Pierdomenico Fiadino

The Telecommunications Research Center Vienna (ftw.)

Vivisecting WhatsApp in Cellular Networks

Servers, Flows, and Quality of Experience

fiadino@ftw.at

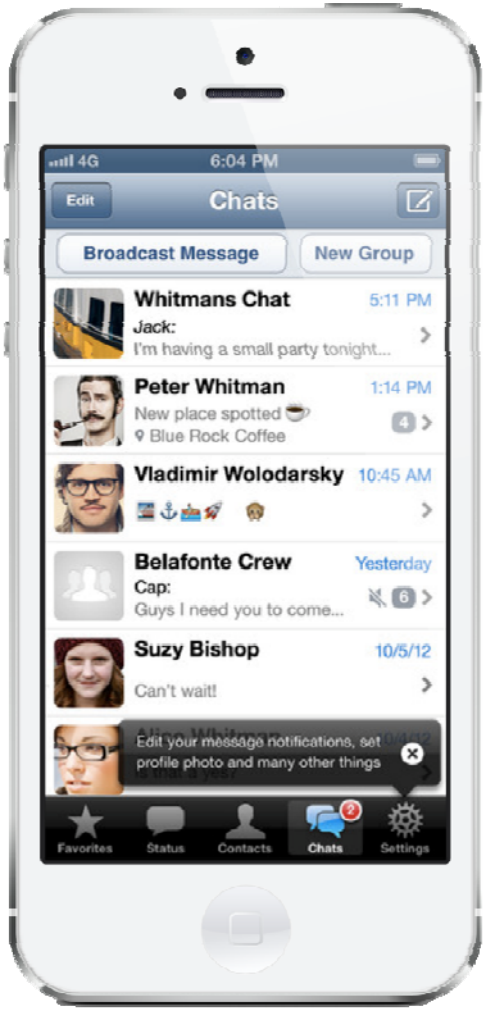
http://userver.ftw.at/~fiadino

WhatsApp
overview

ftw Creating Communication Technologies

Hard facts:

- 64 billion messages per day
  - 700 million photos
  - 100 million videos
- 500 million of daily active users
- Company with the quickest growing user base in history
- Acquired by Facebook for 19 billion $
  - Each user is worth 40$



Operators need to investigate it because:

- It is taking over (or already has...) the SMS/MMS market
- They need to learn how to track its usage
- They need to understand its impact on their networks

Outline

1. Characterization of WhatsApp through hybrid measurements
   - Traffic captures in lab with standard tools
   - Large-scale passive traces (mobile national network)
   - Large-scale active measures (RIPE Atlas)

2. Quality of Experience remarks
   - Build a QoE model with user tests
   - Apply model on passive traces

3. WhatsApp global outage overview
   - Characterization
   - Social media impact

Outline

1. Characterization of WhatsApp through hybrid measurements
   - Traffic captures in lab with standard tools
   - Large-scale passive traces (mobile national network)
   - Large-scale active measures (RIPE Atlas)

2. Quality of Experience remarks
   - Build a QoE model with user tests
   - Apply model on passive traces

3. WhatsApp global outage overview
   - Characterization
   - Social media impact

Reverse engineering Whatsapp naming scheme...

Hybrid measurements

Testbed:

- Traffic (chat and media exchange) actively generated at end devices (Android and iOS)
- Passively captured at a gateway (**Wireshark**)
- Focus on DNS requests

Findings:

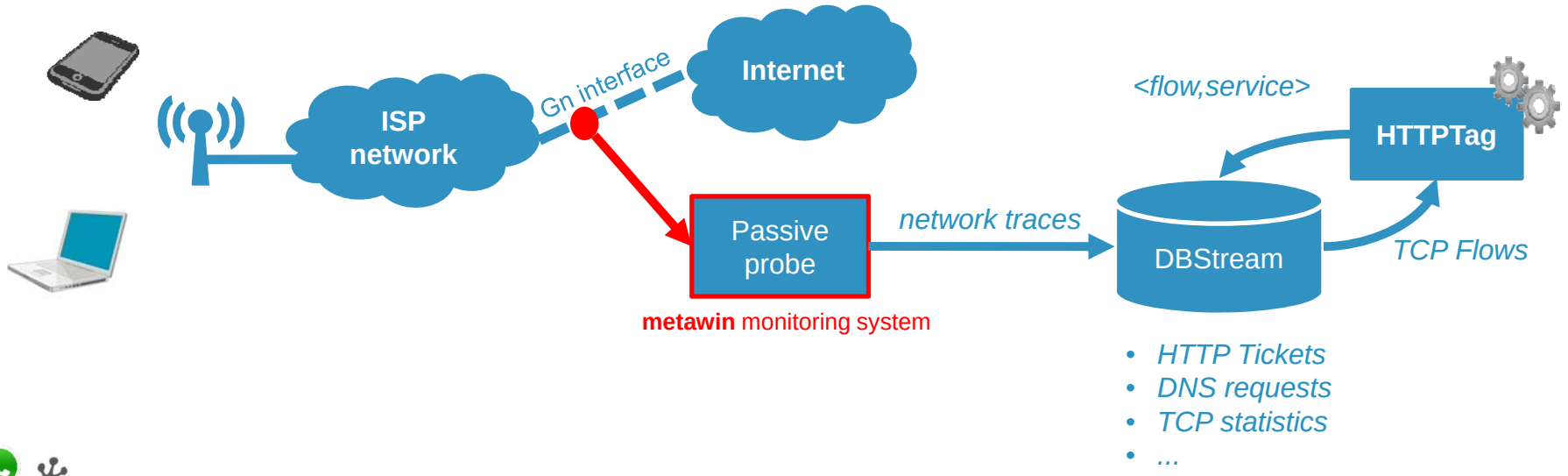
- Whatsapp used custom XMPP protocol
- Media exchange via HTTPS servers
- One persistent SSL connection to XMPP servers while the app is running
- Dedicated TLS connections to HTTPS servers for each media transfer

Servers naming scheme `[*.whatsapp.net]`:

3rd lvl domain	prot. (port)	type
`cX, eX, dX`	XMPP(5222,443)	chat & control
`mmiXYZ,mmsXYZ`	HTTPS (443)	media (photo,audio)
`mmvXYZ`	HTTPS (443)	media (video)

Passive probes
network architecture

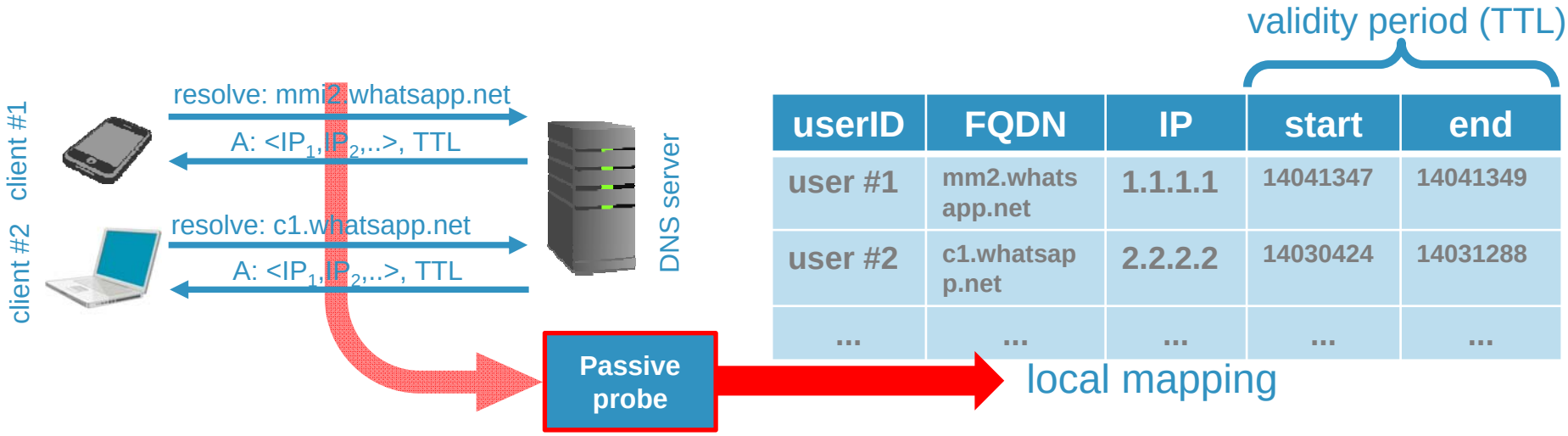
- Capture network traces at the core of an operational 3G/4G network (~5 millions of users)
- Dataset: 1 month long (February 2014)
- Classify flows: only consider Whatsapp-related traffic (details on classification techniques -> next slide)



Classification of TCP flows

Collect DNS requests from passive traces

- Idea: use **passively collected DNS** requests to dynamically map *<services,serverIPs>*

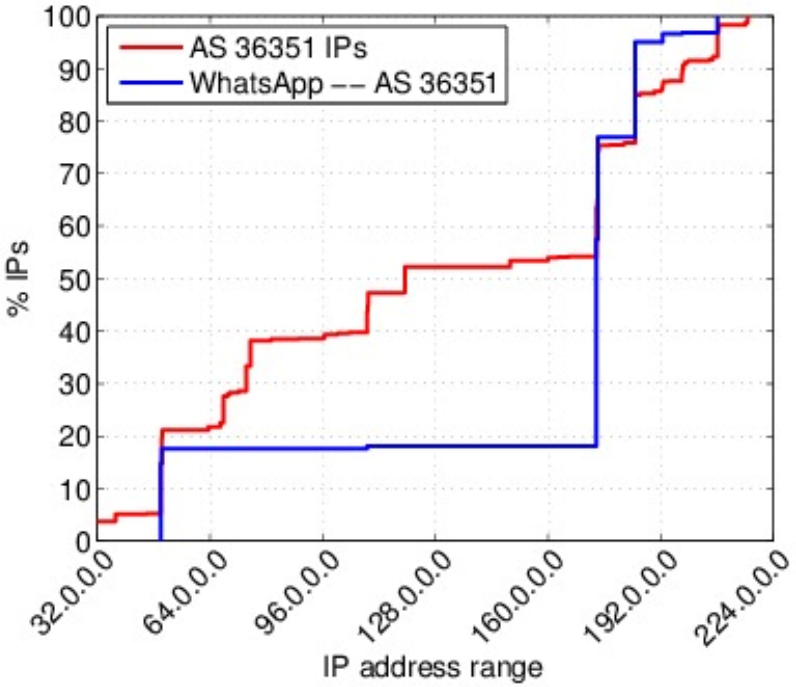


userID	FQDN	IP	start	end
user #1	mm2.whatsapp.net	1.1.1.1	14041347	14041349
user #2	c1.whatsapp.net	2.2.2.2	14030424	14031288
...	...	...	...	...

- Every subsequent flow between a **<user>** and a **<server_ip>** in the validity period **[validity_start:validity_end]** are assigned to **<FQDN>**
- The Fully Qualified Domain Names (FQDN) are assigned to service with usual pattern matching

Revealing Hosting Infrastructure

Through large-scale passive measurements



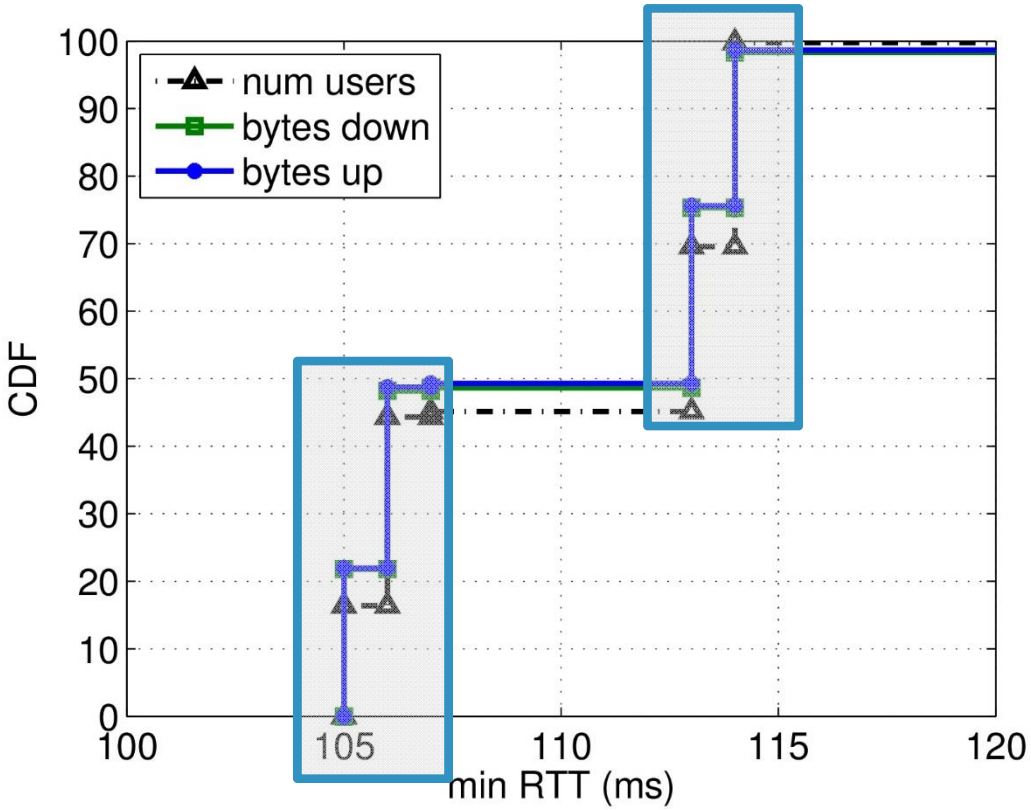
- 386 IP adresses used by Whatsapp (chat and media)
- All in AS36351 (Softlayer)

Service/AS	#IPs	# /24	# /16	# /8
WhatsApp	386	51	30	24
SoftLayer (AS36351)	1364480	5330	106	42

Revealing Hosting Infrastructure

Through large-scale passive measurements

Localization of servers through RTT measurements



~400 IP addresses in Softlayer AS

Two big steps in RTT distribution at:

- 106ms
- 114ms

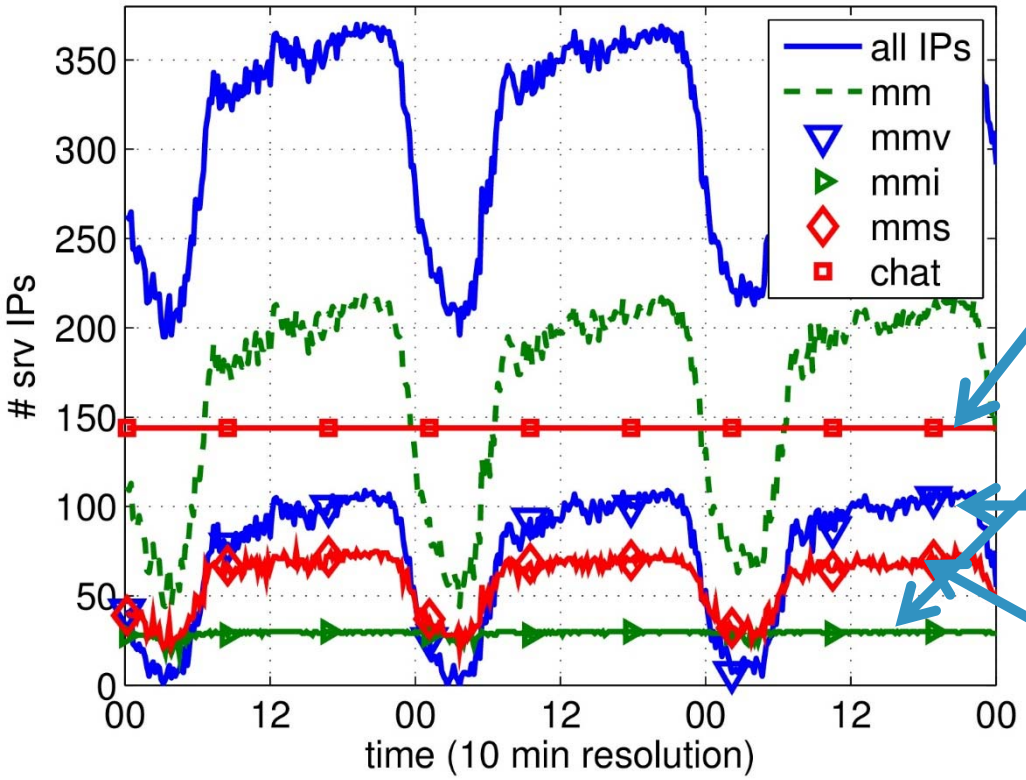
Localized by MAXMIND in:

- **Houston**
- **Dallas**

Revealing Hosting Infrastructure

Through large-scale passive measurements

Active IPs

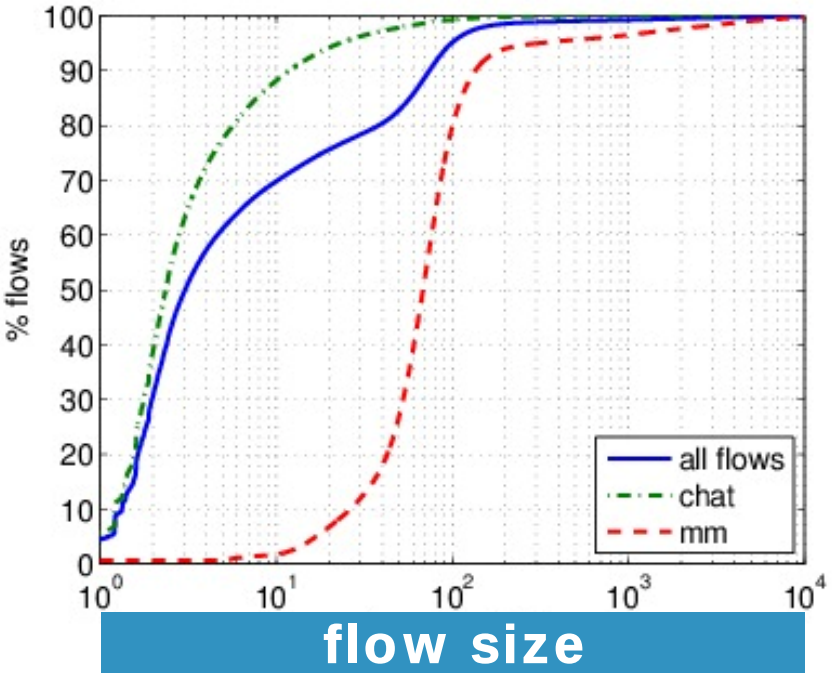


- Chat servers
  - 150 IPs "*always on*"
- Images exch. servers
  - 25 IPs "*always on*"
- Video exchange servers
  - Up to 100 IPs at peak hours
- Audio exchange servers
  - Up to 75 IPs at peak hours

At least 200 IP addresses active at any time

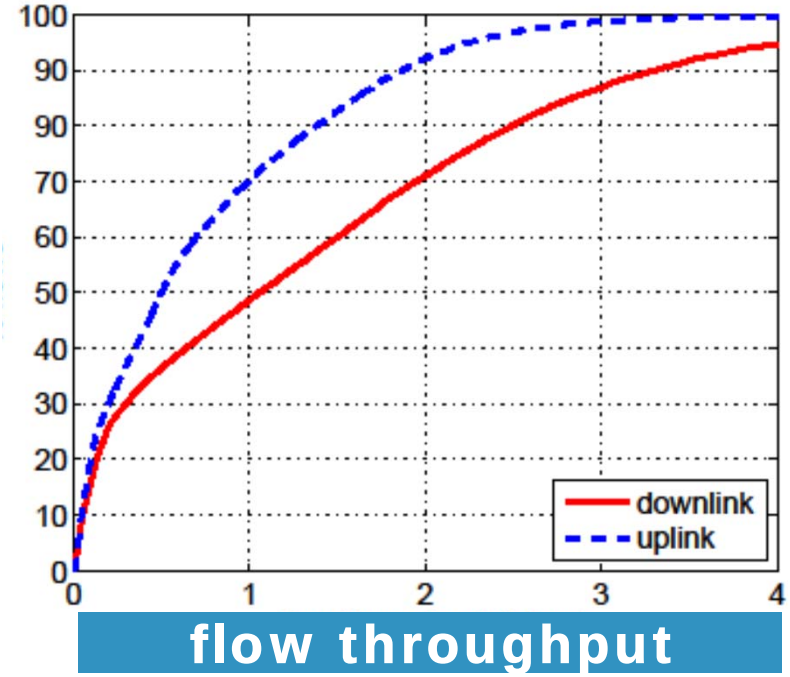
Whatsapp traffic characteristics

flow size and throughput



flow size

- Smaller chat/control flows and heavier mm flows
- 90% of chat flows < 10KB
- 50% of mm flows > 70KB

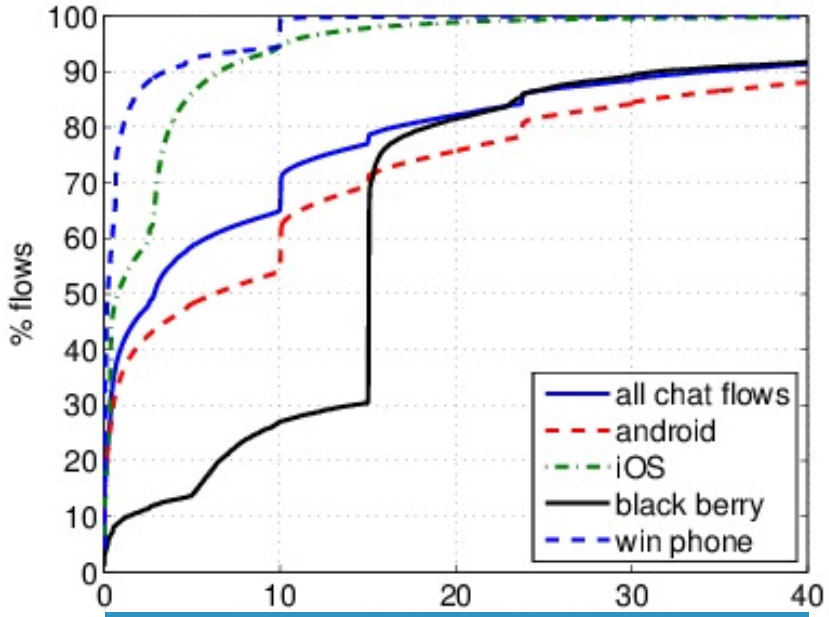


flow throughput

- Only bigger flows (>1MB) considered
- Up to 1.5Mbps in downlink
- Up to 800Kbps in uplink

Whatsapp traffic characteristics

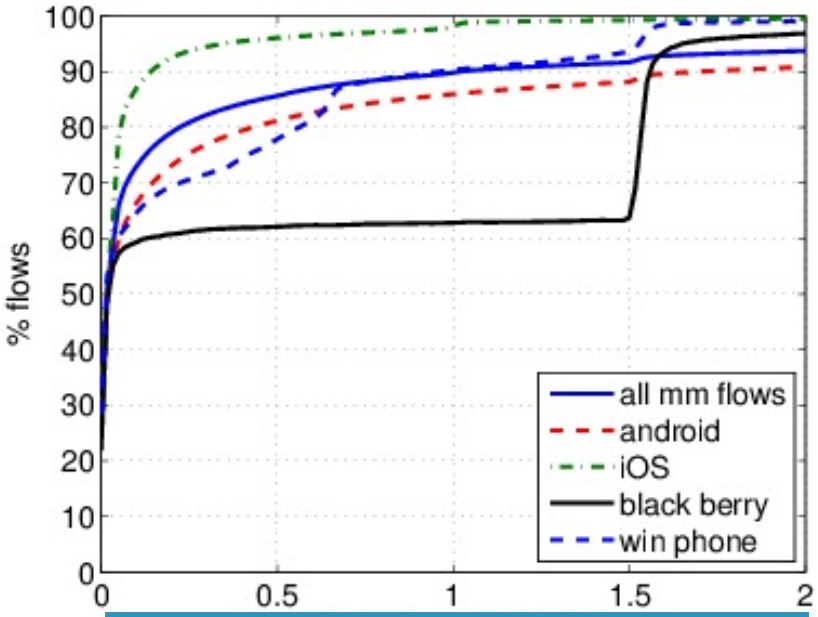
flow duration with OS breakdown



flow duration (chat flows) [m]

Timeouts:

- Android: 10/15/25 min
- iOS: 3 min
- Blackberry: 15 min
- Windows Phone: 10 min



flow duration (mm flows) [m]

Timeouts:

- Blackberry: 90 sec

Flow characteristics depend on terminal operating system

RIPE Atlas infrastructure

for geo-distributed active measurements

- **RIPE NCC**: Regional Internet Registry for Europe (equivalent of ARIN for North America)
- **RIPE Atlas**: a large measurement network composed of geographically distributed active probe used to measure connectability and reachabiltiy

RIPE Atlas probe v3

TP-Link MR3020 router with custom firmware

Hosting infrastructure „stability“

Geographical distributed active measurements

http://atlas.ripe.net

RIPE Atlas probe v3
TP-Link MR3020 router with custom firmware

- My UDM (User Defined Measurement): **600 probes world-wide resolve Whatsapp hostnames** (`{mmX|dX}.whatsapp.net`)
- Result: same set of IP addresses

Previous conclusions for WhatsApp hosting infrastructure are still valid from other VPs

Outline

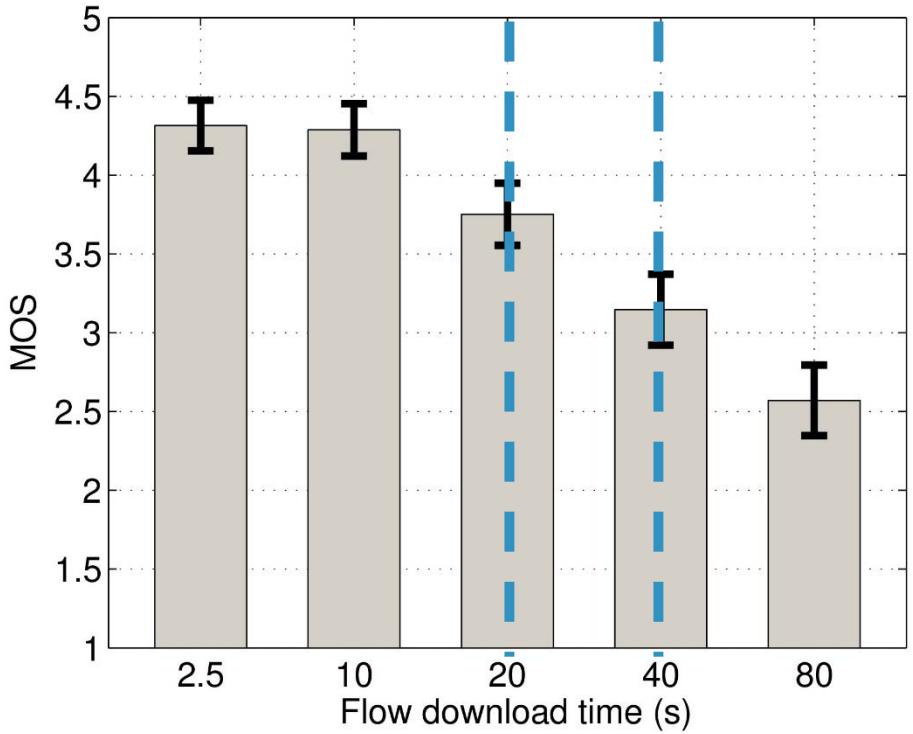
1. Characterization of WhatsApp through hybrid measurements
   - Traffic captures in lab with standard tools
   - Large-scale passive traces (mobile national network)
   - Large-scale active measures (RIPE Atlas)

2. Quality of Experience remarks
   - Build a QoE model with user tests
   - Apply model on passive traces

3. WhatsApp global outage overview
   - Characterization
   - Social media impact

Quality of Experience (QoE)

Measuring Mean Opinion Score (MOS) through user tests



Measurement campaign:

- 50 participants
- 45/50 male/female
- 60/40 students/employees
- 23 average age

Experiment:

- Media file transfer (5MB)
- Emulation of different network conditions (change transfer time)

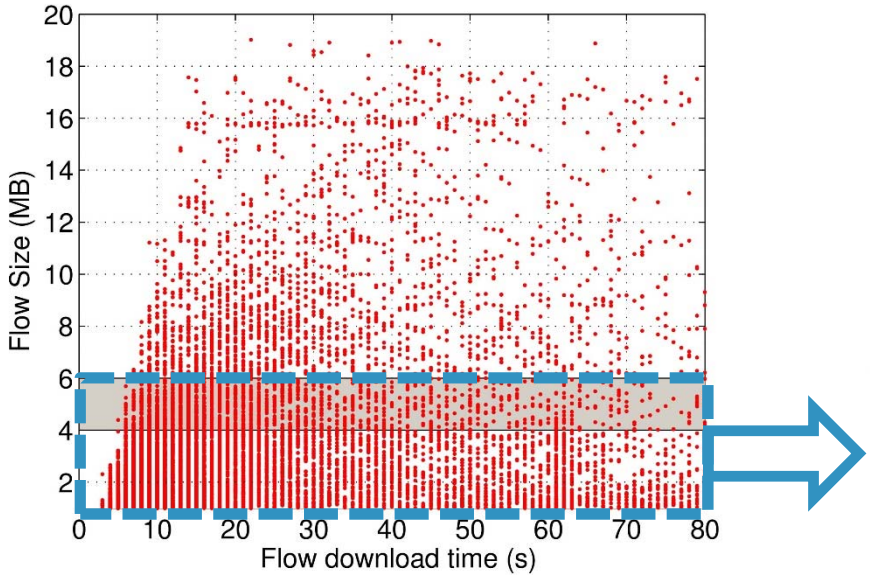
Mean Opinion Score (MOS):

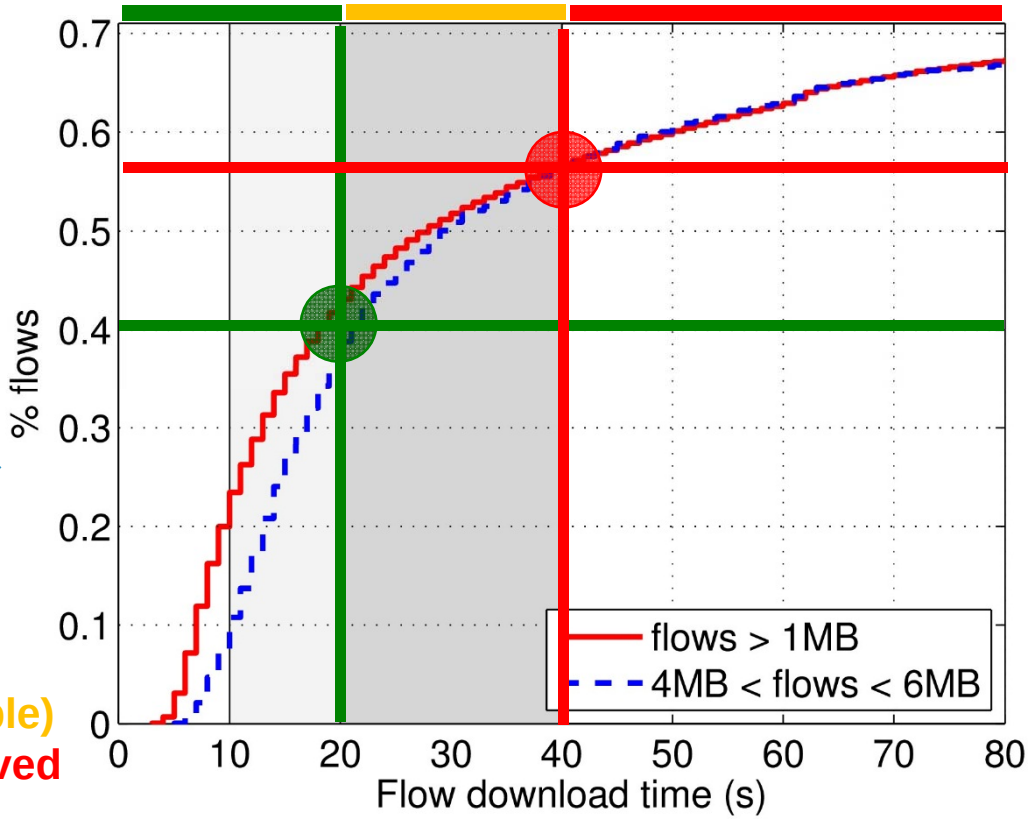
- From 1 (very bad) to 5 (excellent)

- Users tolerate transfers of up to 20 seconds with good experience
- Threshold of 40 seconds discriminates good and bad experience
- The score is undependent of the flow size

Quality of Experience (QoE)
in the large scale measurements

- User tests: 5MB transfers, <20s good, >40s bad (previous slide)
- Assumption: flow size is not taken into account
- Apply same model on large scale network traces





- **40% flows last less than 20s (good)**
- **60% flows last less than 40s (acceptable)**
- **35% flows are badly/very badly perceived**

Outline

1. Characterization of WhatsApp through hybrid measurements
   - Traffic captures in lab with standard tools
   - Large-scale passive traces (mobile national network)
   - Large-scale active measures (RIPE Atlas)

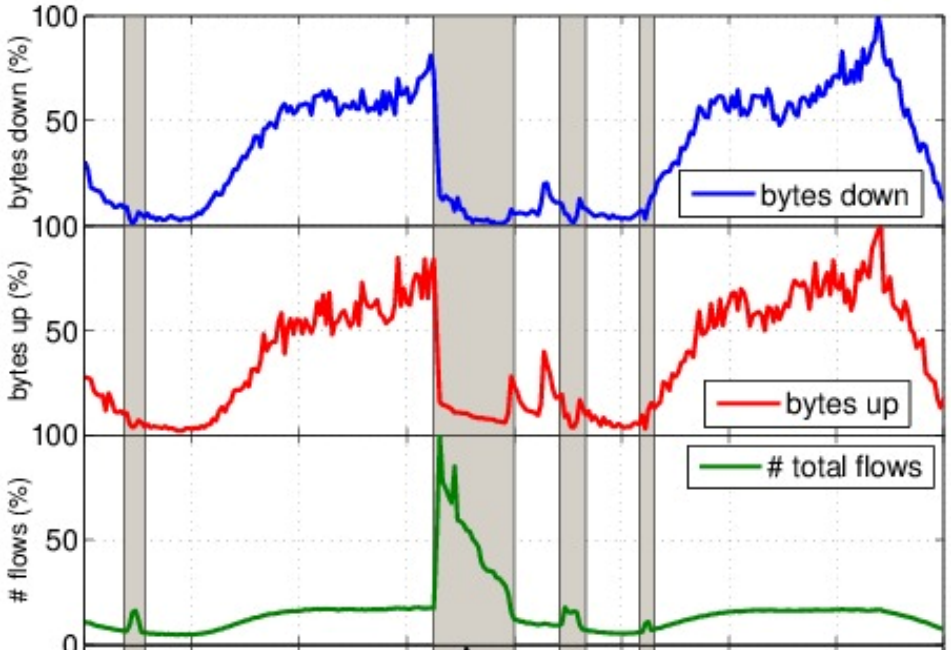
2. Quality of Experience remarks
   - Build a QoE model with user tests
   - Apply model on passive traces

3. WhatsApp global outage overview
   - Characterization
   - Social media impact

The big 5-hours outage (Feb. 22nd, 2014)
press reaction

The big outage (Feb. 22nd, 2014)

as seen from passive measurements and social feeds



drop in volume down
(with small residual traffic)

drop in volume up
(with small residual traffic)

ramp-up on flow counts
(SYN/RST exchange)

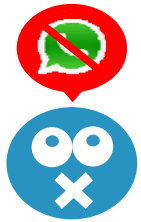
1. Residual traffic during outage
2. Number of flow increase
3. Clear social media reactions

Conclusions

1. Characterization of WhatsApp
   - **Hybrid methodology for service characterization**
   - **Revealed *.whatsapp.net server naming scheme**
   - **Centralized infrastructure by Softlayer in Texas**
   - **75% of volume for media transfers (36% is video)**

2. Quality of Experience remarks
   - **Downlink throughput up to 1.5Mbps, but...**
   - **40% of flows could suffer of bad user experience**

3. WhatsApp global outage overview
   - **Servers still reachable, fail at app layer**
   - **Strong social media impact (correlation)**



http://www.ftw.at

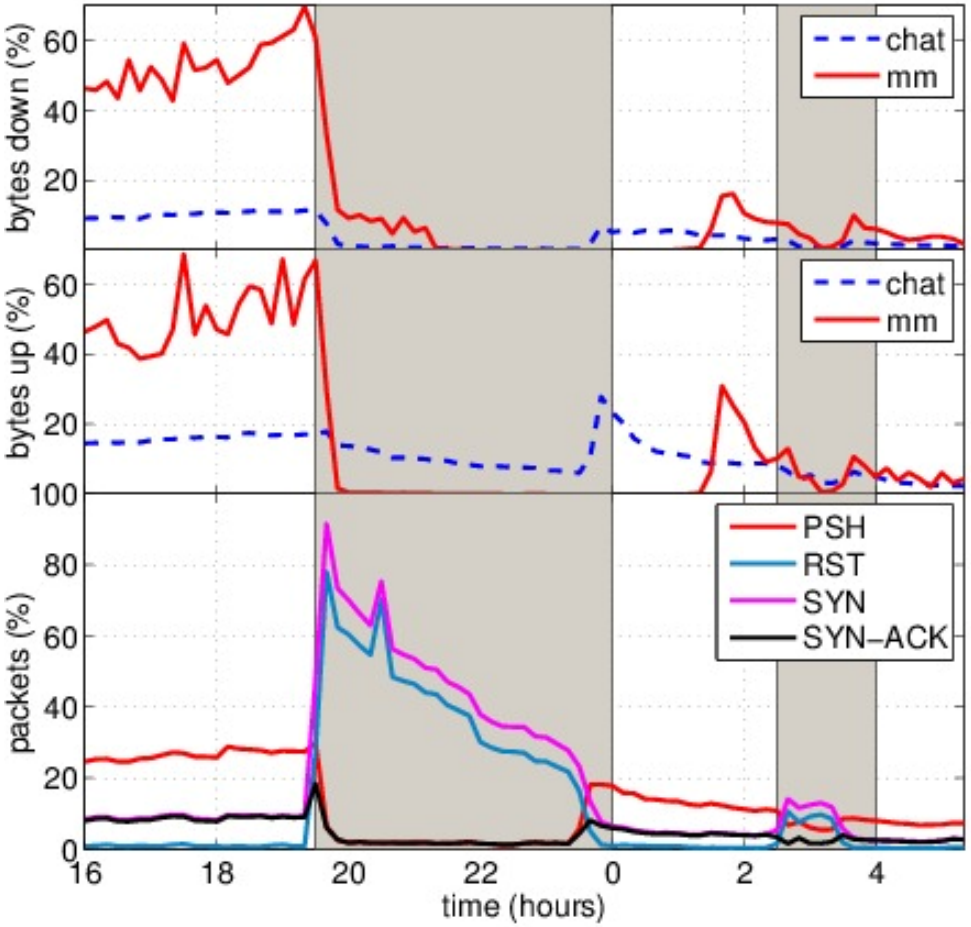
http://www.ict-mplane.eu

thank you for you attention!

Pierdomenico Fiadino
Telecommunications Research Center of Vienna (FTW)
<fiadino@ftw.at>

backup slides =>

The big outage (Feb. 22nd, 2014)
as seen from passive measurements and social feeds



residual volume down (mm)

residual volume up (chat)

TCP flags counters