Tow Center for Digital Journalism
A Tow/Knight Report

DIGITAL SECURITY AND SOURCE PROTECTION FOR JOURNALISTS

SUSAN MCGREGOR

Columbia Journalism School

Funded by the Tow Foundation
and the John S. and James L. Knight Foundation

Acknowledgements

This paper is the product of countless conversations, encounters, recommendations and insights provided to me over the past year by members of the digital security and journalism communities.

Though it would be impossible to name everyone whose insight and expertise has informed this work, my deepest thanks go to Franzi Roesner, Carey Shenkman, Nabiha Syed, Seamus Tuohy, Chris Walker and Carol Waters for providing invaluable feedback and direction on early versions of this paper; your expertise and insight was invaluable. Also to Oktavia Jonsdottir, Scott Klein, Quinn Norton, Jeremy Singer-Vine and Jennifer Valentino-DeVries, who shared their ideas and work processes with me. And a very special thanks to Peter G. Neumann, who offered me the long view on digital security and directed me to essential resources and experts on the subject.

This project would also not have been possible without the financial and intellectual support of my colleagues at Columbia Journalism School, and the Tow Center for Digital Journalism–particularly Emily Bell and Taylor Owen. Likewise, I am very grateful to both OpenITP and TacticalTech–organizations whose generosity in hosting me in the fall of 2013 made possible so many enlightening interactions with the digital security community. Without the openness of these organizations and their associates, this work would not have been possible.

And finally, many grateful thanks to my talented, enthusiastic, and insightful illustrator and collaborator, Matteo Farinella, for truly making the invisible, visible.

This volume produced with the assistance of the tufte-LaTeX class.
First printing, June 2014

Contents

This paper is dedicated to all members of the digital security and journalism communities, whose shared commitment to transparency, skepticism towards power, and dedication to informing and empowering individuals everywhere has been an inspiration. I hope that this will be only the first of many collaborations between our equally scrappy, resilient, and adaptive professions.

Preamble

In August of 2011, the United Kingdom experienced a wave of riots that swept across the country. The first of these took place in the Tottenham neighborhood of London, where local resident Mark Duggan had recently been shot and killed by police. Within hours, riots had erupted in cities around the United Kingdom, from Manchester to Bristol, in neighborhoods and communities with no obvious connection to the events that took place in Tottenham. Violence, vandalism, and understandable panic gripped the nation. In the midst of these events, the *Guardian*–led by Simon Rogers, founder and current editor of the Datablog–engaged in what might best be described as a real-time investigative-reporting effort to cover not only the where, when, and what of these riots, but also their why and how[1]. Using social media reports, as well as court records, extensive data analysis, and on-the-ground reporting, the *Guardian* eventually built the story of how these riots occurred and spread: largely through coordination efforts facilitated by digital communications. Though the *Guardian*'s analysis would eventually reveal–counter to the claims of Prime Minister David Cameron–that Twitter and Facebook had played no role in spreading the mayhem,[2] the rioters had made extensive use of BlackBerry Instant Messenger service to coordinate their activities. While the *Guardian*'s analysis was sufficient to derail Cameron's call for a "red button" for social media, RIM–the company that then ran BlackBerry–chose to cooperate with UK police, turning over the messages and handles related to the events.[3] The

[1] Datablog + UK Riots 2011. *The Guardian*, 2011-12

[2] Reading the Riots. *The Guardian*, 12/4/11

[3] London riots: BlackBerry to help police by Josh Halliday. *The Guardian*, 8/8/11

next month in New York, the Occupy Wall Street movement would lead to its own set of arrests, followed by the NYPD's courtroom pursuit of Malcolm Harris's Twitter "metadata"–especially information about the location of his phone when he posted certain tweets.[4] Even when Twitter attempted to *resist* cooperating with police,[5] they were eventually forced to turn over Harris' data anyway.[6]

These events threw into sharp relief the realities of privacy in the realm of digital communications: Whether messages were privately sent or publicly shared, users of services like BlackBerry Instant Messenger and Twitter had few enforceable rights around the information they communicated through these services, or even around information that the services had *about* them. Somehow, the evolution of digital communication systems had given rise to a strange class of information known as "metadata": the data *about* data that can seemingly reveal almost everything about anyone, and yet, simultaneously, belongs to no one at all.

[4] Twitter Must Hand Over Protester Malcolm Harris' Tweets by Jennifer Peltz. *The Huffington Post*, 7/2/12

[5] Twitter fights back against subpoena by Nilay Patel. *The Verge*, 5/8/12

[6] Twitter Turns Over Protester Posts by Tiffany Kary. *Bloomberg*, 9/14/12

Digital Security for Journalists: A 21st Century Imperative

In the spring of 2013, *Guardian* reporter Glenn Greenwald received a set of classified documents from a former NSA employee who would later be revealed as Edward Snowden. Among the leaked documents eventually published by both the *Guardian* and the *Washington Post* were some revealing that the United States' National Security Agency had for some time been performing bulk collection of digital communications metadata records, allegedly from corporations ranging from U.S. telecom companies to digital service providers like Google and Yahoo. [7] Though met with public outrage, the response of U.S. lawmakers to these revelations was decidedly measured: Not only was the program in question legal, but these collection practices had been taking place for some time. Nevertheless, continued reporting by the *Guardian* and the *Washington Post*–for which they would both eventually win a Pulitzer Prize–indicated a heretofore unconfirmed fact: that the digital communications systems that many Americans believed to be importantly private were, in fact, anything but.

The shocking nature of the Snowden revelations catapulted Greenwald to the center of the ongoing debate about the future of journalism. In October of 2013, Bill Keller, former managing editor of the *New York Times*, invited Greenwald to debate their views on the essential principles

[7] NSA paid millions to cover Prism compliance, *The Guardian*, 8/22/13

"As far as I know, this is the exact three-month renewal of what has been the case for the past seven years," said Feinstein.

of journalism in the 21st century in his *Times* newspaper column.[8]

Yet while Keller and Greenwald's exchange in that column did highlight their philosophical differences, it glossed some of the practical ones that were arguably no less significant to the story: Snowden took his documents to Laura Poitras and Glenn Greenwald in part because they could meet his communication-security requirements.

In the past 15 years, digital publishing and communications have changed the landscape–and even the nature–of journalism in innumerable ways. Old business models have collapsed, and are yet to be reasonably replaced. Private individuals and citizen journalists have access to the same platforms for publication and can cultivate the same profile as reporters at major news organizations. The power of the crowd can be used both to document and condemn. [9] And yet every corner of our industry–from fashion to finance, the national desk to national security–is still driven by a single, essential imperative: Get the story.

There are no stories without sources. Unless researchers, executives, parents, politicians, religious figures, heads of state, whistleblowers, and widowers–unless *people*–are willing to share information with journalists, our profession cannot function. Whether what they share with us is a trove of secret documents, the location of a meeting, or the story of a loved one lost, without them journalism as we know it ceases to exist. And yet the missing acknowledgment in Keller and Greenwald's debate was exactly this: that a perhaps fundamental difference in their journalism was a question of neither form nor philosophy, but of capacity. Greenwald and his colleagues were able to offer Snowden the digital protections he demanded. How many of today's practicing journalists, independent or institutional, can effectively do the same?

That certain professional practices are essential not just to the integrity but the viability of the journalistic enter-

[8] Is Glenn Greenwald the Future of News?, *The New York Times*, 10/27/13

[9] Social media and the search for the Boston bombing suspects. *CBS News*, 4/24/13

prise is already codified into our professional practice. Libel training and editorial review help protect journalists and their institutions from debilitating lawsuits. Reputable news organizations have articulated codes of conduct designed to sharply limit the personal benefit reporters may derive from their professional activities; many also have explicit conflict of interest surveys that reporters must file on a regular basis.[10] And yet in many newsrooms, the consideration given to the systematic protection of our most valuable assets–our sources–is uneven at best.

There can be little dispute at this point that journalism, even within the United States, is under legal and technical attack.[11] The year 2013 saw virtually unprecedented criminal charges leveled against both journalists and their sources. In some cases, members of the press have been forced to risk jail time to defend their sources; in others, they never had the chance.[12] And major news organizations[13] have acknowledged repeated hacking attempts on their systems, at least some of which are known to be direct efforts to uncover sources. Major communications companies have also acknowledged that a significant proportion of digital hacking targets are journalists.

Whatever the dollar cost of a lawsuit or a system recovery, the detriment that these events pose to our industry is incalculable. At the same time that Snowden's conscious choice to share his information with recognized journalists may inspire confidence in the continued importance of professional journalism, the difficulties he experienced in doing so securely[14] point to a significant deficiency in our existing practices. Moreover, his very revelations only confirm how thin is the veil that protects our digital communications from the eyes of others, whether they be governments', lawyers', service providers', or hackers'. As this understanding rightly permeates the public consciousness, the chilling effects will be immeasurable.

In order to maintain the confidence of–and therefore the

[10] NPR Ethics Handbook: Independence. *NPR*, Retrieved: 6/5/14

[11] AP President accuses DOJ of source intimidation, Jennifer C. Kerr. *AP*, 6/19/13

[12] Will Eric Holder Back Off?, Emily Bazelon. *Slate*, 6/2/14; DOJ scrutiny of James Rosen draws fire, Ann E. Marimow. *Washington Post*, 5/20/13

[13] e.g. The New York Times, Washington Post, Bloomberg, Wall Street Journal

[14] No Place to Hide – review, Philippe Sands. *The Guardian*, 5/23/14

access to–our sources, it is imperative that the journalistic profession as a whole develops a coherent set of professional practices around their protection. While judicial decisions and statutes in 49 states and the District of Columbia provide some form of reportorial "privilege,"[15] the legal and technical realities of digital communications systems today are such that many journalists will never have the opportunity to invoke it.

For robust journalistic security practices to be effective, they must both offer the real protections that sources deserve and be reasonable enough to integrate into the process of newsgathering and publication. To achieve these ends, any approach must be grounded in a fundamental understanding of the technical and legal frameworks in which our digital communications exist, and how their sometimes strange intersections influence the way that journalists must operate. The goal of this paper is to provide a coherent and salient introduction to these frameworks, as a foundation for developing supportable security practices for the journalism industry.

The genesis of this research stems directly from recent events: the *Associated Press* phone records scandal and the Snowden revelations that took place in the spring of 2013. Though I came to this topic well-versed in the basics of digital communication technologies, my collaborative development of a mobile application for secure, anonymous, authenticated communication had made me acutely aware that creating better tools for secure digital communications was only a part of the problem, and I was happy to leave the job of offering practical digital security advice to those with more experience than I. Yet as I reviewed existing guides and recommendations, I found that few of these resources were comprehensive in their discussions of the "when" and "why" of digital security. This is with good reason. There is no such thing as generic "security", and even when contextualized, its practices must effectively navigate any number of

[15] Journalists' Privilege, Kathleen Ann Ruane. *CRS*, 1/19/11

legal and technical pitfalls.

As I began this work, I spoke anecdotally with journalism colleagues who employed secure digital communication technologies in their work. In the process, the first outlines of a pattern began to emerge. Those who understood and applied digital security practices to their reporting, even occasionally, were either themselves covering sensitive topic areas like the NSA–and therefore came to these understandings of professional necessity–or, like me, they had a sufficiently technical background to parse these topics for themselves. This paper strives to provide an accessible level of technical and legal understanding for the broader journalism community, so that as an industry we can begin to have an informed conversation about how the realities of today's digital communications systems should be appropriately addressed within our work.

The remainder of this paper is organized into four sections. First and second, I present overviews of the current state of law and technology as they exist in and shape the realities of digital communications, privacy, and security with a focus on the needs of source protection for journalists. Third, I present some models for conceptualizing and implementing digital communications practices for journalists and newsrooms, in the context of current tools and communities. Finally, I offer recommendations for both industry development and academic research in the areas of digital privacy and security.

Interlude: Addressing Complexity

The difficulty of creating simple models to describe digital security risks and solutions stems from the fact that they must operate at the place where two major social systems–whose properties are almost perfectly juxtaposed–intersect. First, there is the law, which is intentionally slow, exhaustive, cautious, and reactive. Then, there is technology, which is inherently fast, emergent, experimental, and constructive.

And while our lives are shaped by both of them on a daily basis, their inner workings remain almost entirely invisible to all but the most highly initiated.

The crafting of laws and rendering of legal decisions often hinge on the byzantine interleaving of statutes, case law, and judicial inference that is argued in courts and described in documents away from the public view, and ultimately codified as binary decisions on the particulars of a given case. Technologies, meanwhile, exemplify the unpredictably complex expression of equally binary decision trees as they interact with the human world, yet their public form is often intentionally not readable by humans. The result is that nearly all of the workings of both systems are inaccessible to the public, expressed as they are in coded language and housed on largely proprietary systems.

In a healthy democratic society, the collective effect of citizens' individual actions in the political, economic, and social spheres constitute cultural "forces of nature." In this ecosystem, individual technologies are like cultivars–while their general features are known, their ultimate forms and behaviors are inextricably tied to their interaction with the broader environment. Law, meanwhile, is the "gardener" of that environment, and its role is necessarily reactionary and unequivocal. Prune here, thin there, tie back some stems and add support to others. Law does not determine what technologies come into being, and only once a "species" is known can law attempt to proactively influence its characteristics. As law and technology react and respond to one another, they create an ecosystem whose state is both dynamic and unpredictable. In order to be successful, digital security practices must be adaptive enough to acclimate to the changing circumstances of this *emergent system*.

Fortunately, our worlds are comprised almost entirely of similarly emergent systems–from the flow of traffic to the workings of party politics–and all of us capably navigate untold numbers of them in the course of our daily lives. The main difference between these and our digital communications systems is that most of these others operate largely in the visible and/or physical world, and follow rules that are accessible to us. The key, therefore, to creating a set of principles that supports journalistic values within digital communications systems is to make these systems at least conceptually "visible," and to translate the rules by which they operate into language that is broadly accessible to our community.

Digital Security is Not Sui Generis

In the physical world, we accept that privacy and security are context dependent. We appreciate that jaywalking is generally less safe than crossing at a stoplight, that postcards and loud public telephone conversations are less private than sealed letters and whispered exchanges. If we latch our yard gate, we do so knowing it will probably not stop a determined criminal, but may deter an opportunistic one. We know our front door deadbolt will not stop a SWAT team, but may delay an intruder long enough for our loved ones to escape to safety. An alarm system cannot extinguish a fire, but it may alert professionals to an emergency.

We are able to make informed judgments about our physical privacy and security because the rules and assumptions of the systems they involve are generally apparent and understandable to us. We also appreciate that these judgments–and the choices we make on their basis–are inherently probabilistic and imperfect. Crossing the street involves making numerous estimates about the speed of traffic, one's own crossing pace, and even the conditions of the road. An incorrect estimate may put you or those around you at substantial risk. Simiarly, a "shoulder surfer" may read our email at a cafe; a fellow passenger on a train may read documents we are holding. At all points we appreciate that neither our security nor our privacy is absolute.

This is exactly the same appreciation we must develop in our interactions with digital communications systems. As the forthcoming sections illustrate, the digital world is subject to all of the same complexities and probabilities as the physical one. And just as we have all learned the skills necessary to cross the street safely despite ever-changing road conditions, so too we can all learn to navigate the digital world in a way that keeps our sources–and ourselves–safe. We just need to learn how to look both ways.

The Law: Security and Privacy in Context

> “The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.”
>
> –U.S. Constitution, Amendment IV

“Privacy” and “security” in digital communications with sources can be thought of as two sides of the same coin: Our ability to offer our sources a sense of *security* in communicating with us is directly proportional to our ability to keep their real identities *private*. Central to effectively maintaining this privacy in the context of digital communications, however, is understanding the “right to privacy” as codified by the legal particularities that govern information retention and exchange in the digital realm.

Understanding the source and the scope of these distinctions begins with the notion of privacy itself. What does “privacy” really mean? While most of us have a loose notion that the Fourth Amendment affords us a “right to privacy,” this isn’t precisely the case. The Fourth Amendment, as written, affords us protection from “unreasonable search” by governmental authorities which, through legal precedent has come to include a notion of “privacy” as well.

The Judicial View of Privacy: The Fourth Amendment and Third-Party Doctrine

Over the course of many decades, the character of the American "right to privacy" has been both enshrined and circumscribed by the nature of the space in which an object or activity exists, and the process through which information about those activities is obtained. Individuals' "persons, houses, papers and effects" constitute private realms that should not be intruded upon by the government except in the presence of probable cause and the issuance of a warrant; where a space–either virtual or physical–is shared with others, its "public" character dictates that no right to privacy exists. This dichotomy is sometimes expressed via the shorthand: "No right to privacy in public."

Over time, the Supreme Court has occasionally interpreted and expanded the list of realms deemed "private," and therefore safe from unwarranted search. It was in the 1967 case *Katz v. United States* that Justice Harlan, in a concurring opinion, coined the term "reasonable expectation of privacy" that has largely come to represent the popular understanding of a constitutional right to privacy. At issue in *Katz* was law enforcement's placement of a recording device on a public telephone booth, through which it captured the contents of the defendant's end of a phone conversation. In ruling for the defendant, the Court held that the Fourth Amendment protected "people, not places." Harlan's same concurring opinion asserted that "electronic, as well as physical, intrusion into a place...may constitute a violation of the Fourth Amendment."

Yet the protections of *Katz* are highly circumscribed. Importantly, the contents of the conversation obtained in *Katz* were procured directly from the defendant himself by recording his own words. This makes Katz a "first party" to the content of the search, which is required for a Fourth Amendment claim. Because of this crucial detail, the ruling in *Katz* has little bearing on situations where information has been knowingly shared with others.

In 1976, this distinction was reinforced in *United States v. Miller*, wherein the court held that there can be no expectation of privacy around information–in this case, financial–that an individual has "voluntarily conveyed and…exposed" to a third party when he or she is aware that it may be shared with still others "in the ordinary course of business." Three years later, the court cited the *Miller* decision in *Smith v. Maryland*, which affirmed law enforcement's right to obtain the numbers dialed by an individual when placing a telephone call, on the basis that "telephone users…typically know that the numbers they dial are transmitted to the phone company and recorded." Under this rubric, which still applies today, telephone users have no "expectation of privacy" around the numbers they dial, even from a home telephone.

Though in *U.S. v Warshak* (2010), the Sixth Circuit asserted the defendants' privacy interest in the contents of their emails, the law supporting the search has not proven unconstitutional.

Together, the rulings in *Miller* and *Smith* help form the basis of what is broadly termed the "third-party doctrine" in American judicial proceedings, which holds that any information shared with a third party is not "private" with respect to the Fourth Amendment. As will be discussed below, only in cases where "privilege" has been established around a particular relationship (such as those with a doctor, lawyer, or, in some jurisdictions, a journalist) is there an exception to this general rule of thumb. Outside of such circumstances, *Miller* asserts, citizens should expect that any time they voluntarily expose information to a company or its equipment there is "the risk that the company would share that information with the police."

Though discussed in terms of privacy, it is worth noting that the decisions in *Katz*, *Miller* and *Smith* still hinge fundamentally on the Fourth Amendment's protection against "unreasonable search": Thus the core of the reasoning relates not to what was obtained, but how it was obtained. In *Katz*, the "search" of the defendant's conversation was made on his own person, whereas in *Miller* and *Smith* the search was of the defendants' financial institution and telephone service provider, respectively. The "reasonableness" of the search is therefore evaluated with respect to the rights of the organization in question, not the account holder(s), regardless of whether the search reveals information that pertains to them. Similar reasoning led to the 2012 decision in *United States v. Graham* , which held that "historical cell site" data–the telecommunication providers' records of the cell towers to which a subscriber's phone has connected–is not protected by the Fourth Amendment.

The Legislative View of Privacy: ECPA & FISA

The Electronic Communications and Privacy Act (ECPA)

In the decades since the decisions in *Miller* and *Katz*, bank transactions and telephone digits have become only two of the many types of metadata generated by the digital communications systems used by the public on a regular basis. And while the Supreme Court's rulings indicate that there is no constitutional right to privacy around this metadata, this in no way limits the ability of Congress to pass laws expanding individuals' privacy rights around these records. This fact has led to several pieces of legislation doing just that, including The Right to Financial Privacy Act (12 U.S.C. §3401), Video Privacy Protection Act(18 U.S.C. §2710) and HIPAA (42 C.F.R. §403.812).

Some of the earliest, and still most relevant, legislation in this area is the Electronic Communications Privacy Act (ECPA) of 1986, which defines the classes of metadata that telephone and electronic service providers may be compelled to share with law enforcement. Though successful in enhancing privacy protections in the context of real-time telephone wiretapping, ECPA was written at a time before email was commonplace, and when electronic storage costs were high and local-only access was the norm. The result is that while it remains the primary piece of legislation that governs data collection and sharing requirements for electronic communications today, the application of its provisions to both wired and wireless mobile communications, as well as online services, seems to expose more than it protects.

Crucially, the Stored Communications Act (18 U.S.C. §2701-12) portion of ECPA, and, in particular, its "Required Disclosure Of Customer Communications or Records" provision (18 U.S.C. §2703), enumerates the data points that service providers must turn over to law enforcement when provided with a subpoena. [16] As enumerated in section (c)(2), this information includes:

(A) name

(B) address

(C) local and long distance records, or records of session times & durations, including type of service used

(D) length of service, including start date

(E) telephone, instrument number or other subscriber number or identity, including any temporary network address

(F) means and source of payment, including credit card or bank account number if applicable

In the context of the technologies then prevalent, the metadata that ECPA makes available to law enforcement was not nearly as revealing of citizens' day-to-day activities as it is currently. According to recent research, reviewing even the relatively broad locational data accessible via GPS can illustrate an individual's movements and activities beyond what would be meaningfully feasible via physical surveillance. The "mosaic theory" concludes that "comprehensive aggregation of even seemingly innocuous data reveals greater insight than consideration of each piece of information in isolation."[17] Yet because we may "voluntarily share" our GPS location (and even more fine-grained cell-site) data

[16] With the exception of 18 U.S.C. §2703(d), which requires some showing before a judge. Both are less stringent than the "probable cause" required to obtain a warrant.

[17] When Enough is Enough, Bellovin et al. *NYU Journal of Law*, 2014

by virtue of our service contract with a provider, this information is not considered private. A similar circumstance is created when we surf the Web, as our Internet protocol (IP) address is shared with virtually every website we visit.

For more detail, see the next chapter.

The implications of ECPA are not limited to metadata, however. Section (a) of the "Required Disclosure" clause discussed above provides that contents of "electronic communications" in electronic storage that are more than 180 days old may be obtained by law enforcement via an administrative subpoena. Although the act explicitly exempts from the definition of "electronic communications" the kind of oral, tone-based, GPS and financial data that in 1986 constituted most of the general public's phone calls and wire transfers, it explicitly *includes* any "transfer of signs, signals, writing, images, sounds, data or intelligence"(18 U.S.C. §2510(12)): an almost perfect description of email. The result is that any email on a provider's server that has been opened or is more than six months old may have its contents accessed via such a subpoena.

18 U.S.C. §2705(a) provides for a renewable delay of notification for a period of 90 days if notification may have an "adverse result."

FISA & the PATRIOT Act–Ambiguity Abounds

Thus far, the policies we have discussed actually apply to anyone's electronic communications, not only those of journalists. In fact, the only part of ECPA that makes journalist-relevant stipulations is the controversial "business records" section of the PATRIOT Act (215–now 50 U.S.C. §1861). Though section 1(a) of the law allows the FBI to "require production of any tangible things" as part of an investigation to obtain "foreign intelligence information," this only applies "provided that such an investigation is not conducted solely upon activities protected by the First Amendment"[50 U.S.C. §1861(1)(a)]. Unfortunately, this latter characterization is described only in the guidelines of the Attorney General. Though these guidelines have been revised and their protections expanded[18] since the AP phone records scandal in the spring of 2013, these guidelines do not carry

[18] Holder Tightens Rules on Getting Reporters' Data, by Charlie Savage. *The New York Times*, 7/12/13

the force of law; journalists and their organizations have no legal recourse if they are breached.

As Eve Burton points out: "The AP cannot march into court and sue the DOJ."

Why Metadata Matters

> "Although the law provides less protection for metadata than content, metadata can be even more revelatory than content itself."
>
> –Susan Landau[19]

While the target of the AP phone records collection situation is still uncertain, the implications of metadata collection for journalists are clearly illustrated in the cases of both James Rosen and James Risen, whose alleged sources were first identified by law enforcement based on the analysis of telephone, email, and other communications' metadata. Coupled with the uncertain standing of reporters' privilege (discussed below), these cases are particularly troubling as there are indications that they are the early threads of a trend, as suggested by a department of justice official:

> "As a general matter, prosecutions of those who leaked classified information to reporters have been rare, due, in part, to the inherent challenges involved in identifying the person responsible for the illegal disclosure and in compiling the evidence necessary to prove it beyond a reasonable doubt."[20]

What makes up metadata?

Email: to, from, subject line, timestamp, attachment names, IP address

Mobile activity: origin number, target number, tower location, time, call duration, account holder information, hardware phone ID

[19] Making Sense from Snowden, Susan Landau. *IEEE Computer and Reliability Societies*, 8/13

[20] A High-Tech War on Leaks, by Adam Liptak. *The New York Times*, 2/11/12

What's Next

In recent months, conflicting conclusions about the constitutionality of metadata collection by the government have been evident in decisions issued by various circuit courts. Specifically, in *ACLU v. Clapper* (2014)[21] the 2rd Circuit held that metadata is not private, and that under FISA following connections up to three links away from the target part is acceptable. In December of 2013, however, the opposite ruling was reached by the D.C. Circuit in the case of *Klayman v. Obama*. As of April, 2014, however, the Supreme Court refused to hear Klayman[22] , meaning that it is likely to be some time before more clarity on these issues is gained.

Privilege & Source Protection

Testifying Against Sources

At the federal level, the question of reporters' privilege remains, in the words of one prominent law firm "in disarray."[23]

The Supreme Court recently rejected James Risen's appeal,[24] though without much explanation. Nonetheless, there are still a number of directly conflicting decisions in courts at the district level.

At the core of these conflicts are ongoing gaps in interpretations of 1972's *Branzburg v. Hayes*–the most recent Supreme Court decision related to reporters' privilege. Though in *Branzburg* the reporter's petition to quash a subpoena for appearance before a criminal grand jury was ultimately denied on the facts of the case, a concurring opinion by Justice Powell is sometimes interpreted as confirming the existence of *some* reportorial privilege, because in it Branzburg v. Hayes, Concurringhe asserts the importance of "striking the proper balance between freedom of the press and the obligation of all citizens to give relevant testimony."

While the Risen decision speaks poorly for the promise

[21] ACLU v. Clapper – Challenge to NSA, *ACLU*, retrieved: 6/6/14

[22] Supreme Court declines look at NSA case, Lawrence Hurley. *Reuters*, 4/7/14

[23] Amicus Brief, *Risen v United States*, LSKS, LLP. 3/26/14

[24] Justices reject reporter's bid to protect source. *AP*, 6/2/14

of any federal reporter's privilege, even in cases where it has been upheld, *Branzburg* makes clear that any privilege applies only to journalists, not their sources:

> "We note first that the privilege claimed is that of the reporter, not the informant, and that, if the authorities independently identify the informant, neither his own reluctance to testify nor the objection of the newsman would shield him from grand jury inquiry, whatever the impact on the flow of news or on his future usefulness as a secret source of information."[25]

Thus, reporters wishing to protect their sources must do everything in their power to prevent their being independently identified by law enforcement, as once this has been accomplished the source may be vulnerable to prosecution. In point of fact, reporter testimony often may be unnecessary to identify sources, as explicitly noted in Risen's case. In granting Risen's motion to quash his subpoena to appear before a grand jury, Judge Leonie M. Brinkema wrote that Risen's testimony was probably unnecessary as the government already had "numerous telephone records, e-mail messages, computer files and testimony that strongly indicates that Sterling was Risen's source."

As noted, forty-nine states and the District of Columbia, meanwhile, have case law or legislation that offer journalists some protection from being compelled to identify their sources. As the foregoing discussion illustrates, however, the authorities rarely need to obtain this information via reporters' testimony when journalists have been communicating with their sources digitally–the metadata associated with these communications is often sufficient to identify who is speaking with whom. Because of this, it is increasingly likely that even where some form of journalist privilege exists, reporters will never get the chance to invoke it;

the kind of metadata collection/analysis being conducted at the federal level is known to take place at the state and local level as well.[26]

[25] Branzburg v. Hayes, Opinion of The Court, Justic White. 6/29/72

[26] Agencies collected data on Americans, Ellen Nakashima. *The Washington Post*, 12/9/13

"This isn't the NSA asking for information." -Senator Edward J. Markey (D-MA)

Data and Testimony–Where Encryption Assists

Metadata and direct testimony, of course, are not the only means by which the authorities can obtain information about a journalist's sources. Searches of digital storage, such as a computers, hard drives and mobile phones, may reveal source identities and more.[27] Where the data in question has been encrypted, however, protecting it is still possible in certain circumstances.

A recent Seventh Circuit ruling held that forcing a defendant to decrypt the contents of a drive to which he had not already admitted having access and control was likely a violation of his Fifth Amendment rights. In some cases, simply requiring the act of decryption has been classed merely as "production" of materials–similar to handing over the key to a lockbox–and not subject to Fifth Amendment protection. Yet the Seventh Circuit pointed out that in some cases the act of production "has communicative aspects of its own":

> "...compliance with a subpoena tacitly concedes: (1) the existence of the documents, (2) their possession or control by the accused, and (3) the accused's belief that the documents are authentic."[28]

Thus, unless the government can "establish its knowledge of the existence, possession, and authenticity of the subpoenaed documents with 'reasonable particularity' " one may be able to resist a subpoena that requires decrypting data in response to a subpoena.

As indicated in this same opinion, however, Fifth Amendment protections can only be applied to communications which "relate a factual assertion or disclose information," and do not generally apply to actions such as standing in a lineup or providing a handwriting sample. As a result, decryption mechanisms that require only a fingerprint or facial recognition may not afford the same level of legal protection.[29]

[27] Armed agents seize records of reporter, Guy Taylor. *The Washington Times*, 10/25/13

[28] Judge Says Giving Up Your Password May Be A 5th Amendment Violation, Mike Masnick. *Techdirt*, 4/25/13

[29] Apple's Fingerprint ID May Mean You Can't "Take the Fifth", Marcia Hoffman. *Wired*, 9/12/13

The Technology: Understanding the Infrastructure of Digital Communications

Prologue: The Design Imperative of the Internet

The essential framework for the design of the modern Internet was first described by Paul Baran in a 10-page paper published in March of 1964. Issued as part of the IEEE's "Transactions of the Professional Technical Group on Communications Systems,"[30] the introduction to the paper–entitled, "On Distributed Communications Networks"–begins with the following sentence:

> "Let us consider the synthesis of a communication network which will allow several hundred major communications stations to talk with one another after an enemy attack."

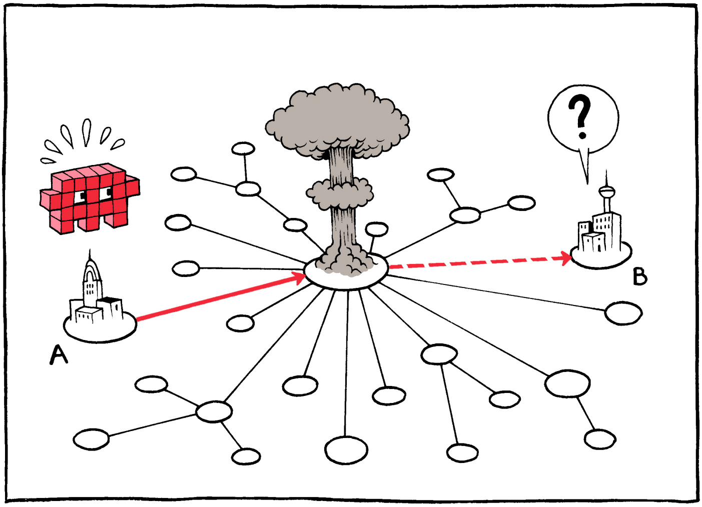
The paper goes on to describe the salient features of such a network, designed as a direct response to the most terrifying threat of the then-ongoing Cold War: a large-scale, long-range, distributed attack on American soil.

Just as George Washington forewent postal mail and employed private couriers to send messages to his troops[31] during the Revolutionary War, the U.S. military sought to guarantee the integrity and availability of its own communications in case of a substantial attack on commercial networks during the Cold War. Achieving this, however,

[30] On Distributed Communications Networks, Paul Baran. *IEEE CS-12, (1)*, 3/64

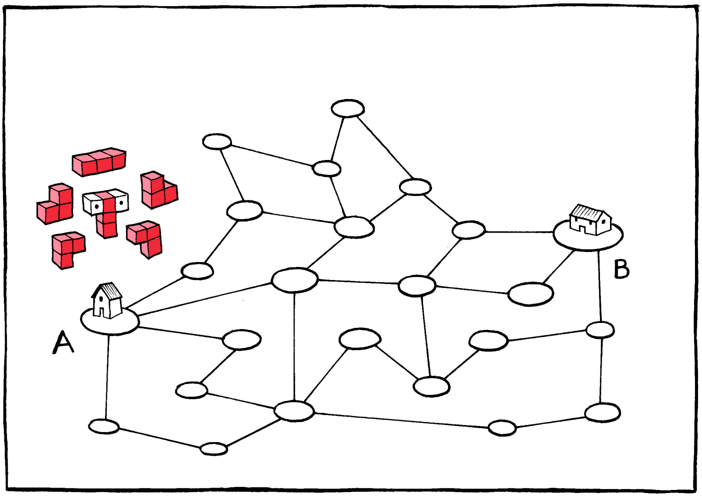
[31] *Spreading the News: The American Postal System from Franklin to Morse.* Richard R. John, 2009.

required that the system be all of the things that the telephone system of the day was not: distributed, redundant, asynchronous, relatively unpredictable, and cheap. Rather than requiring the synchronous, persistent connections of telephony, which transmitted information through the hubs and spokes of a decentralized network, the Internet would function something like an automated, multi-hop telegraph: Digital messages would be broken into pieces and individually addressed to the destination on the sender's end, then collected and reassembled at the other.

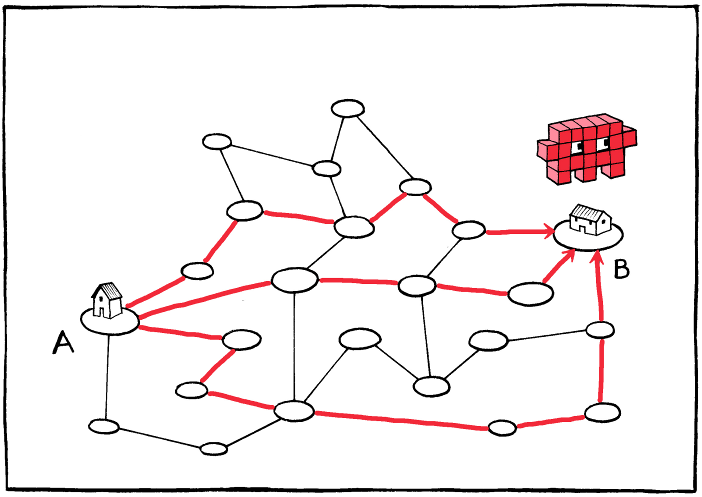


Centralized and decentralized networks fail quickly in the case of malfunction or destruction.

As anyone who has ever chosen the "ship as items become available" option knows, there is advantage to breaking up "deliveries" in this way. Packages (or in the case of the Internet, packets) can be routed opportunistically, moving toward their destination along whichever delivery path is most readily available. In addition to maximizing capacity, this approach also provides a certain level of unpredictability: Not even the operator of the system, much less an adversary, can know with complete certainty what the paths of these packets will be.

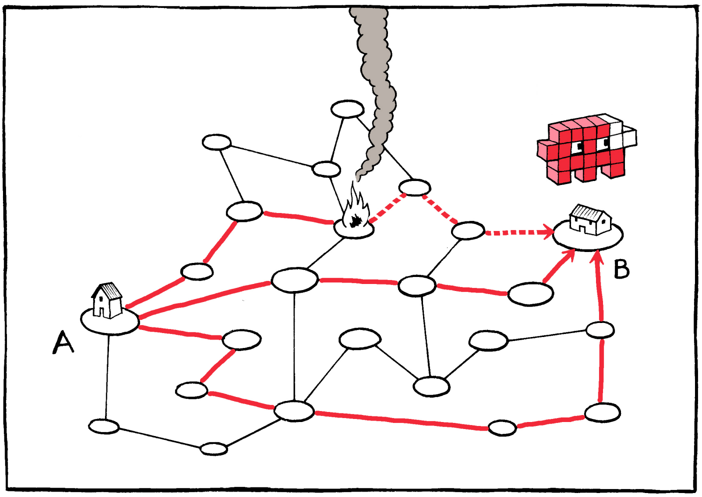


All Internet traffic is broken up into equal-sized packets.



Data packets take independent paths across the Internet once they leave your service provider.

Because the packets are small, there is room for redundancy: The same packets can be cheaply sent multiple times to ensure delivery. This means that the cost of losing any given packet is small; if the first copy doesn't reach the destination, the second or third will. In the worst case scenario, the packets' return address can be used to ask for any found missing.

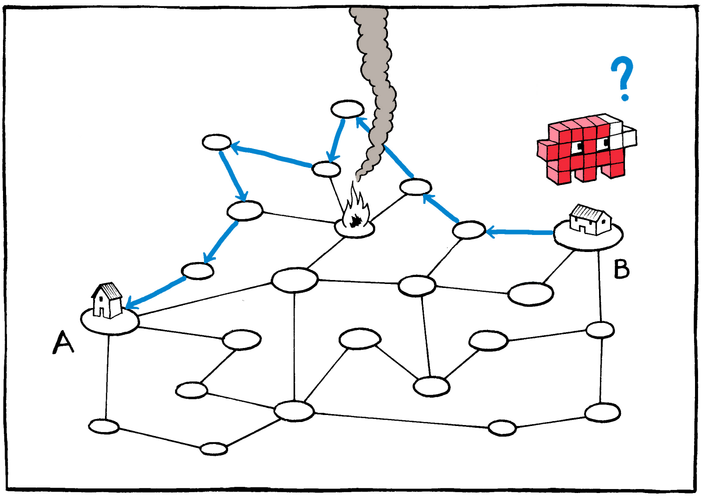


Data packets may be regularly dropped or lost in transit.

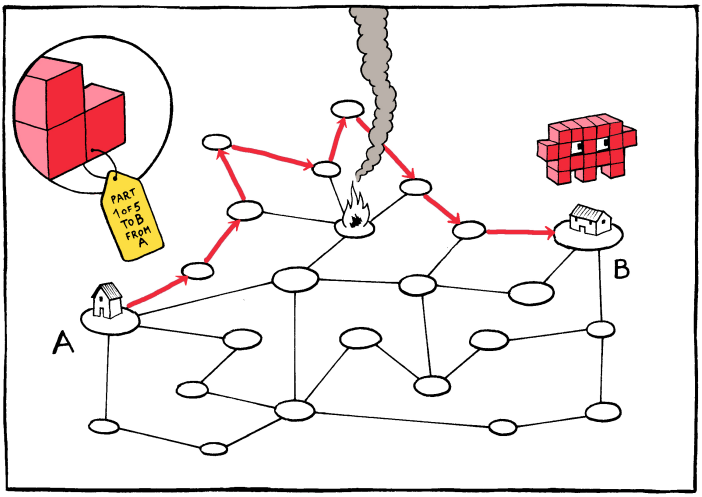
At the infrastructure level, this cheap redundancy–especially coupled with the return-address safety net–ensures that neither the individual routing nodes nor their actual connections to one another need be especially reliable. Connections may be intermittent or interrupted; nodes may malfunction or be destroyed. This means that the stations themselves can be both cheap and plentiful–and the more plentiful the nodes, the more resilient the network. Or, as Baran put it, the distributed network of the Internet is one:

> "...in which system destruction requires the enemy to pay the price of destroying n of n stations. If n is made sufficiently large, it can be shown that highly survivable system structures can be built, even in the thermonuclear era."[32]

[32] On Distributed Communications Networks, Paul Baran. *IEEE CS-12, (1)*, 3/64



If the recipient device discovers a piece missing, it can request another copy.



This request is made possible by the requirement that packets travel with a "return address."

In other words: an atomized communications solution for the atomic age.

Understanding the Internet

The vulnerabilities that your data faces in terms of observation, collection, and manipulation–both legal and illegal–stem directly from some of the essential features of the system described above. In order to function, the network of nodes and links that comprise the Internet requires certain information (target and return addresses, for example) to be attached to every data packet it sends. Moreover, the distributed nature of the system means that this information is seen by a large number of nodes and the parties that operate them.

Yet while we have little control over the exposure of metadata like the "to" and "from" addresses of our emails, choosing the right connection methods and using the right tools appropriately can reduce the general observability of our data as it traverses the Internet. By coupling an understanding of how digital communication systems work with the capacities of secure-communications tools, we can develop and adapt effective strategies for improving the privacy of our communications with sources and colleagues.

Getting to the Web

In the first incarnation of the commercial Web, the number of hands that touched your data on its way to and from the Internet was especially apparent. Accessing the Internet required that you wire your computer to a physical modem which was itself plugged into a telephone line and had it call a dedicated Internet service provider (ISP) when you wanted to connect. Because users typically paid by the minute and Web indices like Yahoo and Google did not yet exist, ISPs like America Online (AOL) often leveraged their gatekeeper status, funneling users through a proprietary landing page, with the broader Web only accessible beyond its threshold. In this configuration, the parties who could "see" our online activity were fairly apparent: the telephone company that received the call, and the ISP whose portal we used to send and receive data from the Web.

Though most of the visible signposts have since disappeared, our digital information passes through the same sets of hands today. In most cases, in fact, only two things have changed. In 2014, our telecom provider and Internet service provider will likely be one and the same, and our actual access to the Web–whether on a computer, phone, or tablet–is probably wireless. While undoubtedly convenient, these changes also mean that today the records of our online activities are concentrated in fewer and fewer hands, and as we walk around with our devices, the default configuration of our digital "return addresses" provides a more and more minute trace of our physical movements.

Let's Get Physical

Computers, mobile phones, and tablets are our tangible interfaces to the networked digital world. At some point, as they send and receive information over the Web, all of those virtual messages have to find their way back to the correct physical device so that we can access that email or load that news story. In the case of wireless routers, any number of devices may be connected to the same data "pipe." In order to correctly send and return all those little data packets that make up a Web page or even a Tweet, a wireless router must be able to uniquely identify each individual device. To make this possible, every Internet-enabled device on earth is assigned a unique machine address code (MAC) at the time of manufacture. By combining pieces of this ID with pieces of its own address, wireless routers are able to accurately exchange information with the Web on behalf of dozens of devices at a time.

The MAC address is a set of six semicolon-separated letter/number pairs: the first three sets comprise the "vendor id" indicating the device manufacturer, such Apple or Sony. The second three sets are the "device" ID.

Getting Connected

When a computer or wireless device is looking for a connection, it essentially does so by shouting its MAC out to the world and waiting for a router to respond, like a digital version of the kids' game "Marco Polo." The wireless router then uses the latter half of that MAC to create a temporary ID for your device, which it will store along with a variation on the Internet protocol (IP) address that the router itself has been assigned by your ISP. While the former is used to send information from the router back to your device, the latter is used as the "return address" for the information your device sends out to the Web. Because each of these is created anew every time your device connects to the router, your device's "return address" will change slightly each time you connect.

In a well-managed system your device's MAC should never move past the router. A malicious party operating the router, however, can capture this information.

Likewise, though IP addresses are designed to describe the geography of the digital world, they have a direct mapping to the physical world as well. This is not an accident; the opportunistic routing of the Internet only works if a node can tell which of its neighbors is the closest to the final destination marked on the particular packet. This is the layer at which digital and physical location correlate: For the ISP, a, IP address can be mapped, not just to a general geographic vicinity, but all the way down an individual router.

This illustrates just one of the ways in which our digital activities can be mapped to our physical identities: Not only does our ISP know the precise location of individual routers, it knows who pays the bills for them, including name, address, and credit card or other financial information. Since all of these are available for subpoena under the ECPA rules mentioned above, it is easy to see how Web browsing–more so from the home–is almost never substantively "private," especially from law enforcement or other government entities.

In point of fact, the wireless router in your home (if you use one) is not the only router through which your Web traffic is likely to pass. Almost all major ISPs have even larger routers that they use to route traffic from their customers in a particular geographic area out to the broader Web. Typically it is actually this latter, more general IP address that is visible to the broader Internet.

You can view your own current IP address by visiting: What's my IP?

Your Data's Identity

In the course of a day, most of us connect to the Web from at least several wireless access points–even more when we consider our mobile devices. Unfortunately, the "return address" IP information our data packets carry with them is by no means the only, or even the most granular, data that our Web activities carry with them. A great deal of information is stored in the software we use to actually access the Web: namely, Web browsers and apps. For example, most browsers share information about your operating system and browser type with websites, primarily so that they can provide you with the best user experience. Added to this, Web "cookies" - small bits of text stored on your hard drive via your browser - can be used to track your Web activity. Though cookies set by one website cannot be detected or read by another, companies like Google can set cookies on so many sites that even their cookies alone can uniquely identify a user;[33] programs that can access browsing history can do this as well.[34] This browser "fingerprint" can be used to draw together otherwise disparate threads of Web activity, and revealing your own identity and potentially that of a source.

Most websites can see - and log - the time, your browser type, and your location.

See "How cookies track you" for more.

Thus, every time we request a Web page, we are not only locating ourselves to our service provider, but fingerprinting and timestamping our traffic across locations, like thousands of pieces of digital registered mail. Moreover, even if the resolution of our "return address" is not particularly granular to the outside Web, there are other ways that our digital "fingerprint" can be connected to our physical identity. This is exactly what we do when we log in to a Web service, be it Facebook, or Gmail, or Twitter, or Amazon.com. To these

[33] NSA uses Google cookies to pinpoint targets, Andrea Peterson. *The Washington Post*, 12/12/13

[34] Why Johnny Can't Browse in Peace, Lukasz Olejnik et al. *5th PET Workshop*, 7/13/12

providers–or anyone who is eavesdropping on their communications or who has access to their logs–the browser fingerprint, digital-geographic "return address" and physical identity of Web activity is not only timestamped and transparent, but highly detailed and often verified, through billing or other personal information.

Though our device's exact "return address" on a network will change slightly each time one reconnects, browser fingerprints tend to change little over the course of weeks or even months. Once linked to your real identity, it can be used fairly effectively to track your activity on the Web across swathes of time and space.

The above applies regardless of which type of wireless device one uses to connect to the Internet: Laptops, tablets, and Internet-enabled mobile phones all employ these same mechanisms. Of course, given the constant signal hand-off among cell-towers required in order for you to walk down the street with a mobile device and carry on a phone conversation, load a Webpage, or provide required emergency services access, your service provider can physically pinpoint your phone to within 50 to 300 meters at any given time.[35]

What About the "Bad" Guys?

Until this point, we have been considering data collection, retention, and access that fall within the current business and legal norms of the United States. Whether these meet our personal preferences or professional requirements, they represent the default operating environment for digital communications when, in effect, all parties are following the rules.

But there are other players on the system as well. Hackers, identity thieves and hostile observers may also wish to track our behavior, identify us, and access or manipulate our accounts. Though we cannot control *what* metadata will be sent along with our information requests over the Web, we do have some choices about *how* it is sent. Making these

[35] Enhanced 9-1-1 Wireless Services, *FCC*, Retrieved: 6/7/14

choices wisely can have an enormous impact on the relative visibility of our Web activity to unauthorized observers.

Nothing But Noise

Whether bridging the physical gap between your device and a wireless router or sending data across the network to and from a website, there are two broad classes of connections available in each case: unsecured and secured. In an unsecured connection, information sent between the two points is transmitted in "clear" or "plain" text–as readable as this sentence to any party watching. With a secured connection, at least some level of encryption–essentially scrambling of the message–is applied before information is sent over the network, drastically reducing the ability of outside observers to infer what's being sent. As we will see below, these situations are analogous to sending and receiving postcards and letters (in envelopes), respectively.

Not surprisingly, the information that an observer can see depends upon his or her position in the network. If you connect to an unsecured wireless network, your Web activity (what pages you are visiting and so on) is potentially observable by any other device on that same network . If you connect to an unsecured website, the information you are sending back and forth to that website is potentially observable by anyone with access to one of the nodes on the network that passes it along. At a given moment, this would include anyone with access to the router you're connected to and the Internet service provider that links that router to the wider Web. Though questionable, "sniffing" packets on the broader Web is more than possible.[36]

For example, see Julian Oliver's "Reconstructing images from Web traffic"

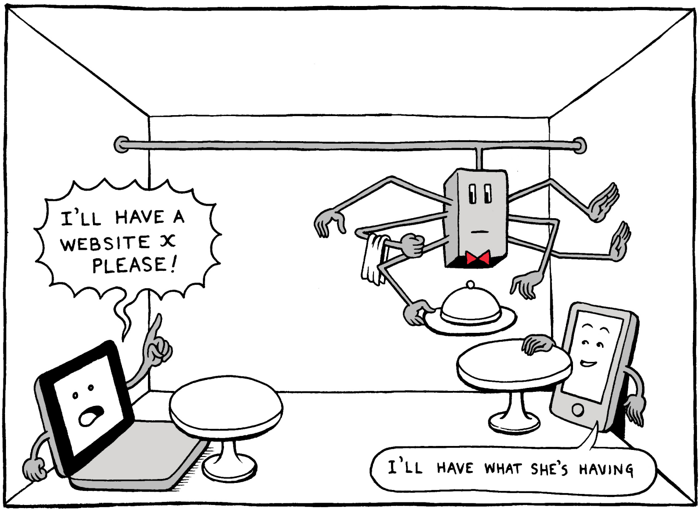
Wireless-level Security

In practice, using an unsecured wireless network is somewhat analogous to shouting your order across a crowded restaurant rather than waiting for someone to come and take your order

[36] AT&T willing to spy for NSA, MPAA, and RIAA, Nate Anderson. *Arstechnica*, 6/13/07

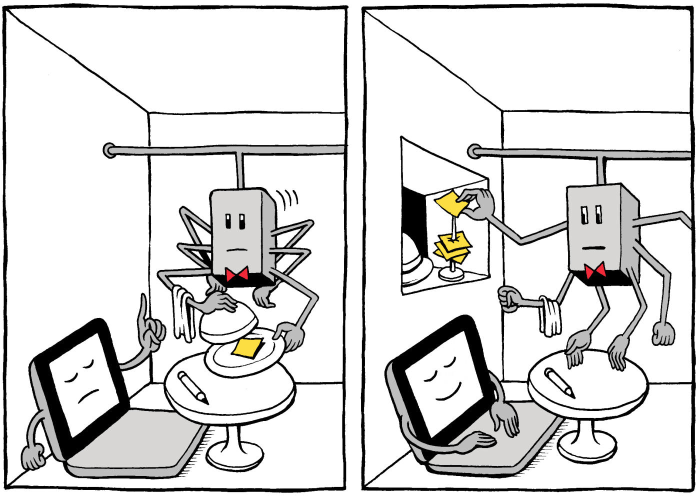
at your table. What you're sending and receiving can be heard by anyone in the room who chooses to listen: Your device is speaking in a language that anyone can understand.

An unsecured wireless connection means other devices can easily "overhear" the data your device sends and receives.



In a secure connection, however, communication begins with a kind of one-to-one greeting process. Your device identifies the router and sends it a short message, indicating that you would like to connect. The router then replies to your message with a randomly generated nonsense string that both devices will use as the basis for a kind of elaborate Pig Latin in which all of your messages will be exchanged from that point on. Thus, while your device is still sending all its messages across the room, those messages are protected in such a way that only your device and the router can "hear" what is being sent.[37]

[37] "WPA2 enterprise" security is most robust; though "WPA2 home" has some vulnerabilities, both are far preferable to an unsecured network.



In a secure connection, the messages that devices exchange with the router are protected.

Small s, Big Significance

When it comes to connecting to a website, one also has the choice between a secure and an insecure connection. At any given moment you can tell which type of connection you are using simply by looking at the url bar of your browser window–the location will start with either `http` or `https`. The significant difference between these two lies in the little `s` at the end of the second; it stands for "secure."

What does it mean for your connection to a website to be secure? In effect, it is much the same as in the case of the wireless connection described above: A secured connection means that the information you are sending back and forth to that website is encoded in such a way that your messages look like gibberish to anyone observing.

At the same time, there is an important difference between communicating with a router that is in the same room as you are (or very nearby) and communicating with a website that may be on the other side of the world. In a coffee shop or library, there is almost always a third party–a person, maybe a sign–that provides the name of the network you should connect to, and (if it is secure) its password. Though most of us assume this information is correct if we're able to successfully connect to the Internet, our decision to use that network also implies a good deal of trust: We trust that the router is properly secured, and even that it is the router it claims to be. If we are familiar with the business or organization we may not give this a second thought, in the same way that we routinely trust waiters and salespeople not to abuse our credit card information. Yet it is important to remember that this expression of trust is implicit in our decision to connect.

On the broader Web, we also need someone to vouch for the fact that the website we're connecting to is actually the one we think it is, because it's also incredibly easy to make one website look like another. In fact, even if you type the URL of the website you want to visit directly into your browser's url bar, it's possible to end up at a website that only *looks* like the one you intended to visit. This is also how some Internet censorship is implemented,[38] as was seen recently in Turkey.

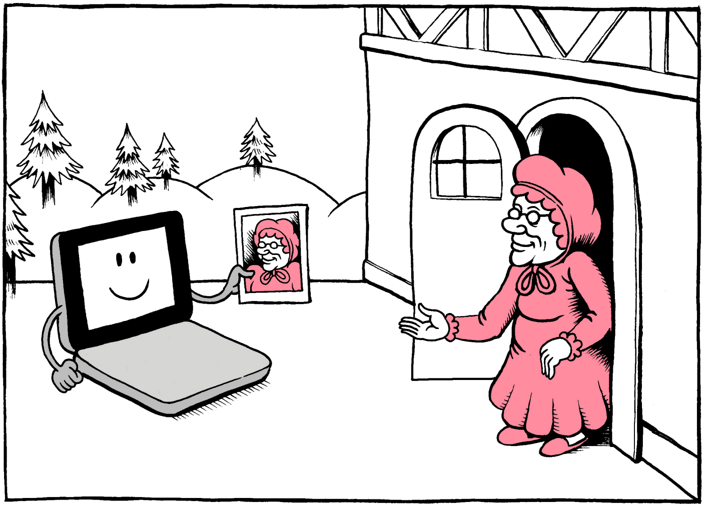
Websites are actually located via IP address, not URL. When you enter a URL in a browser, its IP address gets looked up like a telephone number: by matching the URL "name" to the IP "number." In this case, your browser checks the given URL with each of its Domain Name Servers (DNS). An incorrect or malicious DNS, however, can match your URL to the wrong IP address.

[38] Internet Censorship by DNS Injection, Anonymous. *ACM SIGCOMM, (42)(3)*, 7/12

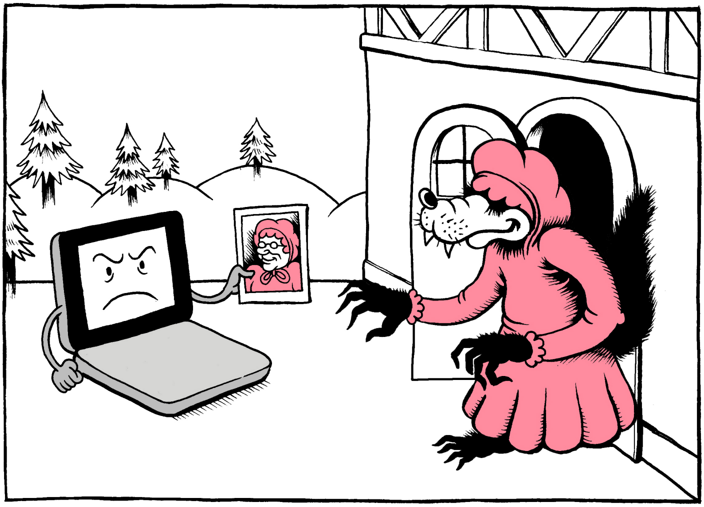
So how can you be sure that the secure connection you establish is to the website you're actually looking for? In practice this is made possible by the fact that https connections don't just establish secure communications willy-nilly. Before agreeing to encode your communications with a website, your browser first asks it for identification, in the form of a *security certificate.* These certificates work something like digital passports for the Web. They are issued by certificate authorities which generally provide them after running a kind of background check on the website, making sure that it is actually owned by the company or individual who claims to run it.[39] Before agreeing to establish a secure connection to a website via `https`, your browser asks the site for this "passport," and then checks with the issuing authority to make sure it is authentic. If the credentials check out, your browser moves on to the next step and actually establishes an encrypted connection. Otherwise, it will throw up a warning, letting you know that something's not right.

[39] Make HTTPS and Email More Secure, Peter Eckersley. *EFF*, 11/18/11

This part of the process is known as "authentication."



A valid security certificate means your browser will establish a secure and *trusted* connection.

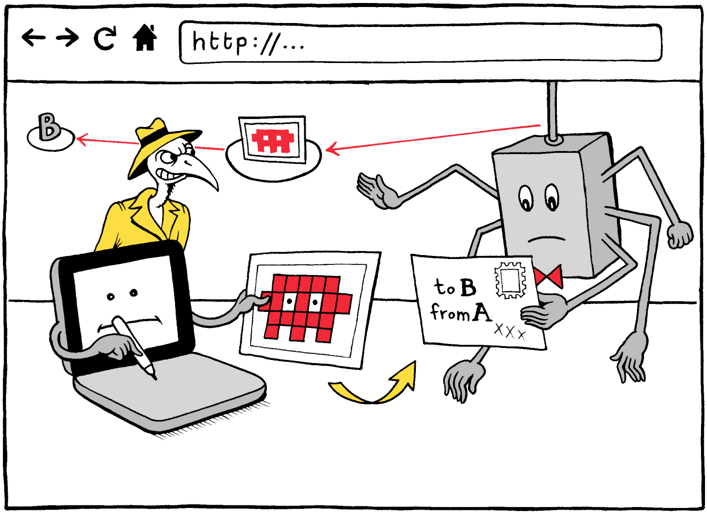


A bad security certificate will cause your browser to throw up a warning.

What an Authenticated, Encrypted Connection Means, and What It Doesn't

So let's say that at this point you've connected to both a router and a website securely. Does that mean that no one can see what you're doing on the Web? Not quite. Remember that the information you send to and from that website still has to find its way from your device to the website and back again. To do this, every node along the route needs access to some information about your messages in order make sure they get to where they're going. This is where our metadata comes back into play: Every node on the network still needs to know where each message is coming from and where it's going. In this sense, your encrypted Web traffic is something like mail in a *very* sturdy envelope; anyone can see where it's coming from and when, as well as where it's going. But only your device and the website you're communicating with can open the envelope to see what's inside.

By contrast, if you connect to a website via `http`, it's more like sending your information via hundreds of postcards; not only can every node that handles it see where it is coming from and where it is going, the contents of your messages are there for anyone to read as well.

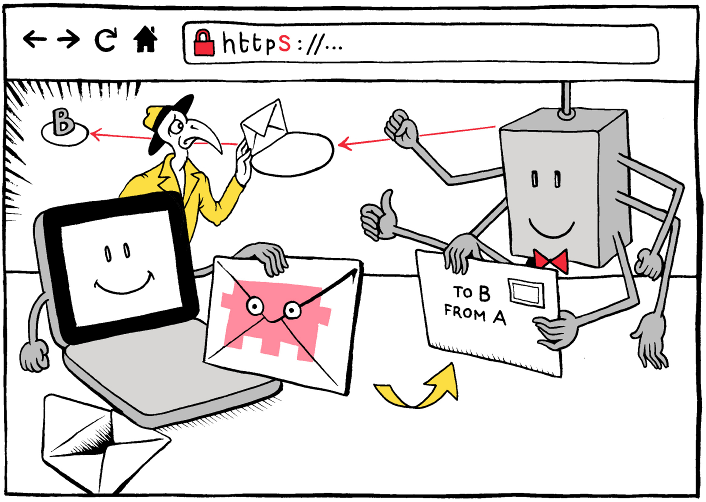


An http connection leaves all the contents of your packets visible to anyone on the network.

This is why using https connections is so important, especially when you are sending sensitive information like usernames and passwords. The same goes for any website where you might enter financial, medical or personally identifying information. In spite of this, many websites that ask for and deal with sensitive information don't always require or provide https connections. Until very recently, for example, Yahoo did not require an https connection to log in to its mail service.[40] Ideally, one would use https connections as much as possible. Fortunately, however, the Electronic Frontier Foundation recently launched a project aptly named HTTPSEverywhere, a free browser plugin that always attempts to make a secure connection, and then only revert to an insecure one if the former is unavailable. Installing this on your browsers can help make using the most secure connection available a more seamless part of your Web browsing activity.

[40] Yahoo finally enables HTTPS encryption for email by default, Liam Tung. *ZDNet*, 1/8/14

An https connection, meanwhile, protects the contents. Only the metadata remains visible.



Beyond the Exchange: Data Security in the Cloud

Understanding Email

Thinking of messages sent over http as postcards and those sent via https as letters in envelopes is actually a fairly good analogy for how your information is and isn't exposed when you're actively sending messages to and from a website. But what about after that exchange has ended? At least some of the information we exchange with websites–be they financial websites, social networking services, or email providers–ends up being stored by them for various purposes. Obviously, any website that requires a username and password will need to store those in order to protect your account. Social networking services host copies of our posts and photos; financial institutions retain records of our accounts and transactions. Email providers that offer Web-based email access maintain copies of our emails.

While all of this is to be expected, the way this information is *treated*, both legally and technically, may not always meet our expectations. For example, it is generally regarded as good practice to store passwords only in encrypted form in case of a security breach, though leak incidents demonstrate that not all service providers adhere to this security rule of thumb.[41]

To protect your information from outside attackers, financial institutions and email providers also typically encrypt the copies they maintain of your account information and emails. At the same time, these companies have the ability to *decrypt* your information as well; if they didn't, opening an email on the Web would only reveal encrypted gibberish, rather than the readable text of the note that your best friend sent you this morning. Obviously, this is part of what makes Web-based email services so convenient–we can read and refer to our email from anywhere.

Google's recent "end-to-end" offering may address some of these issues, but further analysis is needed to assess exactly how the library works.

The flip side of this convenience, however, is that your email provider has as much capacity as you do to access the readable text of the note your friend sent to you. If their servers are successfully hacked, the attacker will be able read all of your emails as well. Perhaps more likely–though arguably equally problematic–there are many cases where your email provider may be compelled to use its access to provide the text of your email messages to requesting authorities.[42] This capacity can also be leveraged by law enforcement to compel an email provider to decrypt and share emails and other information that is more than 180 days old, because these are considered "abandoned" and so can be legally obtained through a subpoena–contents and all (18 U.S.C. 2703(a)).

Understanding Endpoints

Even without recurring headlines of cybercrime, attacks on Web servers, and security breaches, it probably seems commonsensical that businesses would use encryption to protect the data they store. After all, a bank's website is as obvious a target for a criminal attack as a brick-and-

[41] 42 million unencrypted passwords leaked from Cupid Media, Darlene Storm. *Computer World*, 11/20/13

[42] Snowden's E-Mail Provider Defied FBI Demands to Turn Over Crypto Keys, Kevin Poulsen. *Wired*, 10/2/13

mortar branch would be; perhaps even more so given that it can be accessed from anywhere and is likely a conduit to a much greater volume of assets. Not encrypting sensitive organization and customer information would be an obvious security hole.

But the data on your computer doesn't need to be a gateway to millions of dollars in order to have value to an attacker. Journalists and those close to them may be the targeted for the information that their devices contain about both sources and stories: contact lists, interview notes, source documents, etc. Even for non-journalists, gaining access to your device can be valuable to the authorities or a criminal seeking information, since many of us store passwords or sensitive financial and medical information on our computers and phones.

In security-speak, an "endpoint" is any device that stores information. In this sense, Web-service providers' servers are endpoints; so is your laptop or mobile phone. Protecting them generally requires following some simple rules: Don't leave them in a situation where others can gain physical access to them; and to be on the safe side, encrypt them. We are apart from our devices more than we may readily imagine, whether to use the bathroom at a coffee shop or pass through a security checkpoint at an airport. Probably the most common and high-profile example of the latter situation was exemplified when Glenn Greenwald's partner, David Miranda, flew through London Heathrow Airport in August of 2013; his computers and hard drives were taken from him for several hours.[43] Any unencrypted information they contained would have been readily accessible to authorities. Just because you have a password on your computer or phone doesn't mean its contents are encrypted; this is something you need to set up explicitly.

[43] David Miranda:"They said I would be put in jail", Jonathan Watts. *The Guardian*, 8/19/13

Don’t Get “Pwned”

That said, encryption will not protect your data against malicious software that you have voluntarily (if unknowingly) loaded onto your device yourself. This most commonly happens through USB drives and downloads, both of which may contain hidden programs designed to access, manipulate, and/or communicate your private information whenever you next connect to the Internet. Trusting the source of the device or document isn’t enough to protect you. Brand-new USB drives have been known to have malware embedded by their manufacturers, and documents that can run programs–such as PDFS–may contain malware of which even the sender is unaware. This doesn’t mean that working with these resources is impossible, but it does require taking some simple precautions. One such approach is opening potentially problematic filetypes in a service like Google Docs first; these services are built to scan for and eliminate malware. Another approach is to set up an “air gapped” computer, which is simply an old machine that you never let connect to a network. This way, any malicious software is starved of the connection it needs to leak information. Of course, the first option only makes sense if the information is not sensitive; anything you save to Google Docs (or a similar Web-based service provider) falls under the same legal and technical rules of access as those described for email above.

The term“pwn” (pronounced p’ōn) derives either from World of Warcraft or chess.

The Strategies: Source Protection and Digital Security

Preface: Beware the Buzzwords

Since the Snowden revelations of 2013, a seemingly endless supply of app developers and service providers has emerged onto the online landscape promising "private," "anonymous," and/or "secure" communications. Despite the recent SnapChat settlement,[44] however, companies' use of these terms is unregulated and the terms themselves are generally poorly defined.[45]

Yet in order to meaningfully protect our digital communications, clarifying the nuances of these terms is essential. We began that process in an earlier section, where we exposed the legal concept of privacy to be importantly limited to, among other things, information that has not been shared in any way with a third party.

"Anonymity," meanwhile, is a word that is often used as a catchall to describe communications that may, in reality, be anonymous, pseudonymous, or simply unpublished. In journalism, citing "anonymous" sources is usually "a last resort"[46], but such sources are typically anything but: Their physical and/or legal identities are likely well-known to the reporters and editors with whom they work. "Pseudonymity," meanwhile, describes any handle other than our direct physical or legal identity. In this sense, all digital communications are technically pseudonymous, whether or

[44] Snapchat Settles FTC Charges, *FTC*, 5/8/14

[45] Whistleblowers Beware: Apps Will Rat You Out, Andy Greenberg. *Wired*, 5/14/14

[46] The Disconnect on Anonymous Sources, Margaret Sullivan. *The New York Times*, 10/12/13

not we generally think of them as such. Pseudonyms may be persistent or one-time use; an email address of your legal name is a pseudonym as much as a throwaway Reddit account. As we will see, however, the most important characteristic of an effective pseudonym, for journalists and sources, is whether or not it is *linkable* to its user's physical or legal identity.

The concept of *linkability* is crucial precisely because, as discussed above, it is what undermines the practical value of most current shield protections for journalists. Though the concept of source protection is often thought of as not "naming" your source, the reality is that the defaults on most email, chat, text, and telephone systems do exactly that with every exchange through the digital traces these activities leave behind. The metadata records stored by service providers can so efficiently link journalists to their sources that the authorities need rarely make the effort to take journalists to court. This means that where protecting the identity of a source is truly necessary, it is essential that these communications be *unlinkable*.

Though in Risen's case the government argued that "[n]o other person can [identify] Sterling as the individual who disclosed the national defense information", Risen's grand jury subpoena was quashed precisely *because* many metadata records already linked him to Sterling.

As we will see below, effective use of pseudonyms is an essential component of achieving *unlinkability*. Yet pseudonyms themselves present something of a conundrum: How can we know who really "owns" a particular digital identity? This fundamental issue is one of the reasons that so many of us have rushed to "claim" the email addresses comprised of our legal names on major service providers, and why many organizations follow a simple firstname[dot]lastname[at] pattern when generating email addresses as well. These practices recognize the problems of *authentication* and *verification* - determining who "owns" an email address or telephone number. While we often assume that an email address made up of someone's legal name belongs to them, in reality we also take steps to verify digital contact information, though often without thinking about it: Someone gives us their email address at an event, and we continue the in-person conversation digitally; we call a phone number we've been given and the voice that answers sounds like our friend or colleague; we arrange to meet someone via chat and the person we expect shows up at the right place and time. In each case, we use some kind of non-digital communication to verify that the person using the email address, telephone number or chat handle is the same person who provided it to us. In the digital world, this is often described as "out of channel" or "out of band" verification, and it's an important aspect of using all digital pseudonyms, including the "public keys" that we will discuss below. The complement to this process, meanwhile, is *authentication,* where someone proves to us–through their physical self or voice–that they are who they claim to be. If authenticating a person's identity ourselves isn't possible, then we have to rely on the word of someone we trust.

Of course, the easiest part of protecting our information–whether it's footage, photographs, notes or a source list–is through a strong password practice. As we'll see in more detail below, the simplest way to do this is simply to stop thinking pass*word* and start thinking pass*phrase.* This makes for logins that are both easier to remember and harder to crack.

Protecting Your Sources, Protecting Yourself

At the heart of most U.S. shield protections for journalists is a simple premise: If journalists cannot protect their sources, it substantially harms their ability to obtain the information they need to hold government accountable–perhaps the fundamental objective of a free and independent press. Given the current legal and technical realities, however, journalists who communicate with their sources digitally may be rendering these protections essentially moot. In practice the defaults on most email, chat, text, and telephone systems effectively identify our sources with every exchange, so protecting them means successfully scattering these digital traces so they cannot be used to connect the dots.

Of course, we work with many sources whose connection to us can be, and is, acceptably known, through long association or publication. Yet communication with these sources needs protecting too. As every journalist knows, sources sometimes don't appreciate the implications of the information they share; ensuring confidentiality of their communications with us is an important part of source protection even if their identity is public.

Because of this, you'll find the content below broken broadly into two sections: strategies for protected but *linkable* communications, followed by strategies for *unlinkable* communications. First, however, we'll address the three main methods of digital information protection: encryption, obfuscation, and deletion.

Encryption

Put simply, encryption is the process of scrambling or encoding messages in such a way that only someone with the correct "key" can unlock or unscramble the original.

In digital communications, there are two primary types of encryption: *symmetric* and *asymmetric.* In symmetric encryption, a single, secret key is used to both encrypt and decrypt a message. This approach is strong and fast, but requires that sender and recipient somehow agree on–and *securely* share–a single secret key. Symmetric encryption has been used for thousands of years and, at its heart, is very similar to the kinds of alphabetic substitutions that you might find on a decoder ring.

Asymmetric encryption, on the other hand, works by generating a mathematically related "public-private" key *pair.* Though each key can decode a message encrypted by the other, the two keys are asymmetric in the sense that the public key can be generated from the private key, but not the reverse. How is this possible? Asymmetric encryption takes advantage of the fact that it is generally much easier to mix things together than it is derive the original components from the finished product. For example, it is easy to create some color of green paint by mixing together yellow and blue. But the only way to tell what proportion of yellow and blue paints went into making a particular shade of green is through a long and arduous process of trial and error.

In mathematics, multiplication is fast, but factoring is time-consuming–even for a computer. The only way to find the factors of a number is to work your way up the number line, testing every possibility as you go. Public-private key pairs are based on this principle: The private key consists of a unique set of factors that when multiplied together yield the public key. Use enough factors–preferably prime numbers–in your private key, and it would take today's computers decades or more to derive the private key from the public one.

These special properties of public-private key pairs let us do things with them that we cannot do with symmetric keys. For example, we can (as the name suggests) make the public key *public*, and claim it openly on the Web. Anyone who wants to send us an encrypted message can encode it with our public key, knowing that *only* the owner of the private key can decode it. Likewise, by sharing a message encoded with our *private* key, we let others verify that the public key we have indicated truly belongs to us. Of course, knowing that a person controls a particular key pair doesn't actually tell you *who* they are. That step–confirming the "real" identity of the person who controls a particular key–is known as *authentication*.

For individuals, authentication can be done by *securely* sharing the *hash* (sometimes also called called a "fingerprint") of your public key. Most simply, this can be done in person, via business card or QR code. Voice authentication is also a good option, since we tend to recognize individuals' voices. Even postal mail can be an option, if you're confident about where to physically reach the person with whom you're trying to communicate. For websites, third parties

A typical PGP hash is 32 characters long.

vouch for the legitimacy of a public key by "signing" (or authenticating) it with their own. This is the equivalent of believing your friend when he or she gives you someone's email address.

In practice, virtually all digital encryption systems are *hybrid* systems: They use both symmetric and asymmetric encryption. Typically, this means encrypting the actual message with a unique symmetric key, and then encrypting the symmetric key itself with the appropriate public key and transmitting it with the message. This is the process that underlies both secure (https) Internet connections and encrypted email .

A website's "security certificate" is just another name for its public key.

There are cases, however, where no already-known "public key" is available to encrypt that symmetric key; when connecting to a wireless router, for example, or using many secure chat programs. For these, keys must be generated and exchanged on the fly, using a process called Diffie-Hellman key exchange. Despite the fact the fact that the first messages are exchanged "in the clear," this process makes it possible for both devices to derive the same shared secret key.

For a helpful demonstration of this type of exchange, see Chris Bishop's segment for the *Royal Institution Christmas Lectures*

There is a vulnerability here, of course. How does one know that the *first* message really came from the person you think it did? If you had a public key to compare it to, it would be easy to check. Without this, however, it is possible that a third party could intercept your communications and impersonate each side to the other–all the while decrypting and reading all of the messages you exchange. This is known as the "Man in the Middle" (MITM) attack.

Fortunately, the MITM attack is simple to thwart: Simply telephone the person with whom you are chatting to verify that your secret keys match (many programs will display them on screen), or exchange it via encrypted email (provided you've already or *authenticated* that the public key you have really belongs to them). After that, you can chat with confidence.

Obfuscation

Obfuscation is exactly what it sounds like: digitally "hiding" information, whether it's data stored on your computer or your IP address on the Internet. Some forms of obfuscation also provide "plausible deniability"; in other words, the reasonable appearance that nothing is being hidden.

In general, obfuscation is difficult to do on one's own; its effectiveness depends primarily on your data's ability to "blend in." For example, using "hidden volumes" to make an archive of sensitive documents *look* like a movie file works best if you have a reasonable number of movie files on your computer. That way, the chances of an attacker locating the one that is not actually a movie is much lower. Similarly, using a VPN or Tor to mask the geographic origin of your Internet traffic (addressed shortly) works best if there are many other users on the same system. In this sense, obfuscation can be best understood as a kind of herd protection, much like digital security in general. The bigger the crowd your data or operations can blend into, the more difficult it makes you and your sources to target.

One of the common challenges to the suggestion that more people use encryption is that there are environments where it makes one stand out, and thus a target for greater scrutiny. There are many cases where this is true. Obfuscation is exactly the principle of "security by obscurity," which may actually mean *forgoing* encryption in certain contexts.

Deletion

Ultimately, no kind of search can expose information you don't have, both for individuals and service providers. Regularly and securely deleting unnecessary emails and files, especially from hosted services, is a simple and effective protection against having your data rifled through either by the authorities or by hackers. Think of this as cleaning out your file cabinet on a regular basis: Do it once every three months to help keep your exposure in check. And remember that simply "trashing" your information isn't enough. Online you'll need to "delete forever," and on your computer you'll want to use a tool like CCleaner to truly overwrite "deleted" files so the data can't be retrieved.

Recall that the contents of emails over 180 days old may be subject to subpoena, so doing this every six months is a minimum.

When Linkability is Acceptable

In the bulk of day-to-day journalism, source linkability is not a paramount concern. We tend to prefer our sources "on the record" in the first place, and once quoted in an article their connection to us is public–and published–knowledge.

At the same time, concision isn't the only reason we don't publish entire interview transcripts. Most source conversations contain a mix of "on" and "off the record" remarks; not everything we discover in an interview should be discoverable. Yet if these exchanges take place through (unencrypted) email, shared documents, or chat applications, "discoverable"–both legally and technically–is exactly what they are.

As in the physical world, no digital security measure is absolute, and ultimately the onus falls on individual journalist or organization to determine which measures should be baseline and which applied only in certain cases. Whatever the decision, however, its consequences will most substantially be borne by your source, not yourself. A situation need not be life-threatening to be a significant risk: Loss of a job or professional standing, family and marital consequences, financial and/or legal liabilities can all be the fallout of source communications coming to light. In the majority of cases, the risk you take is not your own.

> "We're allowed to make choices about risks to ourselves, not our sources."
>
> –Jonathan Stray

Encrypt to Protect

The main key to protecting your communications with known sources is simple: Encrypt, encrypt, encrypt.[47] Where encryption isn't possible for legal, technical, or resource reasons, consider communication alternatives that may be more familiar to your correspondents or have better legal protections, such as postal mail.

No matter the communication method, there is a good range of solutions out there for all kinds of devices. Android devices tend to have a greater variety of open-source options, but there are several trusted cryptographers making solutions for Apple devices as well. In general, authentication and encryption require that both parties use the particular app in question; this can made Web-based solutions like CryptoCat valuable for journalists trying to move their sources to more secure communications channels, since only a browser plug-in is required.[48]

Web Browsing

If you've ever worked in a large organization, chances are you've had to use a "virtual private network," or VPN at one point or another. VPNs funnel all of your Web traffic through an encrypted tunnel into the network for which they are configured. This provides two types of protection for Web activity, in that it both encrypts your traffic and obfuscates its origin, by sending it first to your organization's network–wherever in the world that is–and *then* out to the broader Internet, no matter where you are. Using your company VPN or a similar commercial solution like Private Internet Access or TorGuard to route your Web traffic through a known network can help protect you from a potentially hostile Web provider.[49] Keep in mind, however, that while the ISP you're on will only know that you're using a VPN, your company or VPN provider will be able to see all of the Web requests you make. An IPsec VPN will route not

[47] Privacy Tools: Encrypt What You Can, Julia Angwin. *ProPublica*, 5/6/14

[48] Crypto for the Masses, Quinn Norton. *The Daily Beast*, 5/12/14

[49] Five Best VPN Service Providers, Alan Henry. *LifeHacker*, 3/23/14

just your browser traffic but all of your Web traffic (e.g. for applications like DropBox etc.) through the VPN.

To truly *anonymize* the location of your Web browsing activity, however, the Tor Browser Bundle is your best bet. The Tor Browser is an open-source, cross-platform browser, described in more detail below, that is an important tool for anonymizing your location online, especially in places where VPNs may be banned outright.

Email

Email presents a special security problem for a number of reasons, but the first step you can take is not to store your emails on any Web server longer than necessary. Set up a local email application like the free, open-source and cross-platform Thunderbird on your computer and connect it to your regular email account. Even if you don't use the application much on a day-to-day basis, you can use it to follow a "download-and-delete" protocol where every three months or so, you download all of your emails that are more than a few months old and archive them locally. After you've backed them up to an external hard drive, delete the copies on your Web-based email provider. You will still be able to save and search your old emails, but since they reside on your own physical computer, they will be better protected against subpoena-based searches.

GPG (GnuPrivacy Guard) is the open-source version of PGP ("Pretty Good Privacy"), which was created by Phil Zimmerman.

For more information on working with GPG, see the GPG Mini How To

Thunderbird also provides support for encrypting and authenticating email, through its Enigmail add-on and GPG. While setting up some of these systems can take several steps, once you've generated your public-private key pair and connected it to your email application, all of this happens in the background. The software can store the public keys of your correspondents, and performs the necessary encryption/decryption and signing of your messages with the click of a button. Though Google recently announced an "end-to-end" encryption option for Gmail, the fact that it's still decrypted in the browser makes it less private than options like Enigmail, where the only decrypted version exists on your physical computer.

Chat

There is a range of encrypted chat tools for both desktop and mobile that can be used with your existing services (e.g., GoogleTalk) while still encrypting and authenticating your conversations. Adium for Mac and Jitsi for PC are open source and work with most major chat services. CryptoCat, meanwhile, is both open source and Web-based, requiring only a simple plugin installation.

Adium also lets you manage contacts for multiple chat services through a single interface.

On mobile, ChatSecure for Android works with GoogleTalk. SilentCircle and Wickr , have apps for both Android and iPhone that support encrypted chat as well as encrypted text messaging and voices calls with other users of the app.

Phil Zimmerman is a founder of Silent Circle. Noted security researcher Dan Kaminsky and Whitfield Diffie are advisors to Wickr.

Text Messaging

A good Android solution for texts communications is TextSecure, which supports message-level encryption and authentication and also encrypts the texts stored on your phone. SilentCircle and Wickr, mentioned above, also offer encrypted text messaging.

Voice

Encrypted voice and video communication over the Web depend on secure “SIP” services, such as those supported by ostel.co or linphone.org, which can handle these more complex data streams. When used with these services, Jitsi supports encrypted voice calls from PCs, while OSTel does the same for Android devices via the CsipSimple service. SilentCircle and Wickr offer encrypted voice call support. RedPhone also provides an encrypted-voice option, but relies on your existing phone number, so should not be used where unlinkability is required.

SIP, or “Session Initiation Protocol” is similar to http, but used for exchanging multimedia and voice data over the Web. Just as secure Web connections need to be https, a secure SIP service is required for encrypted voice services.

Where Unlinkability Is A Necessity

Creating unlinkable digital communications first requires understanding what our digital communications networks see, as was discussed in detail in the previous section. The next step is to evaluate exactly what level of unlinkability a given situation requires. Is it important that the source not be linked to a *particular* journalist, or is it dangerous for the source to be seen communicating with any journalist at all? Is having this person remain truly anonymous–meaning his or her physical and/or legal identity is unknown even to the reporter–an option, either technically or journalistically speaking? What are the physical, technical, legal, financial, and expertise constraints of both parties?

The next step is to consider the data streams that can be used to connect our digital activity to our real identities. Our Web browsers and even operating systems are littered with flecks of digital DNA; accounts that require logins–like email, social networks and online marketplaces–tether these digital identities to our physical identities through recognizable handles and financial details. Our network connection points are often linked to our physical selves as students, cable account holders, and employees.

Creating unlinkable communications, then, means creating an environment where these traces are either eliminated or obfuscated, from our operating systems up through our online accounts. Fortunately, some good, open-source tools exist that can be composed along with some solid strategies to make this feasible.

Unlinkable Web Browsing: Tails + Tor

Tails

The simplest way to make sure that your computer is "sanitized" of any identifiable digital traces is simply to do as a doctor would: Dispose of your operating system after every use. Though impractical for everyday tasks where you need to be able to regularly store, modify, and share files, this “discard after use” approach is exactly the system design of Tails–a Linux-based operating system that lives on a USB drive and can be run directly from that drive on any computer available. When the computer is restarted, Tails starts up only in the host machine’s random access memory (RAM) and deletes itself on shutdown. Because Tails is recreated from scratch every time it is started, it can’t leak identifying bits of digital debris when you connect to the Internet.

Ideally, you should not store files on your Tails USB drive, with the possible exception of your GPG key. The Tails documentation provides a good overview of what Tails + Tor can and cannot protect.

Tor

Unfortunately, the mechanics of the Internet *require* that a fair amount of identifying information be attached to our digital communications, simply in order to function. In some ways the most stubborn of these identifiers is our IP address, which maps to the physical location of our Internet connection. While it may be easy to imagine creating “throwaway” email accounts to use with a source, influencing IP addresses seems generally beyond our control.

Enter Tor, or “the onion router.” Best known for its well-used (and very usable) browser, Tor is a combination of special Internet nodes (“relays”) and software designed to effectively mask the IP address of your Internet traffic. It accomplishes this by wrapping each packet of your Web traffic–including its “to” and “from” metadata–in three successive layers of encryption; it then hands off these encryption-wrapped packets to its exclusive network.

Your packets then move across three nodes in the Tor network, each of which has the ability to “peel off” exactly one layer of encryption; by the time your data reaches the final “exit” node–whose IP address it will take on–it has been “unwrapped” to its original state and can make its way on the open Web just as it would with your regular Internet connection. The result is that, to outside observers, your Web traffic seems to be coming from an IP address (and therefore physical location) other than your own.

Tor is designed not only to protect your data from being traced by outside parties, but also from monitoring by the Tor network itself. Not only is your content encrypted, but each node only knows the location of the immediately previous node in the Tor communication chain. In addition to this, Tor cycles the set of relays your traffic uses approximately every 10 minutes, so that analyzing your pattern of Internet traffic for other identifying information is more difficult to do.

For more detail, see the Tor overview

While Tor is often described as an "anonymity" network, it offers a very particular *type* of anonymity: locational anonymity for your Web traffic, nothing more. While it does this very well, using Tor does not obfuscate *what* you are doing on the Web, only *where* you are doing it from. If you log-in to Facebook–or Google or Yahoo or Twitter or what have you–those services will still have records of your activity as they always would; it's just that the IP address they see won't match where you really are. Likewise, using Tor Browser doesn't automatically protect you from cookie-based tracking; you need to make sure that cookies are turned off.

Two final points about Tor: The addresses of Tor relays are not secret. In fact, having them known is part of what makes it possible to perform checkups that assure the network isn't compromised. This does mean, however, that anyone observing your Web traffic will know that you're using Tor, either because they see your traffic going *in* to a known Tor relay or coming *out* of a known exit node. There is nothing illegal about using Tor. In fact, the more that people use Tor for regular Web browsing, the better its obfuscation properties work. That said, depending on where you are, connecting to the Tor network may make your Web traffic stand out. If you think the network operator (or the state) may be watching your traffic, using Tor may not be a good idea. Remember, all security is situationally dependent; there is no substitute for knowing your context.

Second, Tor is a *low-latecy* (i.e. minimal-delay) network, meaning it passes your packets back and forth as quickly as possible. While this is part of what makes it a viable alternative to more mainstream browsers like Firefox and Chrome, it does mean that someone watching both your IP address and the correct exit node stands a good chance of being able to connect it back to you. Although your connection going into the Tor network is encrypted, if this encryption were penetrated, experiments have shown that over time someone watching both the entering and exiting streams of traffic could statistically connect the two. However, because this type of traffic analysis is at least legally restricted in many places this risk is most salient when you are on a private and/or state-controlled (e.g., company, university and some national) network.[50]

"Mix-nets" are *high-latency* networks that send out messages in batches, making it difficult to identify their destination.

Unlinkable Email: Keep Your Pseudonyms Isolated

Any time you communicate with a source via email or chat, you are both necessarily communicating either with a *pseudonym* which may or may not be *linkable* to your "real," or physical, identity. In some cases, it may be important that both you and your source use *unlinkable* pseudonyms; this will help protect your sources in the case that being known to communicate with a journalist (or with you specifically) may put them at risk. If this is not the case, however, it may be sufficient for your source to use an unlinkable email address or handle for a particular exchange.

In order for unlinkable email to work, the address or handle itself must be created in an unlinkable context (e.g. an email account that you create and access only via Tor Browser), and you must both be vigilant not to include identifying information in the account details, or share any information that might connect the account to real individuals or locations. This means not discussing anything personally identifying: physical location, local stores, workplace name, or friend or family connections. Remember that unless the contents of your chat or email are encrypted, this

Lantanya Sweeney's foundational work on identifiability demonstrated that 87% of Americans could be identified by a combination of zip code, gender and date of birth.

[50] Harvard Student Receives F For Tor Failure, Runa A. Sandvik. *Forbes*, 12/18/13

information could be accessed by third parties (e.g., your email provider or law enforcement) and used to connect these communications to your real identities.

Likewise, your unlinkable identities themselves cannot be exchanged via any linkable communication channel (e.g., unencrypted, linkable email or chat accounts). You *must* agree upon and exchange these identities by some means outside of the communication channel you wish to use. In-person exchanges are best, human networks (trusted mutual acquaintances), voice conversations (for an existing source), and physical mail exchanges are also reasonable options. You will have to judge which of these is the best approach for a given situation on a case-by-case basis.

Postal mail services are a viable option for many reasons: the physical & legal protections are better, as is the obfuscation - a great deal of postal mail still moves through the system each day.

Unlinkable Chat

Unlinkability in synchronous (real-time) communications like chat is easier than with asynchronous communications like email; it actually has the potential for *perfect forward secrecy*. A chat service used with an unlinkable handle that is accessed on a Tails computer and/or through Tor Browser is particularly robust. Any of the encrypted-chat applications mentioned above can be used in this environment, though ones like CryptoCat, which don't require saving any information to Tails, are preferable in these instances.

Unlinkable Text

Many encrypted text programs actually use your mobile phone's data connection to send messages, but most of them read your contact information and will (necessarily) show the phone number of the sender and recipient along with the message. While TextSecure supports perfect forward secrecy, chat is often a better option for unlinkable exchanges.

Unlinkable Voice

As with text, encrypted voice calls are actually carried over your data connection. Apps like OSTel use a number different than your regular phone number, as do apps like SilentCircle and Wickr. Though some reporters may use so-called "burner phones," obtaining and using any kind of mobile device in such a way that it cannot be connected to your identity is nearly impossible. Stick with computer-based voice calls or go back to good old-fashioned postal mail.

Protecting Your Keys and Accounts

Creating unlinkable email accounts, verifying pseudonyms, and using encryption to protect your communications doesn't do any good if your accounts get hacked or someone gets access to your private key, which is why the heart of all good security practice is the use of strong, unique passwords on all your accounts and on the software that protects your private key.

On its surface, this recommendation seems simple enough, and it's advice that most of us have heard before. But in practice, it presents two primary problems. First, what is a "strong" password? Where guidance is offered (often along the lines of "must include at least one number and one uppercase letter"), the results are often difficult to remember and the actual "strength" is not clear. Meanwhile, sites that grade the quality of the passwords we create are often opaque about *how* to create good ones. In this scenario, we're lucky to remember what we typed by the time our entry receives a strong grade.

Fortunately, these issues have both digital and analog solutions, and you can mix and match the strategies according to your preference. Since at least some situations, however, will require that you remember your password, we'll start with simple recommendations for creating strong and memorable passwords.

Forget Pass*word*; Remember Pass*phrase*

The strength of a password comes is determined by two attributes: its length and its complexity. Length, naturally, is the number of characters; complexity relates to the randomness of those characters, in terms of both order and type (e.g., punctuation, numbers, and uppercase characters). Increase the strength of either (or both) characteristics of your passwords enough, and you have a set of characters that would take some number of centuries for a computer to figure out.

The trouble is that increasing complexity is both hard to do well–common number-for-letter substitutions are well known by hackers–and makes passwords hard to remember. Increasing length, on the other hand, is very straightforward when you think in terms of *phrases* instead of words.

Take, for example, your favorite quote from a television show or movie. This is likely to be both pretty long (in terms of characters) and you already have it memorized. As long its source isn't associated with any of your digital profiles and it isn't a well-known "catch phrase," it's probably a pretty good choice. Even better, pick a phrase or quotation from your "guilty pleasure" canon–movies, television shows or songs you don't even like to admit you enjoy, so that even someone close to you might not think to guess it. Longer phrases and irregular capitalization help improve the strength of the password, meaning it will be pretty tough to hack programmatically, as well as hard to guess.[51]

Password Generators & Managers

There's obviously no way for someone to steal a passphrase that only lives in your memory, but trying to remember the dozen or more we may need for all of our various devices and accounts can still be overwhelming–leading to the risky temptation to use the same password in multiple places.

A great way to deal with this is to use a password man-

[51] zxcvbn: realistic password strength estimation, Dan Wheeler. *DropBox*, 4/10/12

ager, like KeePass or LastPass, which stores your passwords (and can also create them for you) in an encrypted file that you unlock with a master passphrase. KeePass is open source and can be stored on a USB key; LastPass is cloud-based but crucially only stores the encrypted file, so the service doesn't have access to any of your passwords.

In some cases, even writing a passphrase down on a piece of paper can be suitable, as long as it's kept in a both legally and physically secure location, like a locked drawer in your desk at home. Obviously this limits your access to them, but as long as the place it's kept is yours (so it can't be searched without a warrant) and it's not just lying around, this is still a better approach than using weak passwords. Just make sure never to carry it with you!

Protecting Your Devices & Stored Data

Generally device and data protection is achieved by thoughtfully combining three techniques: encryption, obfuscation and deletion. Even in the case of email this triumvirate applies. Encrypt your content with GPG, obfuscate your location with the Tor Browser Bundle, and securely delete the traces left by your regular activities by using a disposable operating system like Tails.

Don't forget to regularly download, backup and delete your email.

Encrypt Your Devices

Encrypting your various devices is probably the easiest part. FileVault comes bundled with Macs, and TrueCrypt is free and cross-platform. Both Android devices and iPhones have built-in encryption options and can be set up in under five minutes.

TrueCrypt development ended in May, 2014. An audit of the code is underway, however, and community support may continue.

Encrypt and/or Obfuscate Groups of Files

Even beyond email, you want to make sure you encrypt and obfuscate sensitive data stored on your computer, in case you are compelled to decrypt it. TrueCrypt is especially good for this; you can use it to encrypt files on your computer and then rename the TrueCrypt file to look like something else entirely, like a movie file. Just make sure that the size of the file makes sense for the "cover" you give it–a 500 MB .doc file is pretty unlikely. While similar options don't currently exist for most phones, there are research and development efforts in this area.[52]

If you are traveling, consider whether you need to keep sensitive data on your computer at all. In places with good Internet access, services like Martus can encrypt your files locally and store them remotely; you don't even have to keep the key with you if you don't need access to them until you get home. Likewise, if your encrypted files are small enough, you can even transfer them to a USB drive and hide it somewhere that it's unlikely to be found in a search.

Don't Forget to Delete!

Moving files to the "trash" on your computer–and even "emptying" it–doesn't actually do anything to destroy those files; it simply lets the computer know that the memory they are using can be overwritten if needed. To really destroy the contents of a file, you need to make sure it's overwritten, ideally a few times. Tools like CCleaner will let you do exactly that.

Metadata: More Than Just Communications

Sending data over the Web isn't the only way our information gets tagged with metadata. Many cameras and camera phones attach data to photographs about the time and location of a particular photograph, as well as the equipment on which it was shot. This can be incredibly helpful for photo editors, but dangerous for journalists and sources whom it can be used to locate. Tools like InformaCam encrypts the metadata about photos on your phone, while ObscuraCam lets you blur out anyone (or anything) you need to. While

[52] CleanOS: Limiting Mobile Data Exposure, Yang Tang et al. *usenix;login; (38)(1)*, 2012; Blackphone, The Smartphone That Simplifies Privacy, Natasha Lomas. *Techcrunch*, 2/26/14

computer-based software to eliminate this metadata exists, a quick-fix for posting images to the Web is simply to screengrab them and use the capture instead of the original.

Documents, too, tend to contain metadata about who created them and when (often drawn from your computer username or software registration information), so avoid sharing text in formats that can contain "macros" (e.g., .doc/x or .pdf) where possible. Likewise, do some cleanup before posting any original documents given to you by a source. Some organizations have been known to use digital "watermarking" on sensitive documents to trace leakers. This can take many forms, but can be as subtle as giving each copy of the document slightly different typos and formatting, so that if it appears somewhere the source can be identified. Always discuss the implications with your editor, but consider formatting and spellchecking any source documents before posting to eliminate these patterns.

"Macro" is another term for a small program.

Analog Extras

Finally, a good deal of security comes from taking simple, physical precautions. Get a privacy screen for your laptop, and keep an eye out for "shoulder surfers". Put a Post-it over your laptop and phone cameras when you're not actively using them, and never let your devices out of your sight or loan them to anyone. The one exception to this may be in cases where you are meeting a source in person: Rather than turning your phone off, leave it at home or at work.

Looking Ahead

Source Protection and More

The primary objective of this paper has been to provide an overview of the legal and technical infrastructure that shapes the practical requirements for source protection in the age of digital communications. In the preceding section, I presented some tools and approaches for protecting source communications in the context of the two primary use cases for journalists: *linkable* and *unlinkable* sources. While the above discussion is not exhaustive, evaluating source communications within a framework of *linkability* does offer a valuable mechanism for designing digital communication solutions for journalists.

Integrating the strategies discussed above into the current workflow of journalism is no small task, and the suggestion to do so is not made lightly. I do believe, however, that this integration is both necessary and feasible, and will outline some possible approaches to it below. Before elaborating on these further, however, it is worthwhile to consider some of the additional advantages that these source-protection methods offer to journalists and their organizations.

Verification Protects Your Reputation

Verification–the process of confirming one's ownership of a particular digital identity–is an essential aspect of encrypted digital communications. But verification has an additional value for journalists–the ability to protect their reputations in the case that one or more of their digital identities (e.g., email, Twitter, or Facebook accounts) are compromised. Digital signatures, for example, can be attached to emails whether they are encrypted or not, and, for all practical purposes these signatures cannot be forged. If an account is hacked remotely, the attacker will not be able to replicate the digital signature, immediately tipping off any recipients that the communication is not genuine. Likewise, one can message essential contacts from a new account and they will be able to confirm it is really "you" by checking the signature.

Asymmetric Encryption Supports *True* Anonymity

Conversely, verification also supports the possibility of working with *truly* anonymous sources. Over time, the handles and email accounts used by such a source may change, but as long as the digital signature remains consistent, one can be confident that the person (or, in some cases, organization) on the other end of the exchange is the same. Moreover, a source wishing to make secure contact can do so by encrypting a message with the journalist's public key. This message could be an email or even a file that the journalist is directed to by another means. Whether sent from a throwaway email address or posted anonymously, that information will only be accessible to the journalist for whom it was intended.

On an organizational level, this can be achieved by implementing a tool like SecureDrop

Digital Security: A Journalistic Essential for the 21st Century

> "It should be clear that unencrypted journalist-source communication is unforgivably reckless."
>
> –Edward Snowden[53]

As Snowden, Greenwald, and Poitras' experiences demonstrate, adopting robust digital security measures has largely been the project of individual journalists and personal net-

[53] Edward Snowden Speaks to Peter Maass, Peter Maass. *The New York Times*, 8/8/13

works. Even in large organizations, these practices are often confined to a handful of individuals who have preexisting technical knowledge or work with sensitive sources or material.

This situation is inadequate for a number of reasons. At the individual and organizational level, ad-hoc source protection is incompatible with sensitive–and sometimes essential–reporting. Few organizations could afford the kind of protracted legal battle James Risen has faced, yet the alternative is either to abandon this coverage altogether or allow sources to wither under an atmosphere of increasingly aggressive leak prosecutions. Moreover, those doing sensitive reporting are not the only targets for attack; for an adversary seeking to infiltrate a news organization's network–as experienced by Bloomberg, the *The New York Times*, the *Wall Street Journal* and the *Washington Post*–any point of entry will do. In this sense and others, digital source protection is a "herd" protection. It works best–and, arguably, only truly works at all–if everyone is doing it.

Implementing Digital Security At Scale

Integrating digital security practices into a newsroom workflow will require concerted organizational effort. Editors will need to recognize the importance of digital security in the reporting process, and honor the scheduling implications it may entail. Reporters will need training and technical support for new systems and software. At the highest levels, implementing robust digital security practices will require an institutional commitment both financially and culturally, wherein noncompliance is not tolerated and privacy and security considerations sit at the table alongside legal and editorial ones.

While not insurmountable, the complexities of operationalizing digital security in a newsroom are certainly real. Likewise, it is unrealistic to expect every journalist to become expert in all possible nuances of digital security practice. Yet here is another sense in which digital security is not *sui generis*. In its technical complexity and importance to the industry, digital security is much like libel. Just as news organizations expect that reporters and editors have a basic literacy in libel considerations, so too should they assess and/or train their reporters and editors in the essentials of digital security. And just as news organizations retain experts to consult on particularly thorny libel issues, they should have experts on hand to advise on sensitive digital security measures where needed; some organizations already do.[54]

At a time when most news organizations are overstretched and underfunded, the very idea of adding yet another step to the reporting process or another employee to the payroll may seem laughable, even cruel. Yet, as is often the case, with new approaches also come new opportunities. Some of these are financial, as I will discuss below. No less significant, however, are the editorial opportunities. Reporters well-versed in digital security practices will find themselves with a new array of skills for locating and verifying sources and stories. They may even find themselves with new sources altogether as their networks learn of their better security practices.

Challenges and Opportunities in Digital Security Technologies

> "'Encryption works,' said Snowden. The problem, in real life, is nothing that *runs* encryption works."
>
> –Quinn Norton

The acknowledged usability of much digital-security software at the moment could probably be best summarized as "abysmal."[55] While less generally true of mobile-based offerings–such as SilentCircle, Wickr, TextSecure etc.–the difficulty of creating keys with GPG or installing Tails on a USB drive is still prohibitive to most users. Setting up each

[54] Meet the Man Hired to Make Sure the Snowden Docs Aren't Hacked, Lorenzo Franceschi-Bicchierai. *Mashable*, 5/27/14

[55] Why King George III Can Encrypt, Arvind Narayanan. *Center for Information Technology Policy*, 6/10/14

of these requires several steps and, often, some amount of risk to the user's computer if something goes wrong.

There are several reasons for this overall lack of usability. First and foremost, many secure-communications projects lack the steady revenue stream and well-defined user-base that are prerequisite for effective usability testing and development. Many of the tools recommended here are dependent on periodic government grants that do not prioritize usability, and as the recent "Heartbleed" bug illustrated, the level of donations to even widely used security libraries is often abysmally low.[56] The result is that many of these tools are materially unusable, unstable, or unreliable– and they are liable to disappear entirely if their funding is not renewed.

Yet it is these very challenges that also offer an opportunity for news organizations to diversify their revenue streams. By partnering with existing projects or building their own, large news companies can invest in the development of secure digital tools specifically designed to meet the needs of journalists and then sell or license those solutions to other organizations.

To be clear, my suggestion is not that journalistic organizations begin developing the kind of "black box" software produced by some commercial security vendors; transparency is as essential to software development as it is to journalism.[57] As in journalism, transparency in digital-security software is both an ethical and a practical concern: either you must be able to see the code, trust the community to validate the code, or trust the person who wrote it. If you don't know the code, you don't know what it does.

Being able to "see the code," however, does *not* require that the code also be *free.* Though many of the security projects listed above are both free and open source, alternative economic models are possible. "Source available" software,[58] for example, makes the code available for review and even reuse in other open-source projects, but commercial use or distribution requires a license fee; the widely used

[56] Tech giants finally agree to fund OpenSSL, Jon Brodkin. *Arstechnica*, 4/24/14

[57] "Be transparent" as a guiding journalism principle, Tom Rosenstiel. *Poynter*, 9/16/13

[58] Source Available vs Open Source vs Free Software, Phil Haack. 7/26/06

MySQL employs this type of dual-licensing. Alternatively, a service-and-support model, similar to that offered by RedHat is another possibility for generating revenue.

Engaging the Legal System

> "[T]here are no comparable Federal statutory standards to protect the privacy and security of communications transmitted by new noncommon carrier communications services or new forms of telecommunications and computer technology. This is so, even though American citizens and American businesses are using these new forms of technology in lieu of, or side-by-side with, first class mail and common carrier telephone services...
>
> Most importantly, the law must advance with the technology to ensure the continued vitality of the Fourth Amendment. Privacy cannot be left to depend solely on physical protection, or it will gradually erode as technology advances. Congress must act to protect the privacy of our citizens. If we do not, we will promote the gradual erosion of this precious right."
>
> –cited in *"The New Privacy Interest"*[59]

An ongoing issue in the privacy and security of digital communications is the increasingly revealing nature of metadata and the sheer volume of data that is observable by third parties,[60] especially as location-aware mobile and embedded technologies become more prevalent. Beyond our duty to investigate, report on, and explain the implications of these issues, the journalism industry should evaluate its relationship to efforts like Digital Due Process. In our own work, we must continue to be rigorous in protecting the privacy of our sources and readers, not just in reporting but in publication.

[59] The New Privacy Interest: Electronic Mail, Steven Winters. *Berkeley Technology Law Journal (8)*, 1993

[60] Ars tests Internet surveillance, Sean Gallagher. *Arstechnica*, 6/10/14

Conclusion

> "The only successful, robust way to address problems that involve personal responsibility and behavior is with social rather than technological tools. If we instead try and restrict behavior technologically... the only result will be an arms race that nobody wins."
>
> –Dr. Greg Jackson[61]

The legal and tehnical infrastructure of today's digital communication systems has enormous implications for journalistic source protection, and keeping our industry on stable footing given the ever-shifting ground at the intersection of these two fields will require significant education, organization and innovation from the journalistic profession as whole.

The recommendations presented here can neither be outsourced nor implemented overnight. But by pursuing the education of our colleagues, the coordination of our institutions, and collabortion with the digital security field, we can provide better protection to our sources and ourselves. Only by doing this can we protect our ability to hold power accountable, serve the public, and honor in practice the spirit of existing shield protections.

Perhaps even more than this, the journalism industry can help support those working for the freedom of the press all over the world by using and innovating around secure, usable digital communication tools. Only through long-term partnerships and/or the support of a recognized market can truly sustainable solutions be developed; journalistic organizations can and should be that market, for their own sake as well as the public's. The collective force and voice of our industry can help spur not only the technical but the legal changes that will help all users recapture their fundamental rights to privacy, freedom of association, and freedom of expression. If we are not willing to vigorously and emphatically fight for these rights, we may soon find that we no longer recognize ourselves as either a profession, or a nation.

[61] AT&T willing to spy for NSA, MPAA, and RIAA, Nate Anderson. *Arstechnica*, 6/13/07

Bibliography

Social media and the search for the Boston bombing suspects. *CBS News*, April 24, 2013.

NSA paid millions to cover Prism compliance costs for tech companies. *The Guardian*, August 22, 2013.

Justices reject reporter's bid to protect source. *The Associated Press*, June 2, 2014.

Google says government requests 'up 120%' in four years. *BBC*, March 28, 2014.

Snapchat Settles FTC Charges That Promises of Disappearing Messages Were False. *Federal Trade Commission*, May 8, 2014.

New York v. Harris. *Electronic Frontier Foundation*, Retrieved: June 5, 2014.

NPR Ethics Handbook: Independence, Retrieved: June 5, 2014.

Risen v. United States. *SCOTUSblog*, Retrieved: June 5, 2014.

Nate Anderson. AT&T engineer: NSA built secret rooms in our facilities. *Arstechnica*, April 12, 2006.

Nate Anderson. AT&T willing to spy for NSA, MPAA, and RIAA. *Arstechnica*, June 13, 2007.

Julia Angwin. Privacy Tools: Encrypt What You Can. *ProPublica*, May 6, 2014.

Anonymous. The Collateral Damage of Internet Censorship by DNS Injection. *ACM SIGCOMM Computer Communication Review*, 42(3), July, 2012.

Paul Baran. On Distributed Communications Networks. *IEEE Transactions of the Professional Technical Group on Communications Systems*, CS-12(1), March, 1964.

Emily Bazelon. Will Eric Holder Back Off? *Slate*, June 2, 2014.

Steven M. Bellovin, Renee M. Hutchins, Tony Jebara, and Sebastian Zimmeck. When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning. *8 New York University Journal of Law & Liberty 555*, 2014.

Chris Bishop. Untangling The Web: Key Exchange, 2008.

Justice Blackmun. Smith v. Maryland, June 20, 1979.

Justice Boggs, McKeague, and Keith. U.S. v Warshak, Sixth Circuit, December 14, 2010.

Judge Leonie Brinkema. U.S. v Sterling, Fourth Circuit. July 29, 2011.

Jon Brodkin. Tech giants, chastened by Heartbleed, finally agree to fund OpenSSL. *Arstechnica*, April 24, 2014.

Jon Brodkin. AT&T offers gigabit Internet discount in exchange for your Web history. *Arstechnica*, December 11, 2013.

Federal Communications Commission. Enhanced 9-1-1 Wireless Services, Retrieved: June 7, 2014.

Judge Davis, Traxler, and Agee. U.S. v Graham, Fourth Circuit, March 29, 2013.

Peter Eckersley. Sovereign Keys: A Proposal to Make HTTPS and Email More Secure. *Electronic Frontier Foundation*, November 18, 2011.

Lorenzo Franceschi-Bicchierai. Meet the Man Hired to Make Sure the Snowden Docs Aren't Hacked. *Mashable*, May 27, 2014.

Sean Gallagher. Ars tests Internet surveillance—by spying on an NPR reporter. *Arstechnica*, June 10, 2014.

Eva Galperin. In Turkey, folks are painting IP's of non-Turkish DNS servers onto the posters of the governing party., March 20, 2014.

Andy Greenberg. Whistleblowers Beware: Apps Like Whisper and Secret Will Rat You Out. *Wired*, May 14, 2014.

Phil Haack. Source Available vs Open Source vs Free Software. July 26, 2006.

Josh Halliday. London riots: BlackBerry to help police probe Messenger looting 'role'. *The Guardian*, August 8, 2011.

Alan Henry. Five Best VPN Service Providers. *LifeHacker*, March 23, 2014.

Alex Hern. Did your Adobe password leak? Now you and 150m others can check. *The Guardian*, November 7, 2013.

Parker Higgins. Pushing for Perfect Forward Secrecy, an Important Web Privacy Protection. *Electronic Frontier Foundation*, August 28, 2013.

Kashmir Hill. Lavabit's Ladar Levison: 'If You Knew What I Know About Email, You Might Not Use It'. *Forbes*, August 9, 2013.

Marcia Hoffman. Apple's Fingerprint ID May Mean You Can't 'Take the Fifth'. *Wired*, September 12, 2013.

Jacob Hoffman-Andrews. Forward Secrecy at Twitter. *Twitter*, November 22, 2013.

Lawrence Hurley. Supreme Court declines early look at NSA surveillance case. *Reuters*, April 7, 2014.

Justice John Marshall Harlan II. Katz v. United States, Concurring, December 18, 1967.

Richard R. John. *Spreading the News: The American Postal System from Franklin to Morse*. Harvard University Press, 2009.

Tiffany Kary. Twitter Turns Over Wall Street Protester Posts Under Seal. *Bloomberg*, September 14, 2012.

Bill Keller. Is Glenn Greenwald the Future of News? *The New York Times*, October 27, 2013.

John Kelly, Kevin A. Kepple, Jerry Mosemak, Janet Loehrke, and Jeff Dionise. Cellphone data spying: It's not just the NSA. *USA Today*, December 8, 2013.

Jennifer C. Kerr. AP President Pruitt accuses DOJ of rule violations in phone records case; source intimidation. *The Associated Press*, June 19, 2013.

Susan Landau. Making Sense from Snowden: What's Significant in the NSA Surveillance Revelations. *IEEE Computer and Reliability Societies*, July/August, 2013.

Judge Richard J. Leon. Klayman v. Obama, D.C. Circuit, December 16, 2013.

LLP Levine Sullivan Koch & Schulz. Amicus Brief, *Risen v United States*. March 26, 2014.

Paul Lewis and Tim Newburn. The Reading the Riots project: our methodology explained. *The Guardian*, December 4, 2011.

Adam Liptak. A High-Tech War on Leaks. *The New York Times*, February 11, 2012.

Natasha Lomas. A Closer Look At Blackphone, The Android Smartphone That Simplifies Privacy. *TechCrunch*, February 26, 2014.

Artur Janc Lukasz Olejnik, Claude Castelluccia. Why Johnny Can't Browse in Peace: On the Uniqueness of Web Browsing History Patterns. *5th Workshop on Hot Topics in Privacy Enhancing Technologies*, July 13, 2012.

Peter Maass. How Laura Poitras Helped Snowden Spill His Secrets. *The New York Times*, August 18, 2013.

Peter Maass. Q. & A.: Edward Snowden Speaks to Peter Maass. *The New York Times*, August 8, 2013.

Ann E. Marimow. Justice Department's scrutiny of Fox News reporter James Rosen in leak case draws fire. *The Washington Post*, May 20, 2013.

Mike Masnick. Judge Says Giving Up Your Password May Be A 5th Amendment Violation. *Techdirt*, April 25, 2013.

Declan McCullagh. Yahoo, ICQ chats still vulnerable to government snoops. *CNET*, February 28, 2014.

Susan McGregor. AP phone records seizure reveals telecom's risks for journalists. *Columbia Journalism Review*, May 15, 2013.

Steven J. Murdoch and George Danezis. Low-Cost Traffic Analysis of Tor. *IEEE Symposium on Security and Privacy*, 2005.

Ellen Nakashima. Agencies collected data on Americans' cellphone use in thousands of 'tower dumps'. *The Washington Post*, December 9, 2013.

Arvind Narayanan. Why King George III Can Encrypt. *Center for Information Technology Policy*, June 10, 2014.

Quinn Norton. Crypto for the Masses: Here's How You Can Resist the NSA. *The Daily Beast*, May 12, 2014.

Department of Justice. Report on Review of News Media Policies. July 12, 2013.

Ed O'Keefe. Transcript: Dianne Feinstein, Saxby Chambliss explain, defend NSA phone records program. *The Washington Post*, June 6, 2013.

Julian Oliver. Method for extraction and presentation of image content from captured wireless traffic. April 23, 2014.

Nilay Patel. Twitter fights back against subpoena of Occupy protester's tweets. *The Verge*, May 8, 2012.

Jennifer Peltz. Twitter Must Hand Over Protester Malcolm Harris' Tweets, Judge Rules. *The Huffington Post*, July 2, 2012.

Nicole Perlroth. Hackers in China Attacked The Times for Last 4 Months. *The New York Times*, January 30, 2013.

Andrea Peterson. NSA uses Google cookies to pinpoint targets for hacking. *The Washington Post*, December 12, 2013.

Kevin Poulsen. Edward Snowden's E-Mail Provider Defied FBI Demands to Turn Over Crypto Keys, Documents Show. *Wired*, October 2, 2013.

Justice Powell. United States v. Miller, April 21, 1976.

Justice Powell. Branzburg v. Hayes, Concurring, June 29, 1972.

Tom Rosenstiel. Why 'be transparent' has replaced 'act independently' as a guiding journalism principle. *Poynter*, September 16, 2013.

Kathleen Ann Ruane. Journalists' Privilege: Overview of the Law and Legislation in Recent Congresses. *Congressional Research Service*, January 19, 2011.

Philippe Sands. No Place to Hide: Edward Snowden, the NSA and the Surveillance State by Glenn Greenwald – review. *The Guardian*, May 23, 2014.

Runa A. Sandvik. Harvard Student Receives F For Tor Failure While Sending 'Anonymous' Bomb Threat. *Forbes*, December 18, 2013.

Charlie Savage. Judge Questions Legality of N.S.A. Phone Records. *The New York Times*, December 16, 2013.

Charlie Savage. Holder Tightens Rules on Getting Reporters' Data. *The New York Times*, July 12, 2013.

Stephan Somogyi. Making end-to-end encryption easier to use. *Google*, June 3, 2014.

Justice Stewart. Katz v. United States, Opinion of the Court, 1967.

Darlene Storm. 42 million unencrypted passwords leaked from hacked online dating site Cupid Media. *Computer World*, November 20, 2013.

Margaret Sullivan. The Disconnect on Anonymous Sources. *The New York Times*, October 12, 2013.

Latanya Sweeney. k-Anonymity: A Model for Protecting Privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), 2002.

Yang Tang, Phillip Ames, Sravan Bhamidipati, Ashish Bijlani, Roxana Geambasu, and Nikhil Sarda. CleanOS: Limiting Mobile Data Exposure with Idle Eviction. *usenix;login:*, 38(1), 2012.

Guy Taylor. Armed agents seize records of reporter, Washington Times prepares legal action. *The Washington Times*, October 25, 2013.

Christina Tsuei and Paul Antonson. How Advertisers Use Internet Cookies to Track You. *The Wall Street Journal*, July 30, 2010.

Edward R. Tufte. *Beautiful Evidence*. Graphics Press, LLC, first edition, May 2006. ISBN 0-9613921-7-7.

Liam Tung. Yahoo finally enables HTTPS encryption for email by default. *ZDNet*, January 8, 2014.

American Civil Liberties Union. ACLU v. Clapper – Challenge to NSA Mass Call-Tracking Program. Retrieved: June 6, 2014a.

American Civil Liberties Union. ACLU v. Clapper - Legal Documents. Retrieved: June 6, 2014b.

Jennifer Valentine-DeVries. Sealed Court Files Obscure Rise in Electronic Surveillance. *The Wall Street Journal*, June 2, 2014.

Jonathan Watts. David Miranda: 'They said I would be put in jail if I didn't co-operate'. *The Guardian*, August 19, 2013.

Dan Wheeler. zxcvbn: realistic password strength estimation. *DropBox*, April 10, 2012.

Justice White. Branzburg v. Hayes, Opinion of The Court, June 29, 1972.

Steven Winters. The New Privacy Interest: Electronic Mail in the Workplace. *Berkeley Technology Law Journal*, 8, 1993.