E2-Chat: A Web-Based End-to-End Encrypted Messaging Service
(Technical Paper)

Individuals, Businesses, and Governments: Framing Public Policy for Internet Data Privacy
(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Saiteja Bevara
Fall, 2020

Technical Project Team Members
Ashwin Pathi
Phillip Phan
Rithik Yelisetty

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature  Date 11/24/2020
Saiteja Bevara

Approved _____ Date _____
Yixin Sun, Department of Computer Science

Approved _____ Date _____
Rider Foley, Department of Engineering and Society

Internet Data and Privacy

As of 2018, an estimated 2.5 quintillion bytes of data were created each day across the Internet, and the amount of data generated between 2016 and 2018 alone accounted for 90 percent of all data in the world (Marr, 2018). The amount of data on the internet is increasing, and controlling access to this information and its potential uses is a growing concern. In the United States (US), roughly eight in ten adults believe they have little or no control over the data that companies and the government collect, and a large majority believe that the risks of collection outweigh the benefits (Auxier et al., 2019). Their concerns were in the global spotlight when Facebook admitted that political consulting firm Cambridge Analytica had secretly collected the information of nearly 87 million users of Facebook, using the information to sway voters during the 2016 presidential election (Kang & Frenkel, 2018).

Concerns over protecting internet data are at the heart of user privacy, which can be defined as the ability of an individual to control their personal information and the manner in which it is communicated with others (Hong & Thong, 2013, p.276). The issue of how to protect internet data privacy has become more important recently within the context of the amount of data being generated and the revelations of its misuse. This has led to calls for stricter policy, as current data privacy legislation in the United States is non-comprehensive and narrow in scope (Mulligan, Freeman, & Linebaugh, 2019). However, when considering interactions with internet data, other actors exist, namely companies who have economic interests and governments who are invested in using it to protect national security. Thus, the issue of crafting public policy to address internet data privacy is complex and requires further analysis, as it must be understood from the perspectives of all relevant groups.

In addition to policy, technologies themselves can also be designed to protect internet data privacy, such as end-to-end encrypted (E2EE) services. E2EE services allow for the encryption of text during communication, such that content is only readable by communicating parties. There is much discussion on applications of E2EE specifically for instant messaging platforms. Instant messaging services offer real-time communication and are widely popular. They are also important technologies as they have been shown to encourage daily conversation and promote relationship maintenance among users (Ramirez & Broneck, 2009, p.292). However, most instant messaging platforms that utilize E2EE, such as the popular platform WhatsApp, interface explicitly through smartphones to enable encryption (Greenberg, 2020). Those without access to smartphones are relegated to using unencrypted mediums of communication for instant messaging. This is significant particularly in emerging economies such as India and Brazil, where more than 60% of adults use the internet but only 45% have smartphones, indicating a divide between those who may have access to web-based messaging services but not to smartphones (Silver, 2019).

This project will attempt to build an entirely web-based E2EE messaging service which eliminates the need for an auxiliary mobile device, making encrypted communication more accessible. Further, it will explore the issue of framing public policy around the broader issue of internet data privacy to protect the interests of individuals, businesses, and governments.

E2-Chat: A Web-Based Approach to End-To-End Encrypted Messaging

End-to-end encryption (E2EE) is the process by which messages or other text is encrypted, or converted from readable plaintext to uninterpretable ciphertext when content is stored or in-transit. Unauthorized third parties only have the ability to obtain the encrypted content, which makes E2EE services less vulnerable to outside attacks. E2EE instant messaging

allows users to send and receive these encrypted messages in near real-time. Popular instant messaging platforms today include WhatsApp, Facebook Messenger, and WeChat, and each has over a billion monthly active users (Clement, 2019). E2EE is not a default setting on the majority of these popular messaging services.

WeChat, owned by the Chinese company Tencent, only applies symmetric AES encryption, and encrypts messages between servers and clients individually but not end-to-end. This form of encryption allows controlling parties to still have backdoors into the data to access and read messages in their plaintext form. In the case of WeChat, this has led to surveillance and censorship activity by the Chinese government (Chen, Clayberg, & Li, 2019). Facebook Messenger also does not offer any default end-to-end encryption on messages besides a "secret conversations" feature which allows for temporary encrypted messaging (Greenberg, 2020).

WhatsApp, another platform owned by Facebook, is one of the only popular messaging services with complete end-to-end encryption. At a high level, the encryption protocol itself is based on the Signal Protocol, which generates key pairs for users upon registration and uses them to create master secrets that define encrypted messaging sessions for each chat ("WhatsApp Encryption Overview", 2020). WhatsApp also allows for a wide range of features including group chats and media transfer. However, this application runs on smartphones. Desktop and web versions of WhatsApp exist, but they require smartphones to interface through to proxy encrypted messages (Greenberg, 2020). Thus, there is currently no application which offers encrypted instant messaging that does not require smartphones or other specific hardware.

This project aims to build an entirely web-based E2EE instant messaging service (E2-Chat) with functionality for one-to-one chats, group chats, and various media or file transfer. This web-based service would allow for encrypted communication for people who lack access to

smartphones, but perhaps have other internet access such as through a shared or public computer. Furthermore, it can benefit businesses that require an independent and portable encrypted channel for communication internally or with customers. In this manner, E2EE is extended to a larger population that otherwise would not have access to encrypted communication.

The project will rely on the Rivest-Shamir-Adelman (RSA) cryptosystem as the public-key encryption scheme. Public key encryption schemes are processes used for encryption and decryption of text. Among existing public-key encryption schemes, the RSA system is both widely used and sufficiently secure. Categorized as an asymmetric scheme, it relies on a pair of public and private keys to encrypt and decrypt messages. It guarantees that key generation is efficient for computation, it is impossible to derive the private key from the public key, and it is infeasible to obtain the plaintext from only the encrypted text and the associated public key. This encryption system is reasonable for this project as it allows users to generate key pairs efficiently upon registration which can be used to securely encrypt and decrypt messages (Meelu & Meelu, 2012).

The overall scheme of the project is as follows (see Figure 1). Upon registration each user will be assigned a known public key and a secret private key. For each new one-to-one or group chat that is created there will be a new key pair which is generated specifically for that chat. This chat public key will go in the public-key repository and the chat private key will be sent to each chat member, encrypted using the user's respective public key so only they can decrypt it. All messages to the chat will then be encrypted using the chat public key and all recipients in the chat will use the chat private key to decrypt and view the messages. These private keys can be stored locally, directly in the browser or on some other portable storage, which means the entire process can take place through a web browser without the need for a smartphone.

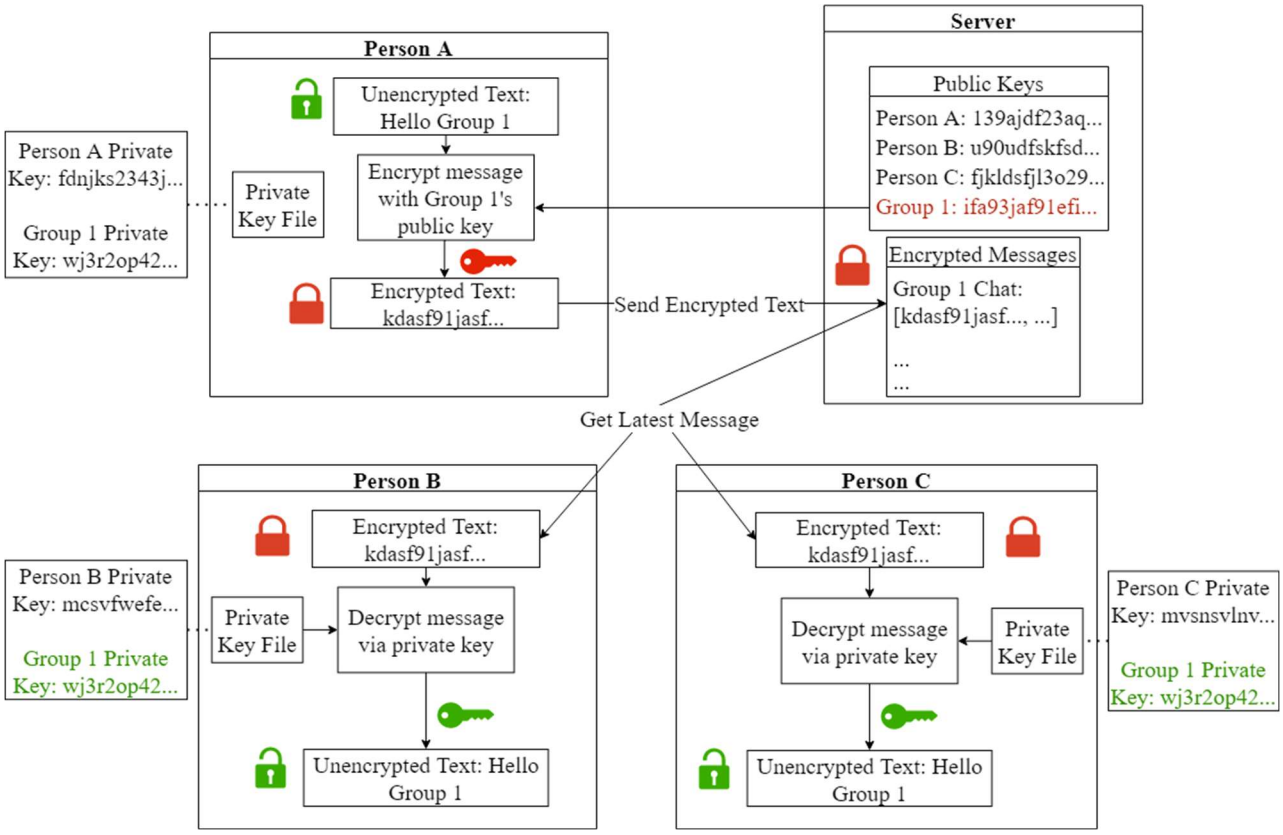

Figure 1. Example Data Flow of Encrypted Messages with Three Parties (Created by Bevara et al., 2020).

The application will be built using GraphQL, Node.js, and React. GraphQL is a querying language developed by Facebook which can be used in this project to query the server for information such as public keys and user messages. GraphQL has the advantages of flexibility with regards to data types and formats, and reduced overhead by removing unwanted data in query results. Due to the expected large amount of data expected to be communicated, GraphQL is an ideal querying framework for this project. React and Node will comprise the front-end and back-end components, as they are both easy to develop and work well with GraphQL (Jeon & Hwang, 2019).

Framing Public Policy for Internet Data

As the amount of data available on the internet increases, it is essential for data privacy policy to keep pace and for technology to evolve. A report by the Congressional Research Service (2019) analyzed the landscape of current privacy legislation within the United States. Findings showed a "patchwork" of individual legislation that was technical and complex, and was targeted at specific fields and industries or specific categories of data. For example, the Gramm-Leach-Bliley Act which targets financial institutions or HIPAA which applies to personal health information. These industry specific acts demonstrate a lack of comprehensive legislation at the federal level (Mulligan, Freeman, & Linebaugh, 2019). However, the human and social dimensions of encryption and internet data reveal the concerns of several major social groups, namely individuals, businesses, and governments, who interact with internet data in different manners.

Individuals, in this context representing regular users of communication services, social platforms, or other internet-enabled devices, use technologies and social platforms on a near-daily basis, and the data they generate is a major privacy concern at a personal level. From their perspective, internet data introduces considerations for a human "right to privacy" and threatens their individual confidentiality and autonomy (Miller & Weckert, 2000, p.257). For these users, individual human rights are the foundation for values they look to uphold, and data privacy directly relates to this interest. Privacy itself as an independent human right is widely argued, but regardless, data privacy can be thought of as a protector of individual autonomy. Autonomy then enables human rights and civil liberties such as freedom of expression and religion. Therefore, with the many ways in which the internet has become an intrinsic component of everyday life for individuals, it is apparent that protecting internet data is critical in defending basic human rights, transcending online identities and impacting real-world experiences (Bernal, 2014).

Businesses are another social group that shape data privacy systems, and represent those who control the major platforms which ultimately collect and manipulate user data. In most cases, the collection of information and data is directly relevant to their financial business model, leading to questions over the monetary value of internet data. The value of data is especially clear when viewing advertising revenues for large corporations who rely on data for targeted content. In a single quarter of 2019, Google and Facebook earned $32.6 billion and $16.6 billion in advertising, respectively (Baca, 2019). Corporations may also use internet data for beneficial tasks, such as during the COVID-19 pandemic to track the spread of the virus or to better understand societal inequalities which contribute to higher rates of sickness in communities of color (Brill, 2020). An ability to maintain economic incentives and pursue specific user-data related tasks, albeit beneficial or harmful, remains the current focus for companies in collecting and processing internet data.

Finally, governments interact with internet data largely for national security. Internet data provides a source for governments to surveil and track security threats. From a regulatory perspective, it can be said that the interests of individual privacy disagree with national security concerns. For example, with end-to-end encrypted technologies such as the one this technical project aims to create, there is a concern that such platforms enable criminals to communicate freely. Governments, who require user data to protect national security, have no method of obtaining unencrypted content from these E2EE services without technical backdoors which breach user privacy (Endeley, 2018). This was apparent in one specific instance following the 2015 Paris attacks, as the Obama administration designated encryption technologies as platforms for adversaries to freely communicate (Sanger & Perlroth, 2015). Governments thus view

internet data as a significant source in identifying security threats, and data privacy becomes a hindrance to these tasks.

To first understand how internet data is defined and constructed by individuals, companies, and governments, the Social Construction of Technology (SCOT) framework will be employed. This framework focuses on the *interpretive flexibility* afforded to technologies as they develop. For any technology, there are social groups who will give unique meaning and function to it based on the group's perspectives and values. Based on these interpretations, problems can be identified that each group hopes to address relating to the technology. Solutions and alternate designs result from these various problems, building a *multi-directional model* of artifacts, social groups, problems, and solutions that display the evolution of technologies and considerations in its development. Eventually, the design of the technology can stabilize and reach closure (Pinch & Bijker, 1984).

The framework first requires the identification of relevant social groups who interact with the system, and through this interpretive flexibility define the technology based on their perspectives and values. In this case, the relevant social actors will be individuals, businesses, and governments, who based on the aforementioned descriptions give different meanings to internet data. The framework then requires the building of a *multi-directional model,* with problems and solutions identified for each of the social groups. For individuals, the problems relate to protecting user privacy. For businesses, the problems are economically or politically motivated. For governments, their concern is protecting national security. From both a technical perspective and policy standpoint, solutions to these problems vary for each group. Solutions can include access to more privacy-focused services such as E2EE technologies, stricter policy regulation on data collection, or a legal requirement for technical backdoors that allow

governments to access data. Applying SCOT will therefore help in fully understanding the considerations in the social construction of internet data and the competing interests of different actors. These considerations will be used to properly frame comprehensive public policy that balances the interests of all relevant social groups.

Research Question and Methods

Internet data privacy requires comprehensive policy and regulation that considers the interactions of individuals, businesses, and governments with internet data. The amount of data being generated, collected, and potentially misused is increasing, and competing interests among participating social groups threaten the landscape of internet data privacy. This motivates the following research question: How can internet data privacy policy be framed around the interests of individuals, companies, and governments?

This question will primarily be researched through policy analysis of secondary sources in existing policy documents. First, the current landscape of data privacy policy will be analyzed within the United States. Current acts such as the Gramm-Bliley Act which specifically focuses on financial institutions and HIPAA which focuses on public health information will set the foundation for legislation that is targeted at specific industries or categories of data. Two specific policy documents will then be viewed which are more comprehensive: the California Consumer Protection Act (CCPA) and the General Data Protection Regulation (GDPR). The CCPA specifically applies to companies that collect the personal information of Californians. At a high level, it provides consumers with three main rights: the right to know, the right to opt-out (of the sale of their information), and the right to delete. The GDPR on the other hand is a model comprehensive legislation from Europe, generally identifying the right to personal data protection of individuals from all potential interferences (Mulligan, Freeman, & Linebaugh,

2019). Furthermore, privacy policies of individual companies and international statements issued by governments will provide more sources.

These policy documents will be analyzed and compared to determine how they have afforded value to, and conformed to the interests of, individuals, companies, and governments and the meanings each of these groups have placed on internet data. Content analysis of specific policy decisions in the context of varying interpretations identified through SCOT will provide insight into how to best frame comprehensive internet data privacy policy. First, policy documents, including the GDPR and CCPA as well as HIPAA and the Gramm-Bliley Act, will be collected. Then, these documents will be read and relevant portions of it identified for further analysis by February 23, 2020, and these documents will then be compared through policy analysis. Finally, the paper will be completed by April.

Conclusion

Internet data is being generated at an incredible rate, and this data is being collected and processed by numerous parties. Potential and observed misuse of user data has led to calls for the identification of methods to protect internet data privacy. This project will identify a technical solution, by researching and building a fully functional web-based end-to-end encrypted instant messaging service that allows users to communicate information securely and independently. All messages and other data stored on the server will be confirmed to be encrypted within the database. The project will also identify methods to frame comprehensive internet data policy around the interests of relevant actors of individuals, businesses, and governments. This understanding will help in recommending an appropriate course of action when implementing federal public policy for internet data.

References

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019, November 15). Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Retrieved from Pew Research Center: Internet & Technology website: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

Baca, M. C. (2019, October 14). What you do on the Internet is worth a lot. Exactly how much, nobody knows. *Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2019/10/14/what-you-do-internet-is-worth-lot-exactly-how-much-nobody-knows/

Bernal, P. (2014). *Internet Privacy Rights: Rights to Protect Autonomy*. Cambridge, United Kingdom: Cambridge University Press.

Bevara, S., Pathi, A., Phan, P., & Yelisetty, R. (2020). Example Data Flow of Encrypted Messages. [Figure 1]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Brill, J. (2020, October 16). Why privacy is essential to equitable recovery. Retrieved from Microsoft on the Issues website: https://blogs.microsoft.com/on-the-issues/2020/10/16/privacy-laws-open-data-economic-recovery/#_ednref1

Chen, M., Clayberg, L., & Li, H. (2019). *Security in the Face of Censorship*. Massachusetts Institute of Technology. Retrieved from: https://courses.csail.mit.edu/6.857/2019/project/3-Chen-Clayberg-Li.pdf

Clement, J. (2019). Most popular global mobile messaging apps 2020. Retrieved from Statista website: https://www.statista.com/statistics/258749/most-popular-global-mobile-

messenger-apps/

Endeley, R. E. (2018). End-to-End Encryption in Messaging Services and National Security—Case of WhatsApp Messenger. *Journal of Information Security*, *09*(01), 95–99. https://doi.org/10.4236/jis.2018.91008

Greenberg, A. (2020, January 10). Facebook Says Encrypting Messenger by Default Will Take Years. Retrieved from Wired website: https://www.wired.com/story/facebook-messenger-end-to-end-encryption-default/

Hong, W., & Thong, J. Y. L. (2013). Internet Privacy Concerns: An Integrated Conceptualization and Four Empirical Studies. *MIS Quarterly*, *37*(1), 275–298.

Jeon, D., LIUHAOYANG, & Hwang, H. (2019). Design of Hybrid Application Based on GraphQL for Efficient Query for PHR. *2019 International Conference on Information and Communication Technology Convergence*. Presented at the Jeju Island, Korea (South). Retrieved from https://ieeexplore-ieee-org.proxy01.its.virginia.edu/document/8940003

Kang, C., & Frenkel, S. (2018, April 4). Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users. *The New York Times*. Retrieved from https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html

Marr, B. (2018, May 21). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#5ebb2e0060ba

Meelu, P., & Meelu, R. (2012). Implementation of Public Key Cryptographic System: RSA. *International Journal of Information Technology and Knowledge Management*, *5*(2),

239–242.

Miller, S., & Weckert, J. (2000). Privacy, the Workplace and the Internet. *Journal of Business Ethics*, *28*(3), 255–265.

Mulligan, S. P., Freeman, W. C., & Linebaugh, C. D. (2019). *Data Protection Law: An Overview*. Congressional Research Service. Retrieved from: https://fas.org/sgp/crs/misc/R45631.pdf

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441.

Ramirez, A., & Broneck, K. (2009). `IM me': Instant messaging as relational maintenance and everyday communication. *Journal of Social and Personal Relationships*, *26*(2–3), 291–314. https://doi.org/10.1177/0265407509106719

Sanger, D. E., & Perlroth, N. (2015, November 16). Encrypted Messaging Apps Face New Scrutiny Over Possible Role in Paris Attacks. *The New York Times*. Retrieved from https://www.nytimes.com/2015/11/17/world/europe/encrypted-messaging-apps-face-new-scrutiny-over-possible-role-in-paris-attacks.html

Silver, L. (2019, February 5). Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Retrieved from Pew Research Center Global Attitudes Project website: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

WhatsApp Encryption Overview (2020). Retrieved from https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf.