FIDIS

Future of Identity in the Information Society

Title:	D13.12: Privacy – Self-valuation of communication traffic data
Author:	WP13
Editors:	Marek Kumpošt (MU, Czech Republic)
	Vashek Matyáš (MU, Czech Republic)
Reviewers:	Stefan Berthold (TU Dresden, Germany)
	Jozef Vyckoč (VaF Bratislava, Slovakia)
Identifier:	D13.12
Type:	[Deliverable-Experiment results]
Version:	1.0
Date:	Thursday, 23 April 2009
Status:	[Final]
Class:	[Public]
File:	fidis-wp13-del13.12.doc

Brief Summary

Description of an experiment conducted within FIDIS, targeting the value people attach to their private information. In this experiment the goal was to evaluate the value people attach to the information describing the use of online communication tools. We describe the experiment and the structure of our online questionnaire and provide a detailed evaluation of received answers.

Copyright Notice:

Members of the FIDIS consortium

1.	*Goethe University Frankfurt*	Germany
2.	*Joint Research Centre (JRC)*	Spain
3.	*Vrije Universiteit Brussel*	Belgium
4.	*Unabhängiges Landeszentrum für Datenschutz*	Germany
5.	*Institut Europeen D'Administration Des Affaires (INSEAD)*	France
6.	*University of Reading*	United Kingdom
7.	*Katholieke Universiteit Leuven*	Belgium
8.	*Tilburg University*	Netherlands
9.	*Karlstads University*	Sweden
10.	*Technische Universität Berlin*	Germany
11.	*Technische Universität Dresden*	Germany
12.	*Albert-Ludwig-University Freiburg*	Germany
13.	*Masarykova universita v Brne*	Czech Republic
14.	*VaF Bratislava*	Slovakia
15.	*London School of Economics and Political Science*	United Kingdom
16.	*Budapest University of Technology and Economics (ISTRI)*	Hungary
17.	*IBM Research GmbH*	Switzerland
18.	*Centre Technique de la Gendarmerie Nationale*	France
19.	*Netherlands Forensic Institute*	Netherlands
20.	*Virtual Identity and Privacy Research Center*	Switzerland
21.	*Europäisches Microsoft Innovations Center GmbH*	Germany
22.	*Institute of Communication and Computer Systems (ICCS)*	Greece
23.	*AXSionics AG*	Switzerland
24.	*SIRRIX AG Security Technologies*	Germany

Versions

Version	Date	Description (Editor)
0.1	07.2.2009	• Results processed, basic figures
0.5	22.2.2009	• First draft, with graphs and some analyses
0.6	9.3.2009	• Second draft, after minor revisions
0.7	13.3.2009	• Third draft for internal review, minor changes, added conclusions
0.99	17.4.2009	• Almost final version for last internal check
1.0	23.4.2009	• Final version

Foreword

FIDIS partners have contributed in this experiment with translations to national languages and active advertisements of the experiment details during the process of collecting answers. Researchers from the following institutions were the main contributors:

- Masarykova univerzita, Brno, Czech Republic

- Technische Universität Dresden, Germany

- Virtual Identity and Privacy Research Center, Switzerland

- VaF Bratislava, Slovakia

- KU Leuven, Belgium

Management Summary

This document describes the survey that was done at the end of 2008 with the primary goal to evaluate the value people attach to their private information. The study was done in a very similar fashion as the survey from the spring 2006 (D13.5). In the new survey the assessed value is the information about the usage of online communication tools like email or instant messaging. Our participants were asked how much money they will require for being tracked by a special, on purpose developed and certified software. In this deliverable, we provide the details about the experiment set-up as well as the structure of the questionnaire and a detailed description of the results.

Introduction

The price of privacy is a notion used quite often, but there are always doubts about the correctness of the price itself. Privacy is a very complex notion and as such it is very difficult to evaluate it in the whole complexity. We therefore took an approach of defining one rather simple facet of privacy and came out with the following study.

When looking for a suitable privacy-assessed value we settled on the information about the use of online communication tools like emails or instant messaging.

The main goal of the study was to find out how people valuate information about their usage of online communication tools. We have realised that this sort of study would be very interesting for the FIDIS project we are part of. This fact allowed us to organise another experiment quite similar to the one organized in spring 2006. The results of this study were described in D13.5 add-on.

Design of the study

It is very hard to find a way that would allow to obtain realistic data about the privacy value. When you let people say a number (in Euros, Pounds, or Crowns) they will quite probably give you a number that is relatively high. Firstly, they know what you are interested in and naturally they do not want to look like someone ignorant about her/his privacy. Secondly, there is no motivation to keep the delivered number low – it is very hard to say – "Hey friend, it is too much what about just half of what you said?" – without introducing another bias.

We have expected that the approach (the similar approach as in our previous experiment targeting the price of location privacy) – auction for participation in an experiment related to a quality of mobile phone networks – could eliminate both problems mentioned above and allow us to obtain quite realistic data.

We have therefore also decided to use a small deception – claim that we are interested in research of usage of online communication tools – namely the email and instant messaging systems. We believed that we would get reasonable data. On the other hand, we were also very concerned with ethical issues of this approach. We have consulted the issue with people with considerable experience with experiments in the social sciences. To make it short, we were told that it is quite usual to use deceptions if these are explained afterwards.

The whole study was thus hidden under a cover story stating that it is going to be a sociological study about the use of online communication. A similar introductory e-mail was sent out to university students in four countries (translated to local languages and with proper name of local partners of the study).

Our potential participants were mainly from the academic environment since we asked our partners to spread the information about the experiment within their institutions. We may also expect some non-academic participant, but there was no clear question in the web questionnaire asking for academic/non-academic status of the participant.

The questionnaire was online for two weeks and we noticed the majority of responses during the first week. We think that there will not be any considerable increase of answers if the questionnaire would have been online for one or two more weeks.

Dear reader,

Masaryk University participates in a European-wide study organised within the FIDIS (Future of Identity in the Information Society - www.fidis.net) project. This study involves gathering traffic data for a number of volunteers over a period of 30 days.

We are looking for people who will be tracked for the purpose of a sociological study about the use of online communication.

Teams from the FIDIS project in each country (mainly research and academic environment) will provide (or install if you want) a special software that will be used to collect desired data. This software will be provided to all participants in both source code and executables, and its functionality is verified by an external auditor. Data will be collected periodically and there will be regular transfers of observed traffic to our collection servers. We will provide all participants with a removal tool (or members of national teams will remove the software manually) once the experiment is over.

Each participant in the study will receive a monetary compensation, and we are running an auction to select those who will take part. We invite you to submit a bid for the amount of money you require to take part in such a study. As our budget is fixed and limited, successful bidders will be those who bid the lowest amounts, and each will be paid the amount of compensation demanded by the lowest unsuccessful bidder.

Please visit the link http://www.buslab.org/fidis/ regardless of your intent to take (or not) part in this study.

Best regards,

<name of sender>

The link in the text pointed out readers to a homepage of the study with another introductory text.

Design of the web questionnaire

The structure of the web questionnaire was as follows. On the first page, the users were able to choose the language of the questionnaire. Available languages were Czech, English, Slovak, German, or Flemish.

The next page contained the text explaining the study (similar description as in the introductory email):

> *This is the homepage of a European-wide study organized within the FIDIS (Future of Identity in the Information Society - http://www.fidis.net) NoE (Network of Excellence). This study involves gathering traffic data for a number of volunteers over a period of 30 days.*
>
> *We are looking for people who will be tracked for the purpose of a sociological study about the use of online communication. The study is somewhat similar to what have been done by e.g. researchers at the Northeastern University (link).*
>
> *Teams from the FIDIS project in each country (mainly research and academic environment) will provide (or install if you want) a special software that will be used to collect desired data. This software will be provided to all participants in both source code and executables, and its functionality is verified by an external auditor. Data will be collected periodically and there will be regular transfers of observed traffic to our collection servers. We will provide all participants with a removal tool (or members of national teams will remove the software manually) once the experiment is over.*
>
> *Each participant in the study will receive a monetary compensation, and we are running an auction to select those who will take part. We invite you to submit a bid for the amount of money you require to take part in such a study. As our budget is fixed and limited, successful bidders will be those who bid the lowest amounts, and each will be paid the amount of compensation demanded by the lowest unsuccessful bidder.*

> *We have deployed several mechanisms to detect cheating and we reserve the right not to pay the participant if any kind of cheating is detected.*
>
> *You are giving consent for this kind of observation by participating in our study in accordance with your national law on data protection.*
>
> *Some additional information is in our FAQ.*

This introductory text was followed by a radio box (Do you want to take part in this study?) of three choices:

1) Yes, with a PC only
2) Yes, with any mobile device(s)
3) Yes, with both PC and mobile device(s)
4) No, I do not want to participate

If a visitor selected any of the first three options the first set of questions was displayed. These were designed to support the goals of the study but also to obtain data we were interested in.

1) Your age?
   a. <18; 18-24; 25-34; 35-44; 45-54; 55-64; 65-74; >75
2) Gender?
   a. Male
   b. Female
3) Is the type of device (you will use in the study) in your possession or do you share it with someone else?
   a. The device is under my full control
   b. The device is shared with another person(s)
4) We are interested in the level of your IT-knowledge. Do you consider yourself having:
   a. basic IT-knowledge
   b. medium IT-knowledge
   c. advanced IT-knowledge
   d. professional in IT
5) How much money do you want for being tracked for two weeks (please put only a number in EUR and specify only those scenarios you are interested in)?
   a. For email traffic data (no message body) (EUR):
   b. For instant messaging traffic data (no message body) (EUR):
   c. For all traffic data (EUR):

The observed information was further explained in our FAQ in order to explain the situation to those participants who are not aware about such details.

After the first set of questions, two special scenarios for the experiment extension were introduced. Participants had the possibility to opt out in any of these two scenarios. We also asked the participant to fill in their email address so that we can contact them in case they will be selected for participation in our experiment.

1) We will consider to process the data within the FIDIS research consortium and provide some answers to queries or summaries in an aggregated form to a commercial subject who entered in a contractual relation with us. (No direct access to the data). What amount of money (in EUR) would you request for your participation in such case?

2) This is a hypothetical issue: As you might be aware, our data collection might be of some use for system training and improvement in detecting terrorist activities. Imagine that we could provide the data to your national government (only) to let them improve their terrorist activity detection and tracking tools. Although we do not plan to do this, how much money (in EUR) would you ask for participating under these circumstances?

Option 4 in the introductory part was for those visitors who did not want to participate in our study. After selecting this option the visitors were asked for reasons why they did not participated in the experiment. The possible options were:

1) I do not have appropriate hardware equipment
2) I do not have time
3) I do not see the value of such study
4) Such study is not ethical
5) I do not trust your intentions with this study -- but I would trust another institution, namely:
6) Other reasons:

The web forms were up for two weeks.

Results

There were 1080 people who saw the introductory text. They may have closed the application after they read the text or decide whether to participate or not. These visitors were given the choice from five language versions: Flemish, Czech Republic, Germany, Slovak Republic and others who choose the English version of the questionnaire. The numbers in the following lists reflect the chosen language version.

- Flemish (31 – 2.8 %)
- Czech (362 – 33.5 %)
- German (71 – 6.6 %)
- Slovak (470 – 43.5 %)
- English (146 – 13.5 %)
- Total sum: 1080

The numbers of visitors who were willing to participate (regardless of the device selection) in the experiment are (the numbers reflect the chosen language version):

- Flemish (13 – 3 %)
- Czech (174 – 40,7 %)

- German (30 – 7 %)
- Slovak (136 – 31,8 %)
- English (75 – 17,5 %)
- Total sum: 428 (39 % of those who visited the intro text)

Visitors who explicitly did not want to participate were (we will evaluate their why-did-you-not-participate answers later) as follows (the numbers reflect the chosen language version):

- Flemish (2 – 3.5 %)
- Czech (19 – 33.3 %)
- German (9 – 15.8 %)
- Slovak (21 – 36.8 %)
- English (6 – 10.5 %)
- Total sum: 57

Visitors willing to participate. The first value refers to the PC-based participation only, the second refers to mobile devices only and the third refers to all devices (the numbers reflect the chosen language version):

- Flemish (7;2;4)
- Czech (114; 16; 44)
- German (21;2;7)
- Slovak (106;12;18)
- English (42;9;24)

The next list shows portions of visitors who were willing to participate and finished at least the first tracking scenario – male (first number) vs. female (second number). This information was acquired in the second form, so there have already been some drop outs between the first question (will you participate) and second set of questions (where we also asked for a gender). The numbers in the following list reflect the chosen language version.

- Flemish (6/1)
- Czech (109/20)
- German (17/1)
- Slovak (78/7)
- English (33/12)

Results from the second form (question about age, gender, IT-knowledge, device possession).

- Age under 18 (3)
- Age 18-24 (243)
- Age 25-34 (29)
- Age 35-44 (9)

- Basic IT knowledge (15)
- Medium IT knowledge (80)
- Advanced IT knowledge (137)
- Professional in IT (52)

- Own hardware (254)
- Shared hardware (30)

In the following list, we provide numbers of visitors that filled at least *n* scenarios (1st scenario – academic use of data, 2nd scenario – results available to commercial subjects, 3rd scenario – data available to national governments). The numbers in the following list reflect the chosen language version.

- Flemish (7;5;5)
- Czech (129;117;116)
- German (18;11;10)
- Slovak (85;56;56)
- English (45;36;31)

Number of participates who saw the introductory part, actually filled the questionnaire and number of those who answered the 'will participate' part and filled the other forms.

	will participate	1st form	2nd form	3rd form
Percentage w.r.t. intro text numbers	46.1%	26.3%	20.8%	20.7%
Percentage w.r.t. participants	100.0%	57.0%	45.2%	44.9%

We can see that only a quarter of visitor who saw the introductory text actually filled at least the first tracking scenario. Such decrease can be caused by two facts: 1) visitors were really sensitive to this kind of private information or 2) extensive similarity with our previous experiment (we noticed several why-did-you-not-participate reasons explicitly stating this fact).

First scenario – academic use of data

	First bids			First bids – male			First bids - female		
1st quartile	10	10	12	10	9.5	12	10	10	15
2nd quartile	30	30	50	32.5	25	50	30	35	50
3rd quartile	100	100	200	100	100	200	275	150	300

Monetary compensation (all values are in EUR) for the first scenario – email tracking, online messaging tracking and all traffic data for the academic purposes – and numbers for males/females.

We decided to use quartiles for calculating bids instead of minimal, maximal and average values. The reason is the existence of extreme values that will negatively influence average values. For instance for the first bid such extreme monetary compensation was EUR 1000000.

Regarding the number of visitors, who filled the first scenario bids, we can say that there is no considerable difference between the bids placed by males and females.

	First bids – own HW			First bids – shared hw		
1st quartile	10	9.25	15	50	35	8.75
2nd quartile	20	25	40	100	100	50
3rd quartile	100	99.75	200	500	175	262.5

The above table shows differences between those who selected "own hardware" option and "shared hardware" option.

Explicit 'No participation in the second scenario' – 23 participants (9,871%) – these were the participants who decided not to participate in case of the commercial use of data.

Second scenario – commercial use of data

	First bids			Second bids		
1st quartile	10	8.25	10.425	10	10	15
2nd quartile	20	22.5	40	40	40	50
3rd quartile	100	80	150	100	100	200

Evaluation of bids of those who filled at least the first and the second scenarios (the second scenario was that we will provide the data to a commercial subject who we have a contract with). We also observed the first scenario bids of those visitors who decided to participate in the second scenario of our experiment.

Second bids - male			Second bids – female		
10	10	15	10	10	10
40	40	50	40	35	40
100	100	200	150	70	127

Male and female bids in the second scenario. In this case, we can say that males tend to place higher bids. But again the number of female participants was too low for this result to be considered conclusive.

Explicit 'No participation in the third scenario' – 41 participants (18 %). These participants explicitly stated that they do not wish to participate in case of the data being provided to national governments.

Third scenario – data will be provided to national governments

	First bids			Second bids			Third bids		
1st quartile	8.75	8.5	11.1375	10	10	15	10	10	15

2nd quartile	20	25	40	40	40	50	50	50	60
3rd quartile	100	80	150	100	100	200	200	200	400

Male vs. female bids in the third scenario. As in the second case, we can say that males tend to place higher bids. It would be nice to have more answers from both groups of participant for being conclusive about this observation.
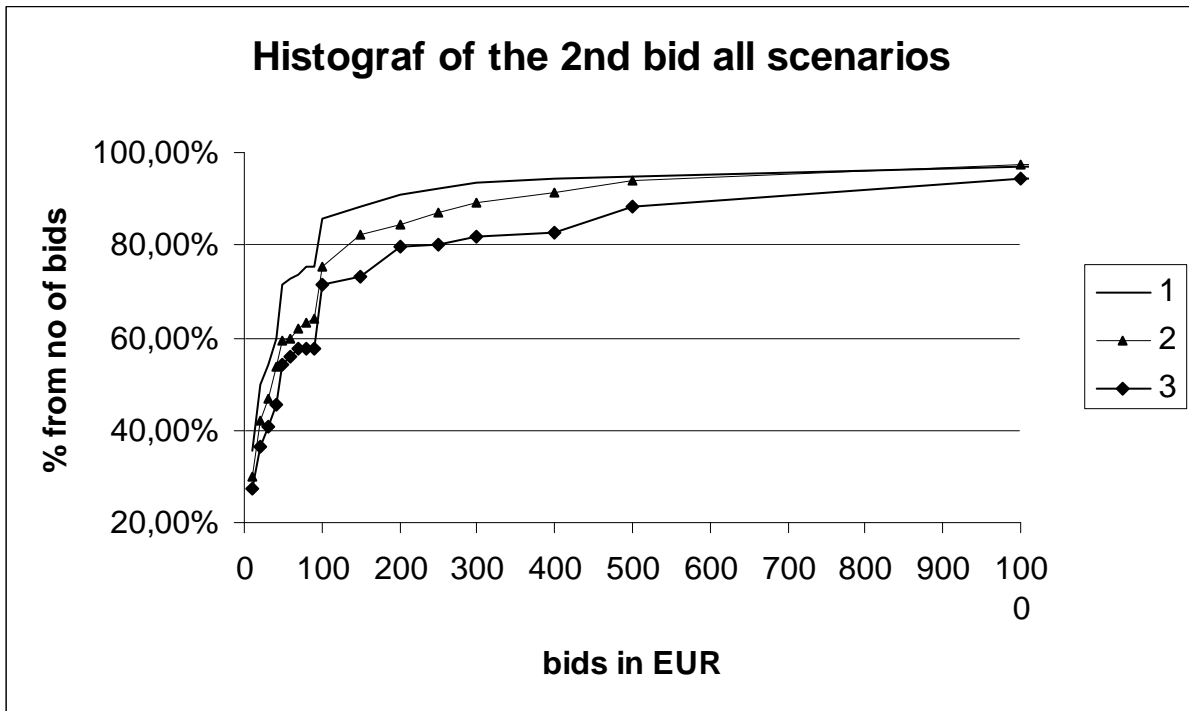
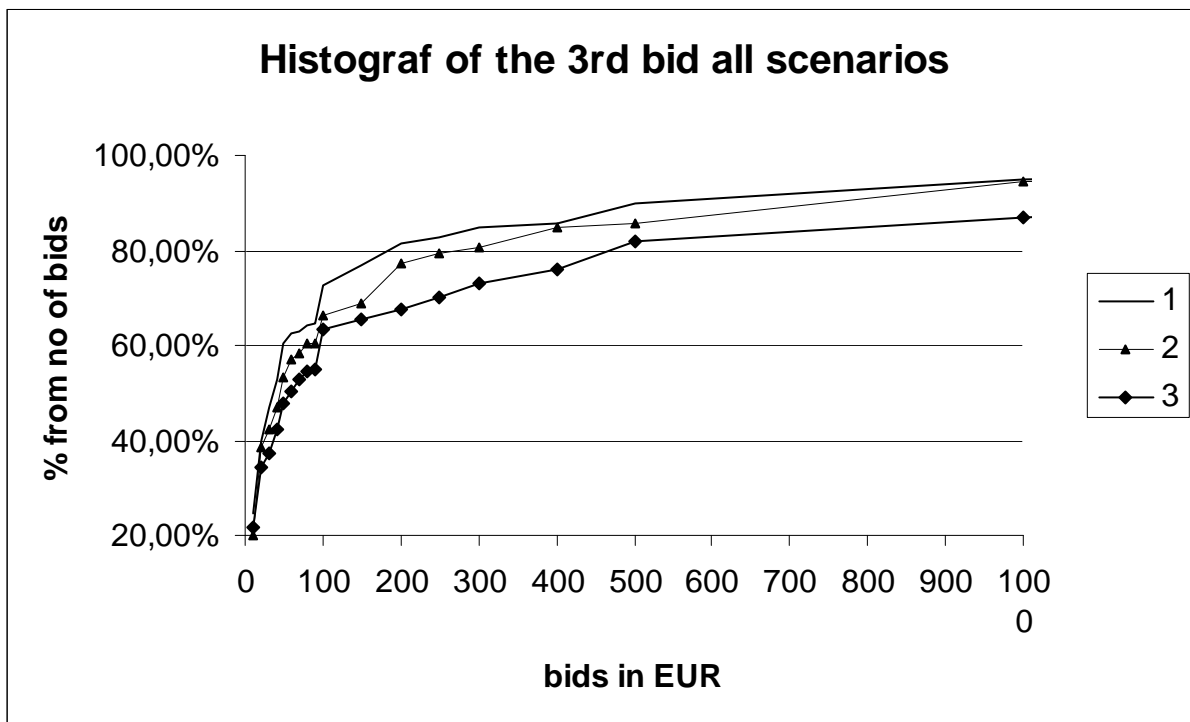| Third bids - male			Third bids - female		
10	10	17.5	10	10	10
50	50	80	40	32	40
200	200	450	200	186	150

Histograms of bids

In this section we provide histograms of bids in all scenarios considered in our experiment. Histograms are quite helpful in observing trends of bids and allow for quick comparisons. We decided not to split all bids according to national variants since those sets are very small for providing conclusive results.

Histogram of first bids (email tracking) in all three scenarios (Academic use of data: neither the data nor the results will be revealed to any other subjects – line 1; Commercial use of data: the data will be provided to a commercial subject we have a contract with – line 2; Data provided to national governments: data will be provided to national governments in order to improve the mechanisms for terrorists detection – line 3) for those visitors, who provided answers in all three scenarios, i.e., visitors who did not explicitly stated the I-do-not-want-to-participate option. We can see that the value of bids is increasing as the use of acquired data changes from academic to national government. But even with this trend, we were expecting higher differences.
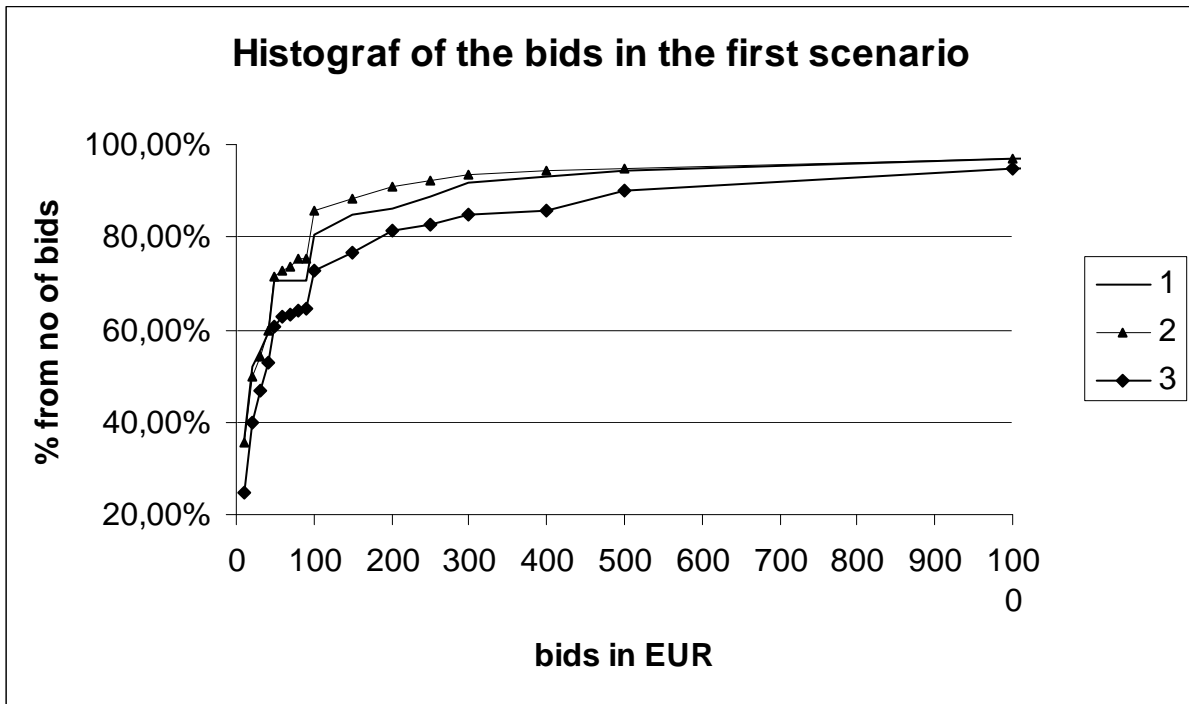
Histograf of the 2nd bid all scenarios

Histogram of second bids (online messaging tracking) in all three scenarios (Academic use of data: neither the data nor the results will be revealed to any other subjects – line 1; Commercial use of data: the data will be provided to a commercial subject we have a contract with – line 2; Data provided to national governments: data will be provided to national governments in order to improve the mechanisms for terrorists detection – line 3) for those visitors, who provided answers in all three scenarios, i.e., visitors who did not explicitly stated the I-do-not-want-to-participate option. The situation with required monetary compensation follows the same trends as in the previous case.

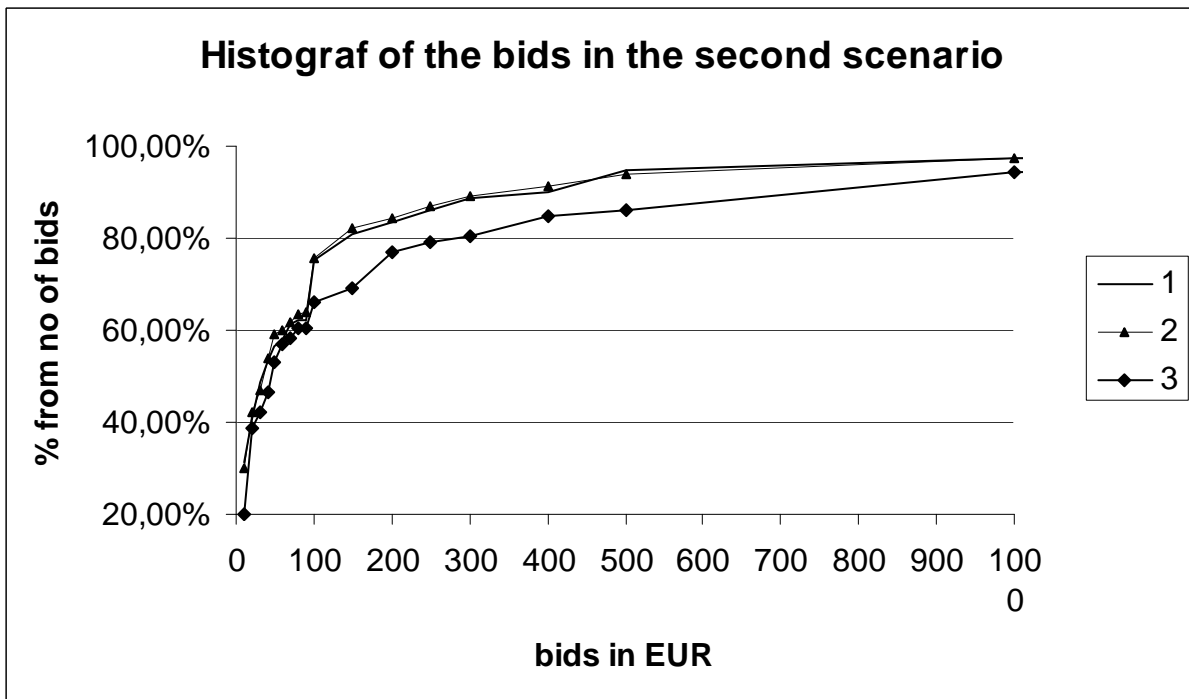## Histograf of the 3rd bid all scenarios

bids in EUR

Histogram of third bids (all tracking data) in all three scenarios (Academic use of data: neither the data nor the results will be revealed to any other subjects – line 1; Commercial use of data: the data will be provided to a commercial subject we have a contract with – line 2; Data provided to national governments: data will be provided to national governments in order to improve the mechanisms for terrorists detection – line 3) for those visitors, who provided answers in all three scenarios, i.e., visitors who did not explicitly stated the I-do-not-want-to-participate option. The situation with required monetary compensation remains the same as in the previous case. Histograms in this case show higher monetary compensation required by the visitors, but the observed value in this case is significantly different when compared to (email tracking or online messaging tracking).

It seems that majority of visitors do not differentiate between the types of tracking data as the differences at the beginnings of all histograms are very small. The following graph depicts the situation in the first scenario (data will be used for academic purposes only) for all three possible tracking situations – line 1 is email tracking; line 2 – online messaging; line 3 all tracking data. It can also be observed that email tracking data are more valuable for our participants.

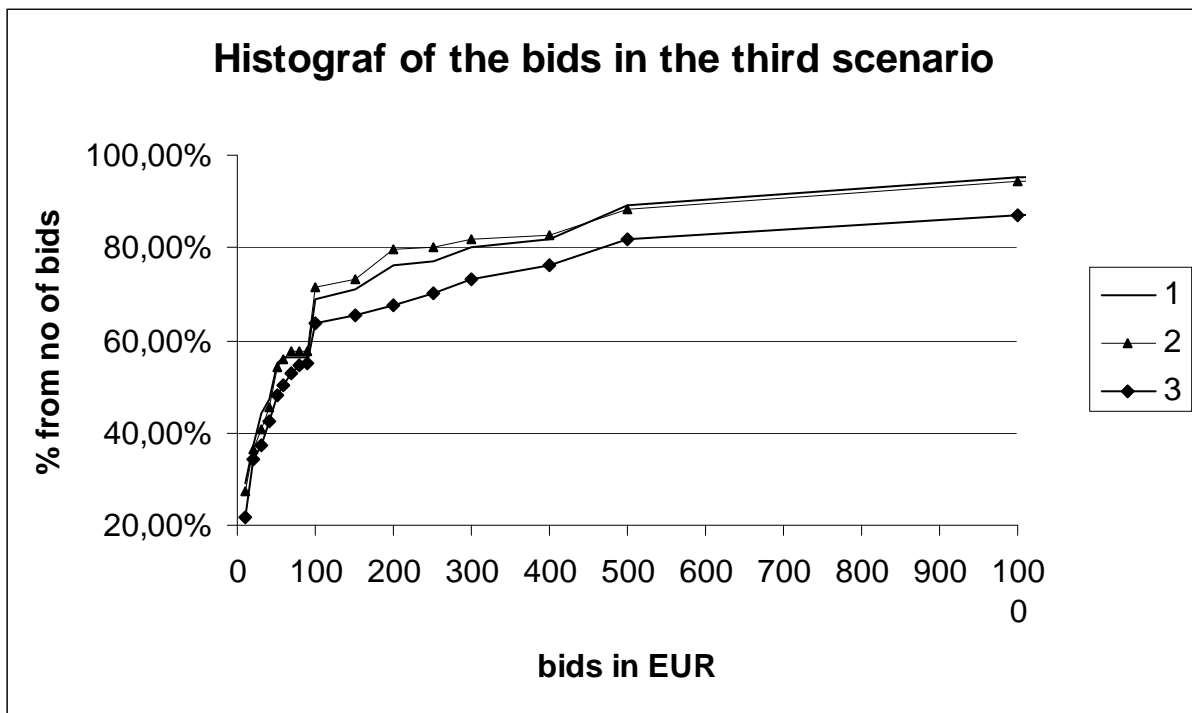**Histograf of the bids in the first scenario**

In case of the commercial use of captured data, the difference between email and online messaging is almost insignificant. Value of all traffic data is then higher as we were expecting.

Histograf of the bids in the second scenario

The last scenario – possible use of captured data by national governments. This histogram shows that people, who will eventually participate, will require more money when compared

to the other scenarios. Again the value of email tracking data and online messaging tracking data is almost the same, while all tracking data is "more expensive".

Histograf of the bids in the third scenario

Highlights of the study results

Let us summarize the most important results of our experiment. We received almost 300 responses for at least the first scenario (academic use of data) from more than four countries (in terms of selected language variant of the web questionnaire). The number is lower than expected but still allows for a solid analysis. The primary goal of our survey was to evaluate the monetary compensation that the visitors require for participating in the experiment. The goal of the experiment was to deploy a special piece of software on participants' hardware that will observe the use of online communication tools. We asked the participants for the monetary compensation they require for taking part in the experiment. We evaluated the results and provided several tables and graphs for better understanding.

Second quartiles in the first scenario may be considered as the main result. The monetary compensation for being observed for email traffic data is EUR 30 and the same price is for instant messaging. All tracking data is "more expensive", namely EUR 50. We provided a comparison between male vs. female bidding and the results showed no considerable differences between these two groups of participants.

We also compared bids of those answering more than the first scenario (the other scenarios were: data will be provided to some commercial subject or to the national governments) and the result confirmed our expectations. The required value for participating has increasing tendency as the scenario changed from academic use of data to commercial use of data and finally to the scenario where we will hypothetically provide the data to national governments.

From the histograms presented in the last part of the text we can observe that in some cases instant messaging tracking data is priced higher than email tracking data and vice versa. All traffic data is always the most expensive case.