serious, perhaps fatal, problems relating to the appropriate value to assign to a crucial parameter called "delta". If these problems can be solved, then equal risk becomes a very attractive approach.

On the other hand, if the problems cannot be solved, we are stuck with having to deal with forgivenesses and K-tables. In this vein, Dr. George Ford, of Z-Tel, has suggested a method for determining the number of forgivenesses the ILEC would owe to the CLEC due to type II error. Dr. Ford has attempted to modify the Texas Plan so as to eliminate some of its more glaring errors. When considering problems arising from forgivenesses, he noted that the K-Table used in the Texas plan to determine the appropriate number of forgivenesses was constructed assuming that the ILEC was always in parity and thus considered only type I errors. Making a reasonable assumption concerning the extent to which the ILEC might diverge from parity, Dr. Ford constructed an "Inverse K-Table", that is, one based on type II error where the value of K tells us the number of "forgivenesses" an ILEC would owe a CLEC for not paying penalties it should have paid, but avoided, due to type II error. Based on his assumptions, Dr. Ford found that for any reasonable number of tests, the number of "forgivenesses" arising from type II errors dwarf the numbers in the traditional K-Table, i.e., those arising from type I errors. Now, clearly, we could change Dr. Ford's assumptions about the extent of the ILEC's divergence from parity and find different numbers for type II forgivenesses. But the lesson he provides us is clear: for reasonable departures from parity, it is likely that the probability of type II errors exceed the probability of type I errors, so from a forgiveness perspective, the ILEC probably owes the CLEC, rather than conversely. Now, nobody truly expects the ILEC to pay more due to type II random variation. Ford's point is that no undue harm is likely to accrue to the ILEC if we drop the notion of random variation and forgiveness altogether. Most CLECs agree with this position.

(iii). Finally, if one wishes to fully understand why some CLECs view forgivenesses as theft, it is important to understand that there are two alternative, and arguably, equally legitimate views of what constitutes "parity in service provision". One view, which we shall call "Parity of Process," holds that parity is achieved if the mean (and variance) of the production process that the ILEC uses to supply its own customers is the same as the mean (and variance) of the production process which it uses to supply the CLEC's customers. As will be explained momentarily, in this approach, the test statistic can be thought of as exhibiting sampling variability. Thus, if one ignores the two criticisms above, a case can be made in support of the legitimacy of forgivenesses.

The second view, which we shall call "Parity of Outcome," holds that the service provision data collected on the CLECs and ILEC each month constitute a population, not a sample. In this approach, the test statistic is not a "statistic" at all; rather it is simply a measure of the extent of discrimination that took place that month. According to this view, since the "test statistic" is not subject to random variation, there is no legitimate statistical justification for forgivenesses. Most CLECs subscribe to this latter view to a greater or lesser degree. Clearly, if that view is correct, then granting a forgiveness to the ILEC -- allowing them to discriminate against the CLEC without penalty -- is tantamount to allowing them to steal a part of the CLEC's local market, both actual and potential. Since the distinction between the two views of parity is fundamental to understanding the CLECs' perspective on forgivenesses, we now turn to a more detailed examination of each.

III. Parity of Process Versus Parity of Outcome

Most PAPs use (some variant of) the LCUG Modified Z statistic as the *deus ex machina* for evaluating the extent of discrimination in service quality provision. The formula for the basic statistic is

$$Z = \frac{\overline{X}_{CLEC} - \overline{X}_{ILEC}}{s\sqrt{\frac{1}{n_{CLEC}} + \frac{1}{n_{ILEC}}}} \qquad (1)$$

where the $\overline{X_j}'s$ are the means and the n_j's are the number of data elements collected on the service for the CLEC and the ILEC, respectively. σ is standard deviation, of the ILEC data if the LCUG approach is used or of the pooled data otherwise. Once this statistic is computed, its value is compared to a critical value to determine whether the deviation from parity is large enough to indicate the presence of discriminatory service provision. Both views of parity conform to this general framework; they differ in their view of the nature of the data used to compute the statistic and the consequent implications on the stochastic nature of the statistic.

The Parity of Process view takes the data to be realizations of a sample from an infinite population. That is, the production process that the ILEC used to supply its own customers last month could have generated an infinity of possible outcomes, as could the production process that the ILEC used to supply the CLECs' customers. The data on these processes can then be thought as simply the outcomes of the processes observed last month. They are therefore samples of all of the observations that could possibly have arisen from each of the respective processes. Their means and variances ($\overline{X}$ and S^2, respectively) of the true measures of location and dispersion (μ and σ^2, respectively) of their corresponding production processes. Note that these production processes could have produced infinitely many other samples, each having a different mean (and variance). Thus both sample means, while certainly estimates of their corresponding population parameters, are themselves random variables that follow statistical distributions. According to the Central Limit theorem, for large samples, the sample mean follows a normal distribution with mean given by the population mean and variance given the population variance divided by the sample size. It is further known that if we create another random variable by taking the difference in the means of the two samples, it will also follow a normal distribution, with mean equal to the difference in the population means and variance given by the sum of the population variances divided by their respective sample sizes. This random variable can be converted to a *standard* normal random variable, i.e., one having zero mean and unit variance, by subtracting out its mean and dividing through by its standard deviation (the square root of its variance). More formally

$$\overline{X}_{CLEC} \rightarrow N\left(m_{CLEC}, \frac{s^2_{CLEC}}{n_{CLEC}}\right) \quad and \quad \overline{X}_{ILEC} \rightarrow N\left(m_{ILEC}, \frac{s^2_{ILEC}}{n_{ILEC}}\right), so$$

$$\overline{X}_{CLEC} - \overline{X}_{ILEC} \rightarrow N\left(m_{CLEC} - m_{ILEC}, \frac{s^2_{CLEC}}{n_{CLEC}} + \frac{s^2_{ILEC}}{n_{ILEC}}\right), and\ hence \qquad (2)$$

$$Z = \frac{(\overline{X}_{CLEC} - \overline{X}_{ILEC}) - (m_{CLEC} - m_{ILEC})}{\sqrt{\frac{s^2_{CLEC}}{n_{CLEC}} + \frac{s^2_{ILEC}}{n_{ILEC}}}} \rightarrow N(0,1)$$

To conduct any statistical test, the test statistic is always computed assuming the null hypothesis is true. For parity testing, the null hypothesis is equality of distribution, that is equality of means and variances, so that $H_0 : \mu_{CLEC} - \mu_{ILEC} = 0$ and $\sigma^2_{CLEC} = \sigma^2_{ILEC}$. Substituting these restrictions into the Z statistic of equations (2) will reproduce the appropriate test statistic of equation (1). It follows that the statistical properties of a parity test are inherited from the statistical properties of its components (means and variances), that are in turn inherited from what we assume about the properties of the data that make them up. Different assumptions about the data will lead to different implications as to the nature of the test statistic, as will soon be shown.

Parity of Process therefore is based on a test statistic derived from a standard normally distributed random variable. This result allows us to easily compute the extent of random variation and, ignoring type II error, provides us with a statistical justification for forgivenesses. For instance, the fact that Z follows a standard normal distribution indicates that there is only a 5% probability of computing a value of it in excess of 1.645 by chance. Now suppose we are analyzing data on order completion interval, or any other service for which larger values indicate worse service, and undertake the parity test at the .05 level of significance. Suppose further that we obtain a value of the test statistic in excess of 1.645, so that we conclude discrimination against the CLEC. There is only a 95% chance, in general, that this is a correct decision. There is a 5% chance that we got a statistic value this large because one of the means came from a sample taken from an extreme or uncharacteristic part of its production process. That is, there is a 5% chance that the processes are actually in parity even though our statistical results suggest otherwise. In this case, according to the parity of process view, the ILEC would be forced to pay a fine when it was in fact providing parity service. The ILEC thus argues that such a "violation" should be forgiven since it is not actually a violation at all. To reiterate, if all tests are undertaken at the 5% level of significance, there is a 5% chance of this error occurring for each test. Thus, if we conducted one hundred tests per month, on average, we would expect five of the resulting outcomes to exhibit this type I error, and hence, so the story goes, we should forgive five violations on the part of the ILEC.

Now let us contrast this view with a Parity of Outcome approach. This approach does not view the data to be analyzed as realizations of outcomes of the output of some unspecified production process. The Outcomes approach does not view the data as a sample at all, but rather as a population. Whether more or different data might have been generated from the process is both esoteric and immaterial; *what we have is **all** of the data on the various service quality measures that were generated that month.* Thus when we compute the means and variances of these data series, we are not estimating the mean and variance of some underlying production process, we are literally computing the parameters of the respective populations. It follows that if the CLEC mean is computed to be larger than the ILEC mean, we already know what we were testing to find out in the Process approach, that $\mu_{CLEC} > \mu_{ILEC}$. This does not mean that the computation of equation (1) is not important from the Outcomes view. But in this view, it is a measure of materiality, not a test statistic. It allows us to address the question of whether the existing means difference if big enough to have an important effect on competition. If we compare it to some critical value to make that decision, and if that critical value happens to be 1.645, so be it. It probably makes more sense to use a statistically determined value to demarcate materiality than a mere guess at the actual means difference that would be marginally competitively significant.

Thus, even though the two approaches are superficially similar, they are fundamentally different. This difference is no more pronounced than in the determination of forgivenesses. For statistical legitimacy, forgivenesses require random variation, specifically, type I error. But in the Parity of Outcomes approach the data constitute populations, not samples, so that "statistics" computed from random variables based on them do not exhibit sampling variability. Thus *there can be no type I error, no random* *variation, and consequently, no justification for forgivenesses.*

The Parity of Outcomes approach is rather extreme and not very many CLECs subscribe to it. However, several CLECs do subscribe to a hybrid of the two approaches which relies on the outcomes view heavily enough to refute the rationale for forgivenesses. This view follows the Parity of Process approach up to the computed value of the test statistic exceeds the critical value, then it adopts (a variant of) the parity of process approach. The argument goes like this: When the ILEC fails a parity test, it has provided the CLEC with inferior service -- type I error or no type I error. They can only fail the test if the computed Z is larger than the critical Z. But this can occur only if the CLEC's mean exceeds the ILEC mean, i.e., only if the CLEC has been given inferior service. Of course, there may be a 5% probability that this outcome was due to chance. But all this suggests is that the ILEC did not discriminate against the CLEC on purpose; that is, they did not employ a discriminatory process, they simply achieved an extreme or uncharacteristic result from an equivalent process. Nevertheless the fact remains that the CLEC received inferior service. CLECs that support this view find no provision in the Telecommunication Act of 1996 that the ILEC be excused from providing parity service simply because it did not intend to discriminate. What they do find is that the law requires service to be of at least equal quality to that which it provides its own customers. When an ILEC fails a parity test, it has not met this requirement.

This section has tried to provide a CLEC perspective on legitimate reasons why parity testing does necessarily require the granting of forgiveness. In fact it should now be clear that the only statistical foundation justifying forgivenesses is a pure Parity of Process view, and even this view ignores mitigation due to type II error. However, given that almost every PAP that does not advocate equal risk requires forgivenesses in one form or another, many CLECs are developing the following philosophy: If forgivenesses must be granted, at least make an effort to grant no more than are justified. The implicit question here leads us directly to the next section.

IV. What is the Appropriate Number of Forgivenesses?

Most of this paper up to now has suggested that the obvious answer to this question is zero, at least from a CLEC perspective. On the other hand, as we noted earlier, ILECs tend to overstate, or simply provide no justification for, their forgiveness demands. It is therefore important to have some accurate analysis based on statistical principles as to the appropriate answer to this question. Since a pure Parity of Process view is necessary for the legitimacy of the granting of any forgivenesses, we assume that it is correct in what follows. We do not, however, advocate it as the correct approach.

Let us consider the following experiment. Suppose we conduct many, say N, parity tests, each at the α level of significance. The outcome of each test can be classified into one of two possible categories: Pass (a failure) or Fail (a success). The probability of failing a test by chance is thus α, so that P (success) = α. Finally, the outcome of each test is independent of that of every other test. Under these assumptions, the number of failed tests is a random variable (call it K), known as a Bernouli variable. As such, it is known to follow a binomial distribution with parameters N and p. N is known as the number of Bernouli trials, the number of tests in this case, and p is the probability of success for any trial, which equals α in this case. Notationally, it is said that

$$K \sim b(N,p) \qquad (3)$$

and the probability distribution function of K is thus

$$P(K \le k) = \sum_{K=1}^{k} \frac{N!}{k!(N-k)!} \quad p^k (1-p)^{N-k} \tag{4}$$

While technical, this information is important because it allows us to compute the probability that we will fail a certain number of tests by chance. For example, suppose we conduct 100 parity tests at the α = .05 level of significance, i.e., N = 100 and p = .05. Now if we wish to know the probability of failing exactly five tests by chance, we have

$$P(K=5) = \frac{100!}{5!95!} p^5 (1-p)^{95} = 0.18 \tag{5}$$

or if we wish to know the probability of failing fewer than, say, four tests a by chance

$$P(K<4) = \sum_{K=1}^{3} \frac{100!}{K!(100-K)!} p^K (1-p)^{(100-K)} = 0.118 \tag{6}$$

Figure 1 and Table 1, below it, (next page) show and tabulate, respectively, the probability distribution of K under these assumptions. It is worth noting that the probability of failing more than ten out of the 100 tests is only about 1.1%.

The mean of any random variable is its expected value; that is, the sum of the values that the random variable can take and times the probability of those outcomes. A Bernouli random variable is typically viewed as taking on a value of zero for a failure and one for a success. Thus the expected value of a Bernouli random variable consists of the sum of N identical elements of the form 0 . (1 - p) + 1 (p). It follows that

$$E[K] = Np \tag{7}$$

Likewise, it can be shown that the variance of K is

$$V[K] = Np(1-p) \tag{8}$$

In the above example with 100 tests, each taken at the 5% level of significance, N = 100, p = .05, therefore the expected (mean or average) number of misses is 5 (= 100x .05). and the variance is 0.475 [5 x (.95)].

Finally note that as the number of trials (N) gets large, the binomial distribution approaches the normal. Thus for large N,

$$K \sim N[Np, Np(1-p)] \quad (9)$$

How large does N need to be before the normal approximation can be used? An often suggested rule of thumb is that the normal approximation is a good one so long of the smaller of the two numbers given by N p and N (1 -p) is greater than or equal to 5. Figure 1 illustrates. Since N = 100 and p = .05, N p = 5, so the normal approximation should be acceptable. From Figure 1 we can see that the mean of K, 5, is also equal to

Figure 1
The Binomial Probability Distribution for N=100 and p=.05
(The vertical axis graphs the probability that K=k and the horizontal axis graphs the categories of K. Category 1 corresponds to K=0,category 2 corresponds to K=1, … , category 11 corresponds to K=10)

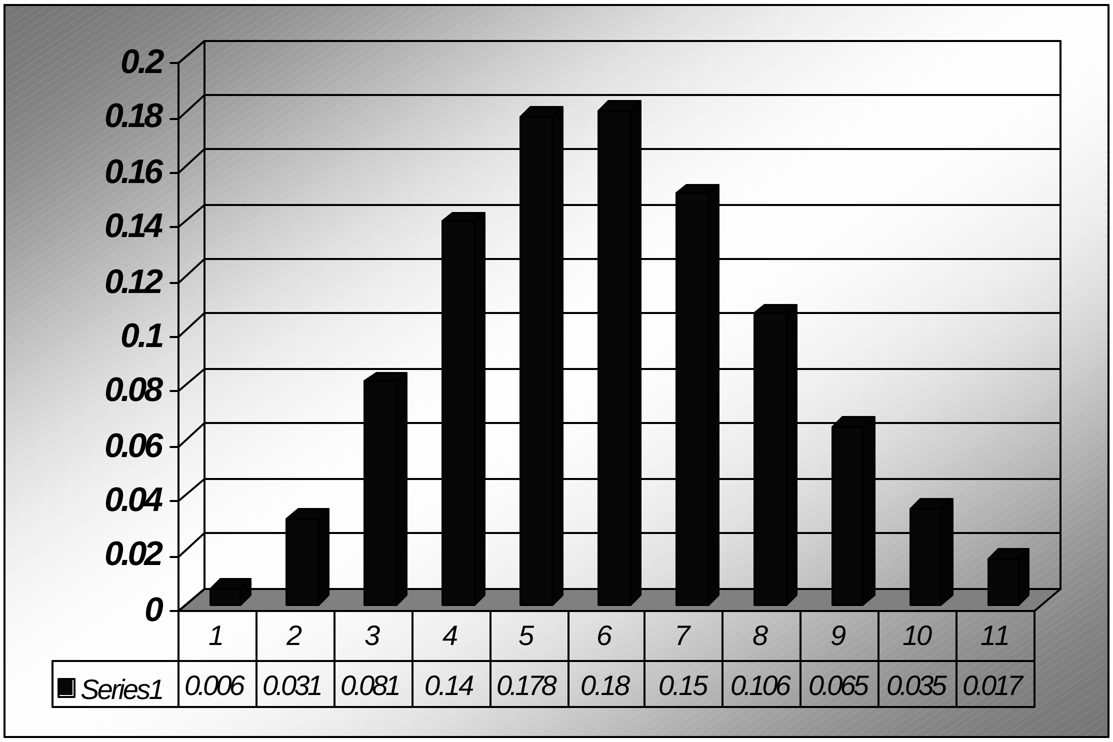

Table 1
The Data corresponding to Figure 1

K	P(K)
0	.006
1	.031
2	.081
3	.140
4	.178
5	.180
6	.150
7	.106
8	.065
9	.035
10	.017

the mode (the most likely value in this case 0.18%) of K, and hence also equal to the median (middle value) of K. Since the mean, median, and mode of K are all equal, the distribution of K is essentially symmetric. Figure 1 also bears out the familiar bell curve shape of the normal.

It is worth nothing that for smaller N, the binomial is skewed to the right so that the mode < median < mean. In this case we are more likely to observe K values smaller than the mean than ones larger than the mean.

All of these technical details are important foundations that must be laid in order of justify the following **key** proposition: *If forgivenesses must be granted, the (maximum) number appropriate to grant is equal to the expected (mean or average) number of chance test failures in N trials (or tests) under taken.* This is the natural measure that we have employed in earlier sections of this paper, and now we see that it has a sound statistical foundation. To be clear, the appropriate number of forgivenesses to grant is E [K] which is computed as N, the number of tests, times p, the level of significance of each test (which we have also called α above). Because it is the mean of the distribution of K, it is a statistically unbiased measure of the number of failures. This means that, in the absence of any further information, it is our best guess at the actual number of test failures, assuming the ILEC always provides parity service. Of course, since K is a random variable, we might on occasion observe more than Np failures, and on other occasions, we might observe fewer. But over time, with many parity tests undertaken each month, the number of failures will average out to Np. This generalization is especially true for large N, where the distribution of K is symmetric, because in this case it is clear that the probability of observing a number of failures greater than Np is exactly equal to observing a number of failures less than Np.

When N is smaller, we are more likely to observe a number of failed tests smaller than the mean (since the mode of the distribution is less than N p). This is one reason why we suggest that the maximum number of forgivenesses: Over time we would be likely to observe fewer failures than the mean value -- at least in the small N case. We do not belabor this point, however, since most PAP's envision monthly parity testing for a large number of submeasures. We conclude that since a large number of parity tests is the norm, symmetry of the distribution of K should be expected. Thus, over time, parity testing should cause the number of tests failed due to random variation to converge to Np tests.

There is, however, one point to be made that suggests that granting Np forgivenesses to the ILEC every month may be - - even on average - - granting too many. When we suggested that we could expect Np failures each month due to random variation, we based their result on the assumption that the ILEC always provided parity service. In other words, the conditional expectation of K, the expected number of failures given the ILEC is always in parity, is Np. It follows that the relevant, or unconditional expectation, of K is Np times the probability that the ILEC is always in parity. A crude measure of this probability is given by

$$P(ILEC\ always\ provides\ parity\ service) = 1 - \frac{number\ of\ \ failed\ tests}{total\ number\ of\ tests} \qquad (10)$$

Thus, we suggest the following modification to the earlier rule. *The appropriate number of forgivenesses to grant the ILEC in any given month is F, where*

$$F = [\frac{number\ of\ \ passed\ tests}{total\ number\ of\ tests}]xNp \qquad (11)$$

To illustrate, we continue with the N = 100 and p = .05 example. That is, we conduct 100 independent parity tests at the α= .05 level of significance. Suppose 20 of those tests fail. Originally, we would have suggested that Np = 5 test failures should be forgiven, so that only 15 failures should be penalized that month. However, we now note that there is not a 100% probability that the ILEC provides parity service for each and every submeasure. A hueristic estimate of the probability that the ILEC provides parity service for any one submeasure is 0.8 (80, the number of tests passed, divided by 100, the total number of tests undertaken). Thus we suggest the ILEC be granted only 4 forgivenesses (0.8 x 5) and that it be penalized for 16 violations if the desire is to grant the statistically appropriate number of forgivenesses.

V. K-Tables and Forgivenesses

A number of ILEC PAP's, mostly in states serviced by SBT, use a K - Table to determine the number of forgivenesses. From our earlier discussion, it may be recalled that a K-Table consists of a set of test numbers and corresponding forgiveness (and critical Z) values. The table basically says to the reader, "You tell me how many tests you are going to conduct, and I will tell you how many parity violations must be forgiven to correct for random variation (and the appropriate Z_{Crit} value to use in the tests)." The number of forgivenesses is called "K" in the table, hence the name. In what follows, we will review the history of the K-Table and discuss how one is calculated. We will then argue that using the K-Table to determine the number of forgivenesses to be granted to the ILEC in a given month is a dramatic overstatement of the amount that they legitimately merit.

Early on (pre-1998) in CLEC/ILEC/state regulatory commission discussions of 251/271 compliance verification, AT&T, with most CLECs' approval, had proposed a three tiered penalty structure: Tier I related to the ILEC providing parity service to the individual CLECS (one by one). Tier II related to the ILEC providing parity service at the industry level, i.e., to all CLECs taken together. Tier III related to service or persistent ILEC violations at the industry level, penalties for which would be paid to the state (a persistent violation is one which occurs for three consecutive months). Tier I thus considered individual tests on individual submeasures for individual CLECs, but Tiers II and III required the consideration of the industry as a whole. Therefore these upper tiers required the aggregation of the results of many tests. In particular, the question arose "How many tests would the ILEC have to fail before we are (95%) sure that their failure to provide parity service is not attributable to chance?" The first K - Tables were early attempts to answer this question. Similarly, the paper submitted by then separate MCI and WorldCom entities in TX contained Dr. Mallow's K table for use in determining 251/271 compliance, not for determining if any remedies should be paid to CLECs when inferior service is received.

While the LCUG literature produced prior to 1998 may contain K-Tables, the first K-Table to be produced in written testimony was provided by Dr. Colin Mallows of AT&T in a document presented to the FCC dated May 29,1998. We refer the reader particularly to pages 18-21 of this document and the attached Exhibit 1. Dr Mallows begins by noting that, in reviewing aggregate results of ILEC's performance, if all tests have "…a Type I error rate of 5%, then we would expect, on average, 5% of these tests to indicate non-compliance even when the ILEC is in full compliance." He further notes that this number is a random variable so, "We need to derive some threshold number of parity tests such that if more than this number are observed to fail, then non-compliance can be deduced." Thus we have his announced purpose for creating the K-Table.

The object of the K-Table is to determine the number of individual violations (K) and the type I error of the individual tests (α) so that the probability of falsely claiming a violation of 251/271 requirements is set at 5%. Assuming that the ILEC is fully in compliance and that we know N, the number of tests to be aggregated, Dr. Mallows suggested the following procedure for setting up a K-Table: (i) Choose a tentative value for α, say α=0.05. (ii) Determine K to be the largest number such that the probability that the overall set of tests violate parity is no greater than .05. (iii) Decrease the value of α until the overall probability of a violation using the K determined in (ii) is exactly .05. The resulting values of α and the implied $Z_{crit,}$ which will be read from the table, determine the values to be used in the individual tests. The corresponding number K, also read from the table tells us the maximum number of tests that can be failed under these conditions such that any additional failures will render us (95%) certain that parity is not being provided at the industry level.

Before providing an example, it is worth noting that that Dr. Mallows proposed the following formula for finding K in step (ii):

$$P(K < k) = 1 - [(1 - \alpha^3)^N * b(k, N, \alpha)]$$

where the first term in brackets is the probability of three consecutive misses, the persistent failures component. The cognoscenti typically ignore this term either because their plan contains no persistent failures component or because the resulting number is so close to unity (for the N=100, α=.05 case, the term is equal to 0.988). The second term in brackets is the probability from the binomial distribution of finding k or fewer successes in N trials when the probability of success is α, which we discussed earlier. (Again Dr. Mallows suggested an adjustment to α relating to the persistence component, which is almost universally ignored in subsequent work because it is so small.) Thus, if we are simply concerned with finding the maximum number of failed tests before lack of parity is assured with 95% confidence -- without regard to persistence -- we simply make use of the binomial distribution. For a given N and trial p we find the largest k such that the probability that the number of failures is less than or equal to k is at most 0.95. Holding this k constant, we reduce p until that overall probability is exactly 0.95. This consequent value of p defines the level of significance, and hence the critical Z value, at which all N individual tests should be undertaken.

A simple illustration using EXCEL may help clarify the procedure. Suppose we wish to conduct 100 tests, and we begin by assuming a p (=α) of 0.05. Using the statistical function CRITBINOM, we set TRIALS=100, PROBABILITY=.05, and ALPHA=.95. The function returns the smallest value of k for which the cumulative binomial probability is greater than ALPHA -- 9 in this case. However, we wish the largest value of k for which the cumulative binomial probability is just less than ALPHA. Thus our desired value of k is the number the function returns minus one -- 8 in this case. Next we use the BINOMDIST statistical function with NUMBER=8, TRIALS=100, PROBABILITY=.05, and CUMULATIVE=true. We then nudge the PROBABILITY entry downward slightly and continue to do so until the function returns exactly .95 -- roughly .048 in this case. Finally, this probability if entered into the NORMINV function with MEAN = 0 and STANDARD DEVIATION = 1 to find the critical Z value at which the 100 tests should be conducted -- 1.67 in this case. A K-Table simply repeats this exercise for various numbers of trials (or tests, N) and tabulates the results.

A further illustration is provided by Dr. George Ford in his paper on "The Modified Texas Plan", page 13. There he reproduces and expands the Texas K-Table. It turns out that it is an exact replica of the one in Dr. Colin Mallows testimony referenced earlier. As such it, presumably unknowingly, corrects for persistence when no correction is justified. Dr Ford recomputes the table without the persistence factor and presents the corrected table on page 13 as well. For our purposes, either table will do (although Ford's corrected table was computed exactly as outlined above). *According to the Texas Plan, one determines the number of tests to be conducted, goes to the K-Table, and finds the corresponding entries for K and Z. The K entry indicates the number of tests the ILEC is allowed to fail before it owes a penalty; the Z entry gives the critical value at which each test must be conducted. It is our contention that this procedure forgives the ILEC far too many failed tests and is therefore unfair to the CLECs.*

As shown above, the value for K from the table tells us the maximum number of tests the ILEC can fail before we are 95% sure that the ILEC is out of parity for the industry for that month. This is exactly what Dr. Mallows designed the Table for and it is exactly what the Table is supposed to tell us. It is also correct that this means that there is a 5% probability of type I error for the testing process that month. That is, for say, the N=100 and p=.05 case, if every test were undertaken at the .048 level, there is a 5% chance that if we observed more than 8 violations that month, that the ILEC would still be in parity. Up to this point everything is fine.

The problem arises because somebody on the Texas Staff or at SBT decided that (for N=100, p =.05, say) because 8 tests must be failed before the ILEC is judged out of parity, the ILEC should be forgiven those 8 failures. This is a *non sequitur*; there is no logical connection between the information in the K-Table and the appropriate number of forgivenesses. What is so amazing is that people were so unfamiliar with the notion of a K-Table and what it was designed to do that they are only just now realizing the fallacy. One way to see the problem is to note that if we, as is typical, equate random variation with type I error, then we should only forgive those errors in excess of 8 because they are the ones that would arise due to type I error. This is clearly incorrect, but it follows the logic of using the K table for forgivenesses.

The problem with the K-Table reasoning is that it ignores the fact that, under the assumptions used to generate it, ***all misses are due to random variation***. Figure 1 of section IV may prove helpful here. It shows that there is about a 6% chance of failing more than 8 tests due to random variation. But it also shows that there is a 38% chance of failing more than 5 tests due to random variation, a 44% chance of failing fewer than 5 tests due to random variation, an 18% chance of failing exactly 5 tests due to random variation, etc. *The point is that when we assume the ILEC always provides parity service, any observed test failure must be due to random variation. Thus if we wish to estimate the actual number of failures arising due solely to random variation, we should not be asking, "What is the maximum number of test failures that could occur before we would be 95% sure that the next failure was not due to random variation (the K-Table question)?" Rather, what we should be asking is, "How many test failures due to random variation would we expect if we conducted 100 tests, each at the 5% level, month after month, after month (the expected value question)?" As we showed in section IV, the* *answer to this question is the expected value of the binomial random variable K. Under the above assumptions, we would expect, over time, on average, 5 tests to fail each month, not 8. Thus forgiving 8 violations instead of five, forgives the ILEC three failures with no statistical justification. Certainly, granting these three additional forgivenesses cannot be justified on the basis of the expected failures due to random variation -- as we have shown above.*

For these reasons, it seems clear to the CLECs that the number of failed tests forgiven the ILEC should be based on the expected value of K = Np, not on the K-Table. Without doubt, more than Np tests will fail due to random variation in some months. But equally, fewer than Np tests will fail due to random variation in others. Statistical theory guarantees us that over time the number of test failures due to random variation will converge to Np and not some number from a K-Table. However CLECs believe that even Np is too many forgivenesses. Recall that Np is the conditional expectation of K (conditioned on the assumption that the ILEC is always in parity). CLECs believe that the more appropriate is the unconditional expectation of K, i.e., Np weighted by the probability that the ILEC passes all of the tests. Since this probability is less than one, this view must imply fewer legitimate forgivenesses. CLECs hasten to add that even this adjusted measure of forgivenesses ignores type II error. Since this probability is non zero, it suggests even further reduction in the number of test failures that can legitimately be granted an ILEC.

VI. Conclusions

This paper presents a CLEC perspective on random variation, forgivenesses, and their manifestation in many PAPs, K-tables. The analysis begins by explaining the ILECs rationale for requesting forgiveness (i.e., being forgiven a fine) for failing parity tests due to sampling variability in the random variable underlying the parity test statistic. We then explain the CLEC view that granting such requests constitutes theft of the CLECs' actual and potential local market. Three tenets support this view: (i)The rationale for forgivenesses is based on an unrealistic hypothetical -- that the ILECs always provide parity service. (ii) Forgiveness arguments and rationales ignore type II error -- if it were taken into account, it would likely more than offset the extent of type I error that serves as the statistical justification for forgivenesses. (iii) Finally it is noted that only an extreme version of one of two alternative views of the parity testing scenario statistically justify the granting of forgivenesses. Next a detailed examination of the two alternative views is offered. It is shown that a pure "Parity of Process" view is the only approach to parity testing that offers ILECs some hope of statistical legitimacy for forgivenesses (and, then only if type II error is ignored). A "Parity of Outcomes" view does not admit to random variation so that forgivenesses have no statistical justification. Even a hybrid of the two views refutes the appropriateness of forgivenesses.

The remainder of the paper assumes that the pure Parity of Process approach has been judged acceptable (a major problem in itself from a CLEC perspective) and asks, "What is the correct number of forgivenesses that should be granted to the ILEC?" We argue that the answer to this question is the expected number of type I errors, which is given by the number of tests undertaken times the level of significance of the tests. This is the appropriate value because it is the value that the number of type I errors would tend toward for a large number of tests conducted month after month. In fact, to be more accurate, this number should be weighted by some measure of the probability that the ILEC is providing full parity service. In addition, many ILEC PAPs, particularly those affected by the "Texas Plan", demand that the number of forgivenesses be given by a "K-Table". We examined the history of the K-Table and its evolution via the Texas plan. We then showed that K-Tables demand considerably more forgivenesses than are justified by sound statistical theory. This result implies that if forgivenesses are to be based on sound statistical principles, they should be calculated as the expected value of a binomial random variable, not drawn from some K-Table.

We conclude by offering the CLEC perspective on random variation, forgivenesses, and K-Tables. In summary, we suggest that there is at best only a limited and uncertain rationale for forgivenesses; the idea should be scrapped. Should some forgivenesses be granted as state policy, at least grant only the statistically justified number. This requires doing away with the K-Table as a calculator of forgivenesses..