

Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks

Diane E. Duffy, Allen A. McIntosh, Mark Rosenstein, Walter Willinger

Bellcore
445 South Street
Morristown, NJ 07690-6438

Abstract

In this paper we report on an ongoing statistical analysis of actual CCSN traffic data. The data consist of approximately 170 million signaling messages collected from a variety of different working CCS subnetworks. The key findings from our analysis concern: (1) the characteristics of both the telephone call arrival process and the signaling message arrival process, (2) the tail behavior of the call holding time distribution, and (3) the observed performance of the CCSN with respect to a variety of performance and reliability measurements.

1. Introduction

Current and future telecommunications network architectures and services are strongly dependent on the Common Channel Signaling Network (CCSN) and the Signaling System Number 7 protocol (SS7) for communication. Important applications include 800 Database Service, Line Information Database (LIDB) applications, National Integrated Services Digital Network (N-ISDN), Advanced Intelligent Network (AIN) applications, and Personal Communications Services (PCS). Rising penetrations of these advanced services, as well as recent regulatory decisions are leading to increases in CCSN deployment and utilization.

CCSNs need to be well designed, carefully evaluated, and strategically managed to ensure adequate performance and quality of service at reasonable cost. Engineering and modeling are two main technologies for designing, evaluating and managing telecommunications networks such as the CCSN. For example, CCSN engineering guidelines have been developed for sizing links, linksets, and Service Control Points (SCPs). In order to meet performance objectives and to utilize the capital deployed in the CCSN efficiently, these engineering guidelines should be based on a solid foundation of knowledge about the network and its traffic. Similarly, models, both analytic models and simulation models, are currently used to compare alternative implementation strategies for new services and applications, to assess network integrity and reliability, and to evaluate network performance. Again, in order to assess network strategies, integrity and performance reliably, models should be based on realistic assumptions about the network and its traffic.

In this paper, we present results of an ongoing statistical analysis of SS7 data from actual working CCS subnetworks of local exchange carriers (LECs). We focus on a careful and comprehensive analysis of SS7 message level CCSN data, and on informed summaries and syntheses of data-analytic results. The insights gained from our analyses provide accurate information about the CCSN and its traffic. While some of our findings are already being used to validate and improve CCS engineering and modeling, possible implications for CCSN planning and engineering of other findings of our analysis require further examination (e.g., sizing D-links to ICs, modeling SCP delay, analyzing transmit buffer congestion thresholds).

Our results are based on a rigorous statistical analysis of tens of millions of individual time-stamped SS7 messages on links monitored at the Signaling Transfer Point (STP). To date we have collected traffic measurements from four different CCS subnetworks, resulting in data from a variety of link types, from a range of different link utilizations, from a number of different geographic areas, and from business, residential and mixed end offices (EOs) — see Table 1. In total, more than 7 gigabytes of binary data have been collected. The size and complexity of these data sets require sophisticated software and methodology for data management, data translation and manipulation, and data analysis, both computational and graphical. We have developed a software library that permits efficient and effective *exploratory* analysis of the data. By *exploratory* analysis, we mean that the data can be manipulated in a flexible fashion to explore factors, such as time of day, system load, link type, and type of network element (i.e., hardware and software configuration), on quantities of interest, such as system response time, performance measures, and delay components. Our analyses range from sophisticated investigations of the nature of the CCSN traffic process, to empirical assessments of network performance and descriptive summaries of message types and frequencies.

The remainder of this paper is organized as follows. In Section 2 we review the data collection process, describe the data sets we have collected, and summarize our analysis environment. In Section 3 we investigate the nature of the CCSN traffic process. Our key findings concern the nonhomogeneity of the traffic rate, the correlation structure in the traffic process, and the heavy-tailed nature of the underlying call holding time distribution. In Section 4 we summarize some interesting empirical findings for a variety of network performance measures. Section 5 contains concluding remarks.

2. CCSN Traffic Data at the SS7 Message-Level

The data presented in this paper were collected by a pair of custom-built monitoring devices, installed at mated STPs, that non-intrusively monitor a maximum of 16 bi-directional SS7 signaling links. The devices are time synchronized to allow the integration of data collected across the mated pair of STPs and are configured to capture and time stamp, with millisecond accuracy, all SS7 messages in both directions of a monitored link except for FISUs. Each device is equipped with a 128 Megabyte disk. The duration of data collection is limited by the available disk space and there is a tradeoff between the number of links monitored and the length of the monitoring period.

To date we have collected data from four different CCS subnetworks of LECs. Table 1 gives a summary description of these measurements. Information is provided about the time period over which data were collected, the number of data sets that were obtained, the approximate number of messages collected, the types of signaling links that were monitored,

typical traffic loads observed, and some features of special interest. Our data are diverse with respect to network element type, in that we have hooked up the monitoring devices to different vendors' STPs and we have monitored links to different EO switches and access tandems (ATs). While we have a great variety of data on A-links to EOs and some variety in data on A-links to ATs and on D-links, our data on A-links to SCPs, on B-links, and on C-links is not yet geographically diverse.

In binary form our data take up more than 7 gigabytes; ASCII translations are considerably larger, up to an order of magnitude bigger depending on the level of detail. In order to support both a wide variety of analyses and sophisticated analyses, where warranted, it is imperative that we are able to manipulate the data flexibly and quickly. To this end, we have developed a software library in C for handling these data. This library is designed on the basic principle of taking a stream of binary packets as input, doing some selection and/or calculation, and writing a stream of output. The output is typically another stream of binary packets. The extensive use of binary files saves considerably on disk space. Since the computing is quite fast, intermediate files can be quickly regenerated and need not be saved.

3. Characterization of CCSN Traffic

SS7 message level data provide a unique opportunity to compare actual CCSN traffic data with some of the commonly made theoretical assumptions about CCSN traffic. In this section we study the message arrival process (i.e., all SS7 messages) and the call holding time distribution (i.e., the time between an ANM message and the end of the call). To simplify the exposition, we will illustrate our results exclusively with data collected on Subnetwork II. However, similar results have been obtained using data sets from the other subnetworks presented in Table 1 and thus our findings are consistent over time, across a range of loads and in different network configurations.

3.1 CCSN Traffic at the Message Level

Traditionally, CCSN traffic models have been based on the Poisson process assumption. It has been recognized (e.g., Skoog 1991), that this assumption is likely to be invalid: even if call arrivals follow a Poisson process, message arrivals to a signaling link will not be Poisson because message arrivals for a particular call are correlated. A straightforward analysis (not shown here) of call arrivals (obtained by extracting IAM messages) shows that they are statistically well described by a time-inhomogeneous Poisson process, but that the observed degree of time-inhomogeneity in the arrival rate does not support the commonly used engineering notion of a stationary "busy hour" or even of a stationary "busy period" of shorter duration. These observations confirm what traffic engineers have known for a long time. However, when message arrivals, rather than call arrivals, to a signaling link are considered, our analysis of the data reveals a more complex picture and requires some novel techniques for studying traffic.

To illustrate this, Figure 1(a) shows the number of SS7 message arrivals per second on one channel of an A-link to an EO for Subnetwork II over an 8.3 hour time period during Friday morning and afternoon. We use non-parametric smoothing (see Cleveland 1979) as implemented in the S statistics language (Becker, Chambers, and Wilks 1988) to estimate the arrival rate (white line in Figure 1(a)). Plotting the empirical autocorrelations based on the raw data (Figure 1(b)), and on the "detrended" data; i.e., after subtracting the estimated arrival rate

from the raw data (Figure 1(c)), shows that a substantial part of the autocorrelations in Figure 1(b) is an artifact of the non-stationary nature of SS7 message arrivals. However, the autocorrelations for the detrended data (Figure 1(c)) do not "look like" those based on i.i.d. data (Figure 1(d)), but rather show small positive autocorrelations which seem to decay very slowly.

We can check whether the small positive autocorrelations observed in Figure 1(c) are statistically significant with a graphical statistical test called the *variance-time plot* (Cox, 1984). Note here that we are not assessing the significance of the individual autocorrelations but their significance when *summed* over all lags. To calculate the variance-time plot, consider a given "block size" m and create a new count process by averaging the original process over non-overlapping time intervals of length m . The variance-time plot is a plot, on log-log scale, of the variance of the new count process as a function of the block size m . If a count process has positive autocorrelations that decay *exponentially* fast, or if it is independent, then the variance-time plot will have an asymptotic slope of -1 . Larger asymptotic slopes mean that positive autocorrelations decay more slowly than exponential (e.g., like a power).

The variance-time plot for the "detrended" counts of SS7 messages per second is shown in Figure 2. Dark plotted circles denote the stable region of the asymptotic slope; the dotted reference line has slope -1 . The asymptotic slope is about -0.5 ; it is clearly different from -1 and suggests that the autocorrelations decay like a power with exponent -0.5 . (Note that the link test messages that are sent at regular intervals do not explain the observed autocorrelations.) Autocorrelations obeying a power-law not only cannot be modeled with a Poisson process, but they also cannot be modeled in a parsimonious manner (i.e., using a small number of parameters) with any of the Poisson-based models (e.g., batched Poisson, interrupted Poisson, Markov-modulated Poisson). On the other hand, such autocorrelations can be parsimoniously modeled using *long-range dependent* processes; long-range dependence has also been observed in recent studies of other packet traffic measurements (e.g., see Beran et al. 1992, Erramilli and Willinger 1992, and Leland et al. 1993).

While the implications for SS7 traffic engineering of long-range dependence in the message arrival process remain largely unknown, simulation studies using similar traffic (see Fowler and Leland 1991) and some approximate analytic results for queues with long-range dependent arrival streams (see Norros 1992) provide initial indications of the drastic impact of slowly decaying autocorrelations on performance. Moreover, Leland et al. 1993 discuss the impact of long-range dependence on traffic rate measures and on "burstiness" measures and comment on their sensitivity to the length of the measurement interval. For example, it is commonly known that the larger the sample (i.e., the longer the time interval) the smaller the variability of the estimate of the traffic rate. In the absence of long-range dependence, this variability typically decreases like the reciprocal of the sample size; with long-range dependence, the decrease is much slower. This issue is of practical importance because CCSN operations systems typically aggregate traffic rate measurements over long time intervals (15-60 minutes, occasionally 5 minutes). Armed with only these aggregate traffic rate measurements, network engineers become increasingly concerned about the variability of traffic rates over shorter periods (3-10 seconds). Valid inferences about the variability on finer time scales requires a fuller understanding of the impact of long-range dependence in SS7 message arrivals.

3.2 The Heavy-Tailed Nature of Call Holding Time Distributions

Traditionally, telephone call holding times (CHTs) have been modeled using exponential distributions (e.g., Erlang 1918). It has long been recognized that the exponential approximation seriously underestimates the actual numbers of very long calls (e.g., data calls that last for many hours). Recent work has taken a more formal look at models for CHT. For example, Bolotin (1993) studied a year's worth of complete call records for hundreds of individual residential lines in one geographical area, and found that the CHT distribution is well described by a mixture of lognormal distributions. Our discussion here is not focused on fitting a particular model to the observed CHT distribution. Instead we identify general properties exhibited by the actual CHTs. Since we have CHTs for millions of calls, we are able to provide very sensitive assessments of the observed CHT distribution. Figure 3(a) plots the histogram of the logarithm (base 10) of CHT (in seconds) for calls that *started* during the high traffic period shown in Figure 1(a). It is critical that the selected sample of calls not be truncated; i.e., that we look at all calls that *started* in a given time period, not those that started and ended in a given time period. There were 302,225 such calls, and their durations ranged from 0.001 seconds to 29.5 hours.

The presence of a significant number of very long CHTs (typically between 1% - 5%) suggests that the underlying distribution has heavy tails. Formally, a probability distribution function F is called *heavy-tailed* if $1-F(x)$ decays like a negative power of x ; i.e., $1-F(x) \approx x^{-\alpha}$, as $x \rightarrow \infty$ ($\alpha > 0$). Examples of such distributions are the lognormal, Weibull, and Pareto. Note that if $\alpha < 2$, then F has infinite variance (e.g., Pareto with Parameter $1 < \alpha < 2$), and if $\alpha < 1$, then F has infinite mean.

Let $\hat{F}(x)$ denote the empirical distribution function of the CHT. We can assess tail behavior by plotting, on the log-log scale, $1 - \hat{F}(x)$ vs x . If the underlying distribution is heavy-tailed, in the sense defined above, then we see an approximately straight line for large x -values, with slope of $-\alpha$. Indeed, Figure 3(b) exhibits such structure as shown with the plotted solid circles. The slope estimate (obtained by eyeballing a straight line through these points) is about 2.0.

A more rigorous method for estimating α is provided in Hill (1975). To describe Hill's estimate, we use the notation (X_1, X_2, \dots, X_n) for the observed CHTs, and we let $X_{1, n}, X_{2, n}, \dots, X_{n, n}$ denote the corresponding order statistics. Hill's estimate $\hat{\alpha}$ is given by

$$\hat{\alpha}_n = (1/k \sum_{i=0}^{i=k-1} (\log X_{n-i, n} - \log X_{n-k, n}))^{-1}$$

Figure 3(c) plots Hill's estimate vs k , where the k largest order statistics are used to calculate the Hill estimate according to the above formula. Note in particular that the estimate is quite stable, with an $\hat{\alpha}$ -value of about 2.0, for k -values ranging from 500 to 5000 (i.e., using up to about 2% of the calls). Hill's estimate agrees with the earlier conclusion drawn from Figure 3(b): the CHT distribution is heavy-tailed and, in fact, at the borderline of having an infinite variance. A similar analysis of low traffic data (not shown here) indicates that the CHT distribution during low traffic periods is also heavy-tailed (capturing between 10% - 50% of the CHTs); moreover, the corresponding Hill estimate hovers around 1, raising the question of an infinite mean. Note that this property of CHTs during low and high traffic periods is not due to non-stationarity in the data; randomly chosen subperiods exhibit the same phenomenon — a typical feature of heavy-tailed distributions.

The question of whether or not to use CHT distributions with infinite variance/mean in engineering models can be approached in two ways. On the one hand, by declaring enough of the long calls "statistical outliers", it is always possible to obtain a good fit to the remaining observations using distributions such as lognormals and mixtures thereof. On the other hand, one can avoid both the subjective decision of which calls to discard and the loss of a significant fraction of calls (1%-5% during high traffic periods, up to 50% during low traffic hours) by considering infinite variance/mean CHT distributions.

The observed heavy-tailed nature of CHTs can have serious implications for moment matching, a popular method for estimating parameters in an assumed model. To illustrate, consider the 305,225 CHTs mentioned earlier, and recall that our analysis indicated the possibility of an infinite variance for the CHT distribution. Figure 4(a)-(b) shows the sample mean and sample standard deviation for these data as a function of the number of observations (i.e., we successively based the sample moment calculations on larger fractions of the data). Clearly, the sample standard deviation is highly unstable — despite the very large sample sizes, the computed sample standard deviation varies widely. The sudden jumps in Figure 4(b) are caused by the occasional very large observations that are a trademark of distributions with infinite variance/mean.

4. Empirical Analyses of Network Performance Measurements

In addition to studying SS7 message traffic, we used our data for empirical analyses of a variety of network performance measurements. One class of network performance issues concerns delays or processing times at network elements. Examples of these are cross-STP delays, 800 query-response delays, and switch processing delays. Empirical information about these quantities is used to validate network engineering models and to assess system performance. Another class of network performance issues concerns network integrity and reliability. Issues in this class range from assessing packet retransmissions to identifying anomalies such as bad packets, protocol interactions, unusual SS7 message sequences, and routing errors. Finally, there is a class of network performance issues involving events that directly impact customer service. Examples of these are dual seizure and dropped TCAP queries. When a trunk is seized simultaneously by two network elements for two different calls, or when a TCAP query does not receive a response, customers experience delays and/or difficulty in making calls. In this section, we illustrate our empirical analyses of network performance issues with 3 examples: 800 query-response times, a routing anomaly, and some unusual SS7 message sequences.

In order to benchmark an algorithm for SCP engineering, information was needed about 800 query-response times. Table 1 shows that in Subnetwork II we collected data on A-links to SCPs handling 800 queries. Matching the query with the response allows us to calculate the 800 query-response delay. Figure 5 shows the observed distribution of 800 query-response delays (in milliseconds) based on data collected during the morning and early afternoon of 1/19/93. A total of 315028 query-response delays are represented. The key features of this distribution are a sharp mode at 457, and a strongly curved left shoulder of times less than 457 with its own mode of 320. The mean delay is 389 and the standard deviation is 81. The median delay is 412; 90% of the delays are less than 469; 99% of the delays are less than 496. Similar looking distributions were obtained after accounting for the non-stationarity of the query rate over the observed time period. The overall shape of the distribution is expected given the specific SCP architecture.

In data collected from Subnetwork II between 12/31/92 and 2/8/93 we observed a routing anomaly between two EOs. All traffic from the first EO to the second EO was routed through one of the two mated STPs; traffic in the reverse direction was properly balanced across the two mated STPs. The error has been traced to corruption of certain information in the first EO switch, and has led to an audit of all EO switches of that kind.

Finally, in-depth empirical analysis of our data can reveal SS7 message sequences that are clearly pathological but do not appear to violate the SS7 protocol. For example, we have identified a few telephone calls which involve anywhere from several hundred to several thousand SS7 messages. Given that an average call involves 5-7 SS7 messages to set up and tear down, the identified calls have quite excessive signaling loads. The worst call of this type that we have seen to date lasted for 31 minutes and had a total of 5976 SS7 messages, including 2985 pairs of Suspend Messages (SUS) and Resume Messages (RES). An SUS is sent when the called party hangs up. An RES is sent if the called party picks up again quickly (6-30 seconds) after hanging up), and resumes the call. To our knowledge, the SS7 protocol does not limit the number of allowable SUS, RES pairs; nonetheless, calls such as these are clearly anomalies.

5. Conclusions

For gaining a thorough understanding of packet networks such as the CCSN, an integrated approach is needed consisting of (1) high quality traffic measurements, (2) sophisticated traffic data analysis methods, and (3) advanced traffic modeling tools that result in useful traffic and network engineering guidelines.

The main purpose of this paper was to describe a current data collection effort that has resulted in exceptional (with regard to size and quality of the measurements) data sets of SS7 messages and to summarize the results of an ongoing statistical analysis of these traffic measurements. We have shown that SS7 traffic at the message level exhibits long-range dependence and that call holding times may be best described using heavy-tailed distributions with possibly infinite variance/mean. In addition, we have illustrated how empirical analyses of a number of different network performance measurements can reveal useful, interesting and at times absurd information and features. While some of our findings are already being used to validate or improve engineering guidelines, the impact on CCS traffic and network engineering of other findings of our analysis require further examination.

Some of the results we have presented here are meant to be illustrative rather than definitive. The CCSN, the services it provides, and the telecommunications business as a whole, are evolving so rapidly that for some analyses we are chasing a moving target. In our opinion, this does not invalidate the usefulness of empirical data, but it does argue for caution in applying certain results.

6. Acknowledgement

Collecting, analyzing, and interpreting data from working CCS subnetworks is a team effort which has benefited from numerous colleagues at Bellcore and elsewhere. We thank them all. We are, however, particularly grateful to Dave McSherry with whom we collaborate on the physical data collection, and to Paul Tukey for analyzing the 800 query data, and to Sid Resnick (Cornell University) for suggesting the use of the Hill estimate.

7. References

1. R. A. Becker, J. M. Chambers, A. R. Wilks, *The New S Language*, Wadsworth, 1988.
2. J. Beran, R. Sherman, M. S. Taqqu, W. Willinger, "Variable-Bit-Rate Video Traffic and Long-Range Dependence", accepted for publication in *IEEE Trans. on Communications*, subject to revisions, 1993.
3. V. A. Bolotin, "Modeling Call Holding Time Distributions for CCSN Network Design and Performance Analysis" — this *JSAC* issue, 1993.
4. W. S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots", *J. Amer. Statist. Assoc.* 74, 829-836, 1979.
5. D. R. Cox, "Long-Range Dependence: A Review", in: *Statistics: An Appraisal*, H. A. David and H. T. David (Eds.), The Iowa State University Press, Ames, Iowa, 55-74, 1984.
6. A. K. Erlang, "Solutions of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges", *The Post Office Electrical Engineers' Journal* 10, 189-197, 1918.
7. A. Erramilli, W. Willinger, "Fractal Properties of Packet Traffic Measurements", *Proc. of the St. Petersburg Regional ITC Seminar*, pp. 144-158, St. Petersburg, Russia, 1993.
8. H. J. Fowler, W. E. Leland, "Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management", *IEEE Journal on Selected Areas in Communications* 9, 1139-1149, 1991.
9. B. M. Hill, "A Simple General Approach to Inference about the Tail of a Distribution", *Annals of Statistics* 3, 1163-1174, 1975.
10. W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic", *Proc. of the ACM/SIGCOMM '93*, pp. 183-193, San Francisco, CA, 1993.
11. I. Norros, "Studies on a Model for Connectionless Traffic, Based on Fractional Brownian Motion", *COST24TD(92)041*, 1992.
12. R. A. Skoog, "Study of Clustered Arrival Processes and Signaling Link Delays", in: *Teletraffic and Datatraffic in a Period of Change* (Proc. 13th ITC, Copenhagen, 1991), A. Jensen, V. B. Iversen (Eds.), North Holland, 61-66, 1991.

8. List of Tables and Figures

Table 1. Qualitative description of the CCSN/SS7 traffic measurements used in the analysis in Sections 3 and 4.

Figure 1 (a) - (d). (a) Message arrivals (all SS7 messages) per 1 second during the high traffic period starting at 9:30 and ending at 17:50 on 1/22/93 and corresponding empirical arrival rate process (white line); (b) empirical autocorrelations for the time series in (a); (c) corresponding empirical autocorrelations for the detrended time series using the empirical call arrival rate (white line in (a)); (d) empirical autocorrelations from a sequence of 30000 i.i.d. normal random variables.

Figure 2. Variance-time plot of the detrended version of the number of all SS7 messages (see Figure 1). The asymptotic slope (determined using the "brushed" points) is clearly larger than the slope -1.0 of the dotted reference line and is readily estimated to be about -0.5.

Figure 3 (a) - (c). (a) Empirical histogram of the logarithm of 305,225 call holding times that started during the high traffic period on 1/22/93; (b) log-log plot of the complementary empirical call holding time distribution function $1 - \hat{F}(x)$ versus x ; (c) Hill's estimate for the slope in (b) using the k largest order statistics, as a function of k .

Figure 4 (a) - (b). Sample mean (a) and sample standard deviation (b) of the 305,225 call holding times during the high traffic period on 1/22/93 as a function of the number of observations used for computing the sample moments.

Figure 5. 800 query-response delays.

CCSN/SS7 Traffic Measurements				
Subnetwork (Collection Period) Number of Data Sets	Number of SS7 Messages (in Millions)	Types of Signaling Links	Load (busiest 1 sec/1 hour)	Special Features
Subnetwork I (March 93 - May 93) 14 data sets	80	A-links to EOs and AT C-links D-links to R-STPs D-links to STPs of IC	49%/32% 1%/.02% 91%/23% 29%/16%	Includes Easter (4/11/93) and Mother's Day (5/9/93) Business, Residential EOs (including 1 ISDN EO)
Subnetwork II (Dec. 92 - Feb. 93) 15 data sets	70	A-links to EOs A-links to SCPs (LIDB and 800)	15%/7% 23%/7%	Includes New Year's Eve (12/31/92)
Subnetwork III (Aug. 92 - Sept. 92) 2 data sets	10	A-links to EOs and AT B-links to STP pair D-links to R-STPs	24%/14% 9%/3% 5%/.5%	Includes day after Labor Day (9/8/92) Business, Residential and mixed EOs
Subnetwork IV (Dec. 90 - Jan. 91) 3 data sets	8	A-links to EOs	21%/6%	Includes New Year's Eve (12/31/90)









