Monitoring the performance of commercial T1-rate transmission service

by D. R. Irvin

This paper gathers the scattered empirical and theoretical elements of the performance-management problem for commercial T1-rate transmission service and integrates these elements in a useful way. We propose two variants of a time-based performance-monitoring algorithm that are insensitive to the arrival pattern of transmission errors. The first variant compares a count of errored seconds accumulated over an interval of time to a fixed threshold, and issues an alert to the network operator indicating degraded transmission performance whenever the count exceeds the threshold before the measurement interval expires. The fixed-threshold test is calibrated with reference to the well-known Neyman model of transmission errors on metallic-conductor systems. This calibration is then shown to be suitable as well for monitoring the performance of fiber-optic transmission systems where errored seconds follow the cumulative binomial distribution. The second variant of the new performance-monitoring algorithm replaces the fixed-threshold test with a dual-threshold test having a lower threshold that remains fixed and a higher threshold that floats in response to changes in error characteristics. An analysis based on the difference equations that describe the movement of the floating threshold shows that the dual-threshold test is more responsive than the fixed-threshold test in detecting nonstationary trends toward degraded transmission and in detecting stable but mediocre performance levels.

Introduction

Network-management systems and digital transmission at T1 rates have become mainstays of private data communication networks. Much has been written about each, but little has appeared in the literature bringing the theoretical and the practical aspects of digital transmission and network management together in a useful manner. This paper gathers these various and scattered elements, melds them, and proposes a new algorithm for monitoring the performance of commercial T1-rate transmission service that operates over a wide range of error-arrival statistics. In short, the algorithm operates by 1) comparing an error count kept over a predefined time interval to a threshold, where the numerical value of the threshold is

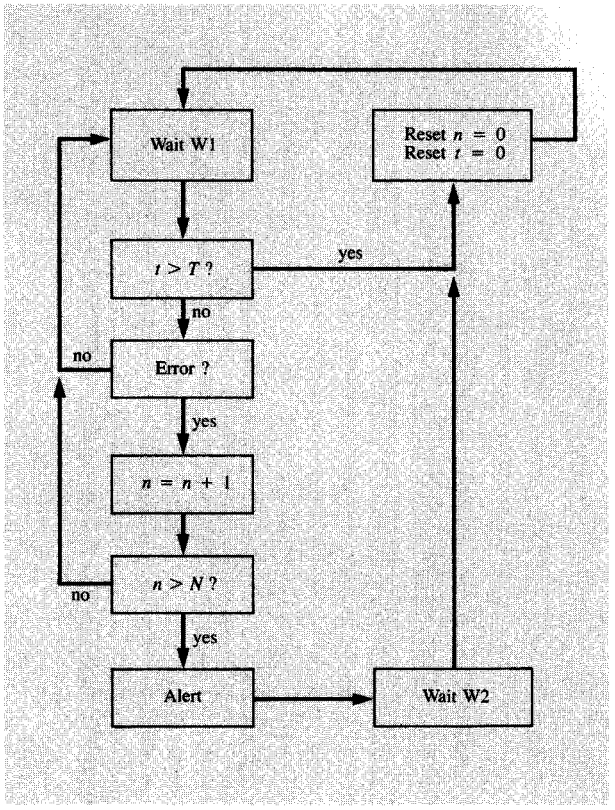

Figure 1

Flowchart showing the counter-management technique: Cycle is driven by error-status interrupts which occur every second.

either held fixed or allowed to float in response to changes in error statistics, and 2) triggering an alert that indicates degraded performance whenever the error count crosses the threshold before the measurement interval expires. Because the algorithm allows an alert to be issued before the measurement interval expires, a report of degraded transmission performance can be promptly dispatched even when long measurement intervals are chosen in the anticipation of good transmission performance.

To select appropriate numerical values for the threshold and measurement interval, three aspects of the performance-monitoring problem must be understood and accommodated: the metrics used in specifying the performance of commercial T1-rate transmission services, the performance objectives set for these services by the common carriers, and the statistical behavior of transmission errors. The questions of metrics and objectives are easily settled, because the telecommunications industry has adopted a set of metrics based on errored seconds of transmission time, and the common carriers often make public disclosures of performance objectives for their services. A search of the body of theory concerning the behavior of transmission errors, however, yields two well-established but distinct trains of thought and two correspondingly disjoint mathematical models. One holds that transmission paths that include copper wire as a medium have error characteristics dominated by impulse noise, and that the resulting bursts of errors can often be described by the Neyman Contagious Type-A probability distribution. The second view holds that fiber-optic transmission paths that include a degraded segment have error characteristics dominated by Gaussian noise, and that the independent bit errors occurring under these circumstances can be described by the Poisson probability distribution.

From an intellectual perspective, the two views are complementary; today's copper-wire transmission plant is evolving into the all-fiber system of the future. From a practical vantage, however, having these two divergent views is not helpful, since the link-management tools available to a particular network may not be able to classify each circuit according to its transmission medium and then accommodate the different prospects with different strategies. The apparent divergence of the two views on transmission errors defines the primary question to be answered here: Can a generic time-based monitoring technique, the different theoretical models of the error-generation process, and the performance levels offered by commercial T1-rate services be accommodated as a useful whole, such that a single algorithm, with a single set of parameters, can determine the health of a circuit without having *a priori* knowledge of the underlying transmission medium?

The remainder of the paper answers this question in the affirmative. First, details of the new fixed-threshold performance-monitoring algorithm are given, followed by a brief discussion of the metrics and objectives for the performance of modern commercial T1-rate transmission services. We then review and contrast two well-established theoretical models of error occurrence, the Neyman Contagious Type-A for copper transmission paths corrupted by bursts of errors, and the Poisson for degraded fiber-optic paths with independently occurring errors. An examination of the behavior implied by these two models suggests that a measurement interval of 900 seconds and a fixed threshold of 90 events are appropriate. In the case of copper-wire circuits, this choice of parameters follows from the numerical evaluation of a bound on the cumulative error statistics, where that bound is derived from the Neyman Type-A distribution as a consequence of Chebychev's inequality. In the case of optical-fiber circuits, the same choice of parameters follows from the direct numerical evaluation of the cumulative error statistics, which are shown to take the cumulative binomial distribution. These numerical evaluations indicate that circuits performing at benchmark

levels of 99% or better error-free seconds (EFS) of transmission can be clearly differentiated from those that have degraded to 90% or worse error-free seconds, and that this differentiation can be made with a low probability of false negatives for circuits with error statistics that follow either the Neyman Contagious Type-A or the Poisson model. Finally, we explore the possibility of allowing the threshold to float in response to changes in error statistics: A one-fixed, one-floating dual-threshold test is proposed. An analytic model describing the behavior of the dual-threshold test is developed and is then applied in an example to show that the dual-threshold test is more responsive than the fixed-threshold test in detecting nonstationary trends toward degraded transmission and in detecting stable but mediocre performance levels.

A time-based performance-monitoring algorithm

The keeping of various sets of counters by equipment located on customer premises has become a key aspect of managing communication networks. Such counters monitor the operation of the communication system and detect degradation or failure of various transmission components and circuits. Having knowledge of the network's statistics, however, is only the first step in ensuring that the network meets established performance targets. A second aspect of the overall problem involves examining the data gathered by the counters and deciding whether the elements of the network are operating within specification. When a network element fails or gives signs of an impending failure, an operator should be notified so that corrective action may begin. In the data processing world, such notifications are known as *alerts*; in the world of telephony, they are called *alarms*. Moore reviews the history, function, and operation of alerts within IBM's Systems Network Architecture (SNA), and shows how the scope of SNA alerts can be extended to include the management of multivendor networks [1].

A general time-based technique [2] for managing link-performance counters and dispatching alerts is given in **Figure 1**. In the flowchart shown there, parameter T is the duration of the measurement period. Parameter t measures the flow of time, and is kept with modulo T; thus t advances from 0 to T, and is reset to 0 on each occurrence of T. Parameter N is a fixed threshold for the event counter; parameter n indicates the present contents of the counter. At the top of the flowchart shown in Figure 1 is state W1, where the algorithm waits in a quiescent state for interrupts that arrive every second to report transmission-error status. On each interrupt, the value of timer t is checked. If t has expired, the event counter n is reset, and t is returned to 0 and allowed to advance; otherwise (t has not expired), the transmission status report is examined. If the report indicates that no errors were detected in the last one-second interval, the algorithm returns to state W1 and waits for the next one-second interrupt; otherwise (transmission errors were detected), the event counter is incremented, and the numerical value of its contents, n, is tested against the threshold value N. If n does not exceed N, the algorithm returns to state W1 and waits for the next one-second report; otherwise (n exceeds N), an alert is issued, and the algorithm then enters state W2, where it waits for a period of time to inhibit the sending of back-to-back alerts in the case of a hard failure. On leaving W2, the event counter n is reset to zero, timer t is reset to zero, and the algorithm returns to quiescent state W1, where it waits for the first one-second report of the next measurement interval. It is important to note that even though a defined measurement interval has been introduced, an alert may be issued as soon as the event counter crosses its threshold—it is not necessary to wait for the T-second timer to expire.

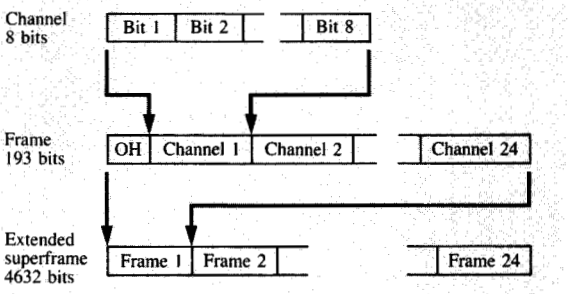

Figure 2

Structure of the extended superframe: OH is the overhead bit.

The extended superframe format and the CRC-6 check

The fundamental unit in T1-rate transmission is the 193-bit frame, which recurs every 125 microseconds. Each 193-bit frame may be subdivided into twenty-four 8-bit slots, each slot corresponding to a DS0 channel, plus one framing bit per frame. The extended superframe format (ESF) imposes a 24-frame superframe over the 193-bit frames, as shown in **Figure 2**. The 193rd bit, recurring every 125 microseconds, consumes 8000 bps of channel capacity; this 8000 bps accounts for the difference between the two often-cited T1 rates: 1.544 or 1.536 Mbps. ESF framing further divides the 8000-bps overhead channel into subchannels for three distinct uses: a 4000-bps channel for network maintenance; a 2000-bps channel for maintaining synchronization; and a

2000-bps channel for a six-bit cyclic redundancy check, known as the CRC-6, which provides the basis for determining the performance of a circuit.

In most of today's configurations, the error counters derived from the extended superframe format are kept by customer-premises equipment (CPE) and read by the central office (CO). Standards are progressing toward defining a symmetric relationship where the CPE can read the CO registers, and toward defining a uniform way to communicate CPE and CO counts to the end-user's network-management application program. As new equipment supporting these functions is developed by the carriers, the set of performance metrics derived from the ESF counters should continue to assume an increasingly important role in defining performance levels for commercial transmission services.

The performance of commercial T1-rate transmission service

An illustration of a commercial performance specification is now given in preparation for a discussion of how the test proposed in Figure 1 may be calibrated. In order to uncouple the problem of specifying performance from the problem of identifying and classifying the principal noise source responsible for introducing errors, metrics describing the performance of commercial service offerings are based on the idea of an errored second of transmission. An errored second is a one-second interval during which either frame synchronization is lost, or one or more bit errors are detected by the CRC-6 check carried in the extended superframe format.

Although much modern technology has been introduced into the public network, local circuits that connect the customer's premises to the service provider's nearest point of presence remain, in many cases, copper-wire twisted pairs. Because of the persistence of twisted-pair access, the performance characteristics of the local-access segment of a connection may differ from the performance characteristics of the other components that make up the connection, as illustrated in the following objectives for a commercial T1-rate service now in widespread use.

Access performance objective The objective for the performance of a T1-rate local-access circuit is a maximum of 170 errored seconds (ES) per 24-hour period [3].

Interexchange performance objective The objective for the interexchange portion is [4]

- A maximum of 86 ES per 24 hours for circuits shorter than 250 miles.
- A maximum of 346 ES per 24 hours for circuits between 250 and 1000 miles in length.
- A maximum of 605 ES per 24 hours for circuits longer than 1000 miles.

End-to-end performance objective End-to-end performance objectives are found by adding the ES contributions of two access circuits (340 ES per 24 hours) and one interexchange circuit, giving

- A maximum of 426 ES per 24 hours (99.51% EFS) for circuits shorter than 250 miles.
- A maximum of 686 ES per 24 hours (99.21% EFS) for circuits between 250 and 1000 miles.
- A maximum of 945 ES per 24 hours (98.91% EFS) for circuits longer than 1000 miles.

According to the tariff covering this offering, the customer receives a credit allowance for interrupted service when 300 or more errored seconds occur in a 900-second interval [5]. The provision for credit allowance suggests an opportunity for a second test, similar in nature to the test shown in Figure 1, to detect service interruptions that meet the formal criterion and to make note of the credit due.

The Neyman model of transmission errors

Let us now turn to the problem of choosing values for T and N to fit the algorithm proposed in Figure 1 to the performance levels offered by commercial T1-rate services using metallic conductors as the transmission medium. To accomplish this, a mathematical model that describes the introduction of transmission errors is needed. Over the years, numerous models of the error-generation process have been proposed; Kanal and Sastry examine sixteen of these models and reference eighty-five related papers in their 1978 overview of the topic [6]. More recently, the Neyman Contagious Type-A probability distribution, borrowed from the biological sciences, has come into prominence for describing the performance of circuits whose error characteristics are dominated by bursts of externally generated noise coupled into the transmission system through copper-wire paths [7–10].

One strength of the Neyman Contagious Type-A model is that it takes into account the possibility that the errored seconds may themselves occur in clusters. The descriptor *contagious* is attributed to an earlier work, where the term was used to indicate that the occurrence of one event increases the probability of occurrence of additional events in the immediate neighborhood [11]. In the Neyman model, the occurrence of clusters of errored seconds is represented by a Poisson process with parameter m_1. The number of errored seconds in each cluster is represented by a second Poisson process with parameter m_2. In principle, the values of m_1 and m_2 may be estimated from the measured statistics of occurrence of errored seconds

on a particular circuit under scrutiny. For the purposes of this estimation, the observation interval is subdivided into a number of equal subintervals, each of length T seconds. The expected value of the count of errored seconds over the duration of the experiment is given by the product $\hat{\mu} = m_1 m_2$; the variance is given by $\hat{s}^2 = m_1 m_2(1 + m_2)$, where the value of m_1 is normalized with respect to the interval T. Now, m_1 and m_2 may be found by solving two equations against the observed mean and variance [8].

The arrival statistics of the distribution provide the basis for computing a numerical value for the threshold N. Let $P\{n = k\}$ be the probability that k errored seconds occur in an observation interval of duration T seconds. Then, from the defining equation of the Neyman distribution,

$$P\{n = k\} = e^{-m_1} \frac{m_2^k}{k!} \sum_{t=0}^{\infty} \frac{z^t t^k}{t!}, \qquad k = 0, 1, 2, \cdots, T, \qquad (1)$$

with the notational convenience $z = m_1 e^{-m_2}$. The probability that the count of errored seconds will exceed the threshold during the measurement interval is given by the cumulative form of Equation (1):

$$P\{n > N\} = \sum_{k=N+1}^{T} e^{-m_1} \frac{m_2^k}{k!} \sum_{t=0}^{\infty} \frac{z^t t^k}{t!}. \qquad (2)$$

The Neyman model, under appropriate scaling assumptions, has been shown to provide an acceptable model of the occurrence of errored seconds for many circuits whose performance is characterized by bursts of errors [9]. Readers who would like to pursue the details of this model further should refer to the source papers. Neyman [7] is the primary source; Herzer [10] gives a good tutorial and overview, and specifically considers how the model relates to the world of telecommunications. Becam et al. [8, 9] show how the Neyman model may be rigorously tested using examples taken from the measured performance of the French telephone network. Beall has proposed a recursive form of Equation (1); this is given in Neyman's paper and recommended there as a computational aid. Ritchie and Scheffler [12] and Brilliant [13] give additional data on the measured performance of T1 circuits.

Calibrating the fixed-threshold test

The preceding section of this paper discussed how parameters m_1 and m_2 may be determined to fit the Neyman Contagious Type-A model to performance data measured on a particular metallic-conductor circuit, and how to evaluate the performance-management threshold accordingly. Let us now address the question of calibration and the question of generality in assuming that a useful threshold can be determined for metallic-conductor circuits whose error statistics follow a Neyman distribution, but where the parameters of that distribution are not known and the means to find them are not available, and where the same parameter values must accommodate fiber-optic circuits whose error statistics follow a different distribution.

A bound can be established on the behavior of the Neyman Contagious Type-A distribution using Chebychev's inequality together with the expressions giving the mean and variance of the distribution. The threshold N can be conveniently restated in terms of an offset N_μ from μ, where μ is the number of errored seconds expected over the measurement interval T, as follows:

$$N = \mu + N_\mu .$$

Now, $n > N$ implies that $n - \mu > N_\mu$, where n is the count of errored seconds during the measurement interval T. Chebychev's inequality holds that

$$P\{|n - \mu| \geq \alpha\sigma\} \leq \frac{1}{\alpha^2} .$$

For the Neyman Contagious Type-A distribution, we know that $\sigma^2 = m_1 m_2(m_2 + 1)$, and $\mu = m_1 m_2$. With the substitution $N_\mu = \alpha \sqrt{\mu(1 + m_2)}$, Chebychev's inequality becomes

$$P\{|n - \mu| \geq N_\mu\} \leq \frac{\mu(1 + m_2)}{N_\mu^2} . \qquad (3)$$

As an example relating this bound to the performance level of commercial T1-rate service, let the measurement interval T be 900 seconds, and let the threshold N be 90 errored seconds; these particular values were selected by trial and error. Furthermore, let satisfactory performance be defined as 99% or better error-free seconds (EFS), and unsatisfactory performance be 90% or worse EFS. Suppose that a circuit under test operates at the level of 99% EFS. This implies that the average number of errored seconds expected during the 900-second test is given by $\mu = 9$. Equation (3) gives the probability of false alert as

$$P\{|n - 9| \geq 81\} \leq \frac{9(1 + m_2)}{81^2} = \frac{1 + m_2}{729} . \qquad (4)$$

A brief digression helps explain the implications of Equation (4). Recall that for the Neyman Contagious Type-A model, the occurrence of clusters of events is represented by a Poisson process with parameter m_1; the number of events in each cluster is represented by a second Poisson process with parameter m_2. In examples taken from the biological sciences by Neyman [7], and from the performance of telecommunication links by

Becam and Brigant [9], the value of parameter m_2 does not exceed 1.5, and normally takes values less than 1.0. The net effect of this can be understood on an intuitive level by noting that values of m_2 less than 1.0 imply that a significant number of the clusters are clusters of zero events, which in turn implies that the remaining clusters are each populated by multiple events. This peculiar property defines the unique personality of the Neyman Contagious Type-A distribution and illustrates its contagious nature.

Returning to Equation (4), we see that imposing an *ad hoc* restriction that $m_2 < 1.5$ gives a numerical bound on the probability of a false alert during a single 900-second interval as

$$P\{|n - 9| \geq 81\} \leq \frac{9(1 + m_2)}{81^2} = \frac{1 + m_2}{729} < 0.004.$$

Under these conditions, the likelihood of having a false alert for the circuit operating at 99% EFS can be bounded from above by one occurrence every several days. If the performance of the circuit deteriorates to 90% or worse EFS, an alert will be issued with virtual certainty over the course of a few contiguous 900-second tests, as the expected value of errored seconds for the 90% or worse level of performance is equal to or greater than the threshold, 90 errored seconds. Thus, to the extent that the two benchmarks 90% EFS and 99% EFS are acceptable, the time-based test for errored seconds is satisfactory for monitoring the performance of T1-rate twisted-pair service.

Optical fibers are practically immune to interference caused by external factors. Furthermore, the statistical description of transmission errors that occur on optical fibers differs from the statistical description of errors on metallic circuits. A recent study [14] gives performance data gathered during a laboratory experiment where the operating margin for a fiber-optic link was deliberately lowered to produce a significant number of transmission errors. In the field, such a failure mechanism might plausibly stem from faulty splices or connectors that cause the light path to become partially obscured or misaligned. During the laboratory experiment, the occurrence of individual bit errors was found to conform closely to the Poisson model. This agreement suggests that the individual bit errors introduced by degradation of a fiber-optic link are statistically independent, much as they would be if they were caused by the injection of a Gaussian noise impairment. Whereas there is evidence that individual bit errors are independent events for degraded fiber-optic links, this idea has been thoroughly discredited as a descriptor of the performance of metallic conductor systems in the presence of impulse noise [15].

Once again, a question arises about how the technique proposed in Figure 1 is to be calibrated to the performance of commercial data communication services, this time for all-fiber links. Under the assumption that bit errors on a degraded fiber-optic link are statistically independent events, the probability that J errors occur in M bits is given by the binomial distribution

$$P\{J \text{ errors in } M \text{ bits}\} = \binom{M}{J}(P_{\text{BER}})^J(1 - P_{\text{BER}})^{M-J}, \quad J = 0, 1, 2, \cdots, M. \tag{5}$$

In Equation (5), P_{BER} is the probability that a bit will be in error, or the bit-error ratio (BER).

If one or more bit errors occur during a one-second period of transmission, that period is an errored second. The probability that a second contains at least one error may be found by subtracting from unity the probability that a second contains no errors. From Equation (5), setting M equal to the number of bits transmitted in a second, and $J = 0$, the probability of an errored second P_{ES} is given by

$$P\{\text{a second is errored}\} = P_{\text{ES}} = 1.0 - (1.0 - P_{\text{BER}})^M. \tag{6}$$

Under the approximation that the occurrences of errored seconds are independent events, the probability of occurrence of more than N errored seconds in T trials (i.e., in a T-second test) is given by the cumulative binomial distribution

$$P\{\text{more than } N \text{ ES in } T \text{ seconds}\} = \sum_{k=N+1}^{T} \binom{T}{k}(P_{\text{ES}})^k(1 - P_{\text{ES}})^{T-k}. \tag{7}$$

Numerical evaluation of Equations (6) and (7) shows that the algorithm proposed in Figure 1, together with a 90-event threshold and a 900-second measurement interval, has remarkable power to distinguish between the performance benchmarks of 99% EFS and 90% EFS for T1-rate transmission under the assumption of independent bit errors for fiber-optic circuits. The probability of a false alert is virtually zero, while the probability of detecting degraded performance in one 900-second interval is virtually unity. In this case, 99% EFS corresponds to $P_{\text{BER}} = 6 \times 10^{-9}$, and 90% EFS corresponds to $P_{\text{BER}} = 7 \times 10^{-8}$.

Although the crispness of the fixed-threshold test improves greatly when the test is applied to optical-fiber transmission systems, fixed-threshold algorithms of the kind shown in Figure 1 are fundamentally and theoretically limited by the possibility that false alerts will be issued in response to the normal fluctuations of a circuit performing at a satisfactory level. Ultimately, any two levels of performance that can be differentiated are separated by a gap, since a test designed to give timely results must be based on a measurement interval that is short relative to

the 24-hour measurement interval specified by the common carriers. Narrowing the gap exaggerates the problem of false alerts. If the values of m_1 and m_2 are known in the case of metallic-conductor systems, a numerical evaluation of Equation (2) allows tighter benchmarks than those used here, since the Chebychev inequality is fairly loose. The final choice in accepting false alerts is subjective, of course, and the benchmarks can be tailored according to taste: As knowledge is gained about the performance of a particular circuit, the threshold can be adjusted heuristically to narrow the gap. Heuristic adjustment is not the only way to minimize false alerts and penetrate the gray area hidden from the fixed-threshold test, however, and we next consider an alternative: a dual-threshold algorithm, where one threshold remains fixed and the other threshold floats. Another alternative, not pursued here, is to select different parameter values or algorithms to match the performance characteristics of different kinds of circuits, perhaps one set for copper wires and another set for optical fibers. Although the benefits of such an approach might be argued convincingly in theory, the practical implications of taking this tack would be to increase the complexity of the performance-monitoring apparatus and to require that the ultimate users of such a system have more intimate knowledge of the ever-changing composition of the common-carrier network. Whether the added complexities would provide significant added value to the ultimate users remains an open question.

A dual-threshold algorithm wherein one threshold floats

So far, we have discussed the general problem of monitoring the performance of T1-rate transmission service and the specific problem of selecting an appropriate numerical value for a fixed-threshold test designed to detect degraded transmission integrity. In practice, the degradation of a link may arise from an abrupt failure of a link component or from the gradual deterioration of the underlying transmission channel. Gradual deterioration is often driven by some form of electromagnetic noise, where that noise is described by a stochastic process. Stochastic noise, however, is always present as an interfering component at work in the background of transmission channels; consequently, there is not always a clear distinction between a healthy transmission channel that is experiencing temporary difficulty as the natural result of the random fluctuations expected of noise and a troubled channel that is experiencing the early stages of noise-driven failure.

Given the ubiquitous presence of random noise fluctuations, the numerical value of the fixed threshold must be chosen so that the need for sending an alert can be determined under five different operating conditions:

- A healthy channel operating within its nominal bounds of operation.
- A healthy channel temporarily perturbed outside its nominal bounds of operation by the chance excursions of noise.
- A channel operating in the gray area of stable but mediocre performance just below the threshold.
- A channel that is driven through stages of degraded performance toward failure by a systematic increase in the interfering noise.
- A channel that is in a state of complete failure.

The desire to pass judgment on these five operating conditions in the presence of random noise puts conflicting requirements on the fixed-threshold test. If the threshold is set too low, the normal noise fluctuations present in the healthy channel result in the sending of false alerts, which distract the network operator and provide incorrect and unneeded information. If the threshold is set too high, states of gray-area performance may not be detected at all, and states of progressive deterioration that drive toward failure may be detected only after a considerable delay. Some of the problems inherent in setting the alert-generation threshold can be solved by using a dual-threshold algorithm wherein one threshold is fixed and the other threshold floats. Adopting such an approach provides the following improvements:

- Increased probability of detecting gray-area performance.
- Decreased probability of issuing false alerts.
- Decreased delay in detecting progressive failure caused by a systematically growing noise impairment.

We now turn to an examination of the dual-threshold approach: The algorithm itself is first described, numerical values for the various components that are appropriate for monitoring commercial T1-rate transmission service are suggested, and the performance of the floating-threshold test is examined.

- *Description of the dual-threshold algorithm*

The dual-threshold algorithm operates as part of a larger system for collecting data to be examined, such as the time-based system described in Figure 1, where a new function, threshold adjustment, is incorporated as shown in **Figures 3** and **4**. In Figures 3 and 4, L_m is the value of the floating threshold on the mth iteration of its adjustment, K is the constant of convergence for the threshold attack algorithm, G is the constant of convergence for the threshold retreat algorithm, L_0 is the initial value of the floating threshold, n_m is the present value of the errored-second counter, and m is the iteration number, or number of times the floating threshold is adjusted before the general reset that accompanies an alert. The symbol N,

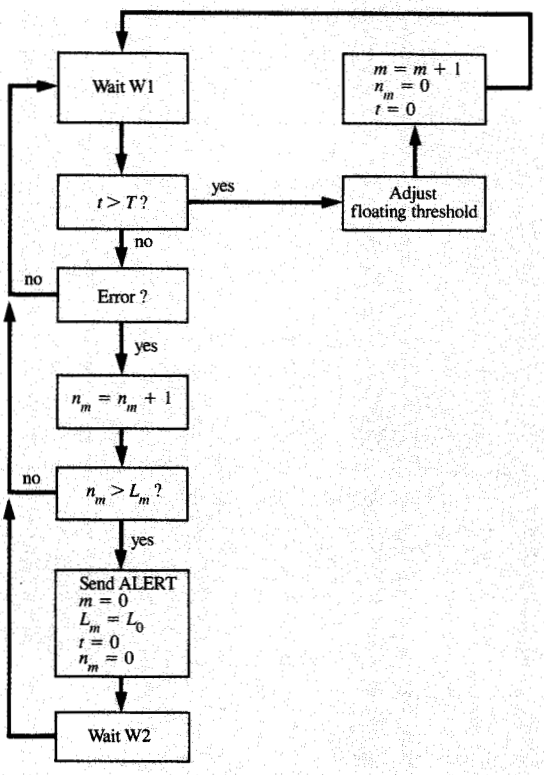

Figure 3

Counter-management technique modified to hold the dual-threshold test: See Figure 4 for a description of the threshold-adjustment algorithm.

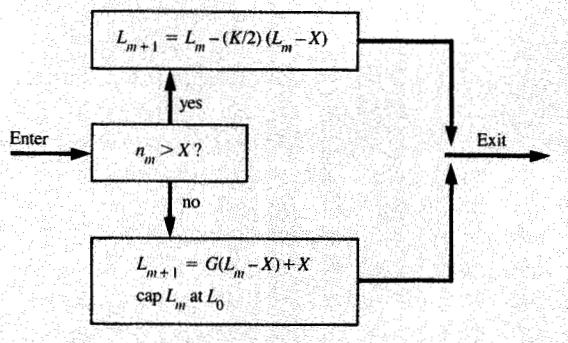

Figure 4

Threshold-adjustment algorithm: Enter from, and exit to, Figure 3.

which denoted the (single) fixed threshold in Figure 1, is not used in describing the dual-threshold variant.

The floating threshold is adjusted as shown in Figure 4: On the expiration of timer T, the count of errored seconds n_m is compared to a fixed lower threshold, here called X, which is set at roughly one standard deviation above the expected value of the errored-second count. If n_m does not exceed this fixed lower threshold, the floating threshold L_m is adjusted upward, the index of iteration m is increased by one unit, and the process returns to the wait-for-data state shown in Figure 3. If n_m exceeds the fixed lower threshold, the floating threshold L_m is adjusted downward, m is increased by one unit, and the process returns to the wait-for-data state.

For commercial T1-rate transmission service, the expected number of errored seconds over a 900-second measurement interval is 9, which corresponds to 99% EFS of transmission. Let σ be a rough estimate of the standard deviation of the count of errored seconds over the measurement interval; then, the fixed threshold is set at $X = \sigma + 9$. For purposes of discussion, let $\sigma = 9$, which gives for this example $X = 18$; the test is not sensitive to the choice of the numerical value of σ. The initial value of the floating threshold is $L_0 = 90$, which corresponds to 90% EFS, in agreement with the value suggested earlier in this paper as the minimum-acceptable-performance benchmark. The constants of convergence for the threshold-adjustment algorithm are set at $K = 0.500$ and $G = 1.344$; these choices are explained subsequently.

• *Analysis of the dual-threshold technique*

Comparing n_m to the (fixed) lower threshold X serves to characterize in some sense the count of errored seconds as "insignificant" or "significant." In the case of insignificance, the numerical value of the floating threshold is increased according to the threshold-retreat algorithm $L_{m+1} = G(L_m - X) + X$, subject to a cap at the initial value L_0. This increase is made so that

- The numerical value of the floating threshold will be progressively restored to its initial value L_0 when a trend toward degraded performance is reversed before becoming severe enough to merit the sending of an alert.
- The numerical value of the floating threshold will not drift aimlessly lower during periods of good performance; during such periods, the floating threshold will be continually reset to the maximum (initial) value by the operation of the threshold-retreat algorithm and the cap.

If the error count has been deemed significant, the floating threshold is lowered according to the threshold-attack algorithm $L_{m+1} = L_m - (K/2)(L_m - X)$. This lowering of the floating threshold makes the test more sensitive on the next iteration: The threshold moves methodically downward to meet a systematically rising trend of increasing error counts, or moves methodically downward to meet a sustained level of gray-area performance that lies in this example between the 18-event and the 90-event thresholds.

The monotonic operation of the attack algorithm for the floating threshold is given in Figure 4 by the first-order difference equation:

$$L_{m+1} = L_m - \frac{K}{2}(L_m - X).$$

This recursion may be unraveled to give

$$L_m = (L_0 - X)\left[\frac{2 - K}{2}\right]^m + X. \quad (8)$$

Figure 5 shows the monotonic attack pattern of the floating threshold as given by Equation (8) for $K = 0.5$, using the parameter values assumed earlier ($X = 18$, $L_0 = 90$). Also shown in Figure 5 is a linear regression illustrating by example a hypothetical, systematic increase in the count of errored seconds, where this count rises from nine errored seconds in the initial 900-second interval to 90 errored seconds in the 16th 900-second interval. This regression is given by

$$\hat{n} = 9 + 5.0625m, \quad (9)$$

where $\hat{n}$ is the expected value of the count of errors in the mth 900-second measurement period. Note from Figure 5 that the expected time-to-detection of this hypothetical trend improves considerably when the threshold floats. The detection time could be improved further by increasing K, but improvement would come at the expense of increasing the likelihood of false alerts. If $K = 1.0$, the convergence pattern of the floating threshold becomes a binary search of the gray area between the fixed threshold X and the initial value L_0 of the floating threshold; setting $K = 0.0$ reduces the dual-threshold system to a special case where both thresholds are fixed.

The monotonic operation of the retreat algorithm for the floating threshold is given in Figure 4 by the first-order difference equation,

$$L_{m+1} = G(L_m - X) + X, \quad (10)$$

subject, since the recursion produces a divergent sequence, to a cap at $L_{m+1} \leq L_0$. If the constant G of the retreat algorithm is related to the constant K of the attack algorithm by

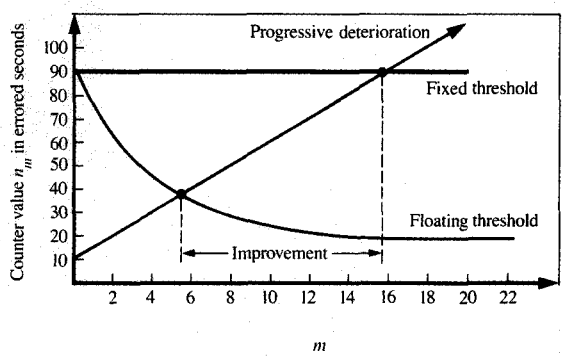

Figure 5

Improved time-to-detection offered by the dual-threshold technique: Shown here is a hypothetical, progressive deterioration of transmission quality. The iteration number is m.

$$G = \frac{2}{2 - K},$$

the attack and retreat algorithms are symmetric; i.e., an application of the retreat algorithm in the next several iterations retraces the steps of the attack algorithm.

The recursive form of the threshold-retreat algorithm given in Equation (10) may be unraveled as

$$L_m = (\hat{L}_0 - X)G^m + X. \quad (11)$$

In Equation (11), the initial value of the threshold-retreat algorithm is denoted by $\hat{L}_0$, which indicates the value of L_m when the retreat algorithm begins its operation; in most cases of interest, $\hat{L}_0$ will not be the same as L_0, since L_0 is the initial value of the threshold-attack algorithm.

Concluding remarks

This paper discusses new ways of monitoring transmission performance. Specifically, we have proposed two variants of a new time-based algorithm that bring together in a practical and useful way the scattered pieces of the performance-management problem for commercial T1-rate digital communication services. These pieces include the performance levels expected of commercial common-carrier offerings, and two well-established but disparate mathematical models, one describing the error-arrival statistics for metallic-conductor systems characterized by bursts of errors, the other describing the error-arrival statistics for fiber-optic systems characterized by independently occurring bit errors. A single set of parameter values is proposed for the fixed-threshold algorithm: $N = 90$ and $T = 900$. Bounds on the

performance of the fixed-threshold monitoring algorithm are derived from Chebychev's inequality applied to the Neyman Type-A distribution, and from the cumulative binomial distribution. Numerical evaluations of these bounds show the fixed-threshold technique capable of differentiating, under a wide range of assumed error statistics, between benchmark levels of performance appropriate for commercial T1-rate transmission service. An analytic model based on recursion shows that the floating-threshold test offers an improvement over the fixed-threshold test in applications that demand higher powers of discrimination and faster convergence times in separating tighter performance benchmarks or in dealing with nonstationary transmission error mechanisms.

Acknowledgments

The author thanks his IBM Research Triangle Park colleagues Kian-Bon Sy and Robert E. Moore for their suggestions concerning the techniques shown respectively in Figures 1 and 3.

References

1. R. E. Moore, "Utilizing the SNA Alert in the Management of Multivendor Networks," *IBM Syst. J.* **27**, No. 1, 15–31 (1988).
2. D. Irvin and K. Sy, "A Technique for Managing Network Counters," *IBM Tech. Disclosure Bull.* **33**, No. 6B, 72–75 (November 1990).
3. "Access Specification for High-Capacity (DS1/DS3) Dedicated Digital Services," *AT&T Technical Reference TR 62415*, AT&T Customer Information Center, 2855 N. Franklin Rd., Indianapolis, IN 46219, June 1989.
4. "ACCUNET T1.5 Service Description and Interface Specifications," *AT&T Technical Reference TR 62411*, AT&T Customer Information Center, 2855 N. Franklin Rd., Indianapolis, IN 46219, December 1988.
5. *Private Line Services; AT&T Tariff FCC No. 9*, AT&T Communications, Bridgewater, NJ, May 4, 1990, p. 82.
6. L. Kanal and A. Sastry, "Models for Channels with Memory and Their Applications to Error Control," *Proc. IEEE* **66**, No. 7, 724–744 (July 1978).
7. J. Neyman, "On a New Class of Contagious Distributions, Applicable in Entomology and Bacteriology," *Ann. Math. Statist.* **10**, 35–57 (1939).
8. D. Becam, P. Brigant, R. Cohen, and J. Szpirglas, "Testing Neyman's Model for Error Performance of 2 and 140 Mbit/s Line Sections," *Proceedings of the 1984 IEEE International Conference on Communications*, May 1984, pp. 1362–1365.
9. D. Becam and P. Brigant, "Poisson and Neyman Models Applied to Errored Seconds on Digital Transmission," *Proceedings of the 1986 IEEE International Communications Conference*, May 1986, pp. 38.2.1–38.2.5.
10. R. Herzer, "Comments on the Neyman and Beall Formulas for Contagious Type-A Probability Distributions Applied to Burst Processes Especially in the Field of Digital Transmission," *Frequenz* **37**, No. 7, 182–186 (1983); in English.
11. G. Polya, "Sur quelques points de la théorie des probabilités," *Annales de l'Institut Henri Poincaré* **1**, 117–162 (1931).
12. G. Ritchie and P. Scheffler, "Projecting the Error Performance of the Bell System Digital Network," *Proceedings of the 1982 IEEE International Conference on Communications,* May 1982, pp. 2D.2.1–2D.2.7.
13. M. Brilliant, "Observations of Errors and Error Rates on T1 Digital Repeatered Lines," *Bell Syst. Tech. J.* **57**, No. 3, 711–746 (1978).
14. "Fiber Bit Errors—Effectiveness of ATM Cell Header Error Correction," *CCITT Contribution COM XVIII, No. D.171*, International Telecommunication Union, Geneva, Switzerland, June 19, 1989.
15. V. Johannes, "Improving on Bit-Error Rate," *IEEE Commun. Mag.* **22**, No. 12, 18–20 (December 1984).

Received July 19, 1990; accepted for publication July 22, 1991

David R. Irvin *IBM Networking Systems, Emerging Carrier Technologies, P.O. Box 12195, Research Triangle Park, North Carolina 27709.* Mr. Irvin is an advisory engineer/scientist at the IBM U.S. Telecommunications Center at Research Triangle Park. Since joining IBM in 1974, he has worked in the fields of telecommunication analysis, communications and network management architecture, system performance analysis, transmission technology, and digital signal processing. Prior to joining IBM, he worked under contract for the United States Navy. Mr. Irvin received a B.E.S. in 1970 from The Johns Hopkins University at Baltimore, an M.E.E. in 1971 from North Carolina State University at Raleigh, and the P.D.E. degree from the University of Wisconsin, Madison, in 1990. He is a member of Phi Beta Kappa, Tau Beta Pi, and Eta Kappa Nu, and a Senior Member of the Institute of Electrical and Electronics Engineers.