Documentation for the Dataset on Tor Users

by: David Harborth* and Sebastian Pape
Chair of Mobile Business & Multilateral Security
Goethe University Frankfurt am Main
*david.harborth@m-chair.de

1 Introduction and Important Information

This dataset was collected for research conducted within the project AN.ON-Next funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371.

The following papers are based fully or partially on this dataset:

1. Harborth, D., and Pape, S. (2020). HOW PRIVACY CONCERNS, TRUST AND RISK BELIEFS AND PRIVACY LITERACY INFLUENCE USERS' INTENTIONS TO USE PRIVACY-ENHANCING TECHNOLOGIES - THE CASE OF TOR. *ACM SIGMIS The DATA BASE for Advances in Information Systems*, (forthcoming).

2. Harborth, D., and Pape, S. (2020). Explaining Technology Use Behaviors of Privacy-Enhancing Technologies: The Case of Tor and JonDonym. In *Proceedings on Privacy Enhancing Technologies (PETS)*.

3. Harborth, D., Cai, X., and Pape, S. (2019). Why Do People Pay for Privacy-Enhancing Technologies? The Case of Tor and JonDonym. In G. Dhillon, F. Karlsson, K. Hedström, and A. Zuquete (Eds.), *ICT Systems Security and Privacy Protection. SEC 2019. IFIP Advances in Information and Communication Technology*, vol 562 (pp. 253–267). Springer, Cham. https://doi.org/10.1007/978-3-030-22312-0_18.

4. Harborth, D., and Pape, S. (2019). How Privacy Concerns and Trust and Risk Beliefs Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor. In *Hawaii International Conference on System Sciences (HICSS) Proceedings* (pp. 4851–4860). Hawaii, US.

The dataset includes – among others – constructs from different established models of the literature like the technology acceptance model (TAM) by Davis (1985) and the Internet Users Information Privacy Concerns (IUIPC) model by Malhotra et al. (2004). Furthermore, there are extensive questions on privacy literacy covered by the online privacy literacy scale (OPLIS) by Masur et al. (2017). See Table 1 for the complete list of questions in the dataset.

Further relevant information:

1. For OPLIS, it is important to note that five questions of the original survey were excluded since they deal with European and German data protection law. These questions are difficult to answer and may not provide any insight about the privacy literacy of JonDonym users who are not necessarily only coming from Europe and Germany (e.g. from the US). Thus, our dataset only contains 15 instead of

20 OPLIS questions. The questions with the abbreviation OP1–OP5 cover participants' knowledge about institutional practices. Questions OP6–OP10 cover knowledge about technical aspects of data protection and questions OP11–OP15 cover knowledge about data protection strategies.

2. Values for experience in the dataset are equal to 21, if participants stated to have an experience of more than 20 years (for EXP and TOREXP).

3. Demographics were not mandatory to fill out due to anonymity reasons and the highly privacy-sensitive target population. Thus, the fragmented pieces of data regarding demographic factors are not included.

Please contact David Harborth in case there are any questions regarding the dataset or the documentation.

2 Survey Distribution Channels

We conducted the study with German and English speaking Tor users in order to maximize the sample size. The translation process of the constructs into German and further details on the two versions are described in several previous research articles by the authors (see for example Harborth and Pape (2018a,b, 2019); Harborth et al. (2019); Harborth and Pape (2020b,a)).

The survey was distributed via the following channels:

1. Mailinglists:
   (a) tor-talk[1]
   (b) liberationtech[2]
   (c) IFIP TC 11[3]
   (d) FOSAD[4]
   (e) GI PET[5]
   (f) GI FBSEC[6]
2. Twitter with #tor and #privacy
3. Boards:
   (a) reddit (sub-reddits: r/TOR, r/onions, r/privacy)
   (b) ubuntuusers.de

[1] https://lists.torproject.org/cgi-bin/mailman/listinfo/tor-talk/
[2] https://mailman.stanford.edu/mailman/listinfo/liberationtech
[3] https://dlist.server.uni-frankfurt.de/mailman/listinfo/ifip-tc11
[4] http://www.sti.uniurb.it/events/fosad/
[5] http://mail.gi-fb-sicherheit.de/mailman/listinfo/pet
[6] http://mail.gi-fb-sicherheit.de/mailman/listinfo/fbsec

4. Tor Hidden Service Boards, Sections posted into:
   (a) Darknet Avengers[7], Off Topic
   (b) The Hub[8], Beginners
   (c) Onion Land[9], Off Topic
   (d) 8chan[10], /tech/
   (e) IntelExchange[11], Unverified Users
   (f) Code Green[12], Discussions
   (g) Changolia[13], overchan.random
   (h) Atlayo[14], Posting

5. Personal Announcements at Workshops

[7] http://avengersdutyk3xf.onion/
[8] http://thehub7xbw4dc5r2.onion
[9] http://onionlandbakyt3j.onion
[10] http://oxwugzccvk3dk6tj.onion
[11] http://rrcc5uuudhh4oz3c.onion
[12] http://pyl7a4ccwgpxm6rd.onion
[13] http://jewsdid.oniichanylo2tsi4.onion
[14] http://atlayofke5rqhsma.onion/

3 Questionnaire Composition

Table 1: Constructs in the Dataset (measured on a seven-point Likert scale ranging from "strongly disagree" to "strongly agree", if not otherwise indicated)

Trust in Tor	$Trust_{Tor}1$	Tor is trustworthy.	Pavlou (2003)
	$Trust_{Tor}2$	Tor keeps promises and commitments.	
	$Trust_{Tor}3$	I trust Tor because they keep my best interests in mind.	
Perceived Anonymity	PA1	Tor is able to protect my anonymity in during my online activities.	Benenson et al. (2015)
	PA2	With Tor I obtain a sense of anonymity in my online activities.	
	PA3	Tor can prevent threats to my anonymity when being online.	
Perceived Usefulness of Protecting Users' Privacy	PU1	Using Tor improves the performance of my privacy protection.	Benenson et al. (2015); Venkatesh and Davis (2000)
	PU2	Using Tor increases my level of privacy.	
	PU3	Using Tor enhances the effectiveness of my privacy.	
	PU4	I find Tor to be useful in protecting my privacy.	
Perceived Ease of Use	PEOU1	My interaction with Tor is clear and understandable.	Venkatesh and Davis (2000)
	PEOU2	Interacting with Tor does not require a lot of my mental effort.	
	PEOU3	I find Tor to be easy to use.	
	PEOU4	I find it easy to get Tor to do what I want it to do.	
Behavioral Intention	BI1	I intend to continue using Tor in the future.	Venkatesh and Davis (2000)
	BI2	I will always try to use Tor in my daily life.	
	BI3	I plan to continue to use Tor frequently.	
Actual Use Frequency	USE	Please choose your use frequency of Tor. (10 point frequency scale from "never" to "all the time".	Rosen et al. (2013)
Risk Propensity	RP1	I would rather be safe than sorry.	Donthu and Gilliland (1996)
	RP2	I am cautious in trying new/different products.	

Construct	Abbreviation	Item	Adapted from
	RP3	I avoid risky things.	
Privacy Victim	VIC	How frequently have you personally been the victim of what you felt was an improper invasion of privacy? (7 point likert scale ranging from "never" to "very frequently")	Malhotra et al. (2004)
Trusting Beliefs	TB1	Online companies would be trustworthy in handling (the information).	Malhotra et al. (2004)
	TB2	Online companies would tell the truth and fulfil promises related to (the information) provided by me.	
	TB3	I trust that online companies would keep my best interests in mind when dealing with (the information).	
	TB4	Online companies are in general predictable and consistent regarding the usage of (the information).	
	TB5	Online companies are always honest with customers when it comes to using (the information) that I would provide.	
Risk Beliefs	RB1	In general, it would be risky to give (the information) to online companies.	Malhotra et al. (2004)
	RB2	There would be high potential for loss associated with giving (the information) to online firms.	
	RB3	There would be too much uncertainty associated with giving (the information) to online firms.	
	RB4	Providing online firms with (the information) would involve many unexpected problems.	
	RB5	I would feel safe giving (the information) to online companies. (R)	
Information Privacy Collection	COLL1	It usually bothers me when online companies ask me for personal information.	Malhotra et al. (2004)

Construct	Abbreviation	Item	Adapted from
	COLL2	When online companies ask me for personal information, I sometimes think twice before providing it.	
	COLL3	It bothers me to give personal information to so many online companies.	
	COLL4	I'm concerned that online companies are collecting too much personal information about me.	
Information Privacy Awareness	AWA1	Companies seeking information online should disclose the way the data are collected, processed, and used.	Malhotra et al. (2004)
	AWA2	A good consumer online privacy policy should have a clear and conspicuous disclosure.	
	AWA3	It is very important to me that I am aware and knowledgeable about how my personal information will be used.	
Information Privacy Control	CONTROL1	Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared.	Malhotra et al. (2004)
	CONTROL2	Consumer control of personal information lies at the heart of consumer privacy.	
	CONTROL3	I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.	
Facilitating Conditions	FC1	I have the resources necessary to use Tor.	Venkatesh et al. (2012)
	FC2	I have the knowledge necessary to use Tor.	
	FC3	Tor is compatible with other technologies and applications I use.	
	FC4	I can get help from others when I have difficulties using Tor.	
Trade-off Effort and Use	EFFORTUSE1	Tor offers a good value for my invested effort (time-wise and monetary).	self-made

Construct	Abbreviation	Item	Adapted from
	EFFORTUSE2	Tor offers a good value for my invested time effort.	
	EFFORTUSE3	Tor offers a good value at the current price.	
Result Demonstrability	RESULTDEMON1	I have no difficulty telling others about the results of using Tor.	Venkatesh and Davis (2000)
	RESULTDEMON2	I believe I could communicate to others the consequences of using Tor.	
	RESULTDEMON3	The results of using Tor are apparent to me.	
	RESULTDEMON4	I would have difficulty explaining why using Tor may or may not be beneficial.	
Consumer Independent Judgement Making	CIJM1	Prior to purchasing a new brand, I prefer to consult a friend that has experience with the new brand. (R)	Manning et al. (1995)
	CIJM2	When it comes to deciding whether to purchase a new service, I do not rely on experienced friends or family members for advice.	
	CIJM3	I seldom ask a friend about his or her experiences with a new product before I buy the new product.	
	CIJM4	I decide to buy new products and services without relying on the opinions of friends who have already tried them.	
	CIJM5	When I am interested in purchasing a new service, I do not rely on my friends or close acquaintances that have already used the new service to give me information as to whether I should try it.	
	CIJM6	I do not rely on experienced friends for information about new products prior to making up my mind about whether or not to purchase.	
Consumer Novelty Seeking	CNS1	I often seek out information about new products and brands.	Manning et al. (1995)
	CNS2	I like to go to places where I will be exposed to information about new products and brands.	
	CNS3	I like magazines that introduce new brands.	

Construct	Abbreviation	Item	Adapted from
	CNS4	I frequently look for new products and services.	
	CNS5	I seek out situations in which I will be exposed to new and different sources of product information.	
	CNS6	I am continually seeking new product experiences.	
	CNS7	When I go shopping, I find myself spending very little time checking out new products and brands.	
	CNS8	I take advantage of the first available opportunity to find out about new and different products.	
Online Privacy Literacy Scale	OP1	The National Security Agency (NSA) accesses only public user data, which are visible for anyone. (True/false/don't know)	Masur et al. (2017)
	OP2	Social network site operators (e.g. Facebook) also collect and process information about non-users of the social network site. (True/false/don't know)	
	OP3	User data that are collected by social network site operators (e.g. Facebook) are deleted after five years. (True/false/don't know)	
	OP4	Companies combine users' data traces collected from different websites to create user profiles. (True/false/don't know)	
	OP5	E-mails are commonly passed over several computers before they reach the actual receiver. (True/false/don't know)	
	OP6	1. What does the term "browsing history" stand for? In the browsing history... A. ...the URLs of visited websites are stored. B. ...cookies from visited websites are stored. C. ...potentially infected websites are stored separately. D. ...different information about the user are stored, depending on the browser type.	

Construct	Abbreviation	Item	Adapted from
	OP7	2. What is a “cookie”? A. A text file that enables websites to recognize a user when revisiting. B. A program to disable data collection from online operators. C. A computer virus that can be transferred after connecting to a website. D. A browser plugin that ensures safe online surfing.	
	OP8	3. What does the term “cache” mean? A. A buffer memory that accelerates surfing on the Internet. B. A program that specifically collects information about an Internet user and passes them on to third parties. C. A program, that copies data on an external hard drive to protect against data theft. D. A browser plugin that encrypts data transfer when surfing online.	
	OP9	4. What is a “trojan”? A trojan is a computer program, that... A. ...is disguised as a useful application, but fulfills another function in the background. B. ...protects a computer from viruses and other malware. C. ... was developed for fun an d has no specific function. D. ... caused damage as computer virus in the 90ies but doesn’t exist anymore.	
	OP10	5. What is a “firewall”? A. A fallback system that will protect the computer from unwanted web attacks. B. An outdated protection program against computer viruses. C. A browser plugin that ensures safe online surfing. D. A new technical development that prevents data loss in case of a short circuit.	
	OP11	Tracking of one’s own internet is made more difficult if one deletes browser information (e.g. cookies, cache, browser history) regularly. (True/false/don’t know)	

Construct	Abbreviation	Item	Adapted from
	OP12	Surfing in the private browsing mode can prevent the reconstruction of your surfing behavior, because no browser information is stored. (True/false/don't know)	
	OP13	Using false names or pseudonyms can make it difficult to identify someone on the Internet. (True/false/don't know)	
	OP14	Even though It-experts can crack difficult passwords, it is more sensible to use a combination of letters, numbers and signs as passwords than words, names or simple combinations of numbers. (True/false/don't know)	
	OP15	In order to prevent the access to personal data, one should use various passwords and user names for different online applications and change them frequently. (True/false/don't know)	
Internet Experience	EXP	How many years of experience do you have with computers? (Answer options range from 0 years to "more than 20 years".)	self-made
Experience with Tor	TOREXP	How many years are you using Tor? (Answer options range from 0 years to "more than 20 years".)	self-made
Donation	DON	Did you ever donated money to the Tor project? (y/n)	self-made
Amount of donation	AMOUNT	How much money did you donate to the Tor project?	self-made
Recommendation of Tor	REC	Would you recommend Tor? (y/n)	self-made
Purpose of Tor Use	PUR	For what purposes are you using Tor? PUR1: Surfing the internet; PUR2: E-Mail Service; PUR3: Audio and Videostreaming; PUR4: Filesharing; PUR5: Instant Messaging; PUR6: Cloud Services	self-made

Construct	**Abbreviation**	**Item**	**Adapted from**
Knowledge about JonDonym	JD	Do you know the anonymization service JonDonym?	self-made

References

Benenson, Z., Girard, A., and Krontiris, I. (2015). User Acceptance Factors for Anonymous Credentials: An Empirical Investigation. *14th Annual Workshop on the Economics of Information Security (WEIS)*, pages 1–33.

Davis, F. (1985). A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. *Massachusetts Institute of Technology*.

Donthu, N. and Gilliland, D. (1996). Observations: The infomercial shopper. *Journal of Advertising Research*, 36(April):69–76.

Harborth, D., Cai, X., and Pape, S. (2019). Why Do People Pay for Privacy-Enhancing Technologies? The Case of Tor and JonDonym. In Dhillon, G., Karlsson, F., Hedström, K., and Zúquete, A., editors, *ICT Systems Security and Privacy Protection. SEC 2019. IFIP Advances in Information and Communication Technology, vol 562*, pages 253–267. Springer, Cham.

Harborth, D. and Pape, S. (2018a). Examining Technology Use Factors of Privacy-Enhancing Technologies: The Role of Perceived Anonymity and Trust. In *Twenty-fourth Americas Conference on Information Systems (AMCIS2018)*, pages 1–10, New Orleans, USA.

Harborth, D. and Pape, S. (2018b). JonDonym Users' Information Privacy Concerns. In Janczewski, L. and Kutyłowski, M., editors, *ICT Systems Security and Privacy Protection - 33rd IFIP TC 11 International Conference, SEC 2018*, pages 170–184, Poznan, Poland. Springer, Cham.

Harborth, D. and Pape, S. (2019). How Privacy Concerns and Trust and Risk Beliefs Influence Users' Intentions to Use Privacy-Enhancing Technologies - The Case of Tor. In *Hawaii International Conference on System Sciences (HICSS) Proceedings*, Hawaii, US.

Harborth, D. and Pape, S. (2020a). Explaining Technology Use Behaviors of Privacy-Enhancing Technologies: The Case of Tor and JonDonym. In *Proceedings on Privacy Enhancing Technologies (PETS) (accepted)*, pages 1–18.

Harborth, D. and Pape, S. (2020b). HOW PRIVACY CONCERNS, TRUST AND RISK BELIEFS AND PRIVACY LITERACY INFLUENCE USERS' INTENTIONS TO USE PRIVACY-ENHANCING TECHNOLOGIES - THE CASE OF TOR. *ACM SIGMIS The DATA BASE for Advances in Information Systems*, (forthcoming).

Malhotra, N. K., Kim, S. S., and Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355.

Manning, K. C., Bearden, W. O., and Madden, T. J. (1995). Consumer Innovativeness and the Adoption Process. *Journal of Consumer Psychology*, 4(4):329–345.

Masur, P. K., Teutsch, D., and Trepte, S. (2017). Entwicklung und Validierung der Online-Privatheitskompetenzskala (OPLIS) [Development and validation of the Online Privacy Literacy Scale (OPLIS)]. *Diagnostica*, 63(4):256–268.

Pavlou, P. A. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, 7(3):101–134.

Rosen, L., Whaling, K., Carrier, L., Cheever, N., and Rokkum, J. (2013). The Media and Technology Usage and Attitudes Scale: An empirical investigation. *Comput Human Behav.*, 29(6):2501–2511.

Venkatesh, V. and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal Studies. *Management Science*, 46(2):186–205.

Venkatesh, V., Thong, J., and Xu, X. (2012). Consumer Acceptance and User of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1):157–178.