



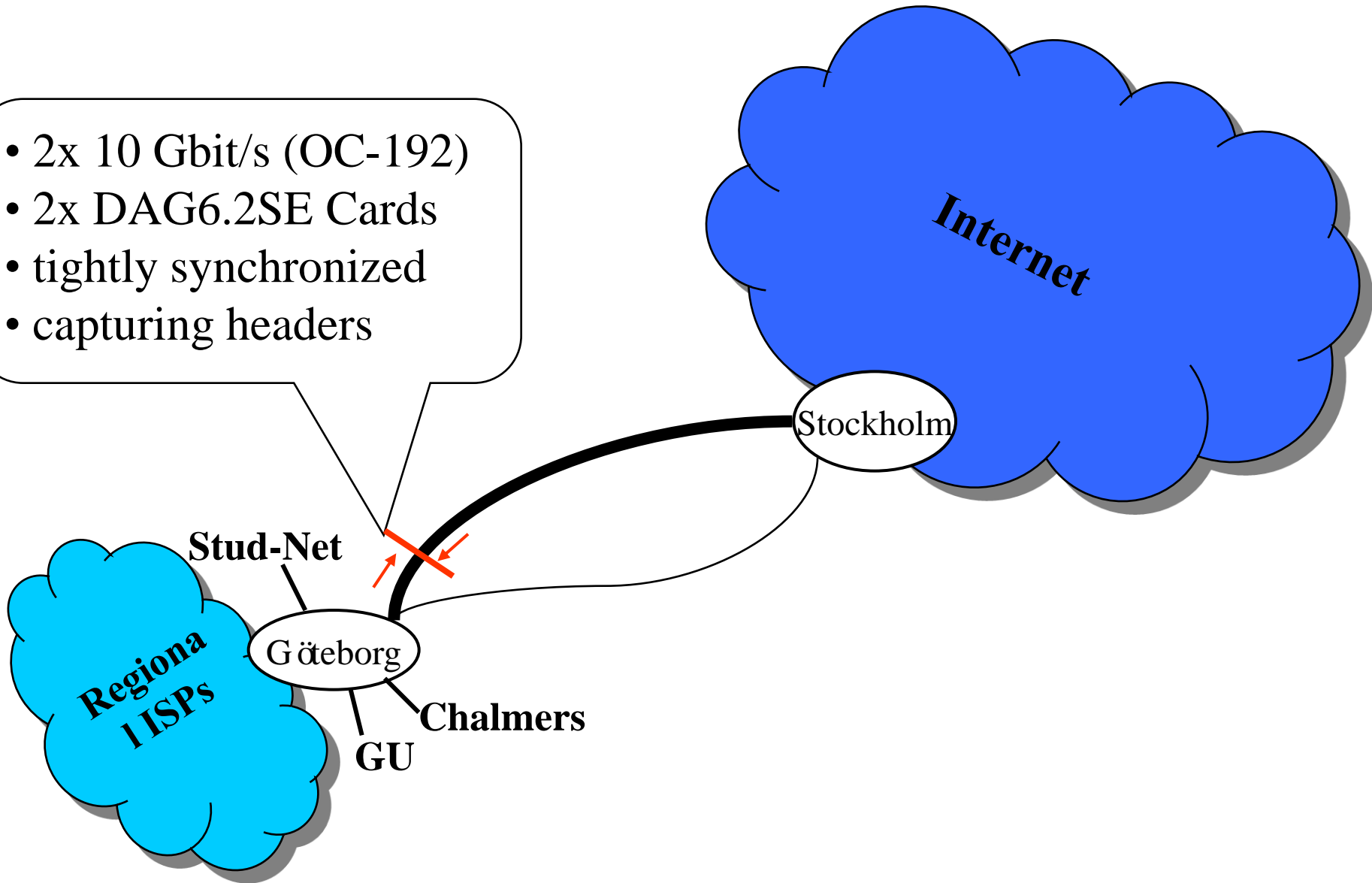
What does the Internet traffic in Sunet look like in the MonNet traces?

Sven Tafvelin

Dept. of Computer Science and Engineering
Chalmers University of Technology
Göteborg, Sweden

Measurement location between Göteborg and Borås

- 2x 10 Gbit/s (OC-192)
- 2x DAG6.2SE Cards
- tightly synchronized
- capturing headers



The MonNet spring 2006 traces

- Data from 20 days in April 2006
- They were taken 20 minutes long in both directions
- The starting times were 02:00, 10:00, 14:00 and 20:00
- 146 traces, 10.7 billion frames, 7.5 TB
- We were only allowed to keep the protocol header data and the IP-addresses should be anonymised.

IPv4 protocol breakdown (in %)

	02:00		10:00		14:00		20:00	
	Pkts	Data	Pkts	Data	Pkts	Data	Pkts	Data
TCP	91.3	97.6	91.5	96.8	93.2	97.1	91.4	97.2
UDP	8.5	2.3	7.6	2.8	6.1	2.7	8.3	2.7
ICMP	0.2	0.02	0.19	0.02	0.20	0.02	0.12	0.01
ESP	0.01	0.00	0.47	0.19	0.35	0.14	0.02	0.02
GRE	0.01	0.01	0.08	0.08	0.04	0.03	0.06	0.04

UDP burst (%)

Outgoing

Date	Time	Packets	Data
2006-04-16	14:00	6.8	1.7
2006-04-16	20:00	40.6	5.1
2006-04-17	02:00	51.9	6.1
2006-04-17	10:00	58.1	7.1
2006-04-17	14:00	5.7	1.8

TCP options

In SYN/ACK
segments

Kind	02:00	10:00	14:00	20:00
2(MSS)	99.0%	98.7%	99.7%	99.1%
3(WS)	21.4%	18.4%	16.6%	16.5%
4(SACK perm)	91.0%	86.6%	88.9%	89.8%
8(TS)	18.2%	15.3%	13.3%	12.8%
No opt	86.5%	85.2%	87.3%	88.6%
SACK	3.1%	2.8%	2.9%	3.1%
TS	9.7%	11.2%	9.0%	7.6%
MD5	0.02%	0.02%	0.01%	0.01%

In all segments

TCP header anomalies

Options

Anomaly	02:00	10:00	14:00	20:00
Undef opt	1062	507	413	388
Inv opt length	1200	399	915	3020
Inv hdr length	71	528	130	119
RST+SYN+FIN	8	35	11	15
RST+SYN	25	70	43	27
SYN+FIN	4	22	8	9
Zero flags	32	78	86	90
RST+FIN	10200	10988	14320	16334

Flags

Malicious traffic (1)

- Distinct IP addresses seen

	Total		Outbound		Inbound	
	Inside	Outside	Source	Dest.	Dest.	Source
Total	0.63E+6	22.0E+6	0.27E+6	19.2E+6	0.49E+6	19.8E+6
TCP	0.41E+6	05.0E+6	0.18E+6	04.3E+6	0.31E+6	04.5E+6
UDP	0.48E+6	19.2E+6	0.18E+6	16.4E+6	0.38E+6	16.9E+6
Rest	0.15E+6	01.9E+6	0.02E+6	01.1E+6	0.15E+6	01.0E+6

Malicious traffic (2)

- Connection attempt breakdown

	total		outbound		inbound	
	Count	%	Count	%	Count	%
TCP connections	72.6E+6	100.00%	28.0E+6	38.56% (100.00%)	44.6E+6	61.44% (100.00%)
rejected	44.3E+6	60.99%	12.3E+6	(44.04%)	<u>32.0E+6</u>	(71.63%)
established	28.3E+6	39.01%	15.7E+6	(55.96%)	12.7E+6	(28.37%)

rejected connections	44.3E+6	100.00%	12.3E+6	27.84% (100.00%)	32.0E+6	72.16% (100.00%)
scanning - no reply	34.8E+6	78.66%	08.2E+6	(66.74%)	<u>26.6E+6</u>	(83.26%)
asymmetric traffic	04.8E+6	10.84%	02.2E+6	(17.94%)	02.6E+6	(8.10%)
scanning - RST reply	04.3E+6	9.81%	01.7E+6	(13.83%)	02.6E+6	(8.25%)

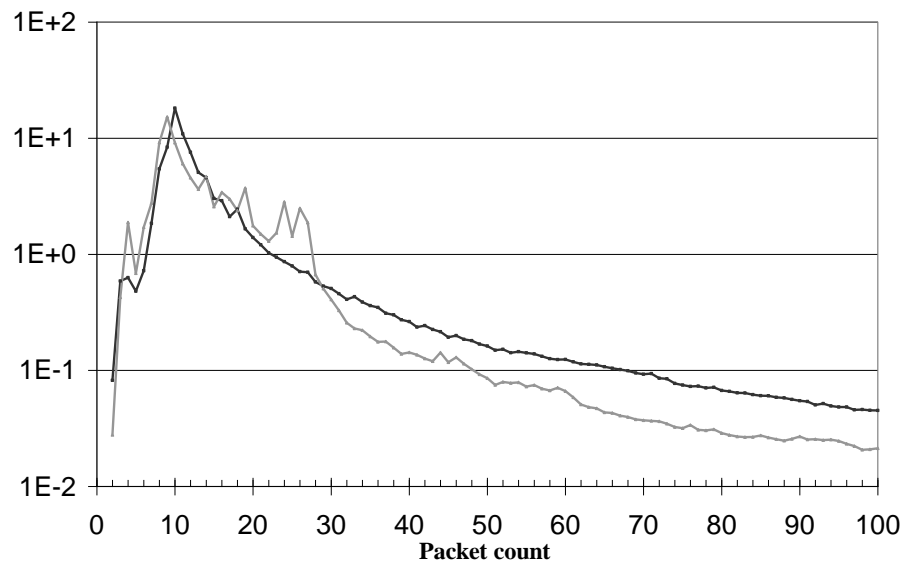
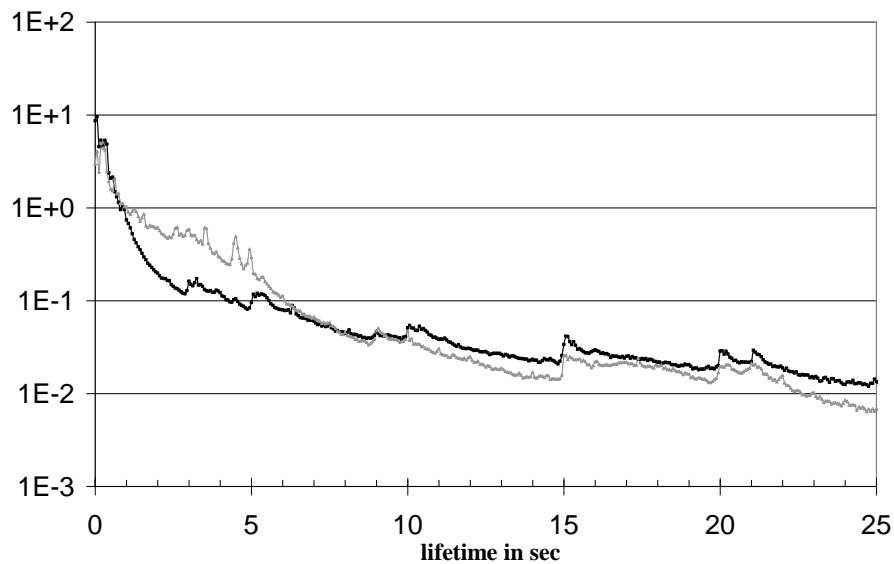
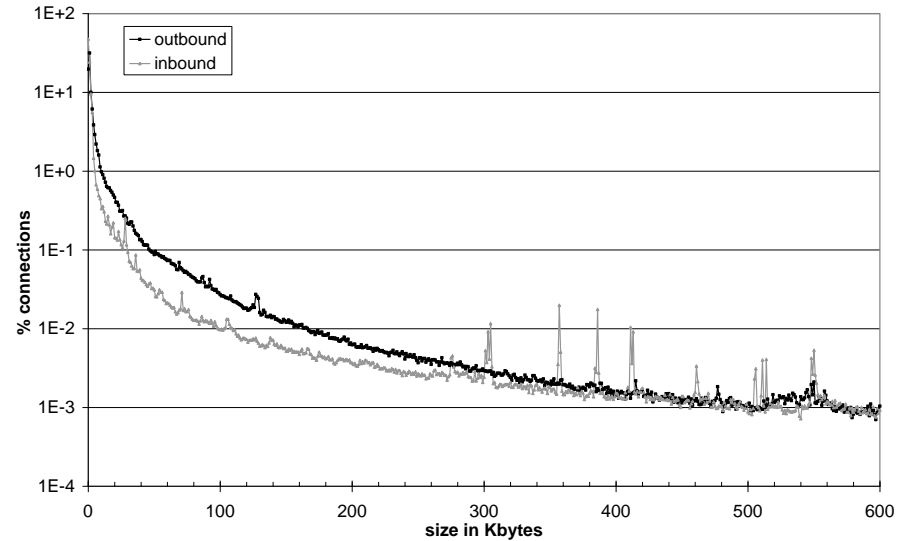
Malicious traffic/ P2P traffic

- Established TCP connection properties
- Inbound connections more likely to:
 - show lifetimes between 1 and 5 seconds
 - be long lasting (>10 minutes)
 - carry more data and more packets
 - show higher asymmetry (client-server pattern)

Property		mean	σ	median	P80
Lifetime in sec	out	18.2	60.7	1.8	16.6
	in	17.3	65.8	0.6	24.8
Size in Kbytes	out	61.0	2362	1.1	2.9
	in	81.5	3298	1.9	8.9
Packet Count	out	81.5	2289	11.5	22.0
	in	113.0	3538	11.5	21.0

Malicious traffic / P2P traffic

- Connection properties



- Quantification according to port-numbers:
 - 13% of data in outbound connections
 - 25% of data in inbound connections
- Missing payload -> underestimates!

Common P2P port numbers

TCP:	
688[0-9]	bittorrent
49200	bittorrent
32459	bittorrent (μ torrent)
49152	bittorrent (μ torrent)
41[1-2]	direct connect (dc++)
1412	direct connect (dc++)
6346	gnutella (limewire)
6348	gnutella (bearshare)
466[0-8]	overnet (edonkey)
14662	overnet (edonkey)
14663	overnet (edonkey)

UDP:	
688[0-9]	bittorrent
49200	bittorrent
32459	bittorrent (μ torrent)
41[1-2]	direct connect (dc++)
1412	direct connect (dc++)
9183	dc++ kademlia
6346	gnutella (limewire)
6348	gnutella (bearshare)
466[0-8]	overnet (edonkey)
4672	overnet (edonkey)
14672	overnet (edonkey)

- Packet sizes

Packet size	total	outbound	inbound
20-39	0.14%	0.18%	0.11%
40-60	39.25%	38.41%	40.02%
576	0.98%	0.63%	1.30%
628	1.79%	2.12%	1.49%
1300	1.13%	1.25%	1.01%
1400-1500	38.53%	38.62%	38.45%

P2P traffic (4)

- TCP termination behavior

	total		outbound		inbound	
	Count	%	Count	%	Count	%
established connections	28.3E+6	100.00%	15.7E+6	55.21% (100.00%)	12.7E+6	44.68% (100.00%)
proper closing (2xFIN)	19.0E+6	66.99%	11.4E+6	(72.87%)	07.6E+6	(59.71%)
FIN and RST outbound	03.2E+6	11.21%	542E+3	(3.46%)	02.6E+6	<u>(20.81%)</u>
FIN and RST inbound	01.7E+6	6.06%	711E+3	(4.54%)	01.0E+6	(7.93%)
single RST	02.2E+6	7.71%	01.6E+6	(9.98%)	620E+3	(4.89%)
FIN, RST in counter dir.	01.2E+6	4.11%	889E+3	(5.67%)	276E+3	(2.18%)
unclosed	01.0E+6	3.63%	487E+3	(3.11%)	540E+3	(4.27%)

P2P traffic (5)

- TCP options

	MSS			WS		
	SYN	SYN/ACK	both	SYN	SYN/ACK	both
outbound	100.00%	99.59%	99.59%	19.36%	15.46%	15.46%
inbound	99.94%	99.92%	99.85%	24.33%	23.85%	23.83%

	SACK			TS		
	SYN	SYN/ACK	both	SYN	SYN/ACK	both
outbound	93.67%	69.70%	69.70%	16.50%	12.32%	12.32%
inbound	97.22%	90.40%	90.38%	19.72%	18.51%	18.50%

P2P traffic (6)

- 68 million UDP flows
- 51 million carry less than 3 packets!
- DNS: 5%; NTP 1.7%
- P2P overlay traffic: at least 20%!
- Distr. Hash table (DHT) like Kademlia
 - Update routing tables in decentralized way
 - Periodic “ping” queries and replies
- P2P overlay networks span entire globe
- High fluctuation in peering partners -> lots of IPs

Conclusions

- Connection analysis revealed the importance of
 - Malicious traffic
 - P2P for transport of data
 - P2P traffic for keeping the overlay network connected
- High level analysis does not necessarily show differences -> detailed analysis does!



Complete articles available:

Wolfgang John and Sven Tafvelin: *Analysis of Internet backbone traffic with focus on header anomalies*, Chalmers, 2007, Submitted for publication

Wolfgang John and Sven Tafvelin: *Differences between in- and outbound Internet Traffic*, to be presented at TNC in Copenhagen in May 2007.

I think that there are at least two topics in traffic measurements which could have scientific significance:

1. Delay measurements between Scandinavian and China
2. Comparison on what the academic network traffic look like in Scandinavia and China

- 1) Delay measurements between Scandinavia and China is interesting because:
 - The distance is really long and not optimal in any way.
 - It will probably be improved if/when Glorriad connection becomes available
 - It is an interesting set up for delay experiments and QoS focusing on delay reduction and more predictable delays.

2) Is the academic Internet traffic in China the same as in Scandinavia? Probably not.

Reasons could be:

- Different penetration levels among students
- Effects of the different cultures
- Rules on how the network may be used
- Political pressure on network users
- ??? ??? ???