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

Analyzing how to Balance Consumer and Company Needs in the California Consumer Privacy Act Through Actor-Network Theory
(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

Ashwin Pathi
Fall, 2020

Technical Project Team Members
Saiteja Bevara
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
Ashwin Pathi

Approved _____ Date _____
Yixin Sun, Department of Computer Science

Approved _____ Date _____
Catherine Baritaud, Department of Engineering and Society

According to Forbes, nearly 90% of all data has been collected in the past two years. What is more alarming is that most of this data has been collected without user consent or knowledge (Fertik, 2020). Data collection companies such as Google and Facebook have profited billions of dollars from user data, and this number is projected to increase in the future as the internet of things and wearable devices increase in usage (Feiner, 2019). However, in the wake of recent data scandals, consumers have begun to question the data collection practices of these companies (Confessore, 2018a), and as a result of these scandals the California Consumer Privacy Act (CCPA) was passed in 2018. This act acknowledged the leverage that technology companies had over consumers on their data, and sought to level the playing field of consumer and technology company data interactions.

As privacy awareness and browser technology has increased in adoption across the globe, it is necessary to provide accessible and portable needs of privacy enhancing technology, such as end-to-end-encryption. The technical portion of this project will implement an end-to-end encrypted messaging service for web browsers, with the goal of allowing encrypted and privacy-aware messaging on a wide array of devices. The technical portion also seeks to investigate and better understand the compromises such systems make to ensure user privacy. The tightly coupled STS project will employ Actor-Network Theory (ANT) to analyze the shortcomings of the CCPA, and what changes can be made to the CCPA to better balance the concerns of individual consumers and the businesses that thrive off consumer data.

The technical project will be completed in collaboration with Rithik Yelisetty, Saiteja Bevara, and Phillip Phan through a single semester capstone course. The technical advisor of the project will be Professor Yixin Sun in the Computer Science Department. The technical paper and the accompanying technical project will be completed between October and November of

2020. The STS research portion of the project will be completed between February and April of 2021 as part of a two-semester long thesis course.

WEB BASED END-TO-END ENCRYPTED MESSAGING SERVICE

According to a 2015 study by Madden and Rainie (2015), nearly 90% of adults believe that knowing who has access to their private data is important. This increased awareness in privacy has manifested in different forms, but the most overt and widely recognized form of privacy protection is end-to-end encryption (E2EE). Companies such as Apple and WhatsApp currently use E2EE to secure their messaging services, which has become a strong business point for both platforms as they seek to increase their userbase by catering to an increasingly privacy-centric society.

Many popular E2EE applications exist on the market. Three of the most popular examples include Apple iMessage, WhatsApp and Telegram. WhatsApp's E2EE protocol, which is derived from the open source Signal Protocol, has been published and proven to be mathematically unbreakable (Li et al., 2016, p. 8). Though these offerings are technically sound, the problem with standard E2EE messaging services is their accessibility through web interfaces. WhatsApp and other similar apps have web-based applications, but ultimately, these act as interfaces for the mobile application (Greenberg, 2020). Additionally, Apple's iMessage works exclusively with Apple devices, and Telegram's desktop offerings do not offer E2EE. In essence, a smartphone is required to participate in E2EE messaging. As privacy issues grow increasingly paramount in society, it is necessary for E2EE applications to be more available across a wider variety of devices. To alleviate this lack of accessibility, the technical project aims to create a

new web-based E2EE messaging platform. This application would be able to operate on any device that supports a modern web browser, which includes phones, tablets, and computers. This would also be beneficial for areas with low smartphone usage. Emerging economies such as India, Nigeria, and Indonesia have high internet availability, but do not own many smartphones (Silver, 2019). Enabling an E2EE application to work on a wider variety of internet-enabled devices will also allow people in such emerging markets to have access to E2EE messaging.

The technical project will leverage several encryption concepts and web frameworks to create an E2EE web application. The key generation and encryption will be done using the RSA algorithm. The RSA algorithm exploits the properties of large prime factors to create mathematically paired public and private keys. These keys can then be used to decrypt and encrypt messages, meaning that only users with the decryption key can see the contents of messages encrypted using its associated encryption key (Rivest, Shamir, & Adleman, 1978, p. 3). In many modern applications, a central server stores the public key, which is then used to encrypt message prior to the message transit (Rastogi & Hendler, 2017, p. 2). The technical project will implement an asymmetric key exchange process, and a symmetric key message encryption process, which is represented in Figure 1 on page 4 with an example of a three-user message exchange. All chats are abstracted as group chats, where each chat has an associated key.

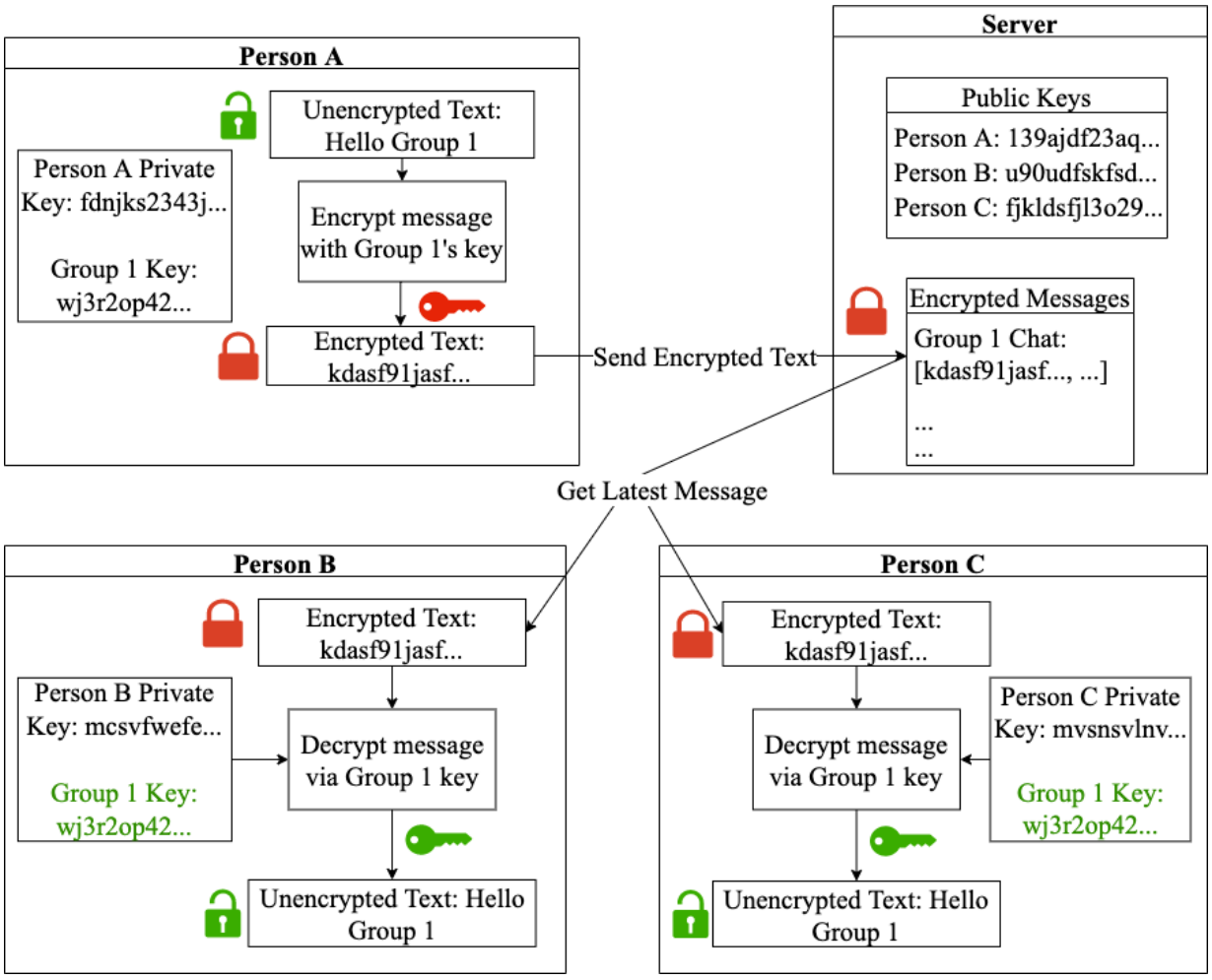

Figure 1: Example Data Flow of Encrypted Messages: An example exchange of messages between three parties using symmetric encryption. Only the three end users with the group key can read the plain text message (Bevara et al., 2020).

The group keys are initially sent to each user through a fan-out process where the group key is encrypted using the RSA public key of each member of the group. Each user can then decrypt the group key using their associated private key. The group key can then be used to encrypt and decrypt subsequent messages that are sent in the group chat. The keys are only present on the physical device of each user, and the server will have no knowledge of how to decrypt the encrypted contents in its storage. Because the server cannot decrypt the messages contained in its database, the application preserves the privacy of all three users.

The technical project can be completed using an array of free and open-source resources and web frameworks. Amazon Web Services will be used to host the NodeJS server, React will be used as the frontend framework, SQL will be used for the database, GraphQL will be used to interface with the SQL database, and Github Pages will be used to host the frontend of the web application. By creating a new E2EE web application from the ground up, this project aims to offer a new, more accessible secure messaging platform. Additionally, this project aims to shed light onto the internal workings of E2EE web applications, and to search for potential areas of improvement or complications in current E2EE offerings. The results of the project will be presented in a conference-style paper.

BALANCING CONSUMER AND BUSINESS INTERESTS IN THE CALIFORNIA CONSUMER PRIVACY ACT

Consumer privacy has become an increasingly relevant topic in the modern era of big data and data analytics. Companies have begun amassing terabytes of data on customers with the intention of selling or leveraging their data to provide meaningful services to various customers. At its core, privacy has two different meanings to technology companies and consumers. One of the turning points for consumer's viewpoints on data privacy occurred in the wake of the Facebook and Cambridge Analytica Scandal. In the scandal, Cambridge Analytica, a data analytics firm in the United Kingdom, used improperly obtained Facebook user data to sway US elections. After it was revealed that nearly 80 million users had their data collected without their consent by Cambridge Analytica, many consumers began to resent technology companies and their handling of personal data (Confessore, 2018b). In the eyes of consumers and the

government, it was Facebook's responsibility to vet data usage, and to protect user's privacy (Confessore, 2018b). However, data privacy pundits argued that Facebook instead prioritized profits over the protection of their user's data (Cyphers et al., 2018). The scandal showed that companies view privacy is a feature that users have to pay for, and a selling point. As Shoshana Zuboff (2015) puts it, consumers are ultimately data points to extract value from in the eyes of technology firms (p.79).

To consumers, privacy has become as important as fundamental rights. With the advent of accessible encryption technology, ad-blockers, and other privacy enhancing technologies, consumers are more aware of their ability to keep information confidential. However, consumers still lack control over how their data is used by other companies. Because of the stark power difference between consumers and companies, legislation has been put into place to even the playing field. In 2016, Europe implemented the General Data Protection Regulation (GDPR) to give consumers more agency in the data collection process. The United States followed suit recently with the California Consumer Privacy Act of 2018 (CCPA), which enables users to view and delete data collected on them, opt out of data collection, and not be discriminated by companies for accessing or deleting data (California Consumer Privacy Act, 2018). The CCPA aims to give consumers more agency in the data collection process (Bensinger, 2020). However, the history of the CCPA and the interaction of various actors during the legislation period has led to social and technical issues for both corporations and consumers.

COMPLICATIONS WITH THE CALIFORNIA CONSUMER PRIVACY ACT

Complications arise from the CCPA due to the way consumers and companies frame the issue of privacy. While companies frame privacy as a commodity or a product, consumers are

beginning to view privacy as a right. Though the CCPA begins to mediate the differing viewpoints between the two, its broad policies and heavily lobbied history has led legislation to be skewed towards technology companies. In his *Washington Post* article, Greg Bensinger (2020) describes scenarios where corporations take advantage of the CCPA. For example, the CCPA states that downloaded personal information should "…be in a portable, and to the extent technically feasible, readily usable format that allows the customer to transmit this information to another entity without hinderance" (California Consumer Privacy Act, 2018). The act itself does not provide a standard format for exported data. Bensinger (2020) states that companies take advantage of this lack of restrictions on data formatting by giving end users hard to parse files. While the files themselves are portable, they are inaccessible to a normal consumer, and often require technical knowledge to understand. This in turn reduces the consumer's ability to understand what data is being collected.

Another example of broad legislation that the CCPA introduces is the concept of personal data. The CCPA describes personal data as any information that can reasonably be attributed to a customer or household (California Consumer Privacy Act, 2018). However, many companies have interpreted the concept of personal data differently. Singer (2020) states that various companies give completely disparate sets of personal data to consumers. This is because the CCPA does not provide a codified guideline on what data companies should disclose to users. This ultimately effects the end user, as companies may withhold data that is considered to be personal data. Ultimately, these loopholes and broad definitions in the CCPA stem from a lack of communication between the parties involved in the legislation and enforcement process of the CCPA. These loopholes highlight the main issue of the CCPA in its current state – its inability to make privacy accessible to all users.

USING ACTOR NETWORK THEORY TO ANALYZE THE CCPA LEGISLATION PROCESS

The STS topic is particularly conducive to analysis with ANT, as ANT seeks to relate the social and the technical components of a system (Law & Callon, 1998, p. 285). In the case of the CCPA, the social component is comprised of interaction and the viewpoints of privacy across multiple groups, and the technical component is comprised of the technical solutions used to create privacy preserving technologies, and the CCPA itself. The STS topic will use Actor-Network Theory (ANT) to analyze how the CCPA legislative process was carried out, and how various actors interacted to produce the initial CCPA legislation. The current legislative process can be represented by the ANT Handoff Model, as presented In Figure 2 below.

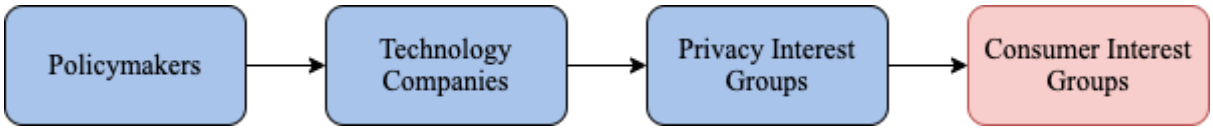

Figure 2: ANT Handoff Model: Linear process depicting the various parties CCPA legislation passed through before its eventual adoption (Adapted by Ashwin Pathi (2020) from W. Bernard Carlson 2009).

legislative process. The bill was first brought to attention by activists, then drafted and brought into law by policymakers. Shortly after, lobbyists from technology companies and privacy interest groups sought to add amendments to the bill. Finally, the legislation reached end users and consumer interest groups. As Tony Romm (2019) accounts, consumer advocacy groups had fairly little say in the CCPA legislation process, and many legislation changes were proposed by technology companies, who contended for a limitation in scope of the CCPA.

However, the handoff network fails to consider the interactions between each actor, and fails to give each actor individual agency. This handoff network can be used to construct a

network of the actors and actants that were involved in the passage and legislation of the CCPA. The key actors in the network are derived from the handoff model, and include technology companies, legislators, privacy advocacy groups, consumers, and the CCPA itself. The STS paper will use ANT to analyze this network, presented in Figure 3 below.

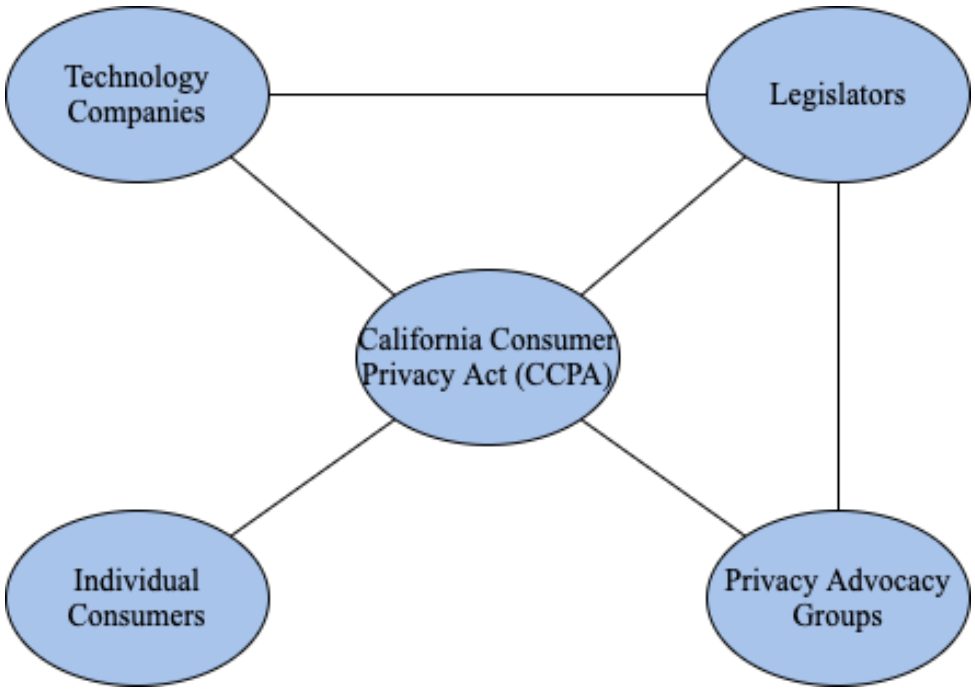

Figure 3: Actor-Network model: Actor-Network model describing the actors and actants involved in the creation of the CCPA (Adapted by Ashwin Pathi (2020), from W. Bernard Carlson 2009).

In this figure, the actors interact with each other through the CCPA, but the consumer group does not have any direct connections with technology companies. The connections between the technology companies, legislators, and privacy advocacy groups represent the lobbying and communication that occurred during the legislation process. This lack of connection between technology companies and consumers ultimately led to the issues that Bensinger (2020) alluded to in his article. However, by creating new networks between the consumer, privacy advocacy

groups, and technology groups, consumers will have increased agency and an ability to form negotiation spaces between other members in the network. This will allow customers to converse and reconcile their values of privacy with the other stakeholders, leading to more customer-centric legislation. Using this method to analyze the CCPA will establish agency for each stakeholder, allowing for the study of how each group exercises its own values and interpretations of privacy. The concept of dispersed agency in ANT has been shown to be successful in previous studies with digital pills, and will act as the main actant in the new model (Hurtado-de-Mendoza et al., 2015). Adding the new network will also allow customers to interact with technology companies and their technologies, ensuring that end users will receive information they want.

Ultimately, both the technical and STS project aim to make privacy more accessible to users. The technical project aims to increase privacy accessibility through technology, while the STS portion of the paper aims to analyze how privacy can be made more accessible through sociotechnical analysis. Current literature has analyzed the CCPA as an individual entity and analyzed the technical reasons as to why current legislation has failed. However, there has been little studies on how sociotechnical aspects have impacted the current CCPA legislation, and how it has reduced the accessibility of privacy to consumers. The STS research will be presented in a scholarly article outlining how the CCPA can better balance the needs of individuals and corporations in a sociotechnical context.

WORKS CITED

Bensinger, G. (2020, January 21). So far, under California's new privacy law, firms are disclosing too little data-Or far too much. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2020/01/21/ccpa-transparency/

Bevara, S., Yelisetty, R., Pathi, A. (2020). *Example data flow of encrypted messages.* [Figure 1]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering And Applied Science, University of Virginia, Charlottesville, VA

California Consumer Privacy Act, Cal. Civil Code §§ 1798.100 – 1798.199. (2018). https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=

Confessore, N. (2018a). The unlikely activists who took on Silicon Valley—And won. *The New York Times*. Retrieved from https://www.nytimes.com/2018/08/14/magazine/facebook-google-privacy-data.html

Confessore, N. (2018b). Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times*. Retrieved from https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Cyphers, B., Gebhart, G., & Schwartz, A. (2018). Data privacy scandals and public policy picking up speed: 2018 in review. *Electronic Frontier Foundation*. https://www.eff.org/deeplinks/2018/12/data-privacy-scandals-and-public-policy-picking speed-2018-year-review

Feiner, L. (2019, August). Facebook and Google's dominance in online ads is starting to show some cracks. Retrieved from https://www.cnbc.com/2019/08/02/facebook-and-googles-ad-dominance-is-showing-more-cracks.html

Fertik, M. (2020, January). CCPA is a win for consumers, but businesses must now step up on cx. Retrieved from https://www.forbes.com/sites/michaelfertik/2020/01/27/ccpa-is-a-win-for-consumers-but-businesses-must-now-step-up-on-cx/

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/

Hurtado-de-Mendoza, A., Cabling, M. L., & Sheppard, V. B. (2015). Rethinking agency and medical adherence technology: Applying Actor Network Theory to the case study of Digital Pills. *Nursing Inquiry*, 22(4), 326–335. https://doi.org/10.1111/nin.12101

Law, J., & Callon, M. (1988). Engineering and sociology in a military aircraft project: A network analysis of technological change. Social Problems, 35(3), 284–297. https://doi.org/10.2307/800623

Li, C., Sanchez, D., Hua, S. (2016). WhatsApp security paper analysis. In *MIT 6.857: Computer and Network Security (Spring 2016).* Retrieved from https://courses.csail.mit.edu/6.857/2016/files/36.pdf

Madden, M., & Rainie, L. (2015). *Americans' attitudes about privacy, security and surveillance*. Retrieved from https://www.pewresearch.org/internet/2015/05/20/americans-attitudes about-privacy-security-and-surveillance/

Pathi, A. (2020). *ANT Handoff Model.* [Figure 2]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering And Applied Science, University of Virginia, Charlottesville, VA

Pathi, A. (2020). *Actor-Network model.* [Figure 3]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering And Applied Science, University of Virginia, Charlottesville, VA

Rastogi, N., Hendler, J. (2017). WhatsApp security and role of metadata in preserving privacy. *ArXiv:1701.06817*. Retrieved from http://arxiv.org/abs/1701.06817

Rivest, R.L., Shamir. A., Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM, 21*(2), 120-126, doi:10.1145/359340.359342

Silver, L. (2019, February). *Smartphone ownership is growing rapidly around the world, but not always equally*. Retrieved from https://www.pewresearch.org/global/2019/02/05 /smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

Singer, N. (2019, December 29). What does California's new data privacy law mean? Nobody agrees. *The New York Times.* Retrieved fromhttps://www.nytimes.com/2019/12/29/ technology/california-privacy-law.html

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89. https://doi.org/10.1057/jit.2015.5