# Measurement and Analysis of Traffic in a Hybrid Satellite-Terrestrial Network

**Qing Shao and Ljiljana Trajković**
**Simon Fraser University**
**Vancouver, Canada**
**{qshao, ljilja}@cs.sfu.ca**

**Keywords:** satellite-terrestrial networks, modeling TCP connections, traffic measurements, traffic prediction, long-range dependence.

**ABSTRACT**

Measurement and analysis of traffic traces are important for better understanding of network behavior. In this paper, we describe measurements of traffic data from a satellite Internet service provider. We present statistical analysis used to characterize traffic loads and distribution of applications and packet sizes in a hybrid satellite-terrestrial system. We investigate long-range dependence as the traffic patterns vary during a day. We propose a traffic model on the Transmission Control Protocol (TCP) connection level. Finally, we use data from billing records to predict traffic using autoregressive integrated moving average model and evaluate its performance.

## 1. INTRODUCTION

Internet has continued to grow and change over the last decade. This evolution has been accompanied by the growth of traffic volume, the development of new protocols, and a variety of new Internet access technologies. Network traffic measurements are useful for network troubleshooting, workload characterization, and network performance evaluation. In order to detect the invariants in a rather dynamic traffic structure, measurement and analysis of genuine network traffic traces are important and continuing tasks.

Collection and characterization of terrestrial Internet traffic has received considerable attention during the past decade [1], [2]. Numerous Web sites offer collected samples of Internet traffic traces [3]. To the contrary, few traffic traces have been collected from wireless or satellite commercial sites. In this paper, we describe measurements of hybrid satellite-terrestrial traffic from ChinaSat, a commercial satellite Internet service provider. Our objective is to analyze and model the collected traffic and to characterize the underlying statistical processes and distributions.

This paper is organized as follows: in Section 2, we briefly describe the satellite traffic measurements. Section 3 presents the analysis of billing records. Section 4 describes traffic analysis and characterization, while Section 5 and 6 introduce traffic models and prediction. Finally, we conclude with Section 7.

## 2. SATELLITE NETWORK AND TRAFFIC MEASUREMENT

The advantage of satellite systems is that they broadcast information to large geographical regions without a need to solve the last mile access problem. ChinaSat, the largest satellite communication service provider in China, operates an asymmetric satellite system called DirecPC that provides Internet access to over 200 Internet cafés across provinces. Each café offers on average 40 PCs to its customers surfing the Internet. A proxy server is configured to enable the PCs to simultaneously share the Internet access through a single connection account. DirecPC utilizes two special techniques: Internet Protocol (IP) spoofing and TCP splitting to improve network performance.

When a customer browses a website in the café, a request is sent through a terrestrial dial-up modem to a local Internet Service Provider (ISP). The DirecPC software installed on the proxy server automatically attaches a "tunneling header" (an electronic addressing mask) to the requested website also called unique resource locator (URL). This "tunneling header" instructs the ISP to forward the URL request to the DirecPC Network Operations Centre (NOC) rather than to the requested URL directly. Once the NOC receives the customer's request, the tunneling header is removed and the request is forwarded to the Internet by a high-speed link. When the desired content is retrieved, the NOC sends the information to the DirecPC satellite that broadcasts it to the DirecPC receiving system at the café's network. This special routing technique is called IP spoofing.

Link delay of a satellite system is ~ 250 ms, much larger than in terrestrial systems because the satellite is located in a geostationary orbit 35,800 kilometers above the earth. This delay causes a large delay-bandwidth product, which defines the number of packets that could be sent without being acknowledged. The throughput of a single TCP connection can be calculated as:

$$throughput = \frac{window\ size}{round\ trip\ time}.$$

The upper bound of the TCP throughput is independent of the channel bandwidth. Hence, larger window size allows TCP to fully utilize higher bandwidth links over long-delay channels. However, the standard TCP version limits the receive window size to 64 Kbytes by using only a 16-bit field in the TCP header. TCP exhibits poor performance in the case of links with large delay-bandwidth product found in satellite networks. The DirecPC system addresses this problem by employing a TCP splitting technique: the hub first acknowledges the downloaded packets from the Internet on behalf of the remote stations and then delivers these packets over the satellite links using a modified TCP with an enlarged window size [4].

The IP spoofing and TCP splitting techniques impose considerable memory requirements on edge routers. Whenever the hub sends downloaded data to remote stations, a copy must be maintained in a retransmit buffer until acknowledgments from the remote stations are received.

The customer's request and the response from the website are re-routed through the satellite hub. Hence, the hub was selected as the location to collect traffic traces. The network topology and the measurement point (monitor) are shown in Figure 1. We used a passive network monitor program tcpdump, rather than commercial collection tools. Traffic traces were collected using a Linux PC equipped with a 100 Base-T Ethernet adaptor and a high-resolution (100 μs) timer. The network access point for the trace collection was a port on the primary Cisco router at the network operation center (NOC). It provided access to the inbound and outbound packets sent among hosts on the NOC LAN using a 10 Mbps Internet connection.

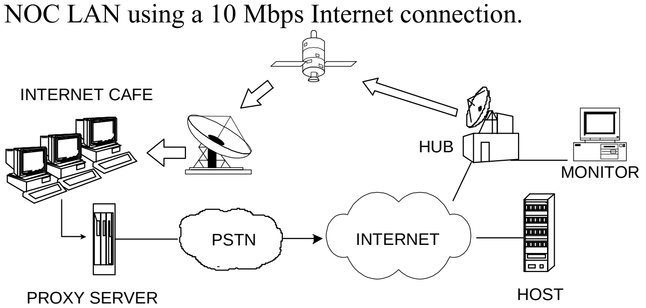
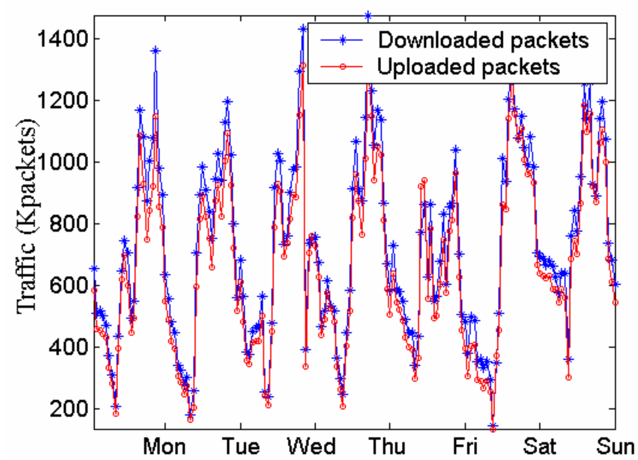



Figure 1. Elements of a hybrid satellite-terrestrial network with a monitoring site for traffic collection.

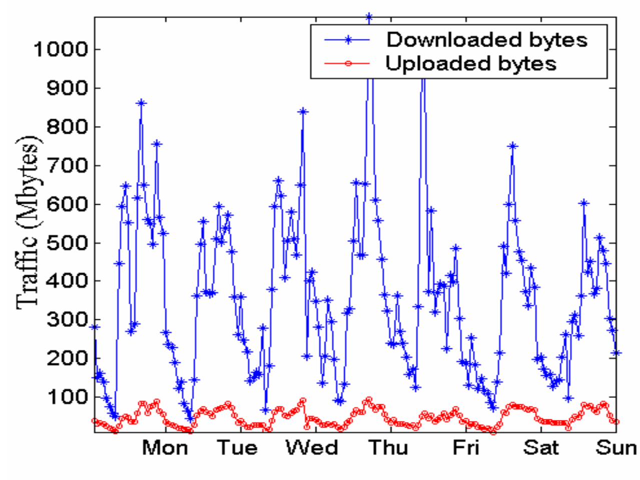
## 3. ANALYSIS OF BILLING RECORDS

We collected billing records from 01-11-2002 to 10-01-2003. Billing records were generated every hour by the satellite billing system. They contain hourly data for connection time, number of downloaded and uploaded packets, number of downloaded and uploaded bytes, and IDs for individual users. Hence, they capture a description of the network workload.

We used the billing records to describe the traffic load in the system. The traffic load information is obtained from billing records by aggregating data for individual users. It exhibits a diurnal cycle. The number of data packets over time is shown in Figure 2(a). The number of uploaded and downloaded packets is similar, with the number of downloaded packets being slightly higher. The gap may be attributed to the small contribution of User Datagram Protocol (UDP) packets. When the traffic load is expressed in terms of bytes, as shown in Figure 2(b), there is a visible difference between the upload and download directions. This difference indicates the asymmetric characteristic of the data transmission.



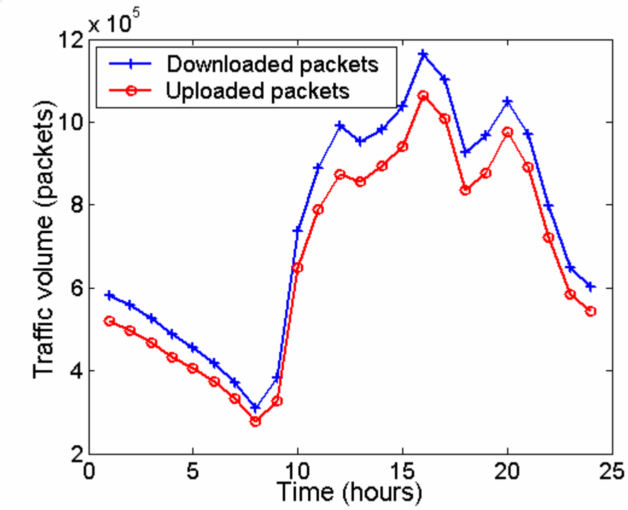

(a)



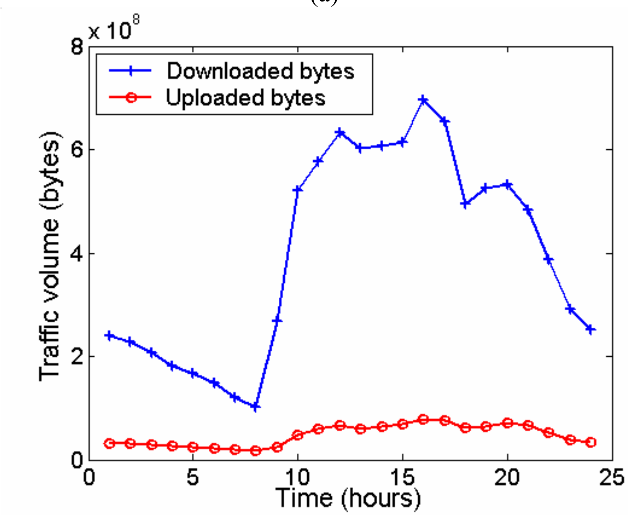

(b)

**Figure 2.** A weekly traffic volume measured in (a) packets and (b) bytes. Traffic data was collected from 09-12-2002 to 15-12-2002.

Figure 3 illustrates the average traffic volume during a single day. The asymmetric pattern appears again. The network usage starts to increase at 8:00 and reaches its maximum about 15:00-16:00. It gradually decreases during the night. This result reflects the customers' behavior in an Internet café.




(a)




(b)

**Figure 3.** Average traffic volume over a single day, measured in (a) packets and (b) bytes. Traffic data was collected from 09-12-2002 to 15-12-2002.

## 4. TRAFFIC ANALYSIS AND CHARACTERIZATION

We recorded traffic traces from 14-12-2002 11:30 AM to 06-01-2003 11:00 AM. The data were stored in 128 collected files, containing ~63 Gbytes of data. *tcpdump* trace has a much finer time granularity (several ms) compared to the billing records (1 h). It contains fields from TCP and IP headers that enable detailed traffic analysis.

### 4.1 Protocols and Applications

The collected traffic traces contain only IP packets. They were collected on the application and transport layers. Properties of the collected data are shown in Tables 1 and 2. We extracted the TCP packets that are used to establish TCP connections between the source and the destination. These packets have either their synchronizing sequence number flags (SYN) or their finishing flags (FIN) set in the TCP header. Packets with SYN or FIN flags were classified according to the applications that generated them. We ignored rare packets with resetting flag (RST) because they imply that the connection has been terminated for reasons such as illegal option length, non-compliant behavior, or firewall options. The type of application is determined from the port number. The contribution of the most common applications is shown in Table 2. Several well-known applications such as email Post Office Protocol version 3 (POP3) and Telnet are present, albeit in small percentages. Note that Web applications dominate the traffic. There are very few File Transfer Protocol (FTP) connections, contributing a large number of bytes. Category "Other" includes applications that use a wide range of TCP and UDP port numbers. The most common port numbers in this category are 1755, 9065, 9013, and 9014. These non-standard ports are used by either new applications or unknown protocols. Playing computer games is very popular in Internet cafés. Hence, we suspect that data from the non-standard ports are most likely generated by online games that use unregistered fixed ports.

**Table 1.** Characteristics of a traffic trace by protocol. Traffic data was collected on 2002-12-22.

| Protocol | Packets (%) | Bytes (%) |
|---|---|---|
| TCP | 84.32 | 94.50 |
| UDP | 14.24 | 5.06 |
| ICMP | 1.45 | 0.45 |
| Total | ~100 | ~100 |

**Table 2.** Characteristics of a traffic trace by application. Traffic data was collected on 12-22-2002.

| Applications | Connections (%) | Bytes (%) |
|---|---|---|
| WWW | 90.06 | 75.79 |
| FTP-data | 0.19 | 10.7 |
| IRC | 0.69 | 0.008 |
| SMTP | 0.17 | 0.01 |
| POP3 | 0.03 | 0.02 |
| Telnet | 0.02 | 0.002 |
| Other | 8.84 | 13.47 |
| Total | 100 | 100 |

### 4.2 WWW Traffic on the TCP Connection Level

With the current growth of the WWW, Web proxy caching has been widely implemented in Web servers. Collected Web data has been used to find the frequency-rank relation of client's requests. The frequency-rank relation is defined as the number of requests vs. the rank of the website in terms of the number of requests. Most traces in the past were collected in local or campus settings, either from Web server logs or from Web proxies [5]. Therefore, it is also of interest to examine the frequency-rank relation of requests for connections in traffic data collected from a geographically broad region. ChinaSat traffic trace consists of end-users' requests in Internet cafés that are located across provinces in China. While past work mainly examined the frequency-rank relation of client's requests, we examine here the number of connections and the volume of downloads by analyzing traffic distribution among the Web servers. This approach provides valuable information for traffic analysis and traffic modeling on the connection level. Our work provides a unique empirical study of the frequency-rank relation of client's connections based on genuine traffic data.

From the collected traffic trace we first extracted data on the TCP connection level and used these data to analyze the Hyper-Text Transfer Protocol (HTTP) connections. These traffic data usually exhibit rather skewed behavior: several websites are very popular while most websites are seldom visited. Common statistics, such as the mean, median, and variance, could not characterize these data.

We first selected Zipf distribution for data characterization because it has been used often in the past to describe skewed data [5]. Zipf's law states that the number of requests (frequency) is inversely proportional to its rank among the requests (the largest number of request corresponds to rank 1). The generalized Zipf distribution is defined as:

$$f_r \sim 1/r^{\theta},$$

where $f_r$ is the number of requests, $r$ is the rank of the website in terms of the number of requests, and θ is a constant. Its log-log plot is linear, with slope θ typically less than 1. We examined frequency-rank relation of requests for connections in the collected traffic trace over several hours or several days. We found that connection frequency-rank relation does not follow the Zipf distribution. As shown in Figure 4 (a), while the Zipf distribution fits well the mid range of the curve, the data exhibit top concavity. This phenomenon appeared frequently in multiple tests of sub-traces. We attributed this phenomenon to the practice that most browsers in the Internet cafés support HTTP/1.1 (rather than HTTP/1.0) because of the increasing usage of Internet Explorer 6.0 and Netscape 7. An HTTP/1.1 connection can be kept open to transfer multiple Web pages from the same server, resulting in multiple requests per connection. Hence, the frequency-rank of client's connection will further deviate from the linear relation as the number of persistent connections increases.

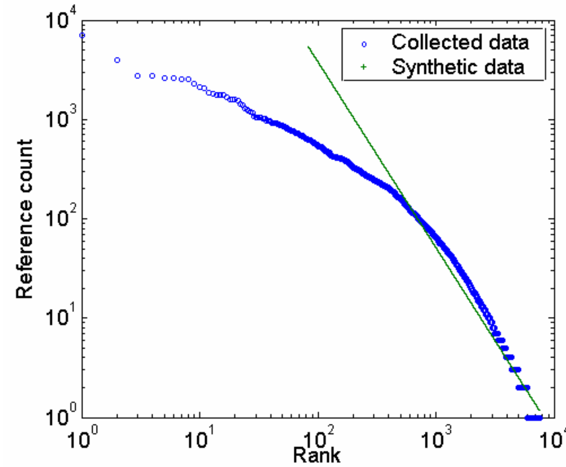
We discovered that collected traffic data follow the Discrete Gaussian Exponential (DGX) [6], as shown in Figure 4 (b). The DGX distribution is a discrete version of the continuous lognormal distribution:

$$p(x=k) = \frac{A(\mu,\sigma)}{k}\exp[-\frac{(\ln k-\mu)^2}{2\sigma^2}], \quad k = 1, 2, ...,$$

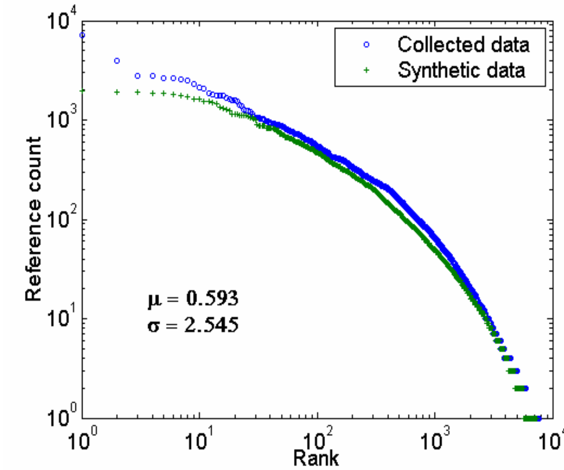
where

$$A(\mu,k) = \{\sum_{k=1}^{\infty}\frac{1}{k}[-\frac{(\ln k-\mu)^2}{2\sigma^2}]\}^{-1}$$

is a normalization constant that depends on the parameters $\mu$ and $\sigma$. Parameters $\mu$ and $\sigma$ are estimated using the maximum likelihood estimator.




(a)



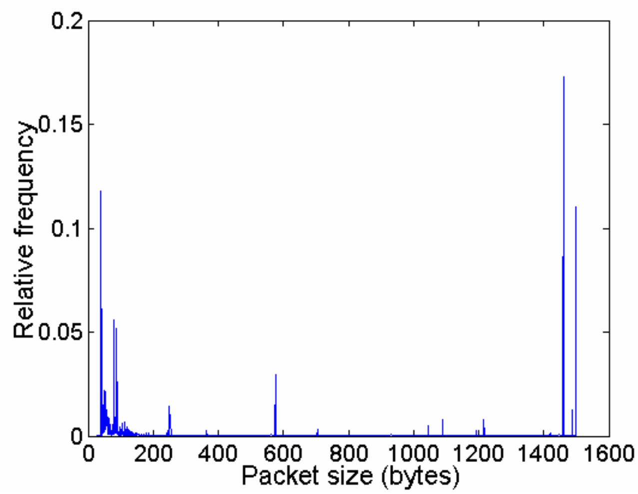

(b)

**Figure 4.** Frequency vs. rank of client's connections: (a) Zipf distribution curve fit and (b) DGX curve fit. Traffic data was collected on 22-12-2002.

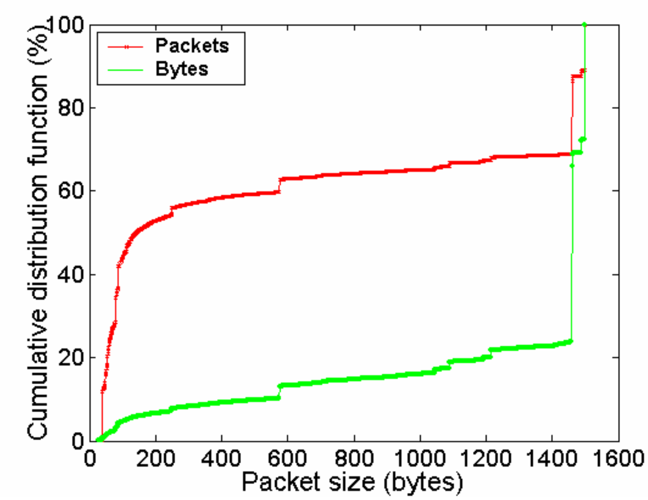
We also examined the top ten busiest websites. It was surprising to find that they are all registered under the Asia Pacific Network Information Centre (APNIC). They account for 60.23 % of the entire traffic load. A Chinese search engine website was the most popular site. This result indicates that the traffic is non-uniformly distributed among the Internet hosts and that language and geographical factors are important for content delivery networks and designing caching proxies.

### 4.3 TCP Packet Sizes

Packet size statistic is shown in Figure 5(a). Packet size distribution is bimodal: there are numerous small acknowledgment packets and many large packets for bulk data file transfer type applications. There are very few in between. Ten most common packet-sizes are (listed in descending order): 1,460, 40, 1,500, 80, 88, 576, 48, 55, 250, and 1,462 bytes. The maximum packet size of 1,460 bytes is difficult to explain: a possible explanation may be the default buffer size in routers. Many 1,500-byte packets result from the limit imposed on maximum packet size in IP networks. Packets of 80 and 88 bytes are UDP packets. The large presence of 576-byte packets reflects TCP implementations without "path MTU discovery", which use packets of 536 bytes (plus 40-byte header) as the default Maximum Segment Size (MSS). This also explains the lack of the common 552-byte packets in the collected trace. The smallest packets, 40 bytes in length, are mainly TCP packets with ACK, SYN, FIN, or RST flags.




(a)




(b)

**Figure 5.** Relative frequency of packet sizes (a) and their cumulative distribution (b). Traffic data was collected from 21-12-2002 22:08 to 23-12-2002 3:28. The total number of packets is 42,939,838.

The cumulative distribution function shown in Figure 5(b) indicates that 50% of the packets are smaller than 200 bytes. They are mainly TCP acknowledgment packets and short HTTP click requests. More than 50% of the bytes are carried in 1,460-byte packets. This is due to the massive transfers of data in the down-stream path. In conclusion, most bytes are transferred in large packets and packet size distribution is bimodal. This concurs with results reported in previous studies [1].

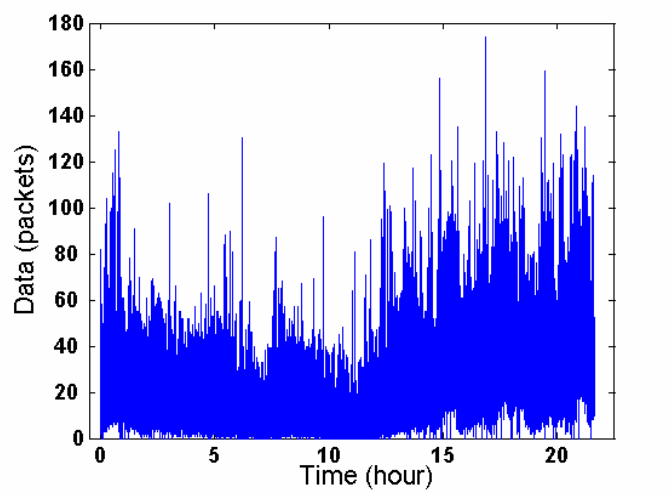
## 5. TRAFFIC MODEL

Self-similarity and long-range dependence have been well document in traffic studies over the past decade [7]. They provide key concepts in analyzing traffic data and modeling the traffic. We address here several practical concerns dealing with long-range dependence and traffic modeling on the TCP connection level.

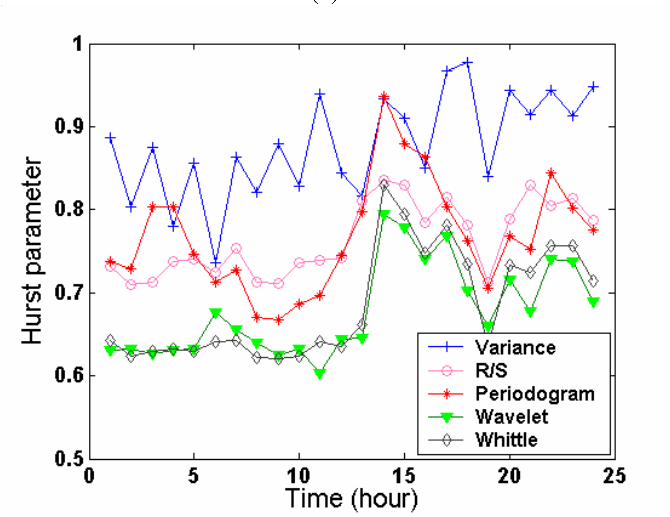
### 5.1 Estimation of Self-similarity

Self-similar traffic models are based on the wide-sense stationarity assumption. Hence, it is important to examine stationarity before characterizing data using various estimators [8]. For example, traffic traces generated from synthetic Fractional Gaussian Noise (FGN) processes [7] are stationary. In genuine traffic traces, effects of non-stationarity such as level-shift and trend are often present. They are more prominent in low-bandwidth networks, such as satellite networks, as shown in Figure 6(a). Testing for stationarity is rather difficult [8]. An alternate approach [9] is to decompose a trace into small sub-traces to eliminate the trend and the effect of level shifts. In practice, this approach may result in short data samples, which leads to unreliable estimates. Hence, a trade-off between finding stationary samples and their small size is often necessary. In order to reduce the effect of non-stationarity, we first decomposed a daily traffic trace into 24 non-overlapping sub-traces. Each sub-trace was then tested for the second order (wide-sense) stationarity. The sub-traces were first visually inspected to locate visible tendencies that imply non-stationarity. Each sub-trace was then divided into equal blocks. We examined the mean, variance, and Hurst parameter of each block by employing various statistical tests [9], [10]. We found that one-hour traces may be treated as stationary, with the exception of several afternoon sub-traces. Hence, these afternoon sub-traces were further decomposed into smaller sub-traces, while keeping a sufficient number of samples.

Figure 6(b) shows the Hurst parameter estimated using various estimators [10]. As expected, estimators generated different estimates of the Hurst parameter. It is also known that estimators may produce unreliable results [11]. Nevertheless, estimated Hurst parameters exhibit similar trend with traffic data. In our experiment, more weight was given to the wavelet estimator because it may help avoid the effects of non-stationarity [12]. To some degree, the wavelet-based estimates reflect the traffic load shown in Figure 6(a). Nevertheless, we cannot conclude that the Hurst parameter increases as the traffic load increases, as observed elsewhere [7]. This also suggests that a single-parameter traffic characterization might not capture the complexity of the network traffic variability. It remains an open question how to test the stationarity and estimate self-similarity in genuine traffic traces.




(a)



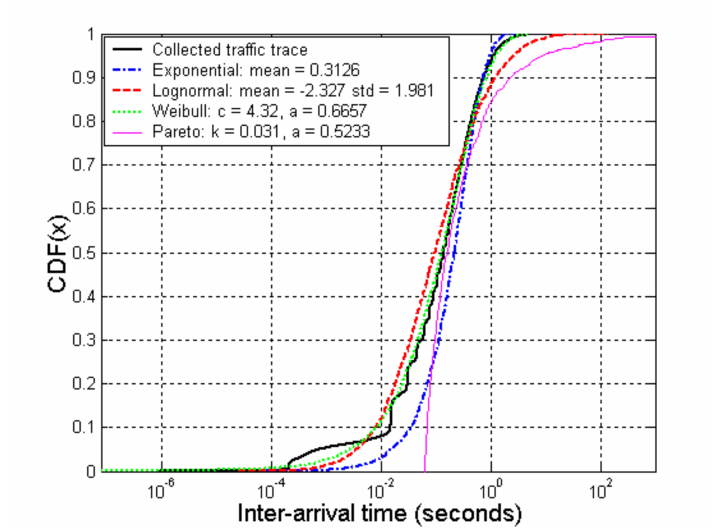

(b)

**Figure 6.** Non-stationarity in (a) daily traffic and (b) variation of the Hurst parameter. Traffic data was collected on 2002-12-09.

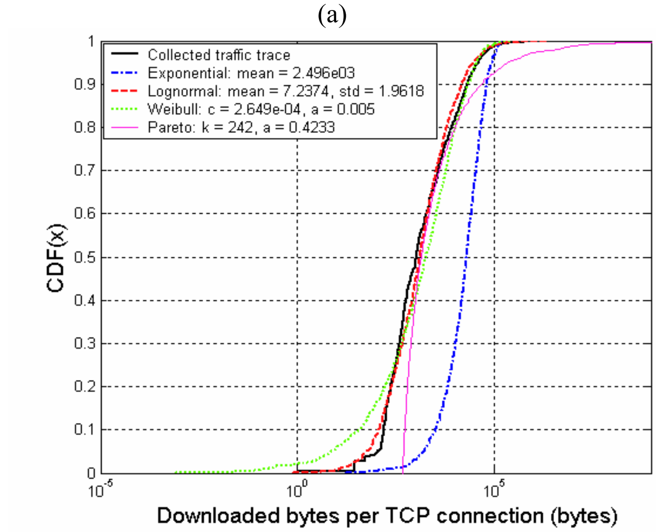
### 5.2 Traffic Model on the TCP Connection Level

TCP splitting technique described in Section 2 may improve the efficiency of a hybrid wired-wireless system. However, it imposes considerable memory requirements on the edge routers. Hence, it is important to model downloaded traffic on a TCP connection level in such hybrid systems. Collected traffic traces can be used to extract individual or aggregate TCP connections. The connection information is extracted from TCP headers (SYN and FIN flags).

The collected traffic trace has long-range dependent (second order self-similar) characteristic. This statistical property does not easily lead to a traffic model. Nevertheless, it is known that processes with such a characteristic are generated by aggregating multiple ON/OFF sources. An ON/OFF source model describes traffic fluctuations between a source/destination pair. The duration of the ON/OFF periods are heavy-tailed distributed with infinite variance. We modeled two important parameters: TCP connection inter-arrival times and the number of downloaded bytes per TCP connection. Assuming that the data are sent at a constant rate, we modeled downloaded data as ON periods and inter-arrival time as OFF periods when there is no data transfer. We examined Pareto, Weibull and lognormal distributions and used exponential distribution for comparison. The traffic data was fitted using the maximum likelihood estimator, as shown in Figure 7.




(a)




(b)

**Figure 7.** Distribution fits vs. empirical CDF for (a) inter-arrival times and (b) number of downloaded bytes per TCP connection.

We evaluated and quantified the performance of various distributions by employing Chi-Square and Kolmogorov-Smirnov goodness of fit tests [13]. We concluded that the

Weibull distribution yields a suitable model for the inter-arrival times, while lognormal distribution was suitable to model the number of downloaded bytes per TCP connection. It is interesting to note that the exponential distribution underestimates both the small and large values in both cases. The Pareto distribution is unique among the three distributions: it is suitable to fit the upper tails of the trace, rather than the entire body.

## 6. TRAFFIC PREDICTION

Traffic prediction plays a crucial role in network management such as resource allocation, capacity planning and buffer management.

Time series data are usually modeled as stochastic processes. Linear models have good physical explanation and are easy to implement. Linear time series models have been the predominant prediction tools over the last 50 years. Classical linear techniques for time series analysis include moving average model (MA), autoregressive model (AR), and their combination Autoregressive Moving Average (ARMA) model. The ARMA has the form:

$$X(t) = \phi_1 X(t-1) + \cdots + \phi_p X(t-p) + e(t) + \theta_1 e(t-1) \cdots + \theta_q e(t-q),$$

where $e(t)$ represents past disturbance, and $\phi_j$ and $\theta_k$ are constant parameters. As shown in Section 3, ChinaSat traffic data has a highly time-dependant pattern that violates the stationarity assumption of the ARMA model. Differencing is used to transform the non-stationary data into stationary. The seasonal form is also added to reflect the seasonal cycle in the data. This leads to the seasonal $\text{ARIMA}(p,d,q)\times(P,D,Q)_s$ model. Parameters *p*, *d*, and *q* are the orders of the autoregressive, differencing, and moving average, respectively. Parameters *P*, *D*, and *Q* are the orders of the seasonal autoregressive, seasonal differencing, and seasonal moving average, respectively. Parameter *s* is the period of the seasonal pattern. The time series data within a season is modeled with $(p,d,q)$, while the seasonal effect of the data is modeled with $(P,D,Q)$ based on data *s* units apart. The ARIMA prediction methodology is widely used in the financial markets to predict daily stock prices and long-term exchange rates. In the past, researchers also used ARIMA model to forecast the aggregate data traffic [14], [15].

We evaluated the ARIMA model for predicting uploaded and downloaded traffic in ChinaSat satellite network. We used the normalized mean squared error (NMSE) to measure the performance of the predictor:

$$NMSE = \frac{1}{\sigma^2 N}\sum_{k=1}^{N}(x(k)-\bar{x}(k))^2,$$

where $x(k)$ is the time series data, $\bar{x}(k)$ is the prediction, and $\sigma^2$ is the variance of the time series. NMSE can be viewed as a noise to signal ratio. The smaller the NMSE, the better the predictor.

We used the Box-Jenkins methodology to determine the $\text{ARIMA}(p,d,q)\times(P,D,Q)_s$ [16]. ChinaSat historical data was obtained by aggregating six weeks of billing records of individual users. Traffic data from the billing records was modeled with $\text{ARIMA}(1,0,0)\times(0,1,1)_{168}$. We used the S-PLUS [17] maximum likelihood and forecast function to obtain the parameter estimation and forecast [18]. Results are shown in Table 3 and Figure 8. They indicate that the predictor cannot capture the bursts of the downloaded traffic, while it is successful in capturing the uploaded traffic. This is due to the highly asymmetric volume of traffic. The burstiness of the downloaded traffic is mainly caused by various types of applications.

**Table 3.** Prediction performance for uploaded and downloaded traffic. The data was collected from 2002-12-14 to 2002-12-20.

| Traffic type | Uploaded traffic | Downloaded traffic |
|---|---|---|
| NMSE | 0.3653 | 0.5988 |

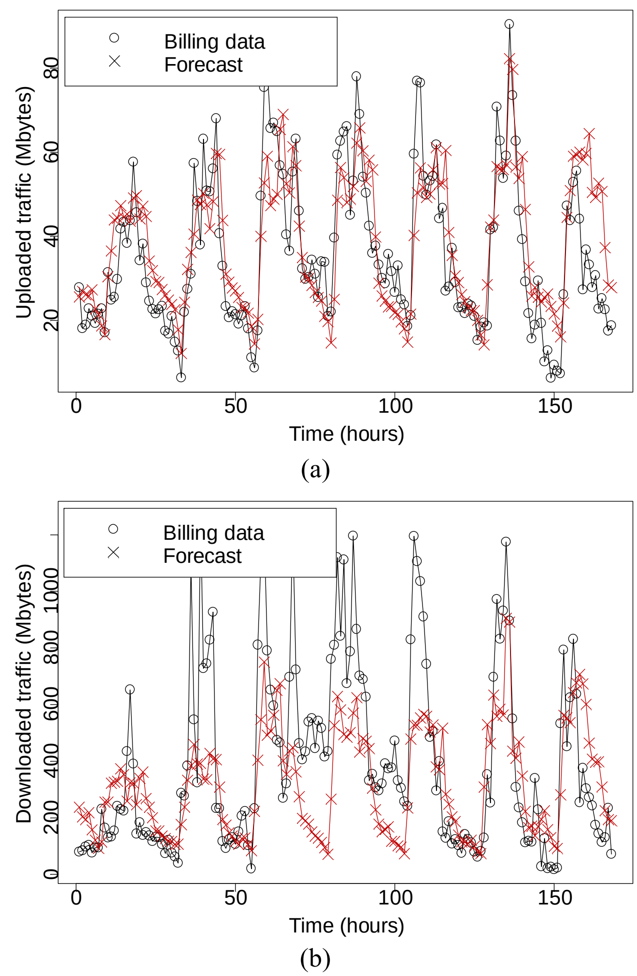



**Figure 8.** One week ahead prediction for (a) uploaded traffic and (b) downloaded traffic. The data was collected from 2002-12-14 to 2002-12-20.

This difference in traffic predictability implies that stationarity may not be achieved by simple differencing for certain traffic data. It may, therefore, be necessary to consider nonlinear models, such as neural networks, when analyzing more complex and fluctuating traffic data where nonlinearity and non-stationarity play a significant role in the forecasting.

## 7. CONCLUSIONS

In this paper, we describe traffic collection from a commercial hybrid satellite-terrestrial network and the statistical analysis of collected traffic traces and billing records.

By examining billing records, we have shown the highly asymmetric traffic transaction pattern. We also characterized the collected traffic trace. Our results show that Web traffic is the dominant application on the hybrid satellite-terrestrial network. Most bytes are transmitted in large packets and packet size distribution is bimodal. The frequency-rank relation of client's connections matches the discrete lognormal distribution. We also estimated the Hurst parameter using various estimators and under different link utilizations. The issue of non-stationarity in investigations of the self-similar behavior of traffic traces was also addressed. We modeled traffic on the TCP connection level and concluded that Weibull and lognormal distributions are suitable for modeling the TCP inter-arrival times and the number of downloaded bytes, respectively. Finally, we used the seasonal ARIMA model to predict one-week of traffic and found that the linear model performed differently for the uploaded and downloaded traffic. The difference implies that the linear model may not be suitable for predicting certain type of traffic behavior.

While the collected traffic data captured only a snapshot of the satellite network, its analysis contributes to better understanding of deployed networks and may be of benefit to commercial network traffic management agencies.


## ACKNOWLEDGMENT

The authors would like to thank Yongxin Shi and Rui Huang, managers of the DirecPC system at ChinaSat, for their help with data collection. Many thanks to Zhiqiang Bi for valuable discussions regarding the "DGX" distribution.

This work was supported in part by the NSERC Grant No. 216844-03 and Canada Foundation for Innovation.


## REFERENCE


[1] S. McCreary, "Trends in wide area IP traffic patterns," in *Proc. of 13th ITC Specialist Seminar on Measurement and Modeling of IP Traffic*, Monterey, California, Sept. 2000, pp. 1–11.

[2] K. Thompson, G. J. Miller, and R. Wilder, "Wide-area Internet traffic patterns and characteristics," *IEEE Network Magazine*, vol. 11, no. 6, pp. 10–23, Nov. 1997.

[3] The Internet Traffic Archive. (2004, May). [Online]. Available: http://ita.ee.lbl.gov/.

[4] V. Jacobson, R. Braden and D. Borman "TCP extensions for high-performance," *Request for Comments: 1323*, May. 1992.

[5] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in *Proc. of IEEE INFOCOM*, Mar. 1999, pp. 126–134.

[6] Z. Bi, C. Faloutsos, and F. Korn, "The "DGX" distribution for mining massive, skewed data," *KDD 2001*, San Francisco, CA, Aug. 2001, pp. 17–26.

[7] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. on Networking*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[8] Y. Zhang, N. Duffield, V. Paxson, and S. Shenker. "On the constancy of Internet path properties," in *Proc. of ACM SIGCOMM Internet Measurement Workshop*, San Francisco, CA, Nov. 2001, pp. 197–211.

[9] D. Veitch and P. Abry, "A statistical test for the time constancy of scaling exponents," *IEEE Trans. on Signal Processing*, vol. 49, no. 10, pp. 2325–2334, Oct. 2001.

[10] M. S. Taqqu and V. Teverovsky, "On estimating the intensity of long-range dependence in finite and infinite variance time series," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA, Birkhauser, 1998, pp. 177–217.

[11] T. Karagiannis, M. Faloutsos, and R. H. Riedi "Long-range dependence: now you see it, now you don't!," in *Proc. GLOBECOM '02*, Taipei, Taiwan, Nov. 2002, pp. 2165–2169.

[12] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffic," *IEEE Trans. on Information Theory*, vol. 44, no. 1, pp. 2–15, Jan. 1998.

[13] R. B. D'Agostino and M. A. Stephens, Eds., *Goodness-of-fit Techniques*. New York: Marcel Dekker, June, 1986, pp. 63–93, pp. 97–145, pp. 421–457.

[14] N. K. Groschwitz and G. C. Polyzos, "A time series model for long-term NSFNET backbone traffic," in *Proc. IEEE Int. Conf. Communications*, May 1994, vol. 3, pp. 1400–1404.

[15] D. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of Internet backbone traffic: observations and initial models," in *Proc. IEEE INFOCOM 2003*, San Francisco, CA, April 2003, pp. 1178–1188.

[16] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*, 2nd edition. Oakland, CA: Holden-Day, 1976, pp. 208–329.

[17] S-PLUS. (2004, May). [Online]. Available: http://www.insightful.com/products/splus/.

[18] E. Zivot and J. Wang, *Modelling Financial Time Series with S-Plus*. Seattle, WA: Insightful Corporation, 2003, pp. 57–103.