SMU Data Science Review

Volume 1 | Number 4 Article 2

2018

It's Your Data: A Blockchain Solution to Facebook's Data Stewardship Problem

Cynthia Alvarado
Southern Methodist University, alvaradoc@smu.edu

Nithya Devadoss
Southern Methodist University, ndevadoss@smu.edu

Rob Rivens
Southern Methodist University, rrivens@smu.edu

Daniel W. Engels
Southern Methodist University, dwe@smu.edu

It's Your Data: A Blockchain Solution to Facebook's Data Stewardship Problem

Cynthia Alvarado[1], Nithya Devadoss[1], Rob Rivens[1], Daniel W Engels[1]
[1] Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75205 USA
{alvaradoc, ndevadoss, rrivens, dwe}@smu.edu

Abstract. Social Media has quickly become a worldwide phenomenon, acting as a repository for capturing the most intimate levels of human behavior. Many social media companies have discovered ways to exploit users' desire and willingness to share information, by not only mining, analyzing, profiling and selling user's data, but also exposing user's data to malicious cyber-attacks via inadequate and/or antiquated data security protocols. As the line between consent and abuse continues to thin, individuals' right to privacy has become an important topic for lawmakers in many developed nations. As stewards of user's data, social media companies have a responsibility to properly use and adequately protect the data entrusted to them by their user communities. In this paper, we examine the impact of Facebook's irresponsible practices of collecting and sharing user's data with third-parties and quantify the impact caused by its latest data breaches. Lastly, we propose a blockchain solution, which requires consent and release by users, to circumvent the abuse of data for Facebook's own purposes.

1 Introduction

In the age of Internet connectivity, people in developed nations are living in "virtual glass houses" that are completely transparent. Almost every activity is being shared, monitored and/or recorded due to the proliferation of the Internet, including (but not limited to) the duration and quality of one's sleep, what one does upon waking, the products which are used for daily hygiene, what and when meals are being consumed and how one feels about everything from the best places to visit to thoughts and feelings about the political climate. Bolstered by the ease of "sharing" content, this knowledge extends to the entire family, friends, neighbors and co-workers of an individual. This is the universe being captured and memorialized by social media, in particular.

While consumers of social media are liberal in sharing their personal data, most don't consider the risks of doing so. Figure 1 shows that since 2013, there have been over 13 billion worldwide data records lost or stolen at the hands of companies entrusted with securing the data. In fact, every day almost 6.2 million records are lost or stolen. This equates to 262,111 records per hour, 4,369 records per minute and 73 records per second[1].

[1] https://breachlevelindex.com/

At this rate, companies must determine more secure ways of ensuring data security for the records in their possession. At an average estimated cost of almost $4 million USD per breach[1], companies not only have the moral obligation, but also the financial incentive to proactively secure data records.

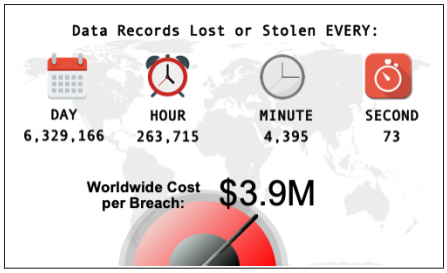

Figure 1: Worldwide Breach Statistics (source: Breachlevelindex.com)

Our approach to understanding data privacy implications began with researching the many recent data breaches that have impacted companies across several industries, worldwide. Our research revealed that Social Media companies have exposed a disproportionate number of consumer's data records in the most recent history, comprising 76% of data records breached in 2018-to-date[1] as depicted in Figure 2.

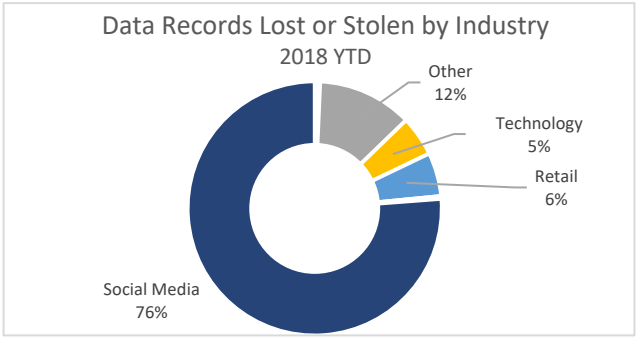

Figure 2: Number of Data Records Lost or Stolen by Industry

Being members of social media communities, we have vested interests in understanding what happens to the data we entrust to social media companies and bear some responsibility for seeking solutions to problems. We also researched the top social media platforms to understand which of them would be best to perform a deep dive analysis. We found that out of the top platforms (Facebook, Instagram, Pinterest, LinkedIn and Twitter) Facebook overwhelmingly had the biggest impact on the social media population, in terms of percent of adults using each platform as seen in Figure 3.

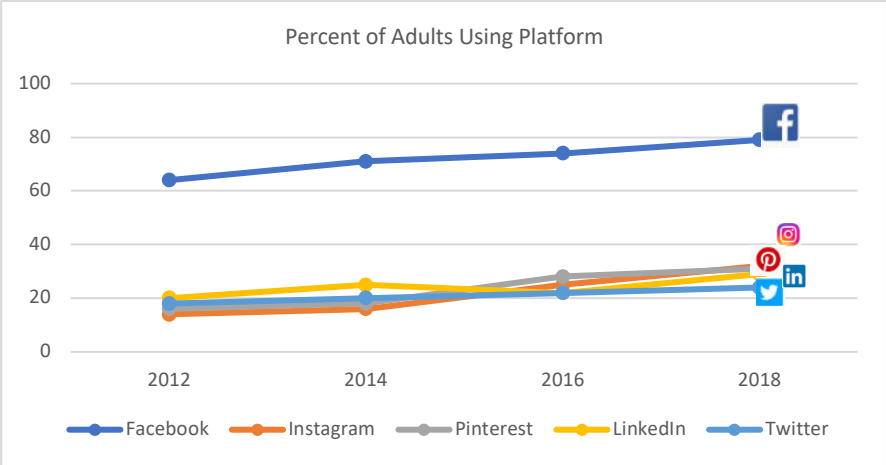

Figure 3: Annual percent of Adults Using Top Social Media Platforms

To ground readers in the basics of Facebook's usage, our paper discusses a brief history of Facebook, as well as the most common uses of the platform. The data presented will help readers understand the magnitude of Facebook's data stewardship role and shed light on the risks resulting from breaches of collected data.

To further define the problem, we began with research regarding Facebook's published Data Policy, Privacy Basics, Privacy Shield, and Cookies & Other Storage Technologies, available via Facebook's Newsroom[2] web page. These documents collectively describe various aspects of Facebook's usage of its member's data, as well as the rights of its user community.

We discuss significant breaches of Facebook's data, one which involved the research firm Cambridge Analytica. We discuss the impact of the breach, which had worldwide implications.

Lastly, we discuss our blockchain solution to Facebook's data stewardship challenges. We examined Private, Public and Hybrid blockchain approaches, deciding that the latter would be best given the structure of social media platforms and the challenges which exist when trying to build a solution for this environment. We also discuss ethical considerations in using blockchain technology. The issue of immutability and the "right to be forgotten" were topics of this discussion.

[2] https://www.facebook.com/about/

2 Understanding the Value of Facebook's Data

Facebook is the largest social media platform in the world, with 2.7 billion Monthly Active Users (MAU) as of September 30, 2018[3]. In terms of its global reach, Facebook captures 38.6% of the global online population, is used by over 5 billion businesses and is the leading advertising platform for both Business-to-Business and Business-to-Consumer companies[2]. Due to its global footprint, it is imperative that Facebook deploys stringent cyber-security methods for protecting the vast amounts of data generated on a daily basis.

2.1 Introduction to Facebook

In 2004, Mark Zuckerberg founded Facebook while studying psychology at Harvard University. Considered a computer programming prodigy, Zuckerberg developed Coursematch and Facemash for fellow students[4]. The former allowed users to see a list of all the students who signed up for a particular course, which allowed them to identify the courses their class-mates were taking, and aided users in choosing desired classmates. The latter allowed users to rate people's attractiveness. These applications formed the basis of Facebook, which allows users to post stories, pictures and commentary, within the bounds of Facebook's guidelines.

2.2 Getting Started

The creation of an account requires a name, email address or mobile phone number, password, date of birth and gender. Users can opt to add personal details such as place of residence, marital status, place of employment, political affiliation, etc. Most users also add a profile picture to aid in identification and recognition by others on the network. Once the profile is created, users can begin adding "Friends", which are connections to other user accounts. Users accomplish connections by sending "Friend Requests," which must be accepted by the other user in order to establish the connection. Users are limited to 5,000 friends, however, one can opt to allow users to "follow" them, which is unlimited.

2.3 Understanding Facebook's Products

Facebook refers to the components of its platform as "Products". The products allow users to interact with various features of the platform, including communication with other users, exposure to advertising by third-parties, and sharing/consumption of content.

Profile. The Profile is the main product in the platform. As the "home page" for an individual's account, it is the first thing other users see when they access the account. The profile contains a main image (if the user has elected to upload one), the user's personal details – subject to visibility settings, and it also serves as the navigation page for the other products offered by Facebook.

[3] https://www.statista.com/

News Feed. The News Feed is a real-time list of content shared by Friends, Pages, Groups and other connections to the platform. Facebook uses algorithms to control the content users see in their News Feed. Based upon exposure to the News Feed, other users are able to interact with content being shared, such as choosing an icon which represents sentiments (e.g. "love," "thumbs up," "thumbs down," "sad face," or "angry face" icons). Users can control the visibility of shared activity through a number of privacy settings.

Messenger. Messenger is the method by which users can privately share content with other users. The Messenger function can access multiple users, enabling group sharing of content. Users are also able to make toll free phone calls worldwide (where an internet connection is available), if they have connected their phone number to their Facebook account. The function displays information regarding a user's availability, such as if they are "active", "inactive" and the duration of their inactivity. Messenger can also import contacts from the user's phone contacts, if such access is granted by the user.

Groups. Groups serve as a virtual "meeting space" providing users the ability to share content and interact with other users as a collective team. Users can access much of the site's functionality such as sharing posts, pictures, files, as well as News Feed updates. Users can be allowed to join by invitation only or the group can be set to allow public access. Groups are used by more than 1 billion people each month, worldwide[2].

Events. Via Facebook Events, users can invite other members to attend scheduled events, manage attendee lists, share content such as maps and media, etc. Events can be private or public. Invitees can be automatically notified if event details change or when others post content in the Event product. Events are used by more than 550 million people, monthly[2].

Video. Each day, users watch over 100 million hours of video via Facebook's Video product. Video is billed as "one of the most engaging and immersive ways to tell your story" according to Facebook's Newsroom web page[2].

Photos. Another of Facebook's main features, Photos are "the most popular place to share photos", according to Facebook. Over 350 million photos are uploaded each day, with the ability to be shared and re-shared at-will. Users are able to upload high resolution photos, on an unlimited basis. Users may also add details to photos such as captions, descriptions and locations. Users can identify subjects in photos via "tagging"[2]. Facebook has image recognition technology that scans faces of subjects in a photo and compares the images to known images in a master database, in order to automatically suggest tagging identified users.

Search. Search allows users to search Facebook across various criteria, such as a user's name, location and place of occupation. Search is not limited to user profiles; users may search for content, such as posts or photos. Users may also search for connections between users and content the user has interacted with, such as "posts liked by" or "photos tagged in". Search allows a user's history to be cleared, however, if a user searches for content previously searched, the search engine will "suggest" the previously searched content in the results.

Pages. Pages are public account profiles that are most often utilized by people who do not want others to have access to their personal Facebook accounts. Most Pages are owned by businesses, public figures or organizations. Owners of Pages can interact with users who post to their Pages, however, they cannot initiate contact with private users.

Facebook's products span a myriad of features enabling users to easily share content with other users. The hidden risk in the utilization of these products is the data which must be collected and stored in order to maintain the functionality of each product. Many personal details are shared via these products, which enables Facebook to have unlimited access to this data in its databases. Facebook is able to mine the data to build user-preference profiles, which are extremely valuable to those seeking to understand or profit from users' online behavior and personal details.

The amount of data collected on Facebook's 2 billion users is staggering. The table below details data generated, either actively by the user through interaction, versus passive data generation that is gathered about the user.

Table 1. Types of user data captured by Facebook.

Active User Generated	Passive Application Generated
Contact information, information you've written in your About You section in your profile, and your Life Events	Your activity on Marketplace
Photos and videos you've uploaded and shared	A history of payments you've made through Facebook
Posts you've shared on Facebook, posts that are hidden from your timeline, and polls you have created	A list of the posts you've saved
Comments you've posted on your own posts, on other people's posts or in groups you belong to	A list of places you've created
The people you are connected to on Facebook	Apps and websites you log into using Facebook and apps you admin
Posts, comments and Pages you've liked or reacted to	Ads topics that are most relevant to you, advertisers who have collected information directly from you and information you've submitted to advertisers
People, organizations or business you choose to see content from, and people who follow you	A history of your searches on Facebook
Messages you've exchanged with other people on Messenger	A history of precise locations received through Location Services on your device
Groups you belong to, groups you manage, your posts and comments	Logs of your calls and messages that you've chosen to share in your device settings
Responses to events and a list of the events you've created	A history of logins, logouts, periods of time that you've been active on Facebook and the devices you use to access Facebook.

2.4 Facebook's Relationship with Third-Party Partners

To understand the vulnerabilities of users with respect to data privacy, it is helpful to understand how third-party partners utilize Facebook's platform in order to advance their own agendas.

Advertisers. The cornerstone of the relationship between Facebook and its advertisers is the advertiser's objective: "what's the most important outcome I want from this ad?". Some examples of objectives are: internet-sourced sales, app downloads, increased brand awareness, etc. Next, advertisers choose the ad objective for the stated goal(s). Thereafter, the target audience is identified, using factors such as age, location or demographics and behaviors that represent the audience. Finally, a budget is set. As one might expect, the larger the budget, the more robust the data collection process will be[4].

Analytics Service Partners. Facebook provides aggregated data to help people and businesses understand how users engage with Facebook's products. This includes posts, listings, Pages, photos, videos and other activities recorded both on and off Facebook's products. Data could be related to the number of accounts that are viewed, interacted with or commented content, including demographic data related to account owners[5].

Partners offering goods and services in Facebook's Products. When users subscribe to premium content or buy something from a seller in Facebook's Products, the content creator or seller can receive public information and other information shared with them, as well as the information needed to complete the transaction (e.g. payment, shipping and contact details) [5].

Vendors/Service providers. Facebook provides data to vendors and service providers who support its businesses (e.g. by providing technical infrastructure services, analyzing how Products are used, providing customer service, facilitating payments or administering surveys) [5].

Researchers and academics. Facebook provides information and content to research partners and academic institutions to conduct research that advances studies as well as innovation that support Facebook's businesses and/or mission. Topics can include general social welfare, technological advancement, public interest, and health/well-being[5].

Law enforcement/Legal requests. Upon request, Facebook shares information with law enforcement or legal entities, subject to certain guidelines[5].

---

[4] https://www.facebook.com/business/learn/facebook-ads-basics
[5] https://www.facebook.com/privacy/explanation/

3 Facebook's Data Policies

With a firm understanding of the depth and breadth of data collected and shared by Facebook – and its intrinsic value – we now present the various policies meant to protect said data from misuse. Facebook has a main document called the Data Policy, which explains the manner in which it collects and uses data. Facebook also shares policies entitled "Privacy Basics", "Privacy Shield" and "Cookies." However, Facebook does not explicitly share a policy explaining its data stewardship practices related to protection from security breaches or malicious attacks.

3.1 Data Policy

The first paragraph in Facebook's Data Policy states "To provide the Facebook Products, we must process information about you. The types of information we collect depend on how you use our Products" [5].

In addition to explaining data collected when users interact with Facebook's products, the Data Policy discusses data with special protections. Data such as religious and political views, health data, and other personal data such as racial or ethnic origin, philosophical beliefs or trade union membership could be subject to special protections under the laws of certain countries[5]. The policy does not elaborate on the protections.

Collection of Device Data. An unexpected use of data described in the Data Policy is the collection of device information. Facebook collects information from and about the computers, phones, connected TVs and other web-connected devices its account owners use that integrate with its products (e.g. data collected about use of our products on a mobile phone could be used to better personalize the content (including ads) or features when using a laptop or tablet). Detailed information is obtained from these devices, including the operating system, hardware/software versions, battery level, signal strength, available storage space, browser type, app and file names and types, and plugins. Information can also be extracted regarding whether a window is in the foreground or background, as well as mouse movements[5].

Network and connections. Facebook's Data Policy grants it access to information such as the name of its user's mobile operator, Internet Service Provider, language, time zone, mobile phone number, IP address, connection speed and, in some cases, information about other devices that are nearby or on the network[5].

Cookie data. Facebook has access to cookies stored on its user's devices, including cookie IDs and settings[6].

[6] https://www.facebook.com/about/basics

3.2 How You're Protected

From the Privacy Basics site, which explains how users can protect their data from abuse by other users, there is a link to a page explaining how Facebook protects data.

Secure Browsing. Facebook notes that it encrypts posts and messages, so that unauthorized entities are unable to access the content[7].

Virus Detection. Facebook deploys tools which can detect if a user has been affected by a virus as a result of accessing a tainted post while using its products. The documentation explains that the tool will help remove the virus, although it does not indicate how[7].

Sharing Personal Data. Facebook states that it will not share personal information without the permission of its users, unless required by law. The statement refers users to the Data Policy for more information[7].

3.3 Privacy Shield

Facebook devotes a section of their Data Policy to the Privacy Shield Principles, which were designed by the U.S. Department of Commerce, the European Commission and the Swiss Administration. The principles provide companies in the US, EU and Switzerland with compliance and data protection requirements related to the transfer of personal data from the European countries to the United States[8].

Facebook's policy describes its adherence to the Privacy Shield Principles in the following three areas:

Access. In accordance with Facebook's commitments under the Privacy Shield, Facebook notes it will work with its Partners to provide individuals access to personal data about them that Facebook holds on behalf of its Partners, taking reasonable steps to enable individuals to correct, amend, or delete personal data that is demonstrated to be inaccurate[9].

Third Parties. Facebook explains that it may transfer data within its family of companies and to third parties, including service providers and other partners. In accordance with Privacy Shield Principles, Facebook notes that it is "liable for any processing of personal data by such third parties that is inconsistent with the Privacy Shield Principles unless Facebook was not responsible for the event giving rise to any alleged damage" [9].

[7] https://www.facebook.com/about/basics/stay-safe-and-secure/how-youre-protected
[8] https://www.privacyshield.gov/welcome
[9] https://www.facebook.com/about/privacyshield

Legal Requests. Personal data transferred to Facebook by its Partners may be subject to disclosure pursuant to legal requests or other judicial/government process, including subpoenas, warrants, or orders[9].

Enforcement. Facebook's policy on Enforcement states that its compliance with the Privacy Shield Principles is subject to the investigatory and enforcement powers of the U.S. Federal Trade Commission[9].

4 Facebook's Data Breaches

Facebook's title as the largest social media company also means it is the largest target for hackers and cyber criminals. This has proven to be true, in that Facebook has now been the subject of the two largest data breaches in the history of social media, compromising over 2 billion data records[1].

Figure 4 depicts the history of data breaches experienced by the top ten social media companies. Facebook has been targeted multiple times, however its most noteworthy breaches occurred in 2015 and 2018[1].

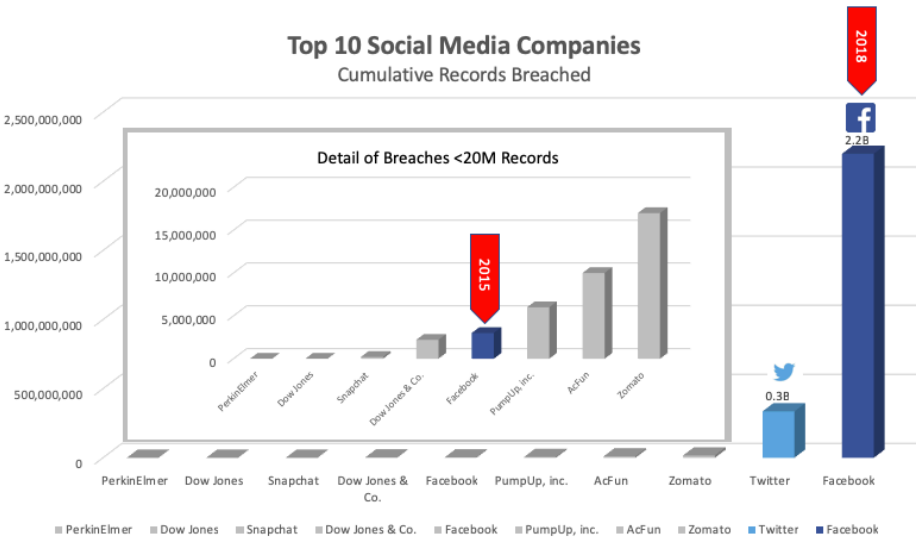

Figure 4: Cumulative Records Breached per Occurrence for Top Ten Social Media Companies

4.1 Cambridge Analytica

In 2013, a Cambridge University researcher named Aleksandr Kogan created a personality quiz app called "This is Your Digital Life". The app was initially installed by around 300,000 people who shared their data as well as some of their friends' data.

Due to the structure of Facebook's platform at the time, Kogan was able to access tens of millions of the app user's friends' data. Unbeknownst to Facebook, Kogan shared these data with the data mining firm Cambridge Analytica[10]. This gave birth to one of the largest data breaches affecting social media companies in history.

The Cambridge Analytica scandal was no ordinary breach. The firm was hired by then U.S. presidential candidate Donald Trump's 2016 election campaign. Gaining access to private information on more than 50 million Facebook users, the firm offered tools that could identify the personalities of American voters, with the intent of influencing their behavior in order to affect the outcome of the 2016 U.S. Presidential election. Cambridge Analytica executives were also caught offering to entrap politicians on video[10].

The data that was breached included details on users' identities, friend networks and "likes." The firm's intent was to map personality traits based on the content users interacted with on Facebook, with the ultimate goal of using that information to target audiences with digital ads. Facebook maintained that no passwords or "sensitive pieces of information" had been compromised, although information about user's locations were made available to Cambridge Analytica[10].

Facebook later revised the estimated number of users affected by the Cambridge Analytica breach to up to 87 million people, mostly in the US[11].

4.2 Token Attack

Still attempting to recover from the Cambridge Analytica attack, which continued to make headlines due to the revelation of more accounts being compromised than previously estimated, Facebook became the target of another attack on September 25 of 2018. In this attack, Facebook discovered that attackers exploited a vulnerability caused by bugs in Facebook's access token process[12]. The access tokens are similar to digital keys that allow a user to remain logged into Facebook, without the need to re-enter their password at every log in instance.

To protect users while Facebook conducted their investigation, they invalidated the access tokens of almost 90 million accounts that were identified as being potentially impacted by the vulnerability. Facebook later determined that between September 14 and 27, the attackers used access tokens to steal certain Facebook account information from the platform by exploiting the "View As" function. The View As function allows users to type in another user's account name to gain a view of how that user sees their account page, given the security settings of the account[12].

The attackers initially controlled a small set of accounts, which were connected to other Facebook friends. Subsequently, they used an automated technique to move from account to account, stealing access tokens of others in the user's network[12].

The attackers accessed two sets of information affecting 15 million users, which included the user's name and contact details (e.g. phone number, email, or both). For an additional 14 million people, the attackers accessed the same set of information

[10] https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html
[11] https://newsroom.fb.com/news/2018/04/restricting-data-access/
[12] https://newsroom.fb.com/news/2018/10/update-on-security-issue/

described above, in addition to other details such as username, gender, language, relationship status, religion, hometown, current city, birthdate, device types used to access Facebook, education, work, the last 10 places they checked into or were tagged in, website, people or Pages they follow, and the 15 most recent searches[12].

4.3 Facebook's Response to Recent Attacks

In light of attacks on its platform, Facebook has implemented a number of changes in Application Program Interfaces (APIs) in order to restrict access to its data. Facebook's Chief Technology Officer, Mike Schroepher, stated "we believe these changes will better protect people's information while still enabling developers to create useful experiences" [13].

Events API. Prior to the security breaches, users could grant an app permission to obtain information about events they host or attend, including private events. While this made it easy to add Facebook Events to calendars, ticketing or other apps, Facebook Events contain information about other user's attendance as well as posts on the event wall, therefore creating a data security vulnerability. In response, Facebook no longer allows APIs access to Event walls[13].

Groups API. All third-party apps using the Groups API now require approval from Facebook and the Group's admin. Additionally, apps no longer have access to the member list of a group. Lastly, Facebook removed personal information (e.g. names, profile photos, posts or comments) that apps can access[13].

Pages API. Prior to the change, any app could use the Pages API to read posts or comments from any other Page, allowing developers to create tools for Page owners to help them schedule posts and reply to comments or messages. As a result of the change, all future access to the Pages API requires approval by Facebook[13].

Facebook Login. Facebook must approve all apps that request access to information such as check-ins, likes, photos, posts, videos, events and groups. Apps no longer have access to personal information (e.g. religious or political views, relationship status/details, custom friend lists, education/work history, fitness activity, book reading activity, music listening activity, news reading, video watch activity, etc.)[13].

Data Providers and Partner Categories. Partner Categories, a product that lets third-party data providers offer targeting directly on Facebook, has been removed[12].

Search and Account Recovery: Prior to the change, users could enter another person's phone number or email address into Facebook search to locate the user's account. The function has now been disabled, as it was routinely identified as a means of exploiting accounts[13].

App Controls. Facebook now shows a link at the top of their News Feed containing the apps they use as well as the information they have shared with those apps. Users

[13] https://newsroom.fb.com/news/2018/04/restricting-data-access/

can remove apps that they no longer want. This section will also indicate if a user's information may have been improperly shared with Cambridge Analytica[13].

Although Facebook has taken steps to tighten the control it allows third-parties to have over its data, there are still vulnerabilities on the platform. Tightening controls still requires human intervention in the form of "approvals", which are only as reliable and trustworthy as the person(s) granting the approvals. We believe blockchain technology is the answer to remove the human "trust" element from the equation.

5 The Basics of Blockchain Technology

As discussed, the problem with social media boils down to a question of trust. Do the billions of social media users trust Facebook to provide due diligence in "authorizing" APIs to access their personal data? Has Facebook's track record been a good indicator of the level of trust users should grant them? By contrast, trust-less security systems are criteria of blockchain-based solutions. Blockchain technology has gained popularity primarily among industries concerned with cybersecurity and finance, due to its ability to execute secure and decentralized transactions. In fact, many industries are beginning to implement blockchain technology due to its ability to enhance the trust-factor for transactions[14].

5.1 An overview of the blockchain

A blockchain is a public ledger of information collected through a network that sits on a distributed network. It has gained popularity because of its transparency, data integrity, auditing, and its decentralized nature. Figure 5 depicts use cases for blockchain technology.

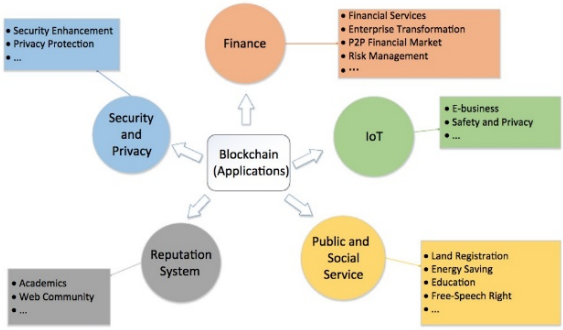

Figure 5: Industries with Use cases for Blockchain

A blockchain is a peer-to-peer (P2P) public ledger of information collected through a network that sits on a distributed network. Distributed ledger technology is a distributed database with several nodes. Each node replicates and saves an identical

[14] https://www.ibm.com/blockchain

copy of the data in the network. Users on each node are allowed to make updates and then the updates are validated and voted to be accepted by all the nodes in the network. This voting mechanism and the rule by which one final copy of the update is accepted is called a consensus, which is automatically conducted by the consensus algorithm. Figure 6 depicts the structural distinction between centralized and decentralized architectures.

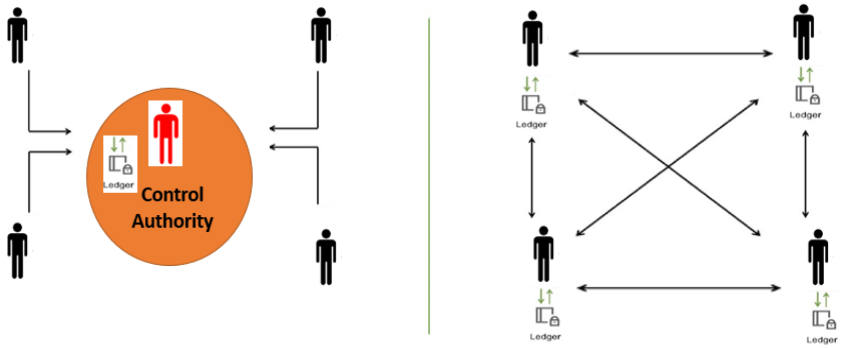

Figure. 6: Centralized (left) typically consists of a single central authority while Decentralized Architecture (right), shares ledger information between all participants.

A blockchain is a sequence of blocks that contains the footprint of the parent block and the actual data, along with the digital hash of the current block. Blockchains also contain a digital signature, which validates identity to ensure changes are from a trusted source using cryptographic hash. The basic structure of a blockchain is shown below in Figure 7.

Figure 7: Basic structure of a Blockchain Transaction

5.2 Blockchain Architecture

A blockchain is a sequence holds a complete list of transaction records such as a public ledger (Lee Kuo Chuen, 2015). Figure 8 illustrates an example of a blockchain. Each block links to the previous block through a hash value of the previous block, called the parent block. This is important to maintain the ordering of the blocks and is also the reason why blocks cannot be tampered. The first block of a blockchain is called the genesis block which has no parent block, hence the data in the genesis block alone is used to generate the initial values[15].

[15] An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. 10.1109/BigDataCongress.2017.85.

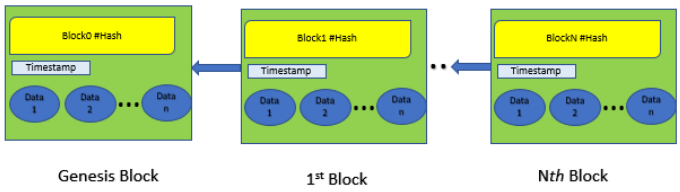

Figure 8: An example of how individual blocks are chained together to create a blockchain

Blockchain uses digital signatures based on asymmetric key cryptography for validating the different users in the network. Users utilize a public-private key pair to authenticate and validate one another (Stallings, 2014). Only users who are authenticated and have the public key of a user, are able to read the encrypted data sent over the network. Other users who don't have the public key of the user who sent the data can only pass the data through the blocks and will be unable to read it. This key also ensures that the data is sent by the actual user. There are various papers on digital signatures to increase the security of the blockchains such as the Elliptic Curve Digital Signature Algorithm (ECDSA) as used by bitcoin, and aggregate signatures (Stallings, 2014).

5.3 Types of blockchains

There are particular features that make blockchains unique as distributed ledgers. First, blockchains have a growing list of records. Second, altering or deleting records is very difficult because any changes within the block modifies the hash value, thus breaking the chain. This means that blockchains are more secure due to the digital hash and digital signature that validates previous blocks. Lastly, all activity in the network is logged. Blockchains are broadly classified as three types: public, private and hybrid models[16].

Public blockchains. Public, also called as permission-less blockchains, allows any user to set up a node in the network and have the privilege to read and write blocks of data. Any new user can start adding data blocks to the ledger without having to seek approval. They also have copies of the ledger history, as a copy is automatically stored in all the nodes. Discrepancies are to be resolved together as a network. Public blockchain allows the user to be anonymous but maintains transparency of the transactions. As long as the transaction is valid, it will move through the chain. Examples of existing public blockchains are Bitcoin, Litecoin and Ethereum.

Private blockchains. Also referred to as permissioned blockchain, are blockchains where each and every user in the network needs to be identified and trusted. Only users who are granted access to write using the access control layer (ACL) will be able to write. Unlike public blockchain, not everyone is granted permission to read the data inside the block, however, they are able to validate the blocks. For this reason, this kind

[16] https://bitsonblocks.net/2015/09/09/gentle-introduction-blockchain-technology/

of blockchain is not truly decentralized. The Linux Foundation hosts a permissioned blockchain called the Hyperledger Fabric[17].

Hybrid blockchains. Also called a consortium, allows only a group of users to have public-level access, which means transparency is limited to a certain group with permissioned access. An entire copy of the blockchain is available only to certain trusted nodes, hence it is partially decentralized. Implemented hybrid blockchains include the payment protocol Ripple, and the firm R3, which supports distributed ledger for more than 200 firms.

The main distinction between a consortium and private blockchain is that a consortium will have a group of entities who maintain consensus, whereas private blockchain models have a single entity maintaining the consensus, leaving it with the possibility of being exploited by the entity or organization. Comparatively, private and consortium models are faster since write and block validation is allowed only by preselected nodes or groups.

5.4 Blockchain in Social Media

Currently, there are a few, albeit small start-up social media companies that have implemented blockchain such as Steemit and Sapien. However, because they are small, they use public blockchain. The characteristics of hybrid blockchain, in particular, digital signature, immutability and permission using ACL as discussed earlier, make it a viable solution to address concerns regarding Facebook's recent data breaches.

There are various blockchains that specialize in certain areas of interest in cybersecurity. For example, Civic is used for identity verification, Prover for video authentication and verification, and User Feeds for news verification. While each one of these excels in a specific aspect of cybersecurity, we propose a hybrid blockchain structure that incorporates specialized blockchains such as the aforementioned entities to solve the lingering issue that persists in social media and achieve the best solution.

Each block in the blockchain is only available to a certain group of users who have permission to view the content of the block. This can be achieved using the access control layer in permissioned blockchain. The access control information can in turn be stored in another consortium blockchain that allows it to be audited at a later point in time. For instance, user A grants access to user B initially, but is able to revoke access at a later point. Hence, this information needs to be stored separately. When a user deletes his or her profile, they can revoke access to users with the help access control layer thus revoking access to their information for any other purposes without their knowledge.

There are a couple of challenges with a blockchain based solution, such as scalability and performance, and are interrelated. Scalability is the ability to allow more users to join the network while the chain grows in length. This is an ongoing issue with most blockchain models currently in use. One solution could be to store the hot and cold data separately. When a node joins the network, the entire history of the chain is downloaded to the node allowing it to act as another peer in a P2P network. However,

[17] https://www.hyperledger.org/projects/fabric

in social media application, the cold data that is less frequently used can be stored separately to allow better performance in the network. Aggregate signatures also improve performance by aggregating the blocks before certifying them, thus improving the speed[18]. Separately, another bottleneck in blockchain is reaching a consensus, which is both resource- and time-intensive. The inclusion of separate specialized blockchains to be interconnected to the hybrid model helps alleviate the bottleneck, since each validation node is not participating in reaching a consensus.

Table 2. Problems and Solutions

Problem	Solution
Identity Theft/Fake identity	Civic Asymmetric key cryptography using public private keys and digital signatures.
Fake news	Dedicated public blockchain user feeds
Fake Video feeds	Dedicated public blockchain Prover for video authentication
Private Message being shared with third parties	Using public private keys, only intended users are able to read the message
Data access after user deletion	Access can be revoked using ACL except for the users who are authorized to view the data even after deletion
Data shared to third parties /transparency	Not possible without user permission. Blockchain data encryption in blocks and use of ACL makes it less possible. In a hybrid blockchain there would be other control bodies as private entities and not one entity being entitled to user's data

6 Ethical Considerations

This paper aims to solve deficiencies found in social media companies' ethics standards. While social media does not fall into software engineering, it does share certain intellectual property traits as software companies who do follow ethics guidelines from one of two entities. Both social media and software companies provide a digital service for its users. The two main entities that address ethics in computing. The Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers' (IEEE) Software Engineering Code of Ethics.

For social media companies, the main aspect of ethical business operation pivots around informed consent. When signing up for an account, terms, conditions, and policies should be made clear and concise using common language. Further, it should

[18] https://doi.org/10.1155/2017/8295275

be made available to the user prior to signing up for services. In order for companies to ensure the consent was indeed adequately provided to the end user, the terms and conditions can be broken down to allow users to interact on the screen before moving on to the next section. The interaction may be a check box, or by prompting for the user's initials.

Moreover, highlighting pertinent parts of the terms and conditions that pertain to implied consent in contrast to expressed consent when a user creates an account would allow new users to have the disclosure from the point of initially creating an account. Expressed consent differs from implied consent in that specific permission is given verbally or in writing. In the context of social media, informing customers that their status posts or messages constitute expressed consent, since the user is providing such information, should be made apparent to the end user. Lastly, as terms and conditions change, the new agreement should be made available to new and existing customers and follow the same interactive prompts to ensure that end users are aware of changes.

Conversely, implied consent is the ability to revoke the consent previously established. Giving users the ability to delete content can be made available, where the information regarding the process of revoking the consent is easy to access and simple to understand. For instance, either providing a form to remove information from a company's database whereby the form is easy to view and access. Similarly, for written content, the post can include a sub menu to give the end user the ability to delete the content without the use of a form. For the scope of using hybrid blockchain to keep data secure, the data deleted by the users can be removed by removing the private key that pertains to a specified blockchain when a user requests or revokes consent.

Finally, there is the matter of feasibility to social media companies to incorporate establishing the agreement with the end user when customers first sign up for a service, or as the terms and conditions change, as well as the revocation of these established permissions. Obtaining informed consent from end users does not exceed the cost to maintain a user's account and has been incorporated by software companies for decades. Similarly, forms and sub-menus to give users the option to revoke permissions can be programmatically incorporated as well as an included feature to the social media interface.

7 Conclusions

While social media companies cannot control user-provided personal information, they have the corporate responsibility to do more to mitigate data breaches when passing information back and forth between third parties as they do so for marketing purposes. We examined how a hybrid blockchain model, compared to other blockchain types, is able to provide that end-to-end security when transferring data to third parties and partners. Further, we discussed the permissions that can be established with the hybrid blockchain model by allowing end users control with the use of private keys. In turn, this will ultimately create a more secure method of using social media, while

ensuring the content provided by the end user is visible to the intended audience specified by the user.

By using the hybrid blockchain model along with focused blockchains for monitoring news feeds, photo and video authentication this will remedy not just the permissioned sharing for a user's content, but it will also alleviate fake news feeds and videos that were used on Facebook. Since Facebook does not directly share content with advertising companies, having the hybrid blockchain implemented would not affect Facebook's source of revenue.

References

1. Zheng, Zibin & Xie, Shaoan & Dai, Hong-Ning & Chen, Xiangping & Wang, Huaimin. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. 10.1109/BigDataCongress.2017.85.
2. Dr. Xue-ming Si (August 2017). Research on a New Signature Scheme on Blockchain Volume 2017, Article ID 8295275, 1 page, from https://doi.org/10.1155/2017/8295275
3. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. In: Cryptography Mailing list at https://metzdowd.com. (03 2009)
4. JIPEL: Micheal J. Kasdan. "Is Facebook Killing Privacy Softly? The Impact of Facebook's Default Privacy Settings on Online Privacy", IN: (Spring 2011), Retrieved from https://jipel.law.nyu.edu/ledger-vol-2-no-2-6-kasdan/
5. Kristoufek, L.: What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. PLOS ONE 10(4) (04 2015) 115 3. HYUNYOUNG, C., HAL, V.: Predicting the present with google trends. Economic Record 88(s1) 29 4. Ettredge, M., Gerdes, J., Karuga, G.: Using web-based search data to predict macroeconomic statistics 5. Miraz, M.H., Ali, M.: Applications of blockchain technology beyond cryptocurrency. CoRR abs/1801.03528 (2018)
6. D. D. F. Maesa, A. Marino and L. Ricci, "Uncovering the Bitcoin Blockchain: An Analysis of the Full Users Graph," 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, 2016, pp. 537-546
7. P. Tasatanattakool and C. Techapanupreeda, "Blockchain: Challenges and applications," 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018, pp. 473-475.
8. H. Halpin and M. Piekarska, "Introduction to Security and Privacy on the Blockchain," 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Paris, 2017, pp. 1-3.
9. Catalini, C., Gans. J. "Some Simple Economics of the Blockchain", MIT Working Paper
10. Walch, Angela, The Path of the Blockchain Lexicon (and the Law) (March 24, 2017). 36 Review of Banking & Financial Law 713 (2017). Available at SSRN: https://ssrn.com/abstract=2940335
11. Stallings, William. *Cryptography and Network Security: Principles and Practice*, 6th edition. Upper Saddle River, NJ: Pearson, 2014.