CLOSED LOOP STABILITY CONTROLS FOR S-ALOHA SATELLITE COMMUNICATIONS

Mario Gerla and Leonard Kleinrock

Computer Science Department
University of California, Los Angeles

ABSTRACT

S-ALOHA channels are intrinsically unstable and must be equipped with proper controls. The function of the controls is to dynamically adjust the ALOHA channel transmission gates in accordance with the dynamic load fluctuations. The purpose of the controls is to protect the channel from unstable behavior while optimizing channel efficiency and performance during normal operating conditions.

Two control algorithms are proposed: the Closed Loop Control-Collision Detect (CLC-CD) algorithm, which assumes the capability of distinguishing collision slots from empty slots at the receiving station; and the Closed Loop Control-Collision Non-Detect (CLC-CND) algorithm, which does not require such capability. The control implementation is distributed among all stations. Channel stability and efficiency is achieved by driving the total transmission and retransmission rate to unity, using a feedback, closed loop control approach.

A family of simulation runs was made to evaluate and compare the performance of the CLC schemes with that of other schemes in a variety of traffic conditions.

Simulation results show that the controlled systems converge to near optimality at steady state. Futhermore, the performance of the CLC-CND algorithms is about equivalent to that of the CLC-CD algorithm, thus indicating that the requirement of distinguishing collisions from empty slots is not critical for the performance of closed loop controls.

The stability properties of the CLC algorithms and their superiority over other schemes for varying load patterns are demonstrated in a series of experiments involving cyclic traffic patterns and pulse patterns. The CLC scheme displays better performance than the uncontrolled schemes as well as the previously proposed control schemes (namely, the Control Limit scheme and the Retransmission Control scheme) even when the latter are specifically tuned to handle the traffic pattern under consideration (the CLC scheme does not require any prior setting of the parameters).

1. INTRODUCTION

The S-ALOHA (slotted ALOHA) channel is a time division, multiple access channel for packet communications. Time is divided into fixed length slots each large enough to contain a maximum size packet. Stations with packets to send will transmit at the beginning of a randomly chosen slot, with no attempt to reserve slots or to preschedule retransmissions.[9]

Due to lack of prescheduling, packets from different stations may destructively collide (i.e. may be transmitted in the same slot) thus requiring subsequent retransmissions and causing an increase in the effective channel load. Because of this behavior, throughput and delay performance of the S-ALOHA channel depend critically on the traffic pattern and the transmission rates. Given the traffic pattern, there is an optimal transmission rate which minimizes delay. If the system is properly adjusted, then the optimal transmission rate yields both maximum throughput and minimum delay. If the transmission rate exceeds the optimum value, channel performance degrades and the channel is not operated cost-effectively. Moreover, the channel may be unstable under particular traffic load conditions. More precisely, a channel with transmission parameters optimized for the nominal load may be driven to a degraded mode by a peak load due to statistical fluctuations, and may remain in such a mode even when traffic conditions return to normal.

For the above reasons it is necessary to introduce transmission control mechanisms in the ALOHA channel. The objectives of the controls are:

- To adjust the channel parameters so that the channel operates at optimal performance for the offered (time varying) traffic pattern.
- To recover from instability generated by peak loads.

The functions of the control procedures are: monitoring of channel status (i.e., empty slots, collision slots, success slots, own retransmissions, etc.); and adjustment of the transmission parameters based on channel observations.

Control procedures may be centralized or distributed. Here we limit our study to distributed control procedures, since centralized procedures are affected by reliability and vulnerability problems.

A variety of distributed ALOHA controls may be found in the literature[2,3,4,9]. Unfortunately, most of the proposed schemes are based on some restrictive assumptions (e.g., only one buffer per station, time invariant traffic load and traffic pattern, large station population, capability of detecting collisions, etc.). Therefore, a direct application of such schemes to the most general S-ALOHA network is not always possible nor cost-effective.

In the following we introduce an S-ALOHA satellite network model. With this model we review the proposed techniques and discuss their limitations. We then present a novel approach to the control of ALOHA channels and propose two algorithms for its implementation.

2. THE MODEL

The satellite network model considered here consists of an arbitrary number of ground stations dynamically competing for access to a broadcast satellite channel in a slotted ALOHA mode.[1] The stations are equipped with large buffer pools to allow for efficient utilization of the channel, and may provide traffic concentration for both batch and interactive users.

Each station maintains two transmission queues: the queue of new packets and the queue of retransmit packets (i.e., packets that must be retransmitted because of error or collision). Service is FIFO in each queue, with the retransmit queue having strictly higher priority than the new queue.

Using the protocol of the S-ALOHA channel, transmissions occur at random times, and the intervals between subsequent transmissions are geometrically distributed. The randomization is obtained by applying probability gates P_N and P_R to new transmissions and retransmissions respectively (i.e., each station will retransmit a packet with probability P_R or, if the retransmit queue is empty, then a new packet will be transmitted with probability P_N at the beginning of each slot). In general $P_N \neq P_R$.

Of critical importance for the design of control procedures is the capability of distinguishing colli-

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract DAHC-15-73-C-0368.

sion slots from empty slots. In our study we consider two models: (1) Collision Detect (CD) Model, and (2) Collision Non-Detect (CND) model. The CD model assumes the capability of distinguishing collision slots from empty slots; the CND model assumes no such capability.

External arrivals are assumed to be Poisson, with time varying rates to reflect changes in the traffic pattern with time.

3. PREVIOUS WORK

The control of unstable random access channels has been the object of considerable research effort in recent years, leading to several contributions that are documented in the published literature. In this section, we summarize the techniques that are most representative of the state of the art in order to provide a framework for the introduction of our proposed closed loop control scheme.

3.1 Control Limit Policy[9]

This policy consists of monitoring channel load and dynamically changing gate values P_N and P_R from their steady state values to more conservative values when the channel load exceeds a given threshold.

Letting L_C be a critical (optimal) channel load (L_C may be expressed in terms of the fraction of empty slots) and P_0 and P_C be the gate probabilities under steady state and heavy load conditions respectively, the control policy selects the gate values as follows:

$$P_N = P_R = P_0 \quad \text{if} \quad L \leq L_C$$

$$P_N = P_R = P_C \quad \text{if} \quad L > L_C$$

While P_0 optimizes channel performance under normal operating conditions, P_C is sufficiently small so as to reduce channel congestion and drive the channel back to normal conditions. The control limit policy was proven to be very effective for systems with:

- A large number of stations with interactive traffic
- Single buffer stations
- Stationary traffic patterns

In order to implement the above procedure, one must first choose (experimentally) the values of P_0 and P_C that are appropriate for the given system configuration. Then, one analytically finds the optimal value L_C for the chosen values of P_0, P_C and the other system parameters.

In our satellite network model we have an arbitrary number of stations (small or large) with a mix of interactive and batch traffic and with multiple buffers. Furthermore, we permit changes in traffic pattern. These new requirements imply some modifications in the original definition of the Control Limit policy before application to our model, as discussed below.

First, a time-varying traffic pattern requires the dynamic adjustment of P_0 to ensure optimal channel performance with different channel load conditions. Second, an appropriate value P_C must be chosen for stability. Here the presence of multiple buffers creates serious complications. Consider what happens when a burst of traffic exceeds the channel capacity: queues will grow large; the channel load will exceed L_C; and the gate value will be reduced from P_0 to P_C. Since $P_C \ll P_0$, the channel load will rapidly fall below L_C. At this point, we restore $P = P_0$; but P_0 will most likely generate $L > L_C$ because of the existing backlog, and so on. The oscillation between P_0 and P_C may persist for a long time and may reduce the channel performance. Such oscillating behavior was verified in the simulation experiments (see Figure 10).

3.2 Retransmission Control Procedure[3,4,9]

This procedure consists of using m + 1 different gate values: P_N for new transmissions; and $P_R^{(i)}$, i = i,..., m for the i^{th} retransmission of the same packet. For retransmissions beyond m, the value $P_R^{(m)}$ is used. Note that, for m = 1, the gates reduce to P_N and P_R.

The value of P_N is chosen so as to optimize steady state performance, while the $P_R^{(i)}$'s are chosen in decreasing sequence (i.e., $P_R^{(1)} > P_R^{(2)} > P_R^{(m)}$), based on the concept that a high number of retransmissions is a symptom of channel congestion and that a reduction of P_R facilitates the recovery from congestion.

A given gate selection, however, is optimal only for a given traffic pattern. If the pattern changes with time (as it does in our model) the gate values must be dynamically readjusted to maintain efficiency and to avoid instability.

4. A CLOSED LOOP CONTROL (CLC) APPROACH

The brief survey presented in the previous section shows that the major limitation of the existing schemes is the inability to adjust to significant traffic pattern changes.

The need for efficient controls that can adjust to time-varying traffic patterns therefore stimulated research in new directions. In particular, the property that the maximum throughput in an S-ALOHA channel is obtained when the sum of transmission and retransmission rates is equal to unity[1] suggested some interesting possibilites for optimal channel control.

More precisely, Abramson showed that in an S-ALOHA channel with random transmissions, with any number of stations and with any traffic pattern, the maximum throughput is obtained with G = 1, where G is the average sum of transmission and retransmission attempts per slot (including conflicts).

A simple control approach based on the G = 1 optimality condition then suggested itself. This approach consists of measuring G over a proper history (window into the past) and adjusting the ALOHA gates upwards or downwards so as to nullify the error $\varepsilon \triangleq G - 1$, using a closed loop mechanism.*

Because of the distributed implementation of the closed loop control procedures, it is required that each station broadcast its current gate value, P_N. This is done by stamping the value P_N in the packet header at the time of transmission (or retransmission). Gate value broadcasting is necessary in order to maintain all P_N's identical in all stations and to avoid capture by stations whose P_N is higher than average.

In heavy traffic the controls maintain the equilibrium condition G = 1, thus achieving optimal channel utilization for any traffic pattern. In light traffic conditions, the value G = 1 obviously cannot be reached; in this case the effect of the closed loop controls will be to open the gates completely ($P_N = 1$) so as to minimize delay. The system is protected from congestion since a sudden channel overload causes prompt reduction of gate values and therefore a reduction of the load.

Fairness is guaranteed by the fact that all P_N's are equal.

Three closed loop algorithms based on different assumptions on channel load information may be considered:

a. Complete Knowledge (CK): Each station can distinguish empty slots from collision slots. Futhermore, the station can determine the number of collided packets in a collision slot. Since the latter assumption is not very realistic, the CK algorithm mainly has theoretical value and serves as a term of

* A closed loop control approach similar to the one described here was proposed independently by Banh-Tri-An[2]. The Banh-Tri-An model, however, assumes collision detection and infinite population in order to estimate the channel load G.

comparison for the other algorithms.

b. Collision Detection (CD): Each station can distinguish empty slots from collision slots.

c. Collision Non-Detection (CND): Stations cannot distinguish empties from collision (except for their own collisions).

These algorithms are defined in detail below.

5. CLC ALGORITHM DEFINITION

5.1 Notation

First we introduce the parameters to be used in the definition of the algorithms:

W: History window (measured in slots) maintained by each station

E: Empty slots in W

S_i: Successful packets from station i in W

$S = \Sigma S_i$: Total successes in W

C_i: Collisions suffered by station i in W

C: Total number of collision slots in W

$\overline{m}$: Average number of collided packets per collision

UC: Interval (in slots) between successive updates of control parameters (note: UC $\leq$ W).

$G = (S+C\overline{m})/W$: Average channel load in window W

DP: Gate value increment

$P_N^{(i)}, P_R^{(i)}$: New transmission and retransmission gate values for station i

$\overline{P} = (\sum_i P_N^{(i)} S_i)/S$: Weighted average of current gate values

5.2 Closed Loop Control (Complete Knowledge) Algorithm

Assumptions: Each station monitors S,C, and $\overline{m}$ and calculates $\overline{P}$.

Algorithm: Every UC slots, station i updates its parameters as follows:

(a) Calculate: $G = (S + C\overline{m})/W$

(b) Let: $\begin{cases} P_N^{(i)} = \overline{P} - (G - 1)DP \\ P_R^{(i)} = \min (P_N^{(i)}, 1/\overline{m}) \end{cases}$

where $0 \leq P_R^{(i)} \leq P_N^{(i)} \leq 1$

Parameters: W: History window

UC: Update interval

DP: Probability increment

$P_{NO}^{(i)}, P_{RO}^{(i)}$: Initial gate values

Notice that we impose $P_R^{(i)} \leq P_N^{(i)}$ in order to exercise the closed loop control also on retransmissions. Furthermore, we impose $P_R^{(i)} \leq 1/\overline{m}$ to insure a reasonable probability of success among the $\overline{m}$ stations engaged in retransmissions after a conflict.

5.3 Closed Loop Control (Collision Detect) Algorithm

Assumptions: Each station monitors S, S_i $\forall$ i and C; and calculates $\overline{P}$.

Algorithm: Every UC slots, station i updates its gate values with the following sequence of steps:

(a) Estimate $\overline{m}$ as follows:

Define the equivalent number of users N_e as:

$$N_e = \left\lceil \frac{S}{\max S_i} \right\rceil \text{ for } S > 0; \; N_e = 1, \text{ for } S = 0$$

Obtain $\overline{m}$ from N_e by table look-up (See Appendix A)*

(b) Estimate G: $G = (S + C\overline{m})/W$

(c) Let: $\begin{cases} P_N^{(i)} = \overline{P} - (G - 1)DP \\ P_R^{(i)} = \min (P_N^{(i)}, 1/\overline{m}) \\ \text{where } 0 \leq P_R \leq P_N \leq 1 \end{cases}$

Parameters: Same as for Complete Knowledge Algorithm (see Section 5.2).

5.4 Closed Loop Control (Collision Non-Detect) Algorithm

Assumptions: Each station monitors S, counts its own successes S_i, and its own collisions, C_i; and calculates $\overline{P}$.

Algorithm: Every UC slots, station i updates its parameters with the following steps:

(a) Estimate collisions:

If $C_i = S_i = 0$, then:

If S = 0, let G = 0, go to (c)
Otherwise, let G = 1, go to (c)

If $C_i > 0$ and $S_i = 0$,
let $G = G_{MAX}$, $\overline{P} = \min (\overline{P}, P_N^{(i)})$
and go to (c)

If $S_i > 0$, let $C' = (C_i/S_i)S$ and go to (b)**

(b) Estimate total channel load G:

$G = (S + C')/W$
where $0 \leq G \leq G_{MAX}$

(c) Let: $\begin{cases} P_N^{(i)} = \overline{P} - (G - 1)DP \\ P_R^{(i)} = \min (P_N^{(i)}, PRMAX) \end{cases}$

where $0 \leq P_R^{(i)} \leq 1$; $0 \leq P_N^{(i)} \leq 1$

Parameters: W: History window

UC: Update interval

DP: Probability increment

G_{MAX}: Ceiling value of G estimate

$P_{NO}^{(i)}, P_{RO}^{(i)}$: Initial probability values

PRMAX: Ceiling value for P_R

5.5 Typical Parameter Values

The following values and ranges of the control parameters were used in the experiments reported below:

W: History window = (16, 64) slots

UC: Update interval = (16, 64) slots

DP: Probability increment = (.1,.5)

---

* A less accurate, but in most cases acceptable, approximation is $\overline{m} = 2$, $\forall N_e$. In the simulation experiments, we will always use $\overline{m} = 2$.

** C_i/S_i is the estimate at station i of the ratio of collisions vs successes in the channel. By multiplying such ratio by S (total successes in W), we obtain the estimate of total number of collisions in W.

$P_{NO}^{(i)}$, $P_{RO}^{(i)}$: Initial gate values = (0.,.5)

G_{MAX}: Ceiling value for G estimate = 2

PRMAX: Ceiling value for P_R = .5

In our experiment the slot is 0.03 sec, and the uplink and downlink propagation delay R is .250 sec or approximately 8 slots. It is desirable to have W > UC > R for a smooth channel load estimation and feedback control.

6. EXPERIMENTAL RESULTS

6.1 General

A family of simulation runs was made to evaluate and compare the performance of different control schemes under various traffic patterns and for different system configurations. The tests included the study of convergence to the optimum steady state solution and the dynamic behavior under pulse traffic pattern and cyclic traffic pattern requirements

System configurations with 5 and 20 stations respectively were considered. Each simulation run lasted on the order of 5,000 to 10,000 slots. This interval was deemed adequate to determine performance trends and to study the dynamic behavior of the various schemes.

The typical variables monitored during a simulation run and displayed in the plots are: the throughput S (successful packet transmissions/slot); the channel load estimate G at the first station (total transmissions/slot); and the gate value $P_N^{(1)}$ of the first station (or the average gate value $\overline{P}_N$ over all the stations).

6.2 Convergence

A series of runs were made to verify that the steady state solution given by the controls is close to the optimal solution predicted by theory, and to measure the time required to converge to such steady state solution, starting from different initial gate values.

Figures 1 and 2 display the convergence of the CLC-CD algorithm in a 5-station configuration under heavy load (new packet generation rate = 1 pkt/slot/station), starting from two different choices of initial gate values, namely P_{Oi} = 0 and P_{Oi} = 1, respectively. For the control parameters DP, UC and W, the following values were experimentally chosen: DP = .1, UC = 16, W = 64.

Letting $P_N = P_R = P$, the equilibrium throughput S predicted by theory is: $S = 5\,P\,(1-P)^4$. The maximum throughput is S = .41 and is obtained by setting $\underline{P} = .2$. The simulation results in Fig 1 and 2 show that at equilibrium the CLC-CD algorithm achieves optimal throughput and optimal transmission gate values regardless of the initial conditions used.

Now consider the start-up period. We notice that the channel throughput S reaches the equilibrium value S = .41 in an interval ranging from 150 to 300 slots for both experiments. For the parameters G and $P_N^{(1)}$, on the other hand, the convergence to the steady state values of 1 and .2 respectively is slower and is characterized by a damped oscillatory pattern. The fluctuations of G and $P_N^{(1)}$ are correlated, as expected, and have a period of approximately 200 slots. There clearly is a relationship between the oscillation period and the parameters and time constants of the closed loop control system. However, such a relationship is very difficult to find analytically due to the non linearity of the system.

It is interesting to observe that the oscillations of S are much less pronounced than those of G and $P_N^{(1)}$. This is explained by the fact that G and $P_N^{(1)}$ are estimates evaluated at one individual station while S, the total throughput, is the sum of several contributions whose fluctuations are only loosely correlated with each other.

Figures 3 and 4 show the behavior of a 5-station system with an unbalanced traffic condition ($R_1 = 0.4$; $R_2 = R_3 = R_4 = R_5 = 0.01$) using the CLC-CD algorithm and starting from different initial gate values. The initial gate value is $P_0 = 1$ in Figure 3, and $P_0 = 0.2$ in Figure 4. The optimal equilibrium gate value predicted by theory is $P_N = 1$ due to the fact that there is only one major traffic source with little conflicting traffic. Both runs converge to the optimal gate value in the same number of slots, i.e., 300 slots. Considering that in the above experiments the channel is not saturated, these results confirm the assertion that the CLC scheme is optimal also for non-saturated channel conditions.

Experiments identical to those shown in Figures 1 through 4 were repeated for the CLC-CK and CLC-CND algorithms. The results of the latter schemes were surprisingly close to the results obtained with the CLC-CD algorithm. This fact indicates that the accuracy of collision estimates does not have a critical impact on control algorithm performance.

6.3 Cyclic Pattern Experiments

In order to test the ability of the controls to adjust to rapid traffic pattern changes we developed a feature known as the "cyclic generator" in our simulator. This cyclic generator is a message generator whose rate can be changed at prescheduled times during the simulation run, thus producing cyclic patterns of desired amplitude and frequency.

The cyclic generator was used in a 20-station configuration to generate the traffic shown in Table 1. In this pattern, two stations generate packets constantly at the maximum rate (i.e., 1 pkt/slot/station), while the remaining 18 stations periodically generate a pulse of maximum rate and duration = 20 slots every 2,000 slots, and are silent for the remainder of the time.

Number of Stations: 20

Traffic Pattern:

R_1	1	1	1	1	1
R_2	1	1	1	1	1
$R_3 = R_4 = \ldots R_{20}$	0	1	0	1	0
	2000	20	2000	20	2000 SLOTS

S-ALOHA SCHEME	THROUGHPUT S (packets/slot)
Uncontrolled: ($P_N = .5$, $P_R = .05$)	.303 (simulated)
CLC-CND Control Algorithm	.413 (simulated)
Optimal (theoretical) Control Algorithm	.435 (theory)

TABLE 1. CYCLIC PATTERN EXPERIMENT: 20 STATIONS

Two different schemes were tested: the uncontrolled scheme and the CLC-CND scheme. The throughput results reported in Table 1 show that throughput is improved by 30% (from S = .303 to S = .413) by introducing controls. The optimal control for the above cyclic pattern consists of using $P_N = P_R = .05$ while 20 stations are backlogged, and $P_N = P_R = .5$ when only 2 stations are active. The theoretical optimum is S = .435. Thus, the performance of the CLC control is only 5% below optimum.

Figures 5 and 6 show the total throughput S, the gate value P_N of the first station, and the channel load G for the uncontrolled and controlled case respectively. Note in Figure 6 that the gate value P_N moves between the value 0.5 (optimum for 2 stations) and the value 0.05 (optimum for 20 stations). It is not surprising that value $P_N = 0.05$ is maintained for approximately 1000 slots, i.e., the time required to deliver the 20 x 20 = 400 packets introduced in the network during the 20-slot burst.

6.4 Pulse Generator

An important performance measure for a stability control algorithm is the time to recover from a pulse burst. A pulse in the input rates may drive a system to instability even though the normal operating conditions are stable. In this section we study the recovery phenomenon.

Table 2 shows a traffic pattern which contains a pulse of duration = 20 slots, between slot 1000 and slot 1020. The pulse is applied to all stations in a 20 station network configuration. A variety of control algorithms were tested. For each algorithm we measured the throughput performance during the first 1000 slots (before the pulse), the performance during the 4000 slots following the pulse and the time to recover from the pulse (i.e., the time required to deliver all packets generated during the pulse). Simulation results as well as theoretical upper bounds on performance are reported in Table 2.

Number of Stations: 20

Traffic Pattern:

	0–1000	1000–1020	1020–5000
$R_1 = R_2$	1	1	1
$R_3 = R_4 = \ldots R_{20}$	0	1	0

0 1000 1020 5000 SLOTS

ALGORITHM	THROUGHPUT BEFORE PULSE S (0-1000 slots)	THROUGHPUT AFTER PULSE S (1000-5000 slots)	TIME TO RECOVER (slots)
Uncontrolled $P_N = P_R = .05$	.1	.158	1,400
Uncontrolled $P_N = P_R = .5$	.5	.003	∞
Uncontrolled $P_N = .5, P_R = .05$	.327	.353	1,400
Control Limit $P_N = P_R = .5$ If G<1.5; $P_N = P_R = .05$ If G>1.5	.5	.367	1,900
CLC - CND	.455	.466	1,400
Theoretical Optimum	.5	.480	1,100

TABLE 2. PULSE PATTERN EXPERIMENT: 20 STATIONS.

The conservative uncontrolled scheme with $P_N = P_R = 0.05$ is stable, but gives poor performance (see Figure 7). The uncontrolled scheme with $P_N = P_R = 0.5$ gives optimal performance before the pulse when only two stations are active, as expected; but it virtually collapses into an unstable mode after the pulse, with zero throughput (see Figure 8). The uncontrolled scheme with P_N optimized for the normal operating condition with two active stations (i.e., $P_N = 0.5$) and P_R optimized for the heavy load condition with 20 active stations (i.e., $P_R = 0.05$) shows better performance than the previous schemes, but is still well below optimum (see Figure 9).

For the control limit scheme, the critical load was set to $L_c = 1.5$, and the gate values were chosen as $P_0 = .5$ and $P_c = .05$; i.e., the optimum at steady state (2 active stations) and the opimum during congestion (20 active stations) respectively. The control limit scheme shows poor performance after the pulse because of the oscillatory behavior of P_N predicted in Section 3.1. The oscillations are triggered by the large backlog (up to 20 packets) at each station. As a consequence, the time to recover is larger than for the other schemes, and the performance after the pulse is substantially less than optimal (see Figure 10).

The CLC-CND scheme shows the best performance among all the other schemes after the pulse, with a throughput only 5% below the theoretical optimum (the optimal control assumes $P_N = P_R = .05$ with 20 active stations, and $P_N = P_R = .5$ with 2 active stations). The fact that the performance before the pulse is not as good as that of other schemes is due to the 300 slot start-up interval required to adjust the gates from the initial value $P_N = 1$ to the optimal value $P_N = 0.5$ (notice that if the initial value $P_N = .5$ instead of $P_N = 1$ were used, optimality would have been achieved also in the pre-pulse period). From Figure 11 we notice that the gate P_N is rapidly reduced from 0.5 to 0.05 after the occurrence of the burst (by decrements of DP = 0.1 every UC = 32 slots), and is restored to the optimal steady state value, $P_N = 0.5$, after the congestion has been cleared. The time to recover (1,400 slots) is slightly higher than the theoretical upper bound (1,100 slots). This is attributed to the fact that the upper bound assumes that the queues of the 18 "bursty" stations clear all at the same time, while in reality there is a gradual phasing out, lasting on the order of a few hundred additional slots.

The above results show that the Closed Loop Control scheme is superior to all the other schemes in handling pulse traffic situations. The result is especially remarkable when we consider that the parameters of both uncontrolled schemes and Control Limit scheme were specifically optimized for the traffic pattern under consideration, while no prior tuning was required for the Closed Loop Control scheme.

7. CONCLUSION

We briefly reviewed the existing ALOHA stability control algorithms and showed that they are not adequate (in their present form) for a general satellite network model because of the following restrictive assumptions:

- Single buffer stations
- Time invariant traffic pattern and traffic rate.

We then proposed a closed loop control approach which attempts to drive the total channel load G to 1 (where G is the sum of transmission and retransmission rates over all stations), based on the well-known result that the performance of a heavily loaded S-ALOHA channel is optimized for G = 1. Finally, we showed that the Closed Loop Control yields optimal performance also in nonsaturated channel situations.

Two closed loop, full distributed control algorithms were introduced, namely:

- Collision Detect (CD) algorithm, which requires distinction of collision slots from empty slots, and
- Collision Non-Detect (CND) algorithm, which does not distinguish collisions

from empties.

Comparing the two algorithms we notice that memory and CPU resources required for the implementation are approximately equivalent. Storage must be provided for the history window in both algorithms. At each update the station must count S, S_i, V and C (for the CK algorithm), or S and C_i (for the CND algorithm). The average $\overline{P}$ is then evaluated, and a few steps lead to the estimate of G and to the updated gate values, $P_N^{(i)}$ and $P_R^{(i)}$.

Channel overhead is slightly increased (with respect to uncontrolled S-ALOHA) by the requirement of carrying the current value of the gate P_N in each packet. Since an 8-bit character is sufficient to encode P_N, the additional overhead is less than 1% with 1000 bit packets.

A family of simulation runs was made to evaluate and compare the performance of different control schemes in a variety of traffic conditions. Simulation results show that the control algorithms converge to near optimality at steady state. Furthermore, the performance of the CND algorithm is about equivalent to that of the CD algorithm, thus indicating that the capability of distinguishing collisions from empties is not critical for the closed loop control of S-ALOHA channels.

Sensitivity studies were carried out to determine the impact of control parameter values (i.e., DP, W, etc.) on performance[5]. The results show that while the behavior of each individual station is rather sensitive to the choice of parameters (e.g. gate value P_N fluctuations became very pronounced if DP = .5 or W = 8), the behavior of global system variables (e.g. total throughput or average gate value $\overline{P}$) is rather insensitive to parameter changes. This fact suggests the possibility, for example, to increase DP in order to reduce the time to converge to steady state without compromising global performance; or to adjust DP dynamically, depending on the nature of the input traffic fluctuations.

The stability properties of the control algorithms and their superiority over uncontrolled schemes for time varying load patterns were demonstrated in cyclic pattern and pulse pattern experiments. Among the controlled schemes, the CLC algorithm displayed a better performance than other algorithms previously proposed (namely, the Retransmission Control and the Control Limit schemes).

The experiments discussed in this paper are based on stations with a large buffer pool and a predominantly batch traffic environment (heavy load). Experimentation is currently under way on system configurations with smaller buffer sizes and various mixes of traffic classes (e.g., interactive, batch, etc.). Partially controlled systems, where only a subset of the stations is controlled while the remaining stations have fixed gate values, are also being considered.

Also, we are exploring the possibility of applying the closed loop control approach to other random access schemes. One promising area is the Unslotted ALOHA channel control[1]. Another important application is the control of the contention subframe in reservation schemes in which reservations are placed in an S-ALOHA mode [6,10]. It appears that a simplified version of the CLC scheme may be suitable for the latter type of applications.

ACKNOWLEDGEMENTS

The authors are grateful to Dave Clemens for his assistance in the development of simulator, and to Lou Nelson for many useful comments and suggestions.

REFERENCES

1. Abramson, N., "Packet Switching with Satellites," National Computer Conference, AFIPS Conference Proceedings, Vol. 42, Montvale, New Jersey: AFIPS Press, pp. 695-702, 1973.

2. Banh-Tri-An and E. Gelenbe, "Étude et Simulation de Réseaux de Diffusion Sous Contrôle," IRIA, Le Chesnay, France, Rapport de Recherche No. 206, December 1976.

3. Ferguson, M., "On the Control, Stability, and Waiting Time in a Slotted ALOHA Random Access System," University of Hawaii, Honolulu, The ALOHA System Technical Report B74-9, December 1974.

4. Fayolle, G., E. Gelenbe, J. Labetoulle, "Stability and Control of Packet Switching Broadcast Channels," IRIA, Le Chesnay, France, Rapport de Recherche No. 116, April 1975.

5. Gerla, M., "S-ALOHA Satellite Channel Stability Control Algorithms and Simulation Experiments,: UCLA Packet Satellite Program Working Note #46, December 3, 1976.

6. Jacobs, I., L. Lee, A. Viterbi, R. Binder,, R. Bressler, N. Hsu, R. Weissler, "CPODA-A Demand Assignment Protocol for SATNET", Fifth Data Communications Symposium, Snowbird, Utah, Sept 27-29, 1977.

7. Kleinrock, L. Queueing Theory, Vol. II: Computer Applications, Wiley Interscience, New York, 1976.

8. Lam, S. and L. Kleinrock, "Dynamic Control Schemes for a Packet Switched Multi-Access Broadcost Channel", National Computer Conference, AFIP Conference Proceedings, Vol. 44, pp. 143 - 153.

9. Lam, S. and L. Kleinrock, "Packet Switching in a Multiaccess Broadcast Channel: Dynamic Control Procedures," IEEE Transactions on Communications, Vol. COM-23, No. 9, pp.891-904, September 1975.

10. Roberts, L., "Dynamic Allocation of Satellite Capacity through Packet Reservation," National Computer Conference, AFIPS Conference Proceedings, Vol. 42, Montvale, New Jersey: AFIPS Press, pp. 711-716, 1973.

APPENDIX A

1. Evaluation of $\overline{m}$ = average number of collided packets per conflict.

Recall

$$G = \frac{S + C\overline{m}}{W} \qquad (A.1)$$

or

$$\overline{m} = \frac{WG - S}{C} = (G - S/W)/(C/W) \qquad (A.2)$$

Let: P_i, i 1,...,N be the transmission probability of station i.

We have

$$\begin{cases} G = \sum_i P_i \\ S/W = \sum_i P_i \prod_{j\neq i} (1 - P_j) \\ E/W = \prod_i (1 - P_i) \\ C/W = 1 - \frac{S + E}{W} = 1 - \sum_i P_i \prod_{j\neq i} (1-P_j) - \prod_i (1-P_i) \end{cases} \qquad (A.3)$$

Thus, $\overline{m}$ can be calculated readily, given the P_i's, by substituting the expressions in (A.3) into Eq. (A.2).

Of interest is the evaluation of $\overline{m}$ for G = 1, since G = 1 is the optimal operating point. For an infinite population and G = 1, recall that S/W = E/W = 1/e.

Thus from Eq. (A.2)

$$\overline{m} = \frac{1 - 1/e}{1 - 1/e - 1/e} = \frac{e-1}{e-2} = 2.39$$

The value of $\overline{m}$ for an infinite population is clearly an upper bound on the admissible values of $\overline{m}$.

The obvious lower bound on $\overline{m}$ is 2. Thus for $G = 1$, we have:

$$2 \leq \overline{m} \leq 2.39$$

Now consider the case of N active users with balanced traffic and note that $P_i = 1/N$ in order to satisfy the condition $G = 1$. We have:

$$\overline{m} = \frac{1 - (1 - 1/N)^{N-1}}{1 - (1 - 1/N)^{N-1} - (1 - 1/N)^{N}}$$

Some values of $\overline{m}$ as a function of N are reported below:

N	2.	3	4	7	10	∞
$\overline{m}$	2	2.19	2.20	2.28	2.31	2.39

The above table may be stored in the station and used by the Collision Detect algorithm to estimate $\overline{m}$, given the equivalent number of users N_e. (Note: if $N_e = 1$, we define $\overline{m} = 2$.)

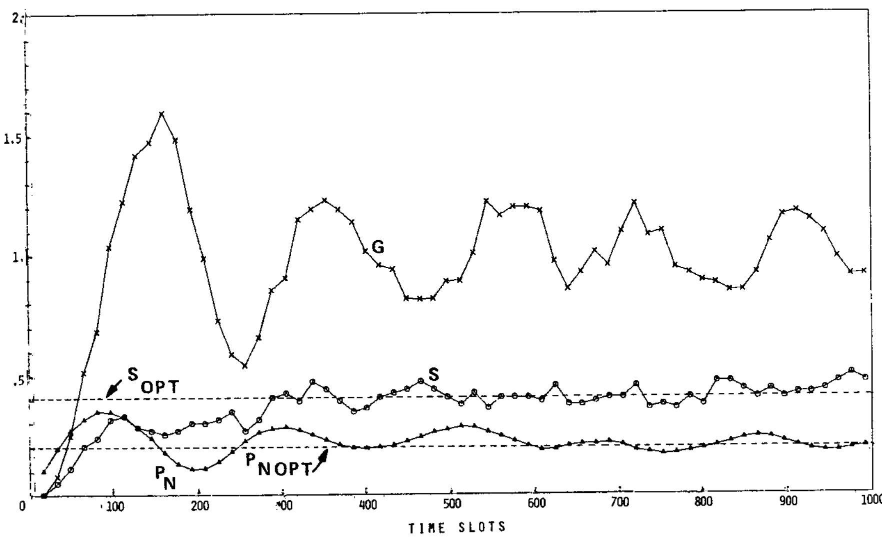

Figure 1. Convergence; CLC-CD Algorithm; 5 Stations; $P_{0i} = 0$, ∀ i.

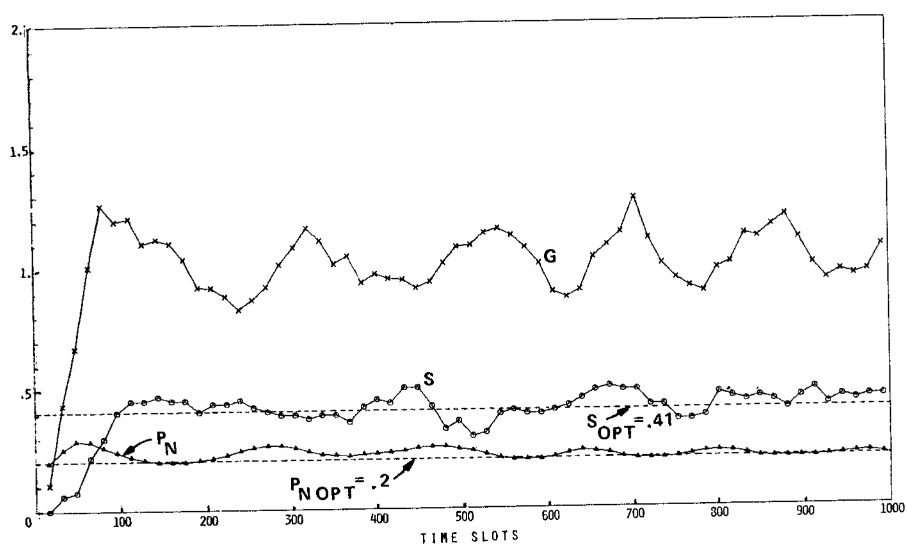

Figure 2. Convergence; CLC-CD Algorithm; 5 stations; $P_{0i} = 1$, ∀ i.

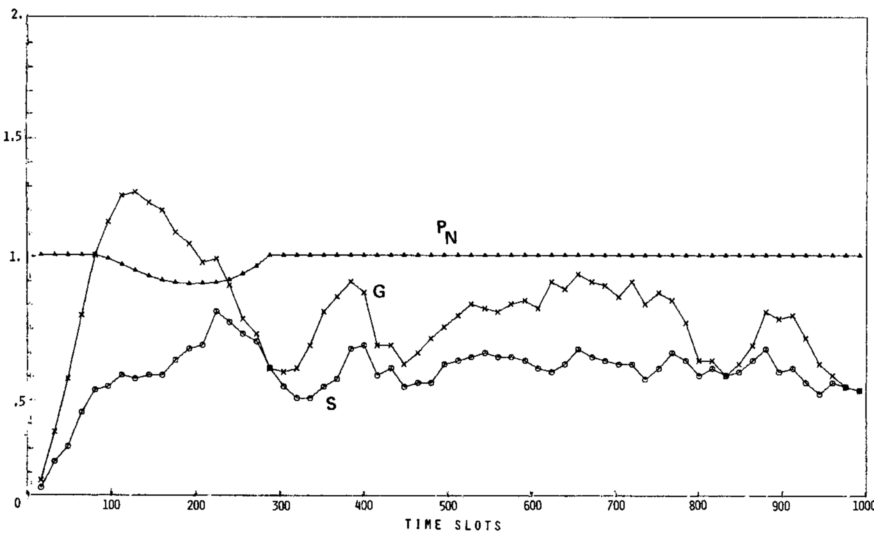

Figure 3. Convergence; CLC-CD Algorithm; 5 Stations; $P_{0i} = 1 \; \forall \; i$

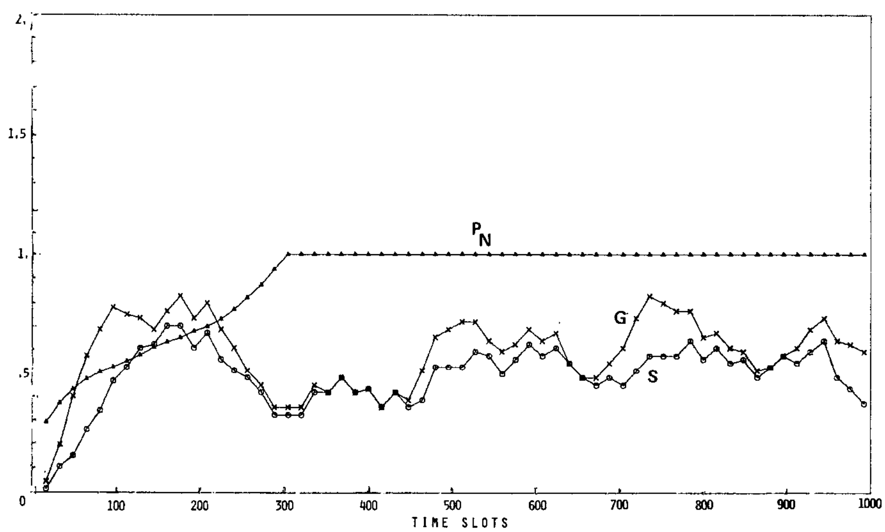

Figure 4. Convergence; CLC-CD Algorithm; 5 Stations; $P_{0i} = .2 \; \forall \; i$

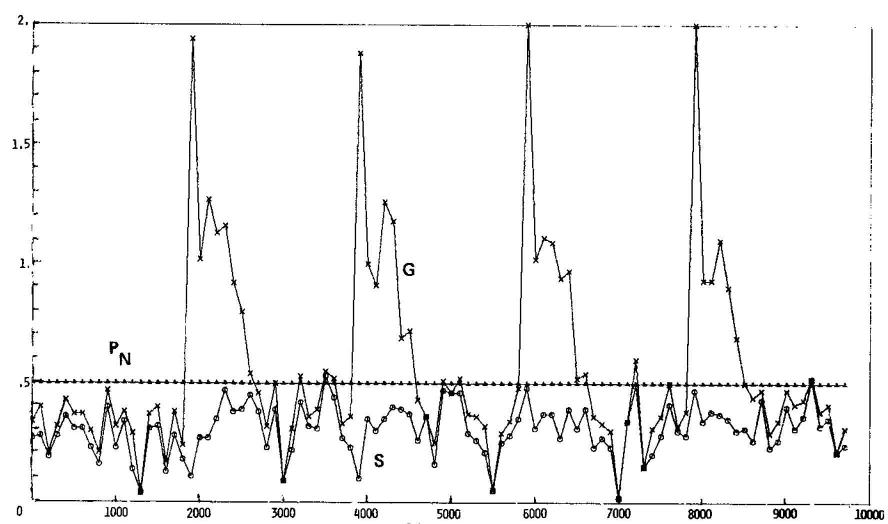

Figure 5. Cyclic Pattern; 20 Stations; Uncontrolled S-ALOHA

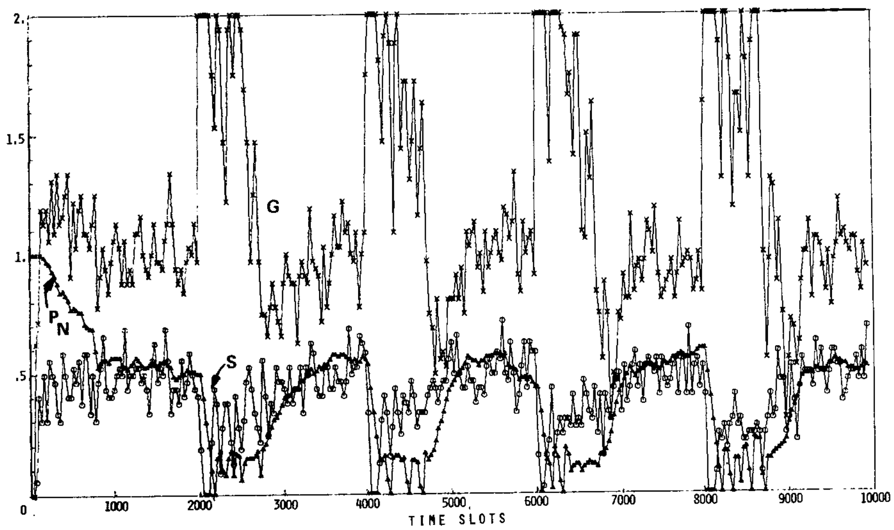

Figure 6. Cyclic Pattern; 20 Stations; CLC-CND Algorithm

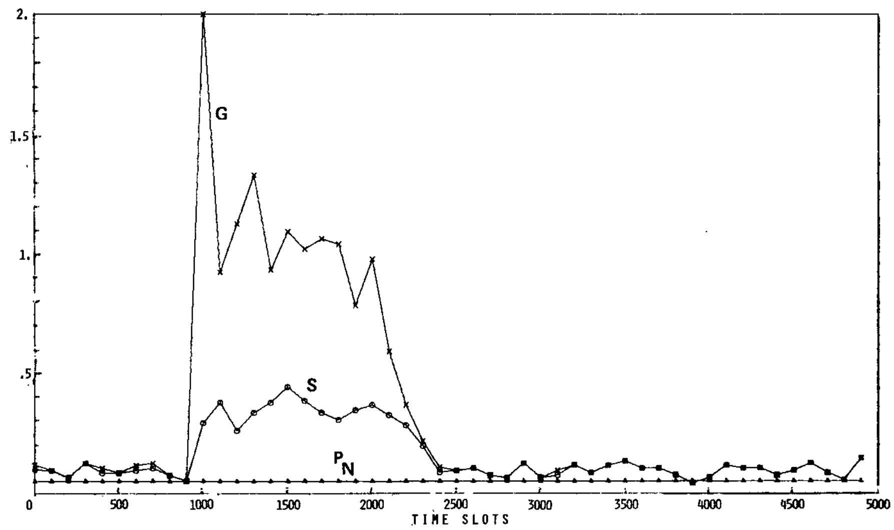

Figure 7. Pulse Pattern; 2 Stations; Uncontrolled ($P_N = P_R = .05$)

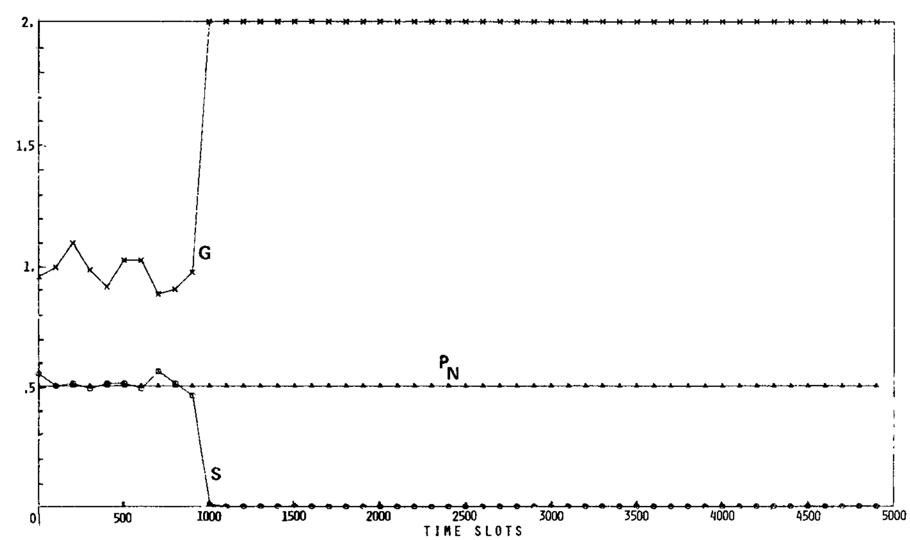

Figure 8. Pulse Pattern; 20 Stations; Uncontrolled ($P_N = P_R = .5$)

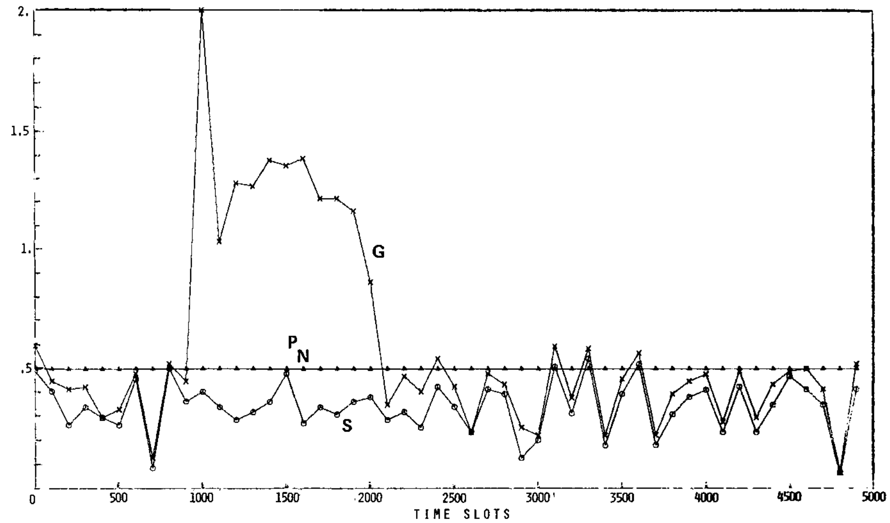

Figure 9. Pulse Pattern; 20 Stations; Uncontrolled ($P_N = .5$; $P_R = .05$)

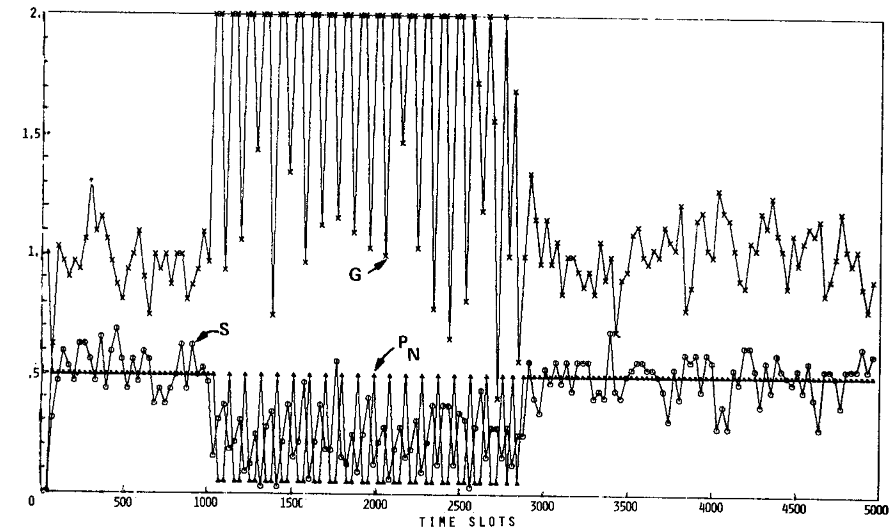

Figure 10. Pulse Pattern; 20 Stations; Control Limit $\begin{cases} P_N = P_R = .5 \text{ if } G < 1.5 \\ P_N = P_R = .05 \text{ if } G > 1.5 \end{cases}$

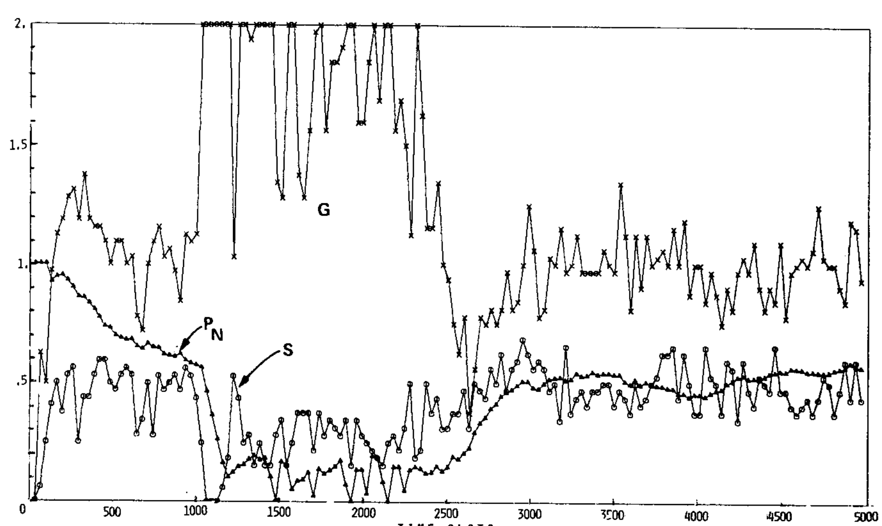

Figure 11. Pulse Pattern; 20 Stations; CLC-CND Algorithm