Privacy as a Currency: Un-regulated?

Vishwas T. Patil and R. K. Shyamasundar

Department of Computer Science and Engineering, Information Security R&D Center,
Indian Institute of Technology Bombay, Mumbai 400076, India

Keywords: Privacy, PII (Personally Identifiable Information), Social Networks, Access Control, Trust.

Abstract: We are living in a time where many of the decisions that affect us are made by mathematical models. These models rely on data. Precision and relevance of the decisions made by these models is dependent on quality of the data being fed to them. Therefore, there is a rush to collect personal data. Majority of the organizations that provide online services are at the forefront of collecting user data. Users, either voluntarily or by coercion, divulge information about themselves in return of personalized service, for example. These organizations' revenue model is based on advertisement where advertisers are paired with user profiles that are built on top of collected data. This data is being used for a variety of purposes apart from delivering targeted advertisements. Mathematical decision models are impartial to the data on which they operate. An error, omission or misrepresentation in data has an irrevocable consequence on our lives, at times, without corrective remedies. This paper touches upon various facets of information gathering; information bias, economics of privacy, information asymmetry – and their implications to our ecosystem if left unaddressed.

1 INTRODUCTION

The advent of communication mediums like newspaper, telegraph, telephone, television and Internet profoundly impacted human lives by allowing humans to trade or exert influence beyond their immediate physical sphere. Internet being the efficient, cost-effective, real-time, two-way communication medium, upended all other preceding mediums of communication and has become the de-facto mode of communication today. It has completely reformed the traditional methods of trade – to the extent that online economy has become *the economy*. In the early days of this transformation from old services into new, there was a lack of an obvious method to charge for these online services. Fee-based and advertisement-based revenue models emerged; the latter has prevailed and is prevalent even today. This is intriguing[1] (Branstetter, 2015) and is the theme of this paper.

Advertisements existed even before Internet but then the central problem was the conflation of audience and traditional media outlets. Big companies with lots of advertising budget needed a way to convince (psychologically manipulate) people to buy their goods. Digital advertising fundamentally altered this model. The computers can watch what one does online and profile users based on their online behavior. Through profiling, service platforms (and in turn the advertisers) aim to know the exact extent to which one is engaged with the service. User profiles grow more accurate with each use of online services and thus shrinks the famous *purchase funnel*[2]. Furthermore, the profiles are enriched using auxiliary information (Calandrino et al., 2011) available from offline platforms/services. In fact, the penchant for data collection is so high that it has become a core objective of many online services – the reasons could be data about user interaction is incidental and the cost to store data is negligible as compared to monetary return the stored data promises.

Data is touted as new Gold (Popper, 2017; Angwin, 2010). Therefore, it is subjected to hoarding, re-sell, barter, etc. There are some obvious downsides and legitimate concerns to data-driven revenue models (Ezrachi and Stucke, 2016). In case of online advertisement, it relies on the profiling of users, which makes many people uncomfortable (Duhigg, 2012), even if the service providers and advertisers say that they do all of this anonymously (Hartzog and Rubinstein, 2017) and without invading privacy (Gao et al., 2011; Manjoo, 2017). On the other hand, this trove

[1] If you do not pay for the product, you are the product.

[2] A marketing model which illustrates the theoretical customer journey towards the purchase of a product.

Patil, V. and Shyamasundar, R.
Privacy as a Currency: Un-regulated?.
DOI: 10.5220/0006478705860595
In *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - Volume 4: SECRYPT*, pages 586-595
ISBN: 978-989-758-259-2

of harvested data has compelling usages beyond advertisement. For example, social networks help track potential spread (thus containment strategies) of contagious epidemics. Unlocking (WEF, 2013) this data needs people's trust. Whoever gets access to the trove of user profiles has advantage over others who do not have it. Therefore the methods to gather and interpret data have become a trade secret. Complex mathematical models are used for decision making, at times, on incomplete data with astonishing precision (O'Neil, 2016). An error in gathered data or the model will have serious repercussions. Application scenarios vary from suitability of a candidate for a job, premium for health insurance, one's political belief, etc. Data-driven mathematical models could lead to empowerment or to discrimination.

Today, huge companies like Amazon, Facebook, Google, and Netflix dominate the web. These corporate giants enjoy an enormous amount of control not only over what people see and do online but over users' private data. Through their privacy policies and settings they (under-) inform and (partially) allow users to see what private data is being collected and how it might be shared with third-parties. Among these giants Facebook is of our special interest because of its application domain – social network. In a social network, a user interacts with other users and the service provider is an intermediary. Therefore, even when one user among the two specifies her privacy policy to be restrictive and the other user specifies her privacy policy liberally, the intermediary does not have to (or cannot) respect the restrictive user's policy – this is what we have observed and report our findings in next section. The amount of data Facebook collects on users has helped it become the world's second-largest advertising company on mobile devices (Economist, 2016a). From time to time it tries to assuage its users about their privacy through privacy settings. However, we have found that certain functions/features of Facebook lead to subtle violations of those settings. In Section 5 we elaborate on how those functions could be implemented to ensure privacy-by-design. We also argue why "privacy as a currency" model is not sustainable, and in order to regulate the use of PII (Personally Identifiable Information) through a cohesive access control model, a common platform for PII is necessary – where users, service providers, and advertisers negotiate PII usage.

2 BACKGROUND

Social network services have an innate appeal to general population as they allow the users to build social relations with other people who share similar personal or career interests, activities, backgrounds or real-life connections. While doing so, users vet each other into audience types; a panacea for advertisers. Service providers innovate (Constine, 2016; Facebook, 2017b) ways to keep users engaged with the platform by providing features and content relevant to a user. User profiles are further enriched by all means possible, either by striking deals with other data aggregators (Halpern, 2016; Dewey, 2016) or through technology (FTC, 2017). The ultimate aim is to know as much as possible about a user and her social neighborhood.

Facebook owns four out of the five most downloaded apps worldwide (SensorTower, 2017). It has become more like a holding company for popular communications platforms than a social network (Economist, 2016b). It appears that human psychology (convenience of socializing online for free, lack of apparent harm, value we place on our privacy) plays a role behind the complicity of users in this massive data aggregation. For the privacy-aware users, Facebook allows to specify privacy settings such that users can decide who all can see or interact with their social persona. The privacy settings safeguard information from other users, not from Facebook. Facebook is implied to be a trusted party. It enforces privacy policies of its users, it assures the users that it internally regulates PII data usage, it also is the consumer of the PII data it regulates, for its business interests. There is a conflict-of-interests!

In the following we list out the reasons and dynamics that are at play behind this sublime currency of privacy. There are 3 pillars of this ecosystem:

Platform Provider – Allows users to form social connections as they do in real-life. Entices users to use the platform by providing compelling features while recording all their interactions. Deploys automated language processes and sentiment analysis on user interactions to categorize users into profiles, affinities, and communities such that the organization is optimized for business practices. It monetizes (Saez-Trumper et al., 2014) the organized data through advertisers and other entities that find value in this data.

Businesses and Organizations – The social media has become an essential medium for brands and organizations because it helps them in;

- advertisement, targeted promotions
- actionable insights for decision making
- brand monitoring & crisis detection
- measurable engagement with communities
- identify emerging trends/markets

- identify undecided voters (linkfluence, 2017)

Users (The Product) – All of the below reveal certain aspect of an individual to the platform.

- targeted feeds: news, entertainment, friends
- product reviews/referrals, redressal
- expression of opinions, start interest groups
- match-making: romance, teachers, plumbers, etc.
- build reputation: seeking work/employment
- location-based assistance, travel

A sense of control and protection is provided to the users through privacy settings and privacy policy of the platform. However, we have observed that even stated policies being violated through legitimate innocuous user actions on the platform of Facebook.

3 GAPS IN PRIVACY SETTINGS

Each user on Facebook is provided with pre-defined relationship categories, called lists, along which users can organize their relationships with others. “Friends” is the basic relationship category to which every user-to-user relationship (friendship) is added. A user is allowed to organize friendship relations into other pre-defined categories like “Family”, “Close Friends”, “Acquaintances” so that a distinct affinity level could be imposed on relations. This is how people, in real-world, intend to organize their relationships. This notion of categorizing (or listing of) friends into affinity levels help users to specify who can have access to their information. Labels are used as access control policies over a user’s information. Any requester who satisfies membership to the label assigned with the post can access the post.

Facebook provides another set of labels for information classification that is intensional, whose members are not due to direct action by the user. “Friends of friends”, “Public”, and any other affiliation-based smart label like “University” or “School” fall under this category. The whole gamut of information labelling in Facebook provides a very rich and flexible access (thus privacy) policy specification over a user’s information. Users are allowed to change labels of their objects as per their discretion. However, this flexibility in policy specification is not well-understood (Liu et al., 2011) by majority of the users and users end up in a state where their policy specification may look innocuous, whereas it may not. We show some such instances with the help of a hypothetical scenario on Facebook as depicted in Table 1. Each row represents a user’s actions in chronological fashion. Therefore, we use A.t_2 to denote an action of user A at time t_2. All other actions, of every user, up to time t_2 is the environment/status of social graph (Bronson et al., 2013) w.r.t. A.t_2. Thus, an action should be analyzed in context with its current state of the environment. Note that, since social graph is a co-creation by its users, an individual has little or no control over the environment in which he/she is operating. An action/setting that seems privacy-preserving can later be compromised by a change in environment. This will become clear as we go through the scenarios below.

Nonrestrictive Change in Policy of an Object Risks Privacy of Others – Consider A.t_2 in context with environment trace B.t_1, B.t_3. As user A is member of B’s Family, through action A.t_2 user A has authored a comment on P_{B_1}. The environment at time t_2 ensured that only Family members of B had access to object P_{B_1}. At t_3 user B has changed policy of his object P_{B_1} from Family to Friends. Thus, *user A’s comment is exposed to friends of B without A’s consent.*

Restrictive Change in Policy of an Object Suspends Others’ Privileges – Consider user action E.t_5 in context with two other events in environment D.t_2, F.t_4. User E has changed policy of her object P_{E_1} from Public to Only Me (restrictive). Prior to policy change, users D and F have liked the object as it was public. *A restrictive change in policy over P_{E_1} locks out users from updating/retracting their own comments or likes. At a later point in time, user E can divulge list of users associated with her post in the past.*

Policy Composition using Intensional Labels is not Privacy-preserving – Consider user actions F.t_3, E.t_4 in context with social list “School” at t_4. Through action in F.t_3, user F has created an object P_{F_3} with a custom policy University-School. Here the intention of the user is to make the object available to his friends from University but not from his School. According to the state of social graph at time t_3 nobody gets access to object P_{F_3} because University $\in$ School. At time t_4, user E disassociates herself from social list School and thus could get access to P_{F_3}. *Disassociation from a social list allow users to bypass the privacy/access intention of a custom policy composed of social labels.*

***Like, Comment* Operations are not Privacy-preserving –** Consider user actions D.t_2 and F.t_4 in context with environment event E.t_1. On Facebook, *List of Friends* is an object of user profile. In its privacy settings, Facebook allows to choose intended audience for this object. We assume all users in this scenario have set their audience to “Only Me” for this object. The intention behind such a setting is not to let the profile visitors know who their friends are, except

Table 1: Snapshots of associations formed in a social graph.

$\xrightarrow{\textit{time}}$

	t_1	t_2	t_3	t_4	t_5
A	$\text{A} \xrightarrow[\text{OnlyMe}]{\text{authored}} P_{A_1}$	$\text{A} \xrightarrow{\text{authored}} P_{B_1}$	$\text{A} \xrightarrow[\text{Public}]{\text{authored}} P_{A_1}$	$\text{A} \xrightarrow{\text{likes}} P_{F_2}$	$\text{A} \xrightarrow{\text{likes}} P_{F_3}$
B	$\text{B} \xrightarrow[\text{Family}]{\text{authored}} P_{B_1}$	$\text{B} \xrightarrow[\text{Public}]{\text{authored}} P_{D_1}$	$\text{B} \xrightarrow[\text{Friends}]{\text{authored}} P_{B_1}$		
C	$\text{C} \xrightarrow[\text{Friends}]{\text{authored}} P_{C_1}$	$\text{C} \xrightarrow{\text{authored}} P_{D_1}$	$\text{C} \xrightarrow{\text{authored}} P_{C_1}$	$\text{C} \xrightarrow{\text{authored}} P_{B_1}$	$\text{C} \xrightarrow{\text{likes}} (\text{A} \xrightarrow{\text{authored}} P_{B_1})$
D	$\text{D} \xrightarrow[\text{FFriends}]{\text{authored}} P_{D_1}$	$\text{D} \xrightarrow{\text{likes}} P_{E_1}$			
E	$\text{E} \xrightarrow[\text{Public}]{\text{authored}} P_{E_1}$	$\text{E} \xrightarrow{\text{likes}} P_{F_1}$	$\text{E} \xrightarrow{\text{likes}} P_{D_1}$	$\text{E} \xrightarrow[\text{FFriends}]{\text{authored}} P_{F_3}$	$\text{E} \xrightarrow[\text{OnlyMe}]{\text{authored}} P_{E_1}$
F	$\text{F} \xrightarrow[\text{School}]{\text{authored}} P_{F_1}$	$\text{F} \xrightarrow[\text{School-E}]{\text{authored}} P_{F_2}$	$\text{F} \xrightarrow[\text{University-School}]{\text{authored}} P_{F_3}$	$\text{F} \xrightarrow{\text{likes}} P_{E_1}$	$\text{F} \xrightarrow{\text{likes}} P_{A_1}$

Assumptions:

at time t_0: (following is the state of different sets)

users = {A, B, C, D, E, F} *objects* = $\{P_{A_1}, P_{B_1}, P_{C_1}, P_{D_1}, P_{E_1}, P_{F_1}, P_{F_2}, P_{F_3}\}$

friendship edges = {(A,B), (B,C), (C,D), (D,E), (E,F), (F,A)} Family$_B$ = {A}

University = {E, F} School = {A, E, F}

at time t_4: (user E has disassociated from list School) School = {A, F}

at time t_5: University = {A, E, F} School = {F}

their mutual friends.

The way Facebook works, Newsfeed of a user is supplied with relevant content from user's social circle. With a high probability friends posts appear in Newsfeed to which the user may interact by making a comment or like. These interactions get consumed by the underlying social graph. *When a user interacts with objects with access policy set to Public, those interactions also become public. Social graph allows queries to public content.* For example, `fb.com/search/FBID/photos-commented` returns all the photo type of objects on which Alice has commented. Similarly, `fb.com/search/FBID/photos-liked` returns all photos liked by Alice. *For a typical user, these queries return objects from their friends. Any user of Facebook can make these queries to social graph for any other user of Facebook.*

List of friends is a sensitive object of any user and its privacy is important because knowing one's friend helps a social engineer to devise identity theft/cloning, phishing attacks on the user (Jagatic et al., 2007; Bilge et al., 2009). There are numerous other objects (Facebook, 2017a) that are associated with each user on Facebook; e.g., email, date of birth, mobile OS type, currency, timezone, etc. Only a portion of the complete object set is guarded by owner's access control/privacy settings. We observed that the notion of privacy conferred upon its users by Facebook is limited to human subjects alone, i.e., privacy from fellow users or public. However the apps that are integrated with Facebook's ecosystem have a quasi unfettered access to users' objects. The ecosystem is indeed a colossal social experiment of our time involving platform owners (a few – controllers), advertisers & tracker (many – influencers & aggregators), and the users (a lot – voluntary reporters or powerless subjects). This leads us to the analogy of "law of jungle" to aptly describe the current state of PII ecosystem where subjects pay a privacy tax to derive benefits from Internet.

4 THE LAW OF THE JUNGLE

Browsers and mobile apps have become our interfaces to online services. These interfaces along with their underlying software & hardware platforms have a first hand access to our PII. Evidently enough, this space is controlled by service providers whose revenue model is based on online targeted advertising. By providing features, which are usually turned on by default, e.g., security feature of website screening for malware detection on browsers and feature of better data connectivity with the help of GPS on mobiles, they turn these interfaces as sensors that continuously report about user activity. By knowing the locale of these interfaces and their time-zones it is straight-forward to guess about users' sleeping patterns – an important health factor that insurers may like to use to evaluate risk of an individual.

Next in the pecking order is: the content publishers. Content publishers have first hand access to transactional attributes of its users. Most of the content is delivered via HTTP, which is a state-less protocol. To maintain state of a user across sessions over the period of time, cookies are used. This helps publishers to track and learn the user behavior. Publishers often support their content through the advertisement revenue they collect by allowing advertisers target their users. Using third-party cookies, a cross-publisher tracking of users is performed. Interfaces accept third-party cookies by default as a matter of convenience to users.

Next in the pecking order is: the connectivity providers. ISPs can (have to) track their users for various reasons. Though there is a wide acceptance of HTTPS protocol they can still observe the meta-data about the communication and at times the communication itself (Upturn, 2016; Dubin et al., 2016). ISPs have to collect PII like home/office address for billing, and details of the financial instrument used for payment. This auxiliary data has potential to reveal an individuals credit rating, for example. Mobile ISPs have advantage of knowing location of its customer in real-time and call history. Mobile apps like Truecaller and Facebook Messenger are capable of collecting location and call logs, which can be turned-off by the user unlike Mobile ISPs collecting logs unconditionally.

Next in the pecking order is: off-grid service providers. Service providers like Citibank, Visa, Uber, Walmart, record transactional data about their customers. These data sets along with its meta-data are of interest to themselves and others from same industry domain or across the domains. For example, (Cranshaw et al., 2010) establishes relationship between mobility patterns (say from Uber) and structural properties of their underlying social network.

Next in the pecking order is: advertisers & aggregators. They are responsible for targeting advertisements to their intended audience. The evolved revenue model is how many actual potential customers get converted into real customers. This greatly depends on accurately discovering the audience. Aggregators help link data points from different web events and off-grid events (BlueKai, 2010) and enrich the profiles of intended audience.

Next in the pecking order is: analytics & intelligence providers. They work on huge trove of PII and off-grid auxiliary data sources in order to calculate/guess intentions, future behaviors, values, preferences, habits, aspirations, etc. of an individual or an audience. It only takes a few data points to track a set of web & off-grid events back to a real person. Facebook, in one of its experiments (Kramer et al., 2014), showed its capability to play with the emotional quotient of its users by tweaking its Newsfeed algorithm. A social network analytics and intelligence startup (linkfluence, 2017) offers developing campaign tactics to identify pockets of undecided voters to try to win them over (300 million sources and 100 million posts harvested on a daily basis by linkfluence to derive the actionable insights).

Last in the pecking order is: governments & law enforcement. They are the lawful users of PII and also the lawful collectors of it. In many countries they do not have the technical, financial wherewithal to collect and derive actionable insights. They frequently rely on the decision to subpoena (accessnow, 2017) corporate giants for that shortcoming. They find these corporate giants doing a complementary[3] job of surveillance and intelligence gathering. And therefore, plausibly, may not have intentions to disturb the established pecking order.

The ability for individuals to interact online without sacrificing their PII is a vital part of the Internet's value, and is intimately related to its trustworthiness. Users are left with limited or inconvenient options at their disposal to conduct themselves in a PII-preserving fashion on Internet.

> *Now these are the Laws of the Jungle, and many and mighty are they; But the head and the hoof of the Law and the haunch and the hump is – Obey!* [Rudyard Kipling]

5 HOW TO PROTECT THE PII?

Today there is a constant battle between privacy advocates and advertisers, where advertisers try to push new personalization technologies, and privacy advocates try to stop them. However, as long as privacy advocates are unable to propose an alternative person-

[3] Facebook is turning into something of an extra-territorial state run by a small, unelected government that relies extensively on privately held algorithms for social engineering (Bershidsky, 2017).

alization system that is private, this is a battle they are destined to lose (Guha et al., 2011).

Privacy is about retaining the ability to disclose data consensually, and with expectations about the context and scope of sharing. Identifiability, linkability of data, and the mining of vast quantities of aggregated information all erode the individual's ability to manage disclosure, context and scope. Networks depend on the use of unique (and often identifying) numbers, and facilitate the instant global dissemination of information; increasingly, devices and applications gather and use geolocation data that builds up into a unique profile for each user. A growing commercial ecosystem based on targeted and behavioral advertising results in an inexorable financial pressure for service providers to exploit personal data. The privacy implications of the current Internet represent a significant and growing concern (ISOC, 2017). The concerted, persistent efforts of organizations to track & target users will have detrimental consequences: erosion of public trust (WEF, 2012), inhibits freedom of expression and freedom of action online (even offline, due to off-grid actors and extent of technology integration in our daily lives). Before we delve into the solutions, we must understand the nature of information and constraints it imposes.

5.1 Nature of PII

Issue of privacy comes to fore as soon as an unintended observer observes an information and learns something more, which later could be associated with that subject under observation. We have seen a range of intended and unintended observers in Section 4. User actions on online services are transactional therefore both parties involved in the transaction have access and ownership to the incidental data arising out of the transaction. Both parties must adhere to common expectations about usage of this co-created data. It is difficult to verify or validate whether the expectations are respected, therefore enforcement is difficult. A refuge in legislation is taken where service provider through its privacy policy commits to the contexts in which user's PII will be used, shared. In the presence of multiple observers (informed or lawful), same events might be recorded by all the observers. Therefore, tracing a leaker of recorded events becomes challenging. This fact indirectly encourages the observers to be lax/exploitative with users' PII.

Due to the architecture of our online ecosystem and the composition of modern interwoven web-services, a typical user request for a service spans across different administrative domains that are participating in the delivery of the web-service (for example, Amazon and FedEx). Each domain operates under its own stated privacy policy and geographic legislative assurances. There is no universal legislation for Internet. The collected user data resides in different administrative domains with respective privacy oversights that are often difficult to verify until thoroughly investigated after a huge breach occurs.

User data is scattered across silos controlled by different entities of the ecosystem. To enforce a uniform usage policy that is acceptable to all stakeholders of the ecosystems the data need to be available for a uniform, consistent treatment. The incumbents who are greatly benefiting from the status quo have no incentive to contribute towards this effort.

5.2 Promising Approaches

5.2.1 Data Exchanges

Tapping on the potential perils of PII violations, online services like `respectnetwork.com` have tapped into the business of providing privacy to PII. It provides a platform for individuals where they can store and control their PII. In (Riederer et al., 2011), a mechanism called transactional privacy that can be applied to personal information of users is provided. Users decide what personal information about themselves is released and put on sale while receiving compensation for it (e.g., `datacoup.com`). Aggregators purchase access to exploit this information when serving advertisements to the user. Governments being the biggest lawful collectors, consumers, and producers of data, are facilitating exchange of data (Eggers et al., 2013) under Open Data Initiatives. These platforms have the potential (WEF, 2014) to allow users to store, control, curate, and monetize data as per their preferences and intents. Such platform brings in a level playing field but the *pecking order* will continue to harness auxiliary data, meta-data about users to get and edge over their competitors. Legislation can play a role to rectify this behavior.

5.2.2 Legislation

There is a school of thought that believes that Governments should nudge corporations to realize the role of public trust in the success of online economy. Through enacting legislations, governments can force corporations to join hands and negotiate a strategy that contributes in bringing back trustworthiness to the Internet. Governments have methods (as it has done for television industry) and tools (as it has done for telecommunications industry) at their disposal to bring in innovation through competition (Bergstein, 2017) and accessibility (Taplin, 2017). Legislations (e.g., EU GDPR – The EU General Data Protection Regulation) have geographic restrictions, it does not have universal reach. Legislation definitely has its utility but it often falls short of coherent enforcement on Internet.

5.2.3 Privacy-by-design

Unlike legislation, technology has a universal reach. Privacy-by-design (Cavoukian, 2012) is an approach where privacy expectations, out of the underlying platform, are guaranteed through its design. Solutions based on this approach are verifiable and enforceable. Privad (Private Verifiable Advertising) is such an alternative system (Guha et al., 2011) for today's advertising ecosystem. FAITH (Facebook Applications: Identification, Transformation & Hypervisor) is another such approach (Lee et al., 2011) designed for Facebook to mitigate privacy breaches and manage social data. In (Juels, 2001; Hardt and Nath, 2012; Guha et al., 2009; Cox et al., 2007) one can find how systems can be designed to accommodate privacy concerns of users and audience measurement requirements of platform owners. Security and privacy enhancing measures usually come at a cost, which is measured in terms of inconvenience. Users use information systems for certain primary objectives and when pressed against time, users tend to ignore/bypass the security or privacy enhancing measure in order to quickly achieve the primary objective (Kainda et al., 2010). Therefore, it is an important design consideration to keep privacy enhancing measures invisible from users' primary objectives.

Privacy-as-a-Service

In privacy enhancing services like πBox (Lee et al., 2013), an external, communication agnostic trusted platform is provided; where a user's PII is aggregated in a vault on user's behalf and made available to advertisers. The platform takes responsibility to safeguard user's privacy. In a natural extension of this approach, there could be multiple, independent PII vault providers on a single platform. Extending it further, there could be multiple, independent platforms. Assuming interoperability of PII across vaults/platforms, there will be a competitive spirit among privacy service providers to innovate and offer user-driven privacy protections. Privacy-as-a-service approach moves the users' trust away from a few powerful incumbents (like Facebook/Google) to a platform where users can observe, curate, delete PII. One can imagine such platforms are offered by consortiums involving governments with regulatory oversights. Envisioning such a federated platform; guarantees about data access control, usage control, integrity, trustworthiness will be desired.

Context-aware control via Blockchain

Owing to the design guarantees about trust and transparency emanating from blockchain algorithms, there are systems like Enigma (Zyskind et al., 2015), and MedRec (Azaria et al., 2016) that allow users to organize & store their PII on vaults in an encrypted format, whereas access rights and access policies (the contexts in which rights should be honored) are specified through smart contracts on the blockchain. Though smart contracts provide flexible control over user's PII through programmable privacy policies, a user would still require certain guarantees from PII consumer regarding post-access usage of PII. There could be scenarios in which users would like to participate in applications/surveys that require users to selectively declassify and desensitize their PII. This brings to life following interesting challenges:

1. who should own and control the data generated out of interactions between two or more subjects?
2. how to express (Wilton, 2013) subtle interpersonal concepts like trust and discretion?
3. how to help a user in determining whether her intended online activity is privacy-preserving w.r.t. her current environment/state?

Since users do not trust PII consumers, the underlying platform has to safeguard PII against: conflict-of-interest, information leakage due to declassification.

Information flow control via RWFM

Information flow security assures that there is no reverse flow (i.e., transfer of *high* information to a *low*-level subject). RWFM model (Narendra, N. V. and Shyamasundar, R. K., 2017) assures information flow security in a decentralized way and further introduces robust declassification rules (as opposed to other similar approaches) in the sense, a subject will not have authority to declassify information to subjects who have not participated in creating the information thus far. These properties shall enable to preserve privacy in the context of multi-level secure databases as well in terms of views, transaction desensitization, declassification, etc. An important desired property is to ensure that any platform that is collecting and monetizing PII data cannot be both consumer and regulator of the collected data, simultaneously – conflict-of-interest (Brewer and Nash, 1989; Narendra K and Shyamasundar, 2016).

Privacy-by-design approaches shift the onus of privacy guarantees of a policy from legislative assurance to technological assurance. And due to technology's universal reach we believe these are the most promising and viable approaches.

To summarize, assuming that, in the near future, our ecosystem converges towards tools and methods for *fair* collection-and-usage of PII through a common platform, there will be a need for a cohesive access control model for managing the PII throughout its life-time. Some of the foreseeable expectations from this platform would be: (i) assurance of integrity of intent and usage of PII, (ii) feedback loop to users alerting about privacy implications of their online activities against their stated privacy preferences on the platform. PII has become a currency and its usage needs to be controlled. Once such control measures are available, PII will be traded as a commodity on this platform as any other commodity is traded in financial markets.

6 CONCLUSION

Data *should not be* treated as Gold. Data *should be* treated as Oil because it not only has an intrinsic value but also has an utility as a fuel to run the decision making algorithmic engines. Its unfettered usage has a potential to pollute the online ecosystem and turn it into an unhealthy one. The fuel needs to be regulated and the engines need to be verifiably standardized to acceptable emission levels.

ACKNOWLEDGEMENT

The work was carried out as part of research at ISRDC (Information Security Research and Development Center), supported by 15DEITY004, Ministry of Electronics and Information Technology, Govt. of India.

REFERENCES

accessnow (2017). Transparency Reporting Index. https://www.accessnow.org/transparency-reporting-index/. accessnow.org.

Angwin, J. (Jul 30, 2010). The web's new gold mine: Your secrets. https://www.wsj.com/articles/SB10001424052748703940904575395073512989404. Wall Street Journal.

Azaria, A., Ekblaw, A., Vieira, T., and Lippman, A. (2016). MedRec: Using Blockchain for Medical Data Access and Permission Management. In 2^{nd} *International Conference on Open and Big Data*, pages 25–30.

Bergstein, B. (Apr 10, 2017). We Need More Alternatives to Facebook. https://www.technologyreview.com/s/604082/we-need-more-alternatives-to-facebook/. MIT Technology Review.

Bershidsky, L. (Feb 17, 2017). Facebook Plans to Rewire Your Life. Be Afraid. http://bv.ms/2lSQAtx. Bloomberg.

Bilge, L., Strufe, T., Balzarotti, D., and Kirda, E. (2009). All your contacts are belong to us: Automated identity theft attacks on social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 551–560. ACM.

BlueKai (2010). Take the mystery out of targeting. http://www.bluekai.com/files/mediakitBlueKai.pdf. The BlueKai Data Exchange.

Branstetter, G. (Feb 25, 2015). Why your privacy should be a currency. https://www.dailydot.com/via/your-privacy-should-be-a-currency/. The Daily Dot.

Brewer, D. F. C. and Nash, M. J. (1989). The chinese wall security policy. In *Proceedings of 1989 IEEE Symposium on Security and Privacy*, pages 206–214.

Bronson, N., Amsden, Z., Cabrera, G., Chakka, P., Dimov, P., Ding, H., Ferris, J., Giardullo, A., Kulkarni, S., Li, H., Marchukov, M., Petrov, D., Puzar, L., Song, Y. J., and Venkataramani, V. (2013). TAO: Facebook's Distributed Data Store for the Social Graph. In *USENIX ATC 13*, pages 49–60.

Calandrino, J. A., Kilzer, A., Narayanan, A., Felten, E. W., and Shmatikov, V. (2011). "you might also like: " privacy risks of collaborative filtering. In *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, S&P '11, pages 231–246. IEEE Computer Society.

Cavoukian, A. (Oct 2012). Privacy by Design and the Emerging Personal Data Ecosystem. https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-pde.pdf. Information and Privacy Commissioner, Ontario, Canada.

Constine, J. (Sep 06, 2016). How Facebook News Feed Works. https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/. TechCrunch.

Cox, L. P., Dalton, A., and Marupadi, V. (2007). Smokescreen: Flexible privacy controls for presence-sharing. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*, MobiSys '07, pages 233–245. ACM.

Cranshaw, J., Toch, E., Hong, J., Kittur, A., and Sadeh, N. (2010). Bridging the gap between physical location and online social networks. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, UbiComp '10, pages 119–128. ACM.

Dewey, C. (Aug 19, 2016). 98 personal data points that facebook uses to target ads to you. http://wapo.st/2bz22Cb. The Washington Post.

Dubin, R., Dvir, A., Pele, O., and Hadar, O. (2016). I know what you saw last minute - encrypted HTTP adaptive video streaming title classification. *CoRR*, abs/1602.00490.

Duhigg, C. (Feb 16, 2012). How companies learn your secrets. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html. New York Times.

Economist (Apr 7, 2016a). Facebook, the world's most addictive drug. http://www.economist.com/blogs/graphicdetail/2016/04/daily-chart-5. The Economist.

Economist (Apr 9, 2016b). The new face of facebook: How to win friends and influence people. http://www.economist.com/news/briefing/21696507-social-network-has-turned-itself-one-worlds-most-influential-technology-giants. The Economist.

Eggers, W. D., Hamill, R., and Ali, A. (2013). Data as the new currency: Government's role in facilitating the exchange. http://deloitte.wsj.com/ riskandcompliance/files/2013/11/DataCurrency_report.pdf. Deloitte Review.

Ezrachi, A. and Stucke, M. (2016). *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*. Harvard University Press.

Facebook (2017a). Graph API Explorer. https://developers.facebook.com/tools/explorer/. Facebook for developers.

Facebook (2017b). How News Feed Works. https://www.facebook.com/help/327131014036297. Facebook.

FTC (Jan 2017). Cross-device tracking: A federal trade commission staff report. https://www.ftc.gov/system/files/documents/reports/cross-device-tracking-federal-trade-commission-staff-report-january-2017/ftc_cross-device_tracking_report_1-23-17.pdf. Technical report, FTC, USA.

Gao, H., Hu, J., Huang, T., Wang, J., and Chen, Y. (2011). Security issues in online social networks. *IEEE Internet Computing*, 15(4):56–63.

Guha, S., Cheng, B., and Francis, P. (2011). Privad: Practical privacy in online advertising. In *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, NSDI'11, pages 169–182. USENIX Association.

Guha, S., Reznichenko, A., Tang, K., Haddadi, H., and Francis, P. (2009). Serving ads from localhost for performance, privacy, and profit. In *In Proc. of the 8th Workshop on Hot Topics in Networks (HotNets 09)*.

Halpern, S. (Dec 22, 2016). They have, right now, another you. http://www.nybooks.com/articles/2016/12/22/they-have-right-now-another-you/. The New York Review of Books.

Hardt, M. and Nath, S. (2012). Privacy-aware personalization for mobile advertising. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 662–673. ACM.

Hartzog, W. and Rubinstein, I. (2017). The anonymization debate should be about risk, not perfection. *Commun. of the ACM*, 60(5):22–24.

ISOC (2017). Your digital footprint matters. https://www.internetsociety.org/your-digital-footprint. Internet Society.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. (2007). Social phishing. *Commun. of the ACM*, 50(10):94–100.

Juels, A. (2001). Targeted advertising ... and privacy too. In *Proceedings of the 2001 Conference on Topics in Cryptology: The Cryptographer's Track at RSA*, CT-RSA 2001, pages 408–424. Springer-Verlag.

Kainda, R., Flechais, I., and Roscoe, A. W. (2010). Security and usability: Analysis and evaluation. In *ARES 2010, Fifth International Conference on Availability, Reliability and Security, 15-18 February 2010, Krakow, Poland*, pages 275–282.

Kramer, A. D. I., Guillory, J. E., and Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790.

Lee, R., Nia, R., Hsu, J., Levitt, K. N., Rowe, J., Wu, S. F., and Ye, S. (2011). Design & implementation of faith, an experimental system to intercept and manipulate online social informatics. In *Int. Conf. on Advances in Social Networks Analysis & Mining*, pages 195–202.

Lee, S., Wong, E. L., Goel, D., Dahlin, M., and Shmatikov, V. (2013). πbox: A platform for privacy-preserving apps. In *Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2013*, pages 501–514.

linkfluence (2017). Social media intelligence for brands and agencies. https://linkfluence.com/en/.

Liu, Y., Gummadi, K. P., Krishnamurthy, B., and Mislove, A. (2011). Analyzing facebook privacy settings: User expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 61–70. ACM.

Manjoo, F. (Apr 5, 2017). The online ad industry is undergoing self-reflection. that's good news. https://nyti.ms/2oHtaWc. New York Times.

Marshall, C. C. and Shipman, F. M. (2017). Who owns the social web? *Commun. of the ACM*, 60(5):52–61.

Narendra K, N. V. and Shyamasundar, R. K. (2016). Decentralized information flow securing method and system for multilevel security and privacy domains. US Patent 9,507,929.

Narendra, N. V. and Shyamasundar, R. K. (2017). A Complete Generative Label Model for Lattice-based Access Control Models. In *The 15th International Conference on Software Engineering and Formal Methods, SEFM 2017, Trento, Italy, September 4-8, 2017*, pages xx–xx. Springer.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown/Archetype.

Patil, V. T. and Shyamasundar, R. K. (2017). Undoing of Privacy Policies on Facebook. In *Proceedings of* 31^{st} *Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec 2017)*, 10359, pages xx–xx. Springer.

Popper, N. (March 23, 2017). Banks and tech firms battle over something akin to gold: Your data. https://nyti.ms/2mTx7ov. New York Times.

Riederer, C., Erramilli, V., Chaintreau, A., Krishnamurthy, B., and Rodriguez, P. (2011). For sale: Your data, By: You. In *Proc. of the* 10^{th} *ACM Workshop on Hot Topics in Networks*, HotNets-X, pages 13:1–13:6. ACM.

Saez-Trumper, D., Liu, Y., Baeza-Yates, R., Krishnamurthy, B., and Mislove, A. (2014). Beyond CPM and CPC: Determining the Value of Users on OSNs. In *Proceedings of the Second ACM Conference on Online Social Networks*, COSN '14, pages 161–168. ACM.

SensorTower (2017). Top Non-Game Apps by Downloads and Revenue - Worldwide, Q1 2017. https://sensortower.com/blog/top-apps-q1-2017. SensorTower.

Taplin, J. (Apr 22, 2017). Is It Time to Break Up Google? https://nyti.ms/2p7Emhp. New York Times.

Upturn (March 2016). What ISPs Can See. Technical Report, Upturn.

WEF (Feb 2013). Unlocking the Value of Personal Data: From Collection to Usage. http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf. World Economic Forum & The Boston Consulting Group.

WEF (May 2012). Rethinking Personal Data: Strengthening Trust. http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf. World Economic Forum & The Boston Consulting Group.

WEF (May 2014). Rethinking Personal Data: Trust and Context in User-Centred Data Ecosystems. http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf. World Economic Forum.

Wilton, R. (2013). The Language of Privacy. http://www.internetsociety.org/es/node/186003. Internet Society.

Zyskind, G., Nathan, O., and Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops*, pages 180–184.