



# Analysis of traffic data from a hybrid satellite-terrestrial network

---

Savio Lau and Ljiljana Trajkovic  
{saviol, ljilja}@cs.sfu.ca

Communication Networks Laboratory  
<http://www.ensc.sfu.ca/research/cnl>  
School of Engineering Science  
Simon Fraser University

A decorative graphic on the left side of the slide features overlapping yellow, red, and blue squares with a black crosshair.

# Roadmap

---

- Introduction
- **ChinaSat**: network architecture and TCP
- Analysis of **billing** records
- Analysis of **tcpdump** traces:
  - general characteristics
  - TCP options
  - network anomalies
- Conclusions

A decorative graphic in the top left corner features overlapping yellow, red, and blue squares with a black crosshair.

# Introduction and motivation

---

- Analysis of traffic data enables:
  - understanding of traffic dynamics
  - characterization and development of new traffic models
  - evaluation of network performance
- Most traffic data are collected at research institutions or from research networks:
  - traffic data from commercial networks are rare
  - commercial network traffic may have different characteristics compared to research networks
- Analysis of traffic data from a commercial network such as the ChinaSat DirecPC network is important

A decorative graphic in the top left corner features overlapping yellow, red, and blue squares with a black crosshair.

# Previous work

---

- Previous analysis of network traffic focused on:
  - characteristics of TCP connections
  - network traffic patterns
  - statistical and cluster analysis of traffic
  - anomaly detection:
    - statistical methods
    - wavelets
    - principle component analysis



# Previous work on the ChinaSat data

- ChinaSat traffic is self-similar and non-stationary
- Hurst parameter depends on traffic load
- Modeling TCP connections:
  - inter-arrival time is best modeled by the Weibull distribution
  - number of downloaded bytes is best modeled by the lognormal distribution
- The distribution of visited websites is best modeled by the discrete Gaussian exponential (DGX) distribution

Q. Shao and Lj. Trajkovic, "Measurement and analysis of traffic in a hybrid satellite-terrestrial network," *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 329–336.



# Previous work on the ChinaSat data

---

- Traffic prediction:
  - autoregressive integrative moving average (ARIMA) can be used to predict uploaded traffic but not downloaded traffic
  - wavelet + autoregressive model outperforms the ARIMA model

A decorative graphic in the top left corner features a vertical black line intersecting a horizontal black line. To the left of the intersection are three overlapping squares: a yellow one at the top, a red one on the left, and a blue one at the bottom.

# Roadmap

---

- Introduction
- **ChinaSat**: network architecture and TCP
- Analysis of billing records
- Analysis of tcpdump traces:
  - general characteristics
  - TCP options
  - network anomalies
- Conclusions



# ChinaSat hybrid satellite network

---

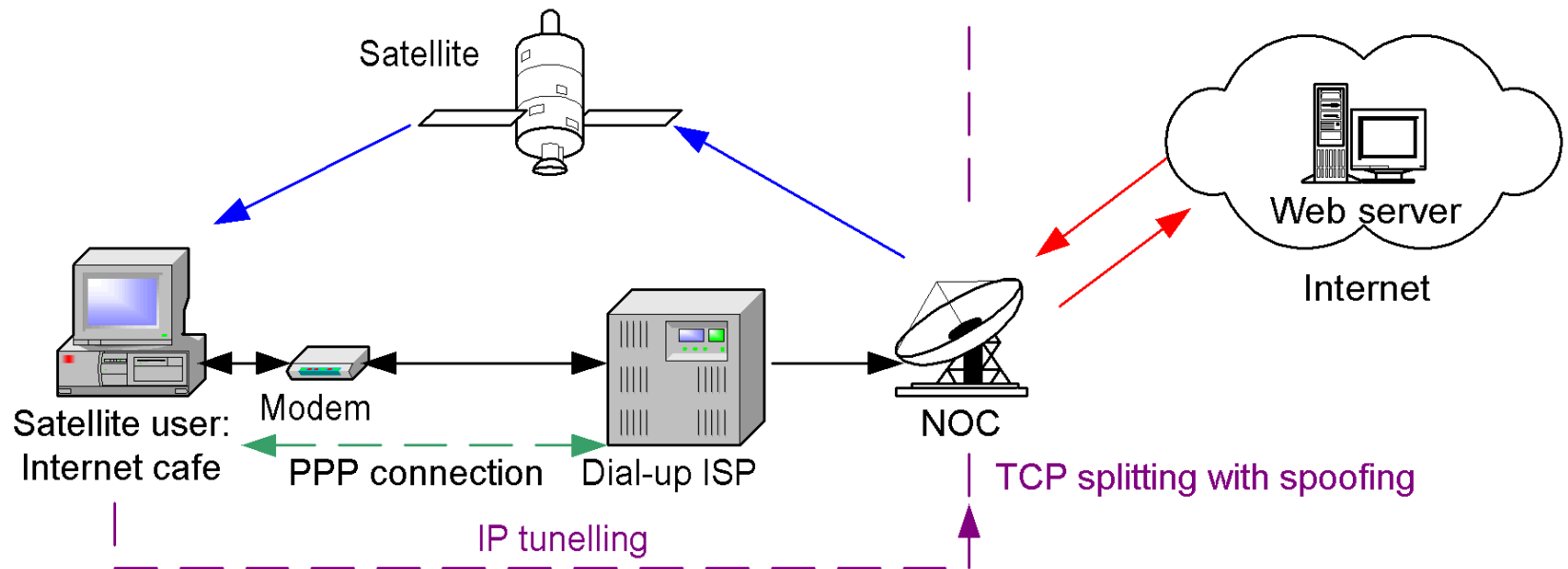
- Employs geosynchronous satellites deployed by Hughes Network Systems Inc.
- Provides data and television services:
  - DirecPC (Classic): unidirectional satellite data service
  - DirecTV: satellite television service
  - DirecWay (Hughnet): bi-directional satellite data service that replaces DirecPC
- DirecPC transmission rates:
  - 400 kb/s from satellite to user
  - 33.6 kb/s from user to network operations center (NOC) using dial-up
- Improves performance using TCP splitting with spoofing

# Characteristics of geosynchronous satellite links



- Large coverage area
- High bandwidth
- Long propagation delay
- Large bandwidth-delay product
- High bit error rates:
  - $10^{-6}$  without error correction
  - $10^{-3}$  or  $10^{-2}$  due to extreme weather and interference
- Path asymmetry

# DirecPC system diagram



NOC: Network operations center  
 PPP: Point-to-point protocol

# TCP extensions for satellite environments



- Increasing initial TCP congestion window (**cwnd**)
- Selective acknowledgement (SACK) option:
  - enables a TCP receiver to acknowledge out-of-order packets
  - allows a TCP sender to identify and retransmit lost segments
  - avoids the performance penalty associated with retransmission timeouts
- TCP sliding window scale option:
  - expands default TCP window from 16 bits to 32 bits
  - allows greater number of unacknowledged packets

# TCP extensions for satellite environments



- Path maximum transmission unit (MTU) discovery:
  - determines the maximum allowable size in links between source and destination
  - enables TCP senders to reach maximum throughput earlier
- Performance enhancing proxies (PEPs):
  - improve TCP performance in specific link environments
  - violate TCP end-to-end semantics
  - technique: **TCP splitting** with **spoofing**



A decorative graphic in the top left corner features overlapping yellow, red, and blue squares with a black crosshair.

# Network anomalies

---

- Scans and worms:
  - packets are sent to probe network hosts
  - used to discover and exploit resources
- Traffic volume anomalies:
  - significant deviation of traffic volume from usual daily or weekly patterns
  - classified as:
    - outages: caused by unavailable links, crashed servers, or routing problems
    - short term increases in demand: caused by short term events such as holiday traffic
  - involve multiple sources and destinations

A decorative graphic in the top left corner features a vertical black line intersecting a horizontal black line. To the left of the vertical line are three overlapping squares: a yellow one at the top, a red one in the middle, and a blue one at the bottom. The text 'Network anomalies' is written in a large, blue, sans-serif font to the right of the vertical line.

# Network anomalies

---

- Flash crowd:
  - high volume of traffic destined to a single destination
  - caused by breaking news or availability of new software
- Traffic shift:
  - redirection of traffic from one set of paths to another
  - caused by route changes, link unavailability, or network congestion

A decorative graphic on the left side of the slide, featuring overlapping yellow, red, and blue squares with a black crosshair.

# Network anomalies

---

- Alpha traffic:
  - unusually high volume of traffic between two endpoints
  - caused by file transfers or bandwidth measurements
- Denial of service:
  - large number of packets directed to a single destination
  - makes a host incapable of handling incoming connections or exhausts available bandwidth along paths to the destination

A decorative graphic in the top left corner features overlapping yellow, red, and blue squares with a black crosshair.

# Roadmap

---

- Introduction
- ChinaSat: network architecture and TCP
- Analysis of **billing** records
- Analysis of tcpdump traces:
  - general characteristics
  - TCP options
  - network anomalies
- Conclusions



# Billing records

- Records were collected during the continuous period from 23:00 on Oct. 31, 2002 to 11:00 on Jan. 10, 2003
- Each file contains the hourly traffic summary for each user
- Fields of interests:
  - **SiteID** (user identification)
  - **Start** (record start time)
  - **CTxByt** (number of bytes downloaded by a user)
  - **CRxByt** (number of bytes uploaded by a user)
  - **CTxPkt** (number of packets downloaded by a user)
  - **CRxPkt** (number of packets uploaded by a user)

Download: from NOC to user through satellite  
Upload: from user to NOC through dial-up



# Billing records format

RecLen	RecTyp	SiteID	Start	Stop	Cmin
Bill	CTxByt	CRxByt	CTxPkt	CRxPkt	
00100	001	0003809504	20030106130005	20030106140005	060
2	0000000414	0000017240	0000000007	0000000227	
00100	001	0004477001	20030106130005	20030106140005	060
2	0000000396	0000006084	0000000006	0000000117	
00100	001	000456EB01	20030106130005	20030106140005	060
2	0015844812	0002903556	0000027471	0000034200	
00100	001	00045C0002	20030106130005	20030106140005	060
2	0003061014	0000397334	0000003789	0000004521	
00100	001	000455B103	20030106130005	20030106140005	008
2	0000000120	0000001021	0000000002	0000000009	

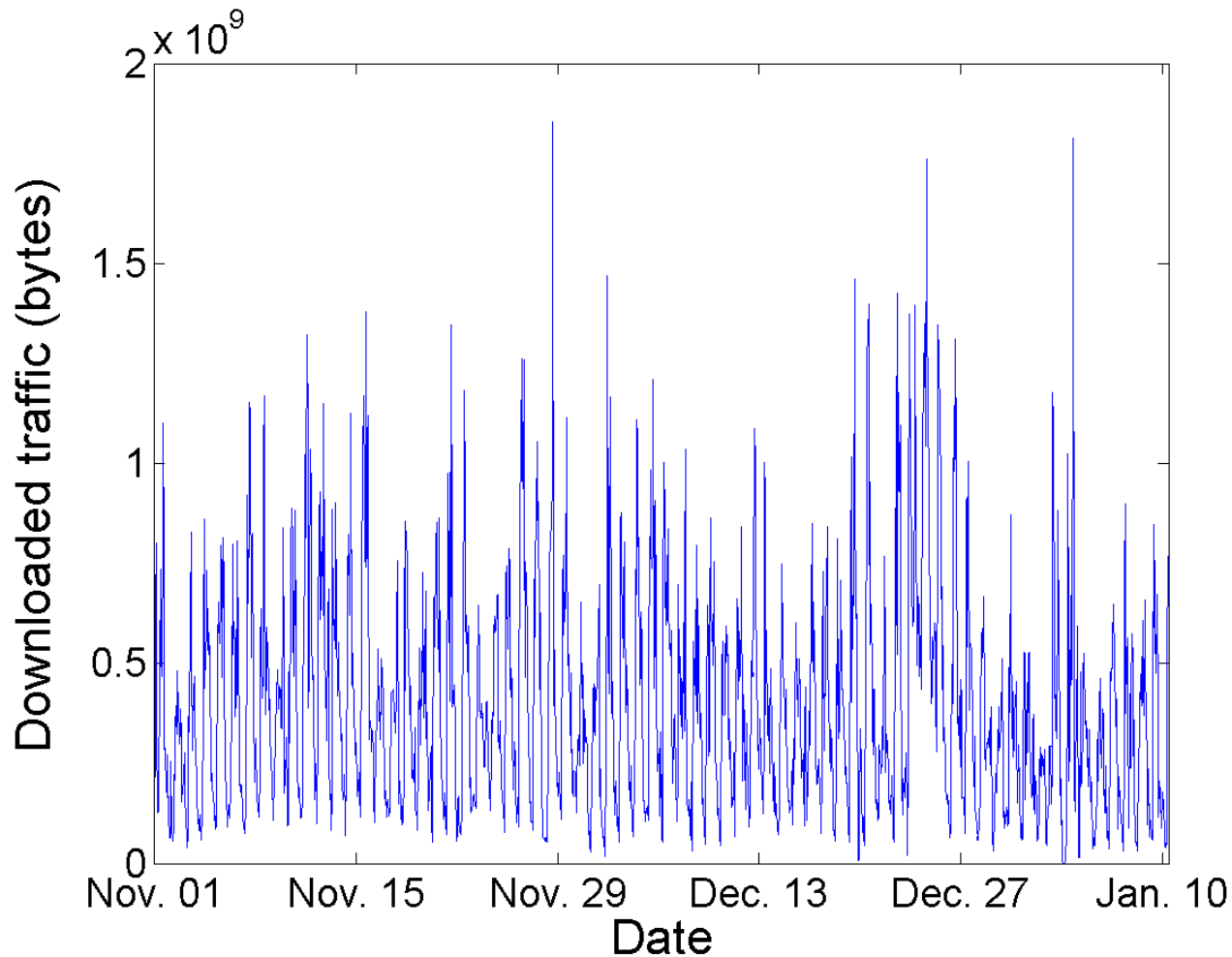
A decorative graphic on the left side of the slide features a vertical black line intersecting a horizontal black line. To the left of the intersection, there are three overlapping squares: a yellow one at the top, a red one in the middle, and a blue one at the bottom.

# Billing records: characteristics

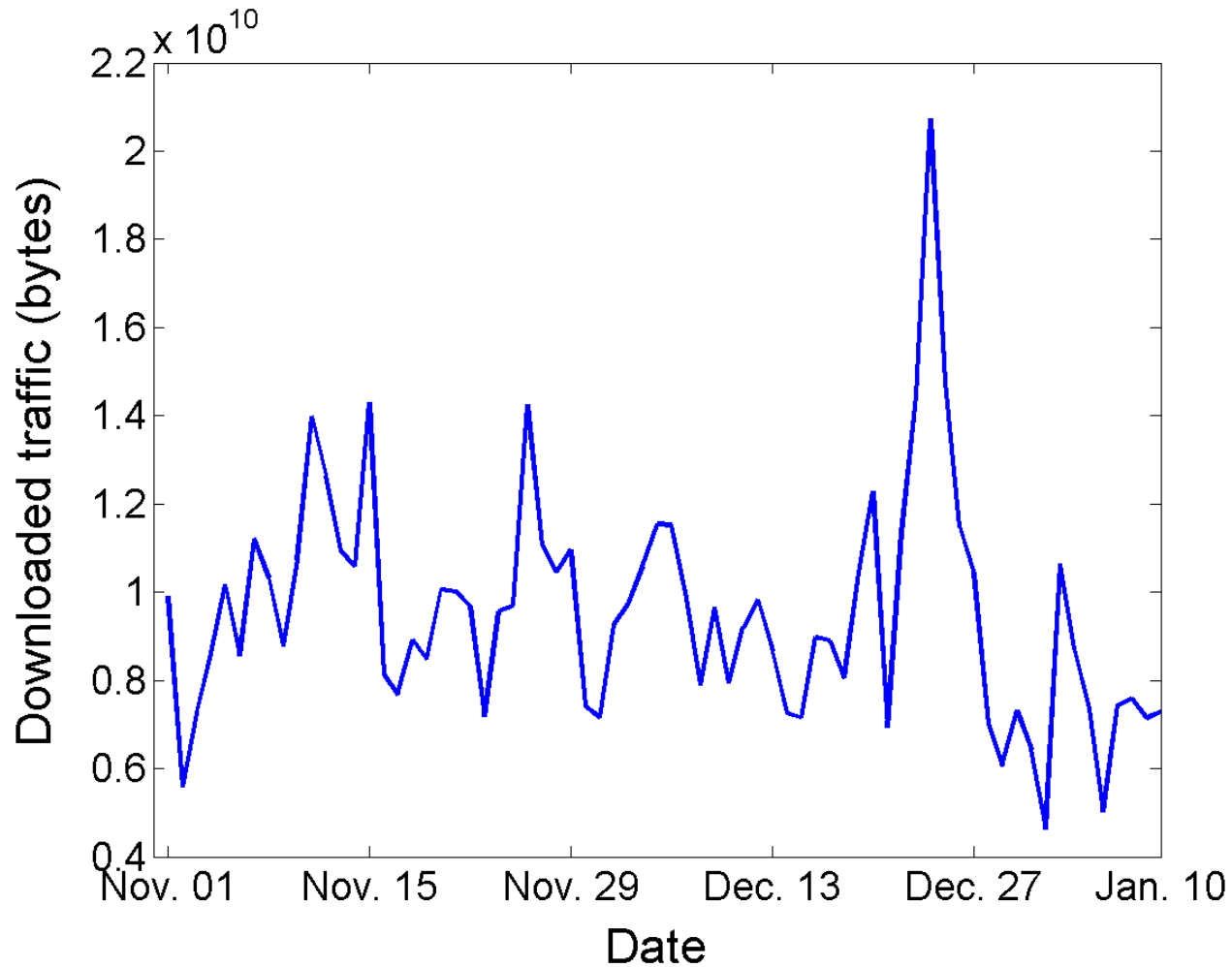
---

- 186 unique [SiteIDs](#) (users)
- Daily and weekly cycles:
  - lower traffic volume on weekends
  - daily cycle starts at 7 AM, rises to three daily maxima at 11 AM, 3 PM, and 7 PM, then decreases monotonically until 7 AM
- Highest daily traffic recorded on [Dec. 24, 2002](#)
- Outage occurred on [Jan. 3, 2003](#)

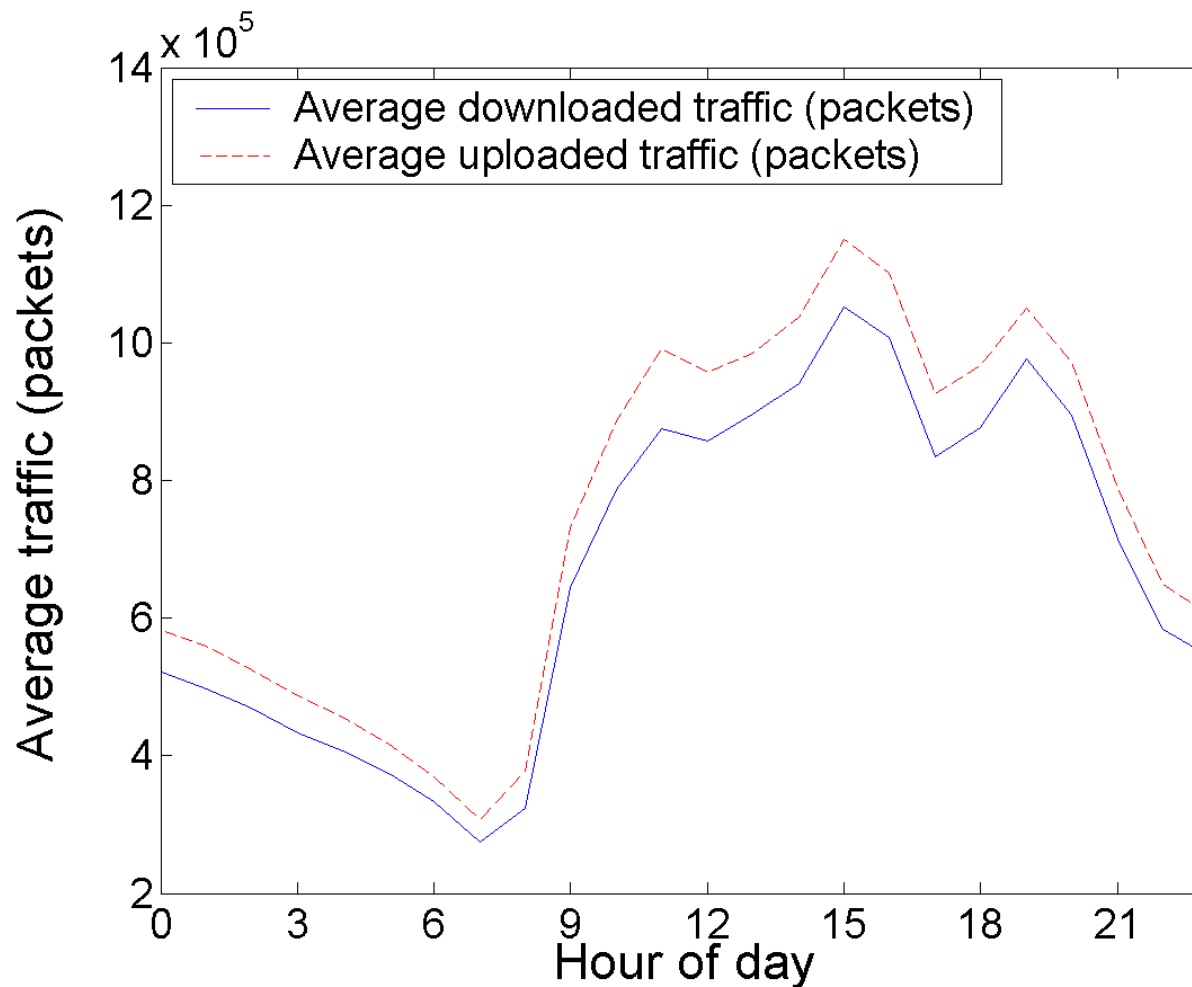
# Aggregated hourly traffic



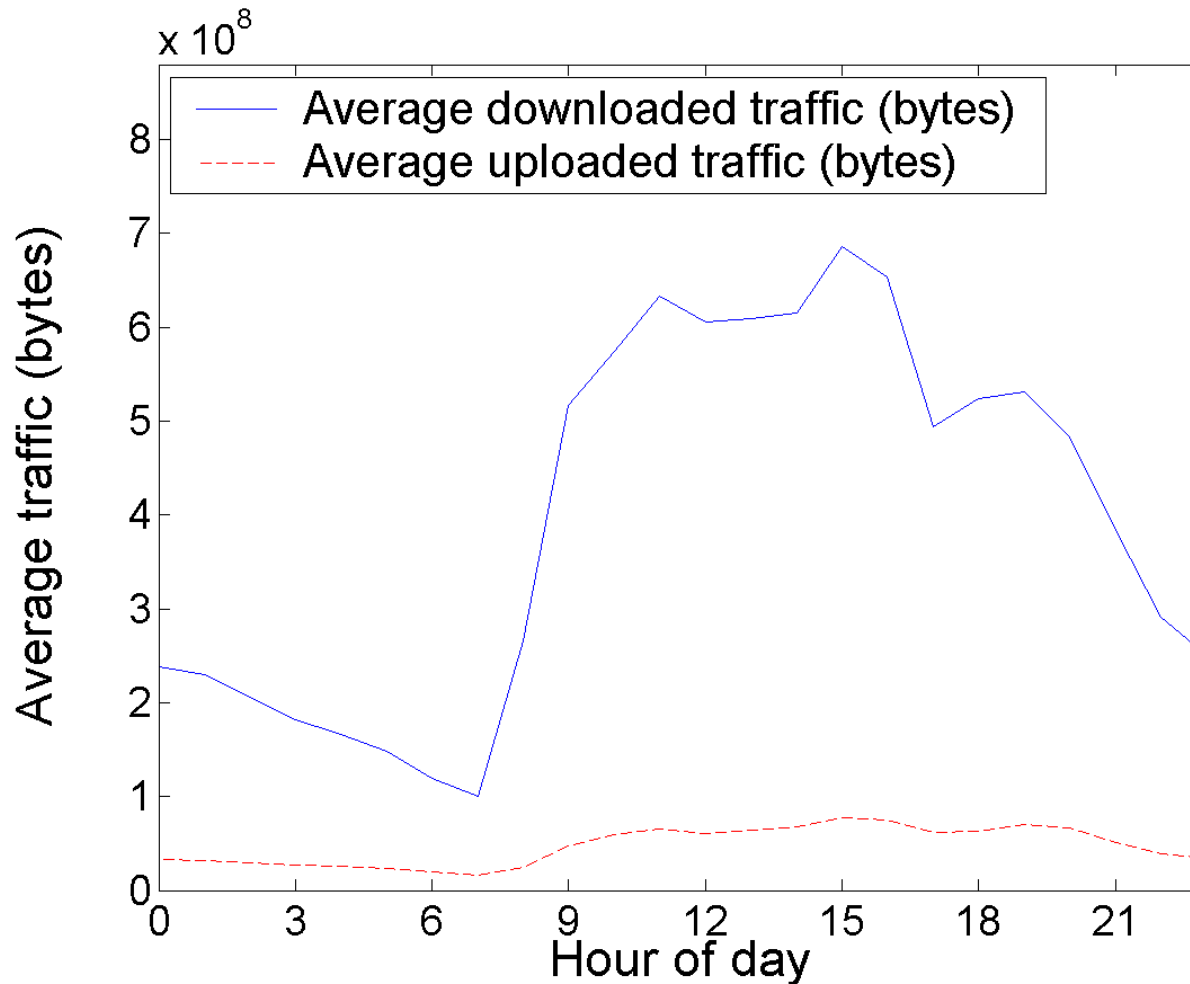
# Aggregated daily traffic



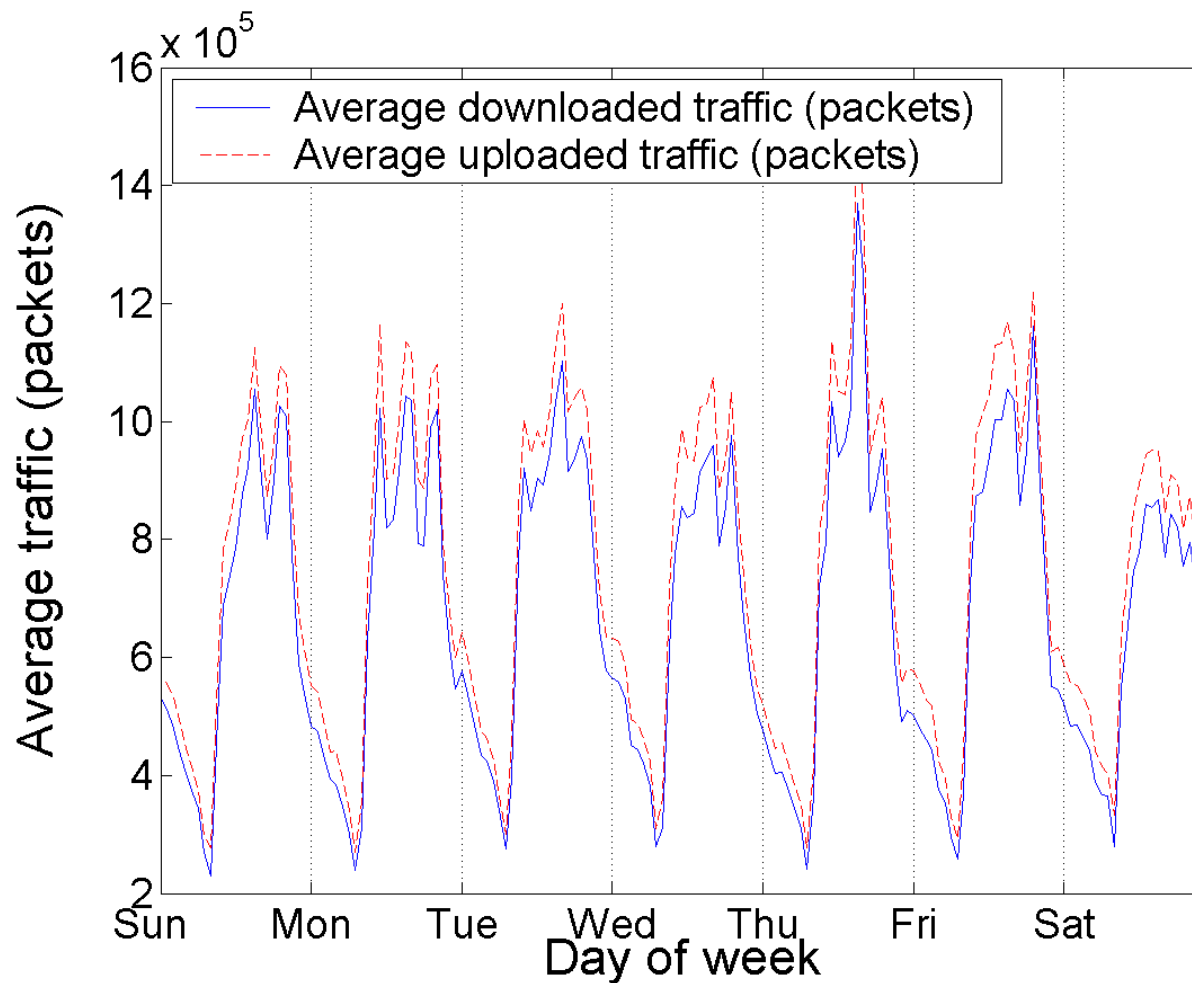
# Daily (diurnal) traffic: average traffic (packets)



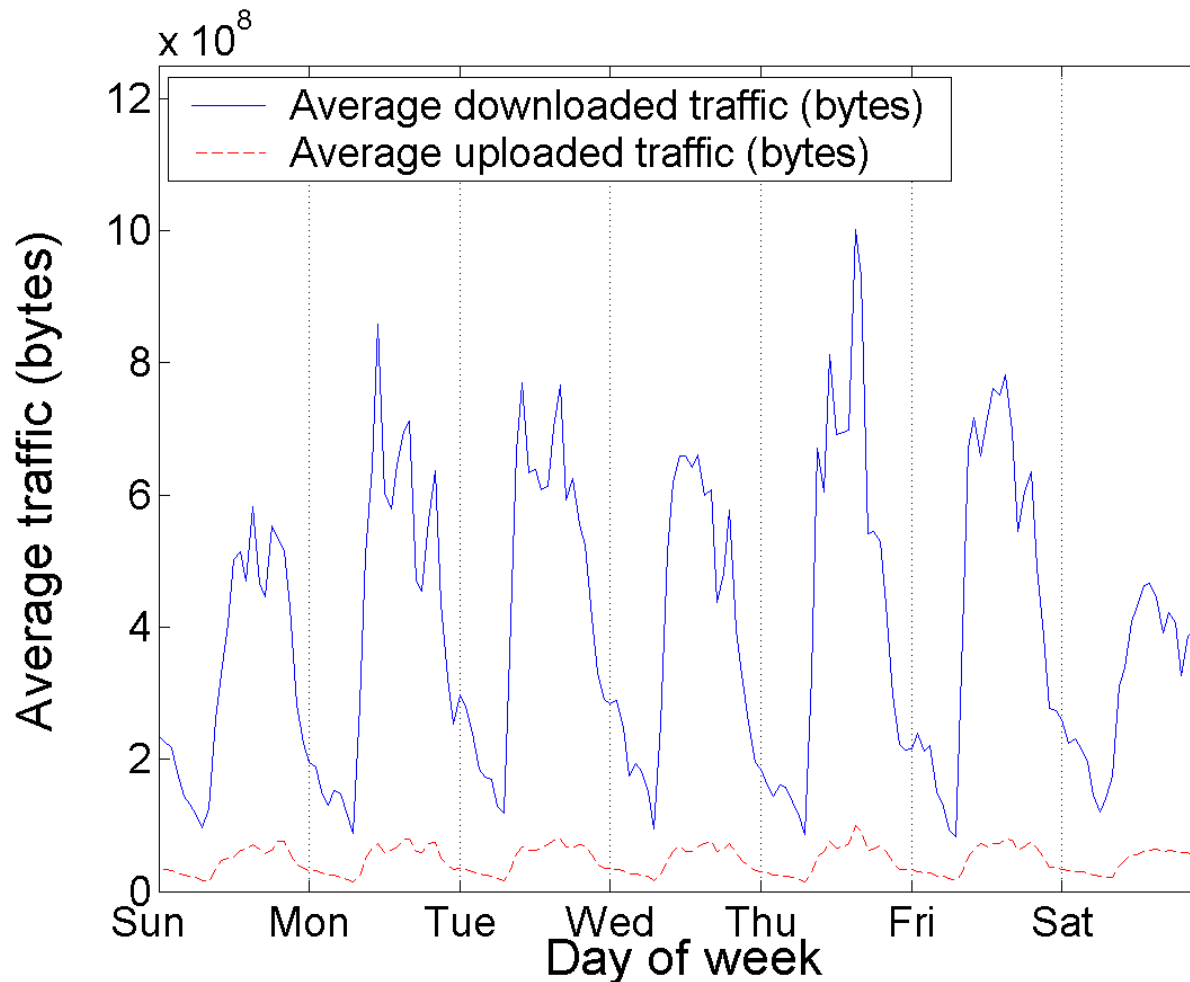
# Daily (diurnal) traffic: average traffic (bytes)



# Weekly traffic: average traffic (packets)



# Weekly traffic: average traffic (bytes)



A decorative graphic in the top left corner features overlapping yellow, red, and blue squares with a black crosshair.

# Roadmap

---

- Introduction
- ChinaSat: network architecture and TCP
- Analysis of billing records
- Analysis of `tcpdump` traces:
  - general characteristics
  - TCP options
  - network anomalies
- Conclusions



# tcpdump trace

- Trace were continuously collected from **11:30** on **Dec. 14, 2002** to **11:00** on **Jan. 10, 2003** at the NOC
- The first 68 bytes of each TCP/IP packet were captured
- **~63 GB** of data contained in 127 files
- User IP address is not constant due to the use of the private IP address range and dynamic IP
- Majority of traffic is TCP:
  - **94%** of total bytes and **84%** of total packets
  - HTTP (port 80) accounts for **90%** of TCP connections and **76%** of TCP bytes
  - FTP (port 21) accounts for **0.2%** of TCP connections and **11%** of TCP bytes





# tcpdump output example

```
12/15/2002 04:27:05.328455 192.168.1.83.63260 > 211.167.92.197.6732: . ack 489 win 8192
12/15/2002 04:27:05.331020 211.100.18.48.80 > 192.168.1.164.41842: S
2928120965:2928120965(0) ack 3324468 win 64240 <mss 1460,nop,nop,sackOK> (DF)
12/15/2002 04:27:05.331612 61.135.137.66.9013 > 192.168.1.164.41806: P
3091059901:3091060177(276) ack 11834706 win 5840 (DF)
12/15/2002 04:27:05.343507 192.168.1.164.41806 > 61.135.137.66.9013: . ack 276 win 8192
12/15/2002 04:27:05.343748 192.168.1.242.45045 > 210.51.17.96.9065: P
25309490:25309522(32) ack 1436759200 win 8192 (DF)
12/15/2002 04:27:05.359048 192.168.1.242.44991 > 211.167.92.226.6732: P 17:25(8) ack 16
win 8192 (DF)
12/15/2002 04:27:05.359218 192.168.1.83.64228 > 61.242.153.168.11745: udp 92
12/15/2002 04:27:05.359383 192.168.1.164.9668 > 211.150.186.218.4000: udp 60
12/15/2002 04:27:05.359537 192.168.1.83.64228 > 61.242.153.168.11745: udp 92
12/15/2002 04:27:05.359693 192.168.1.83.64228 > 61.242.153.168.11745: udp 92
12/15/2002 04:27:05.359694 61.152.252.11.55901 > 192.168.1.242.45311: P 48:56(8) ack 1
win 62851 (DF)
12/15/2002 04:27:05.362315 210.51.17.96.9065 > 192.168.1.242.45045: . ack 32 win 32120
(DF)
12/15/2002 04:27:05.366415 61.135.137.26.9013 > 192.168.1.242.45533: P 112:138(26) ack 1
win 6432 (DF)
```



# tcpdump trace: TCP options

- Selective acknowledgement (SACK) option: supported by > 60% of connections
- Sliding windows scale option: supported by < 5% of connections
- No instances of path MTU discovery
- Most connections use initial `cwnd` size: 4 segments or greater
- Observations agree with the TCP implementation in Microsoft Windows

MTU: maximum transmission unit

A decorative graphic on the left side of the slide features overlapping yellow, red, and blue squares with a black crosshair.

# Network anomalies

---

- `Ethereal/Wireshark`, `tcptrace`, and `pcapread`
- Four types of network anomalies were detected:
  - invalid TCP flag combinations
  - large number of TCP resets
  - UDP and TCP port scans
  - traffic volume anomalies

# Invalid TCP flag combinations

- TCP SYN flag: signal to establish connections
- TCP FIN flag: signal to terminate connections regularly
- TCP RST flag: signal to terminate connections when error occurs
- TCP PSH flag: signal to transmit all outstanding packets in the buffer without delay
- Invalid combinations are **SYN+FIN**, **SYN+RST**, **RST+FIN**, **RST+PSH**, and **RST+FIN+PSH**
- A single invalid packet may cause a vulnerable TCP/IP implementation to exhibit unexpected behavior



# Analysis of TCP flags

TCP flag	Packet count	% of Total
SYN only	19,050,849	48.500
RST only	7,440,418	18.900
FIN only	12,679,619	32.300
*SYN+FIN	408	0.001
*RST+FIN (no PSH)	85,571	0.200
*RST+PSH (no FIN)	18,111	0.050
*RST+FIN+PSH	8,329	0.020
*Total number of packets with invalid TCP flag combinations	112,419	0.300
Total packet count	39,283,305	100.000



# Large number of TCP resets

- Connections are terminated by either **TCP FIN** or **TCP RST**:
  - 12,679,619 connections were terminated by **FIN** (63%)
  - 7,440,418 connections were terminated by **RST** (37%)
- Large number of **TCP RST** indicates that connections are terminated in error conditions
- **TCP RST** is employed by Microsoft Internet Explorer to terminate connections instead of **TCP FIN**

M. Arlitt and C. Williamson, "An analysis of TCP reset behaviour on the Internet," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 1, pp. 37–44, Jan. 2005.



# UDP and TCP port scans

- UDP port scans are found on UDP port 137 (NETBEUI)
- TCP ports scans are found on these TCP ports:
  - 80 Hypertext transfer protocol (HTTP)
  - 139 NETBIOS extended user interface (NETBEUI)
  - 443 HTTP over secure socket layer (HTTPS)
  - 1433 Microsoft structured query language (MS SQL)
  - 27374 Subseven trojan
- No HTTP(S) servers were active in the ChinaSat network
- MS SQL vulnerability was discovered in Oct. 2002, which may be the cause of scans on TCP port 1433
- The Subseven trojan is a backdoor program used with malicious intents

# UDP port scans originating from the ChinaSat network



192.168.2.30:137 - 195.x.x.98:1025  
192.168.2.30:137 - 202.x.x.153:1027  
192.168.2.30:137 - 210.x.x.23:1035  
192.168.2.30:137 - 195.x.x.42:1026  
192.168.2.30:137 - 202.y.y.226:1026  
192.168.2.30:137 - 218.x.x.238:1025  
192.168.2.30:137 - 202.y.y.226:1025  
192.168.2.30:137 - 202.y.y.226:1027  
192.168.2.30:137 - 202.y.y.226:1028  
192.168.2.30:137 - 202.y.y.226:1029  
192.168.2.30:137 - 202.y.y.242:1026  
192.168.2.30:137 - 61.x.x.5:1028  
192.168.2.30:137 - 219.x.x.226:1025  
192.168.2.30:137 - 213.x.x.189:1028  
192.168.2.30:137 - 61.x.x.193:1025  
192.168.2.30:137 - 202.y.y.207:1028  
192.168.2.30:137 - 202.y.y.207:1025  
192.168.2.30:137 - 202.y.y.207:1026  
192.168.2.30:137 - 202.y.y.207:1027  
192.168.2.30:137 - 64.x.x.148:1027

- Client (**192.168.2.30**) source port (**137**) scans external network addresses at destination ports (**1025-1040**):
  - > 100 are recorded within a three-hour period
  - targets IP addresses are variable
  - multiple ports are scanned for a single IP
  - may correspond to Bugbear, OpaSoft, or other worms

# UDP port scans direct to the ChinaSat network



210.x.x.23:1035 - 192.168.1.121:137  
210.x.x.23:1035 - 192.168.1.63:137  
210.x.x.23:1035 - 192.168.2.11:137  
210.x.x.23:1035 - 192.168.1.250:137  
210.x.x.23:1035 - 192.168.1.25:137  
210.x.x.23:1035 - 192.168.2.79:137  
210.x.x.23:1035 - 192.168.1.52:137  
210.x.x.23:1035 - 192.168.6.191:137  
210.x.x.23:1035 - 192.168.1.241:137  
210.x.x.23:1035 - 192.168.2.91:137  
210.x.x.23:1035 - 192.168.1.5:137  
210.x.x.23:1035 - 192.168.1.210:137  
210.x.x.23:1035 - 192.168.6.127:137  
210.x.x.23:1035 - 192.168.1.201:137  
210.x.x.23:1035 - 192.168.6.179:137  
210.x.x.23:1035 - 192.168.2.82:137  
210.x.x.23:1035 - 192.168.1.239:137  
210.x.x.23:1035 - 192.168.1.87:137  
210.x.x.23:1035 - 192.168.1.90:137  
210.x.x.23:1035 - 192.168.1.177:137  
210.x.x.23:1035 - 192.168.1.39:137

- External address (210.x.x.23) scans for port (137) (NETBEUI) response within the ChinaSat network from source port (1035):
  - > 200 are recorded within a three-hour period
  - targets IP addresses are not sequential
  - may correspond to Bugbear, OpaSoft, or other worms

A decorative graphic on the left side of the slide features overlapping yellow, red, and blue squares with a black crosshair.

# Wavelet transforms

---

- A time series signal is decomposed into different time scales using wavelet transforms
- Each time scale expresses the original signal at different frequencies
- Coarser time scales contain lower frequency approximations of the signal
- Finer time scales contain higher frequency approximations

# Detection of traffic volume anomalies using wavelets



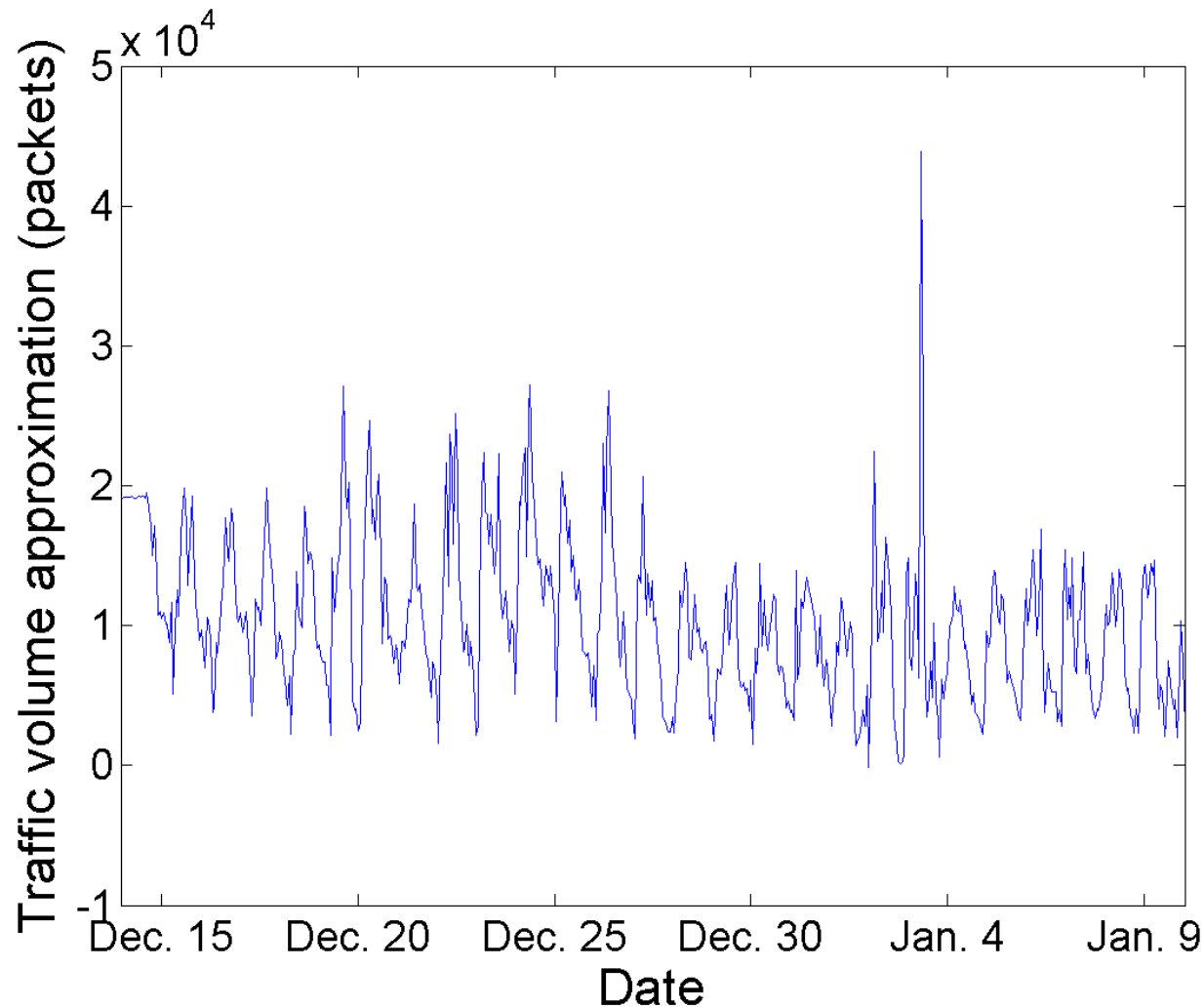
- Traffic is decomposed into different frequencies using the wavelet transform
- Traffic volume anomalies are identified by large variations in wavelet detail coefficient values
- The coarsest scale level where the anomalies is found indicates the time scale of an anomaly

# Detection of traffic volume anomalies using wavelets

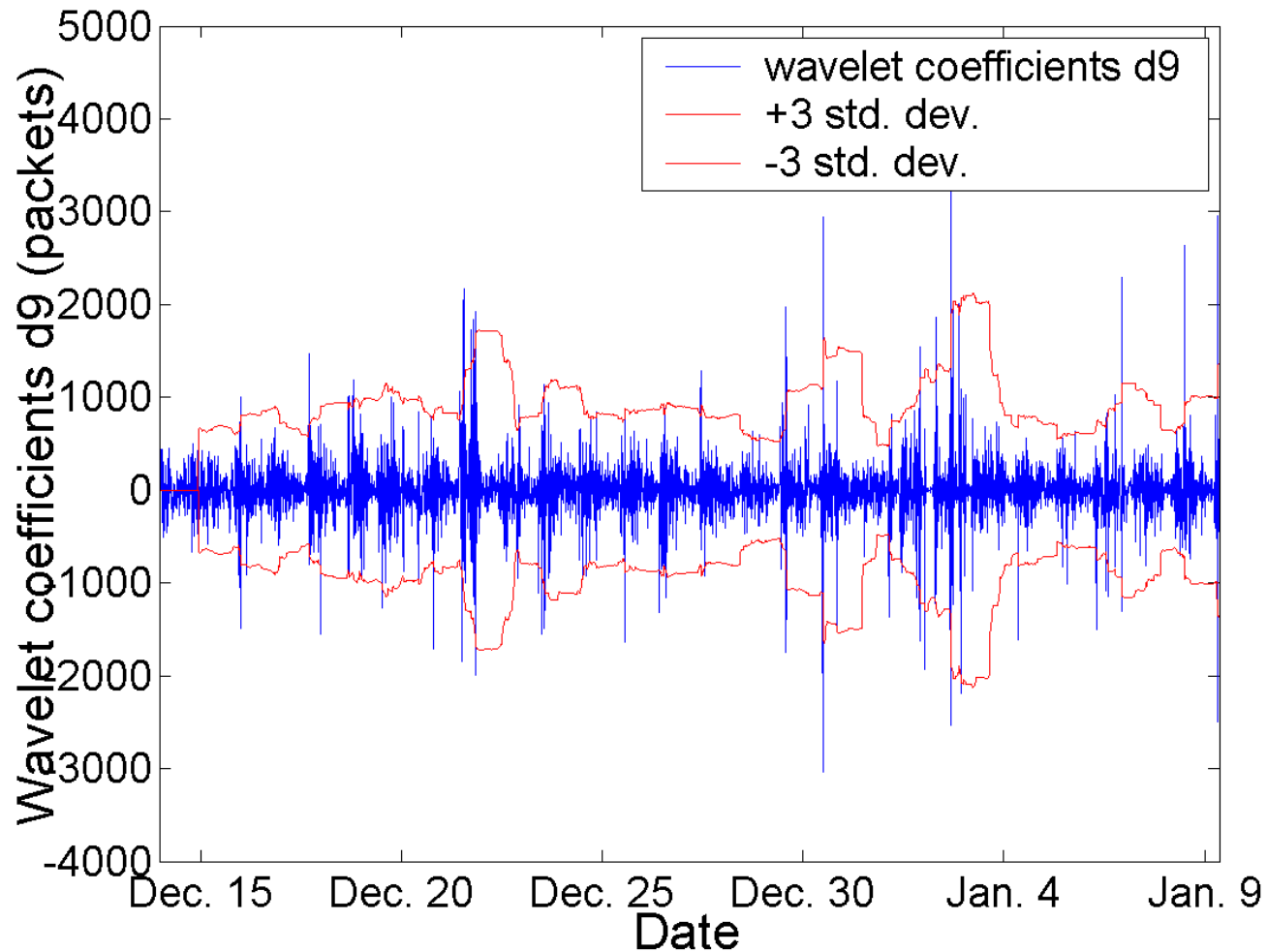


- `tcpdump` trace is binned in terms of packets or bytes (each second)
- Wavelet transform of 12 levels is employed to decompose the traffic
- The coarsest level approximately represents hourly traffic
- Anomalies are:
  - detected with a moving window of size 20 and by calculating the mean and standard deviation ( $\sigma$ ) of the wavelet coefficients in each window
  - identified when wavelet coefficients lie outside  $\pm 3\sigma$  of the mean value

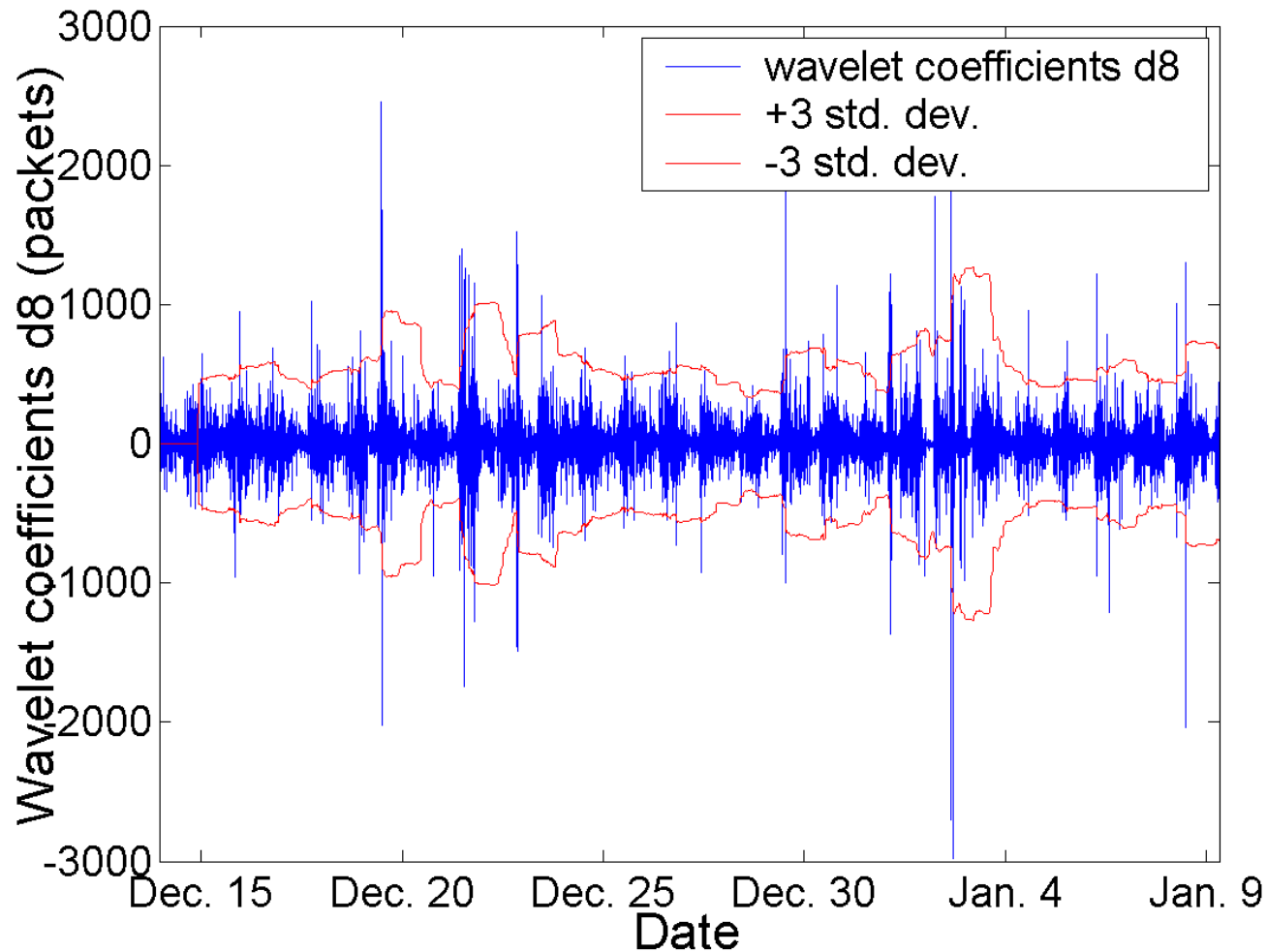
# Wavelet approximation coefficients



# Wavelet detail coefficients: $d_9$



# Wavelet detail coefficients: $d_8$



A decorative graphic on the left side of the slide features overlapping yellow, red, and blue squares with a black crosshair.

# Roadmap

---

- Introduction
- ChinaSat: network architecture and TCP
- Analysis of billing records
- Analysis of tcpdump traces:
  - general characteristics
  - TCP options
  - network anomalies
- Conclusions

A decorative graphic on the left side of the slide features overlapping yellow, red, and blue squares, intersected by a black crosshair.

# Conclusions

---

- We analyzed **billing** records and **tcpdump** traces from a hybrid satellite-terrestrial network operated by ChinaSat
- **Billing** records:
  - daily and weekly cycles
  - number of packets: similar number for uploaded and downloaded traffic
  - number of bytes: downloaded traffic is an order of magnitude higher than uploaded traffic
  - minority of users contributed most of the traffic

A decorative graphic on the left side of the slide features overlapping yellow, red, and blue squares, intersected by a black crosshair.

# Conclusions

---

- `tcpdump` trace:
  - TCP accounts for majority of traffic
  - TCP options most widely used to improve performance are SACK and increasing initial windows size
  - ChinaSat DirecPC hosts may be optimized by:
    - ensuring the SACK option is enabled on all hosts
    - enabling the sliding window scale option
  - network anomalies are detected using open source tools and wavelet decomposition



# References

- Q. Shao and Lj. Trajkovic, "Measurement and analysis of traffic in a hybrid satellite-terrestrial network," in *Proc. SPECTS 2004*, San Jose, CA, July 2004, pp. 329–336.
- J. Postel, Ed., "Transmission Control Protocol," RFC 793, Sept. 1981.
- J. Postel, "TCP and IP bake off," RFC 1025, Sept. 1987.
- J. Mogul and S. Deering, "Path MTU discovery," RFC 1191, Nov. 1990.
- V. Jacobson, R. Braden, and D. Borman, "TCP extensions for high performance," RFC 1323, May 1992.
- M. Allman, S. Floyd, and C. Partridge, "Increasing TCP's initial window," RFC 2414, Sept. 1998.
- M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow, "TCP selective acknowledgment options," RFC 2018, Oct. 1996.
- M. Allman, D. Glover, and L. Sanchez, "Enhancing TCP over satellite channels using standard mechanisms," RFC 2488, Jan. 1999.
- M. Allman, S. Dawkins, D. Glover, J. Griner, D. Tran, T. Henderson, J. Heidemann, J. Touch, H. Kruse, S. Ostermann, K. Scott, and J. Semke, "Ongoing TCP research related to satellites," RFC 2760, Feb. 2000.
- J. Border, M. Kojo, J. Griner, G. Montenegro, and Z. Shelby, "Performance enhancing proxies intended to mitigate link-related degradations," RFC 3135, June 2001.
- S. Floyd, "Inappropriate TCP resets considered harmful," RFC 3360, Aug. 2002.



# References

- D. E. Comer, *Internetworking with TCP/IP, Vol 1: Principles, Protocols, and Architecture*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2000.
- W. R. Stevens, *TCP/IP Illustrated (vol. 1): The Protocols*. Reading, MA: Addison-Wesley, 1994.
- R. Beverly, "A Robust Classifier for Passive TCP/IP Fingerprinting," in *Proc. Passive and Active Meas. Workshop 2004*, Antibes Juan-les-Pins, France, Apr. 2004, pp. 158–167.
- C. Smith and P. Grundl, "Know your enemy: passive fingerprinting," The HoneyNet Project, Mar. 2002. [Online]. Available: <http://www.honeynet.org/papers/finger/>.
- Passive OS fingerprinting tool ver. 2 (p0f v2). [Online]. Available: <http://lcamtuf.coredump.cx/p0f.shtml/>.
- B. Petersen, "Intrusion detection FAQ: What is p0f and what does it do?" The SysAdmin, Audit, Network, Security (SANS) Institute. [Online]. Available: <http://www.sans.org/resources/idfaq/p0f.php>.
- T. Miller, "Passive OS fingerprinting: details and techniques," The SysAdmin, Audit, Network, Security (SANS) Institute. [Online]. Available: <http://www.sans.org/readingroom/special.php/>.



# References

---

- P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies," in *Proc. ACM SIGCOMM Internet Meas. Workshop 2001*, Nov. 2001, pp. 69–73.
- P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in *Proc. ACM SIGCOMM Internet Meas. Workshop 2002*, Marseille, France, Nov. 2002, pp. 71–82.
- Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan, "Network anomography," in *Proc. ACM SIGCOMM Internet Meas. Conf. 2005*, Berkeley, CA, Oct. 2005, pp. 317–330.
- A. Soule, K. Salamatian, and N. Taft, "Combining filtering and statistical methods for anomaly detection," in *Proc. ACM SIGCOMM Internet Meas. Conf. 2005*, Berkeley, CA, Oct. 2005, pp. 331–344.
- P. Huang, A. Feldmann, and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," in *Proc. ACM SIGCOMM Internet Meas. Workshop 2001*, San Francisco, CA, Nov. 2001, pp. 213–227.
- A. Lakhina, M. Crovella, and C. Diot, "Characterization of network-wide anomalies in traffic flows," in *Proc. ACM SIGCOMM Internet Meas. Conf. 2004*, Taormina, Italy, Oct. 2004, pp. 201–206.
- A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 4, pp. 219–230, Oct. 2004.
- M. Arlitt and C. Williamson, "An analysis of TCP reset behaviour on the Internet," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 1, pp. 37–44, Jan. 2005.