Undergraduate Thesis Prospectus

A Mobile Application for Robust and User-Friendly Steganography: Discreet Communication Over Publicly Visible Channels

(technical research project in Computer Science)

Data Hiders and Finders
A Balance Between Privacy Rights and Public Good

(sociotechnical research project)

by

Dylan Cao

December 9, 2020

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.
Dylan Cao

Technical advisor: David Wu, Department of Computer Science

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

How may privacy be optimized?

Americans disagree about privacy. In 2013, after the Edward Snowden revelations, a Gallup poll found that 53 percent of U.S. adults disapproved of telephone or internet data collection by U.S. government agencies; 37 percent approved (Newport, 2013). What is the optimum balance of privacy and security? Should Americans have legal access encrypted messaging, and may use of encryption constitute legal grounds for suspicion?

A Mobile Application for Robust and User-Friendly Steganography: Discreet Communication Over Publicly Visible Channels

How can messages be both discreetly and securely passed over a public communication channel?

This is a Computer Science independent research project with Professor David Wu (dwu4); no other collaborators. Our program will hide text inside JPEG images (a technique known as image steganography) such that the text resists digital damage, is hard to find by third-party observers, and is ideally only readable by the intended recipient. Sharing information in this manner strikes a balance between security and discreetness: Encrypted file sharing is very secure but very obvious; steganographic encoding is less secure and less obvious.

This project works to improve upon two flaws in mainstream tools: lack of robustness, and low ease of use. A file's data is "robust" when the file can be modified while retaining most of its information. To use image steganography in a social media outlet (or many other public image upload services), this is a requirement since most sites compress images in a lossy way to save space; this destroys data embedded using fragile techniques. Much of the consumer usable

state-of-the-art in the field relies on least-significant bit (LSB) steganography (but LSB techniques are extremely fragile and useless for JPEG files, which is unacceptable as JPEG is the de-facto standard for photo sharing) or transient data which is likely to be lost as soon as the file is modified (for example, JSteg and F5 embed data in transient JPEG fields) (*OpenStego*, n.d.) (Westfeld, 2001). Second of all, most more complex steganographic tools (i.e. not using LSB) are rarely available outside of desktops, which is highly unfortunate when a lot of picture sharing and communication occurs on mobile devices today.

We will improve upon this by developing a mobile application using advanced steganography techniques that create robust JPEG files suitable for use on social media and photo sharing sites. The rationale for focusing on such sites is that they are likely candidates for practical steganography use: if a secure, non-public channel is available, one does not need to use steganography. Notable technical constraints are that our technique must be computationally efficient to target mobile devices, and that the available programming languages on mobile devices are limited as well. We plan to use the Discrete Wavelet Transform to embed data in the low-frequency parts of an image along with error-correcting codes for more robustness; an initial prototype shows this performs acceptably well. Experiments will be conducted by trying to transmit image files over social media with arbitrary text content, and trying to successfully extract the text. Metrics will be collected on discreetness. Once the technique is validated, a mobile application user-interface will be built. The resulting program will be one of few tools available to mobile device users for discreetly and robustly sharing information without the scrutiny of using end-to-end encryption services.

Data Hiders and Finders: A Balance Between Privacy Rights and Public Good

How do data collectors and data hiders compete to advance their respective agendas?

The International Data Corporation has estimated that by 2018, 33 zettabytes of digital data had been collected worldwide (Reinsel et al., 2018), equivalent to the capacity of about 33 billion mid-sized consumer hard drives at about a terabyte each. COVID-19 contact tracing apps may omit useful data to preserve privacy (Fahey & Hino, 2020). Data collecting wearables could help persons with autism read social cues, but issues of consent arise (Kirkham & Greenhalgh, 2015). How accessible should data be? Who should have access to data or to data encryption, and under what conditions? How are the advantages of data access balanced against the value of privacy?

Data collectors seek to limit data hiding; some argue that unlimited data hiding would be socially undesirable. Two main subgroups comprise collectors: first, there are those who believe unlimited data hiding is a detriment to society. Attorney General Barr has warned that "unlimited privacy" would shield criminals (Barr, 2019). Some object to paywalls for research publications as an obstacle to research. Open access journals charge readers nothing, but instead charge authors or their institutions; fees can be in the thousands of dollars (Open Access, n.d.). The Budapest Open Access Initiative maintains that "By 'open access'" to peer-reviewed research literature, "we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles" (2012). Others illegally distribute even paywalled research freely. Through Sci-Hub, readers can get free access to about 85 percent of paywalled research articles (Himmelstein et al., 2018). To some data collectors, such as Facebook, data are merely opportunities for profit. Facebook has warned its clients that iOS 14 privacy standards will impair its Audience Network business. It has advised

clients that because of the rules, they will be less able to "accurately target and measure their campaigns on Audience Network"; hence they will "monetize on Audience Network" less. (Facebook, 2020).

Data hiders include businesses that want to protect proprietary data. For example, John Deere applies "Technological Protection Measures" to limit third-party repairs and modifications of its tractors (Bartholomew, 2015). Sony BMG covertly modified consumers' operating systems to prevent copying of their music CDs (Halderman & Felten, 2001). Other data hiders are activists seeking to protect their activities (Nierenberg, 2020). Privacy advocacies defend data hiding. The Electronic Frontier Foundation (EFF) claims its mission is "to defend free speech online, fight illegal surveillance, advocate for users and innovators, and support freedom-enhancing technologies" (EFF, n.d.). Mozilla implemented encrypted domain name lookups for its US Firefox users (Deckelmann, 2020). The Internet Security Research Group (ISRG) Let's Encrypt program issued one billion free TLS certificates by 2020, because "Nothing drives adoption like ease of use," (Aas & Gran, 2020). Media outlets support encryption to "protect your identity, location, and the information you send us," (The New York Times, 2016). Because ecommerce depends upon data hiding, ecommerce companies are major defenders of encryption. During the second quarter of 2020, ecommerce accounted for about 16 percent of all U.S. retail sales (Commerce, 2020). The Shopify platform added free TLS encryption to its sites as "the right thing for ecommerce in 2016" and sponsors the ISRG Let's Encrypt initiative (Cornu, 2016).

References

Aas, J., & Gran, S. (2020, February 27). Let's Encrypt Has Issued a Billion Certificates. https://letsencrypt.org/2020/02/27/one-billion-certs.html

Barr, W. (2019, October 4). Attorney General William P. Barr delivers remarks at the Lawful Access Summit. https://www.justice.gov/opa/speech/attorney-general-william-p-barr-delivers-remarks-lawful-access-summit

Bartholomew, D. (2015). Long comment regarding a proposed exemption under 17 U.S.C. 1201. http://copyright.gov/1201/2015/comments-032715/class%2021/John_Deere_Class21_1201_2014.pdf

Budapest Open Access Initiative. (2012, September 12). Ten years on from the Budapest Open Access Initiative: Setting the default to open. Budapest Open Access Initiative. https://www.budapestopenaccessinitiative.org/boai-10-recommendations

Commerce (2020). U.S. Department of Commerce. Quarterly Retail E-Commerce Sales. https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

Cornu, D. (2016, February 2). All Shopify Stores Now Use SSL Encryption Everywhere. Shopify. https://www.shopify.com/blog/73511365-all-shopify-stores-now-use-ssl-encryption-everywhere

Deckelmann, S. (2020, February 25). Firefox continues push to bring DNS over HTTPS by default for US users. The Mozilla Blog. https://blog.mozilla.org/blog/2020/02/25/firefox-continues-push-to-bring-dns-over-https-by-default-for-us-users

EFF (n.d.). Electronic Frontier Foundation. About EFF. https://www.eff.org/about

Facebook. (2020, September 15). How we're preparing businesses for the impact of iOS 14. Facebook for Business. https://www.facebook.com/business/news/preparing-our-partners-for-ios-14-launch

Fahey, R. A., & Hino, A. (2020). COVID-19, digital privacy, and the social limits on data-focused public health responses. *International Journal of Information Management*, *55*, 102181. https://doi.org/10.1016/j.ijinfomgt.2020.102181

Halderman, J. A., & Felten, E. W. (2001). Lessons from the Sony CD DRM Episode.

Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Greshake Tzovaras, B., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *ELife*, *7*, e32822. PubMed. https://doi.org/10.7554/eLife.32822

Kirkham, R., & Greenhalgh, C. (2015). Social Access vs. Privacy in Wearable Computing: A Case Study of Autism. *IEEE Pervasive Computing*, *14*(1), 26–33. https://doi.org/10.1109/MPRV.2015.14

Newport, F. (2013, June 12). Americans disapprove of government surveillance programs. Gallup.Com. https://news.gallup.com/poll/163043/americans-disapprove-government-surveillance-programs.aspx

Nierenberg, A. (2020, June 12). Signal downloads are way up since the protests began. The New York Times. https://www.nytimes.com/2020/06/11/style/signal-messaging-app-encryption-protests.html

Open Access. (n.d.). History of the Open Access Movement. Open Access. https://open-access.net/en/information-on-open-access/history-of-the-open-access-movement

OpenStego. (n.d.). https://www.openstego.com/features.html

Reinsel, D., Gantz, J., & Rydning, J. (2018). The Digitization of the World from Edge to Core. International Data Corporation. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Tips. (2016, December 14). The New York Times. https://www.nytimes.com/tips.

Westfeld, A. (2001). F5—A Steganographic Algorithm: High Capacity Despite Better Steganalysis. In I. S. Moskowitz (Ed.), *Information Hiding* (Vol. 2137, pp. 289–302). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45496-9_21