UNIVERSITY OF CAMBRIDGE
Computer Laboratory





Ethical issues in research using datasets of illicit origin

Daniel R. Thomas Sergio Pastrana Alice Hutchings
Richard Clayton Alastair R. Beresford

Cambridge Cybercrime Centre, Computer Laboratory, University of Cambridge, UK

Internet Measurement Conference (IMC) 2017

GPG: 5017 A1EC 0B29 08E3 CF64 7CCD 5514 35D5 D749 33D9

Think before using data of illicit origin

- Include an ethics section
- “human participants” not “human subjects”
- Talk to your Research Ethics Board (REB) (IRB / Ethics committee)

We had questions about ethics in our research[1]

- UDP honeypot DDoS sensors
- Developed statistical method for estimating coverage
- Verified using leaked booter databases...

[1] Daniel R. Thomas, Richard Clayton, and Alastair R. Beresford. 2017. 1000 days of UDP amplification DDoS attacks. In *APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Scottsdale, AZ, USA.

Case studies

- Malware & exploitation (Carna, ...)
- Password dumps
- Leaked databases (Booters, Patreon, ...)
- Classified materials (Snowden, Manning)
- Financial data (Panama papers)

We drew out common justifications, safeguards, potential harms, and benefits.

Ethics are norms of conduct

- Distinguish between acceptable and unacceptable behaviour
- Guidance in Menlo Report[2]
- Enforced by REBs and Program Committees (e.g. IMC[3])
- Minimise harm, maximise benefit.

[2] David Dittrich, Michael Bailey, and Erin Kenneally. 2013. Applying ethical principles to information and communication technology research: A companion to the Menlo Report. Tech. rep. U.S. Department of Homeland Security, (Oct. 2013) .

[3] Mark Allman and Vern Paxson. 2007. Issues and etiquette concerning use of shared measurement data. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 135–140.

We should consider these ethical issues

- Identification of stakeholders
- Informed consent (where possible)
- Identify harms
- Safeguards
- Justice
- Public interest

Leaked databases: Ethical issues

Sources	Ref	Year	Identify stakeholders	Identify harms	Safeguards	Justice	Public interest
6 underground forums	[6]	2011	✓	✓	✗	✓	✓
3 carding forums	[12]	2013	✗	✗	✗	✓	✓
TwBooter	[4]	2013	✓	✓	✓	✗	✗
TwBooter, 14 others	[9]	2013	✓	✓	✓	✓	✓
Asylum, Lizard, Vdos	[5]	2015	✓	✓	✓	✓	✓
Patreon	[7]	2016	✓	✓	✓	✓	✓
Vdos, CMDBooter	[10]	2017	✓	✓	✓	✓	✓
4 underground forums	[8]	2017	✓	✗	✗	✓	✓

Summary of the ethical issues and the identified by the authors for each paper. ✓ means the issue was discussed, ✗ means it was not.

Legal issues may arise

- Computer misuse
- Copyright
- Data privacy (e.g. GDPR)
- Terrorism
- Indecent images
- National security
- Contracts

IANAL. Just pointers: not even the beginning of advice.

AT&T iPad users database breach

- "Researchers" from Goatse Security found vulnerability in AT&T website
- Obtained email addresses for 114 000 iPad users
- Shared vulnerability, exploit, and email addresses with third parties
- Did not report vulnerability to AT&T
- Went to jail

Unethical and illegal.

Illegal but ethical research?

- *Mens rea*: lack of criminal intent
- Not in the public interest to prosecute
- REB approval! REB protection?
- Get the law fixed

Bad and good justifications

- Not the first
- Public data

- No additional harm
- Fight malicious use
- Necessary data

Leaked databases: Justifications and oversight

Sources	Ref	Year	Not the first	Public data	No additional harm	Fight malicious use	Necessary data	Ethics section	REB approval
6 underground forums	[6]	2011	✗	✓	✓	✗	✗	✗	✗
3 carding forums	[12]	2013	✗	✗	✗	✗	✗	✗	✗
TwBooter	[4]	2013	✓	✓	✓	✗	✗	✓	✗
TwBooter, 14 others	[9]	2013	✓	✓	✓	✗	✗	✓	✗
Asylum, Lizard, Vdos	[5]	2015	✓	✗	✓	✗	✗	✓	E
Patreon	[7]	2016	✗	✓	●	✗	✓	✓	∅
Vdos, CMDBooter	[10]	2017	✗	✗	✓	✗	✓	✓	E
4 underground forums	[8]	2017	✗	✗	✗	✗	✗	✗	✗

● means that the authors decided that the work could not be justified.

∅ means not applicable and E means exempt.

Patreon database leak[4]

- Researching crowdfunding on Patreon by scraping the website (complete scrape?)
- Patreon is compromised and full database + source code leaked
- Authors decided there would be additional harm
  - Private vs. public data
  - Legitimize criminal activity
  - Violate privacy

[4] Nathaniel Poor and Roei Davidson. 2016. The ethics of using hacked data: Patreon's data hack and academic data standards. Data and Society. Tech. rep. Council for big data, ethics and society, (Mar. 2016), 1–7. Retrieved Sept. 28, 2017 from http://bdes.datasociety.net/council-output/case-study-the-ethics-of-using-hacked-data-patreons-data-hack-and-academic-data-standards/.

Safeguards against potential harm

SS Secure storage

P Privacy

CS Controlled sharing

Potential harms

I	Illicit measurement
PA	Potential Abuse
DA	DeAnonymisation
SI	Sensitive Information
BC	Behavioural Change
RH	Researcher Harm

RH: Researcher Harm

- Researchers are participants
- Ethics also protects researchers
- Maltese journalist Daphne Caruana Galizia killed by car bomb used Panama Papers data also used by researchers

Benefits provided

R	Reproducibility
U	Uniqueness
DM	Defence Mechanisms
AT	Anthropology & Transparency

Leaked databases: Safeguards, harms and benefits

Sources	Ref	Year	Safeguards	Harms	Benefits
6 underground forums	[6]	2011			U,DM,AT
3 carding forums	[12]	2013			DM,AT
TwBooter	[4]	2013	P	SI	
TwBooter, 14 others	[9]	2013	P	SI	
Asylum, Lizard, Vdos	[5]	2015	P	SI	
Patreon	[7]	2016		SI,RH	U,AT
Vdos, CMDBooter	[10]	2017	P,CS	SI,BC	U,AT
4 underground forums	[8]	2017			R,DM,AT

Safeguards: SS Secure Storage, P Privacy, CS Controlled Sharing.
Harms: I Illicit measurement, PA Potential Abuse, DA De-Anonymization, SI Sensitive Information, RH Researcher Harm, BC Behavioural Change.
Benefits: R Reproducibility, U Uniqueness, DM Defence Mechanisms, AT Anthropology and Transparency.

Safeguards, harms, and benefits[5]

Privacy: No individuals identified

Controlled Sharing: Under legal agreement

Sensitive Information: Names, email addresses, criminal activity

Behavioural Change: Don't advertise (working) booters

Uniqueness: Ground truth on booter activity

Anthropology & Transparency: Real behaviour of booters

[5] Daniel R. Thomas, Richard Clayton, and Alastair R. Beresford. 2017. 1000 days of UDP amplification DDoS attacks. In *APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Scottsdale, AZ, USA.

Lets do ethical research

- Explain why our work is ethical
- Learn from each other
- Improve Research Ethics Board (REB) process

Thank you! Questions?

Daniel.Thomas@cl.cam.ac.uk
https://www.cl.cam.ac.uk/~drt24/
GPG: 5017 A1EC 0B29 08E3 CF64 7CCD 5514 35D5 D749 33D9

Cambridge Cybercrime Centre has data to share
https://www.cambridgecybercrime.uk/

University support for complex ethics workshop
Oxford 12th December

Daniel R. Thomas was supported by a grant from ThreatSTOP. All authors are supported by the EPSRC [grant number EP/M020320/1].

References I

[1] Mark Allman and Vern Paxson. 2007. Issues and etiquette concerning use of shared measurement data. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 135–140.

[2] David Dittrich, Michael Bailey, and Erin Kenneally. 2013. Applying ethical principles to information and communication technology research: A companion to the Menlo Report. Tech. rep. U.S. Department of Homeland Security, (Oct. 2013).

[3] David Dittrich and Erin Kenneally. 2012. The Menlo Report: Ethical principles guiding information and communication technology research. Tech. rep. U.S. Department of Homeland Security, (Aug. 2012).

References II

[4] Mohammad Karami and Dammon McCoy. 2013. Rent to pwn: Analyzing commodity booter DDoS services. *Usenix login;* 38, 6, 20–23. USENIX.

[5] Mohammad Karami, Youngsam Park, and Damon McCoy. 2016. Stress testing the Booters: Understanding and undermining the business of DDoS services. In *Proceedings of the 25th International Conference on World Wide Web (WWW)*. ACM, 1033–1043.

[6] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. 2011. An analysis of underground forums. In *Proceedings of the ACM SIGCOMM Internet measurement conference (IMC)*. ACM, 71–80.

References III

[7] Nathaniel Poor and Roei Davidson. 2016. The ethics of using hacked data: Patreon's data hack and academic data standards. Data and Society. Tech. rep. Council for big data, ethics and society, (Mar. 2016), 1–7. Retrieved Sept. 28, 2017 from http://bdes.datasociety.net/council-output/case-study-the-ethics-of-using-hacked-data-patreons-data-hack-and-academic-data-standards/.

[8] Rebecca Portnoff, Sadia Afroz, Greg Durrett, Jonathan Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. 2017. Tools for automated analysis of cybercriminal markets. In *Proceedings of 26th International World Wide Web conference (WWW)*. ACM, 657–666.

References IV

[9] José Jair Santanna, Romain Durban, Anna Sperotto, and Aiko Pras. 2015. Inside booters: An analysis on operational databases. *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015*, 432–440.

[10] Daniel R. Thomas, Richard Clayton, and Alastair R. Beresford. 2017. 1000 days of UDP amplification DDoS attacks. In *APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Scottsdale, AZ, USA.

[11] Daniel R. Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R. Beresford. 2017. Ethical issues in research using datasets of illicit origin. In *Proceedings of the Internet Measurement Conference (IMC)*. ACM. London, UK, (Nov. 2017).

References V

[12] Michael Yip, Nigel Shadbolt, and Craig Webber. 2013. Why forums? An empirical analysis into the facilitating factors of carding forums. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 453–462.