

## **Impasse Issue on Benchmark for PO-19 Stand-Alone Test Environment (SATE) Accuracy**

### **Overview**

With Qwest's decision to offer a Stand-Alone Test Environment (SATE), the CLECs proposed a PO-19 performance measure to evaluate Qwest's ability to provide accurate production-like tests to CLECs for testing both new releases and between releases in the SATE environment. After extensive discussions, the parties reached agreement on the performance indicator definition (PID) for the new measure. These negotiations took place in the August to December 2001 timeframe so the resulting PID was not part of the 4.0 PID distributed on October 25<sup>th</sup>. Because of that, it is inserted here for easy reference.



"PO-19 SATE New  
PID 05Oct01DRAFT.d

Although the TAG has reached collaborative agreement on the PID description, it has been unable to reach agreement on the standard to be used for the PID. The parties agree that a benchmark is the proper type of standard but disagree on the benchmark level.

Qwest originally proposed a benchmark in the low 90's that would become effective in March 2002. The CLECs proposed a standard of 98%. In further discussions, Qwest offered 92% and then lastly 95% in March 2002 and 96.5% for June 2002 and beyond. The CLECs' proposal remains at 98% beginning in March 2002.

This issue is now at impasse and has been referred to the Steering Committee for resolution. This is in line with the recent resolution by the Executive Committee of the jurisdictional issue concerning administration of non-test impacting PIDs during the duration of the ROC OSS test.

Below are statements prepared by Qwest and AT&T/CLECs, outlining their positions and rationales on this issue, for consideration by the ROC Steering Committee in resolution of this impasse.

### **Process**

The positions should be submitted to the ROC Project Manager, Denise Anderson of MTG, by Noon MT on Friday, January 25<sup>th</sup>. MTG will consolidate and distribute the position document to the TAG on January 25<sup>th</sup>. The ROC

Steering Committee will hold a resolution discussion of this topic at its next meeting on Monday, January 28<sup>th</sup>. The TAG will be notified of the ROC Steering Committee's decision and rationale by COB on the following day.

This timeline represents an acceleration of the schedule for PO-19 impasse resolution based on a request from Qwest in consideration of related discussions/actions that are transpiring in Arizona. The Arizona collaborative would like to understand where the ROC benchmark will be set prior to its further consideration of PO-19 which is imminent. The ROC Steering Committee was agreeable to accelerating the schedule by one week.

### **Qwest's Position on the Benchmark for PO-19**

Qwest's original proposal for the PO-19 standard was diagnostic. When the CLECs indicated they would not agree to a diagnostic standard, in the spirit of compromise Qwest offered to establish a benchmark in the low 90's beginning in March 2002, as indicated in the above Overview. Qwest continues to assert that, given the newness and ongoing changes of both the SATE process and the measurement of the process, the appropriate standard continues to be "Diagnostic," subject to some additional evidence and comments below.

Qwest notes that PO-19 is not in any PAP. Nevertheless, provisions in all the various PAPs under consideration would provide PO-19 to be included in the six-month review. This is the clear and appropriate approach for any process and measurement that is so new and dynamic. In the mean time, Qwest continues to work with CLECs in developing and updating the Stand Alone Test Environment. We note that it is in the interest of both Qwest and CLECs for SATE to be useful and successful, with or without a standard applied to PO19.

Qwest bases its positions on several factors as follows:

- Qwest originally proposed this measurement be diagnostic to allow a period of time for the code to stabilize. SATE is a relatively new process that has been in place for only a few months. As such, it is subject to the initial refinement common when any new software is deployed.
- In addition, in response to CLEC requests, Qwest anticipates further development of SATE capabilities over the next several months leading to continued relative instability in the code. Specifically the next several months will bring the following developments in the SATE environment:
  - When Qwest deploys version 9.0V on 1/28/02, we will be installing VICKI and its automated post-order transactions. We will be introducing another layer of complexity to SATE and to the PID test. Version 9.0V will require the testing of post-order scenarios as well as the existing transactions in the Data Document. This increases the number of transactions that Qwest tests, in addition to creating a reliance on a new system that may be modified to meet additional CLEC needs in its first few weeks. We feel

- we have a fairly stable environment now, but that will change drastically after release 9.0V is deployed.
- In addition to the 9.0V release we have several major SATE releases planned in the first half of 2002:
    - SATE 9.0F in February, introduces Flow through and changes much of the data in SATE, thus resulting in significant changes to the Data Document that is used to verify the PID.
    - SATE 9.01 in March, introduces some new systems that SATE needs to interact with, including Facility Check.
    - SATE 9.0F phase 2 in May, includes additional significant data changes.
  - Not only is the SATE test environment a new environment, but the PID measurement is also new. Typically, for new measurements, an initial period is provided during which the PID remains diagnostic to allow for fine-tuning of the measurement reporting process. In this case, however, as noted above, the process being measured is facing changes that the parties agree should be made. Thus, implementing any benchmark initially would not provide a reasonable period for the measurement to reach stability.
  - In previous correspondence, CLECs have indicated Qwest should have a high benchmark because we require that they have a transaction accuracy of 100% before they are certified. Comparing these two “benchmarks” is not an appropriate comparison for the following reasons – basically making it an “apples-to-oranges” comparison:
    - A CLEC passes a given test if their transaction works as desired. The multitudes of data values within a request or response are not verified. PO-19 measures the accuracy of Qwest’s data, including detailed fields on a pre-order response as well as the transactions.
    - The CLECs typically certify to a smaller and more finite set of products. The PID tests all products and transactions contained in the Data Document.
    - The CLECs use an iterative process to certify. That is, if they are not at 100% when they first execute their test, they can correct their transactions and re-execute.
      - They can re-execute their test as many times as necessary in order to certify.
      - They have approximately a 9-month window to gain certification. Qwest runs the PID test once per month and posts the results from that one test.
      - Refinements will continue to be made to the code after the test transactions are completed but before the code is released to production. This by itself, while providing better code to production, could result in deviations from published data and business rules causing PID scenario failures. Thus, setting a high benchmark can adversely impact the responsiveness of the process.
  - Generally, industry interface development projects are iterative in nature. This means that clarifications, questions, and resulting changes are to be

expected. If Qwest were to attempt to meet a high PID benchmark, we could not nimbly make the changes required for an efficient certification process.

- Qwest began reporting results for PO-19 with November data. In November, the results were 94.46% based on 358 out of 379 transactions completing successfully. December results are higher, at 98.73%, based on 390 of 395 transactions, but this does not reflect the increasing complexity of future releases. Thus, recent high results are, by no means, justification for implementing any benchmark at the present time, let alone an overly-high benchmark such as the CLECs propose.

Given these facts, it is clearly premature to establish any standard other than diagnostic at the present time. Going forward, Qwest proposes that the PAP six-month review is the most reasonable point for considering a benchmark. By that time, which could be about nine to twelve months from now, the SATE process should be stable and, following achievement of that stability, the measurement process should also be stable, with a few months of data to demonstrate it.

Nevertheless, if the Steering Committee believes it is necessary to specify a standard now, Qwest recommends that the following schedule for applying a benchmark:

- Through May 2002      Diagnostic (based on above significant changes)
- June 2002 – forward       $\leq$  95 percent

In no case should the benchmark be set above 95 percent. When Qwest previously suggested a compromise with different timing and a somewhat-higher benchmark, it was before the magnitude of the above-described future changes were clear. Now that there is more clarity as to the upcoming SATE changes, as well as to what it will take to remain flexible in responding to CLECs' changing needs in a dynamic environment, it is very clear that the ultimate benchmark should be no higher than 95 percent.

Verizon has an EDI Testing Environment and PID. While Verizon's environment is different than Qwest's, the benchmark for SATE errors is  $\leq$  5% errors, which translates to a 95% success rate. Qwest's ultimate proposal is consistent with this. Moreover, the FCC's acceptance of the Verizon benchmark would indicate their acceptance of the 95% level of performance as sufficient to provide a competitor with a meaningful opportunity to compete.

In conclusion, this is, fundamentally, an issue of timing and of what is ultimately and minimally necessary to provide an efficient CLEC a meaningful opportunity to compete. The best timing, which would permit a decision based on a stable process and measurement, would be to await the PAP six-month review before considering a standard other than diagnostic. Otherwise, based on more recent understanding of the dynamics of the environment, a benchmark clearly should not be established before June 2002 and should ultimately be no higher than 95

percent, which is more than minimally sufficient to provide an efficient CLEC with a meaningful opportunity to compete.

### **AT&T's/CLECs' Position on the Benchmark for PO-19**

The TAG has been working on the PO-19 measurement since mid-August, endeavoring to develop consensus on all aspects of the measure and to guide its implementation. Negotiations have succeeded except for the issue of the standard of performance. CLECs recommend a 98% accuracy standard to be in effect at March, 2002; and Qwest has offered to be measured against a standard of 95% accuracy. No other disputes remain.

PO-19 "Evaluates Qwest's ability to provide accurate production-like tests to CLECs for testing both new releases and between releases in the Stand-Alone Test Environment ("SATE") environment."<sup>1</sup> Qwest has implemented the SATE as the means by which CLECs can test EDI interfaces for pre-ordering and ordering against Qwest IMA and back-end system business rules and interface specifications. Qwest established the SATE because the FCC has required such a testing environment to approve applications under Section 271. Applications made by Verizon point out the strength of the CLEC testing environment made available starting with the New York case. The test environment has been part of each Verizon OSS showing in all subsequent applications. SBC applications, beginning with the Texas case have demonstrated its CLEC testing environment to meet the needs of CLECs with electronic interfaces. A Qwest application without a SATE would have been risky.

The purpose of the SATE is to provide CLECs with a testing environment that is separate from the Qwest production environment, which would avoid production impacts caused by concurrent test processing. By its nature, testing can cause systems to fail, and the production users should not have to suffer the consequences of either successful, or unsuccessful testing. Moreover, the SATE is to provide a mirror of the production environment, such that a CLEC that has processed test orders successfully should expect that those same orders, placed into the production system, would yield like results.

The measurement of the SATE processing is designed to show that Qwest has achieved mirroring in principle, and not mirroring by the literal meaning of an exact copy of the production systems. CLECs and Qwest share the opinion that if Qwest implements the SATE with sufficient attention to the modules for business rules in the LSR interfaces, processing systems, and the requisite legacy systems, CLECs can have reasonable assurance that order processing mirroring has been achieved. The implementation of Qwest's SATE on July 31, 2001, began prior to the negotiations for the PO-19 measurement. There is no

---

<sup>1</sup> PO-19 – Stand-Alone Test Environment (SATE) Accuracy – 05 Oct 01 DRAFT

reason to believe Qwest would intend not to implement mirrored edit, validation, and format processes in the SATE than are in the production environment. A CLEC building an EDI interface is reliant upon Qwest IMA User Guides, implementation guidelines and other Qwest-supplied documentation and the SATE assists in validating whether the CLEC's interpretations of Qwest's business rules and specifications are correct. The CLEC wants to test its implementation successfully and implement smoothly.

When CLEC EDI orders are processed by Qwest, the application of the business rules and processes that determine whether CLECs have used the proper interface specifications, is a pass/fail level of LSR evaluation; i.e., an EDI LSR must be 100% consistent, or it will be removed from electronic processing as a fatal reject or set aside for manual processing. Qwest provides no tolerance for less than accurate LSRs. For the mirroring principle to be equitably evaluated, the same standard should apply.

The CLECs have worked effectively with Qwest on the implementation of the SATE and the PO-19 measurement. The relative newness causes a realization that the first six months of operation may find weak spots that need adjustment, or unforeseen implementation problems. The 2% tolerance level each month is a reasonable recognition by the CLECs that Qwest may find difficulties that need to be corrected to the SATE so that it operates just like the production environment. PO-19 *does not evaluate the results of CLECs using the SATE*, but rather it reports on Qwest's own execution of the Qwest-developed "test deck" in the SATE for one transaction (pre-order and order steps) per defined scenario for each of the IMA releases using all current data definition documents. PO-19's description states:

For this measurement, **Qwest will execute the test transactions** in the Stand-Alone Test Environment. Release related test transactions will be executed when a full or point release of IMA is installed in SATE. These transactions will be executed within five business days of the numbered release being originally installed in SATE. This five-business day period will be referred to as the "Testing Window." Mid-release monthly performance test transactions will be executed in the months when no Testing Window for a release is completed. These transactions will be executed on the 15<sup>th</sup>, or the nearest working day to the 15<sup>th</sup> of the month, in the months when no release related test transactions are executed. [Emphasis added.]

CLEC transactions have no influence on the results that Qwest would produce for PO-19. This measurement is an evaluation of Qwest's provision of the test data and the Qwest implementation of the SATE edit, validation, and business rules. If Qwest fails to properly develop the test data, resulting in PO-19 failures, the same data would already have been used by CLEC SATE users, necessitating re-work by the CLECs. If Qwest fails to properly implement the SATE edits, validations, and business rules, CLECs that believed they

successfully tested using SATE would have to retest once the corrections to SATE are implemented.

The SATE should enable CLECs to conduct their testing in a reliable environment and then migrate to the production environment, assured that the testing was rigorous and thorough. KPMG's Exception 3077 underscores this issue: "KPMG Consulting's expectation is that test environment transaction responses should mirror those from the related production environment. Accordingly, CLECs can gain a reasonable level of assurance that they will receive the same results for the transactions that they are testing, once they migrate into production. This should facilitate a smooth transition into production for CLECs, and minimize problems for both the CLEC and Qwest."<sup>2</sup> Poor PO-19 results would demonstrate that CLECs do not have access to a test environment that supports effective implementation of electronic interfaces.

To conclude, Qwest's preference for a 95% accuracy standard allows a higher level of errors in the SATE with the consequence of rejects and additional work for CLECs upon production turn-up. Given that CLECs are reliant on the Qwest-provided test bed for use in the SATE, and the demand of 100% accuracy for the EDI LSRs in Qwest's production environment, CLECs believe a 2% error-rate is not only reasonable, but fair in consideration of the fact that PO-19 measures Qwest's use of its own testing processes.

### **Steering Committee Resolution**

By a unanimous vote of nine (9) to zero (0), with one abstention, the Steering Committee (SC) determined that the benchmark to be used for the ROC PO-19 PID should be 95% beginning in March 2002 and should be revisited within six months of that time.

The SC considered the following key aspects in its determination:

- A benchmark of 95% does not seem unreasonable based on current results
- Implementation of this interim benchmark starting in March 2002 coupled with a 6-month review allows time for enhancements to the SATE platform to reach maturity and stability before a final benchmark is established
- A 95% benchmark in the interim should encourage Qwest to not release future upgrades of SATE until such time as the release is performing at least to a 95% level of accuracy, thus furnishing the CLECs with reasonable assurance of a stable platform

Voting on the issue: All states in attendance except Minnesota that abstained. North Dakota, New Mexico and Wyoming were not represented on the call.

---

<sup>2</sup> Exception 3077, KPMG 1st Response, 1-8-02 pp 11

