Characteristics of Real Open SIP-Server Traffic

Jan Stanek, Lukas Kencl, Jiri Kuthan

CZECH TECHNICAL UNIVERSITY
IN PRAGUE

Tekelec
For What's Next

Outline

- Brief SIP introduction
- SIP server & SIP dataset description
- Analysis of SIP traffic
- Future plans
- Discussion

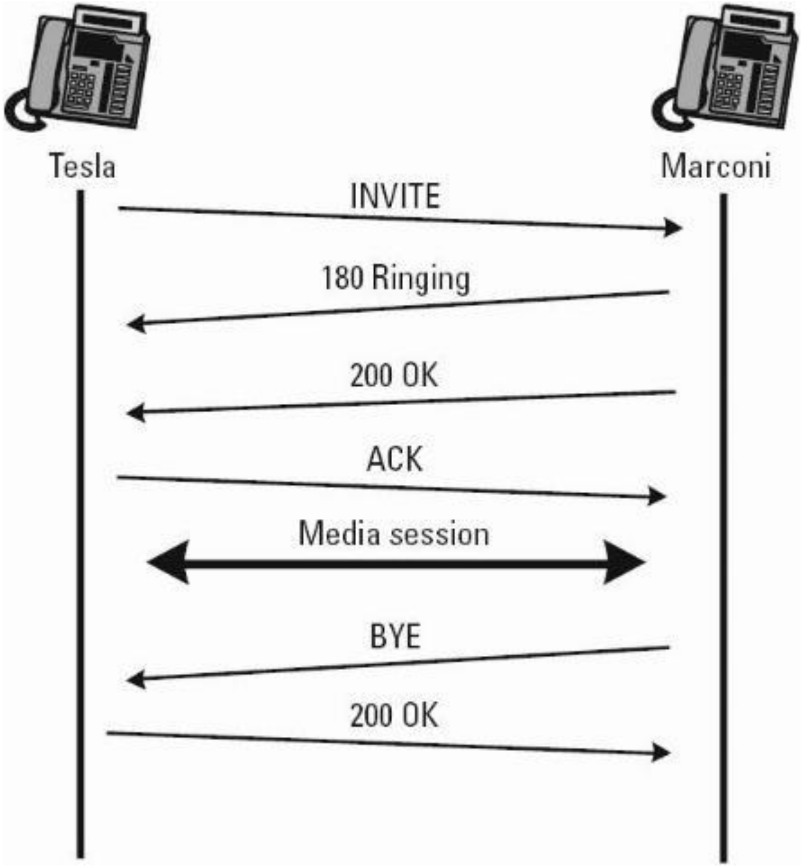
Session Initiation Protocol (SIP) [1]

- **Signaling** protocol designed for controlling multimedia sessions
- Widely used in VoIP
- Emerging in mobile core networks
- Tested in content delivery networks (CDNs)
- Structurally similar to HTTP

[1] SIP: Session Initiation Protocol - http://www.ietf.org/rfc/rfc3261.txt

SIP messages

- Requests
  - Starting with keyword
  - 6 basic + 8 extensions
  - INVITE, ACK, BYE, CANCEL,
  OPTIONS, REGISTER
- Responses
  - Starting with keycode
  - Six classes



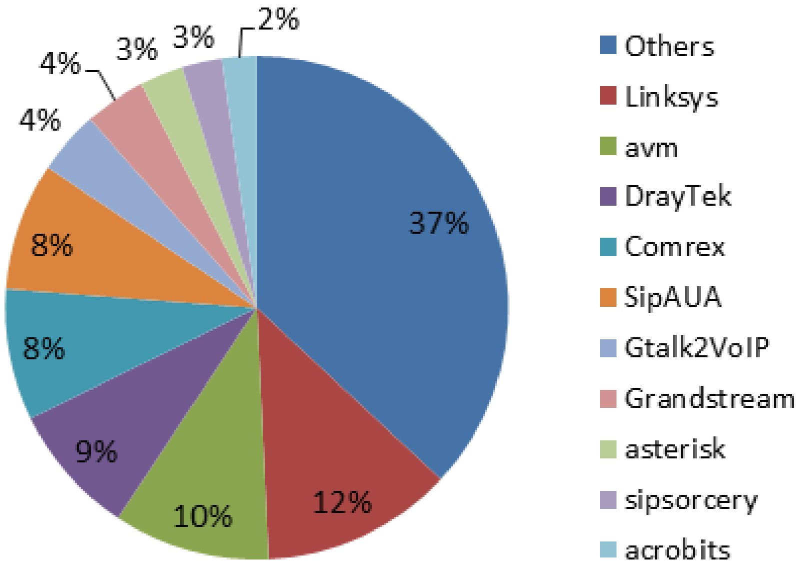
1XX	Provisional	2XX	Success	3XX	Redirection
4XX	Client Failure	5XX	Server Failure	6XX	Global Failure

SIP server & SIP dataset

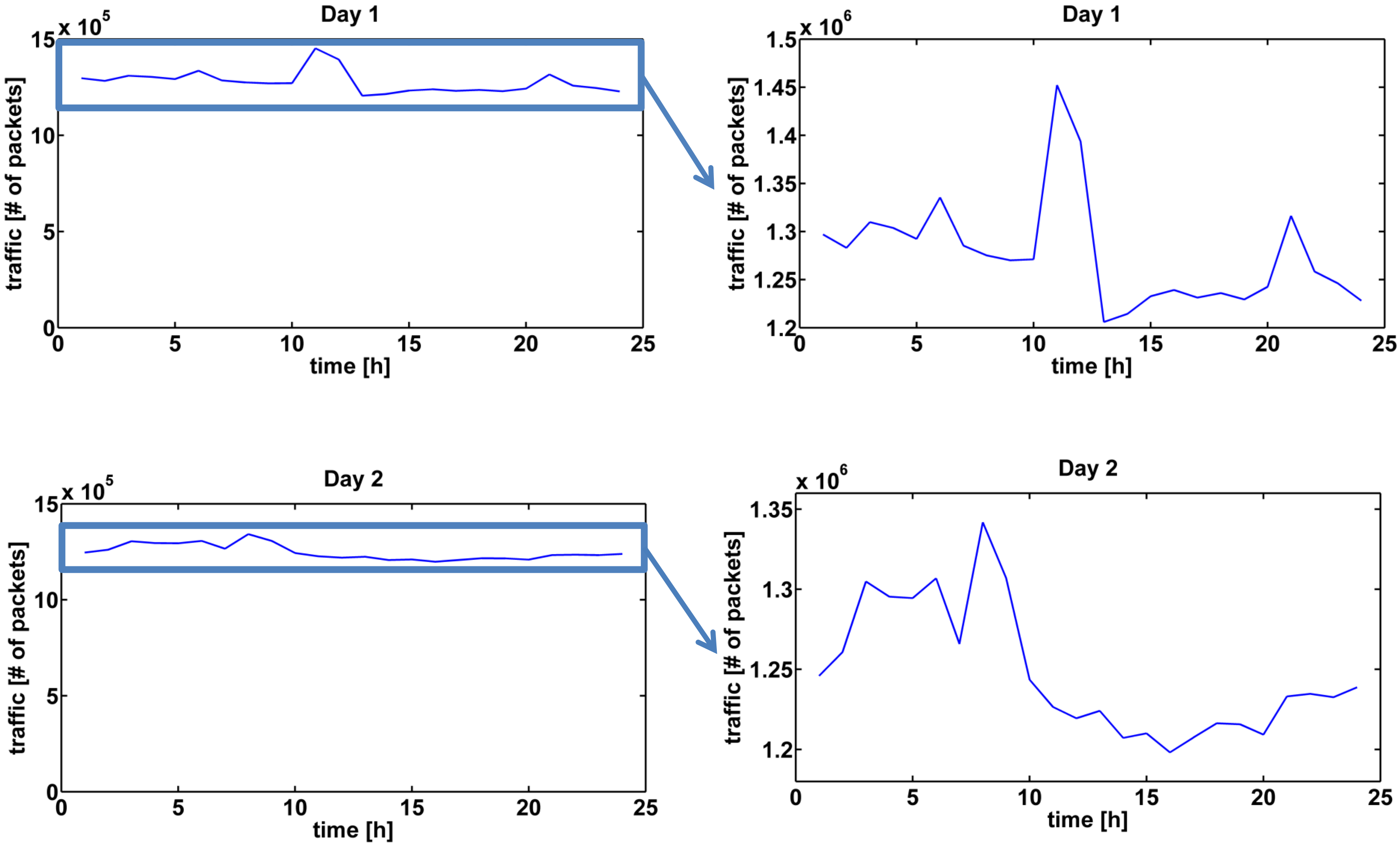
iptel.org

- SIP server
  - Open, public & free experimental SIP service
  - SIP Express Router on a single host blade server

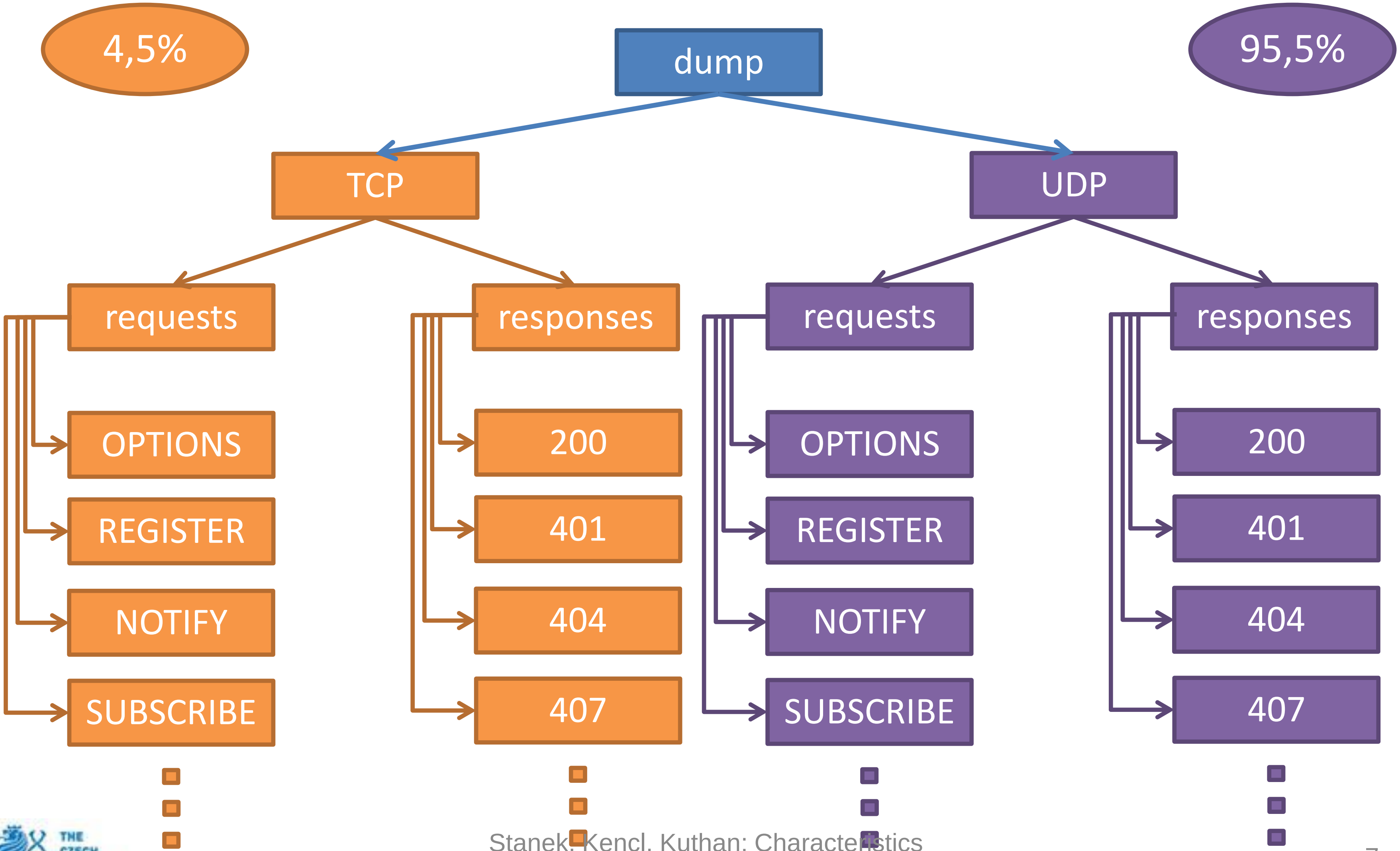
- Dataset
  - 67 hours of full SIP traffic capture
  - Over 40GB in total
  - ~3400 users
  - 280 distinct SIP clients



SIP traffic during a day



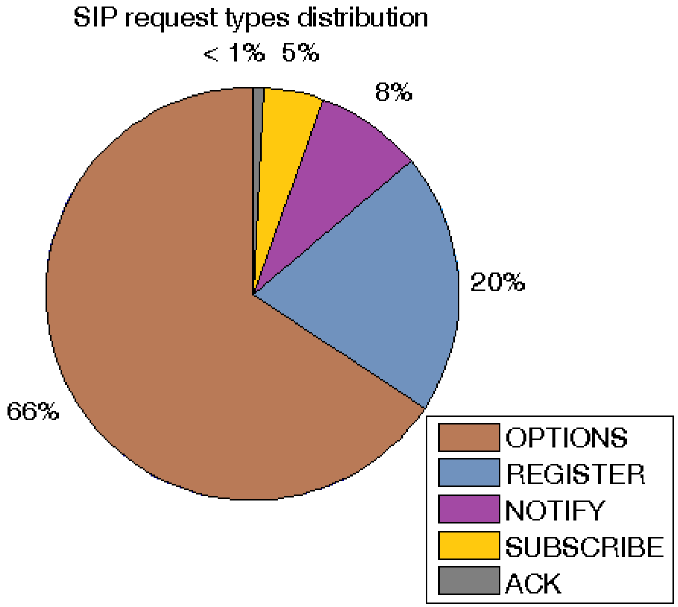
Structure after processing



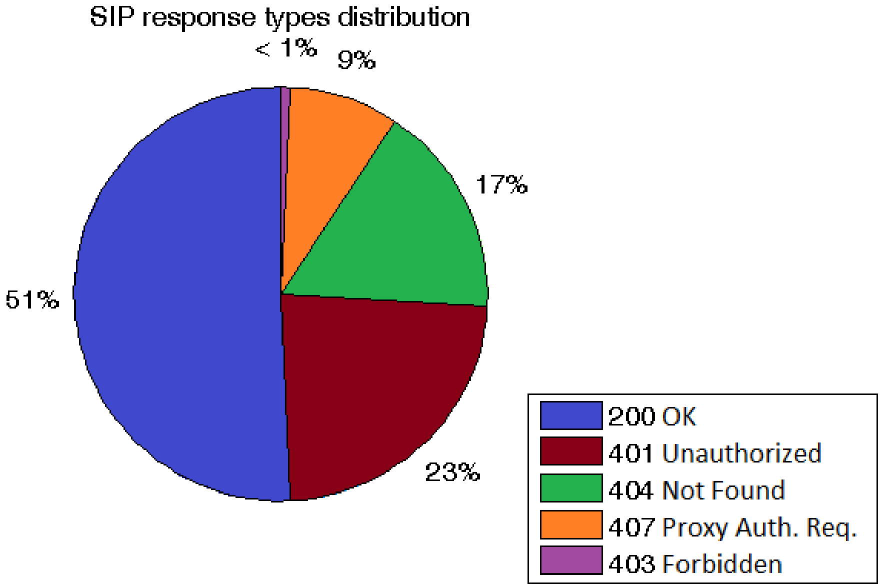
Requests and responses

Twice as many requests as responses!

REQUESTS

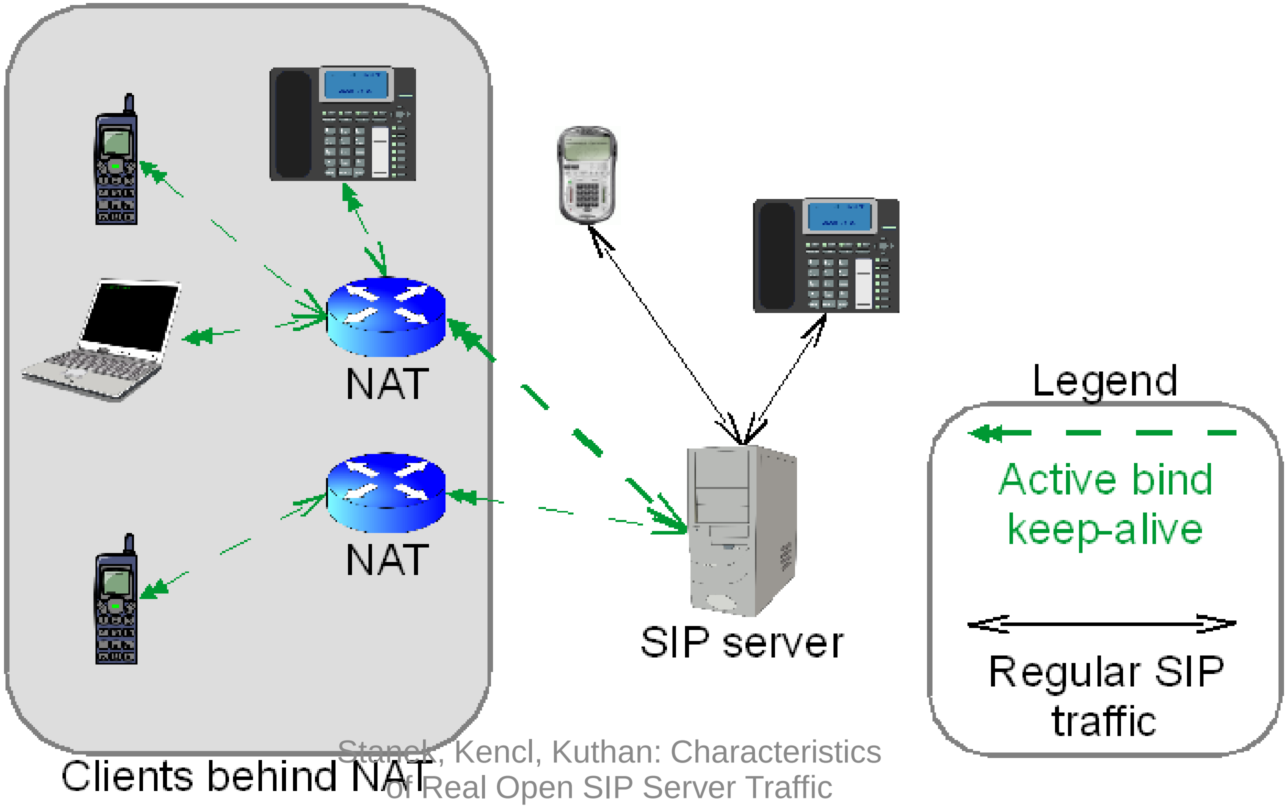


RESPONSES



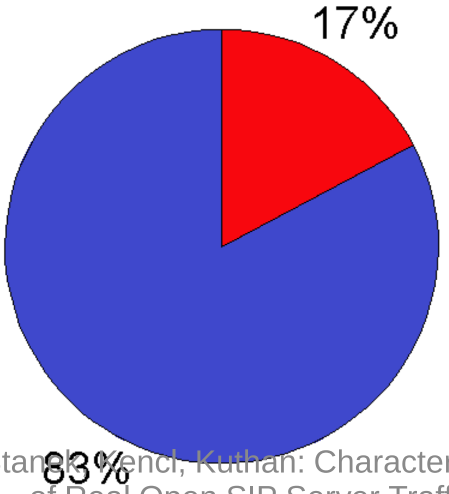
Why so many OPTIONS requests?

- **OPTIONS** form about 45% of the whole traffic
- Answer: proactive server NAT keepalive policy



Still it does not match

- 1 day stats:
  - – 1 500 clients assumed to be behind NAT
  - – 10 447 776 OPTIONS requests captured
  - – 4 x 60 x 24 x 1 500 = 8 640 000
  - – Excessive **1 807 776 OPTIONS requests**

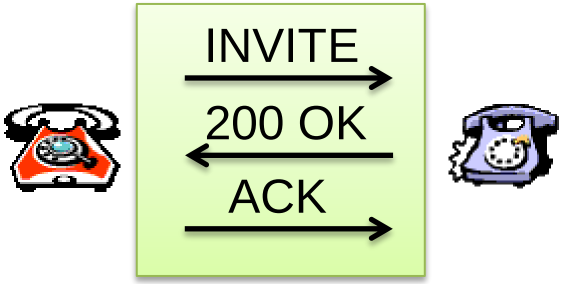


NAT keepalive overhead

- Not only from server but also from **clients**
- Not only OPTIONS but also REGISTER, dummy SUBSCRIBE etc.
- It forms **more than 50%** of the total SIP traffic

The ACK-INVITE anomaly

- Three-way handshake



INVITE

200 OK

ACK

- Obviously, there should be at most as many ACKs as INVITEs

	ACKs	INVITEs
Day 1	229 130	20 910
Day 2	90 282	18 900
Day 3	81 491	15 547

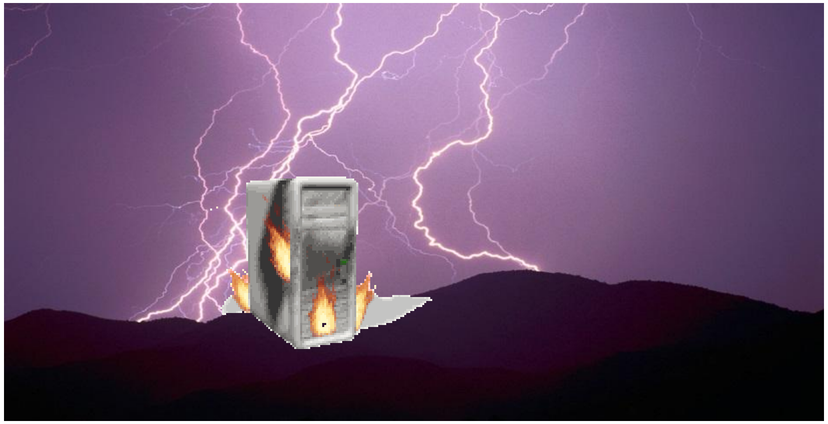
Registration storm

1) SIP server becomes inaccessible for a short time period

   - Clients find out that they cannot re-register, they keep trying

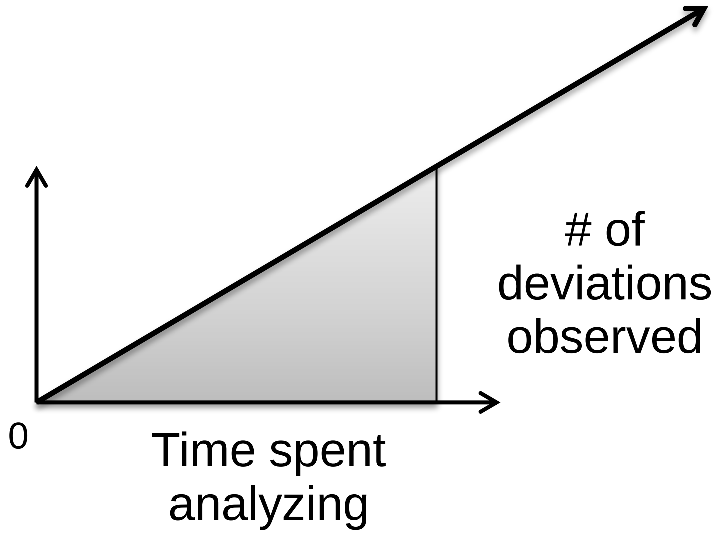
2) SIP server recovers

   - All clients try to register in a short time



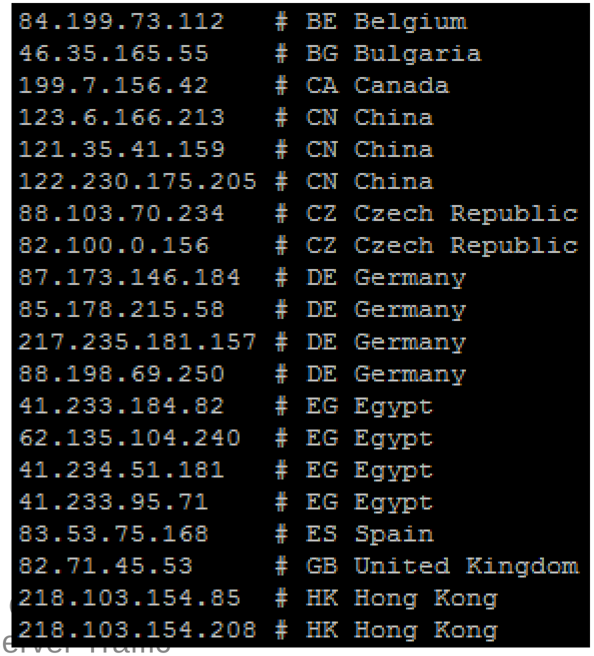
Other deviations observed

- Malformed messages
  - 'RE:50004GISTER'
  - 'RE:50037GISTER '
- Disappearing clients
- Zero-length calls
- …



Are callers and callees collocated?

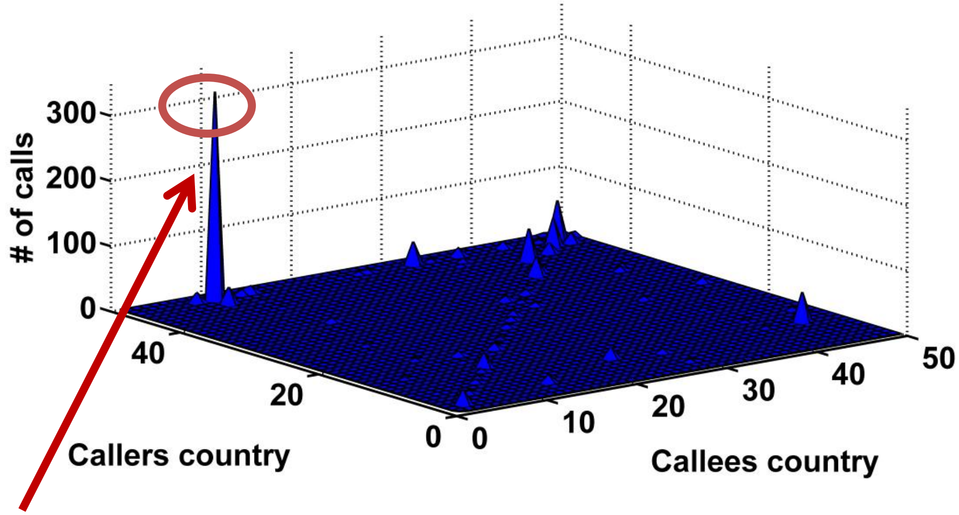
1. Extract individual calls
   - Filter out re-INVITES, unsuccessful set-ups
2. Find geographic location for caller and callee
   - Based upon their IP, used online IP-to-country mapper
3. Plot the results



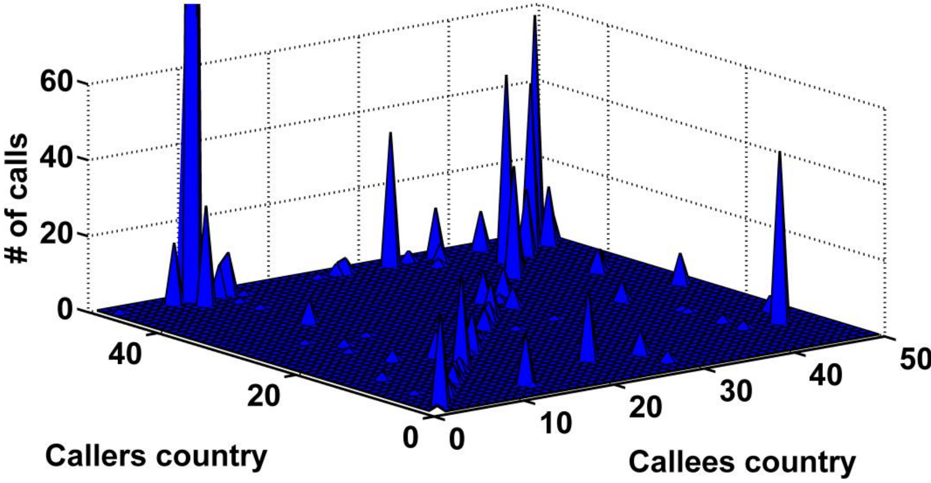
```
84.199.73.112   # BE Belgium
46.35.165.55    # BG Bulgaria
199.7.156.42    # CA Canada
123.6.166.213   # CN China
121.35.41.159   # CN China
122.230.175.205 # CN China
88.103.70.234   # CZ Czech Republic
82.100.0.156    # CZ Czech Republic
87.173.146.184  # DE Germany
85.178.215.58   # DE Germany
217.235.181.157 # DE Germany
88.198.69.250   # DE Germany
41.233.184.82   # EG Egypt
62.135.104.240  # EG Egypt
41.234.51.181   # EG Egypt
41.233.95.71    # EG Egypt
83.53.75.168    # ES Spain
82.71.45.53     # GB United Kingdom
218.103.154.85  # HK Hong Kong
218.103.154.208 # HK Hong Kong
```

Calls “crossing country borders”

10	China
12	Czech Republic
19	Hong Kong
46	Taiwan
48	USA

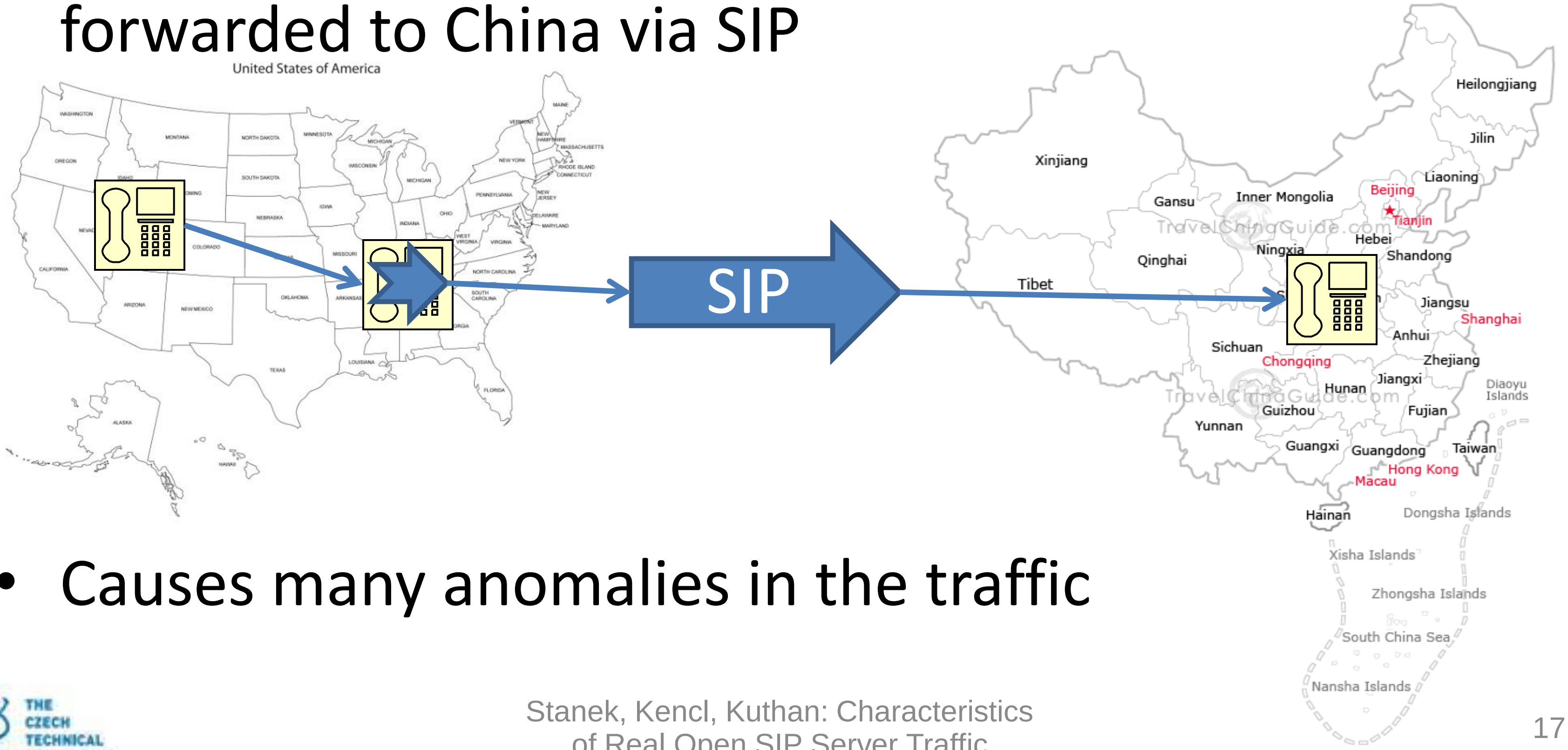


USA → China



The Virtual Phone Line service

- http://www.virtualphoneline.com/
- Call to an American number in USA (local call) is forwarded to China via SIP



- Causes many anomalies in the traffic

Conclusion

- Analyzed SIP traffic is multiplied by
  - Keeping NAT bindings alive
  - Open nature of SIP
  - Unexpected uses of the service
- Concrete traffic will differ, though there are likely to be unexpected anomalies
  - It is necessary to analyze and filter out „crap“ ☺

Conclusion remarks

- Analysis showed interesting findings
- One experimental server dump analysis is isufficient
- We need more dumps from various SIP servers

Publishing the dataset, obtaining more datasets

- Cannot publish without proper anonymization
  - Current anonymization approaches are not sufficient
- Fully automatic tool for traffic dump anonymization is a necessity
  - Must be able to handle large dumps
  - Must handle well partial/unfinished sessions
  - Must avoid destroying important relations
- We are working on it!

Looking for sponsors

Thank you for your attention!

Questions?

jan.stanek@fel.cvut.cz