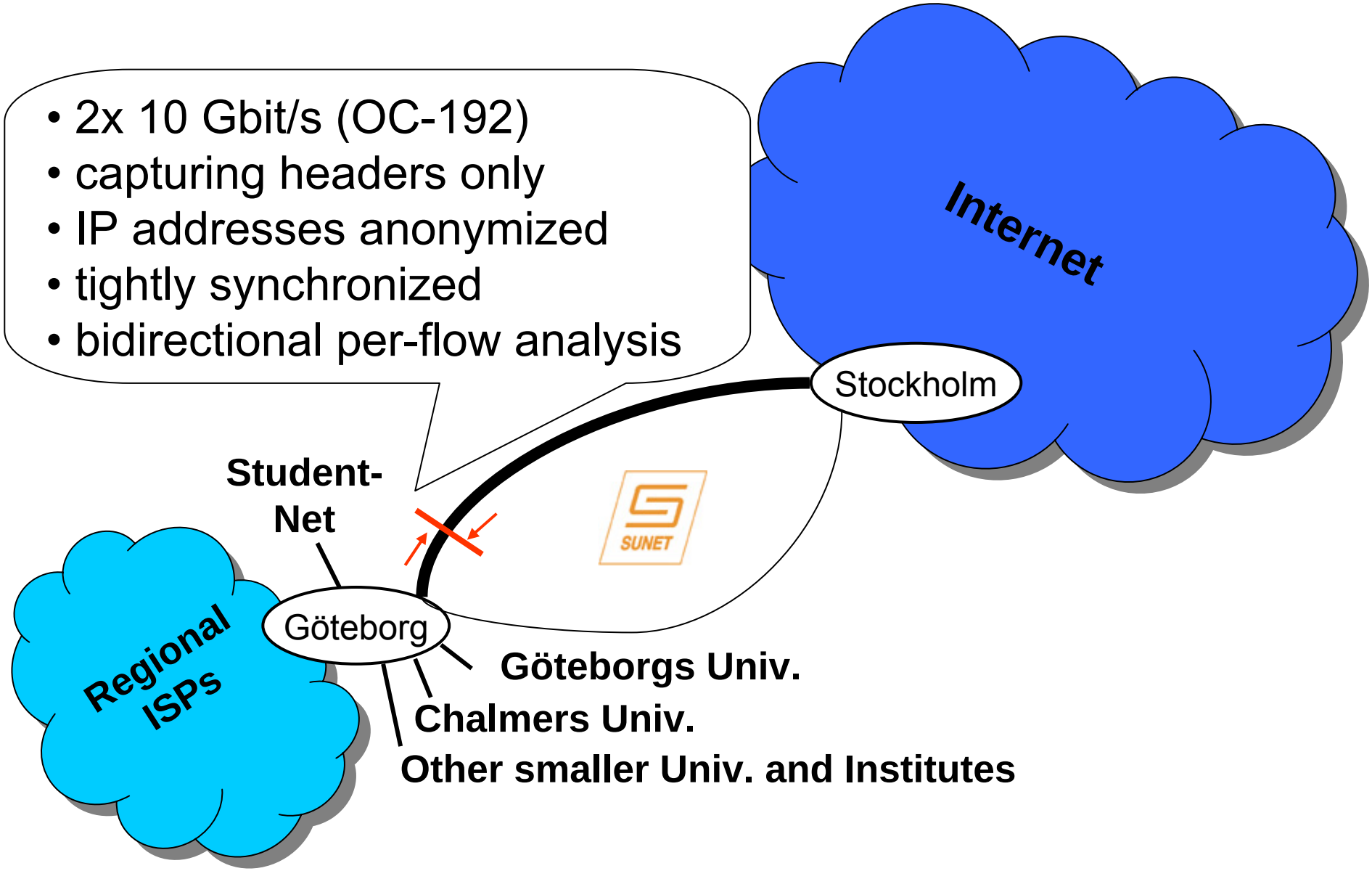
Trends and Differences in Connection-behavior within Classes of Internet Backbone Traffic

Wolfgang John, Sven Tafvelin and Tomas Olovsson
Department of Computer Science and Engineering
Chalmers University of Technology
Göteborg, Sweden

Introduction: Overview

1. **Background**
   - Dataset
   - Traffic classification
2. Results
   - Traffic volumes
   - Diurnal patterns
   - Signaling behavior
   - Option deployment
3. Summary and Conclusions

Background: Measurement location



Background: Datasets

Resulting traces (10 minutes duration)

- *April 2006*
  146 bidirectional traces, 7.5 TB of data
  81 million TCP connections
  91 million UDP flows
- *Fall 2006* (Sep. – Nov.)
  65 bidirectional traces, 5.0 TB of data
  49 million TCP connections
  70 million UDP flows

More Info: CAIDA‘s Datcat, “SUNET OC 192 Traces”

Background: Motivation

- Previous studies

  "*Analysis of Internet Backbone Traffic and Anomalies observed*" (IMC 07)
  "*Differences between in- and outbound Internet Backbone Traffic*" (TNC 07)

  →Influence of P2P and malicious traffic

- How are different types of traffic behaving 'in the wild'?
  - Improving simulation models
  - Developing infrastructure, applications and protocols
  - Finding trends and changes in network applications

Background: Traffic Classification

- Traffic classification necessary
  - Four approaches in literature:
    1. Port numbers
       + easy to implement
       - unreliable (P2P, malicious traffic)

    2. Packet payloads
       + accurate
       - requires updated payload signatures
       - privacy and legal issues
       - data encryption

Background: Traffic Classification (2)

- Traffic classification (contd.)

3. Statistical fingerprinting
   + no detailed packet information needed
   - depending on quality of training data
   - promising, but still immature

4. Connection patterns
   + no payload required
   + no training data required
   - not perfect accuracy

Background: Proposed Heuristics

- Rules based on connection patterns and port numbers

  Inspired by
  Karagiannis et al. 2004: "*Transport layer identification of P2P traffic*"
  Perenyi et al. 2006: "*Identification and analysis of P2P traffic*“

  – 5 rules for P2P traffic
  – 10 rules to classify other types of traffic

Background: Proposed Heuristics (2)

- Main traffic classes
  - P2P file sharing traffic
  - Web traffic (HTTP, HTTPS)
  - Malicious traffic (scans, sweeps and DoS)
  - Other traffic (mail, messenger, ftp, dns …)

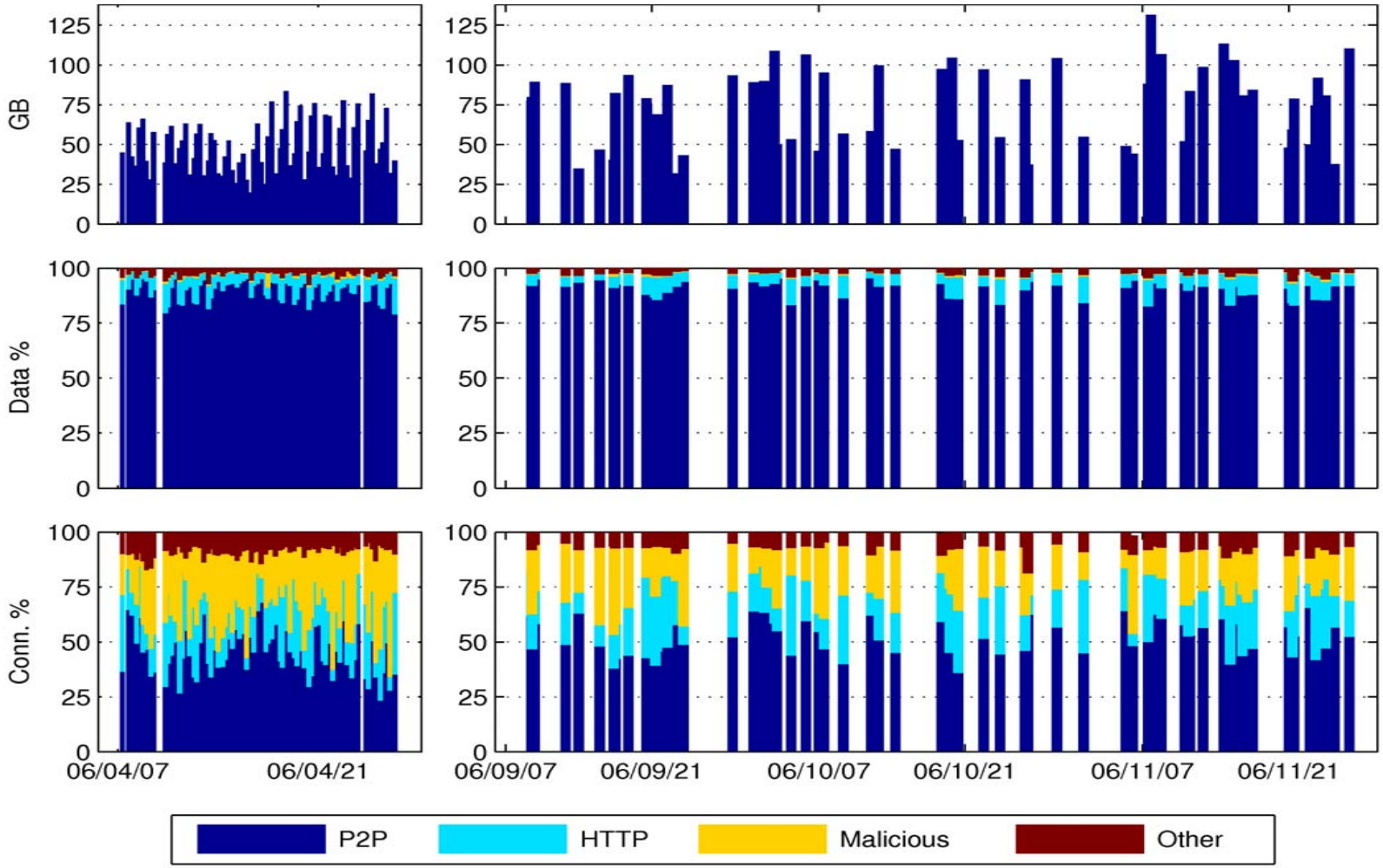
More Info: "*Heuristics to Classify Internet Backbone Traffic based on Connection Patterns*" (ICOIN 08)

Overview (2)

1. Background
   - Dataset
   - Traffic classification
2. **Results**
   - Traffic volumes
   - Diurnal patterns
   - Signaling behavior
   - Option deployment
3. Summary and Conclusions

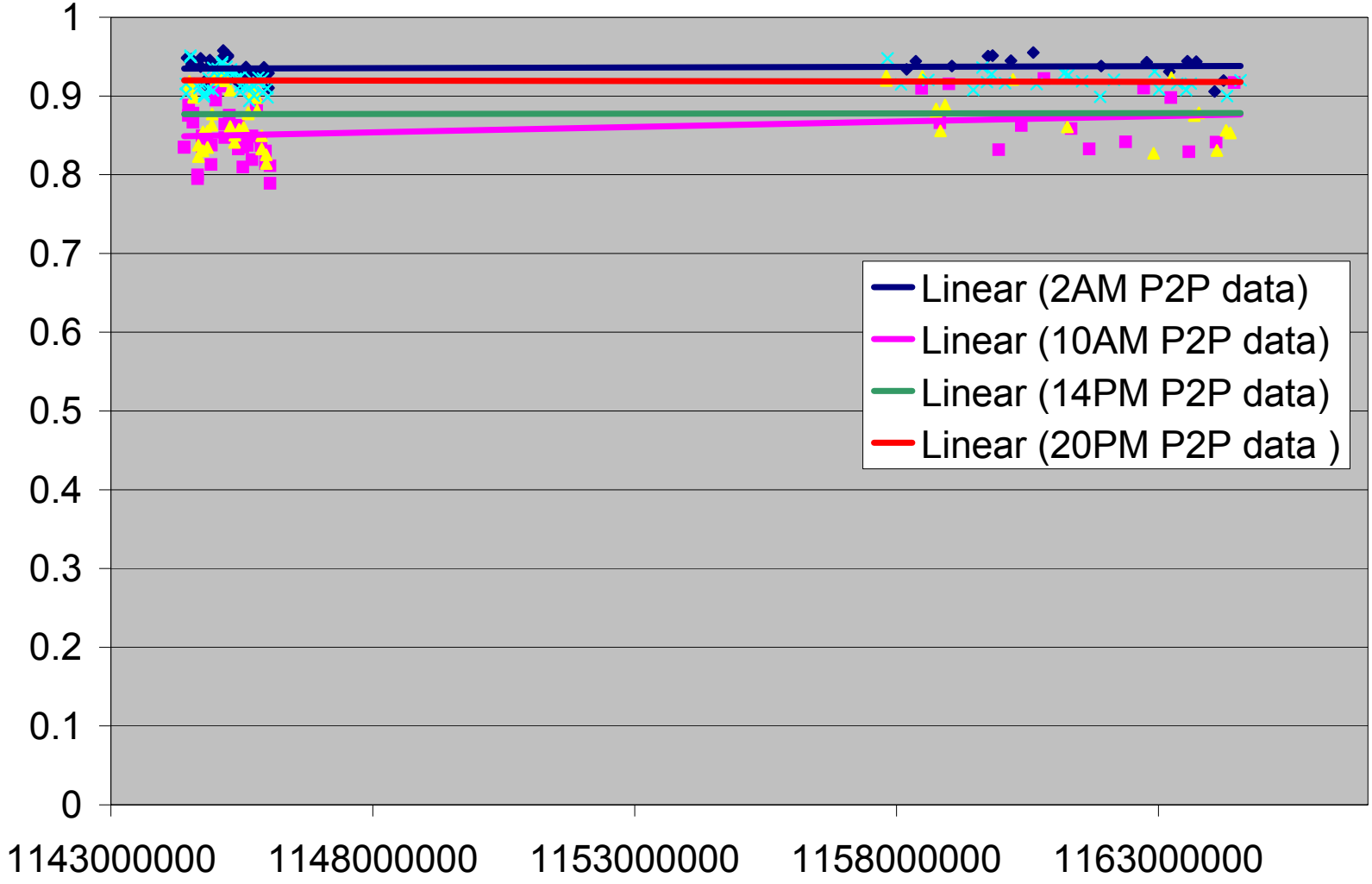
Results: Traffic Volumes

- Application Breakdown April till Nov. 2006



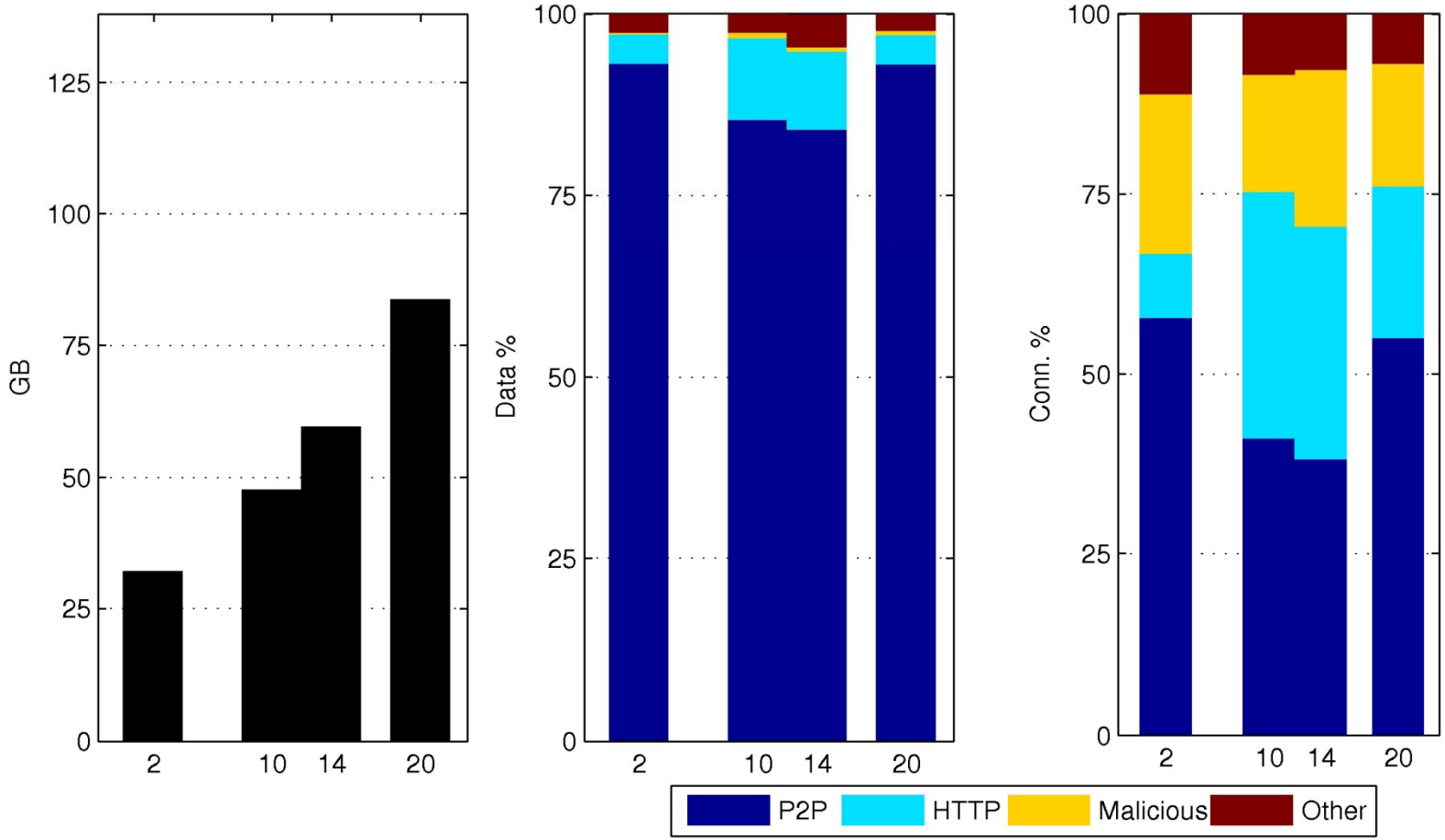
Results: Traffic Volumes (2)

- Fractions of P2P data, April till November



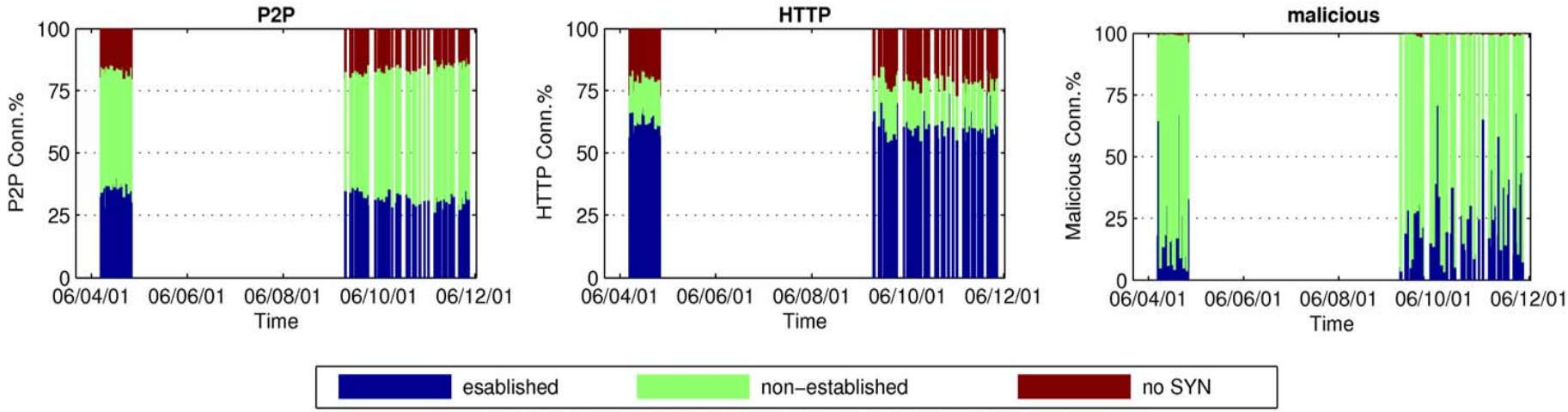
Results: Diurnal Patterns

- Tuesday, 18.04.2006



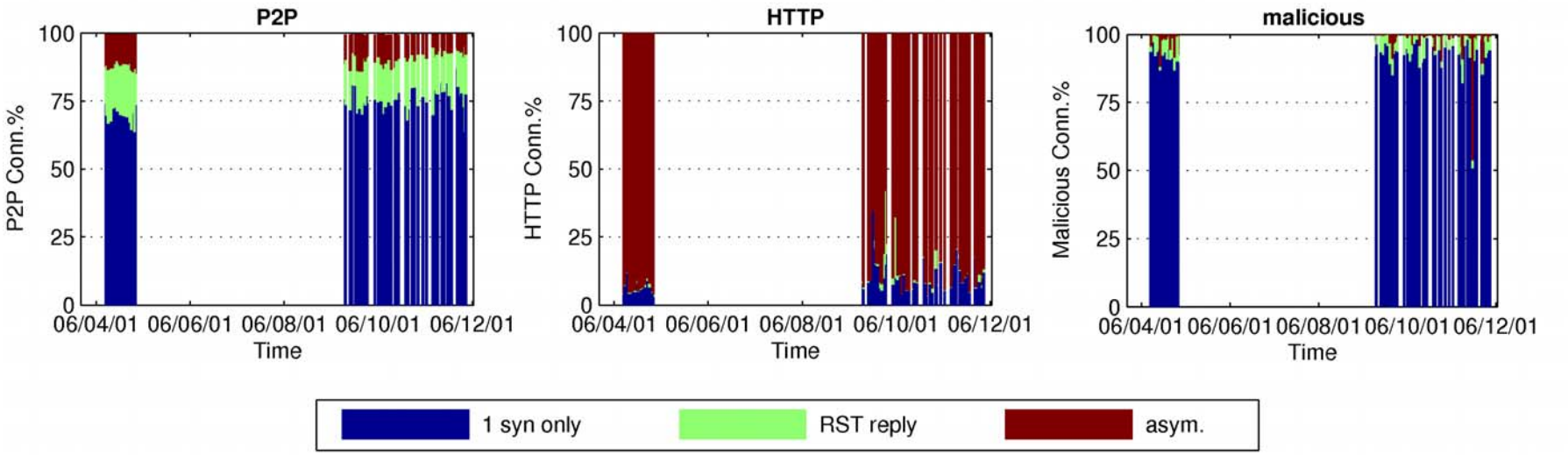
Results: Signaling Behavior

- Connection establishment for P2P, Web and malicious traffic



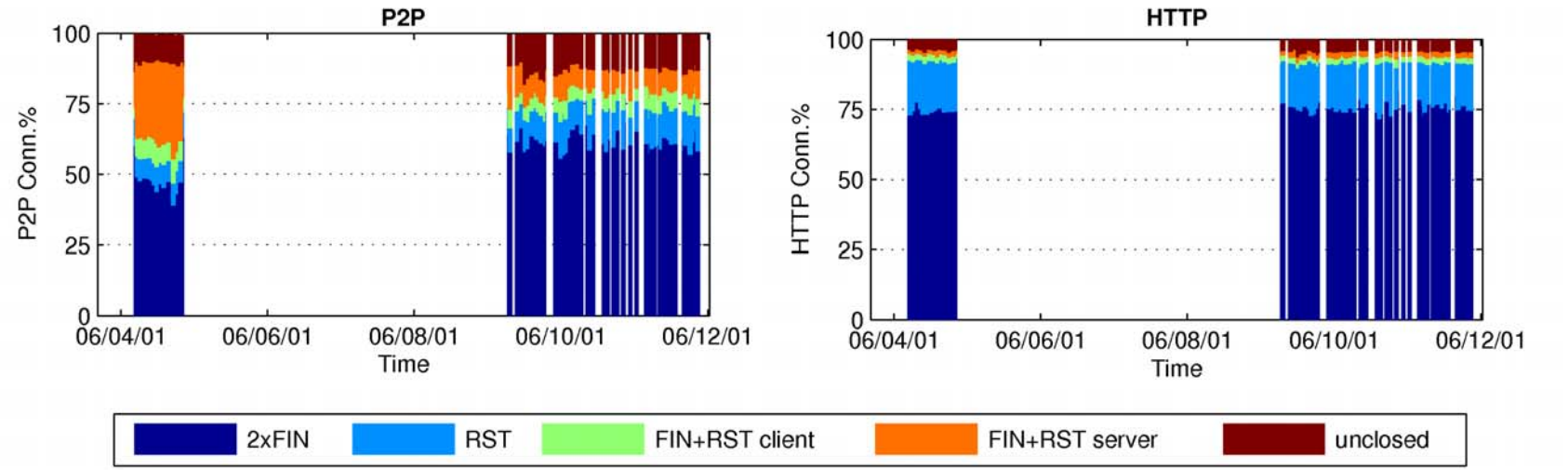
Results: Signaling Behavior (2)

- Breakdown of non-established TCP conn.



Results: Signaling Behavior (3)

- Breakdown of established TCP connections



Results: Option Deployment

- Differences in TCP option deployment

	MSS	SACK	WS	TS
estab.	99.9%	91.0%	14.9%	8.8%
neglected	0.1%	6.5%	0.6%	1.0%
none	0.0%	2.5%	84.5%	90.2%

(a) TCP Options in P2P Conn.

	MSS	SACK	WS	TS
estab.	99.6%	65.7%	16.0%	13.4%
neglected	0.4%	27.9%	4.3%	4.3%
none	0.0%	6.4%	79.7%	82.3%

(b) TCP Options in HTTP Conn.

Summary and Conclusions

- P2P dominating (~90 % of data volume)
  - P2P peak time at evening and night-time
  - Web peak time during office hours
- P2P connections carry large amounts of data
- Traffic is increasing for TCP and UDP
- Fractions of P2P and Web constant
- Malicious traffic constant in absolute numbers
  →'background noise'

Summary and Conclusions (2)

- Major differences in signaling behavior
  - 43% of TCP P2P connections 1-packet flows (attempts)
  - 80% of malicious TCP traffic 1-packet flows (scans)
  - Web traffic behaving ‘nicely‘
- Different TCP options deployment
  - P2P behaves as expected
  - Web traffic shows artifacts of client-server patter
    e.g. popular web-servers neglecting SACK option

More Information:

http://www.chalmers.se/cse/EN/people/john-wolfgang

or Email: johnwolf@chalmers.se

Questions?