CSC2552 Project: The Digital Deluge

Thomas Hollis

University of Toronto

thollis@cs.toronto.edu

February 16, 2019

Overview

I) Background: Motivation

As we approach April 2019, dawn of the first anniversary of the General Data Protection Regulation (GDPR) written into law by the EU in 2018, questions regarding consumer data analytics have reached an all time high. The trust of data collection and privacy in Silicon Valley has plummeted [1], following scandals such as Facebook's Cambridge Analytica debacle. In response to such growing global concern, it seems fitting to exploit GDPR legislation, introduced last year, to uncover the current state of digital data privacy. The main research question that is tackled here is therefore: *"How much can GDPR queries reveal about the use of consumer data in tech companies in 2019?"*
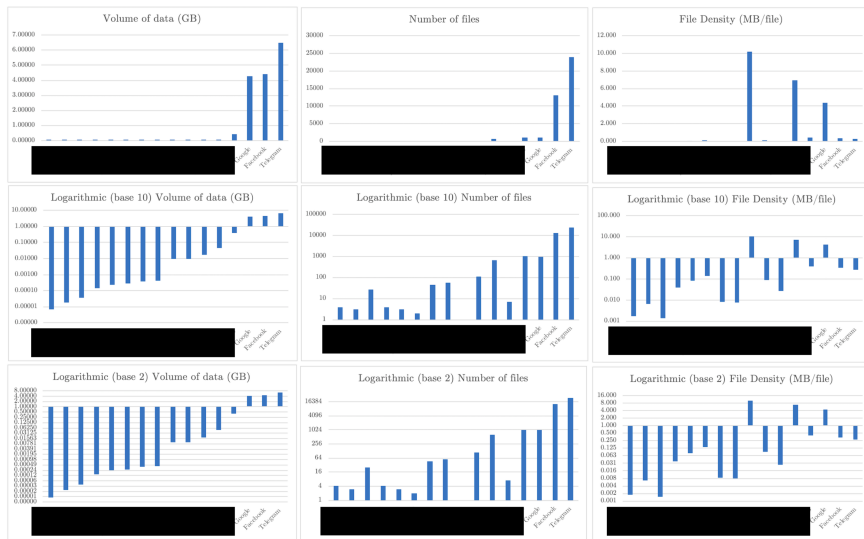
I) Background: Methodology

- Identification of all websites that store data on me (complete)
- GDPR requests to all websites (complete)
- Take a pre-analysis survey (complete)
- Preliminary analysis of size, depth and contents (complete)
- Multiple individual analysis of specific companies and the potential of the data stored
- Take a post-analysis survey to see how opinions have changed

Company	GDPR compliance	Response speed (in hrs)	Volume of data (bytes)	Volume of data (GB)	Number of files	File Density (MB/file)	IP logging	Location logging	Sentiment Ad Sensitivity	Fingerprinting (device info)	Search history logging	Contact info	Total Score	Comments
	compliant	< 1		0.00001	4	0.002	No	No	No	No	No	No	50	Minimal, great
	compliant	< 1		0.00002	3	0.000	Yes	Yes	Yes	Yes	No	No	30	Excessive for usage
Microsoft	compliant	< 1		0.00004	28	0.001	No	No	Yes	Yes	No	No	40	Minimal, great
	compliant	< 1		0.00015	4	0.038	No	Yes	No	No	Yes	No	40	Bad, hiding info
	compliant	60		0.00034	3	0.081	Yes	No	No	No	Yes	Yes	35	Minimal, great
	compliant	< 1		0.00028	2	0.142	No	No	No	No	No	No	50	Logged all code
	compliant	< 1		0.00039	41	0.009	Yes	Yes	No	Yes	Yes	Yes	20	No chat history?
	compliant	14		0.00043	50	0.008	Yes	No	Yes	Yes	No	Yes	30	Mostly profile info
	compliant	< 1		0.00098	1	10.225	No	No	No	No	No	No	50	
	compliant	< 1		0.01068	113	0.091	No	Yes	No	No	No	Yes	40	
	compliant	121		0.01751	645	0.026	Yes	Yes	Yes	Yes	Yes	Yes	20	
	compliant	< 1		0.04763	7	6.987	Yes	Yes	No	Yes	No	No	35	
GitHub	compliant	< 1		0.40378	1048	0.397	No	No	No	No	No	No	50	
Google	compliant	< 1		4.20081	989	4.418	Yes	Yes	Yes	Yes	Yes	Yes	20	Extensive/Invasive
Facebook	compliant	< 1		4.42603	13,182	0.344	Yes	Yes	Yes	Yes	Yes	Yes	20	Extensive/Invasive
Telegram	compliant	< 1	6,951,734,078	6.47433	24,884	0.270	Yes	Yes	No	Yes	No	Yes	30	Only message/security data
	compliant	-	-	-	-	-	-	-	-	-	-	-	50	-
	excellent	-	-	-	-	-	-	-	-	-	-	-	100	-
	excellent	-	-	-	-	-	-	-	-	-	-	-	100	-
	excellent	-	-	-	-	-	-	-	-	-	-	-	100	-
Amazon	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
	poor	-	-	-	-	-	-	-	-	-	-	-	0	-
Slack	poor	-	-	-	-	-	-	-	-	-	-	-	0	-

Questions?