School of Law
UNIVERSITY OF GEORGIA
Prepare.
Connect.
Lead.

Journal of Intellectual Property Law

Volume 28 | Issue 2 Article 2

July 2021

A Deep Dive into Technical Encryption Concepts to Better Understand Cybersecurity & Data Privacy Legal & Policy Issues

Anthony Volini
DePaul University, avolini@depaul.edu

Recommended Citation

Anthony Volini, *A Deep Dive into Technical Encryption Concepts to Better Understand Cybersecurity & Data Privacy Legal & Policy Issues*, 28 J. INTELL. PROP. L. 291 (2021).
Available at: https://digitalcommons.law.uga.edu/jipl/vol28/iss2/2

A Deep Dive into Technical Encryption Concepts to Better Understand Cybersecurity & Data Privacy Legal & Policy Issues

Cover Page Footnote

Senior Professional Lecturer at DePaul University College of Law, Registered Patent Attorney, M.S. Cybersecurity (Networking & Infrastructure) (Anticipated 2021), Certified Information Privacy Professional/United States (CIPP/US), CIPP/Europe (CIPP/E), Cybersecurity Fundamentals Certificate (CSXF). Many thanks to Christopher Boyd (3L) and Ashley Weringa (3L), DePaul University College of Law for their assistance, as well as Professor Joshua Sarnoff, Mat Kresz, Karen Heart, David Habich, Colin Black, Thomas Combs, and Brian Barnes for offering their insights.

A DEEP DIVE INTO TECHNICAL ENCRYPTION CONCEPTS TO BETTER UNDERSTAND CYBERSECURITY & DATA PRIVACY LEGAL & POLICY ISSUES

*Anthony G. Volini**

* Senior Professional Lecturer at DePaul University College of Law, Registered Patent Attorney, M.S. Cybersecurity (Networking & Infrastructure) (Anticipated 2021), Certified Information Privacy Professional/United States (CIPP/US), CIPP/Europe (CIPP/E), Cybersecurity Fundamentals Certificate (CSXF). Many thanks to Christopher Boyd (3L) and Ashley Weringa (3L), DePaul University College of Law for their assistance, as well as Professor Joshua Sarnoff, Mat Kresz, Karen Heart, David Habich, Colin Black, Thomas Combs, and Brian Barnes for offering their insights.

Lawyers wishing to exercise a meaningful degree of leadership at the intersection of technology and the law could benefit greatly from a deep understanding of the use and application of encryption, considering it arises in so many legal scenarios. For example, in FTC v. Wyndham[1] *the defendant failed to implement nearly every conceivable cybersecurity control, including lack of encryption for stored data, resulting in multiple data breaches and a consequent FTC enforcement action for unfair and deceptive practices. Other examples of legal issues requiring use of encryption and other technology concepts include compliance with security requirements of GLBA & HIPAA, encryption safe harbors relative to state data breach notification laws and the CCPA, the NYDFS Cybersecurity Regulation, and PCI standards. Further, some policy discussions have taken place in 2020 regarding encrypted DNS over HTTPS, and lawyers would certainly seem to benefit from a better understanding of relevant encryption concepts to assess the privacy effectiveness of emerging encryption technologies, such as encrypted DNS. Finally, the need for technology education for lawyers is evidenced by North Carolina and Florida requiring one or more hours in technology CLE and New York in 2020 moving toward required CLE in the area of cybersecurity specifically.*

This article observes that there is a continuing desire for strong encryption mechanisms to advance the privacy interests of civilians' online activities/communications (e.g., messages or web browsing). Law enforcement advocates for a "front door," requiring tech platforms to maintain a decryption mechanism for online data, which they must produce upon the government providing a warrant. However, privacy advocates may encourage warrant-proof encryption mechanisms where tech platforms remove their ability to ever decrypt. This extreme pro-privacy position could be supported based on viewing privacy interests under a lens such as Blackstone's ratio. Just as the Blackstone ratio principle favors constitutional protections that allow ten guilty people to go free rather than allowing one innocent person suffer, individual privacy rights could arguably favor fairly unsurveillable encrypted communications at the risk of not detecting various criminal activity. However, given that the internet can support large-scale good or evil activity, law enforcement continues to express a desire for a front door required by legislation and subject to suitable privacy safeguards, striking a balance between strong privacy versus law enforcement's need to investigate serious crimes. In the last few decades, law enforcement appears to have lost the debate for various reasons, but the debate will likely continue for years to come.

For attorneys to exercise meaningful leadership in evaluating the strength of encryption technologies relative to privacy rights, attorneys must generally understand encryption principles, how these principles are applied to data at rest (e.g., local encryption), and how they operate with respect to data in transit. Therefore, this article first explores encryption concepts primarily with regard to data at rest and then with regard to data in transit, exploring some general networking protocols as context for understanding how encryption can applied to data in transit, protecting the data payload of a packet and/or the routing/header information (i.e., the "from" and "to" field) of the packet.

Part 1 of this article briefly explores the need for lawyers to understand encryption. Part 2 provides a mostly technical discussion of encryption concepts, with some legal concepts injected therein. Finally, Part 3 provides some high level legal discussion relevant to encryption (including arguments for and against law enforcement's desire for a front door). To facilitate

[1] F.T.C. v. Wyndham Worldwide Corp., 799 F.3d 236 (3d Cir. 2015).

understanding for a non-technical legal audience, I include a variety of physical world analogies throughout (e.g., postal analogies and the like).

TABLE OF CONTENTS

I. THE NEED FOR LAWYERS TO UNDERSTAND ENCRYPTION

A. ABOUT *FTC V. WYNDHAM*: A LAUNCHING POINT FOR DISCUSSION OF ENCRYPTION

FTC v. Wyndham is a 2015 Third Circuit opinion, which upheld the FTC's authority to fine the Wyndham hotel chain for cybersecurity shortcomings. On three occasions in 2008 and 2009, hackers successfully accessed Wyndham Worldwide Corporation's computer systems.[2] "In total, they stole personal and financial information for hundreds of thousands of consumers leading to over $10.6 million dollars in fraudulent charges. The FTC filed suit in federal District Court, alleging that Wyndham's conduct was an unfair practice and that its privacy policy was deceptive" for misrepresenting that it had good security controls (including representing it had encryption in place).[3]

The court identified a laundry list of cybersecurity shortcomings, including: (1) lack of firewalls, (2) lack of encryption, (3) lack of an incident response plan, (4) easily guessed passwords, (5) failure to maintain an inventory of assets, and (6) failure to implement software updates.[4] This article does not address all of the cybersecurity shortcomings in *Wyndham*. Rather, it focuses on the very important encryption component as a helpful starting point for understanding cybersecurity.

1. *The Need for Lawyers to Understand Encryption*

Understanding encryption is certainly relevant in light of ABA Rule 1.1 (duty of technological competence), Rule 1.6 (duty of confidentiality), a variety of statutes directly or indirectly requiring encryption, encryption as a means to reduce negligence exposure, etc.[5] The ABA has commented that attorneys must understand and use encryption in practice.[6] By doing so, lawyers then can understand how their companies/firms utilize encryption, as well as understand the errors, use, and decisions regarding encryption exercised by their adversaries in litigation.[7] Perhaps more importantly, understanding the underlying mechanics and usage of encryption will help lawyers maximize protection of

[2] *Wyndham Worldwide*, 799 F.3d at 241.

[3] *Id.* at 240.

[4] *Id.* at 241.

[5] MODEL RULES OF PROF'L CONDUCT r. 1.1 cmt. 8 (2020); MODEL RULES OF PROF'L CONDUCT r. 1.6 (2020).

[6] Daniel Garrie & Rick Borden, *Encryption for Lawyers,* 2016 A.B.A. SEC. PUB. BUS. L. TODAY 1, 1, https://www.americanbar.org/content/dam/aba/publications/blt/2016/06/full-issue-201606.pdf.

[7] *Id.*

client information and protection of themselves.[8] Due to the tremendous amount of highly sensitive information today that gets transmitted over the internet, law firms cannot only rely on their Information Technology (IT) department, but also must teach and make sure their employees understand how to encrypt their data and maximize password security to secure the company's data overall and ultimately save money.[9] Despite these somewhat recent pronouncements, the majority of attorneys are not appropriately using encryption.[10] A 2019 ABA Cybersecurity Survey showed 26% of the survey respondents' law firms experienced some type of data breach.[11] Further, the survey results suggest, "less than half of respondents use file encryption (44%), slightly more than a third use email encryption (38%), and even fewer use whole/full disk encryption (22%)."[12] There was a slight increase in the use of encryption by law firms from 2018 to 2019,[13] however these numbers still remain too low.

State bars are now starting to require technology and/or cybersecurity CLE in furtherance of the ABA rules, with North Carolina and Florida requiring one or more hours in technology CLE and New York in 2020 moving toward required CLE in the area of cybersecurity specifically.[14] These factors evidence the need for attorneys to learn about encryption and related tech as this is highly relevant to protecting data in the attorney's care or in the context of the attorney advising clients on data protection strategies for client held data. Examples of legal issues requiring use of encryption and other technology concepts include compliance with security requirements of GLBA & HIPAA, encryption safe

[8] *Id.* (stating, "Even if a company scrupulously follows the requirements of the jurisdiction in which personal data is collected, and provides appropriate notices of the use of the information, and the rights of the individuals who provide the information, unless the information is appropriately protected in transit, during processing, and at rest, the entities who collect or hold such personal data may be liable for significant liabilities and penalties"). It is therefore critical that lawyers not only generally understand encryption and how it should be used, but also how to appropriately protect data in transit, during processing, as well as at rest to avoid liability.

[9] *Id.* at 1-2.

[10] John G. Loughnane, *2019 Cybersecurity*, 2019 A.B.A. SEC. PUB. TECH. REP. (Oct. 16, 2019), https://www.americanbar.org/groups/law_practice/publications/techreport/abatechreport2019/cybersecurity2019/.

[11] *Id.*

[12] *Id.*

[13] *Id.*

[14] Bob Ambrogi, *North Carolina Becomes Second State to Mandate Technology Training for Lawyers*, LAWSITES (Dec. 5, 2018), https://www.lawsitesblog.com/2018/12/north-carolina-beco mes-second-state-mandate-technology-training-lawyers.html; Brian G. Cesaratto & Shawndra G. Jones, *New York Could Become the First State to Require Cybersecurity,* EPSTEIN BECKER GREEN WORKFORCE BULLETIN (Sept. 3, 2020), https://www.workforcebulletin.com/2020/09/03 /new-york-could-become-the-first-state-to-require-cybersecurity-cle/.

harbors relative to state data breach notification laws and the CCPA, the NYDFS Cybersecurity Regulation, and PCI standards.[15] In addition, a lack of encryption can support a negligence claim.[16] Also, encryption can be one useful mechanism to protect an organization's trade secrets or other sensitive business data.

Attorneys and technologists will continue to work together on assessment of technology relative to cybersecurity and privacy issues, digital forensics in litigation, etc. Accordingly, knowledge of encryption and the environment in which it operates should help attorneys bridge communication gaps with technologists. Regarding privacy, encryption is potentially the principal tool to increase privacy, so such knowledge should likely benefit attorneys in assessing the effectiveness of encryption technologies in terms of improving privacy.[17] For example, some policy discussions have taken place in 2020 regarding encrypted DNS over HTTPS.[18] Lawyers would certainly seem to benefit from a better understanding of relevant encryption concepts to assess the effectiveness of emerging encryption technologies, such as encrypted DNS.

2. *Caveat: Encryption Promotes Privacy but is not a Privacy Panacea*

Use of encryption technologies is typically not a guarantee of privacy. For example, as explained in section C below, the HTTPS protocol encrypts the payload of data packets but does not hide the identity of the sender and receiver (e.g., perhaps like a phone record revealing dialed and received phone numbers without access to the contents of conversations). Also, DNS leaks and browser fingerprinting are described below, which may negate a user's intended anonymous browsing when using a VPN service. Accordingly, attorneys are better positioned to assess privacy effectiveness of new or existing encryption technologies by having a basic understanding of encryption concepts and the context in which encryption is implemented in order to understand the pros and potential shortcomings of various encryption technologies.

[15] PCI stands for Payment Card Industry standards, which constitute a private agreement between merchants and credit card issuers, requiring encryption and other protections for credit card information. *See* PCI SECURITY STANDARDS COUNCIL, https://www.pcisecuritystandards.org/ (last visited Oct. 15, 2020).

[16] In collecting and storing employees' data on its computer systems, UPMC owed employees a duty to exercise reasonable care to protect them against an unreasonable risk of harm arising out of that act. Dittman v. UPMC, 196 A.3d 1036, 1047 (Pa. 2018).

[17] *See* Lauren C. Williams, *Edward Snowden Says Encryption Is the Only Way to Counter Mass Surveillance,* THINK PROGRESS (Mar. 10, 2014, 6:57 PM), https://archive.thinkprogress.org/edward-snowden-says-encryption-is-the-only-way-to-counter-mass-surveillance-ee450433dca8/ (noting Snowden's comment that "[e]ncryption is the defense against the dark arts for the digital realm.").

[18] *See infra* Section II.E.

II. A MOSTLY TECHNICAL DISCUSSION OF ENCRYPTION CONCEPTS

A. EXPLORING ENCRYPTION, AND RELATED TECHNOLOGY CONCEPTS, AT REST

In my prior work, I described that a 30,000-foot-high view of tech educational instruction is essentially worthless to attorneys, while a 10,000-foot-high view provides meaningful depth to understand the tech that relates to the legal issues (and 0 foot/in the weeds is overkill for those not doing IT work).[19] I also discussed the importance of visual aids to understand IT concepts. Below, I attempt to provide a 10,000-foot level of education relative to encryption and its contexts, embedding a variety of visual aids to enhance understanding.[20]

1. *Encryption Generally*

To begin the encryption discussion, I first describe what encryption is, its application to stored data, and some methods used to crack/decrypt data. Later, I discuss general networking protocols for transmitting data as context followed by different implementations of encryption.

Encryption is essentially the conversion of plain text to a meaningless string of gibberish (called cipher text) so that an eavesdropper cannot easily decipher it.[21] For context, the earliest known form of encryption is likely the Julius Caesar Cipher, which Caesar used to send military messages. The Caesar Cipher works by using an algorithm, such as shifting every character once to the next letter in the alphabet (or probably a much more complicated algorithm that would be harder to crack, such as shifting the first letter of each word twice, shifting the second letter of each word four times, etc.).[22] A simple example of a Caesar Cipher is to shift each letter once.[23] Therefore, the word "Hello" becomes "Ifmmp." Obviously, an eavesdropper would have to spend at least minimal effort to decrypt "Ifmmp" to the word "Hello."

Modern computers can very easily and quickly crack a Caesar Cipher, so a more sophisticated key/cipher is needed than mere letter substitution. A variety of modern encryption standards have replaced obsolete standards, and currently

[19] Volini, Anthony, *A PERSPECTIVE ON TECHNOLOGY EDUCATION FOR LAW STUDENTS*, 36 SANTA CLARA HIGH TECH. L.J. 165, 173 (2020), https://digitalcommons.law.scu.edu/chtlj/vol36/iss2/2.

[20] I created most of the diagrams herein using draw.io, a free tool available on the Google Chrome browser.

[21] *See Encryption*, WIKIPEDIA, https://en.wikipedia.org/wiki/Encryption (last visited Dec. 2, 2019) (discussing encryption generally).

[22] *Caesar Cipher*, PRACTICAL CRYPTOGRAPHY, http://practicalcryptography.com/ciphers/caesar-cipher/ (last visited Feb. 24, 2021).

[23] *Id.*

used standards may likely remain strong enough for several years.[24] The strengthening of encryption standards can be observed by organizations periodically increasing the length of an encryption key (e.g., from 40 to 56 to 128 bit to 256 bit) to make the encrypted data less vulnerable to brute-force attack (brute-force attack and encryption complexity are discussed in greater detail below). Encryption standards must be improved every few years given Moore's law; Moore's law is an empirical observation that computer processing power roughly doubles every two years.[25] Therefore, according to Moore's law, a computer's power and speed to decrypt encrypted data increases substantially every two years. Consistent with Moore's law, a variety of past encryption standards have been replaced with stronger encryption algorithms.[26]

2. *Local Encryption*

A defensive use of encryption for Wyndham would have been to encrypt all stored personal data, such that the data could only be decrypted and used by Wyndham with a password (i.e., the password essentially operating as the decryption key).[27] Data stored on Wyndham's servers, whether local on-site servers or remote leased servers in the cloud (such as third-party Amazon Web Service servers), thus should have been encrypted.

To further understand local encryption, one can consider a personal laptop. In a laptop running Microsoft (MS) Windows, a consumer can adjust his settings to turn on local encryption. The effect of turning on local encryption is that all stored data on the laptop is turned into unintelligible gibberish when it is logged off or powered off. When the consumer logs back into his laptop, using his password, the password essentially acts as the decryption key that decrypts the gibberish (e.g., MS Word documents) back into "plain text" (also referred to as "clear text"). If a forensics professional or bad actor encounters a powered off/logged off laptop, he may attempt to physically plug into the hard drive to

[24] *See Cryptography Standards*, WIKIPEDIA, https://en.wikipedia.org/wiki/Cryptography_standards (last updated June 5, 2020 2:31 PM) (noting various encryption standards that are now obsolete).

[25] Shara Tibken, *Moore's Law is dead, says Nvidia's CEO*, CNET (Jan. 9, 2019, 11:46 AM), https://www.cnet.com/news/moores-law-is-dead-nvidias-ceo-jensen-huang-says-at-ces-2019. Nvidia CEO (an industry leader in graphics processing units) Jensen Huang has reported on multiple occasions, ". . . as the scale of chip components gets closer and closer to that of individual atoms, it's gotten harder to keep up the pace of Moore's Law. It's now more expensive and more technically difficult to double . . . the processing power for a given chip every two years."

[26] *See 40-bit encryption,* WIKIPEDIA, https://en.wikipedia.org/wiki/40-bit_encryption (last updated Oct. 28, 2019, 3:25 PM) (describing both 40 and 56 bit encryption key lengths as obsolete and replaced by key lengths of 128 bit or longer).

[27] Chris Hoffman, *Why a Windows Password Isn't Enough to Protect Your Data*, HOW-TO-GEEK (Sept. 22, 2016, 4:02 PM), https://www.howtogeek.com/161444/htg-explains-why-a-windows-password-doesnt-protect-your-data/.

extract the data. If the laptop does not have local encryption turned on, then the forensics professional will easily see all of the data in unencrypted form (without the need to log in).[28] With a running logged-on laptop, the local encryption setting is irrelevant: a forensics professional (or a bad actor) physically or remotely accessing the laptop can easily see all the data in unencrypted form, since local encryption occurs upon logging off/powering off.

3. *Local Encryption and Password Protection are not Necessarily the Same*

It is important to note that password protection and local encryption are not necessarily the same thing. For example, if a computer has local encryption turned off, a password may be used to log on to the computer (e.g., access a Windows environment) but no local encryption occurs when logging off. As noted above, with local encryption turned off, a forensics professional or bad actor can physically plug into the computer's hard drive and retrieve all data even though he does not have the password that provides access to the user's computer by authenticating with the computer's operating system. Hence, plugging into the computer's hard drive effectively bypasses the operating system's password gate. In simple terms, the operating system's password gate is like a padlock on a storage locker. An intruder can gain access to the storage locker without going through the padlocked front door (perhaps by unscrewing a back panel of the storage locker). And when access is successful, the files and items in that storage locker are freely accessible – despite the padlock on its front door. Thus, turning on local encryption can be viewed as additional protection (in addition to the password/padlock), perhaps conceptualized as keeping one's data within a locked box inside of the padlocked storage locker.

4. *Local Encryption of Thumb Drives*

In addition to turning on local encryption on a laptop or PC, thumb drives containing sensitive information can also be locked/encrypted with a password. As an example, Microsoft's Bitlocker tool can be used for this purpose and is included with most versions of Windows 10.[29] Certainly, an attorney may encrypt a thumb drive containing sensitive client information to facilitate compliance with ABA Rule 1.6 of the Rules of Professional Conduct (confidentiality).[30]

[28] See Penny Hoelscher, *How to encrypt Files and folders in Windows 10, 8 or 7*, COMPARITECH (June 5, 2018), https://www.comparitech.com/blog/vpn-privacy/encrypt-windows-files/, for a general discussion about implementing local encryption on Microsoft Windows files and folders.

[29] *Overview of BitLocker Device Encryption in Windows 10*, MICROSOFT (Feb. 28, 2019), https://docs.microsoft.com/en-us/windows/security/information-protection/bitlocker/bitlocker-device-encryption-overview-windows-10.

[30] MODEL RULES OF PROF'L CONDUCT r. 1.6 (2020).

5. *Encryption of Data Held by a Cloud Provider*

Organizations are increasingly storing data in the cloud, with a cloud provider such as Amazon holding the data for them. An organization should insist on having such stored data encrypted and should pay attention to which parties have access to the relevant decryption keys for data (as well as other relevant cybersecurity controls).[31] Another strategic decision may involve whether to encrypt data before uploading it to a cloud provider's platform or after.[32]

6. *Encryption of Emails*[33]

Lawyers and other professionals subject to a duty of confidentiality should consider use of email software supporting encrypted communications, including free software or software sold by a particular vendor.[34] The failure to encrypt sensitive emails can constitute an ethical violation. For example, in its Formal Opinion 2010-179, the State Bar of California urges attorneys to encrypt email, stating that it is a "reasonable step . . . when the circumstance calls for it, particularly if the information at issue is highly sensitive and the use of encryption is not onerous."[35]

Various email platforms support encrypted communication, which lawyers can consider when sending sensitive information to a client. For example, the Electronic Frontier Foundation describes PGP as the standard open source encryption software to install for encrypted emails.[36] Various third-party encryption services also exist, such as Virtru, which is advertised as capable of integrating with Microsoft Outlook and Gmail.[37] Besides the need for attorneys to encrypt sensitive emails per ABA Rule 1.6 (duty of confidentiality), HIPAA

[31] See Nate Lord, *What is Cloud Encryption*, DIGITAL GUARDIAN (Sept. 11, 2018), https://digitalguardian.com/blog/what-cloud-encryption, for a general discussion about cloud encryption and decryption keys.

[32] *Id.*

[33] *See infra* Part II (focusing primarily on encryption of data at rest. Certainly, after an email is sent it is at rest (although encrypting emails also provides in transit protection from eavesdroppers)).

[34] Lori Kaufman*, The Best Free Ways To Send Encrypted Email and Secure Messages,* HOW-TO-GEEK, https://www.howtogeek.com/135638/the-best-free-ways-to-send-encrypted-email-and-secure-messages/ (last updated July 11, 2017, 8:56 PM); *see Encryption in Transit in Google Cloud*, GOOGLE CLOUD, https://cloud.google.com/security/encryption-in-transit/ (last updated Oct. 12, 2020) (discussing the encryption of data in transit in general).

[35] Cal. State Bar Formal Op. 2010-179 (3)(a)(ii) (2010) cited in 21 TORTSOURCE 3, 4, 21 NO. 3 TORTSOURCE 3, 4.

[36] *Communicating with Others,* SURVEILLANCE SELF-DEFENSE, https://ssd.eff.org/en/module/communicating-others (last updated June 9, 2020).

[37] Editorial Team, *Why Email Encryption is Necessary for Lawyers,* VIRTRU (June 5, 2020), https://www.virtru.com/blog/law-firm-data-security/.

and GLBA requirements regarding protection of sensitive health or financial information would be furthered by email encryption.

Another benefit of the encryption of emails or using encryption email services is encrypted email can potentially stop an email provider from scanning email contents for targeted ad purposes.[38] Services like Tutanota, Protonmail, and Posteo, just to name a few, offer end-to-end encryption which minimizes the extent to which emails can be scanned, which ultimately eliminates targeted ads resulting from scanned emails.[39]

7. *Encrypted/Password-Protected Documents Attached to Emails*

Professionals sending sensitive information can consider sending password protected documents, so the recipient needs the appropriate password to decrypt and view the contents. Password protection can be readily applied to PDF documents, MS Word documents, and other files.[40] Encrypting email attachments can thus serve as a way to encrypt sensitive information if the attorney and client do not have an email encryption mechanism in place to encrypt contents of the entire email message itself. In such a scenario, the attorney may share the password with the client by telephone or other means.

8. *Brute-Force Attack of Encrypted Data (e.g. passwords or numeric passcode on a smartphone)*

A brute-force attack can involve trying every possible key *upon the data directly* until the data is decrypted (i.e., until the plain text is revealed).[41] Alternatively, a brute-force attack can involve trying every possible password to gain access to a computer. For simplicity, this section focuses on the brute-force attack of a password. Assuming data on a hard drive is locally encrypted, a password must be entered to cause the hard drive's data to be decrypted so that it can be accessed by programs and users. Therefore, the forensics professional or a bad actor may attempt a "brute-force attack" to crack a password and thereby decrypt the data. A brute-force attack involves attempting every possible combination of

[38] Editorial Team, *Do I Need Email Encryption Software?*, VIRTRU (Sept. 5, 2020), https://www.virtru.com/blog/email-encryption-software/ (noting both the government and your email provider may have access to your email communications and service providers frequently harvest users emails to show ads, but encryption software can protect both your individual messages and from normalizing surveillance methods).

[39] Michael Grothaus, *These 4 Gmail alternatives put your privacy first*, FASTCOMPANY (Aug. 21, 2019), https://www.fastcompany.com/90392612/these-4-gmail-alternatives-put-your-privacy-first.

[40] Chris Hoffman, *How to Password Protect Files and Folders With Encryption*, HOW-TO-GEEK (July 30, 2016 11:59 AM), https://www.howtogeek.com/170352/how-to-password-protect-files-and-folders-with-encryption/.

[41] *Brute-Force Attack*, WIKIPEDIA, https://en.wikipedia.org/wiki/Brute-force_attack (last updated Feb. 13, 2021 at 2:22 PM).

characters until the password is successfully cracked, allowing the forensics professional to log in. As a highly simplified example, consider a four-integer password to unlock a smartphone: X X X X. Each integer X could have a value between 0 and 9 (i.e., 10 possible values). Therefore, the password has 10^4 possible number combinations (i.e., 10,000 possible combinations). A brute-force attack would involve trying every number combination until arriving at the correct password.

In contrast to a numeric smartphone passcode, with an unknown laptop password the length of the password could vary. The password complexity is potentially far greater given that the laptop owner could use a variety of letters, numbers, and special characters in the password.[42]

9. Courts Compelling a Defendant to Provide a Password or Biometric

Should the government's attempt to forensically crack a password prove impractical, the government may seek a court order compelling a defendant to disclose his password.[43] Some courts may view this as a Fifth Amendment violation, essentially viewing this as compelled self-incriminating testimony.[44] (As a side note, courts typically view compelled production of a physical

[42] See *Password must meet complexity requirements,* MICROSOFT (Sept. 7, 2017), https://docs.microsoft.com/en-us/windows/security/threat-protection/security-policy-settings/password-must-meet-complexity-requirements, for a 2017 general discussion of Microsoft's complexity requirements.

[43] For example, forensic crime labs may have smartphones or other devices subjected to brute force encryption for many months without success. *See Appleinsider Staff,* APPLE INSIDER, https://appleinsider.com/articles/18/04/17/researcher-estimates-graykey-can-unlock-a-6-digit-iphone-passcode-in-11-hours-heres-how-to-protect-yourself (last visited Oct. 14, 2020) (describing how some phone passcodes can be cracked in mere hours while longer passcodes may take years to successfully brute force).

[44] *See generally* State v. Andrews, 457 N.J. Super. 14 (N.J. Super. Ct. App. Div. 2018), *aff'd*, 243 N.J. 447, 234 A.3d 1254 (2020) (finding compelled production of passcodes did not violate defendant's Fifth Amendment privilege against self-incrimination under the "foregone conclusion" exception to privilege). *But see* Pollard v. State, 287 So.3d 649 (Fla. Dist. Ct. App. 2019), *reh'g denied* (Dec. 23, 2019), *cert. dismissed*, No. SC20-110, 2020 WL 1491793 (Fla. Mar. 25, 2020) (explaining that the state's generalized requests for multiple categories of communications, pictures, and social media activity did not describe contents of defendant's cellphone with particularity, as required to compel production of defendant's cellphone password under the foregone conclusion exception to the Fifth Amendment). *See also, e.g.*, United States v. Kirschner, 823 F. Supp. 2d 665, 669 (E.D. Mich. 2010) (finding compelled disclosure of defendant's password would violate the Fifth Amendment: "In this case, the government is not seeking documents or objects—it is seeking testimony from the Defendant, requiring him to divulge through his mental processes his password—that will be used to incriminate him."). *Cf.* State v. Stahl, 206 So. 3d 124 (Fla. Dist. Ct. App. 2016) ("requiring defendant to produce passcode did not compel defendant to communicate information that had testimonial significance."). See Fern L. Kletter, *Construction and Application of "Foregone Conclusion" Exception to the Fifth Amendment Privilege Against Self-Incrimination*, 25 A.L.R. Fed. 3d Art. 10 (2017), for further discussion of cases on either side of this issue.

biometric, such as a fingerprint, as non-testimonial, and thus, not Fifth Amendment prohibited.)[45] Therefore, if a cell phone can be unlocked by both a fingerprint and a passcode, presentation of a fingerprint to unlock the cell phone could often be compelled, but disclosure of the passcode to unlock the cell phone could not be compelled in some jurisdictions. Other courts may view compelled production of a password as not Fifth Amendment prohibited, however, relying on the foregone conclusion doctrine to justify the compelled production: for example, the compelled production seems to fit this doctrine if the government knows the existence, possession, and authenticity of the incriminating evidence because the production involves no testimonial attribute.[46]

10. *Strategies to Inhibit Brute-Force Decryption or Other Attacks on Passwords*

a. *Avoid Easily Guessed Passwords*

Wyndham's servers had easily guessed passwords, which supported liability under the FTC Act. While the dominant focus of this article is on encryption concepts, discussion of password concepts is merited in this article because, as noted above, passwords are often essentially used as a trigger to encrypt/decrypt data.

An example of an easily guessed password on a server or other computer could be "admin" used as both the username and the password. Historically, manufacturers of servers, routers, and other infrastructure have provided default passwords for their products, which one can locate with a quick web search.[47] For example, a particular model of server could have "admin" as the default username and "cisco" as the password.[48]

Historically, network administrators (and software developers) did not think about security upfront. Instead, the goal was to get the network up and running perhaps with the view that they could always go back and change the manufacturer's password later (and never do that). In recent years, particularly

[45] At the time of writing, courts have split on whether the compelled use of biometric authentication is permissible under the Fifth Amendment. *E.g.,* United States v. Wright, 431 F. Supp. 3d 1175, 1186-87 (D. Nev. 2020). *Compare* Matter of White Google Pixel 3 XL Cellphone in a Black Incipio Case, 398 F. Supp. 3d 785 (D. Idaho 2019) (finding that compelled use of a fingerprint to unlock a cellphone was not testimonial) *with Wright,* 431 F. Supp. 3d at 1187.

[46] *Wright*, 431 F. Supp. 3d at 1186-87 (finding that an officer holding a defendant's phone to their face in order to unlock the device was a violation of the defendant's Fifth Amendment rights).

[47] *See Default Passwords*, DATARECOVERY.COM (June 23, 2014), https://datarecovery.com/rd/default-passwords/ (listing passwords for different devices and applications).

[48] Tim Fisher, *Cisco Default Password List,* LIFEWIRE, https://www.lifewire.com/cisco-default-password-list-2619151 (last updated Oct. 1, 2020).

with the enactment of Europe's GDPR in 2018, implementing security by design and by default is the new norm to comply with regulatory requirements or at least to reduce legal exposure (and US organizations and legislators have been influenced by GDPR concepts).[49] Article 25 of the GDPR requires security by design and by default for parties launching a new IT product or service.[50] Therefore, GDPR compliance would require a network administrator to think about security upfront and change any default passwords immediately when designing and setting up a network for the first time. Obviously, security by design should be implemented with a U.S. organization's systems to reduce exposure regarding the FTC Act or other laws, including GPDR if EU personal data is at issue.

b. Implement Sufficient Password Complexity

In addition to easily guessed passwords, password complexity is another topic of discussion in the security world. Various sources describe preferred minimum lengths of passwords to reduce the risk of an attacker guessing the password or cracking the password via a brute-force attack (discussed above) or a dictionary attack (discussed below). Microsoft recommends a password length of at least eight characters, refraining from using common passwords, and registering for multi-factor authentication (discussed in the next section).[51] Other sources recommend using a long phrase to achieve a high level of complexity.[52]

Thanks to the *Wyndham* case and other similar decisions, hotel chains and other organizations commonly have a password complexity requirement, which employees or other users must comply with to set up a password for network access.

[49] *See* Lisa R. Lifshitz, *United States: Security by Design: California's New IoT Security Laws*, MONDAQ (Nov. 22, 2018), https://www.mondaq.com/unitedstates/security/757388/security-by-design-california39s-new-iot-security-laws (explaining California has a law requiring security by design when launching IoT devices).

[50] Commission Regulation 2016/679, 2016 O.J. (LI 19) [GDPR] at Art. 25, available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG.

[51] *Password policy recommendations*, MICROSOFT (Oct. 13, 2020), https://docs.microsoft.com/en-us/microsoft-365/admin/misc/password-policy-recommendations?view=o365-worldwide.

[52] *See Oregon FBI Tech Tuesday: Building a Digital Defense with Passwords,* FBI (Feb. 18, 2020), https://www.fbi.gov/contact-us/field-offices/portland/news/press-releases/oregon-fbi-tech-tuesday-building-a-digital-defense-with-passwords (noting "extra length of a passphrase makes it harder to crack").

c. Adopt Two-Factor/Multi-Factor Authentication

As of this writing, two-factor authentication is essentially becoming the norm for protecting sensitive data. For example, when a consumer attempts to log in to his online bank account from a new device (or from a new IP address not used previously), the banking site will typically text (or email) a one-time password to the consumer's phone. This way, a bad actor would need not only the consumer's username and password, but also his smartphone (or access to his email) to gain access to the consumer's online bank account. Regarding *Wyndham*, an organization could have password protection on its servers, and require two-factor authentication for remote access to organizational servers.[53]

d. Implement Non-Obvious Usernames

Many computing systems require a combination of a valid username and a valid password to provide access to the user attempting to access it. Frequently, a valid username is easy to guess. For example, a bad actor can notice that an organization's email accounts follow the standard FirstInitial+LastName@Domain.com and that the organization's domain usernames perhaps follow the standard LastName+FirstInitial. Consequently, a bad actor can reliably construct valid usernames —and leave his brute-forcing efforts to guess only that user's password to impermissibly gain access to his email account or laptop. Therefore, implementing non-obvious usernames can, at the very least, increase a bad actor's burden to brute-force his way to access an account. For example, the username standard LastName+FirstInitial+4DigitNumber both lengthens the username and makes it less obvious to guess to make a brute-force discovery of the username more difficult. Thus, while John Miller's domain username might ordinarily be MillerJ, the username might become MillerJ8264. This username is non-obvious and longer than MillerJ, and it may be less likely to be guessed by brute-force attempts. Of course, many organizations might resist such username complexity as it may be inconvenient for users to remember a more complex username.

e. Implement Automatic Account Disabling After Several Failed Login Attempts

Brute-forcing access to a system requires a bad actor try a series of passwords (and potentially usernames) until they find a successful combination, as discussed above. Implementing automatic account disabling whereby the account is automatically disabled (or all data wiped/erased) after a reasonable number of login attempts helps thwart brute-force attempts to access a computer system. (As an interesting side note, it's conceivable that attackers could use easily-

[53] *See Multi-Factor Authentication*, ATLANTIC.NET, https://www.atlantic.net/multi-factor-authentication/ (last visited Oct. 14, 2020 2:59 PM) (noting multifactor authentication as "one of the best ways to protect against remote attacks").

discovered usernames to trigger automatic disabling of accounts, thereby causing a form of denial of service attack.[54])

f. Provide Secure Password-Protected WiFi that Adheres to a Reasonably Strong Encryption Standard

It seems that there is decreasing prevalence of consumers and businesses setting up open Wi-Fi networks that have no password or encryption protection. An open network with no password protection/encryption could easily allow a bad actor to park his car near the Wi-Fi network and sniff the airwaves for unencrypted traffic. As of this writing, the most advanced Wi-Fi encryption standard is WPA3, but WPA2 may continue to be implemented for several years before becoming obsolete.[55] It's conceivable that someone could foolishly choose an obsolete (and thus vulnerable) encryption standard, perhaps out of a list of drop-down encryption options, when configuring a Wi-Fi router (e.g., selecting the obsolete WEP standard).[56]

11. Encryption Key Length

As noted above, organizations tend to increase the length of encryption keys every few years (e.g., from 40 bit many years ago to 128 bit to 256 bit to larger values) to make encrypted data less vulnerable to brute-force attack.[57]

To illustrate how increasing the key length decreases brute-force attack vulnerability, I provide a highly vulnerable and impractical but simple example. First, consider a two-bit encryption key. A bit can have a value of 1 or 0. Therefore, consider a two-bit key XX where X can have a value of 1 or 0. A two bit key therefore has four possible values: 00, 01, 10, and 11 (alternatively expressed as 2^2 possible combinations). Therefore, an attacker who knows there are only four possible keys would try each one until successfully guessing the correct key, thereby decrypting the data. As explained above, in a brute-force attack, the attacker tries all possible two-bit keys (all four of them) until he successfully decrypts the string of cipher gibberish into the original clear text data. A computer could very easily attempt all four decryption keys, so this two-bit key provides almost no security.

[54] *See Understanding Denial of Service Attacks,* CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY (Nov. 4, 2009), https://us-cert.cisa.gov/ncas/tips/ST04-015 (discussing denial of service attack).

[55] *See* Rahal Gupta, *What's the Difference Between WPA3 and WPA2 What's WPA3 Exactly,* GUIDING TECH (Jan. 23, 2018), https://www.guidingtech.com/wpa3-vs-wpa2/ (noting WPA3 implementation beginning in 2018).

[56] *See WEP vs WPA Encryption,* NETGEAR, https://kb.netgear.com/20043/WEP-vs-WPA-Encryption (last updated Nov. 28, 2016) (noting that the "Wi-Fi alliance highly recommends against using WEP and plans to make it obsolete.").

[57] *40-bit encryption*, *supra* note 26.

A key's complexity in this example could be increased exponentially by adding additional characters. For example, a five-character key XXXXX, where X equals 1 or 0, would have 2^5 possible combinations (i.e., 32 possible combinations). Using a realistic modern example, a 128-bit key has 2^128 (or 3.4 * 10^38 possible key values), which in theory could take many years to break. Encryption can be conceptualized as a multiplication process where the original data is multiplied by the key to encrypt and divided by the key to decrypt. Therefore, a brute-force attack can be conceptualized as dividing the encrypted data by every possible key value until successfully decrypting the data to plain text.

As of this writing, organizations have implemented symmetric encryption keys of 128-bit length or greater (for symmetric keys). Given Moore's law, encryption key lengths are likely to increase over time. It should be noted that where encrypted data is encrypted with a 128-bit key, it would be much easier to crack a short password (e.g., several characters in length) to decrypt the data rather than attempting to brute force-decrypt the data directly.[58] Put another way, attempting all possible passwords for a password several characters long is likely much faster than applying 2^128 possible keys directly against the data (especially if one employs a dictionary attack on the password as described below).

12. *A Dictionary Attack as a Faster Alternative to Brute-Force Attack*

As a faster alternative to a brute-force attack, the forensics professional may employ a "dictionary attack."[59] This works by collecting a dictionary of commonly used passwords (obtained from the internet) and trying those common passwords first in an effort to quickly crack the password.[60] A web search of common password cracking tools reveals many software programs, such as "John the Ripper" (JTR) and "Cain and Abel," which are used by forensics professionals and others to crack passwords.[61]

[58] *Why brute-force the password instead of the key directly?,* STACK EXCHANGE, https://security.stackexchange.com/questions/167868/why-brute-force-the-password-instead-of-the-key-directly (last visited Oct. 24, 2020).

[59] *See* Orin S. Kerr & Bruce Schneier, *Encryption Workarounds*, 106 GEO. L.J. 989, 997 (2018) (describing a "guess the key" encryption workaround). A guess the key strategy might involve law enforcement quite literally guessing a suspect's key and manually entering a few guesses. However, a dictionary attack is a more automated species of guess the key where an entire file of commonly used passwords is attempted by an automated program. An automated dictionary attack is described on page 999 of this article without using the term dictionary attack.

[60] *Dictionary Attack*, TECHOPEDIA, https://www.techopedia.com/definition/1774/dictionary-attack (last updated Oct. 21, 2011).

[61] Howard Poston, *10 Most Popular Password Cracking Tools*, INFOSEC (Sept. 25, 2020), https://resources.infosecinstitute.com/10-popular-password-cracking-tools/ (noting John the Ripper as popular tool and referencing Cain and Abel as a topic in this organization's

Returning to the smartphone example, a forensics professional could locate a dictionary of common integers used in four-digit passcodes. For example, one might locate a list, starting with the most common four-digit passcodes to the least common four-digit passcodes (in descending order from #1 to #10,000).[62] A dictionary attack would involve starting with the most commonly used passcodes first (i.e., the most common is 1234 and the second most common is 1111) in an effort to more quickly crack the passcode than mechanically iterating through 0000, 0001, 0002, 0003, etc. until reaching 9999.

A dictionary attack of a password on a laptop or PC follows similar principles. Human behavior, including password selection, can be fairly predictable. Therefore, certain passwords are commonly used on a laptop just as humans have common integer passcodes on a 4- or 6-digit smartphone passcode. With a dictionary attack, a forensic examiner locates a file containing thousands of hacked passwords (i.e., a dictionary of common passwords) from prior data breaches and then applies those passwords first. For example, "password," "monkey," and "superman," are often observed as within the top twenty-five most commonly used passwords.[63] Therefore, a dictionary attack works by attempting these and hundreds of other commonly used passwords first in an attempt to more quickly crack the laptop's password in comparison to trying every random combination of alphanumeric characters.

13. Using Encryption as an Attack Mechanism

While the discussion thus far has focused on encryption of stored data as a defensive measure by the data owner or the computer owner to protect information from a bad actor, it should be briefly noted that encryption can be used offensively in a ransomware attack to attack stored data. In a ransomware attack, the attacker locks the victim's data via encryption and promises to decrypt the data upon payment of a ransom.[64] One might naively believe that maintaining data backups will prevent or mitigate a ransomware attack. For example, an organization maintaining a weekly backup might operate on the premise that following a ransomware attack, it could revert back to the backup,

training course). *See also* United States v. Martin, 2019 WL 5098748, at *1 (A. Ct. Crim. App. 2019) (describing the usage of a "Greykey" program used to unlock password-protected electronic devices, such as smartphones).

[62] Akemi Iwaya, *The Most Common and Least Used 4 Digit Pin Numbers [Security Analysis Report]*, HOW-TO GEEK (Sept. 27, 2012, 10:30 AM), https://www.howtogeek.com/125378/the-most-common-and-least-used-4-digit-pin-numbers-security-analysis-report/.

[63] *List of the most common passwords*, WIKIPEDIA, https://en.wikipedia.org/wiki/List_of_the_most_common_passwords (last updated Oct. 15, 2020, 10:12 AM).

[64] *Ransomware*, CYBERSECURITY AND INFRASTRUCTURE SECURITY AGENCY, https://www.us-cert.gov/Ransomware (last visited Oct. 14, 2020, 3:15 PM).

forego paying the ransom, and lose up to a week of data in lieu of paying the ransom. However, some attacks may target data backups.[65] Therefore, there are several questions to explore. One is whether backups can be sufficiently secured on the local network or on a cloud provider's network. Also, an organization could consider air-gapping, a physical backup stored off the network for sensitive information.[66] Perhaps the most common attack vector for ransomware or other attacks is through email phishing,[67] so employee security awareness training and marking outside emails as "external" are common security measures in a robust, layered security program.

As a side note, ransomware attacks may involve a threat of data exfiltration. For example, an attacker may first extort a victim to decrypt the data. Then, the attacker may extort a second sum from a victim with the threat of publishing the data to others. If an organization has its data securely encrypted, and the attacker is unable to decrypt it, this would reduce the threat of data exfiltration as stolen gibberish is not a target for publication.

B. UNDERSTANDING HASHING VS ENCRYPTION OF STORED DATA

As a starting point, encrypting data changes its form, and the encrypted data is convertible back to its original form by decrypting it. However, a hash value of data generally cannot be converted back to the original data. Hashing is thus loosely analogous to generating a word count of a document. A word count of, say, 10,256 is not readily convertible back to the original text. Put another way, if I gave a person the value of 10,256 words and asked her to recreate my original document, it would be essentially impossible to do without more information. That being said, the value 10,256 words could be used as a tool to distinguish that document from other documents on my computer. This is the essence of

[65] David Bisson, *Ransomware Attacks Targeting Organizations' Backup Storage*, SECURITY INTELLIGENCE (Dec. 9, 2019, 12:30 PM), https://securityintelligence.com/news/ransomware-attacks-targeting-organizations-backup-data-storage/; *see also* Maria Korolov, *How to Protect Backups from Ransomware*, CSO (Jan. 14, 2019, 3:00 AM), https://www.csoonline.com/article/3331981/how-to-protect-backups-from-ransomware.html; Rod Matthews, *Ransomware Will Target Backups: 4 Ways to Protect Your Data*, DARKREADING (Oct. 4, 2017, 10:30 AM), https://www.darkreading.com/endpoint/ransomware-will-target-backups-4-ways-to-protect-your-data/a/d-id/1330029.

[66] *Air gap (networking),* WIKIPEDIA, https://en.wikipedia.org/wiki/Air_gap_%28 networking%29 (last visited Sept. 13, 2020).

[67] *See Glossary,* NAT'L INST. OF STANDARDS AND TECH., https://www.nist.gov/itl/smallbusinesscyber/cybersecurity-basics/glossary, (last updated Feb. 28, 2019) (defining "Phishing" as: "A technique for attempting to acquire sensitive data, such as bank account numbers, or access to a larger computerized system through a fraudulent solicitation in email or on a web site. The perpetrator typically masquerades as a legitimate business or reputable person.").

hashing—it is a value that can be used to identify or represent the data, such as a file, hard drive, or password, but it does not reveal the data itself.

1. *Saving Stored Encrypted Passwords vs. Saving Hash Values of Passwords*

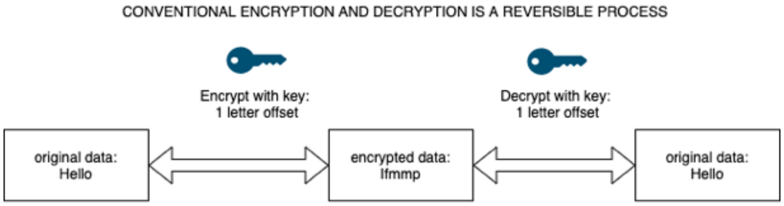
Storing a hash value of a password can be considered more secure than storing encrypted passwords in a database because the hash value is not readily convertible back to the original data. In addition, "salting" a password (as described below) prior to hashing can further improve security by making it even more difficult to ascertain a password from the hash value. For example, in *In re LinkedIn Privacy Litigation*, LinkedIn responded to a data breach of actual subscriber passwords, releasing a statement on its blog that it had switched its password encryption method from a system that stored member passwords in a hashed format to one that used both salted and hashed passwords for increased security.[68]

Generating a hash value of data (i.e., hashing) can be thought of as a cryptologic cousin of encrypting data. However, as discussed, a major difference between encrypting data and hashing data is that a hash value generally cannot be used to reconstruct the original data, while an encryption key can be used to encrypt or decrypt data. The below section describes the difference between encryption and hashing and then further discusses relevant hashing principles.

2. *Conventional Encryption/Decryption is Generally a Two-Way Reversible Process (unlike hashing)*

Conventional encryption (also called "symmetric" encryption) can be visualized as a reversible process as seen in FIG. 1 below, where the same key can be used to either encrypt or decrypt the data. (The sender and receiver use the same key to encrypt or decrypt.)

Figure 1



[68] *In re* LinkedIn, 2014 No. 5:12-cv-03088-EJD, at *1 (N.D. Cal. Mar. 28, 2014).

3. Generating a Hash Value of Data is Generally a One-Way Function

While encryption is a reversible process, generating a hash value is not. Instead, generating a hash value can be described as a one-way function. To illustrate the concept of hashing, I provide a fictitious example. Assume a hashing algorithm creates a hash value that simply counts the number of characters in a word (see FIG. 2A below).

Figure 2A

User	Original Data (password)	Hash Value Algorithm = number of characters
user1	hello	5
user2	Hello	5
user3	major	5
user4	hi	2

In light of the key hashing principle that the original data cannot be readily reconstructed from the hash value, one generally cannot reverse engineer the words "Hello" or "major" from the hash value "5". For example, if a hacker gained access to a database of stored hash values, he could not reconstruct the original password "hello" armed only with a hash value of "5", because there are many words (or random letter strings) that are five characters in length.

However, a major flaw with my fictitious hashing algorithm is that a hacker could create a random password of five characters in length that would gain access to the system for many users having five character passwords. For example, if a hacker accesses a database of stored hash values for each user, he will observe that user1 has a hash value of "5". He could enter the username user1 and any password five characters in length (e.g., 12345) to access the system. The system would calculate a hash value of the password, "5", based on his entered password (12345) and determine that this hash value matches user1, and therefore grant access (assuming two-factor authentication is not in place). Expanding on this problem, the above hashing algorithm needs to produce unique hash values. My conceptual hashing algorithm is defective because three identical hash values (i.e., 5) are created for different data inputs. Therefore, a better hashing algorithm is needed to avoid a hashing "collision" where two identical hash values are generated for different data inputs.

Real-world hashing algorithms include the MD5 and SHA1 hashing algorithms, and these algorithms generally have sufficient mathematical complexity to avoid hash value collisions for different data inputs. Either of these algorithms can be used to generate a hash value of some data input (e.g., a file to be hashed, an entire hard drive to be hashed, or some smaller data input such as a password).

Figure 2B: MD5 Hash Values for Particular Data Inputs

Original Data	Hashing algorithm	Hash value generated
Hello	MD5	8b1a9953c4611296a827abf8c47804d7
major	MD5	aa6df57fb6fe377d80b4a257b4a92cba
hello	MD5	5d41402abc4b2a76b9719d911017c592
Hello there my friends. It's great to see everyone.	MD5	272638daff573f18183dab4570bd3bc0

As seen in FIG.2B, applying an MD5 hash value to "Hello" and to "major" generates a unique MD5 hash value for each original data input.[69] The MD5 hash value is thus superior to my conceptual hashing algorithm in FIG. 2A because there is no hash value collision, meaning that the generated hash values are unique for different data inputs. Also, the MD5 hash value changes dramatically for even a minor change in the data input. For example, a lowercase "h" in the word "hello" results in a very different hash value than upper case "H" for the data input "Hello."

Another feature of a hashing algorithm is that the generated hash value has a fixed length irrespective of the size of the data to be hashed. For example, the MD5 algorithm is designed to generate a 32-character hash value, as seen in the above table, regardless of the size of the data hashed. Therefore, a unique 32-character hash value is generated whether the data input is the word Hello or the data input is all of the text of a novel, such as *War and Peace* by Leo Tolstoy. This can be seen in the bottom row where the relatively long phrase "Hello there my friends . . ." generates a 32-character MD5 hash value despite being a larger data input than the first three rows. Further, the MD5 hashing algorithm, like any

[69] These hash values were generated using an online hash generator. Dan's Tools, *MD5 Hash Generator*, https://www.md5hashgenerator.com/ (last visited Sept. 13, 2020).

other hashing algorithm, is designed so that the original data cannot be readily reconstructed from the hash value: generally speaking, just like a word count of a document is not convertible back to the text, a hash value does not convert back to the original data.

4. *How are Hash Values Used?*

Two common purposes of hash values are data fingerprinting in the forensics context or protection of stored passwords.

a. Digital Fingerprinting

Regarding data fingerprinting, a file such as a photo or other file has a hash value generated for that file. For example, the federal government maintains a database of hash values for known child pornography images; Google and other search engines generate hash values for images on their systems, and if any hash value matches a value of known child pornography image, law enforcement may be alerted, and the image is likely taken down.[70] File fingerprinting can likewise be used to determine whether a plaintiff's stolen file or other data is present on a defendant's computer. A forensics professional can prepare a hash table of all files on a defendant's computer and then look for any hash values that match the plaintiff's hash value of interest.

Fingerprinting is also used in digital forensics as a chain of custody tool.[71] For example, the government or a plaintiff in a civil suit may have a forensics professional take a digital image (i.e., make a copy) of a defendant's hard drive (or a relevant portion thereof). At the time of collection, the forensics professional typically makes multiple copies and records an MD5 or other hash value of the copied hard drive. If the defendant were to later allege that certain hard drive evidence was fabricated or falsified, the forensics professional can provide a copy of the hard drive to the defendant having an MD5 hash value that matches the hash value at the time of collection. The plaintiff or government can then invite the defendant to retrieve the same evidence from that hard drive copy. Because any minor change in data dramatically alters the MD5 hash value, the matching hash value serves as strong evidence that the hard drive's data was not changed post-collection.

[70] *See* United States v. Reddick, 900 F.3d 636 (5th Cir. 2018) ("Private businesses and police investigators rely regularly on 'hash values' to fight the online distribution of child pornography.").

[71] *See* United States v. Bout, 651 Fed. Appx. 62, 64 (2d Cir. 2016) (where defendant tried to assert defects in the chain of custody, including failure to run a hash value comparison between the mirrored drive and the original, arguing therefore that the forensic copy of the hard drive and all exhibits that relied upon its data should have been excluded from evidence as improperly authenticated).

b. Hashing Stored Passwords

As noted above, it's a common practice to store hash values of passwords (e.g., customer passwords) in a database rather than the actual passwords.[72] This way, if a hacker is able to extract password information from a database, it consists of hash value representations of the passwords that are generally not usable to reconstruct the actual passwords. However, in a rainbow attack, discussed below, a hacker or forensics professional might successfully decipher actual passwords from hash values.

In usage, a customer may log in to a website using his password. The website may apply a hashing algorithm to the customer's inputted password to generate a hash value. Then, if that hash value matches the stored hash value for that customer's username, the customer is authenticated and granted access to the system. In addition to the obvious benefit that the plaintext versions of the passwords are not available for theft, this also creates a double-blind environment wherein only the user knows the password; the service only has the hash stored (as opposed to the actual password) available to it.

Regarding hashing algorithms, it should be noted that the MD5 hashing algorithm is not sufficiently secure[73] from a cryptologic standpoint for storing hash values of passwords because the MD5 algorithm is vulnerable to a rainbow attack, (discussed below) which can potentially reconstruct original passwords from their hash values. However, the MD5 hashing algorithm may be considered suitable for data fingerprinting purposes.[74]

5. Salting a Hash Value to Inhibit Rainbow Attacks of Hashed Passwords

In 2012, Linkedin suffered a data breach of over six million passwords.[75] Although hashed values of the passwords were stored, these values were

[72] *See In re LinkedIn*, 2014 No. 5:12-cv-03088-EJD, at *1 (noting usage of storing salted hash values for user passwords).

[73] As of this writing, SHA2 (of varying key lengths) hashing algorithms are considered more secure than MD5. *See* Patrick Nohe, *Rehashed: The Difference Between SHA-1, SHA-2, and SHA-256 Hash Algorithms,* HASHEDOUT (Nov. 9, 2018), https://www.thesslstore.com/blog/difference-sha-1-sha-2-sha-256-hash-algorithms/; *SHA-1,* WIKIPEDIA (Oct. 10, 2020, 7:48 AM), https://en.wikipedia.org/wiki/SHA-1#Attacks; *3 Reasons Why MD5 is not Secure,* MD5, https://www.md5online.org/blog/why-md5-is-not-safe/ (last visited Oct. 24, 2020).

[74] For example, cases within the last few years describe forensics professionals using MD5 hash values for fingerprinting purposes. *See, e.g.*, Campbell Alliance Group, Inc. v. Dandekar, 2014 WL 145037, at *1 (E.D.N.C. 2014) (discussing the process of a forensic examiner collecting an MD5 hash value when making a copy (i.e., forensic image) of a party's hard drive).

[75] Brid-Aine Parnell, *LinkedIn admits site hack, adds pinch of salt to passwords*, THE REGISTER (June 7, 2012, 10:03 AM), https://www.theregister.com/2012/06/07/linkedin_admits_data_breach/.

vulnerable to attack because they were not "salted."[76] Salting a hash value is a further encryption step to inhibit rainbow attacks of hash values.[77] A rainbow attack involves generating a rainbow table, which can be simplistically thought of as a table of hash values corresponding to particular passwords.[78] While this is not entirely accurate (the math is much more complicated), this is conceptually what the attack involves. To create a salted hash value, an organization can concatenate (i.e., combine) a random string (i.e., a salt) with a consumer's password and then apply the hashing algorithm to the combined salt + password as seen below.[79]

Figure 2C

Username	**Password**
user1	password123
user2	password123

Username	**Salt Value**	**String to be hashed**	**Hashed value= SHA256 (password + hash value)**
user1	E1F53. . .	password123E1F53. . .	72AE25495A7981C40622 D49F9A52 . . .
user 2	84Bo3. . .	password12384Bo3. . .	B4B6603ABC670967E99 C7E7F1389. . .

[76] *Id.*

[77] *Salt (cryptography)*, WIKIPEDIA, https://en.wikipedia.org/wiki/Salt_(cryptography) (last updated Aug. 26, 2020, 7:29 AM). *See* In re LinkedIn, No.: 5:12–CV–03088–EJD, 2014 WL 1323713, at *1 n.1 (N.D. Cal.) (2014) ("'[S]alting' is an encryption process that protects information by concatenating a plaintext password with a series of randomly generated characters prior to hashing.").

[78] For a description of a rainbow table, see *Rainbow Table*, WIKIPEDIA, https://en.wikipedia.org/wiki/Rainbow_table (last updated Feb. 16, 2021, 8:40 PM).

[79] The table is copied from Wikipedia, but the various salt values, string to be hashed, and hashed value are shortened/truncated for ease of illustration. *Salt (cryptography)*, WIKIPEDIA, https://en.wikipedia.org/wiki/Salt_(cryptography) (last updated Aug. 26, 2020, 7:29 AM).

The principle here is that the attacker would need to know the salt value and then use this value in constructing a rainbow table to determine actual passwords. The salt is therefore an additional barrier to a rainbow attack. An organization could increase resistance to rainbow attacks by increasing the length of the salt string, using different salt values for different users, and other methods.

C. DATA IN TRANSIT: UNDERSTANDING CORE NETWORKING PROTOCOLS AS CONTEXT FOR UNDERSTANDING ENCRYPTION OF DATA IN TRANSIT

Protecting stored data with encryption is a worthwhile goal, but encrypting data in transit is another helpful protection measure. The following discusses general concepts relative to common data transmission concepts (i.e., data in transit).

Understanding the difference between data in transit and data at rest can be helpful from a legal standpoint, such as courts assessing the application of the Electronic Communications Privacy Act (ECPA) 18 U.S.C. §§ 2510-2523, for data in transit for wiretap orders and application of the Stored Communications Act (SCA) 18 U.S.C. §§ 2701-2712 for stored data.[80] At least one court has generally addressed the concept of data at rest versus data in motion in the context of the Communications Assistance for Law Enforcement Act (CALEA) 47 U.S.C. §§ 1001-1010.[81] When considering whether law enforcement should have a front door to encrypted data (the debate outlined in Part 3), it is helpful to understand the difference between an ephemeral symmetric key for encryption/decryption of data transmitted during a single browsing session and a longer-life key for stored data.[82]

1. *How Data is Transmitted: Data Packets*

As a first general point (as noted in *In re DoubleClick Inc. Privacy Litigation* and as noted previously*),* computers communicate by exchanging information packets.[83] Typically, a large file is split into multiple packets to facilitate transmission, and the packets are reassembled at the receiving end. Each packet has a header and a payload. The header can be thought of as an envelope or a label with "to" and "from" information (source computer IP address and destination computer IP address). The payload is the data. Packets can be

[80] *See* David W. Opderbeck, *Encryption Policy and Law Enforcement in the Cloud*, 49 CONN L. REV. 1657, 1668-1669 (2017) (discussing the legislative history and applicability of the ECPA and the SCA); *see also*, *In re Yahoo Mail Litigation*, 7 F. Supp. 3d 1016, 1026 (N.D. Cal. 2014) (discussing application of the ECPA to data in transit and the SCA to stored data).

[81] *In re Order Requiring Apple, Inc. to Assist in the Execution of a Search Warrant Issued by This Court,* 149 F. Supp. 3d 341, 355-363 (E.D.N.Y. 2016).

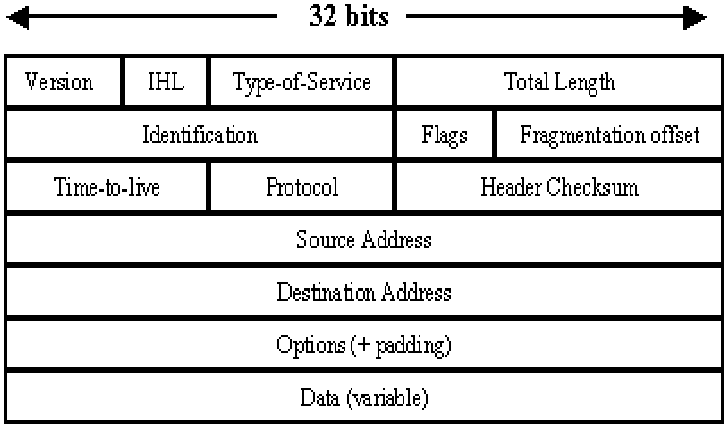
[82] *See* Opderbeck, *supra* note 80, at 1667 (discussing sessions keys being discarded in ephemeral communications).

[83] 154 F. Supp. 2d 497, 501 (S.D.N.Y. 2001).

classified as either "connectionless" or "connection-oriented" as described in detail below. Data is sent in data packets across the internet and into local networks because files are often too large to be transmitted in a single stream of bits. For example, an ethernet cable can only transmit roughly 1500 bytes in an ethernet frame, so a large file greater than 1500 bytes in size must typically be split up into data packets and then reassembled at the receiving computer.[84]

When teaching law students about data packets, which consist of an IP header and data, it is helpful to provide a 10,000-foot view of data packets rather than teaching all details. For example, the following data packet has too much detail:

A Data Packet with Too Many Details



<table>
<tr><td>Version</td><td>IHL</td><td>Type-of-Service</td><td colspan="2">Total Length</td></tr>
<tr><td colspan="3">Identification</td><td>Flags</td><td>Fragmentation offset</td></tr>
<tr><td colspan="2">Time-to-live</td><td>Protocol</td><td colspan="2">Header Checksum</td></tr>
<tr><td colspan="5">Source Address</td></tr>
<tr><td colspan="5">Destination Address</td></tr>
<tr><td colspan="5">Options (+ padding)</td></tr>
<tr><td colspan="5">Data (variable)</td></tr>
</table>

Testing law students on all details of the above data packet is likely overkill. It is certainly acceptable to display and discuss briefly, but extensive teaching and testing of all data packet components seem more suitable for IT students.[85] A more simplified view of a data packet is shown below and is likely more appropriate as a context for understanding the IT concept generally.[86]

[84] See *Maximum Transmission unit,* WIKIPEDIA, https://en.wikipedia.org/wiki/Maximum_transmission_unit (last updated Sept. 26, 2020, 4:35 PM), for a discussion of maximum transmission unit and reference to ethernet frame.

[85] See *Network packet*, WIKIPEDIA, https://en.wikipedia.org/wiki/Network_packet (last updated Sept. 22, 2020, 12:19 PM), for a general discussion of data packet components.

[86] All details of a data packet are omitted as this would result in a 0 foot/in the weeds level of detail more appropriate for an IT professional. A more complete view of a data packet may be viewed at Michael Mullins, *Exploring the anatomy of a data packet,* TECHREPUBLIC (July 2, 2001, 12:00 AM), https://www.techrepublic.com/article/exploring-the-anatomy-of-a-data-packet/.

A Simplified Data Packet

HEADER					**DATA/ PAYLOAD**
Source IP address	Destination IP address	Source Port	Destination Port	Protocol or other technical details	here is a recipe for the perfect margarita. . .

As discussed above, the header can generally be thought of as an envelope, and the data payload can be thought of as a letter within the envelope. Much like a physical envelope needing sender and recipient street address, a data packet needs both a sender address (source IP address) and a recipient (destination IP address). Standard data transmission protocols are discussed below, starting with protocols that do not use encryption to understand better sending encrypted data mechanisms.

2. *Common Encryptionless Protocols Involved in Data Exchange: IP, UDP & HTTP*

Several primitive networking protocols are briefly introduced here for context and expanded upon in later sections; none of them use encryption.

As a first general point, a "protocol" is essentially an agreed-upon procedure for how two computers will communicate.[87] A lawyer could thus view a protocol as a contract between two computers, such as agreeing on a mechanism for uniquely identifying each computer (i.e., the internet protocol discussed below) or an agreement as to further technical steps for sending and receiving an email (e.g., SMTP: simple mail transfer protocol). When transmitting data, two computers will typically rely on multiple protocols to transfer data according to a particular process (e.g., a computer communicating with a web server or two computers exchanging emails). For example, the TCP and UDP protocols discussed below rely on the Internet protocol (IP) as an underlying protocol (e.g., perhaps viewing IP as a first agreement upon which to build further agreements).

Various protocols used by computers have been developed in large part by the Internet Engineering Task Force, which publishes protocols via RFCs (Requests for Comments).[88] Having such protocols or standards in place is critical for computers to interact. Otherwise, an Apple computer might not

[87] *Protocol,* TECHOPEDIA (Apr. 24, 2020), https://www.techopedia.com/definition/4528/protocol.

[88] *RFCs*, IETF, https://www.ietf.org/standards/rfcs/ (last visited Feb. 4, 2021).

communicate appropriately with a Dell computer if it operated according to its protocols rather than common protocols.

Perhaps the most important protocol is the Internet Protocol (IP), which requires each internet-connected computer to utilize a common addressing scheme.[89] IP is thus the foundation of the internet. An IP Address is an internet subscriber's unique identifier for browsing the web. Much like a driver's license number for surfing the web, the IP Address is a series of digits (such as 78.192.192.244) that identifies the user's computer to make connections with other computers.[90] To illustrate the need for a standardized form of unique IP addresses, imagine if computer users named their internet-connected computers by first and last name. The problem with this approach is that users with common names, such as John Smith, would be unable to uniquely identify their computers, and web servers and other computers would end up sending data to the wrong John Smith.

Internet Service Providers ("ISPs") lease unique IP addresses/identifiers for home internet connections, and these leased IP addresses may remain the same for many years.[91] This background on IP addresses is relevant to privacy because an IP address can be used to identify an individual and his browsing history and other data. In fact, in copyright BitTorrent cases, groups of John Doe defendants have been initially listed by their IP address (where plaintiff's attempt to later discover each defendant's actual identity entailed issuing a subpoena to an ISP to disclose name and address associated with that ISP subscriber's IP address).[92]

Regarding data packet transmission, a sending computer could choose to send either a UDP (User Datagram Protocol) packet or a TCP (Transmission Control Protocol) packet. UDP, as explained below, is a "connectionless" protocol, which might be described as "send it and hope it gets there" (because the receiving computer sends no confirmation of receipt or notice of failed delivery), while TCP is connection-oriented such that the two computers confirm that each data packet was successfully received.

[89] *Internet Protocol*, INFO. SCIENCES INST. (Sept. 1981), https://tools.ietf.org/html/rfc791.

[90] Steven J. Vaughan-Nichols, *Static vs. Dynamic IP Addresses*, AVAST (Sept. 23, 2019), https://www.avast.com/c-static-vs-dynamic-ip-addresses.

[91] For example, observe that while "many homes . . . have theoretically dynamic IP's, [they tend to hold] the same IP for multiple years." *How Long Does an IP Address Stay Attached to a Home or Business?,* ELTORO, https://www.eltoro.com/how-long-does-an-ip-address-stay-attached-to-a-home-or-business/ (last visited Oct. 12, 2020).

[92] For general discussion, *see* Eric Goldman, *Strike 3's Copyright Litigation Campaign Completely Strikes Out,* TECHN. & MKTG. LAW BLOG (Nov. 2, 2019), https://blog.ericgoldman.org/archives/2019/11/strike-3s-copyright-litigation-campaign-completely-strikes-out.htm.

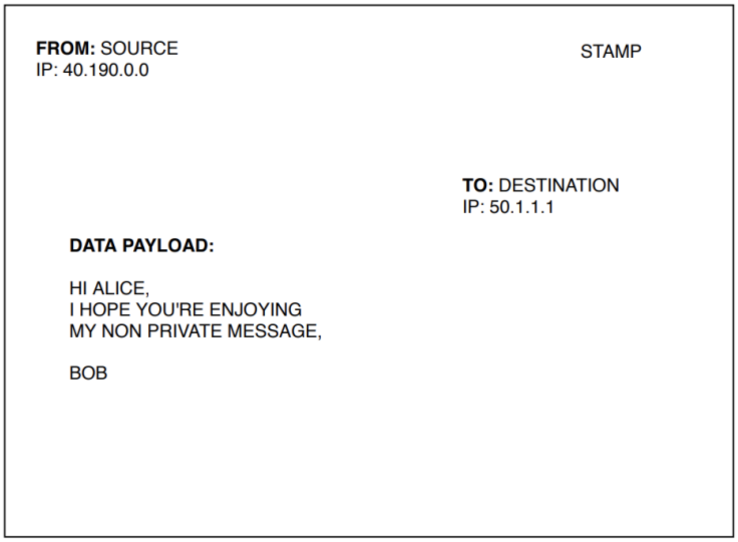
3. UDP is Like Sending a Postcard and Hoping it Gets There

A very simple and common protocol for sending data packets is UDP. It is a simple, no frills "connectionless" protocol for simple, fast, unencrypted data transmission.[93] It is similar to mailing a postcard (because the postal carrier, and any eavesdropper, can easily see the sender and receiver's identities and the message (i.e., the data payload)). Physical postal analogies, while imperfect, are helpful to understand relevant IT concepts and are used in various sections below.[94]

UDP is considered "connectionless" because the sender does not receive automated notice of whether packets failed to successfully arrive. This can be compared to mailing a postcard and not caring or checking whether it successfully arrives. For example, imagine mailing a postcard from a vacation spot and not sticking around for any return to sender notice; the sender simply hopes it arrives.

Figure 3 below conceptually depicts a UDP Packet.

Figure 3



FROM: SOURCE
IP: 40.190.0.0

STAMP

TO: DESTINATION
IP: 50.1.1.1

DATA PAYLOAD:

HI ALICE,
I HOPE YOU'RE ENJOYING
MY NON PRIVATE MESSAGE,

BOB

[93] J. Postel, *User Datagram Protocol,* INFO. SCIENCES INST. (Aug. 28, 1980), https://www.rfc-editor.org/rfc/rfc768.txt.

[94] *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F.Supp.3d 1099, 1112 (N.D. Cal. 2017) (noting "[a]n analogy to physical space, while inevitably imperfect when analyzing the digital world, may be helpful.").

4. TCP, a Component of HTTP/HTTPS, is Somewhat Like Using Certified Mail (and Carefully Tracking Notification of any Failed Delivery)

Another very common protocol introduced above is TCP, characteristically used for both HTTP and HTTPS transmissions.[95] TCP is different from UDP in that TCP is "connection-oriented" and thus "reliable." This means that the sender and receiver have error checking, checking whether all data packets have been received and not lost or corrupted in transmission. As explained more fully later, TCP is helpful to ensure accurate receipt of all data packets, such as an attorney sending a file to a client without lost, missing, or altered words (even after that large file has been split up into multiple packets and reassembled at the receiving end).[96]

5. HTTP, which Integrates TCP, is thus like Sending a Postcard via Certified Mail (and Carefully Tracking Notification of any Failed Delivery).

HTTP has historically been a very common protocol for web browsing (i.e., exchanging data between a user's computer and a web server computer).[97] HTTP data exchange typically occurs over TCP. Therefore, sending an HTTP/TCP data packet is like sending a postcard via certified mail (and looking out for delivery problems). In the HTTP/TCP packet, the sender, receiver, and data are visible to an internet eavesdropper (just like a UDP packet), but the sender and receiver have the reliable (connection-oriented) TCP connection, which notifies the sender of any lost or corrupted data packets, so that he may resend.

As an analogy, suppose someone wishes to send a long message split into two certified postcards and wants to confirm that each postcard arrived to the desired sender. Assume he can use certified mail for the postcards, which alerts the sender of successful or unsuccessful delivery, and which the sender carefully tracks. He marks the first postcard "1 of 2" and the second postcard "2 of 2" and sends them certified. The certified mail process will alert him to any delivery problems, and he can expect to receive some alert if one of the packets (i.e., postcards) did not arrive. TCP is like this certified mail scenario. Unlike the "connectionless" UDP, TCP is referred to as providing "reliable" (i.e., connection-oriented) delivery. As TCP's name implies (transmission *control* protocol) the transmission of data is *controlled* and monitored for delivery

[95] *Transmission Control Protocol*, RFC, https://tools.ietf.org/html/rfc793 (last updated July 29, 2020) (discussing the mechanics of TCP).

[96] *Important Application Layer Protocols: DNS, FTP, SMTP, And MIME Protocols*, SOFTWARE TESTING HELP, https://www.softwaretestinghelp.com/dns-ftp-smtp-mime-protocols/ (last updated Sept. 13, 2020) (describing the MIME application layer protocol, which supports email attachments, running over TCP).

[97] Henrik Nielson, et al., *Hypertext Transfer Protocol—HTTP/1.1*, RFC, https://tools.ietf.org/html/rfc2616 (last updated Jan. 21, 2020).

problems. Both HTTP and its encrypted variant HTTPS use TCP's reliable delivery system, which causes lost or corrupted data packets to be resent. As noted previously, multiple protocols may be involved in a particular type of data transmission. The HTTP protocol is one good example of a protocol that incorporates other protocols as it relies on both IP and TCP.

6. *Modern Web Servers Tend to use HTTPS Rather than HTTP*

Before discussing the details of HTTPS, I note that "HTTPS" protocol is now the norm for web servers, replacing usage of the HTTP protocol.[98] Generally, the "S" of HTTPS stands for "secure" and requires an *encrypted* data payload transmission between a host device (e.g., a consumer's laptop) and a website/web server and also requires authentication of the web server through a digitally signed third-party certificate (discussed in more detail in section VII below). This push for increased HTTPS usage was encouraged in part by a 2016 campaign by the Electronic Frontier Foundation to increase online privacy protection.[99]

7. *HTTPS over TCP is like Sending a Letter in an Envelope via Certified Mail and Requiring the Recipient to Show ID for Receipt*

To begin distinguishing HTTP from HTTPS, continued postal analogy may be helpful. As discussed, HTTP is the internet equivalent of sending a postcard (but with careful delivery tracking), because any internet eavesdropper can readily see: who the card is from (source IP address); who the card is to (destination IP address); and the contents of the postcard's message (the data payload). However, HTTPS adds encryption, encrypting the data as it is divided into packets, such that the payload portion of each data packet is encrypted. One can visualize writing a long letter, encrypting it, and then cutting it up into separate pieces, with each piece having its own envelope to hide the data. At the receiving end, the pieces are then reassembled in the proper order and decrypted. Technically, the HTTPS protocol improves on HTTP by adding an encryption layer, referred to as a transport security layer (TLS) on top of HTTP.

Transport security layer is a reference to the OSI model of computing, the open system interconnection model, which is an abstract model to describe how computers process information and communicate with each other.[100] The transport layer of the OSI model is the layer describing two computers agreeing

[98] *Hypertext Transfer Protocol*, WIKIPEDIA, https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol (last updated Oct. 12, 2020, 8:05 AM).

[99] *Encrypting the Web,* ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/encrypt-the-web (last visited Oct. 14, 2020).

[100] *OSI Model*, WIKIPEDIA, https://en.wikipedia.org/wiki/OSI_model (last updated Oct. 5, 2020, 6:16 AM).

to communicate together, such as a laptop establishing a communication session with a website/webserver.[101]

As alluded to above, the encryption of an HTTPS packet can thus be visualized as an envelope that hides the payload, but the outside of the envelope (i.e., the source and destination IP address) is visible to eavesdroppers. The HTTPS protocol also includes authentication of the web server, via a website certificate, as an initial step before exchanging encrypted data (this web certificate authentication process is explained in section VII below). This could be visualized as requiring the recipient of a certified letter to show identification for receipt. HTTPS is therefore more secure than HTTP, because HTTPS authenticates that data exchange is occurring with a legitimate web server.

Checking whether a website/web server is using HTTPS simply requires examining the URL in the browser bar. For example, typing "depaul.edu" into the browser bar triggers a visit to DePaul's web server, and the browser bar displays "https://www.depaul.edu/Pages/default.aspx." The HTTPS connection ensures that the data payload transmitted between a user's laptop and DePaul's web server is encrypted. This way, if an eavesdropper were to intercept data packets between the two parties, he would see mostly encrypted gibberish rather than clear text data. As an example of an eavesdropping attack, a bad actor could visit DePaul's campus and connect to its local Wi-Fi network. He could then use a free program such as Wireshark to "sniff" (i.e., see) all traffic in real-time on the local network.[102] Any HTTP traffic from other users on the network would be readily visible to the eavesdropper (just like postcards in plain view). For this reason, as noted above, most sites will not allow HTTP connections to their web servers to reduce the risk of having sensitive data visible to such an eavesdropper (e.g., account information, user names, passwords). Organizations using the HTTPS protocol for their web servers are essentially reducing their legal exposure (e.g., negligence claims and potential regulatory actions) by reducing eavesdropping attacks on their website visitors.

A visual of HTTPS communication is provided in Figure 4 below:

[101] *See* Peter Swire, *Privacy and Security A Pedagogic Cybersecurity Framework*, VIEWPOINT (Oct. 2018), https://peterswire.net/wp-content/uploads/Pedagogic-cybersecurity-framework.pdf, for a more detailed discussion of the OSI model as it relates to legal education; *see* Volini, *supra* note 19 at 38-39, my article offering ideas related to Swire's framework.

[102] WIRESHARK, https://www.wireshark.org/ (last visited Oct. 12, 2020).

Figure 4 HTTPS Packets (Conceptual Illustration)

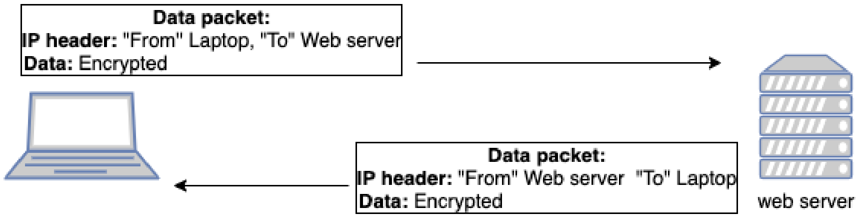
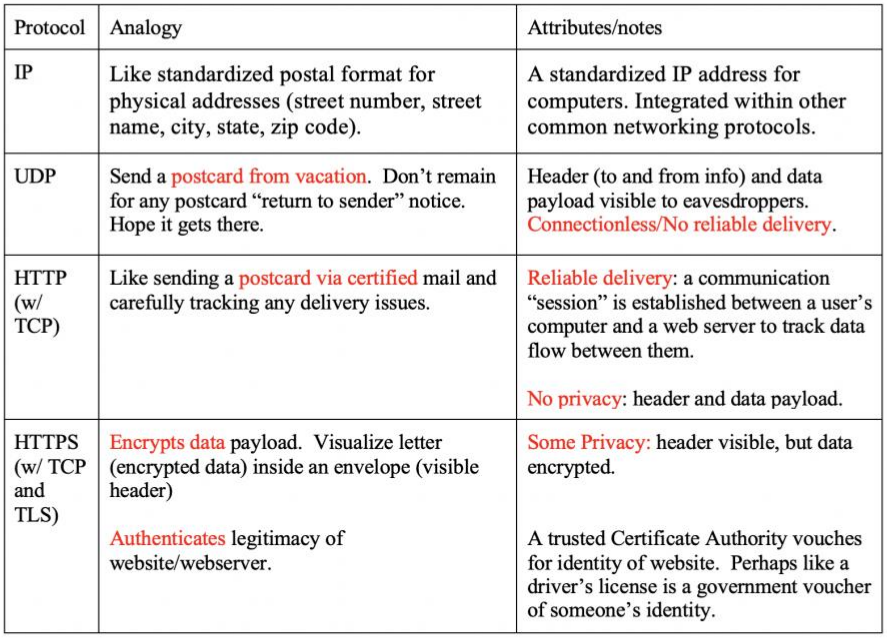

Data packet:
IP header: "From" Laptop, "To" Web server
Data: Encrypted
Data packet:
IP header: "From" Web server "To" Laptop
Data: Encrypted
web server

Figure 4 shows that the header information is visible (unencrypted), but the data is encrypted. In this regard, the header information needs to remain unencrypted so that routers on the internet know where to send the data packets. Further, the return address is needed to facilitate responsive data packets. Accordingly, using the HTTPS protocol encrypts the data from an internet eavesdropper's view, but cannot hide the identity (i.e., IP address) of the sender and receiver from his prying eyes.

A summary of the foregoing protocols is provided in the table below:



Protocol	Analogy	Attributes/notes
IP	Like standardized postal format for physical addresses (street number, street name, city, state, zip code).	A standardized IP address for computers. Integrated within other common networking protocols.
UDP	Send a postcard from vacation. Don't remain for any postcard "return to sender" notice. Hope it gets there.	Header (to and from info) and data payload visible to eavesdroppers. Connectionless/No reliable delivery.
HTTP (w/ TCP)	Like sending a postcard via certified mail and carefully tracking any delivery issues.	Reliable delivery: a communication "session" is established between a user's computer and a web server to track data flow between them. No privacy: header and data payload.
HTTPS (w/ TCP and TLS)	Encrypts data payload. Visualize letter (encrypted data) inside an envelope (visible header) Authenticates legitimacy of website/webserver.	Some Privacy: header visible, but data encrypted. A trusted Certificate Authority vouches for identity of website. Perhaps like a driver's license is a government voucher of someone's identity.

8. *Further Context for Data in Transit: Granular Technical Steps Involved in Visiting a Website*

Knowing what goes on "behind the scenes" enhances understanding of security and privacy concepts. When someone types a web address into the browser on her laptop (e.g., CNN.com), her laptop first performs some form of DNS (domain name service) query because the laptop needs an IP address to connect with the CNN web server (CNN has several web servers to handle its significant website traffic, and the IP address for one of those servers at the time of this writing is 151.101.193.67).[103] If CNN.com is one of her commonly visited sites, her browser probably has the IP address saved in a temporary browser memory (i.e., an unexpired DNS record is available in the temporary browser cache).[104] If the an unexpired IP address/DNS record of CNN.com is not in the cache, however, her laptop needs to reach out to a DNS server. In her home environment, this could be her ISP's DNS server (or in her law school, it may be the school's local DNS server).

When establishing an HTTPS connection, the laptop connects to port 443 of the CNN web server in order to establish a communication "session" between the laptop and the web server. After the browser authenticates the legitimacy of the web site via a third-party web certificate (see section VII below), the laptop and web server essentially agree to use the HTTPS protocol for exchanging encrypted data packets. Once the laptop and web server agree to communicate,[105] the web server sends HTML and other commands to the user's

[103] In this regard, all internet connected computers must use the "Internet Protocol" (IP), which requires all internet connected computers to use a common addressing scheme when communicating with each other. *See What is the Internet Protocol?*, CLOUDFLARE, https://www.cloudflare.com/learning/network-layer/internet-protocol/ (last visited Oct. 14, 2020) (discussing IP addresses and other aspects of the IP protocol); *see Internet Protocol*, WIKIPEDIA, https://en.wikipedia.org/wiki/Internet_Protocol (last updated Oct. 12, 2020, 9:23 AM) (for additional general discussion).

[104] If the Laptop is running a MS Windows operating system, then there is typically a Windows DNS cache (in addition to a browser DNS cache), and the laptop can check the Windows cache before having to query a DNS server. Bradley Mitchell, *DNS Catching and How It Makes Your Internet Better*, LIFEWIRE, https://www.lifewire.com/what-is-a-dns-cache-817514 (last updated May 1, 2020). DNS records in a cache expire frequently, having a TTL (time-to-live), ranging anywhere from perhaps thirty seconds to 24 hours; this means that DNS queries are still frequent even with DNS caches. *TTL Best Practices: the Long and Short of It,* DNS MADE EASY (Aug. 18, 2017) https://social.dnsmadeeasy.com/blog/long-short-ttls/. As a side note, items in temporary cache memory can be a helpful source of forensics information in a legal proceeding. *See, e.g.*, United States v. Romm, 455 F.3d 990, 993 (9th Cir. 2006) (involving forensic analysis of cached web pages in connection with a child pornography charge).

[105] This process of agreeing to establish a session is referred to as a three-way handshake. *See, e.g.*, *Explanation of the three-way handshake via TCP/IP*, MICROSOFT (Sept. 21, 2020), https://docs.microsoft.com/en-us/troubleshoot/windows-server/networking/three-way-handshake-via-tcpip.

laptop. These commands instruct the laptop what colors, text, and functional features to display on the laptop's screen.

Port 443 is one of the "well-known" ports in computing (ports 1-1023 are referred to as well-known ports).[106] One can think of a port like a channel on a CB radio. For example, if one wishes to communicate with truck drivers, he would select channel nineteen on the CB radio (or channel nine for emergency usage).[107] Similarly, a laptop can connect with Port 443 of a web server for an HTTPS connection or Port 80 for an HTTP connection (as discussed, most sites now prohibit Port 80/HTTP connections).[108]

With CB radio communication both parties tune to the same channel (e.g., channel nineteen). With HTTPS connection, a user's laptop connects to port 443 of the web server from a random ephemeral port of her laptop (the random ephemeral port typically has a value within the range of 1025 to about 65,000 in MS Windows operating systems).[109] The user's laptop selects a specific random ephemeral port (such as Port 22,123) to facilitate maintaining a specific port-to-port connection between the user's laptop and the web server during their browsing session. In this example, port 22,123 can be useful to distinguish two computers on the same network that are both visiting CNN.com from the same public IP address.

For a more in-depth discussion, I provide a detailed half hour video lecture on my YouTube channel of the above steps relative to a computer visiting a website.[110] This is one of the first week's lecture videos in my online Data Privacy Law course at DePaul University.[111]

D. VPNS CAN ENCRYPT BOTH THE HEADER AND THE PAYLOAD OF DATA PACKETS

Before discussing a virtual private network (VPN), it may be helpful to discuss the simple concept of a *private* network. For example, two internal

[106] *List of TCP and UDP port numbers*, WIKIPEDIA, https://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers (last visited Oct. 12, 2020).

[107] *CB Radios, Frequencies, & Channels*, CB WORLD, https://www.wearecb.com/cb-radio-frequencies-channels.html (last visited Oct. 12, 2020).

[108] Likewise, a laptop could attempt to connect with port 22 of the web server to send administrative commands to the web server via the secure shell protocol (the SSH protocol is associated with port 22). As a side note, the organization's firewall should block such an attempted port 22 connection and only allow such administrative connections from an authorized IP address.

[109] *See Ephemeral port*, WIKIPEDIA, https://en.wikipedia.org/wiki/Ephemeral_port (last updated Sept. 9, 2020) (discussing random ephemeral ports generally).

[110] Cybersecurity Patent Professor, *How the internet works*, YOUTUBE (Sept. 6, 2019), https://youtu.be/UESJCDX40Wk.

[111] I teach Data Privacy Law: US & EU, which surveys US and European privacy law, and about 25-30% of the course content is relevant technology education.

phones or computers in a home or business that never connect to the outside world could be considered a private network. A VPN is considered "virtual" because it creates a private network connected over essentially public internet infrastructure. A VPN creates privacy virtually by using encryption to hide data as it travels through public space. A VPN allows a sender of data to hide not only the data in the data packets, but also the sender's identity and potentially the ultimate receiver's identity. A VPN can essentially function as a messenger who carries the data packet in a hidden manner across the internet. A VPN typically improves on HTTPS transmission from a privacy standpoint because HTTPS will encrypt the data payload of a data packet, but not encrypt the source and destination IP address (see the unencrypted header in FIG. 4 above).[112]

Before further technical discussion, it's important to note that not all VPN services are created equal and some will essentially sell the VPN user's data (e.g., sites visited) to third parties, thereby defeating the whole purpose of using the VPN service to anonymously browse the web.[113] In addition, some VPN service providers maintain logs of their users' internet activity — thus defeating the purpose of using the VPN service to anonymously browse the web if these logs are discoverable. A computer user may wish to hide his IP address via VPN for legitimate defensive purposes or to hide criminal activity.[114] A consumer can accomplish this typically by using a VPN service, which typically provides proxy service.[115] Another option, often used by organizations, is to utilize what is called a site-to-site VPN (see FIG. 6 below), to encrypt data packets exchanged between two locations (e.g., between two corporate networks).

[112] Steven J. Vaughan-Nichols, *Static vs. Dynamic IP Addresses*, AVAST (Sept. 23, 2019), https://www.avast.com/c-static-vs-dynamic-ip-addresses.

[113] *See* Rob Mardisalu, *How FREE VPNs Sell Your Data*, THEBESTVPN (May 3, 2018), https://thebestvpn.com/how-free-vpns-sell-your-data/ (noting that free VPN services in particular are prone to sell the user's data to third-parties).

[114] Steve Symanovich, *What is a VPN?*, NORTON (Jan. 14, 2021) https://us.norton.com/internetsecurity-privacy-what-is-a-vpn.html.

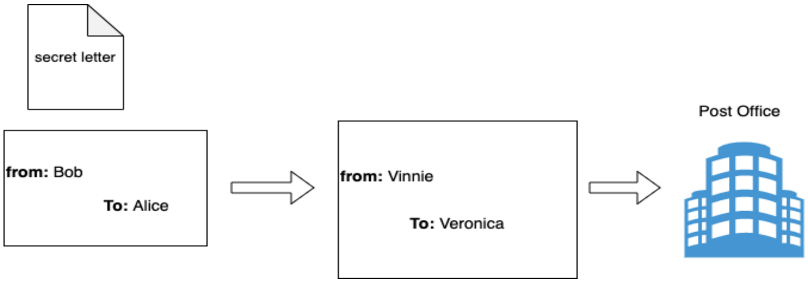
[115] For example, suppose that Bob wishes to view sumo wrestling on WatchSumo.com but does not want anyone to know that he is a sumo fan. He may sign up with BadAssVPN (https://www.badass.sx/product/badass-vpn-service-prepaid/) to use their proxy service. *See BadAss, Sx. VPN service*, BADASS.SX, (https://www.badass.sx/product/badass-vpn-service-prepaid/ (last visited Sept. 13, 2020). Instead of logging in directly to WatchSumo.com, he connects to BadAssVPN, which then forwards his login packet to WatchSumo.com. A traffic analyzer can see the packets between him and BadAssVPN. Another traffic analyzer (or perhaps the same one) can see the traffic between BadAssVPN and WatchSumo.com. Because the data is encrypted, no one can readily match Bob's packets to the ones between BadAssVPN and WatchSumo.com. However, it should be noted that if only a handful of people are connected to BadAssVPN at the time, the timestamps of the packets can be used to prove that Bob is using BadAssVPN as a proxy to connect to WatchSumo.com. This precise situation has occurred when a user was using TOR (The Onion Router) but only a few people were connected, thereby permitting law enforcement to catch him.

1. *Using a VPN is like Hiding the Sender's Letter Within an Intermediary Messenger's Envelope*

Continuing with postal analogies, using a VPN is typically like sending a letter inside of an envelope and then placing that envelope inside a messenger's larger envelope. This way, an eavesdropper can only see the messenger as the sender. Also, all of the data is hidden by the messenger's large envelope. The messenger (i.e., VPN device or service) is referred to as a "proxy."

To illustrate the analogy of sending US mail through an intermediary, consider FIG. 5 below. Suppose Bob in Chicago wishes to send a letter to Alice in New York, but he doesn't want the U.S. Postal Service to know he is communicating with Alice. Bob creates an envelope with Bob as the sender and Alice as the recipient. He then places his letter into his envelope. Next, Bob asks Vinnie the VPN service to mail the letter to Alice. Vinnie creates his own larger envelope, labelling the sender as Vinnie and the recipient as Veronica, his New York associate. Vinnie then places Bob's letter inside this larger envelope and drops it at the post office. All the post office officials can see is the envelope from Vinnie to Veronica, and it cannot see Bob's envelope inside nor the letter to Alice inside Bob's envelope. In IT terminology, this process of essentially placing an original envelope within a larger envelope is referred to as "encapsulation," shown conceptually in FIG. 5 below.

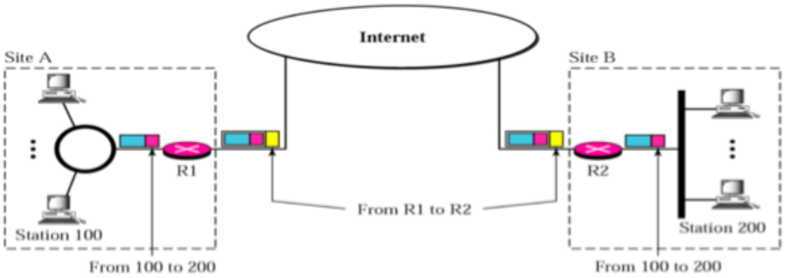
Figure 5 Physical Envelope Analogy to VPN



A VPN service essentially provides this functionality of hiding the original sender and ultimate receiver. The VPN service is the intermediary (Vinnie) which handles sending the message, and the message is sent through the internet rather than the U.S. postal network. Rather than hiding the original sender, the ultimate receiver, and the data with physical envelopes, the VPN service encrypts this information within its own larger data packet. The larger data packet shows a source IP address owned by the VPN, so eavesdroppers on the internet only see the VPN's source IP address. Further, with the VPN site-to-site arrangement shown in FIG. 6 below, an eavesdropper would see another VPN's address as the destination IP address rather than Alice's address.

To illustrate VPN concepts in a more direct sense, the following Figure 6 shows a site-to-site VPN.

Figure 6



In FIG. 6, the goal is to send a data packet from Station 100 on the left to Station 200 on the right. Station 100 wants to hide that the message comes from Station 100 and also wants to hide that the destination is Station 200. This goal is accomplished by sending the data packet through VPN routers R1 and R2. The original data packet on the far left has two key components: data (shown in blue) and source and destination IP address (from Station 100 to Station 200) shown in red.

VPN router R1 encrypts the entire data packet, encrypting both the data and the source and destination IP address. R1 then "encapsulates" the original data packet by adding its own IP header (shown above as "From R1 to R2").[116] This encapsulation step is essentially the IT equivalent of placing an original envelope from Bob to Alice within a larger envelope labeled from R1 to R2. The communication between R1 and R2 is often described as traveling through an "encryption tunnel," and this phrase provides a good visual for the concept of a VPN protecting data packets from eavesdroppers.

2. *Consumer Usage of a VPN Service to Encrypt Header and Data Payload of Packets*

The above corporate site-to-site VPN discussion involved two VPN devices (one proxy VPN device for each end or endpoint of the communication). However, a consumer typically signs up for a free or paid VPN service with the

[116] This VPN scenario demonstrates an example of "encapsulation," a fundamental concept in computer networking. In a VPN context, the sender's and receiver's IP addresses are encapsulated within the source and destination IP addresses of perhaps two VPN routers. *See generally* admin, *What is Encapsulation in computer networking?*, COMPUTER NETWORKING DEMYSTIFIED (July 12, 2013), http://computernetworkingsimplified.in/category-1/layering/encapsulation-decapsulation/ (describing encapsulation generally).

primary goal of hiding data packets from her ISP, perhaps rather than ensuring complete encryption along the entire path between sender and receiver.[117] As discussed, using a VPN service could be for legitimate or illegal purposes.[118] For example, in the 2019 Capitol One breach, the defendant, charged under the CFAA, used a VPN service as an obfuscation technique (i.e., a hiding technique) to make it difficult or impossible for others to identify her actual IP address in connection with her hacking activity. Despite the use of this obfuscation technique, her hacking activity was nonetheless traced back to her by the comments she made on social media, bragging about her exploit.[119]

Figure 7: A VPN Service

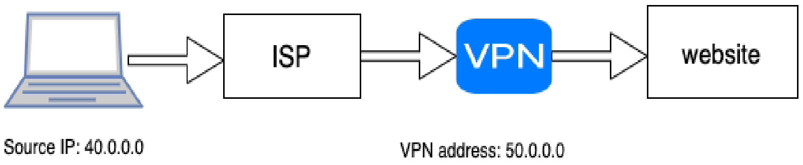


FIG. 7 above conceptually shows a consumer connecting to a website through a VPN service. The idea here is that the ISP, such as AT&T or Comcast, only sees the consumer connecting to the VPN. The VPN then acts as an intermediary between the laptop and the website (the VPN service acting as a proxy to send and receive traffic to/from the website). To use the VPN service, the consumer downloads software from the VPN service, and the software allows the consumer to encrypt data packets between her laptop and the destination website.

When the laptop is using the VPN service, the ISP may see data packets that conceptually look as follows:

[117] *How to Really Hide Your IP address with a VPN-2021*, VPNMENTOR (Jan. 26, 2021), https://www.vpnmentor.com/blog/how-to-really-hide-your-ip-address-with-a-vpn/.

[118] *See* Adam Marshall, *The best VPN service 2021*, TECHRADAR (Jan. 28, 2021), https://www.techradar.com/vpn/best-vpn (giving examples of various VPN services and describing VPN usage as a tool for privacy to keep one's online life "anonymous from prying eyes").

[119] *Capital One Data Theft Impacts 106M People*, KREBS ON SECURITY (July 30, 2019), https://krebsonsecurity.com/2019/07/capital-one-data-theft-impacts-106m-people/.

Source IP	Destination IP	Ultimate Destination Website and Data
Laptop 40.0.0.0	VPN 50.0.0.0	Encrypted Gibberish

As noted previously, an individual can use a VPN service for legitimate or illicit purposes. For example, a consumer may use a VPN service to inhibit eavesdropping attacks when connecting to her banking website.[120] Some organizations, however, will block known VPN IP addresses, which would prevent legitimate users or bad actors from visiting the site through a VPN.[121] In another example, a bad actor may wish to use a VPN service to hide the fact (from his ISP) that he is visiting a child pornography site or a site facilitating illegal downloads of copyrighted movies.

3. *Possible Privacy Shortcomings of VPNs*

a. *Subpoena Resistance by VPN Service and/or Absence of Records*

One issue arising in civil and criminal litigation, noted above, is whether the VPN service will honor a subpoena. Some do provide information, while other overseas VPN services will not respond to a subpoena.[122] In the latter circumstance, a civil plaintiff or the government may need other evidence to prove the defendant visited some illicit site, such as through seizure of the defendant's computer for evidence of browsing history. Another issue with VPNs is that some VPN providers market themselves as not maintaining customer logs of their browsing activity, attracting users in search of this high level of privacy.[123]

[120] Eric Geier*, Protecting Against Wi-Fi Eavesdropping*, TECHGENIX (July 10, 2012), http://techgenix.com/protecting-against-wi-fi-eavesdropping/.

[121] Patrick Marshall, *Why do some banks not allow for me to log on through a secure VPN?*, SEATTLE TIMES (Jan. 17, 2020, 10:00 PM), https://www.seattletimes.com/business/technology/why-do-some-banks-not-allow-me-to-log-on-through-a-secure-vpn/.

[122] And, in some circumstances a particular VPN service might not maintain any logs of customer browsing activity helpful to law enforcement. Ray Walsh, *10 best no log VPNs to use in 2021 | Zero-logs and no tracking*, PROPRIVACY, https://proprivacy.com/vpn/comparison/best-no-logs-vpns (last updated Jan. 08, 2021)*;* Nick Pearson, *Can Commercial VPNs Really Protect Your Privacy?*, TECHDIRECT (Apr. 3, 2013, 11:54 PM), https://www.techdirt.com/articles/20130402/02421422545/can-commercial-vpns-really-protect-your-privacy.shtml.

[123] Paul Bischoff, *Best Logless VPNs in 2021*, COMPARITECH, https://www.comparitech.com/blog/vpn-privacy/best-logless-vpns/ (last updated Nov. 30, 2020).

b. DNS Leaks

A first technical issue to assess with VPN usage is whether the VPN service is susceptible to DNS leaks. DNS leaks may reveal the user's browsing activity to her ISP thereby defeating the privacy goal of using the VPN service to hide browsing activity from the ISP.[124] With a DNS leak, the user's browser makes direct DNS queries to the ISP's DNS server rather than making such requests through the VPN's server.[125] An example of a DNS leak might involve a user's computer directly asking the ISP for the IP address of stdwatch.com rather than having the VPN proxy send the DNS request for this IP address: the effect is that the ISP is now aware that the user visited a site concerning STDs.

c. Browser Fingerprinting

A second technical issue to assess is whether the government or other party could use "browser fingerprinting" to reveal a user's identity despite the usage of a VPN.[126] For a simple example, consider the following. Assume law enforcement track a user's VPN browsing activity (e.g., sites visited within a time period). They then compare that VPN activity to the browsing activity of a previously used public IP address (i.e., the IP address of a defendant's home internet). Law enforcement may statistically demonstrate a very high likelihood that the same user performed both instances of browsing. The preceding two technical issues, along with issues of whether a VPN service maintains logs or honors subpoenas, demonstrate that the use of a VPN service may not guarantee the privacy of a user's browsing.

E. USING ENCRYPTION TO IMPROVE PRIVACY AND SECURITY OF DNS

In recent years, security and privacy discussions have arisen over DNS, including options to potentially improve the privacy of DNS queries.[127] Before assessing potential DNS improvements, attorneys would benefit from first understanding traditional DNS and its privacy shortcomings.

[124] *DNS leak*, WIKIPEDIA, https://en.wikipedia.org/wiki/DNS_leak (last updated Oct. 13, 2020, 7:32 PM); *How to Test for a DNS Leak with Legitimate Results*, THE ST. OF SECURITY (Mar. 15, 2018), https://www.tripwire.com/state-of-security/featured/test-dns-leak-legitimate-results/.

[125] *DNS leak, supra* note 124.

[126] Jon Watson, *How to protect yourself against invisible browser fingerprinting*, COMPARITECH, https://www.comparitech.com/blog/vpn-privacy/what-is-browser-fingerprinting-how-to-protect-yourself/ (last updated Oct. 9, 2020).

[127] Gareth Tyson & Tim Böttger, *DNS-over-HTTPS: why the web's latest privacy tech is causing an outcry*, THE CONVERSATION (Oct. 29, 2019), https://theconversation.com/dns-over-https-why-the-webs-latest-privacy-tech-is-causing-an-outcry-125188.

1. *How does DNS Work Generally?*

People often describe DNS as the phone book of the internet.[128] For example, recall from section IV a consumer visiting a website that she has not recently visited. She types the domain name into her browser bar (e.g., CNN.com). If a temporary memory cache, such as her browser's DNS cache, does not store the IP address for CNN.com on her laptop, her laptop must request the IP address from a DNS server (e.g., likely her ISP's DNS server). Just as one uses a person's name to look up their number in a phone book, the DNS server uses the name "CNN.com" to look up CNN's numeric IP address.[129] The laptop then uses the IP address to visit the website (i.e., establish a session with CNN's web server).[130]

2. *Understanding the DNS Hierarchical Structure*

Usually, a user's computer or DNS server caches the IP addresses of commonly visited websites. These DNS servers are typically either an ISP's DNS server or a local DNS server, such as an enterprise server on a university's network. This caching of IP addresses helps to prevent bombarding the internet's hierarchical DNS system with traffic.

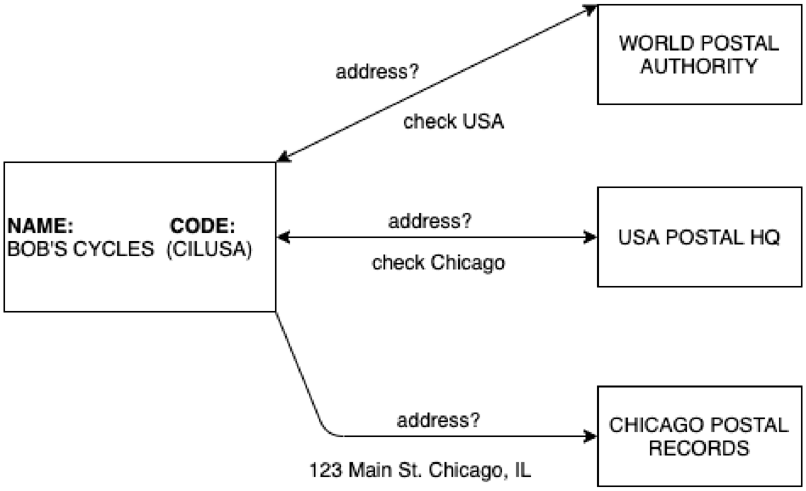
To understand at a high level how an uncached DNS query works, it seems helpful to start with a hypothetical postal analogy as seen in Figure 8A below.

[128] Tim Fisher, *What is DNS (Domain Name System),* LIFE WIRE (Nov. 27, 2019) https://www.lifewire.com/what-is-dns-domain-name-system-2625855.

[129] Like asking Siri to call one's mother, the phone is not dialing M-O-M. Your phone instead uses the 9-digit number (i.e., the IP Address) associated with the name (i.e., domain name) to connect you to the other phone; *What is DNS?*, AMAZON, https://aws.amazon.com/route53/what-is-dns/ (last visited Oct. 13, 2020).

[130] *See What is DNS?*, AMAZON, https://aws.amazon.com/route53/what-is-dns/ (last visited Oct. 13, 2020) (presenting DNS from Amazon Web Services*).*

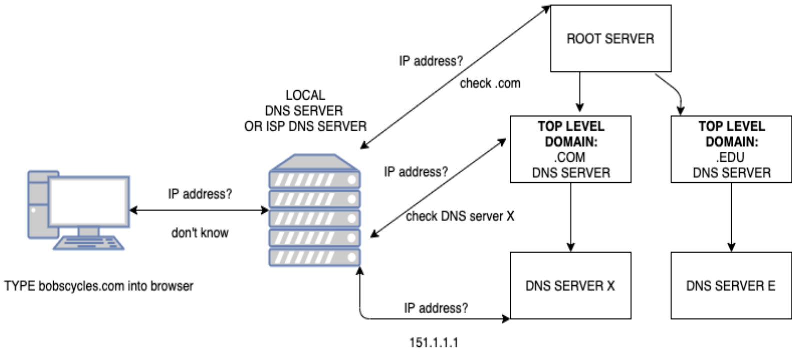
Figure 8A: A Hypothetical Hierarchical Postal Address Lookup Process



In Figure 8A, Alice wants the postal address for Bob's Cycles. Assume there is only one "Bob's Cycles" in a fictitious worldwide business registration system. The only information Alice has is "Bob's Cycles" followed by a code CILUSA. She queries the world postal authority for the address. The world postal authority sees the USA portion of the postal code and tells her to check the USA Postal HQ. The USA Postal HQ then directs her to Chicago Postal Records, which sends her the address, 123 Main St. Chicago, IL.

DNS queries follow a somewhat analogous process, as seen in Figure 8B below.

Figure 8B Hierarchical DNS Query



In Figure 8B, assume Alice types bobscycles.com into her browser bar. Her computer does not cache the IP address, so the computer asks the local DNS server for the IP address. The DNS server, however, may also not know the IP address. The local DNS server must then look for the address through a hierarchical process starting with the root server. The root server observes the ".com" in the domain name and points to the top-level domain DNS server for .com. This in turn directs a query to DNS server X, which has the bobscycles.com IP address of 151.1.1.1.[131]

3. Improving DNS Privacy with Encryption

Traditional DNS queries rely on UDP, and thus, are inherently not private: recall that one can visualize a UDP message as a postcard, which is viewable by the postal carrier and other prying eyes along the route (see FIG. 3 above). More specifically, one can visualize the DNS query as a consumer with a home IP address of say 50.1.1.1 sending a postcard to his ISP's DNS server with the data payload "what is the IP address for someembarrassingsite.com?" The consumer's ISP might be AT&T, Comcast, Verizon, etc. The ISP then has the ability to easily aggregate this data. The ISP can then sell the data to a third-party or use it to deliver targeted ads to the consumer (e.g., advertise embarrassing product A on the ISP's search engine or other ISP controlled sites).

Various privacy articles have discussed DNS over HTTPS (DoH) as a possibility of improving privacy.[132] For an attorney to offer meaningful leadership in this privacy discussion, she would at least need a basic understanding of traditional DNS, including the UDP protocol which traditional DNS relies on, as well as the basic operation of the HTTPS protocol. Fortunately, both the UDP and HTTPS protocols are discussed above, so the discussion here of these basic concepts might not be overwhelming.

DoH attempts to solve the problem of an ISP easily viewing a consumer's visited sites by hiding the data payload (e.g., "what is the IP address for someembarrasingsite.com?") with HTTPS payload encryption. The goal is to hide the DNS query from the ISP and from any other internet eavesdroppers. The consumer can send encrypted DNS queries to a private DNS server rather

[131] This process of the local DNS server checking each server can be accomplished in a recursive or iterative manner. *See What is Recursive DNS?*, CLOUDFLARE, https://www.cloudflare.com/learning/dns/what-is-recursive-dns/ (last visited Oct. 11, 2020) ("[T]he client does a form of delegation in a recursive DNS query. It tells the DNS resolver, 'Hey, I need the IP address for this domain, please hunt it down and don't get back to me until you have it.' Meanwhile, in an iterative query, the client tells the DNS resolver, 'Hey, I need the IP address for this domain. Please let me know the address of the next DNS server in the lookup process so I can look it up myself.'").

[132] *See* Tyson & Böttger, *supra* note 127 (generally discussing DoH mechanics and some concerns raised by it).

than the ISP's DNS server. If the ISP or other eavesdropper looked at the consumer's DNS packet, all they will see is the header information (i.e., from: consumer's home IP address, to: the DNS server's IP address) and the encrypted payload.

A lawyer with a basic understanding of HTTPS packets would understand that DoH does not provide perfect privacy regarding an ISP's ability to determine a consumer's visited websites. In this regard, while the ISP will not be able to see the requested website in the consumer's DNS query, the ISP still has the ability to see header information in the subsequent data packets when the consumer actually visits the site (i.e., when visiting someembarrasingsite.com after the DNS query). The result is that the ISP will need to do more work to assess which sites a consumer is visiting by analyzing header info in data packets between the consumer and various sites. This is conceivably more work than printing out (or otherwise tracking) a listing of traditional unencrypted DNS queries the consumer has made in a given time period.

Another interesting issue to next consider is whether the consumer's privacy with regard to DNS queries is still fully protected with the new DoH system. Mozilla's Firefox browser was the first browser to offer DoH by default.[133] Certainly, the DoH feature makes it more difficult for third-parties, such as the consumer's ISP or other prying eyes, to easily build a list of visited sites. However, a question arises of what Mozilla will do with this data now that it stands in the shoes of where the ISP formerly was? For example, Mozilla is in a position to decrypt the DNS queries and theoretically sell the consumer's visited website data (or alternatively sell targeted ads that Mozilla can deliver to the consumer).

4. *Side Note on DNS Attacks*

As a side note on DNS security, a variety of DNS attacks are possible that direct a user's computer to a malicious website rather than the intended website typed into the browser bar. For example, an attacker could infect the browser's DNS cache or infect a DNS server to cause the user's computer to visit a malicious site. In this scenario, typing CNN.com into the browser bar could direct the browser to the malicious IP address rather than IP address for CNN's web server. In another scenario, a user's computer could send DNS queries to a bad actor's computer that is an impostor of a legitimate DNS server.

Different encrypted DNS mechanisms can be used to mitigate DNS attacks as discussed in later sections.

[133] Yash Wate, *What is DNS over HTTPS and how to enable it on all browsers?*, TECHPP, https://techpp.com/2020/07/21/dns-over-https-guide/ (last updated July 21, 2020).

5. BGP Attacks: A Type of DNS Attack

In 2008, the Pakistani government shut down YouTube for several hours by having its telecom authority announce to the internet that its location was the destination for YouTube's IP address space.[134] This essentially created a black hole where users around the globe attempting to connect to YouTube were routed to defunct Pakistani servers. The Pakistani motivation behind advertising this false route was to prevent local Pakistanis from viewing certain offensive content at the time related to the Prophet Mohammed.[135] This was essentially a BGP (Border Gateway Protocol) attack, discussed below.[136]

While DNS addresses the question of "where do I go" (i.e., what's the numeric IP address for a domain name?), BGP addresses the question of "how do I get there"? Once a computer performs a DNS query for an IP address (say, of some website), the system of routers constituting the internet must determine where to send the data packets to reach the IP address. This can be loosely analogized to a consumer writing an address on an envelope, and then a postal worker, functioning like a router, has to determine how to get there using street signs.

Routers on the internet share/broadcast routing information with each other via the BGP protocol (e.g., Router X says "come to me for IP address range 150.1.1.1 to 150.1.1.255"), and routers trust each other's information. Large organizations, such as commercial ISPs and universities, "peer" with each other essentially informing/broadcasting to each other of which IP addresses they are able to send data packets to. These peering organizations are known as autonomous systems (ASs), and they are registered through ARIN (the American Registry for Internet Numbers).[137] Peering is performed by each ASs having specialized BGP peering routers which inform other ASs of which IP addresses they can connect to. Thus, the internet can be viewed as a collection of ASs worldwide connected by miles of fiberoptic cables that span the world's oceans and land masses.

The Pakistani YouTube shutdown referenced above was caused by ASs trusting each other (i.e., their BGP routers trusting each other). This raises a question of how to avoid this issue in the future. One mechanism could involve organizations selecting specific ASs that they can trust and authenticating that

[134] Declan McCullagh, *How Pakistan knocked YouTube Offline (and how to make sure it never happens again)*, CNET (Feb. 25, 2008, 4:28 PM), https://www.cnet.com/news/how-pakistan-knocked-youtube-offline-and-how-to-make-sure-it-never-happens-again/.

[135] *Protests: Pakistan PM Orders YouTube Shutdown,* SKY NEWS (Sept. 17, 2012, 2:53 PM) https://news.sky.com/story/protests-pakistan-pm-orders-youtube-shutdown-10469856.

[136] Yakov Rekhter, et al., *A Border Gateway Protocol 4 (BGP-4)*, RFC 4271 (Jan. 2006), https://tools.ietf.org/html/rfc4271.

[137] *Autonomous System Numbers*, ARIN, https://www.arin.net/resources/guide/asn/ (last visited Oct. 11, 2020).

they are indeed communicating with the trusted AS. A trusted AS could be authenticated with public key encryption (and untrusted ASs could be ignored). The authentication process could work much like third-party certificates used with the HTTPS protocol as discussed in connection with FIG. 12 below. In addition, vendor products and services could be explored, such as Akamai's Prolexic solution designed to protect against BGP and other attacks.[138]

6. *Further Background: Traditional DNS Relies on UDP for its Speed and Simplicity*

UDP is a very simple, fast protocol for transmitting data. UDP may have been chosen as the protocol for DNS queries to reduce traffic congestion on the internet (given the high volume of DNS requests from so many computers worldwide). Further, another potential advantage of traditional DNS using UDP is that UDP does not require the overhead of establishing three-way handshakes; thus a DNS server can potentially serve more computers' IP address queries than if it were burdened by TCP.[139]

For a small data packet, any speed difference between TCP and UDP is likely negligible. However, speed differences may become more significant for large data packets. UDP is faster than TCP for large data packets for two key reasons.[140] First, as discussed, UDP is a connectionless protocol and can be thought of as "send the data and just hope it gets there." In contrast, TCP (used by HTTP/HTTPS) takes the time (perhaps a few milliseconds) to confirm with the receiver that all data packets arrived and will resend any lost data packets. Second, UDP does not authenticate the identity of the receiver. Thus, HTTP's/HTTP's initial step of confirming the other party's identity (with TCP) can likewise add fractional seconds to the data transmission.

For these reasons, UDP has traditionally been a good option for sending live video data packets.[141] For example, resending live video data often makes no sense. Thus, error checking has been traditionally considered inappropriate for live video transmission because it often does not make sense for a recipient's computer to request a resend of a fractional second of video data. Resending packets of lost video data does not make sense as it is often more appropriate to simply accept the lost video data and continue watching the live video feed. Therefore, UDP is still an often-favored protocol for live video transmission.[142]

[138] *See generally* AKAMAI, https://www.akamai.com/us/en/products/security/prolexic-solutions.jsp (last visited Oct. 14, 2020) (discussing usage of BGP route advertisement change to mitigate particular attacks).

[139] Martin Pramatarov, *Why Does DNS use UDP?*, CLOUDNS (Jan. 23, 2018, 1:51 AM), https://www.cloudns.net/blog/dns-use-udp/.

[140] Divya Varnwal, *Why UDP is preferred for Live Streaming*, OODLES TECHNOLOGIES (May 27, 2016), https://www.oodlestechnologies.com/blogs/Why-UDP-is-preferred-for-Live-Streaming/.

[141] *Id.*

[142] *Id.*

Video transmission involves a massive quantity of audio and visual data transfer. Therefore, encryption can also be problematic in terms of slowing down video transfer (certainly live, real-time transfer).

Historically, in the context of DNS requests, simple unencrypted UDP requests were considered appropriate for simple IP address requests. After all, when DNS over UDP was developed, transmission of an IP address likely did not seem to constitute sensitive data requiring encryption. Regarding error checking, certified mail requires a postal delivery person to confirm receipt, which requires time and energy. If speed is the goal, then ordinary U.S. mail (i.e., UDP) is a good option. UDP was thus chosen for DNS queries years ago as computers needed to quickly look up many IP addresses when a user browses the web.

Another concept is data size. A DNS query has very little data: it is simply a request for an IP address, so there is not much need for error checking if all data was received. In contrast, other data transmissions, such as between a laptop and a web server, may involve substantially more data exchange (e.g., files split into multiple packets) and may thus benefit from TCP's reliable delivery with error checking. UDP was a sensible option for DNS queries during the development of the internet. UDP provided fast and easy data transmission for simple small packet DNS queries. UDP was certainly a practical option before society developed substantial security and privacy concerns.

Encrypted DNS solutions use a process of encapsulation, conceptually similar to the encapsulation principles described above with respect to VPNs. As discussed previously, with DoH, the plain text DNS request is encrypted within an HTTPS data packet, and the HTTPS packet is then sent to the DNS server. As of this writing, there is considerable online debate of what type of improved DNS is optimal from a privacy and security standpoint. From a policy standpoint, rather than engaging in a highly technical debate of which specific solutions are best, perhaps it is best for attorneys to consider the goals of an improved DNS system in terms of privacy and security principles (and then discuss options with the engineering community). Likely the two most important principles for attorneys to raise in such discussions are *authentication* and *confidentiality*, discussed below.

7. *Authentication of DNS Servers*

A first question to consider for future DNS systems is whether there is a substantial need for a user's computer to authenticate that it is communicating with a legitimate DNS server (to reduce instances of DNS poisoning or hijacking attacks). Authentication is provided by DNS using TLS associated with the HTTPS protocol, so DoH is one form of improved DNS to consider.[143]

[143] *See DNSSEC-What Is It and Why Is It Important?*, ICANN, https://www.icann.org/resources/pages/dnssec-what-is-it-why-important-2019-03-05-en

An argument can be made that authentication of many DNS servers is unnecessary or perhaps not as important as confidentiality. In this regard, if the user's computer received an illegitimate IP address from an illegitimate DNS server, HTTPS with its website certificate function, would hopefully alert the user that the website lacks a valid certificate and warn the user not to continue establishing a session with the website. DNS authentication could provide an additional layer of defense. Another argument for lower priority of DNS server authentication is that if a user's computer is already so compromised to have its DNS configuration altered, then that computer may already be so severely compromised that it is past the point of no return from a security standpoint.

8. *Confidentiality of DNS Requests*

Confidentiality/privacy of DNS requests may be a more important goal than authentication of DNS servers from a privacy standpoint (and authentication of the DNS server could be considered more of a cybersecurity protection as opposed to consumer privacy protection). DNS over HTTPS (DoH) increases confidentiality of consumer DNS requests but, as explained above, DoH does not ensure perfect confidentiality because even if the consumer's ISP cannot see the DNS queries, it can still examine other HTTPS packet headers to determine the sites the consumer is visiting.

It would seem that a privacy advantage of encrypted DNS (DoH or other variants of encrypted DNS) is that an ISP would need to sift through and examine various packets to determine sites visited rather than *easily* seeing all unencrypted DNS requests from a consumer's computer. Historically, ISPs could easily aggregate unencrypted DNS requests from consumers' computers and then sell that information (e.g., a list of IP addresses visited by each consumer). With DoH, it is more difficult for the ISP to aggregate that information because the ISP would need to survey and sift through each consumer's data packets, looking for destination IP addresses in those packets and then perhaps translating those IP addresses into domain names. This seems like considerably more work than simply maintaining a regular log of DNS requests for each consumer and having that information ready for sale.

In essence, any form of encrypted DNS will not provide perfect anonymity/confidentiality with respect to websites visited, but the extra work required to extract this information might help consumers to some extent with maintaining some greater level of privacy. Of course, as noted above, an encrypted DNS mechanism may increase privacy relative to the ISP, but the consumer is then subject to any privacy protections or possible abuses of the provider of the encrypted DNS service.

(last visited Oct. 19, 2020) (explaining that DNSSec is another form of security augmented DNS that authenticates the DNS server using public key encryption).

F. EXPLORING DETAILS OF ASYMMETRIC ENCRYPTION

Prior discussion has briefly noted that asymmetric encryption (also known as public key encryption) is used for authentication (e.g., authenticating a DNS server or authenticating a digital signature, such as a third-party website certificate in connection with HTTPS). Further details are provided in this section.

For a thorough understanding of encryption, it is helpful to understand the difference between conventional/symmetric encryption and asymmetric encryption (also known as public key encryption). Recall from FIG. 1 that in conventional encryption, the same key (such as a password) is used to both encrypt and decrypt data. For example, a password on a laptop is a symmetric key: assuming the laptop has local encryption turned on, the password is used as the key to encrypt all the data and that same key is used to decrypt data.

Recall that a Caesar Cipher is likewise symmetric. Both the sender and receiver of the message use essentially the same key to encrypt and decrypt (e.g., + three letters for each character to encrypt and - three letters to decrypt).[144] Put another way, the encryption key and the decryption key are the same.

1. *Asymmetric Encryption (also known as Public Key Encryption)*

Asymmetric encryption differs from conventional encryption in that two different keys are used: a private key and a public key that are mathematically paired together. Either of the keys is used to encrypt the data, and the other key is used to decrypt the data; however, data is generally encrypted using the public key and decrypted using the private key in order to reduce cryptographic attacks.[145] The two unique paired keys are generated with an algorithm that is sufficiently complex to generate a pair of keys unlike any other keys in the world (e.g., such that it's lottery odds for another computer anywhere in the world to generate an identical pair of public and private keys).

Public key encryption was created as an alternative to symmetric key encryption for convenience and security purposes. Suppose an attorney has 100 clients and wishes to exchange encrypted messages with each of them. Without public key encryption, the attorney could share one conventional symmetric key with all 100 clients. (The attorney and all of his 100 clients would use this one shared key to encrypt all messages.) 100 people sharing a secret runs the risk that the shared key might fall into the wrong hands, compromising confidentiality for all 100 clients. To reduce this risk, the attorney could create 100 shared keys

[144] Peter Swire & Kenesa Ahmad, *Encryption and Globalization*, 13 COLUM. SCI. & TECH. L. REV. 416, 427 (2012) (noting in symmetric encryption the same key is used on both ends of the communication).

[145] *Public-key cryptography*, WIKIPEDIA, https://en.wikipedia.org/wiki/Public-key_cryptography (last updated Oct. 8, 2020, 3:37 PM).

(one conventional key for each client), however this is burdensome to manage, and there is still some risk of an individual client having her key stolen. Public key encryption overcomes these problems: the attorney can share one public key with all 100 clients. His clients can encrypt messages with the attorney's public key and send those messages to the attorney. The only party in the world who can decrypt these messages is the attorney because he owns the corresponding private key mathematically paired with his public key.

Besides the above advantage of avoiding the need to create 100 conventional keys for each client, asymmetric encryption is also very useful for authenticating the sender of a message. From a legal standpoint, authenticating the sender of a message is useful in regard to the legal concept of nonrepudiation.[146] Repudiate means to deny, and nonrepudiation means that the sender of a message is unable to easily deny that he sent the message (message could encompass sending a digital signature relative to a contract).[147] With nonrepudiation, case law typically assesses the effects of repudiation in a contract setting and provides that the non-repudiating party may treat the repudiation as a breach and allow the non-repudiating party to stop performing.[148] However, asymmetric encryption from a tech standpoint is aimed at preventing a repudiation in the first place (such as limiting the ability of a party to deny that he digitally signed a contract) because asymmetric encryption mathematically proves that the party's private key was used to digitally sign.

This section will discuss asymmetric encryption in two contexts: (1) authenticating the sender of an encrypted message in a generic sense, and (2) authenticating website certificates (a component of the HTTPS protocol) as a specific example of the process. Turning to the first example, assume the sender of a message, Bob, has two unique keys. One is Bob's private key, and the other is Bob's public key. Bob generates this pair of keys with a suitably complex algorithm, such that the two keys are unique: as noted above, there are no other keys in the world with matching values.

[146] Bernstein v. U.S. Dept. of State, 974 F.Supp. 1288, 1292 (N.D. Cal. 1997) ("NRC identified four major uses of cryptography: ensuring data integrity, authenticating users, facilitating nonrepudiation (the linking of a specific message with a specific sender) and maintaining confidentiality.").

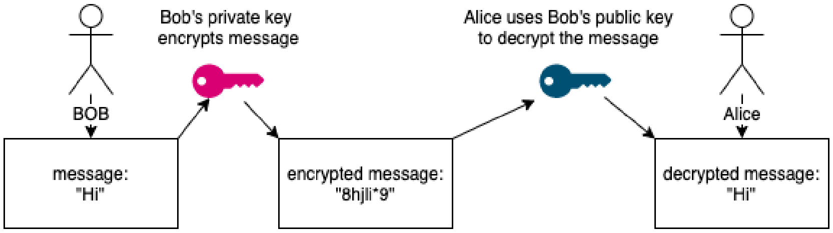
[147] *Advantages and Disadvantages of Asymmetric and Symmetric Cryptosystems*, U. OF BABYLON, www.uobabylon.edu.iq/eprints/paper_1_2264_649.pdf (last visited Dec. 3, 2019).

[148] *See* Tower Investors, LLC v. 111 East Chestnut Consultants, Inc., 864 N.E.2d 927, 940 (Ill. App. Ct. 2007) (citing Yale Development Co. v. Aurora Pizza Hut, Inc., 95 Ill.App.3d 523, 526, 51 Ill.Dec. 409, 420 N.E.2d 823, 825 (1981)); 23 R. Lord, *Williston on Contracts* § 63:33, at 561–62 (4th ed. 2002)) ("[W]hen one party repudiates a contract, the nonrepudiating party is excused from performing (*see, e.g., Curtis Casket Co. v. D.A. Brown & Co.,* 259 Ill.App.3d 800, 806, 198 Ill.Dec. 145, 632 N.E.2d 204, 209 (1994)) or may continue to perform and seek damages for the breach.")

2. *Encrypting with Sender's Private Key Authenticates the Message Sender*

Assume Bob wishes to send a message to Alice and authenticate to Alice that he is the sender. Bob will encrypt the message with his private key as seen in FIG. 9 below.[149] Building off of FIG. 9, further basic concepts of asymmetric encryption are explained below.

Figure 9


Bob's private key encrypts message
Alice uses Bob's public key to decrypt the message
BOB
Alice
message: "Hi"
encrypted message: "8hjli*9"
decrypted message: "Hi"

3. *The Encryption Step in Asymmetric Encryption is a "One-Way" Function. The Key used to Encrypt Cannot then be Used to Decrypt*

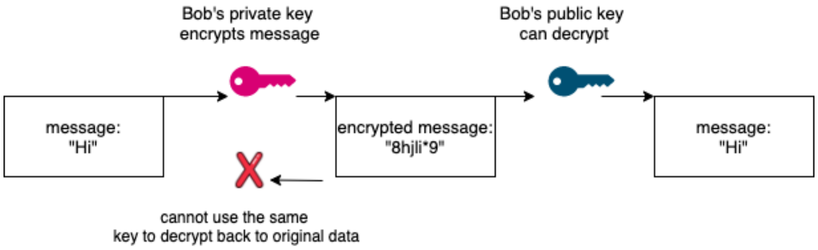
Once Bob encrypts his message with Bob's private key, his private key cannot be used to decrypt the message (see FIG. 10 below). The encryption algorithm is designed so that encryption with the private key is mathematically a "one-way function."[150] The only way to decrypt the message is by using Bob's corresponding public key. Bob has shared his public key with Alice and potentially the general public. Learning the underlying mathematics supporting this one-way function is likely overkill for most attorneys, but various sources describe the underlying mathematics.[151]

[149] *See generally Public Keys and Private Keys in Public Key Cryptography*, SECTIGO (June 9, 2020) https://sectigo.com/resource-library/public-key-vs-private-key (discussing private key usage as a means to authenticate).

[150] *See generally Public Key Encryption*, TUTORIALSPOINT, https://www.tutorialspoint.com/cryptography/public_key_encryption.htm (last visited Nov. 26, 2019); Peter Swire & Kenesa Ahmad, *Encryption and Globalization*, 13 COLUM. SCI. & TECH L. REV. 416, 427 (2012) (noting that with public key/asymmetric encryption, the one-way function is a "calculation that is much easier to execute in one direction than it is to reverse.")

[151] *See generally The science of encryption: prime numbers and mod* n *arithmetic*, UC BERKELEY, https://math.berkeley.edu/~kpmann/encryption.pdf; Burt Kaliski, *The Mathematics of the RSA Public-Key Cryptosystem*, RSA LABORATORIES, https://www.nku.edu/~christensen/the%20mathematics%20of%20the%20RSA%20cryptosystem.pdf (last visited Oct. 11, 2020) (describing relevant mathematics concepts).

Figure 10 The Encryption Step is a One-Way Function (in asymmetric encryption)

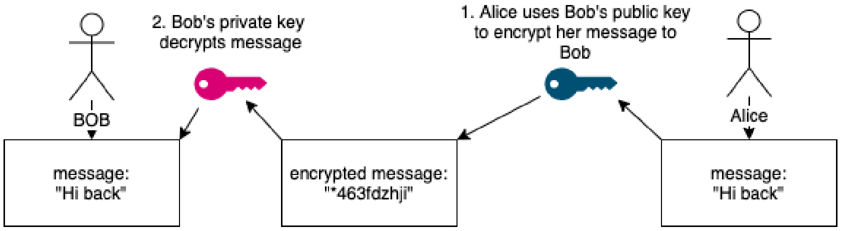


Another concept of asymmetric key encryption is that Bob is the only person in the world with access to Bob's private key (otherwise his private key would not be "private"). Another concept is that only Bob's corresponding public key can be used to decrypt the message. Therefore, Bob's public key and private key are considered complementary. Given that Alice was able to decrypt the message using Bob's public key, this means mathematically that Bob's private key was used to encrypt the message. This is strong evidence that Bob sent the message, and it will be difficult for Bob to argue that he did not send the message (Bob could try to argue that someone else used his computer to send the message, but this may be an uphill argument for Bob).

4. *Encrypting with a Public Key Provides Confidentiality but not Authentication of Sender*

To further understand public key encryption, consider a scenario below (FIG. 11) in which Alice wishes to send Bob an encrypted communication, using one of Bob's keys. Her only option is to use Bob's public key because Bob's private key is private to Bob. If Alice encrypts the message to Bob using Bob's public key, she can be assured that only Bob can decrypt the message because only Bob's private key can decrypt the message (only Bob has Bob's private key).

Figure 11: Asymmetric Encryption Requires Two Keys



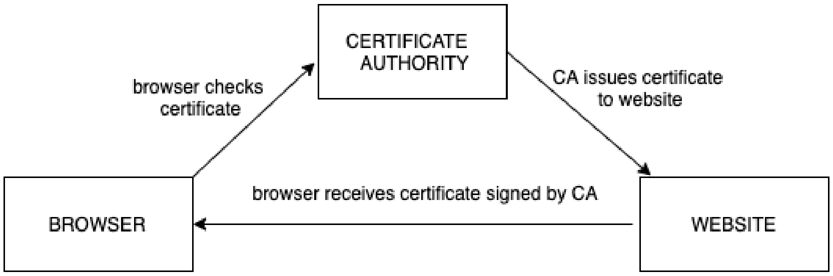
In the above FIG. 11, Alice can be fairly assured of the confidentiality of her message because only Bob can decrypt the message using Bob's private key. However, in the above scenario Bob cannot authenticate that Alice was the sender of the message because Alice used Bob's public key. Bob can presumably share his public key with the general public, so there is no way to know that Alice was the party using the key. Important lessons from FIGS. 10 and 11 are that (1) a private key is used to authenticate a sender and (2) encrypting with a public key merely provides confidentiality but not authentication of the sender. (Granted, using either key provides confidentiality.)

5. *Using Asymmetric Encryption for Website Certificates*

As noted previously, with the HTTPS protocol, the user's browser authenticates the legitimacy of the website/web server as an initial step (and the website certificate can typically be viewed by clicking a padlock icon in the browser bar). This authentication process relies on asymmetric encryption to assess whether the website is trustworthy. If the website is unable to authenticate itself to the browser with a valid certificate, then the browser will warn the user to proceed at her own risk rather than immediately connecting to the website.[152]

The below FIG. 12 is a conceptual diagram:

Figure 12: Authenticating a Website via a Trusted Third-Party Certificate


CERTIFICATE AUTHORITY
browser checks certificate
CA issues certificate to website
browser receives certificate signed by CA
BROWSER
WEBSITE

The browser receives a certificate showing the public key of the certificate authority (CA). The browser checks whether the certificate is valid essentially by querying the CA. As a conceptual example, assume the browser uses the CA's public key to encrypt the certificate number and sends the resulting ciphertext to the CA. Mathematically, the only party in the world that can decrypt the message is the CA that owns the private key. Conceptually, if the CA can decrypt the certificate number and send it back to the browser, this would demonstrate that

[152] One can view a website's certificate by clicking on the lock icon in the browser bar of most browsers.

the CA signed the certificate and that the website can thus be trusted as legitimate. This example conveys a general, conceptual understanding of how a certificate authority digitally signs a certificate with its private key for authentication; however, more specific technical details can be explored as to the exact mechanisms/digital signature standard used by particular certificate authorities (which can involve use of hash value matching).[153]

6. Using Asymmetric and Conventional / Symmetric Encryption Together

Asymmetric encryption requires greater processing power in comparison to conventional encryption.[154] This should be intuitive because asymmetric encryption requires use of two keys for the encryption/decryption process while conventional encryption requires use of one key that is shared by both the sender and receiver. Asymmetric encryption is therefore worthwhile for encrypting small quantities of data, such as authenticating a digital signature, and conventional encryption is preferred for encrypting and decrypting large quantities of data.

In some instances, a system may allow authentication with asymmetric encryption and also share a conventional key via asymmetric encryption.[155] This is what occurs with the very common HTTPS protocol, more specifically the TLS component thereof. When a computer, such as a laptop, connects to a remote web server via an HTTPS connection, the communication "session" between the browser and the web server is established by a three-way handshake: an initial step is for the browser to authenticate the web server is legitimate through a third-party trusted certificate authority (as discussed previously) using public key or asymmetric encryption. The TLS component of HTTPS then

[153] *See Security Tip (ST05-010): Understanding Web Certificates*, CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY (last revised Nov. 19, 2019), https://us-cert.cisa.gov/ncas/tips/ST05-010; Elaine B. Barker, *Digital Signature Standard (DSS)*, NIST (July 19, 2013), https://www.nist.gov/publications/digital-signature-standard-dss-2 (linking to a technically detailed discussion of digital signatures, including discussion of hashing in this context).

[154] *See The Difference Between Symmetric and Asymmetric Encryption*, SSLS.COM BLOG (Sept. 15, 2021), https://www.ssls.com/blog/the-difference-between-symmetric-and-asymmetric-encryption/ (describing symmetric encryption as faster than asymmetric).

[155] *Pretty Good Privacy*, WIKIPEDIA, https://en.wikipedia.org/wiki/Pretty_Good_Privacy#cite_note-3 (last updated Sept. 19, 2020, 9:40 AM) ("PGP can be used to send messages confidentially. For this, PGP uses hybrid cryptosystem by combining symmetric-key encryption and public-key encryption. The message is encrypted using a symmetric encryption algorithm, which requires a symmetric key generated by the sender. The symmetric key is used only once and is also called a session key. The message and its session key are sent to the receiver. The session key must be sent to the receiver[,] so they know how to decrypt the message, but to protect it during transmission it is encrypted with the receiver's public key. Only the private key belonging to the receiver can decrypt the session key[] and use it to symmetrically decrypt the message.").

provides a shared symmetric key for the browser and web server to encrypt the bulk of data they exchange during the session.[156]

Accordingly, a general principle to glean from the foregoing example is that when one computer communicates with a second computer, it can first authenticate the second computer using public key encryption. Because public key encryption employs two keys, its processing overhead is too high for bulk data exchange, so a symmetric key is preferable after authentication is done. Therefore, after authentication, the two computers can share a symmetric key for encrypted data exchange. Rather than sending the symmetric key in plain text over the internet, the symmetric key can be encrypted with public key encryption and sent to the other party. Thereafter, the two computers can use the shared conventional key to encrypt and decrypt data shared between them.

III. SOME HIGH-LEVEL DISCUSSION RELEVANT TO ENCRYPTION

A. SOME HIGH-LEVEL LEGAL DISCUSSION OF ENCRYPTION

1. *Some U.S. History of Surveillance*

Regarding the history of surveillance in the United States, government and private-party electronic eavesdropping of unencrypted data has been around for a long time. Regarding private eavesdropping, in 1864 a stockbroker was convicted of insider trading after he eavesdropped on corporate telegraph lines in order to sell insider information.[157] Regarding telephone lines, in the 1950s it was relatively easy to hire a private detective to listen in on phone lines to investigate adultery or the like.[158] In the mid-1960s, government wiretapping of phone lines was considered a necessary evil for matters of national security, but civil rights concerns had certainly arisen with regard to routine eavesdropping on phone lines for domestic law enforcement purposes.[159] This history is relevant because telegraph and telephone lines, much like the internet, were initially designed without much thought given to security and privacy. Similarly, with both the internet and these prior technologies, the public demand for security and privacy arose largely after the open infrastructure was in place. The de facto

[156] *See generally* Eric Rescorla, *The Transport Layer Security (TLS) Protocol Version 1.3*, INTERNET ENGINEERING TASK FORCE (Aug. 2018), https://tools.ietf.org/html/rfc8446 (providing a detailed discussion of TLS).

[157] April White, *A Brief History of Surveillance in America*, SMITHSONIAN MAGAZINE (Apr. 2018), https://www.smithsonianmag.com/history/brief-history-surveillance-america-180968399/.

[158] *Id.*

[159] *Id.*

implementation of security and privacy likely inspired the security and privacy by design provisions of the EU's GDPR.[160]

In recent years, the government has focused and expanded the use of warrantless, mass surveillance techniques on unencrypted data to target and arrest illegal immigrants.[161] A major concern is the increasingly expansive methods of mass surveillance Immigration and Customs Enforcement ("ICE") is permitted to use, without oversight, to arrest illegal immigrants.[162] Recently, ICE has used social media, specifically Facebook, as a way to gather information on an illegal immigrant where the information was ultimately used to secure an arrest.[163] Further, ICE is permitted to obtain information from local agencies across the country, such as local and state jails, to access arrestees' fingerprints and cross-reference the fingerprints against immigration records.[164] ICE also used state DMV databases to run face recognition searches on driver's license photos for comparison against the photos of undocumented immigrants.[165] The surveillance methods do not end here, but also have consisted of ICE scanning license plates in parking lots,[166] obtaining addresses from utility companies,[167] and partnering with a facial recognition company to compare photos of illegal immigrants to photos on social media.[168] This is just a small glimpse into the increasing danger people might face when it comes to their security/privacy

[160] *See* GDPR Art. 25; *see also supra* text accompanying note 50 (describing the GDPR's privacy by design provisions).

[161] *See* Taher Kameli and Chathan Vemuri, *Immigrant Surveillance – The DHS' Proposal to Expand Biometric Collection To Limit Immigration*, KAMELI & ASSOCIATES (Oct. 5, 2020), https://kameli.com/2020/10/05/immigrant-surveillance/ (describing that internet surveillance of immigrants has "metastasized considerably over the last two decades").

[162] *Id.*

[163] Max Rivlin-Nadler, *How Ice Uses Social Media To Surveil and Arrest Immigrants*, THE INTERCEPT (Dec. 22, 2019, 8:00 AM), https://theintercept.com/2019/12/22/ice-social-media-surveillance/.

[164] Alvaro M. Bedoya, *The Cruel New Era of Data Deportation*, SLATE (Sept. 22, 2020, 1:40 PM), https://slate.com/technology/2020/09/palantir-ice-deportation-immigrant-surveillance-big-data.html/.

[165] *Id.* *See also* George Joseph, *Where ICE Already Has Direct Lines to Law Enforcement Databases with Immigrant Data*, NPR (May 12, 2017, 1:44 PM), https://www.npr.org/sections/codeswitch/2017/05/12/479070535/where-ice-already-has-direct-lines-to-law-enforcement-databases-with-immigrant-d (describing ICE agents tapping into "local law enforcement and drivers' license databases").

[166] *See* Bedoya, *supra* note 163; Russell Brandom, *Exclusive: ICE is about to start tracking license plates across the US*, THE VERGE (Jan. 26, 2018, 8:04 AM), https://www.theverge.com/2018/1/26/16932350/ice-immigration-customs-license-plate-recognition-contract-vigilant-solutions.

[167] Bedoya, *supra* note 165.

[168] *See* Bedoya, *supra* note 165; Kim Lyons, *Ice just signed a contract with facial recognition company Clearview AI*, THE VERGE (Aug. 14, 2020, 3:19 PM), https://www.theverge.com/2020/8/14/21368930/clearview-ai-ice-contract-privacy-immigration.

being compromised and the expansive power and control the government may exert.[169]

2. *Some Legal History of Encryption*

For centuries, encryption has been considered a tool of war. This dates back as early as 405 B.C.E. when Lysander of Sparta sent encoded messages to military personnel.[170] In U.S. history, encryption likewise played a key role in both World War I and II, including the Allied war effort cracking German and Japanese cipher systems.[171]

In an effort to safeguard encryption technology for exclusive use by the U.S. military, Congress passed the Arms Export Control Act of 1976, 22 U.S.C. § 2278, which characterized encryption technology as a "munition," requiring a license from the U.S. Government prior to export to other countries in order to avoid criminal penalties.[172] In 1993, a federal grand jury in California opened a criminal investigation against Philip Zimmerman to investigate whether the worldwide distribution of his PGP (Pretty Good Privacy) encryption software violated the Act, with the government ultimately dropping the charge.[173] Ultimately, the U.S. government in 1999 communicated a concrete change in policy, allowing export of encryption products without restriction.[174] This was based in part on First Amendment challenges to the Act, as well as the U.S. government ultimately conceding that strong encryption was vital to the operation of the Internet by consumers and businesses to protect data.[175]

Regarding a First Amendment challenge to regulating the export of encryption, *Bernstein v. U.S. Department of State*[176] provides a helpful discussion of the history of the government's attempt to regulate encryption, initially through the Arms Export Control Act and its accompanying regulations, and then through the Export Administration Act of 1979 and its accompanying regulations.[177] In *Bernstein*, a mathematician sought a declaratory judgment that

[169] *Snowden Leaks: a Summary of the NSA Programs*, COGIPAS, https://www.cogipas.com/snowden-leaks-summary-of-nsa-programs/ (last updated July 11, 2018).

[170] Ronald J. Stay, *Cryptic Controversy: U.S. Government Restrictions on Cryptography Exports and the Plight of Philip Zimmermann*, 13 GA. ST. U. L. REV. 581, 582 (1997), available at https ://readingroom.law.gsu.edu/gsulr/vol13/iss2/14.

[171] *Id.*

[172] *See id.* at 586 (noting the enforcement of the statute by the Secretary of State via the International Traffic in Arms Regulations (ITAR) under 22 C.F.R. 120.1 (1995)).

[173] Stay, *supra* note 169.

[174] Swire & Ahmad, *supra* note 143.

[175] *Id.* at 425.

[176] 974 F. Supp. 1288 (N.D. Cal. 1997).

[177] As explained in *Bernstein*, On December 9, 1996, President Clinton, via Executive Order 13026, transferred jurisdiction over the export of nonmilitary encryption products to the Department of Commerce under the Export Administration Act of 1979, 50 U.S.C. §§ 2401–

his publication of encryption algorithms was First Amendment protected, thereby negating enforcement under the regulations of the Arms Export Control Act and the regulations of the Export Administration Act.[178] The court sided with the Plaintiff, holding that the encryption regulations were unconstitutional prior restraints in violation of the First Amendment.[179] The decision in *Bernstein* and related cases, along with other public debate, likely paved the way for the U.S. Government in 1999 to ultimately forego regulation of commercial encryption products.[180]

3. Encryption and EU-US Data Transfers

In 2020, the European Court of Justice invalidated the EU-U.S. Privacy Shield program, which was one mechanism allowing participating U.S. organizations to import data from the EU concerning EU data subjects.[181] The Privacy Shield program was essentially invalidated over concerns that EU residents' data did not have adequate privacy protections to protect from U.S. government surveillance in the wake of the Edward Snowden leaks concerning widespread U.S. government surveillance of civilians.[182] Snowden's leaks revealed the U.S. government's collection of vast amounts of civilian data, one example being the government's project prism[183] program which collected massive amounts of emails and search histories.[184] Interestingly, Snowden has reported that widespread use of encryption is likely the primary tool that could

2411 (1991) and the Export Administration Regulations, 15 C.F.R. §§ 730.1–.10 (1997). *Bernstein*, 974 F. Supp. at 1291.

[178] *Bernstein*, 974 F. Supp. at 1291.

[179] *Id.* at 1308.

[180] *See* Swire & Ahmad, *supra* note 143 (noting at page 416 that "[i]n 1999, . . . the administration shifted position to allow largely unrestricted export of encryption technologies. Encryption law and policy discussions largely faded from view.")

[181] Court of Justice of the European Union Press Release 91/20, Court of Justice of the European Union, the Council and the Parliament on Invalidating Decision 2016/1250 (July 16, 2020), https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf.

[182] Adam Satariano, *E.U. Court Strikes Down Trans-Atlantic Data Transfer Pact*, NEW YORK TIMES (July 17, 2020), https://www.nytimes.com/2020/07/16/business/eu-data-transfer-pact-rejected.html.

[183] The Prism program was authorized under Section 702 of the Foreign Intelligence Services Act (FISA). *See* United States v. Mohammad, 339 F. Supp. 3d 724, 744 (N.D. Ohio 2018) (discussing Section 702 data collection under Prism, generally).

[184] Another example is the NSA's Bullrun program designed to decrypt HTTPS and other communications. *Bullrun (decryption program)*, WIKIPEDIA, https://en.wikipedia.org/wiki/Bullrun_(decryption_program) (last updated Jan. 15 2021, 4:36 PM).

be used to curb much of this massive governmental and private party dataveillance.[185]

In August 2020, in the wake of the Privacy Shield invalidation, the German Data Protection Authority (DPA) of the German federal state of Baden-Württemberg issued guidance on international data transfers for ensuring safe transfer of EU residents' data into the U.S. The German DPA guidelines provided a number of suggested safeguards to protect EU data flowing into the U.S., including encryption for which "only the data exporter has the key" and which "cannot be broken by U.S. [intelligence] services."[186] This suggested safeguard reflects an interesting shift because U.S. government policy may now require support of un-surveillable communications, if it wishes to allow U.S. companies to engage in substantial online commerce with Europe. In a sense, the strong European desire for personal privacy is continuing to influence U.S. policy.[187]

4. *Privacy Interests Promoted by Strong Encryption versus Law Enforcement's Desire for a "Front Door"*

a. Arguments for a Front Door

For decades, a debate has been ongoing over whether federal law enforcement should be provided a "back door" (today called a front door) to encryption products provided by tech companies, such as Apple's iPhone.[188]

[185] *See* Lauren C. Williams, *Edward Snowden Says Encryption Is The Only Way To Counter Mass Surveillance,* THINK PROGRESS (Mar. 10, 2014, 6:57 PM), https://archive.thinkprogress.org/edward-snowden-says-encryption-is-the-only-way-to-counter-mass-surveillance-ee450433dca8/ (in which Snowden notes that "mass surveillance by businesses and governments is the biggest threat to national and individual security. During the hour-long discussion, Snowden focused on how mass surveillance and indiscriminate data collection by both erodes individual privacy and national security. Given the intelligence community's unwillingness to reform their tactics, the only way to combat surveillance is for consumers and the tech world to expand the use of encryption and other cybersecurity tools." Similarly, at a presentation hosted by the American Civil Liberties Union, Edward Snowden described encryption as "the defense against the dark arts for the digital realm").

[186] *See German DPA Issues Guidance on Data Transfers Following Schrems II*, HUNTON ANDREWS KURTH (Sept. 2, 2020), https://www.huntonprivacyblog.com/2020/09/02/german-dpa-issues-guidance-on-data-transfers-following-schrems-ii/ (describing other technical safeguards: anonymization or pseudonymization, where "only the data exporter can re-identify the data.").

[187] As an example, the California Consumer Privacy Protection Act is widely regarded as inspired by the European General Data Privacy Regulation (GDPR), considering its post GDPR enactment with various similarities in privacy protections (e.g., requiring consumer consent to personal data collection, right to deletion, etc.).

[188] Kif Leswing, *Apple's Fight with Trump and the Justice Department is about more than two iPhones*, CNBC (Jan. 16, 2020, 1:57 PM), https://www.cnbc.com/2020/01/16/apple-fbi-backdoor-battle-is-about-more-than-two-iphones.html.

Currently, a law enforcement front door exists with respect to wiretap orders under CALEA (Communications Assistance for Law Enforcement Act), which applies to telecommunications carriers, including providers of telephone service, Broadband Internet Service, and providers of VoIP.[189] Such providers are statutorily required to modify and design their technologies with sufficient capabilities to assist such orders. Therefore, law enforcement's current push is for essentially a larger front door, to encompass other technologies not reached by CALEA, such as smartphone devices or cloud providers (e.g., Dropbox or encrypted WhatsApp messages).[190]

In an October 2019 speech, FBI Director Christopher Wray raised concerns over warrant-proof, end-to-end encryption and disclosed that it wants a "front door" to lawfully access communications with a warrant from a neutral judge upon meeting Fourth Amendment requirements.[191] Similarly, in 2016 FBI Director James Comey publicly pushed for a backdoor to Apple's iPhone while seeking a court order under the All Writs Act, 28 U.S.C. § 1651, to compel Apple to unlock an encrypted iPhone 5C used by a San Bernardino attacker in 2015.[192]

The desired front door would require tech companies to facilitate lawful access upon law enforcement obtaining judicial approval. Perhaps the best-selling point for providing a front door is the important need to prevent or prosecute cases of child sexual exploitation or terrorism. It's conceivable that a free society could potentially tolerate allowing some unmonitored drug trafficking communications, essentially weighing society's need for privacy as greater than the need to uncover certain low-level crimes. However, it would seem that a civilized society should not tolerate widespread undiscoverable distribution of videos and other imagery of infants and toddlers subjected to sex acts.[193] FBI Director Christopher Wray noted in his October 2019 speech that

[189] *See Communications Assistance for Law Enforcement Act*, FEDERAL COMMUNICATIONS COMMISSION, https://www.fcc.gov/public-safety-and-homeland-security/policy-and-licensing-division/general/communications-assistance (last updated Oct. 20, 2020) (noting "[c]ommunications services and facilities utilizing Circuit Mode equipment, packet mode equipment, facilities-based broadband Internet access providers and providers of interconnected Voice over Internet Protocol (VoIP) service are all subject to CALEA.").

[190] *See* Opderbeck, *supra* note 80, at 1671 (noting many cloud providers, such as Dropbox, do not provide internet access, which brings them outside the scope of CALEA.).

[191] Christopher Wray*, Finding a Way Forward on Lawful Access Bringing Child Predators Out of the Shadows*, FBI (Oct. 4, 2019), https://www.fbi.gov/news/speeches/finding-a-way-forward-on-lawful-access.

[192] *See* Matthias Schulze, *Clipper Meets Apple v. FBI-A Comparison of the Cryptography Discourses from 1993 and 2016*, 5 COGITATIO 1 (2017), *available at* https://www.cogitatiopress.com/mediaandcommunication/article/view/805; Opderbeck, *supra* note 80, at 1671 (discussing the government's usage of the All Writs Act to fill in the gap is not covered by CALEA).

[193] A fairly horrific example is described in Olivia Solon, *Child sexual abuse images and online exploitation surge during the pandemic*, NBC NEWS (April 23, 2020, 3:01 PM), https:

Facebook has provided more than 90% of the referrals received by the National Center for Missing & Exploited Children, which "now receives more than 18 million referrals" per year.[194] Director Wray expressed concern that if Facebook and other tech companies end up providing warrant-proof end-to-end encryption, they will willfully blind themselves to all content and eliminate the possibility of lawfully accessing particularly egregious criminal content.[195] He notes that Facebook moving in this direction would transform Facebook "from the main provider of child exploitation tips to a dream come true for predators and child pornographers."[196]

One source indicates that policymakers generally favor strong encryption with exceptional, warrant-based access, while the tech community generally rejects this view, essentially arguing that strong encryption technologies should have no back door vulnerability built-in.[197] Certainly, the Internet is a powerful tool that can be used for good or evil purposes. Therefore, this debate over an encryption front door will likely continue indefinitely (perhaps gaining the greatest support in the wake of any significant terror attack), with legislators butting heads with the tech community.

b. Arguments Against a Front Door

Various arguments against a law enforcement front (or back) door have been advanced, perhaps based on a growing distrust in the society of governments (and businesses) having unfettered access to personal data.[198] Some key arguments against a front door include: (1) concern over law enforcement creep where the front door is established for very serious crimes or terrorism but eventually creeps into surveillance of lower level crimes, thereby creating a police

//www.nbcnews.com/tech/tech-news/child-sexual-abuse-images-online-exploitation-surge-during-pandemic-n1190506 (explaining that a zoom bomber delivered explicit video of a sexual assault on an infant to attendees of a virtual conference on climate change).

[194] Wray*, supra* note 190.

[195] Wray, *supra* note 190.

[196] David Shortell, *FBI director claims encryption plan would make Facebook a 'dream come true' for child pornographers*, CNN POLITICS, https://www.cnn.com/2019/10/04/politics/fbi-facebook-child-encryption/index.html (Oct. 4, 2019, 3:22 PM).

[197] *See* Schulze, *supra* note 191, at 59 (noting "[p]olicymakers in general favor strong encryption with exceptional, warrant-based access while the tech community replies that the mathematics either support secure encryption without government backdoors or exceptional access with significantly less security.").

[198] *See* Ryan Budish, Herbert Burkert, & Urs Gasser, *Encryption Policy and Its International Impacts: A Framework for Understanding Extraterritorial Ripple Effects*, HOOVER INSTITUTION (2018), https://dash.harvard.edu/bitstream/handle/1/36291726/budish_webreadypdf%202.pdf?sequence=1 ("Apple's decision to offer end-to-end encrypted messaging and full device encryption, both enabled by default, could be seen as a direct response to declining consumer trust and concerns over NSA surveillance.").

state;[199] (2) concern over any safeguards being ignored by law enforcement and a lack of law enforcement transparency with the public in terms of currently used forensics tools and their effectiveness;[200] (3) concern over a backdoor being leaked and exploited by bad actors;[201] (4) a related concern that it's technologically difficult to design breakable encryption products that are still considered to provide strong protection; (5) concern over U.S. consumers purchasing foreign encryption products, thereby reducing sales of potentially inferior U.S. products, thereby creating anticompetitive effects for U.S. products;[202] (6) concern that overseas sales of U.S. IT hardware or software may decline if a perception exists that they provide weak security on account of U.S. mandated vulnerability;[203] (6) recognition that law enforcement has essentially lost the cryptowars of the 1990s; (7) concern that backdoors are often mandated by authoritarian regimes to surveil a population in violation of their fundamental human rights;[204] (8) concern that on balance, society (both businesses and individuals) is safer when provided with the strongest encryption products;[205] (9) an argument that law enforcement can use other investigative techniques to discover high priority terrorist activity and child molestation without infringing privacy of encrypted communications.[206]

[199] *See* Schulze, *supra* note 191, at 58-59 ("[A]gencies might dig up cases to mandate companies to build in backdoors for more trivial reasons than fighting terrorism, a phenomenon called function creep.").

[200] *See* Ellen Nakashima, *FBI and NSA violated surveillance law or privacy rules, a federal judge found,* MSN (Sept. 4, 2020), https://www.msn.com/en-us/news/us/fbi-and-nsa-violated-surveillance-law-or-privacy-rules-a-federal-judge-found/ar-BB18IVqI; *see also Redacted*, 402 F. Supp. 3d 45 (Foreign Intel. Surv. Ct. 2018) (finding "the FBI's querying and minimization procedures, as implemented, to be inconsistent with statutory minimization requirements and the requirements of the Fourth Amendment").

[201] *See* Schulze, *supra* note 191, at 57 (noting "[b]usiness actors are more afraid of the potential future effects of the government regulating encryption, which might result in the widespread use of inferior technology.").

[202] *See* Ryan Budish, Herbert Burkert, & Urs Gasser, *Encryption Policy and Its International Impacts: A Framework for Understanding Extraterritorial Ripple Effects*, HOOVER INSTITUTION (2018), https://dash.harvard.edu/bitstream/handle/1/36291726/budish_webreadypdf%202.pdf?sequence=1 at page 9: For example, analysts predicted that the economic impact on U.S. companies attributable to the Snowden leaks was in the range of $35 billion to $180 billion in lost revenue.

[203] *Id.*

[204] *See* Schulze, *supra* note 191, at 57 (noting the argument that "control of encryption technology is a norm of authoritarian regimes and police states and therefore inappropriate in democracies.")

[205] *See* Schulze, *supra* note 191, at 59 (statement of Form NSA Director Michael Hayden) ("America is simply more secure with unbreakable end-to-end encryption.").

[206] *See* Christopher Soghoian, Book Note, *Caught in the Cloud: Privacy, Encryption, and Government Back Doors In The Web 2.0 Era*, 8 J. TELECOMM. & HIGH TECH L. 360, 399 (2010) ("[I]f a suspect is important enough, let the police dedicate the significant manpower to break into her home in order to install bugs. Given the finite limit to the financial and human

One example that exemplifies some of the arguments against a front (or back door) for the government or law enforcement is shown by the recent SolarWinds hack. A few sources suggest that there may have been a weakness on account of government backdoors that took part in allowing the success of the attack.[207] The SolarWinds hack raises a question of whether a mandated backdoor does more harm than good. For example, one would need to balance the backdoor's effectiveness in detection/prevention of some level of criminal activity versus the harm caused by a massive data breach potentially enabled by the backdoor.

Regarding a lack of public transparency over existing tools, law enforcement is not informing the public that law enforcement across the country is regularly able to break into phones; even small police departments that have the capability themselves or that simply send the phones to a lab.[208] Perhaps it's understandable for law enforcement not to publicly disclose the investigative tools in its arsenal and its success rate. However, law enforcement's sales pitch to the public makes it sound like the police are never able to break into encrypted phones, and this is not true. Jennifer Granick, a cybersecurity lawyer at the American Civil Liberties Union explains "Law enforcement at all levels has access to technology that it can use to unlock phones. That is not what we've been told."[209] The truth it would seem is that sometimes they can break in and sometimes they can't, with one Manhattan prosecutor explaining "we may unlock it in a week, we may not unlock it for two years, or we may never unlock it."[210]

If law enforcement continues to lose the cryptowars with privacy interests outweighing crime detection, then perhaps a fundamental privacy principle is taking shape equivalent to the English and U.S. criminal concept of Blackstone's

resources available to law enforcement agencies, such a change in the balance of power, by raising the effective cost of such surveillance, would force investigators to prioritize their targets, and shy away from fishing expeditions.").

[207] Shawna Chen, *Dozens of Treasury email accounts breached in SolarWinds hack*, AXIOS (Dec. 22, 2020), https://www.axios.com/solarwinds-hack-treasury-email-accounts-breached-e6a24240-2795-4c09-9056-b53f20e47f37.html (statement of Ron Wyden, Treasury Finance Committee Ranking Member) ("Finally, after years of government officials advocating for encryption backdoors, and ignoring warnings from cybersecurity experts who said that that encryption keys become irresistible targets for hackers, the USG has now suffered a breach that seems to involve skilled hackers stealing encryption keys from USG servers."); *see also*, Glyn Moody, *The widening SolarWinds debacle shows why the reckless idea of backdooring encryption must be dropped forever*, PRIVACY NEWS ONLINE (Dec. 24, 2020), https://www.privateinternetaccess.com/blog/the-widening-solarwinds-debacle-shows-why-the-reckless-idea-of-backdooring-encryption-must-be-dropped-forever/ ("Key to the intrusion was the insertion of malicious code into the Orion network monitoring software from SolarWinds – a backdoor in software that was very widely used and trusted.").

[208] Jack Nicas, *The Police Can Probably Break Into Your Phone*, N.Y. TIMES (Oct. 21, 2020), https://www.nytimes.com/2020/10/21/technology/iphone-encryption-police.html.

[209] *Id.*

[210] *Id.*

ratio: "it is better that ten guilty persons escape than that one innocent suffer." Blackstone's ratio is credited with the development of the beyond reasonable doubt standard in criminal law and likely influenced the development of other constitutional protections within the United States.[211] Roughly applying this same principle to privacy, one could argue that it is preferable to allow a certain number of criminal communications that cannot be surveyed rather than undermine the privacy interests of millions of law-abiding civilians. If the United States supports un-surveillable communications, this could potentially overrule law enforcement's long-standing request to have a readily available encryption backdoor, which companies such as Apple have resisted.[212] While this extremely pro-privacy view is appealing, it could pose some dangers as noted below. If a strong terrorism concern exists regarding EU-US transfers, however, then law enforcement could potentially order a tech provider to hand over the key via the court system. That is, assuming the tech providers maintains the key or the ability to decrypt.

If privacy interests favor warrant proof encryption technologies, this raises a fundamental question of the value of privacy in a free society balanced against the need or desire for law enforcement to investigate criminal or terrorist activity. Likewise, this raises a larger related question, beyond the scope of this article, of what it means to live in a free society and how privacy rights fit within that free society. Certainly, if the government or private organizations (or some combination of the two) can easily review everyone's communications from both a routing standpoint (i.e., the "from" and "to" fields of a data packet) and a content standpoint (i.e., the payload of a data packet), then this unfettered access to emails, text messages, voicemails, search history, and current and past location may exemplify a society that is not free. Strong encryption can thus be a helpful tool to increase freedom through privacy of communications, hiding not only the routing info and identity of the communicating parties (e.g., "from" and "to" info) but also the contents of the messages themselves (e.g., payload of data packets). U.S. government policy, both legislatively and judicially, may thus continue to shift toward stronger protection of personal data.[213]

Interestingly, the New York Civil Liberties Union issued a report in 2020 discussing how attorneys can increase their use of encryption to protect client

[211] *See* Daniel Epps, *One Last Word on the Blackstone Principle*, 102 VA. L. REV. ONLINE 34 (2016) (discussing and critiquing Blackstone's ratio).

[212] Lauren Feiner, *Republican senators introduce bill that tech advocates warned would weaken privacy*, CNBC (June 24, 2020, 9:47 AM), https://www.cnbc.com/2020/06/24/gop-senators-introduce-bill-that-would-create-a-backdoor-for-encryption.html.

[213] Riley v. California, 134 S. Ct. 2473, 2490 (2014) (noting that "many of the more than 90% of American adults who own a cell phone keep on their person a digital record of nearly every aspect of their lives-from the mundane to the intimate." In my view, this evidences the judicial awareness of highly personal data maintained by many people in the United States and the need to safeguard this same data.).

information from being sucked up by mass government surveillance particularly in criminal defense representation.[214] The report notes that "Privileged and confidential attorney-client communication is illegally being recorded as a matter of course by contractors working for the government, like the telephone system contractor Securus."[215] This report supports that there is a real concern in both the EU and the U.S. over warrantless, and technologically easy, government surveillance of communications and the need to protect communications from same. Interestingly, the New York Civil Liberties Union report encourages the use of open-source encryption tools that would seem to remove the possibility of a private party providing a back door to its encryption products.[216]

c. Three Hypothetical Privacy Scenarios to Facilitate Discussion

Three hypothetical scenarios are provided to stimulate discussion with regard to privacy of online communications and stored data: (1) a no privacy extreme; (2) a pro privacy extreme; and (3) a balanced privacy scheme (shown in Tables 1-3 below).

1. Hypothetical **"No Privacy"** Extreme (i.e., a transparent internet)				
Data in Transit or Stored Data?	Information type	Encryption state	Effect	Notes
Data in Transit	Routing information* (header to/from info in data packets)	Unencrypted	Sender and receiver easily observed (by private parties or government)	Similar to posting to/from messages on a public bulletin board

[214] *See* Jonathan Stribling-Uss, *Legal Cybersecurity in the Digital Age*, NEW YORK CIVIL LIBERTIES UNION at *3 (Sept. 29, 2020), https://www.nyclu.org/sites/default/files/field _ documents/20200924_nyclu_legalcybersecurity_final_2.pdf (the report outlining "concrete, accessible steps that legal organizations, especially criminal defense lawyers, can do right now to ensure their attorney client communications are not sucked up into a government surveillance database or stolen by hackers.").

[215] *Id.* at 14-15.

[216] *Id.* at 19 ("[A]ttorneys should seek open-source products whose source code can be publicly accessed and vetted –to ensure there are no secret, government-prompted flaws that risk revealing client information.").

	Content (i.e., data payload of packets)	Unencrypted	All contents easily observed	
Stored Data	Data stored in the cloud	Unencrypted	Data can be easily viewed by the cloud provider or anyone else	This complete lack of confidentiality is generally unacceptable.

2. Hypothetical **“Pro Privacy”** Extreme				
Data in Transit or Stored Data?	Information type	Encryption state	Effect	Notes
Data in Transit	Routing information* (header to/from info in data packets)	Encrypted. Tech platform has no key and maintains no records	Sender and receiver not observable by eavesdroppers	Imagine all internet users using a VPN service that encrypts routing info and payload of packets and maintains no records of activity
	Content (i.e., data payload of packets)	Encrypted. Tech platform has no key and maintains no records	Contents not observable by eavesdroppers	
Stored Data	Data stored in the cloud	Encrypted. no decryption key held by the provider	Data can’t be viewed by the cloud provider, government, or other parties	Analogous to leasing a safe at a bank that maintains no key (only the tenant has the key)

3. Hypothetical **Balanced Privacy** Scenario

Data in Transit or Stored Data?	Information type	Encryption state	Effect	Notes
Data in Transit	Routing information* (header to/from info in data packets)	Encrypted, but tech platform has a key	Sender and receiver not observable by eavesdroppers unless decryption key used	An encryption back door is problematic (inherently vulnerable). Government and private actors tempted to use it unethically?
	Content (i.e., data payload of packets)	Encrypted, but tech platform has a key	Contents not observable by eavesdroppers unless decryption key used	
Stored Data	Data stored in the cloud	Encrypted, but tech platform has a key	Data not observable unless decryption key used	Consider a bank and its customer each having a key to a safety deposit box. In theory, the key is accessed through legal process only where serious crimes are suspected (access possible in civil litigation too?).

Privacy advocates might encourage completely private communications (Table 2) that have no possibility of ever being viewed by government or private parties. A completely private communication might involve no retrievable record of who contacted whom as well as no access to the contents of the communication. This would constitute the ultimate in privacy where for example an email provider or other tech platform removes its ability to access routing information or content. This way, even if the tech platform is ordered to provide such information, it will not have that capability. However, such ultimate privacy may arguably have the potential to cause too much harm.

While 100% private communications (Table 2) would advance privacy, some might argue for a back door (Table 3), asserting that the internet, as a powerful tool for large-scale good or evil activity, requires regulation. Perhaps a balanced view (Table 3) is that society would benefit from development of encrypted online communication platforms that are not easily susceptible to eavesdropping by government or private parties. Ideally, excellent privacy could be provided where civilians could use a platform that provides adequate assurance that neither the platform nor the government will see their communications without judicial process and that the platform will only hand over the decryption key upon receipt of a judicial order to do so. In balancing privacy interests, courts should continue to take into account whether the suspected crime is serious enough to warrant the privacy invasion. Perhaps judicially ordered decryption would be most appropriate where the suspected criminal activity would support a felony rather than misdemeanor charge to reduce the potential for governmental overreach with respect to the bulk of society's private communications and stored data?

Another concept to explore another day is whether civilians might be notified when their data has been examined at the conclusion of an investigation where no charge has been filed. This level of transparency might reduce unnecessary warrantless snooping of civilian data. Developing some form of transparency, perhaps through independent third-party auditing, might ease concerns over the potential for unlawful eavesdropping by federal agents.[217]

As discussed, many policy makers seem to favor legislative reform that would support development of a front door with suitable safeguards, while technologists seem overwhelmingly opposed to any legislative front door requirement. This article does not take a strong position on the issue, but offers a final thought that while a balanced approach sounds ideal, it might not be feasible if mandated backdoors end up doing more harm than good. Therefore, privacy interests may continue to outweigh law enforcement's desire for a front door.

IV. CONCLUSION

The foregoing should provide attorneys with some helpful background on encryption concepts from a technical and legal standpoint, which should be helpful in their efforts to provide effective leadership at the intersection of technology and law.

[217] For an example of alleged government abuse, see Trevor Aaronson, *A Declassified Court Ruling Shows How the FBI Abused NSA Mass Surveillance Data*, THE INTERCEPT (Oct. 10, 2019, 6:00 AM), https://theintercept.com/2019/10/10/fbi-nsa-mass-surveillance-abuse/.