HELSINKI UNIVERSITY OF TECHNOLOGY
Networking Laboratory

Mobile Internet Usage Measurements

Results

Antero Kivi
24.4.2006

Agenda

- Operator Reporting System –Based Measurements
  - Measurement description
  - Mobile terminal base
  - Mobile subscriber population
  - Summary
- TCP/IP Header Collection –Based Measurements
  - Measurement description
  - General packet data traffic patterns
  - Packet data traffic by application protocol
  - Web protocol traffic patterns
  - Summary
- Handset –Based Measurements
  - Measurement description
  - Description of the panel
  - General handset usage patterns
  - General packet data usage patterns
  - Packet data generating application usage
  - Smartphone browsing patterns
  - Summary
- Conclusions
  - Major findings
  - Observations on measurement process
  - Further information

HELSINKI UNIVERSITY OF TECHNOLOGY
Networking Laboratory

Operator Reporting System –Based Measurements

Mobile Internet Usage Measurement

Contents

- Measurement description
- Mobile terminal base
  - Description of source data
  - Terminal distribution by model
  - Terminal distribution by feature
  - Terminal distribution by manufacturer and smartphone type
  - Terminal distribution by year of introduction
- Mobile subscriber population
  - Mobile subscribers by type of subscription
  - Consumer subscribers by GPRS tariff alternative
  - Consumer subscriber GPRS usage and roaming activity
  - Mobile subscriber GPRS usage by terminal radio technology
- Summary

Measurement Description

- Data collected using mobile operators' charging-oriented reporting systems
  - Ticket (CDR) and subscriber information systems of three major Finnish mobile operators' (Sonera, Elisa, DNA)
- Data of one week and/or one month on fall 2005
  - Measured data mainly from weeks 34 and 38, and September 2005
  - In some cases there is variation as data of different operators was not always from the same period
- About 80-90% of Finnish mobile subscribers included
  - Most data from all three operators → 80-90% of Finnish mobile subscribers
  - In some (rare) cases data was available only from two operators

Operator Reporting System –Based Measurements

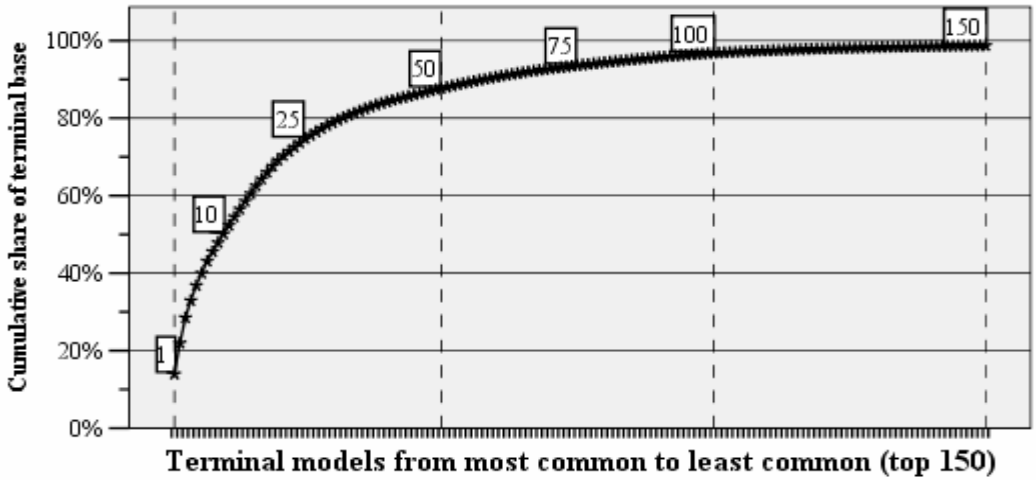
Mobile Terminal Base

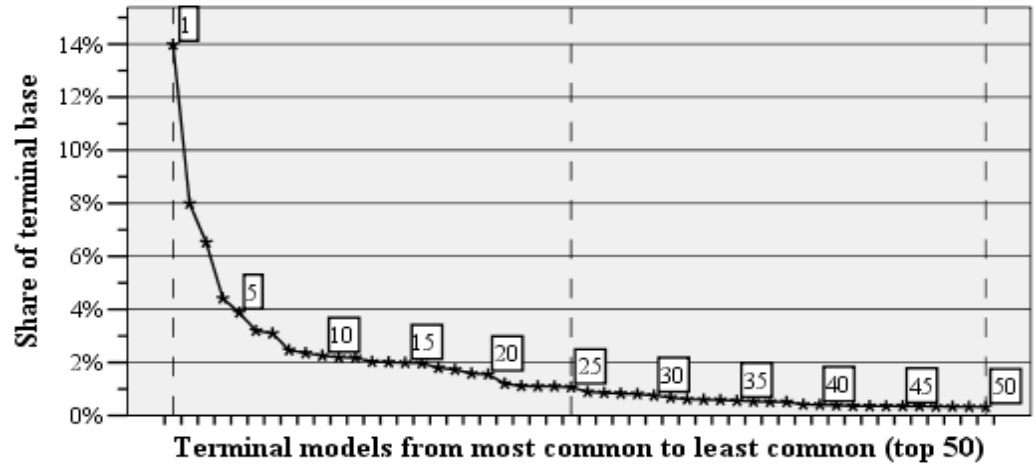
Description of source data

- MSO's included: Sonera, Elisa (+Kolumbus), DNA
  - No data on: Saunalahti, TeleFinland, others
  - Over 4 million terminals in the sample → 80 - 90% of all Finnish mobile terminals
- A data set describing mobile terminal installed base from each operator
  - All MSO's subscribers' terminals with made voice calls / sent SMSs on September
  - All MSO's subscribers' terminals with at least one transaction (phone call, SMS, other) on September
  - All terminals observed at the network on week 34, no particular transactions nor subscription required as merely turning on the phone creates a ticket
- Error due to churn and differences in data sets
  - <3% churn during longer measurements, and between week 34 and end of September
    - Numpac: 115 000 mobile numbers ported on September 2005 → <3% of measured mobile terminal base
    - No major error, as terminal profile of number porting subscribers supposedly does not differ much from general terminal profile
  - Max. 2,5% excess in measured terminal base due to foreign roamers and emergence call readiness
    - No major error from emergency call readiness, as it is not (primarily) related to specific terminal models
    - Some error due to foreign roamers, whose terminals indeed do not belong to Finnish mobile terminal base
- Error due to unidentified terminals and terminal features
  - 4,4% of terminals were unidentified → somewhat more advanced terminals than identified terminal base in general
    - TAC codes not identified (2 data sets): Manufacturers don't deliver TAC code – terminal model mappings to TAC allocating organizations in real time → model information is not available for most recent terminals
    - Terminal models outside top 100 models (1 data set): most recent features not evenly represented among top 100
  - 1 - 1,5% of identified terminals without some specific information on terminal features

Mobile Terminal Base

Terminal distribution by model



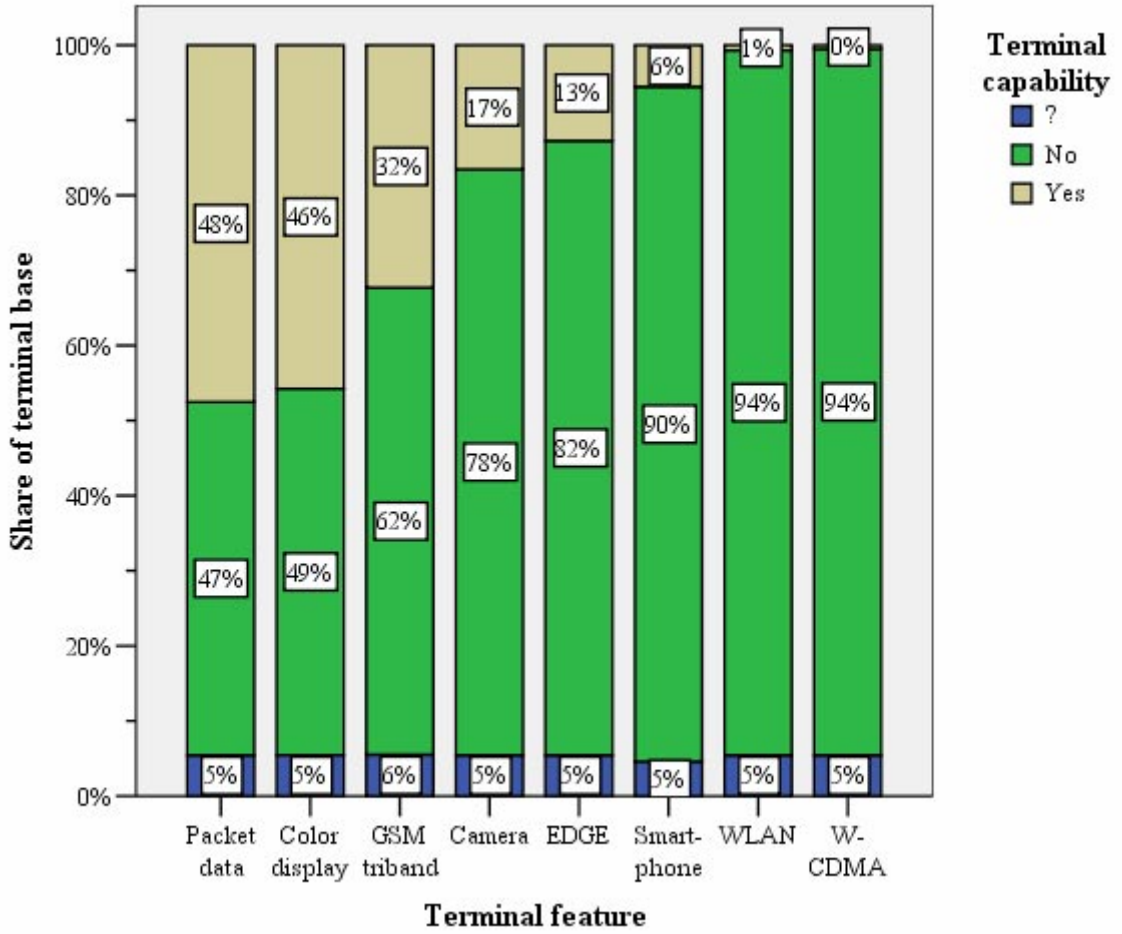


- Terminal base is fairly concentrated
  - 88% of all terminals in top 50
  - 97% of all terminals in top 100
  - 99% of all terminals in top 150
- Roughly 1000 different terminal models identified in total

- Most popular terminal is Nokia 3310 with a 14% share of all terminals
- First camera phone 11th
- First smartphone 21st
- First WLAN terminal 37th
- First WCDMA terminal 54th

Mobile Terminal Base

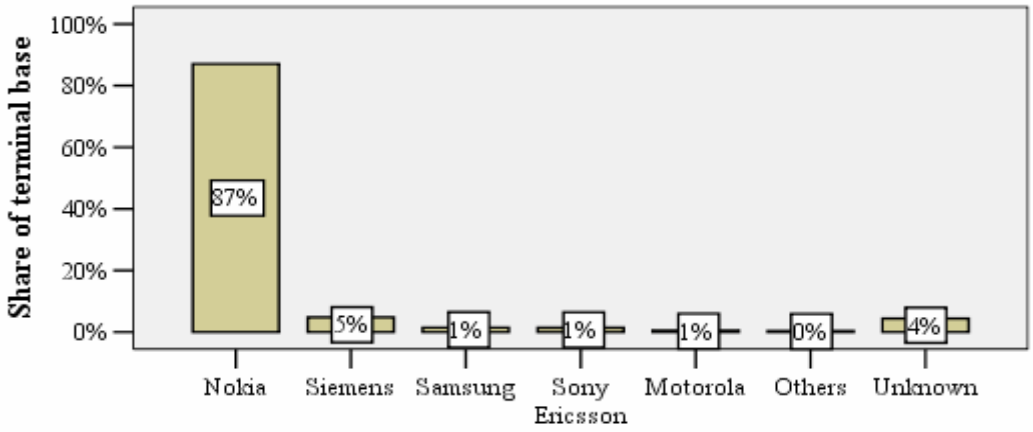
Terminal distribution by feature

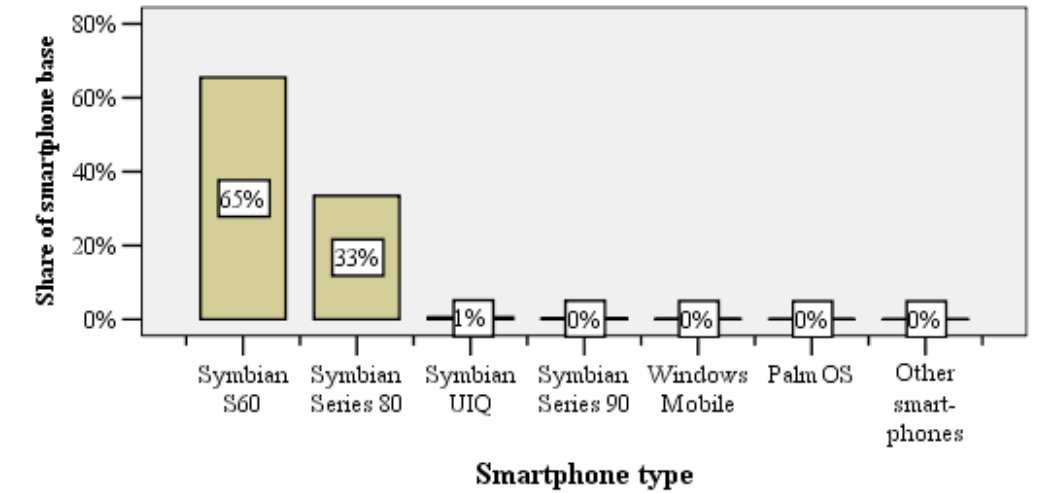


- Background / method
  - Data on features of specific terminals collected from manufacturers' web sites
- Key features for packet data usage are not widely spread
  - Packet data 48%
  - EDGE 15%
  - Smartphones 6%
  - WCDMA 0,5%
- Unknown 5-6% somewhat increases the figures of all features
  - Unknown probably has similar or somewhat more advanced profile than identified terminal base
- Other remarks
  - 0,6% of mobile terminals are not mobile phones
  - Single band 4%, Dualband 58% Triband 32%
  - WLAN 0,7%

Mobile Terminal Base

Terminal distribution by manufacturer and smartphone type

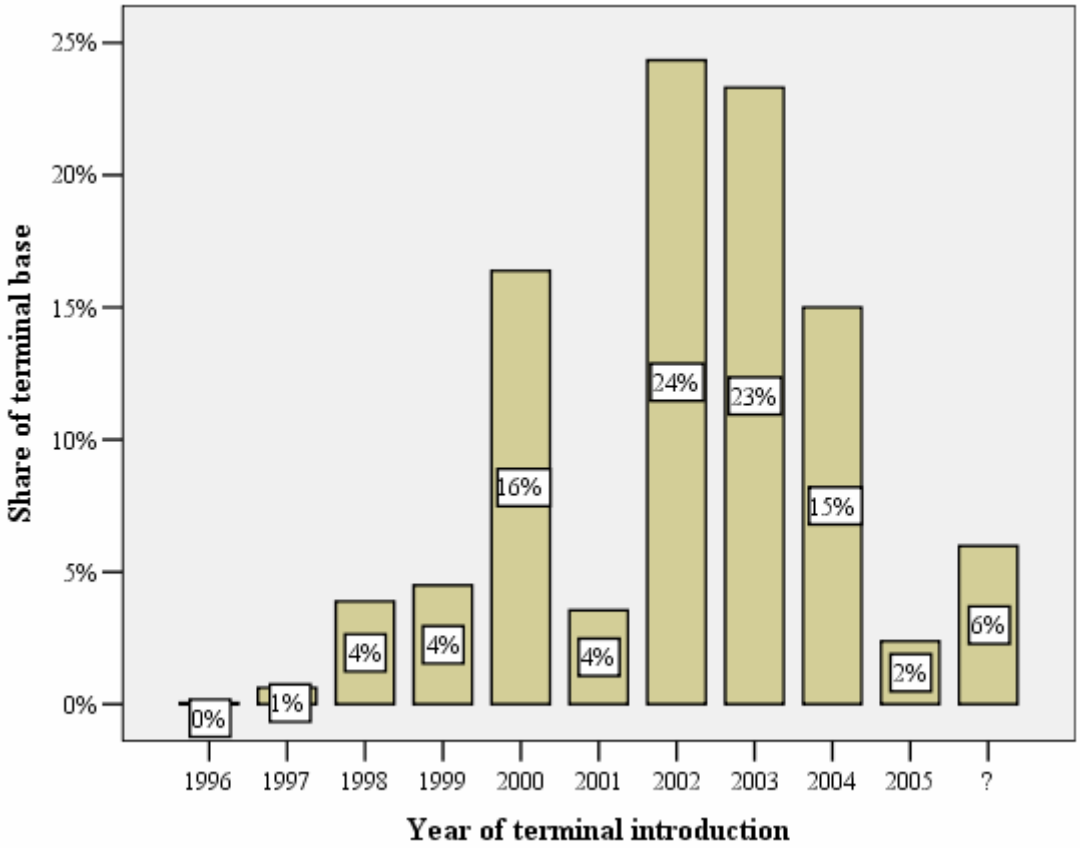




- Nokia 87% market share remarkable
  - First non-Nokia terminal is ranked 30th
  - Siemens possesses the clear 2nd place
- >99% of smartphones Nokia Symbian
  - Nokia communicators (Series 80) have a notable 33% share
  - Other smartphone types seem to be marginal
- Smartphone shares not entirely reliable
  - Smartphones among unknown terminals (4% of total) could significantly change the balance between smartphone types
  - Symbian terminals will nevertheless represent the clear majority of smartphones

Mobile Terminal Base

Terminal distribution by year of introduction



- Finnish mobile terminals are old
  - Average year of introduction is 2002
- Nice bell curve conforming to theory (product life cycle), apart from year 2001
- Something happened in year 2001
  - Burst of economic bubble?
  - GPRS / WAP introductions close to 2001?
- Reliability issues
  - Data mainly from www.mobile.softpedia
    - How reliable is the source?
    - Softpedia data on other terminal features conforms to manufacturer-originated data
  - "Year of introduction" not well defined
    - Official or "accidental" introduction?
    - Delay from terminal introduction to start of sales has increased lately, and depends on e.g. manufacturer and market

Mobile Subscriber Population

Mobile subscribers by type of subscription

Type of subscription	Share of all subscriptions	Type of subscription	Share of postpaid subscriptions	Share of postpaid subscribers' packet data traffic
Postpaid	92 - 94%	Consumer	75%	38%
		Business	25%	62%
Prepaid	6 - 8%	N/A	-	-

- Share of postpaid subscribers is very high
- Consumers represent the majority of postpaid subscribers
- Packet data usage volume is significantly higher among business subscribers

Operator Reporting System –Based Measurements

Mobile Subscriber Population

Consumer subscribers by GPRS tariff alternative

- Consumer subscribers divided into three groups based on different GPRS tariff alternatives
  - No fixed fee (subscriber with operators' default GPRS tariffs)
    - Purely usage-based tariffs or GPRS not activated at the time of study
  - Small fixed fee (subscriber paying some small fixed fee)
    - Usage-based tariff with a fixed fee, subscribers with a chargeable GPRS activation on study period, block-based tariff with block size 2-50 MB
  - Large fixed fee (subscribers paying a clearly higher fixed fee)
    - Block-based tariff with block size 100-500 MB, partly or fully flat-rate tariff

GPRS tariff group	Share of subscribers	Share of packet data traffic	Average chargeable packet data volume per subscriber
No fixed fee	99,0%	18%	0,01 MB / week
Small fixed fee	0,6%	44%	3,93 MB / week
Large fixed fee	0,4%	38%	7,33 MB / week

- 99% of subscribers have the default usage-based tariff alternative
- Subscribers with fixed fee alternatives create the majority (82%) of packet data traffic
  - Volume of packet data usage increases as it gets relatively cheaper

Operator Reporting System –Based Measurements

Mobile Subscriber Population

Consumer subscriber GPRS usage and roaming activity

GPRS usage by consumer subscribers during study week	
Share of consumer subscribers using GPRS	8 – 9%
Average packet data volume per GPRS using consumer subscriber	0,8 MB / week

Roaming by consumer subscribers during study week	
Share of consumer subscribers using voice (and SMS) roaming	3 – 4%
Share of consumer subscribers using GPRS roaming	0,1%
Ratio of GPRS roamers to voice roamers	2 – 6%
Average packet data volume per GPRS roaming using consumer subscriber	0,4 – 0,8 MB / week

- Almost 10% of consumer subscribers used GPRS during study week
  - Only 1% had non-usage based GPRS tariff
- While roaming, GPRS is less (2 - 6%) used by consumers
  - Roaming GPRS usage volumes seem somewhat smaller than GPRS usage at home network
- Multiple reliability issues with roaming figures (very heterogeneous source data)
  - Voice call roaming included vs. voice call and SMS roaming included
    - Partial data showed no major differences between the two figures, although other research suggests differently
  - Originated voice calls / SMSs included vs. both originated and received included
  - One week study period vs. one month study period
  - Subscribers actually roaming vs. subscribers billed for roaming during study period
    - Billed roaming not the same as actual roaming, but still gives a usable average for roaming
    - (Different) delays related to roaming data delivery from foreign operators
    - Billing delay, some data available only when subscriber is billed (monthly or after a euro threshold is reached)
  - Possible variation in roaming activity between random weeks

Operator Reporting System –Based Measurements

Mobile Subscriber Population

Mobile subscriber GPRS usage by terminal radio technology

Terminal radio technology	Share of GPRS capable terminals	Share of terminals actually used for GPRS	Share of postpaid subscriber packet data traffic*
GSM/GPRS capable terminals	100%	100%	100%
EDGE capable terminals	27%	>27%	35%
WCDMA capable terminals	2%	>2%	16%

* Distribution of traffic by terminals of different radio capabilities, not the bearers actually used

- Users of more capable terminals use packet data more actively than those using less capable terminals
  - Terminals with higher radio capability represent a relatively higher share of terminals actually used for packet data than implied by their share of terminal installed base (exact figures for column 3 are not shown due to sensitivity reasons)
  - This is especially clear for WCDMA users → Is WCDMA capability currently acquired specifically for data usage, EDGE capable terminals possibly (also) for other reasons?
- WCDMA capable terminals' share of packet data traffic is particularly high
  - Data card usage increases this share

Summary

- Data from operators' CDR and subscriber information systems including 80-90% of all Finnish mobile terminals/subscribers in fall 2005

- Terminal base old, key features for packet data usage not widely spread (packet data 48%, EDGE 13%, smartphones 6%, WCDMA 0,5%)
- Nokia's terminal market share 87%. Over 99% of smartphones Symbian based, 1/3 of which Nokia communicators
- Mobile terminal installed base concentrated, 88% of all terminals among top 50 models

- 99% of consumer subscribers on usage-based packet data tariff category, creating 18% of consumer packet data traffic
- 92-94% of mobile subscribers postpaid, 75% out of which consumers. Business subscribers create 62% of packet data traffic

HELSINKI UNIVERSITY OF TECHNOLOGY
Networking Laboratory

TCP/IP Header Collection –Based Measurements

Mobile Internet Usage Measurement

Contents

- Measurement description
  - Scope of the measurements
  - Measurement setup
  - Identification of terminal operating system
- General packet data traffic patterns
  - Traffic distribution by operating system
  - General traffic patterns by operating system
- Packet data traffic by application protocol
  - Traffic distribution by application protocol
  - Traffic distribution by application protocol category
  - Traffic distribution by application protocol category and operating system
- Web protocol traffic patterns
  - Most popular web sites
  - Most popular web sites by category
  - Most popular web sites by category and operating system
- Summary

Measurement Description

Scope of measurements

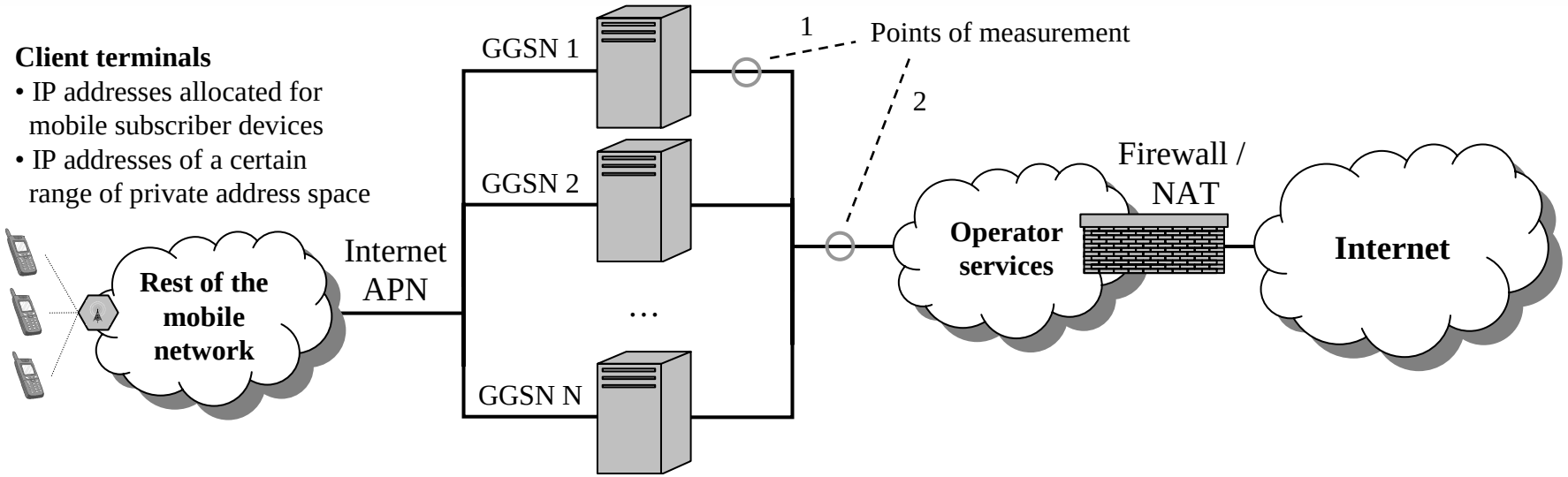
- MSOs included: Sonera and DNA
  - No data on: Elisa, Saunalahti, TeleFinland, others
  - 50 – 60% of all Finnish mobile subscribers (business and consumer, postpaid and prepaid)

- Packet data traffic at Internet APN measured during one week
  - TCP, UDP and IP headers captured
  - ≈90% of all packet data traffic (all APNs) goes via Internet APN
  - Measurements not simultaneous, difference 1,5 weeks (weeks 38-40/2005)

→50% of all Finnish mobile network packet data traffic during one week included

Measurement Description

Measurement setup



- Comparable points of measurement used
  - Traffic quantities (bytes, flows) of measurement 1 multiplied by the actual number of GGSNs in order to have proper weight for the operator's traffic
- Measured traffic not influenced by roaming, as home GGSN roaming is used by both operators
  - All roaming traffic by operators' subscribers routed via home network's GGSN → all packet data roaming traffic by operators' subscribers included, no foreign roamers' traffic included
- Clients and servers identified using terminal IP addresses
  - Client terminals were always on one side of traffic, all other IP addresses considered servers
  - Problem: one operator has public IP addresses for mobiles → client-server roles are sometimes reversed

Measurement Description

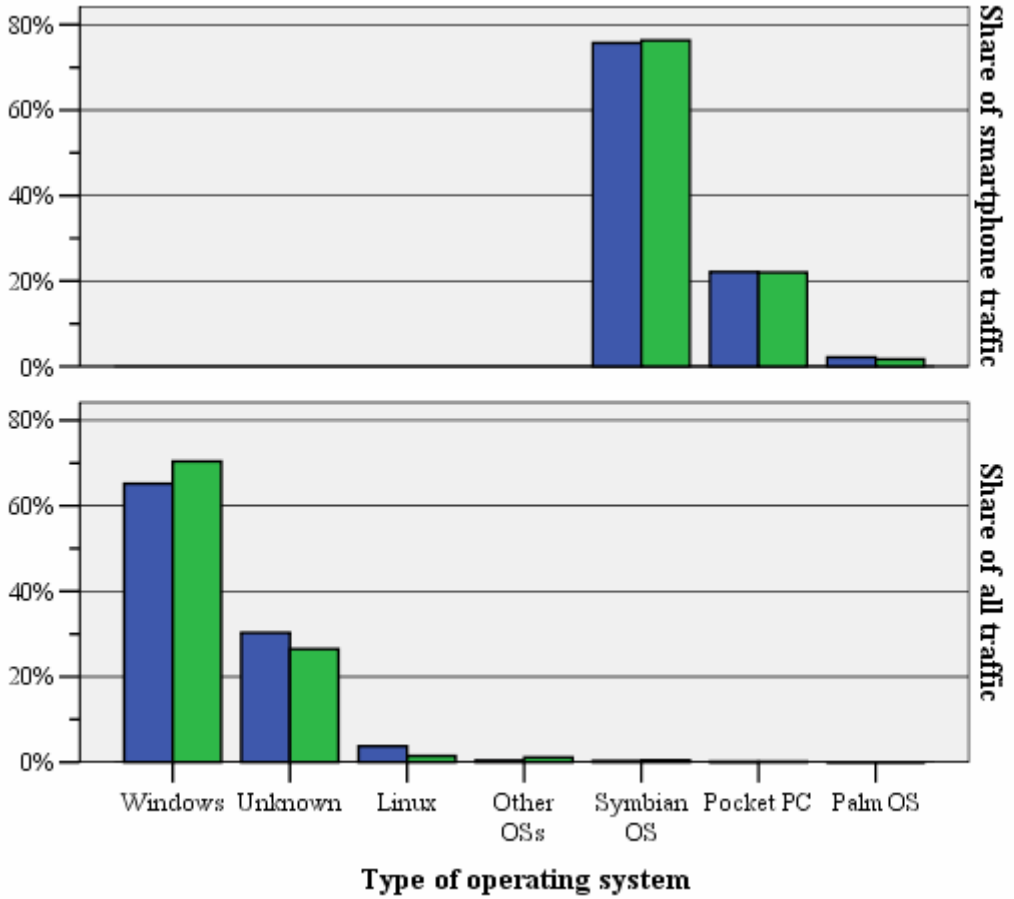
Identification of terminal operating system

- Terminal operating system (OS) identified using TCP fingerprinting
  - Differences in implementation of TCP/IP stack in different OSs → distinct TCP ”fingerprints”
  - Traffic traces are compared to the fingerprints of previously identified OSs
  - Most of the common PC and smartphone OSs can be identified with reasonable accuracy
  - TCP fingerprinting identifies the OS of ≈80% of traffic from fixed Internet traffic traces

- Operating system identification process includes some possible bias
  - Only client OSs identified → both uplink and downlink traffic accounted for the client OS
  - OS identification is based on uplink TCP traffic only (37% of flows, 5% of bytes)
    - OS of uplink TCP flows identified → OS resides at a certain client IP address at a certain time frame
    - Downlink TCP flows, and all UDP flows accounted for different OSs based on this information
  - What is the effect of the 63% of non-identified flows on OS identification accuracy?
    - OS identified correctly following the first uplink TCP flow as long as the user has the same IP address
    - VPN traffic should be mostly allocated on the correct OS, if VPNs are not on as a default (i.e. if there is at least one uplink TCP flow before VPN usage is started)
    - OS identification of TCP based application protocols (e.g. web, email) is more reliable

General packet data traffic patterns

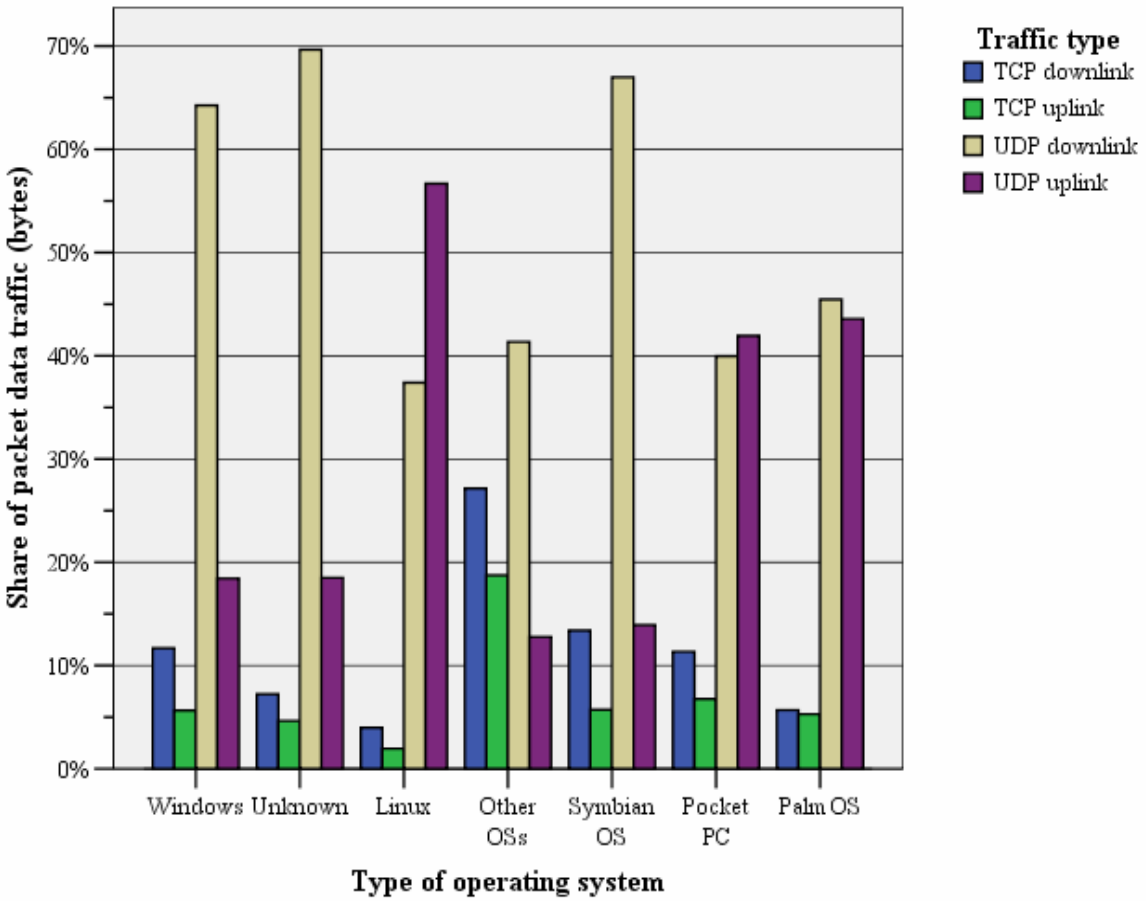
Traffic distribution by operating system



- Windows originates 65% of traffic in mobile network
  - Data cards, GPRS modems, handsets via Bluetooth/cable…
  - A few PCs create more traffic than many mobiles → OS identification necessary to uncover mobile usage
- Unknown 30% of traffic problematic
  - All non-smartphone handsets, possibly additional laptop and smartphone traffic
  - Telematics, machine-to-machine (M2M) comm., alarm terminals, remote cameras…?
  - Could an intelligent modem / GPRS module, VPN or firewall alter the TCP fingerprint?
- Surprisingly large share of smartphone traffic by Pocket PCs
  - Pocket PCs without GSM capability use other devices for network access→ more Pocket PCs than the <1% of terminal base mobile operators see
  - Even 1% of Symbian traffic in Unknown would alter traffic distribution considerably

General packet data traffic patterns

General traffic patterns by operating system



- 85% of traffic volume UDP
  - Fixed Internet: TCP>>UDP
  - UDP used by VPN protocols (for NAT traversal), DNS, WAP, MMS, streaming…
- Windows and Unknown have very similar profiles
  - Is Unknown dominantly Windows as well?
- Symbian profile is different from other smartphones
  - Pocket PC and Palm OS profiles are very similar
  - Symbian profile is rather similar to Windows profile
- Other remarks
  - Downlink > uplink mostly
  - Linux >90% UDP, mostly uplink

Packet data traffic by application protocol

Traffic distribution by application protocol

Rank*	TCP port	Share of TCP traffic		UDP port	Share of UDP traffic	
		Bytes	Flows		Bytes	Flows
1.	80	80 %	74 %	53	7 %	54 %
2.	443	7 %	8 %	2746	17 %	0 %
3.	135	1 %	3 %	4500	15 %	0 %
4.	110	1 %	2 %	10000	13 %	0 %
5.	143	1 %	1 %	370	12 %	0 %
6.	445	1 %	1 %	500	8 %	0 %
7.	8080	1 %	1 %	4672	0 %	7 %
8.	1863	0 %	1 %	0	3 %	0 %
9.	25	0 %	1 %	5003	2 %	0 %
10.	7171	0 %	0 %	32555	2 %	0 %
11.	6346	0 %	0 %	6346	0 %	2 %
12.	4662	0 %	0 %	39273	1 %	0 %
13.	4283	0 %	0 %	9183	0 %	1 %
14.	28467	0 %	0 %	45991	1 %	0 %
15.	139	0 %	0 %	123	0 %	1 %
Others	Others	6 %	7 %	Others	16%	35%

- TCP/UDP server port based application protocol identification is ambiguous
  - About 64000 UDP and 55000 TCP server ports → client ports observed due to the use of public IP addresses
- TCP traffic mainly web and email
  - HTTP (80, 8080), HTTPS (443)
  - POP3 (110), IMAP (143), SMTP (25)
  - Windows “self-initiated” (135, 445)
- UDP traffic mainly DNS and VPN
  - DNS (53)
  - VPN (2746, 4500, 10000, 500)
    - 2746: CheckPoint UDP Encapsulation
    - 4500: IPsec / NAT-Traversal
    - 10000: Network Data Management Protocol (NDMP), also Cisco IPsec VPN
    - 500: ISAKMP / IKE
  - F-Secure updates / BackWeb (370)?
  - High share of other protocols/ports

* Ranked by the TCP/UDP port’s combined share of bytes and flows

Packet data traffic by application protocol

Traffic distribution by application protocol category

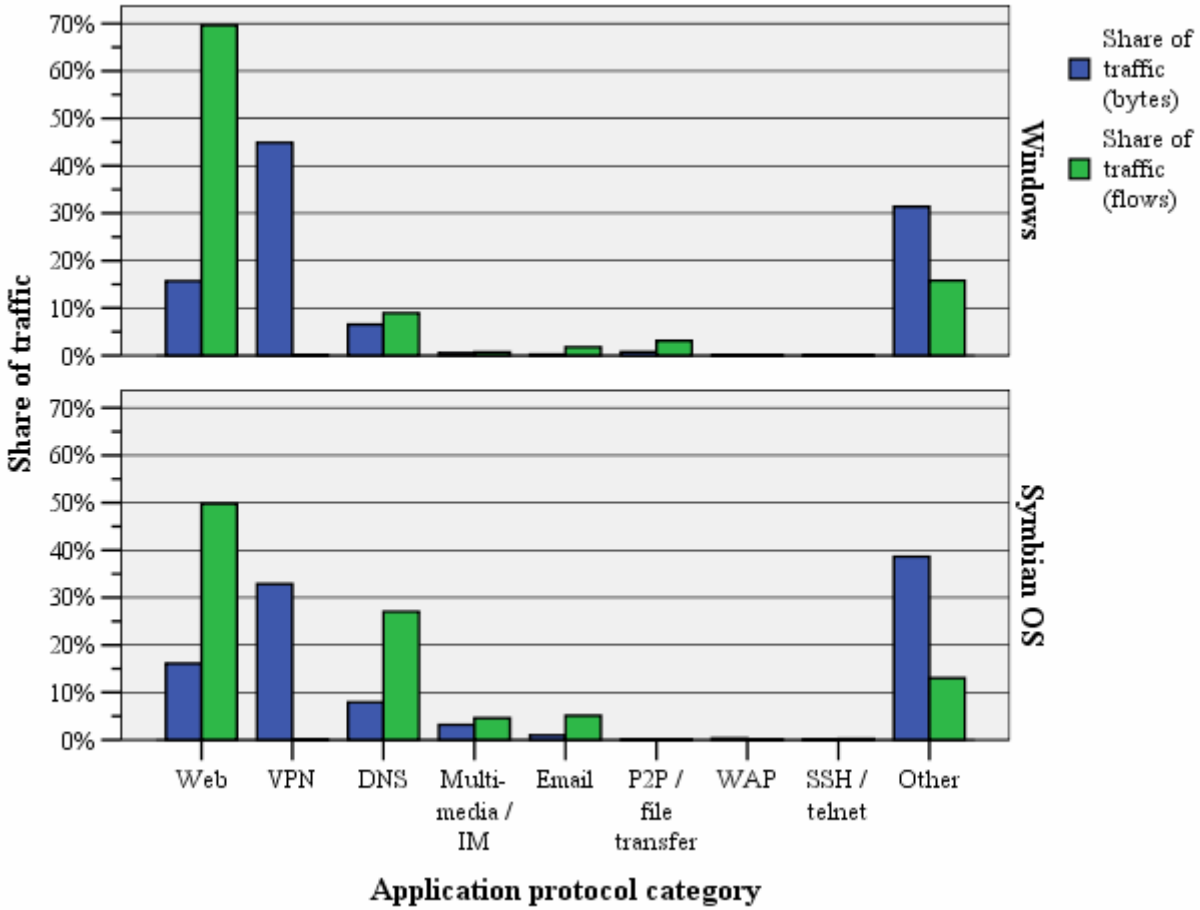
Rank	Application protocol category	Major protocol ports included*	Share of traffic	
			Bytes	Flows
1.	Web	TCP: 80, 443, 8080	13,6%	64,0%
2.	VPN	TCP: 10000 UDP: 500, 2746, 4500, 10000, 1194	45,5%	0,1%
3.	DNS	UDP: 53	6,2%	12,0%
4.	Multimedia / IM	TCP: 1863, 5001, 6667, 554 UDP: 5003, 5001, 5000	2,7%	0,8%
5.	Email	TCP: 110, 143, 25, 993, 995	0,4%	2,7%
6.	P2P / file transfer	TCP: 6346, 4662, 21, 20, 1214 UDP: 4672, 6681	0,5%	2,4%
7.	WAP	TCP: 9200, 2949, 2805, 2923, 4035, 2948, 4036, 9202, 9201 UDP: 9201, 9203, 9202, 9204, 2805, 9200, 2923, 4036, 4035	0,1%	0,0%
8.	SSH / telnet	TCP: 22, 23	0,0%	0,0%
	Other	TCP: 135, 445, 7171, 4283, 139, 28467, 19977, 8081, 50123, 1435 UDP: 370, 0, 32555, 39273, 45991, 6346, 48000, 49000, 10001, 9183, 9181, 8889, 123, 434, 137, 9872, 12345	31,1%	17,8%

* TCP and UDP server ports with at least 0,5% of the total bytes or flows in the category

- VPN 46% of traffic volume
  - Very few flows, as should be
  - Inside VPN is another and possibly different protocol profile
- Web and DNS major applications
  - 25% and 11% of non-VPN traffic
- >30% uncategorized protocols
  - Self-initiated Windows traffic, client ports, malware, P2P…
  - Categorization simplifies, but is very ambiguous
- Other remarks
  - Email share small, but VPN and Web likely include lots of email
  - P2P much smaller than in fixed Internet
  - Multimedia / IM (audio/video conferencing protocols, streaming, IM, IRC) protocol traffic probably partly in ”Other”
  - WAP share is small, as WAP APN traffic was not measured

Packet data traffic by application protocol

Traffic distribution by application protocol category and operating system



- Windows imposes itself on general traffic profile
  - 65% of traffic Windows…
- VPN also used on Symbian
  - Nokia Mobile VPN Client, related to Check Point
- Lots of uncertainty
  - Combined error of both OS identification and protocol categorization
- Other remarks
  - Windows and Symbian seem to have fairly similar profiles
  - Symbian shares for email, DNS and Multimedia / IM are higher
  - Symbian share of “Other” category is higher

Web protocol traffic patterns

Most popular web sites

Rank*	Domain name of site	Share of web traffic volume	Share of web site visits**
1.	mtv3.fi	4,3 %	3,9 %
2.	doubleclick.net	4,9 %	2,4 %
3.	basefarm.net	2,5 %	3,4 %
4.	irc-galleria.net	3,1 %	2,6 %
5.	luukku.com	1,6 %	1,5 %
6.	google.com	0,9 %	1,5 %
7.	hotmail.com	1,4 %	0,9 %
8.	nebula.fi	1,0 %	1,1 %
9.	yahoo.com	1,2 %	0,9 %
10.	sihteeriopisto.net	1,1 %	1,0 %
11.	adtech.de	1,0 %	1,0 %
12.	sampo.fi	0,9 %	0,8 %
13.	htv.fi	0,8 %	0,8 %
14.	yle.fi	0,9 %	0,8 %
15.	akamaitechnologies.com	0,7 %	0,8 %
	Other identified domains	45,4%	47,7%
	Unknown addresses	22,1%	24,3%
	Private addresses	6,2%	4,9%

- Server IP addresses of all TCP flows with server ports 80, 8080, 8000, 8888, and 443 included
  - Might include P2P or malware traffic as well (traffic to e.g. port 80 goes through firewalls)
  - → 75000 web server IP addresses
  - → 73000 domain / sub domain names
    - First PTR record listed taken into account
  - → 19000 domain names
- Web traffic not very concentrated to few domains
  - 48% in top 10, 77% in top 100, 89% in top 400
- Unknown addresses >22% of traffic volume
  - Server (web or other) IP addresses for which no reverse DNS entry was available
- Private addresses 6% of traffic volume
  - Web/WAP servers in operator internal network?

* Ranked by the domain's combined share of bytes and flows

** Share of TCP flows to/from the domain

# of web site visits <= # of flows <= files downloaded from site

Web protocol traffic patterns

Most popular web sites by category

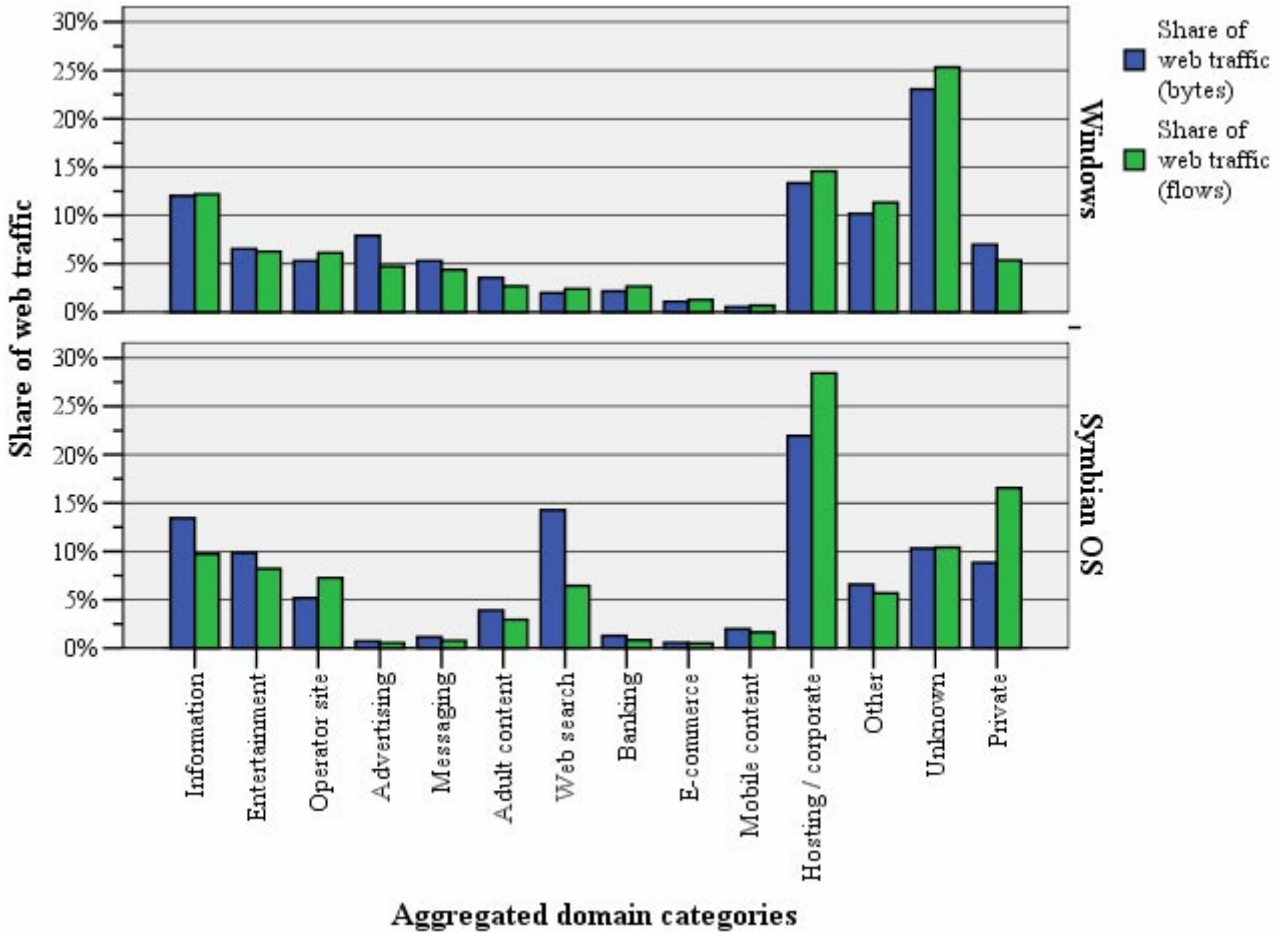
Rank*	Site category	Major sites included**	Share of web traffic volume	Share of web site visits
1.	Information	mtv3.fi, yle.fi, sanomawsoy.fi, almamedia.fi, helsinginsanomat.fi	12,9%	12,8%
2.	Entertainment	irc-galleria.net, veikkaus.fi, sm-liiga.fi, telkku.com	6,0%	5,9%
3.	Operator site	-	5,6%	6,3%
4.	Advertising	doubleclick.net, adtech.de, tradedoubler.com	7,4%	4,5%
5.	Messaging	luukku.com, hotmail.com, gmail.afraid.org, msn.com, passport.com, passport.net	5,3%	4,4%
6.	Adult content	sihteeriopisto.fi, seksitreffit.fi	3,6%	2,7%
7.	Web search	google.com, yahoo.com	2,1%	2,5%
8.	Banking	sampo.fi, eQonline.fi, op.fi, nordea.fi	2,1%	2,4%
9.	E-commerce	huuto.net, mobile.de, infosto.fi, thomann.fi, ebay.com, verkkokauppa.com	1,0%	1,1%
10.	Mobile content	jippii.net, jamster.com, mobilenator.com, buumi.net	0,6%	0,8%
	Hosting / corporate site	basefarm.net, nebula.fi, akamaitechnologies.com	14,2%	15,3%
	Other	-	10,8%	11,8%
	Unknown	-	22,4%	24,5%
	Private	-	6,2%	4,9%

- Information (13%) and entertainment (6%) both significant
- Advertising >7% share notable
  - Pop up windows etc.
- Multiple sources of error
  - 40% share of unknown and uncategorized sites
  - Non-web traffic possibly included
  - Domain name resolving method giving hosting service providers and not the hosted services
  - Categorization based on domain name, not sub domain name
  - Categorization itself is subjective, ambiguous, and error prone
  - Overlapping categories (Information and Entertainment, Web search and Messaging, Operator site and Mobile content)
  - Web traffic inside VPN possibly having an entirely different profile

* Ranked by the domain's combined share of bytes and flows
** Sites with at least 5% of the total bytes or flows of the category

Web protocol traffic patterns

Most popular web sites by category and operating system



- Windows imposes itself on general web profile
  - 65% of traffic Windows…
- Symbian profile could indicate services usable on a handset display
  - Web messaging, banking, e-commerce not really usable?
- Lots of uncertainty
  - Combined error of both OS identification and site categorization
- Other remarks
  - Share of Windows higher in Advertising, Messaging, Unknown
  - Share of Symbian higher in Hosting, Mobile content, Web search, Private addrs.

Summary

- 50% of Finnish mobile network packet data traffic captured at two mobile operators' Internet APN during one week in fall 2005

- Windows originates 65% of traffic in mobile networks → mobile usage profile hidden by Windows traffic
- VPN usage creates 46% of traffic → very high 85% share of UDP compared to that of fixed networks
- 90% of all packet data traffic (all APNs) goes via the Internet APN
- Web also a major application with 14% of traffic volume, 25% of non-VPN traffic volume

HELSINKI UNIVERSITY OF TECHNOLOGY
Networking Laboratory

Handset –Based Measurements

Mobile Internet Usage Measurement

Contents

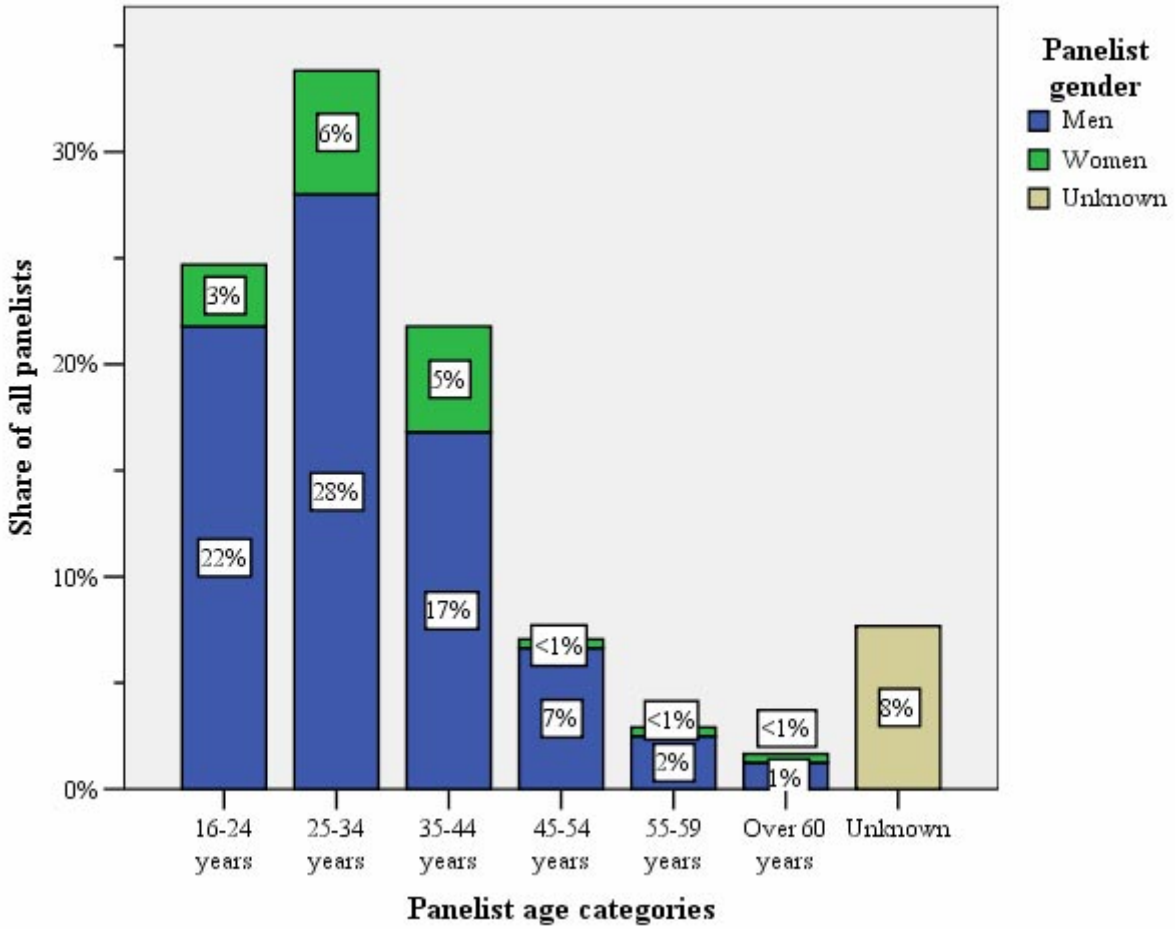
Handset –Based Measurements

Measurement description

- Smartphone360 (SP360) research platform enabling monitoring of usage events on Symbian/S60 handsets
  - Monitoring software installed into Symbian S60 handsets with user agreement
  - Registering all handset usage events to log files
  - Log files automatically sent to server for further analysis
  - No information on the content of communication registered
- Participants recruited using SMS campaign and web based registration process
  - Targeted number of participants (panelists) recruited through various means
    - >20000 SMSs invitations sent to smartphone subscribers with a permission on SMS contacting
    - Additional mobile operator customer service phone calls to secure adequate participation
  - Registration and software installation by the user using web interface
  - Panelist background information (demographics etc.) collected via web questionnaires
  - Incentive for participation was a 20€ compensation for expenses and a possibility to win a Nokia N70 handset
- A panel of 500 handsets monitored for 3 months during fall 2005
  - 8 first weeks (56 days) of panel participation for each panelist studied
  - 482 sufficiently active panelists included in data representing some 180 000 Finnish Symbian/S60 users
    - Sufficiently active: at least 28 active days during the 56 day panel period
    - Active day: messaging, calling, data usage or application launches during the day
  - Participating operators (Sonera, Elisa and DNA) equally represented
  - Geographic distribution of panelists unknown
- Privacy issues central during all phases of the research
  - Recruitment (mobile marketing), monitoring, processing and analysis, …

Description of the panel

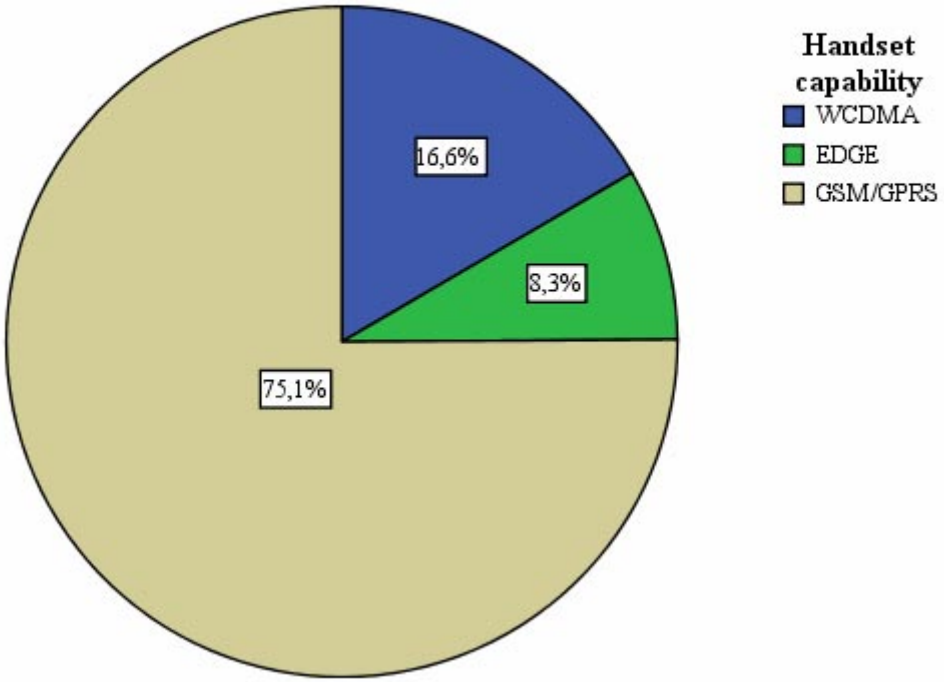
Panel by gender and age



- Panelists dominantly young to middle-aged men
  - 77% of panelists men
    15% of panelists women
  - 59% of panelists under 34 years old
  - Does this correspond to smartphone users' general demographic distribution?
  - Is this partly a result of the recruiting method?
- 8% of panelists with unknown demographics
  - Excluded when analyzing effect of demographics to usage

Description of the panel

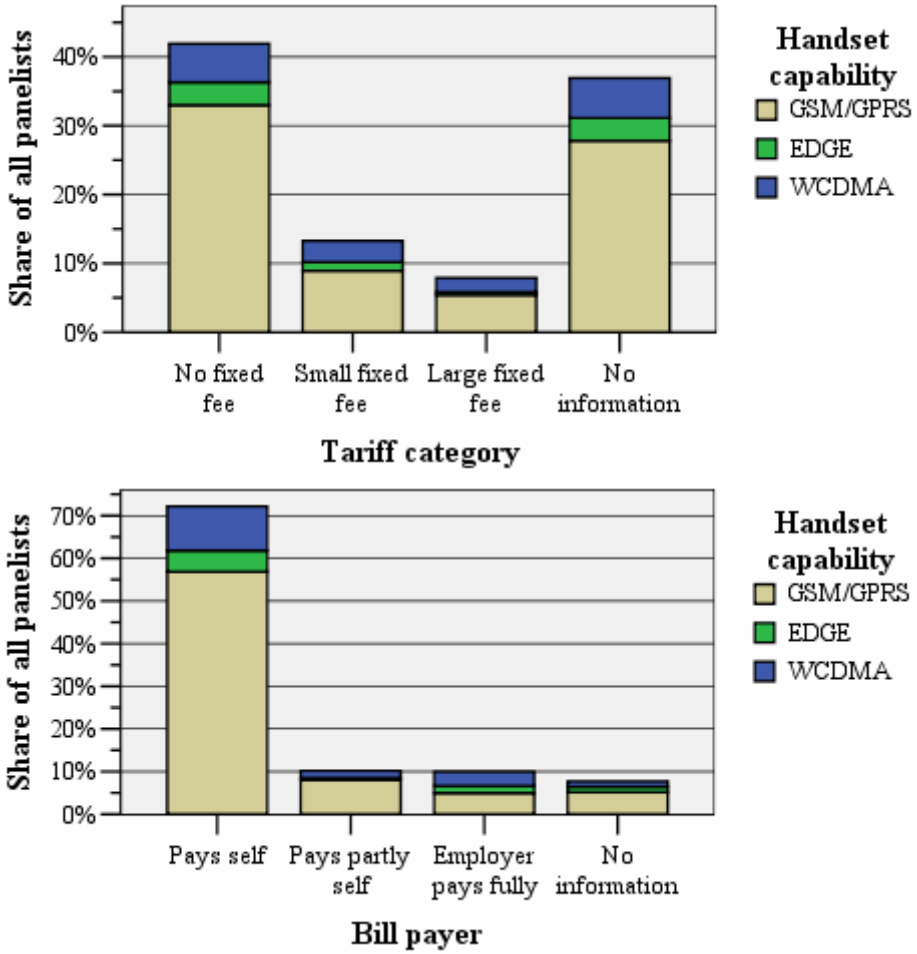
Panel by handset type



- Panelists used 8 different smartphone models
  - 8 Nokia Symbian/S60 handset models
  - Handset capability refers to handset's highest radio capability
    - WCDMA → handsets with WCDMA, EDGE and GSM/GPRS capability
    - EDGE → handsets with EDGE and GSM/GPRS capability
  - In all subsequent analyses WCDMA, EDGE and GSM/GPRS refer to the capability of the handset, not to the bearer actually used
- Panelists are dominantly users of GSM/GPRS capable handsets
- Share of EDGE capable handsets is relatively small

Description of the panel

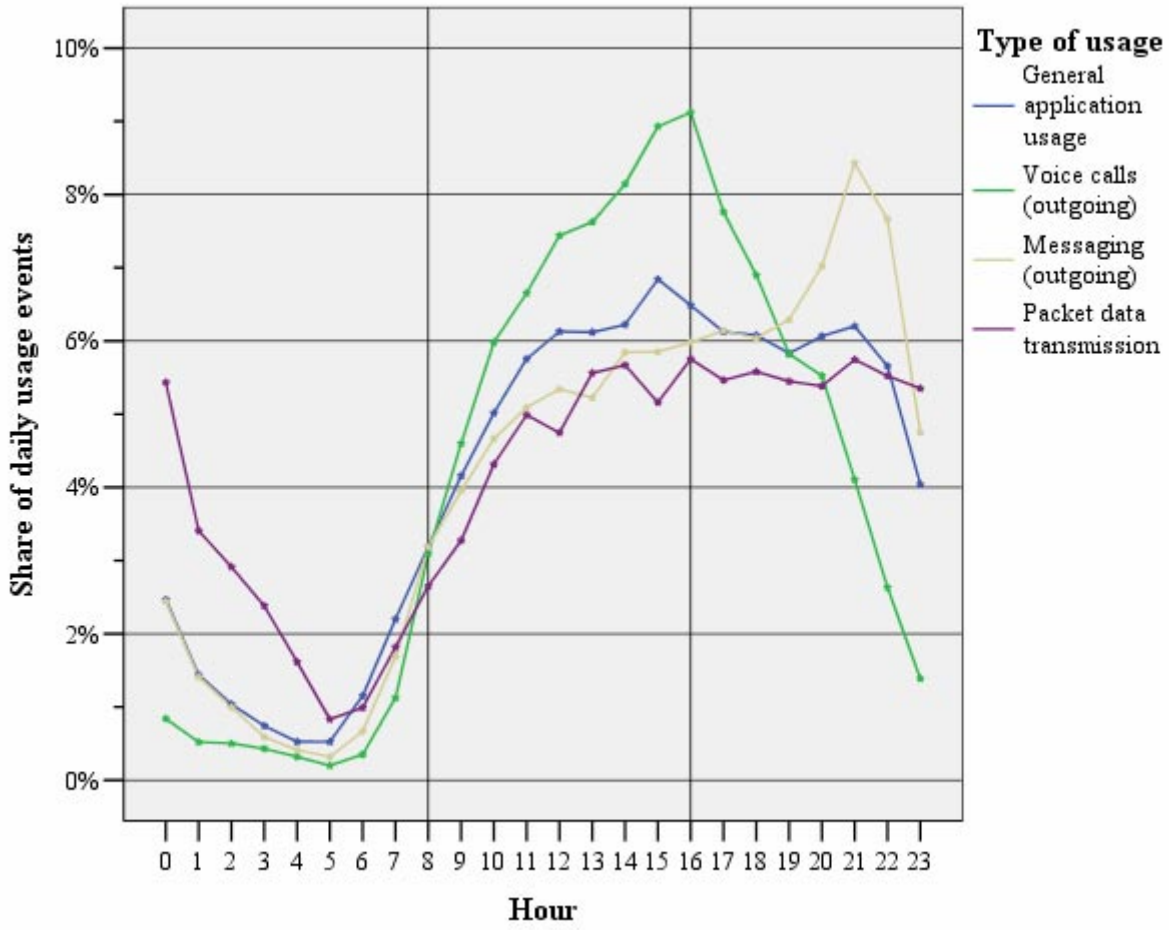
Panel by panelist packet data tariff category and bill payer



- 3 aggregated packet data tariff categories
  - No fixed fee (purely usage-based)
  - Small fixed fee (block-based: 2-50 MB/month)
  - Large fixed fee (block-based: 100-500 MB/month, or flat-rate)
- 42% of panelists with “No fixed fee” packet data tariff category
  - 37% of panelists without background information on tariff category
  - Large fixed fee was proportionally highest in “Employer pays fully” bill payer category
- 72% of panelists pay handset bills themselves
  - Consumer subscribers targeted in recruiting
  - Corresponds to distribution of all mobile subscribers (75% consumers). What is the distribution for smartphone users?
  - Category ”Pays partly self” includes cases where bill payer was somebody in the family
- Handset capabilities distributed to alternative categories fairly equally
  - Less effect on handset capability specific analyses from tariff category and bill payer

General handset usage patterns

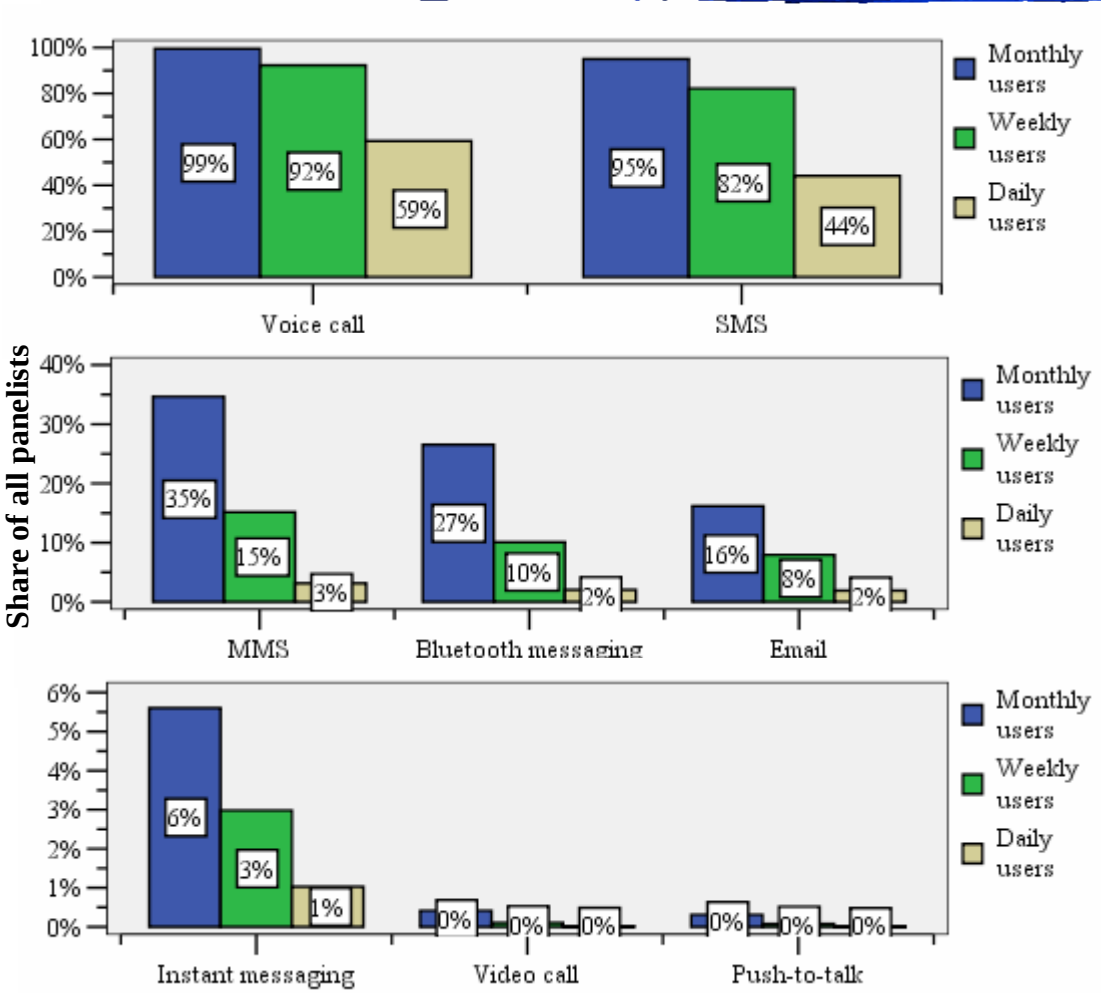
Daily handset usage patterns



- Calling at work
  - Voice call activity most concentrated on office hours, peak at 4 PM
- Messaging activity increases in the evening
  - Most active messaging (SMS and MMS) at 9PM
- Data usage also during night time
  - Data transmission activity also present throughout the night

General handset usage patterns

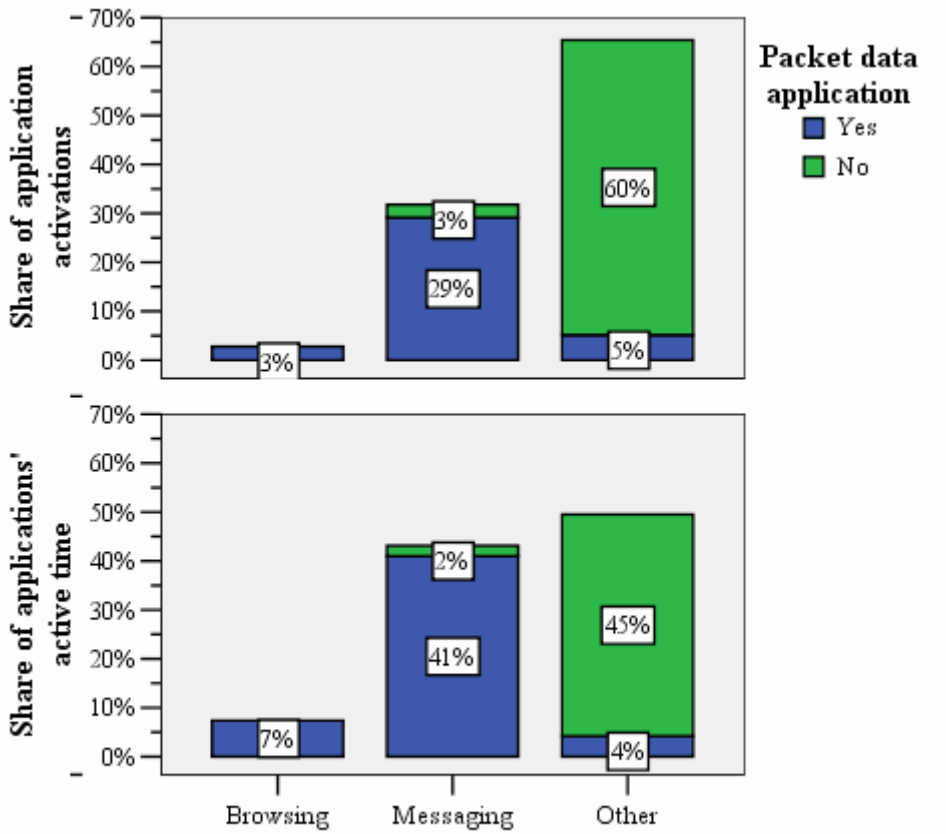
Communication service usage frequencies



- Average share of monthly/weekly/daily communication service users
  - Only outbound voice calls, SMSs, and MMSs included
  - Bluetooth messaging includes e.g. business cards, calendar updates, image sending, not e.g. modem or hands free usage
  - Email includes platform email applications, not 3rd party email apps. nor webmail
  - Instant messaging (e.g. Agile Messenger) and push-to-talk counted from packet data traffic generated with IM and PoC applications
- Traditional calling and messaging dominate
  - Voice calls used more often than SMSs
- Other communication services used seldom
  - MMS used most, but clearly less than voice call and SMS
- Instant messaging used monthly by 6%
- Other remarks
  - Video call and push-to-talk usage marginal, as their availability was limited during panel
  - Emails more received than sent on handsets
  - What is the temporary effect of marketing on comm. service usage? E.g. MMS campaigns?

General handset usage patterns

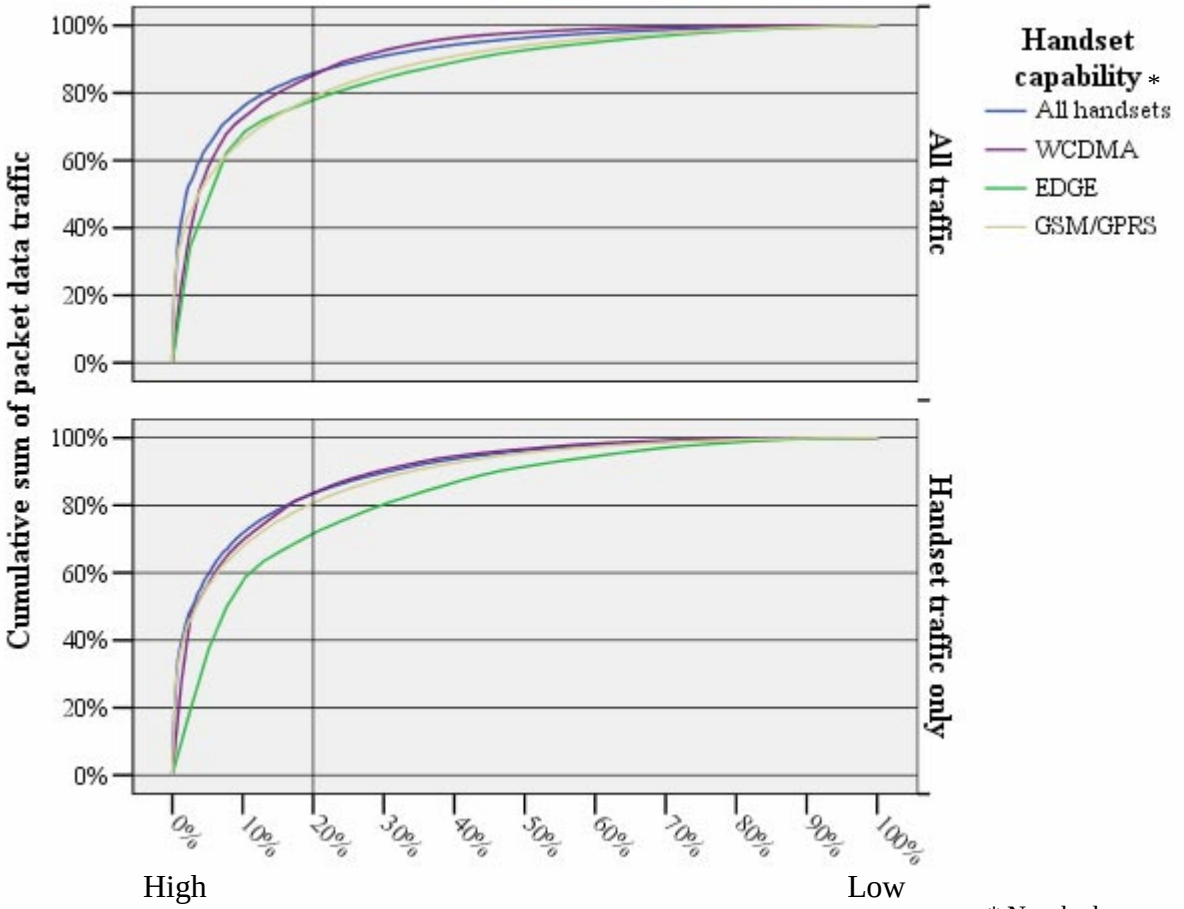
Handset application usage activity



- Packet data generating applications account for 40-50% of all handset application activity
  - 37% of application activations
    53% of application active time
  - Mostly messaging and browsing related applications
  - All activity on packet data generation capable applications is not necessarily related to packet data traffic
- Rest of usage mainly personal information management (PIM), multimedia, and utility applications
  - PIM: phonebook, calendar, clock apps. …
  - Multimedia: camera apps., media players …
  - Utility: file browsers, GPS apps. …
- Other remarks
  - All ”always on” applications (Phone, Menu) have been removed from analysis

General packet data usage patterns

Accumulation of packet data traffic by panelist

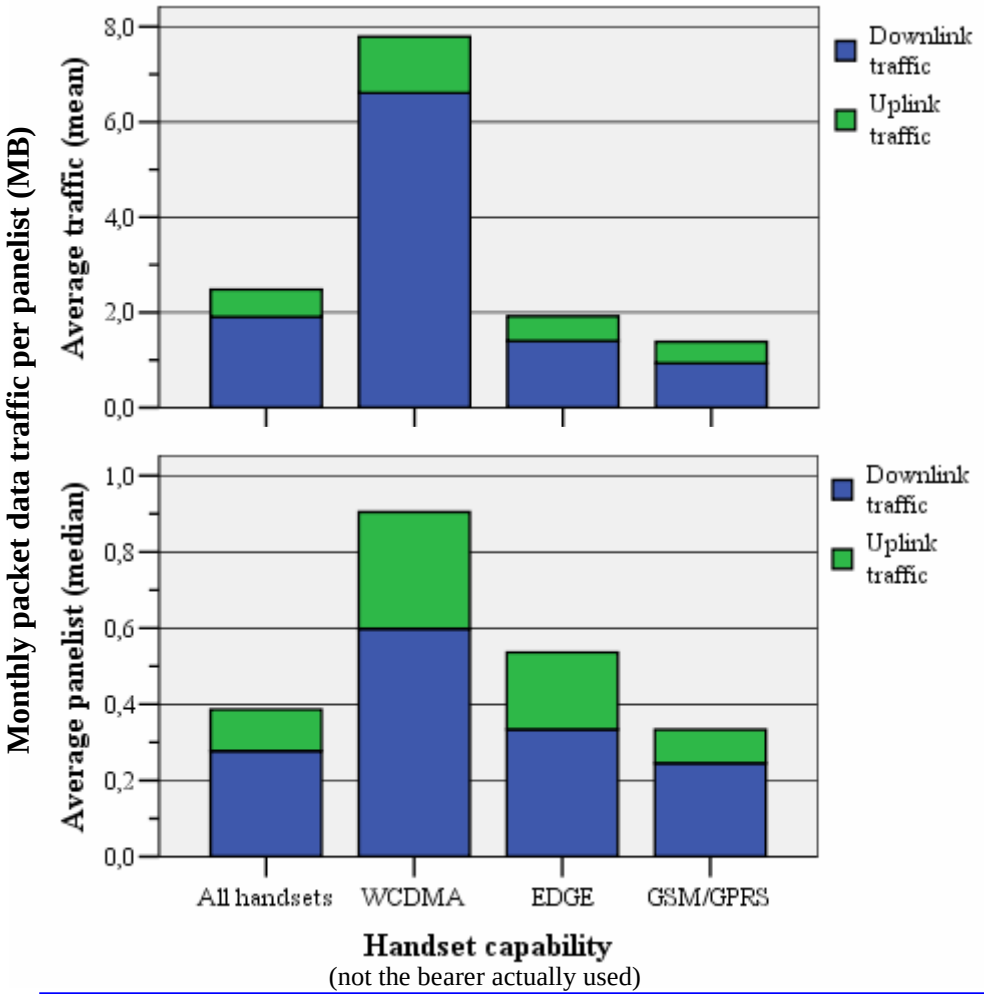


* Not the bearer actually used

- 20% of panelists create 80% of packet data traffic
  - Panelists using WCDMA handsets?
  - Panelists using handset as a modem?
  - Panelists with flat / large fixed fee GPRS tariff plan?
  - Panelists whose employer pays the bill?
  - Other true heavy users?
- The 20/80 – rule applies also when modem usage is filtered
  - Relation weaker for EDGE handsets → smaller sample?
- Using averages gives too much weight on the heavy using 20% of panelists
  - Despite this, average i.e. (arithmetic) mean used in most analyses
  - Median describing ”average panelist” gives more truthful results in some cases

General packet data usage patterns

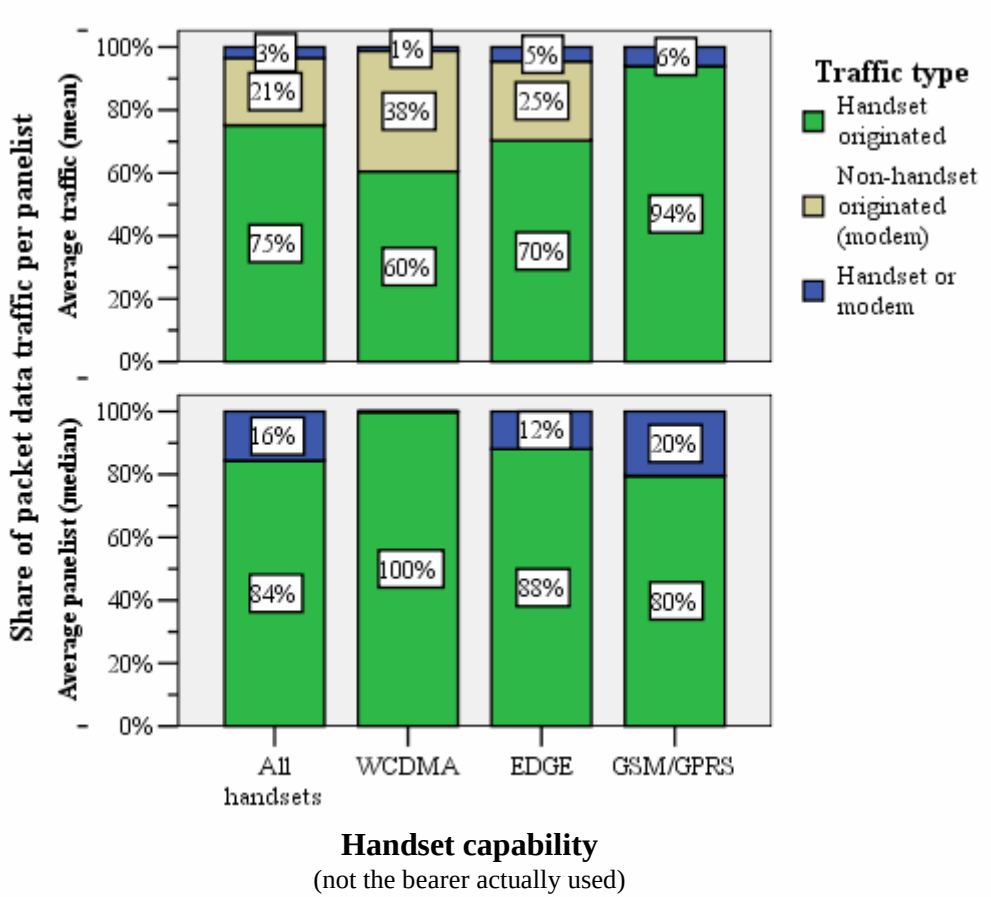
Packet data traffic volumes



- Average traffic volume is higher among users of more capable handsets, especially with WCDMA
  - High usage leads people to acquire capable handsets
  - Using a capable handset increases data usage volume
  - Chicken and egg…
- Average volumes per panelist (mean) are much larger than volumes of ”average panelist” (median)
  - Overweighted heavy users…

General packet data usage patterns

Significance of modem usage



- Identifying modem traffic ambiguous
  - Three traffic categories
    - Handset originated traffic
    - Non-handset originated traffic (modem traffic)
    - Handset or modem traffic (ambiguous cases)
  - Number of modem users not accurately determinable
- 21-25% of panelist's packet data traffic is modem traffic on average (mean)
- "Average panelist" (median) has less/no modem traffic
  - Few modem users create a large share of all traffic, especially while using WCDMA capable handsets
  - Higher figures for EDGE and GSM/GPRS result from ambiguous identification of modem traffic

General packet data usage patterns

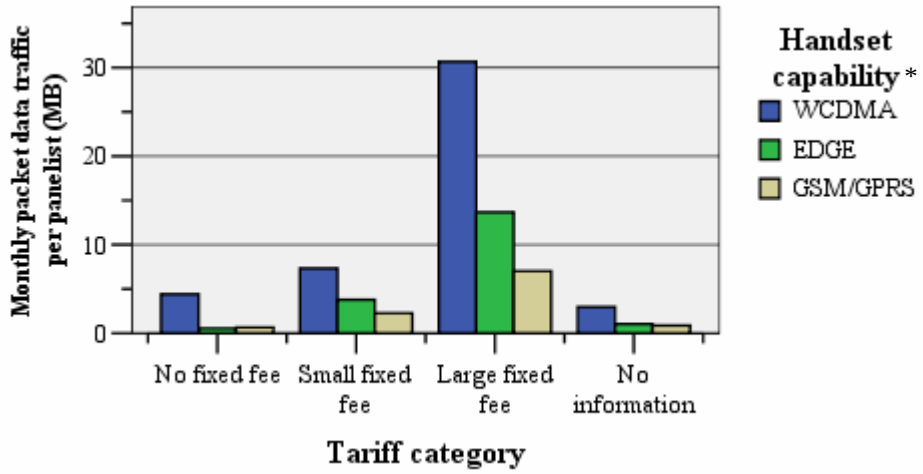
Usage of bearers on packet data traffic by handset capability

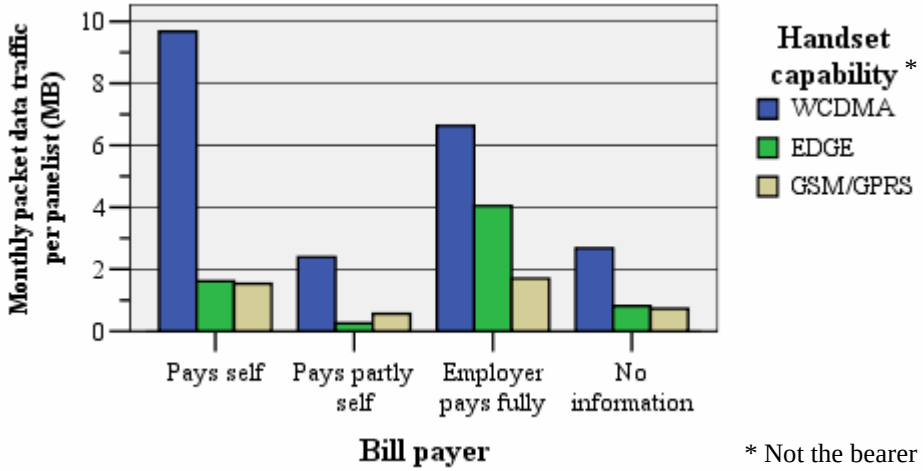
Bearer used	WCDMA capable handsets	EDGE capable handsets	GSM/GPRS capable handsets
WCDMA	45%	-	-
EDGE	10%	30%	-
GSM/GPRS	45%	70%	100%

- Most capable bearer not fully used, probably mostly due to network coverage reasons
  - Network coverage and non-coverage
    - WCDMA (and EDGE) networks do not cover neither entire geographic area nor all mobile subscribers in Finland
    - WCDMA network access was not provided by all operators during the panel
  - Effect of subscription and operator
    - Older SIM cards possibly do not support WCDMA usage?
    - Do operators prioritize some subscription types regarding 2G/3G network usage?
    - Load balancing or prioritizing certain services (e.g. voice calls) by operators?
  - Effect of handset user
    - User can select the preferably used network
- Usage share of EDGE bearer on EDGE capable handsets is surprisingly small
  - EDGE network coverage should be better than WCDMA coverage → share should be at least 55%?
  - Is EDGE capacity small, i.e. not enough room for all EDGE users under EDGE coverage
  - Could there be handset type specific problems in EDGE usage?
  - Sample size in this analysis is even smaller than the already small EDGE sample

General packet data usage patterns

Packet data traffic by pricing category and bill payer





* Not the bearer actually used

- Handset capability clearly drives data usage, especially with WCDMA
  - Similar results obtained when considering medians instead of means
- Panelists with higher fixed fees have more data usage
  - High volume data usage leads people to choose relatively cheaper larger fixed fees
  - Effectively flat fee pricing on data increases usage volume further
  - Chicken and egg…
- Effect of bill payer on data usage is not clear
  - Heavy users use data in high volumes regardless of who pays the bill?
  - Usage volume might increase when employer pays the bill
    - What if the bill is part of salary? Is it still “paid by somebody else”?

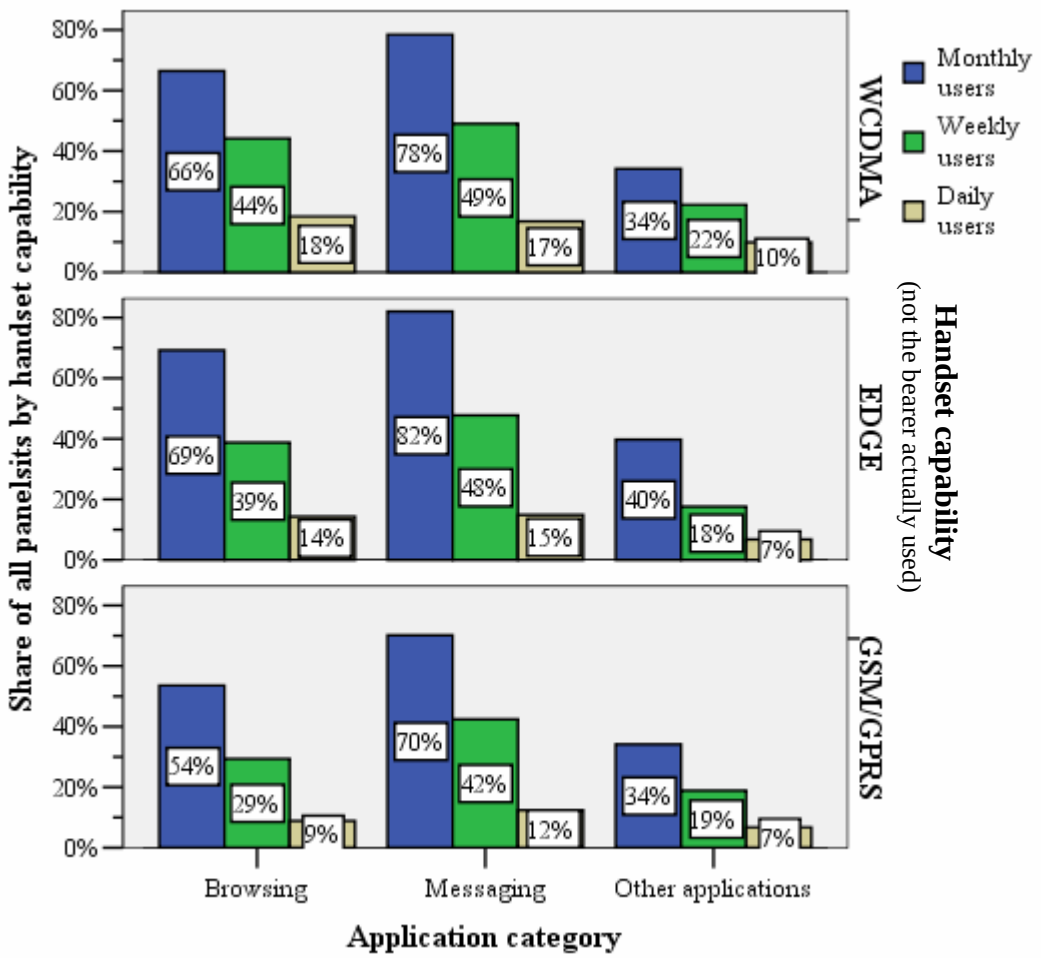
Packet data generating application usage

Categorization of packet data applications

- Packet data applications divided into functional categories
  - Categorization based on application name, or its absence (non-Symbian originated traffic)
  - Only packet data traffic generating applications included
- Browsing
  - Nokia platform browser, Opera, NetFront
- Messaging
  - Platform messaging applications
  - Instant messaging (Agile Messenger, WirelessIRC, IM+…)
  - Email (Profimail…)
- Other applications
  - All other applications with an application name not already in ”Browsing” or ”Messaging”
  - E.g. Anti-virus, PuTTY, Symella, Nowire Teletext…
- External applications (modem usage)
  - All applications without an application name
  - Combines two modem usage categories seen before (“Non-handset originated traffic” and “Handset or modem traffic”)

Packet data generating application usage

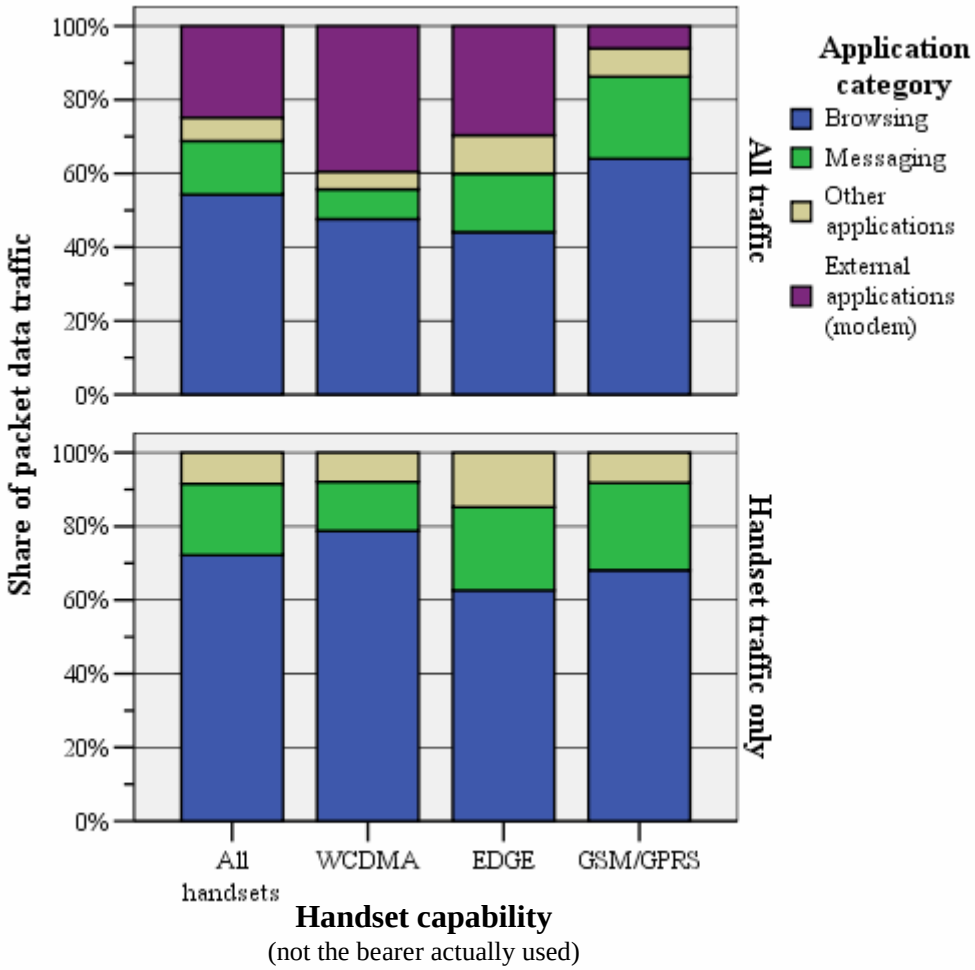
Packet data application usage frequencies



- Messaging applications most frequently used
- Browsing frequency increases with handset capability
- Messaging frequency not as clearly driven by handset capability
- Other remarks
  - "Other applications" is a heterogeneous category
    - No major differences between handset capabilities
    - In general less used than messaging and browsing
  - Browsing is the key to other usage?
    - Other applications are found via browsing

Packet data generating application usage

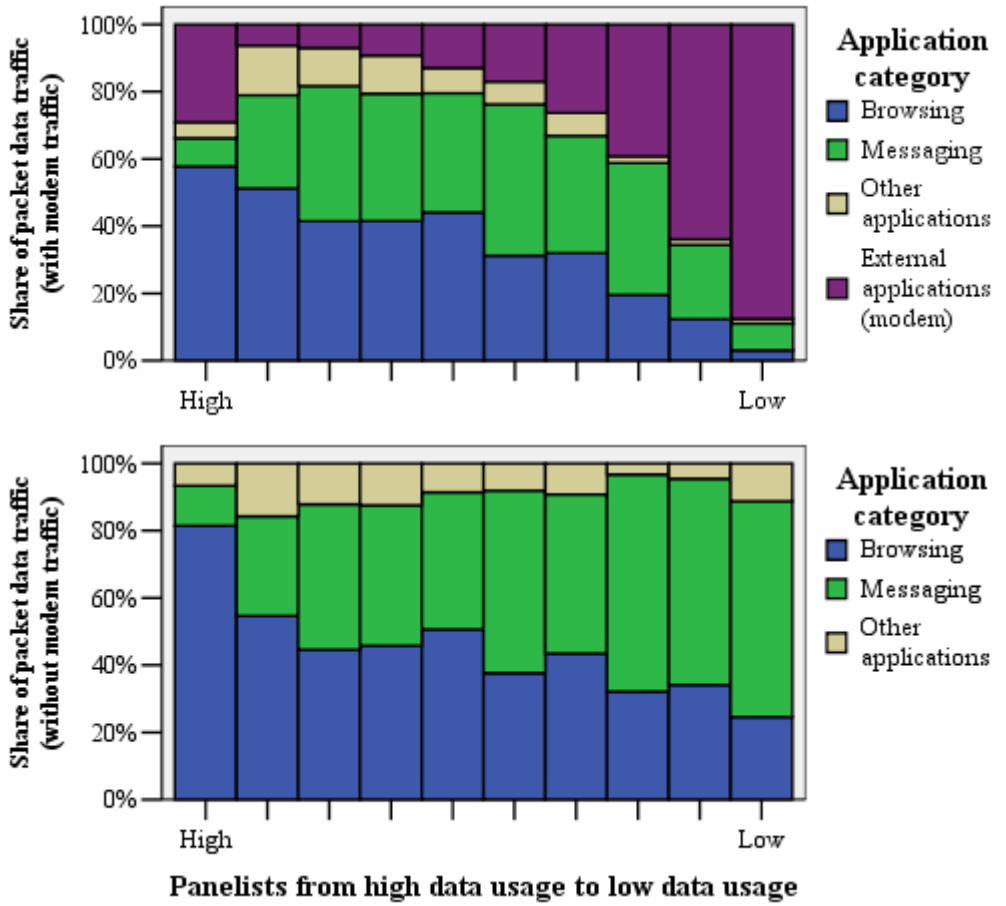
Packet data application usage volumes



- Browsing usage highest for WCDMA handset users when modem traffic is excluded
  - Distribution of traffic between applications is fairly similar for EDGE and GSM/GPRS capable handset
- Browsing generates 70% of packet data traffic, while its share of packet data generating applications' active time was only 13%
  - Browsing sessions active and intensive
  - Volume-intensive non-text content

Packet data generating application usage

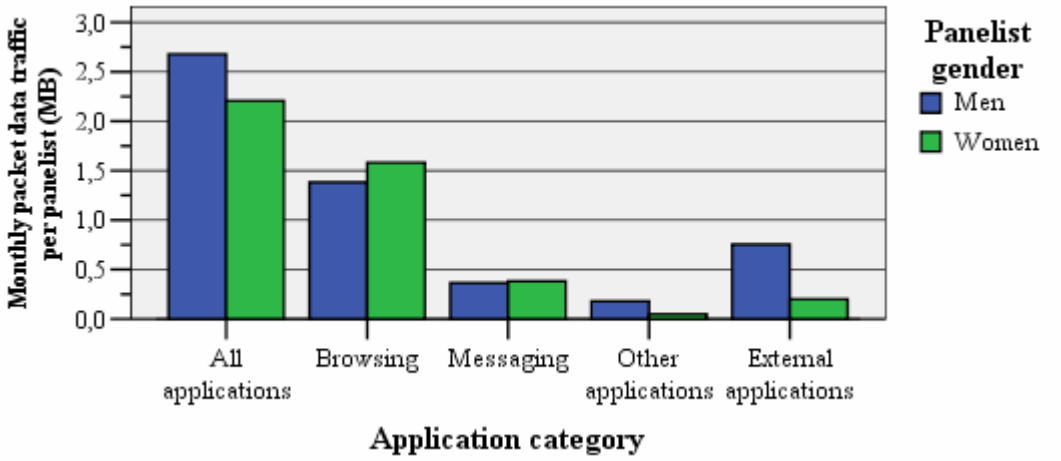
Packet data application traffic by panelist activity

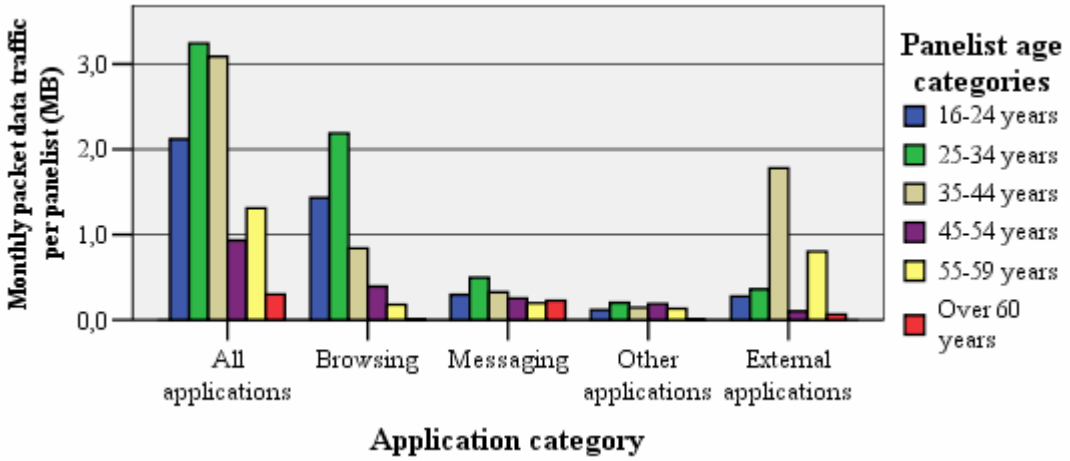


- Share of browsing increases as data usage volume increases
- Share of messaging grows as data usage volume decreases
- Other remarks
  - Share of heterogeneous other applications at fairly constant level among all user groups
  - External applications / modem
    - High share for heavy users probably true modem usage
    - Very high share for low volume usage groups might results from bad/erroneous usage event logs

Packet data generating application usage

Packet data application traffic by gender and age





- Men are slightly more active data users than women
  - Women's more active browsing explained by small sample size and outliers in data
  - No significant differences in messaging
  - Other data applications more used by men
  - Men use external applications (modem) more
- Young people are more active data users than old people, especially in browsing
  - 25-44 year-olds clearly most active data users
  - Age does not have a major impact on messaging activity
  - There are significantly less panelists in higher age categories
- When present, modem usage is in major role
  - Usage in a small number of sessions by a small number of people → results not generalizable
- A few heavy users distort the results due to small sample size
  - One woman created >70% of women's browsing → high browsing averages
  - One 55-59 year-old created >85% of all 55-59 year-olds' modem traffic in one session → high modem usage averages

Smartphone browsing patterns

Most popular web sites by domain

Rank	Domain name of site	Share of web site visits*
1.	mtv3.fi	9,5 %
2.	suomi24.fi	3,6 %
3.	iltasanomat.fi	2,2 %
4.	google.fi	1,8 %
5.	google.com	1,6 %
6.	weatherproof.fi	1,5 %
7.	nokia.com	1,3 %
8.	yle.fi	1,2 %
9.	doubleclick.net	1,0 %
10.	maf.fi	0,9 %
	Others	75,4%

* Web site accesses by handset default browser or Opera, not transferred data volumes

- Top 10 non-operator browsing destinations by domain name
  - Operator-specific sites are removed due to sensitivity reasons
  - Sites visited by less than 10 panelists removed due to privacy reasons
- Web traffic is not very concentrated to few domains
  - 44% in top 10 destinations
    72% in top 50 destinations
    80% in top 100 destinations
    (when including all sites)

Smartphone browsing patterns

Most popular web sites by category

Rank	Site category	Major sites included*	Share of web site visits	Share of panelists **
1.	Operator site	-	31,8%	68,9%
2.	Information	mtv3.fi, iltasanomat.fi, weatherproof.fi, yle.fi	21,5%	58,3%
3.	Entertainment	suomi24.fi, wamli.net, mbnet.fi, subtv.fi, veikkaus.fi	11,4%	36,5%
4.	Adult content	-	8,7%	12,3%
5.	Web search / portal	google.fi, google.com, motionbridge.com	3,7%	35,4%
6.	Mobile content	maf.fi, buumi.net, inpoc.com, zed.fi, funman.fi	3,2%	28,1%
7.	Advertising	doubleclick.net	1,6%	6,0%
8..	E-commerce	huuto.net	0,8%	6,0%
9.	Banking	nordea.fi, sampo.fi, op.fi	0,8%	11,2%
10.	Messaging	-	0,5%	6,5%
	Other	-	9,6%	53,7%
	Hosting / corporate site	nokia.com, opera.com	6,4%	41,4%

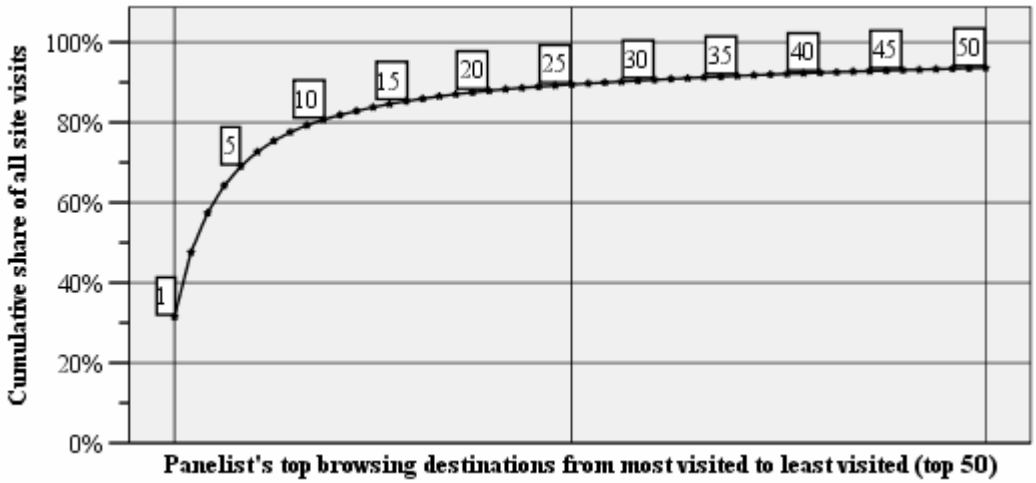
- Subjective domain name based categorization of web/wap sites
  - Overlapping categories
    - Information and entertainment
    - Web search and messaging
    - Operator site, mobile content, and entertainment
- Operator sites actively accessed
  - Handset embedded bookmarks?
  - Includes lots of ”mobile content”?
- Share of infotainment significant, adult content as well
  - Panelists dominantly consumers
- Banking, e-commerce and messaging usage is marginal
  - Using these services with the handset browser is still too awkward?

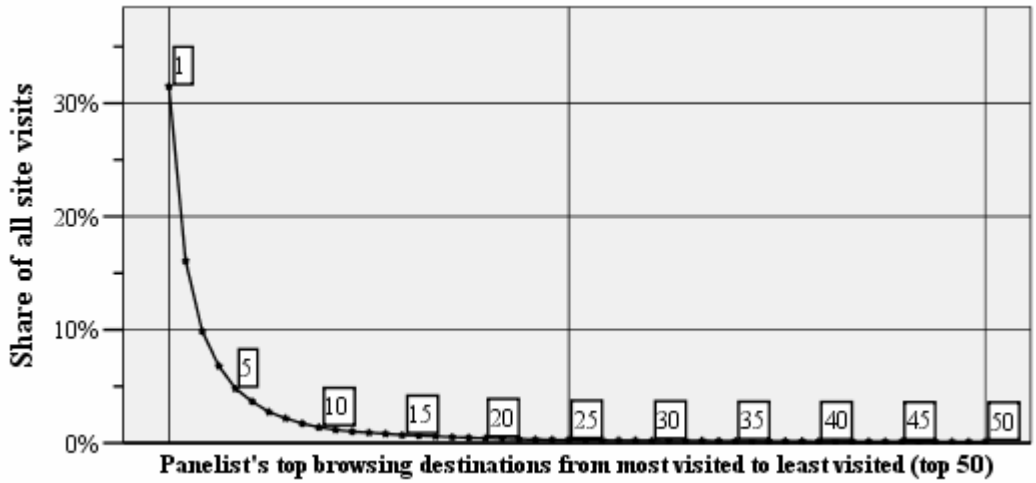
* Sites visited by at least 10 panelists, and with at least 5% of all site visits of the category

** Share of browsing panelists having visited sites of the category during the panel

Smartphone browsing patterns

Individual level browsing patterns





- On average, a panelist made 41 site visits to 9 different domains during the panel
- Individual browsing is concentrated into few domains
  - 69% of visits to top 5
  - 81% of visits to top 10
  - 90% of visits to top 25
- When studying an individual panelist, browsing is even more concentrated

Summary

- Usage of 500 Finnish Symbian / S60 users measured with SP360 research platform monitoring software installed on the handset in fall 2005

- Handset radio capability drives packet data usage frequency and volume (GSM/GPRS → EDGE → WCDMA)
- Data usage volumes are higher for users with larger fixed fee packet data tariffs
- Operator sites 32% and infotainment 33% of web/wap site visits, individual browsing concentrated into few sites
- Using handset as a modem forms a significant 21-25% part of all packet data traffic volume
- 20% of users create 80% of traffic, even when modem usage is excluded
- Browsing most important data application area with 72% of non-modem traffic. Relative share increases with usage volume

HELSINKI UNIVERSITY OF TECHNOLOGY
Networking Laboratory

Conclusions

Mobile Internet Usage Measurement

Conclusions

Major findings

- Operator reporting system –based measurements (80-90% of all Finnish mobile terminals/subscribers)
  - Terminal base old, key features for packet data usage not widely spread (packet data 48%, EDGE 13%, smartphones 6%, WCDMA 0,5%)
  - Nokia's terminal market share 87%. Over 99% of smartphones Symbian based, 1/3 of which Nokia communicators
  - Mobile terminal installed base concentrated, 88% of all terminals among top 50 models
  - 99% of consumer subscribers on usage-based packet data tariff plan, creating 18% of consumer packet data traffic
  - 92-94% of mobile subscribers postpaid, 75% out of which consumers. Business subscribers create 62% of packet data traffic
- TCP/IP header collection –based measurements (50% of Finnish mobile network packet data traffic on one week)
  - Windows originates 65% of traffic in mobile networks → mobile usage profile hidden by Windows traffic
  - VPN usage creates 46% of traffic → very high 85% share of UDP compared to that of fixed networks
  - 90% of all packet data traffic (all APNs) goes via the Internet APN
  - Web also a major application with 14% of traffic volume, 25% of non-VPN traffic volume
- Handset –based measurements (≈500 Finnish Symbian / S60 users)
  - Handset radio capability drives packet data usage frequency and volume (GSM/GPRS → EDGE → WCDMA)
  - Data usage volumes are higher for users with larger fixed fee packet data tariffs
  - Operator sites 32% and infotainment 33% of web/wap site visits, individual browsing concentrated into few sites
  - Using handset as a modem forms a significant 21-25% part of all packet data traffic volume
  - 20% of users create 80% of traffic, even when modem usage is excluded
  - Browsing most important data application area with 72% of non-modem traffic. Relative share increases with usage volume

Conclusions

Observations on measurement process

- Operator reporting system –based measurements
  - \+ Large scope, representative sample (data on national level)
  - – Comparability of data is a problem, depending on usage of custom/existing reports
  - – Data collection is resource intensive for operators, depending on usage of custom/existing reports
  - \+ Comparability issues could be largely fixed with more customized reports
  - \+ Collected raw data enables lots of additional analyses
- TCP/IP header collection –based measurements
  - \+ Large scope, representative sample (data on national level)
  - \+ Good comparability of data, provided that point of measurement is uniform
  - \+ Measurement setup technically straightforward, standard open source tools can be used
  - – Unidentified operating systems (30% of traffic) a problem, although fingerprinting accuracy improves in time
  - – Identification of application protocols based on TCP/UDP server port numbers is not entirely accurate
  - – Almost half of usage masked by VPN usage, although situation should improve as mobile/consumer data usage increases
  - \+ Similarly collected data enables more technical/traditional analyses (e.g. to uncover application protocols more reliably)
- Handset –based measurements
  - \+ Usage data at individual level not attainable otherwise
  - \+ Possibility to link background variables (demographics, subscription type…) to usage
  - – Scope limited to a small sample of smartphone users, although larger samples are possible in the future
  - – Sample size too small when panelists are divided into groups using several background variables
  - – Panelist recruiting phase complicated as this was the first multi-operator measurement. This should be easier in the future
  - – Lots of caution required due to the sensitive nature of almost all measurement phases
  - \+ Collected data enables a multitude of different analyses, presented descriptive results only scratching the surface

Further information

- TKK Networking laboratory master's thesis
  "Mobile Internet Usage Measurements – Case Finland"
  (published in April 2006)
  - http://www.netlab.tkk.fi/~jakivi/publications/Kivi_Thesis_Final.pdf

- Contact antero.kivi(at)tkk.fi