

Usable Cryptographic Provenance: A Proactive Complement to Fact-Checking for Mitigating Misinformation

Emily Sidnam-Mauch¹, Bernat Ivancsics², Ayana Monroe^{1,4}, Eve Washington², Errol Francis II¹, Kelly Caine¹, Joseph Bonneau³, Susan E. McGregor²

¹ Clemson University, ² Columbia University, ³ New York University, ⁴ University of North Carolina at Chapel Hill
esidnam@clemson.edu, bi2155@columbia.edu, ayanam@live.unc.edu, esw2175@columbia.edu, errolf@clemson.edu, caine@clemson.edu, jcb@cs.nyu.edu, sem2196@columbia.edu

Abstract

This paper describes how cryptographic provenance can serve as a proactive, partial solution for mitigating misinformation. Drawing on literature from human-centered computing and usable security, journalism, and cryptography, we discuss the advantages and limitations of both content-based and technical approaches to the problem of online misinformation. We argue cryptographic provenance systems designed for usability can reduce the spread of misinformation by surfacing provenance information and making this information salient and acceptable to information consumers. We highlight challenges and open research areas related to designing usable cryptographic provenance systems, specifically concerning two key stakeholder groups: journalists and news consumers.

Introduction

“Falsehood flies; the truth comes limping after it.” Jonathan Swift’s version (Swift and Roscoe 1841) of the well-known idiom highlights what has been observed by modern research: false news travels faster and farther than the truth, particularly online (Vosoughi, Roy, and Aral 2018). The growing prevalence and harms resulting from misinformation have drawn substantial public and research attention in recent years. From the 2016 US presidential election to the COVID-19 pandemic, the negative impacts of misinformation have ranged from election interference to exacerbating public health crises. In response, there have been numerous efforts to “combat” digital misinformation, but none has yet proved a significant or durable success.

Unlike disinformation, which constitutes false and misleading information created and circulated with the intent to cause harm, researchers characterize *misinformation* as false or misleading content circulated by actors who lack understanding that the information shared is spurious (SciLine 2021). Understandably, many applied misinformation interventions have focused on visually flagging inaccurate, fabricated, or misleading messages in hopes this will stop recipients from unintentionally propagating false content. While the fact-checking and debunking efforts behind such “warning labels” can help mitigate misinformation spread when effectively designed (Lewandowsky et al. 2020), they repre-

sent an inherently reactive approach to misinformation that cannot effectively scale to the magnitude of the problem.

In this paper, we argue that cryptographic provenance systems are a vital component of a misinformation-resistant online ecosystem. We use the term *cryptographic provenance system* to refer to a system that combines cryptographic protocols with a usable interface to 1) confirm the provenance of digital content by generating a secure, unalterable record of that content and any alterations, and 2) surface this information in a meaningful manner to end users. Cryptographic protocols provide a secure infrastructure for transparency in online publishing; in turn, this transparency lays a foundation for accountability and trust-building in digital publishing. At the same time, we note that cryptographic authentication and provenance tools often struggle to be salient and acceptable to end users. To be effective, designers of cryptographic provenance systems must partner with key stakeholders and design for usability. Usable cryptographic provenance systems provide a robust digital infrastructure for transparency, accountability, and trust and serve as a scalable, proactive bulwark against misinformation.

Fact-Checking Fails to Address Key Drivers of Misinformation Spread

Many recent efforts to reduce the spread and impact of misinformation focus on complicating the apparent truthfulness of messages—typically through visual indicators backed by some form of flagging and/or fact-checking process (Walter et al. 2020). For example, Twitter’s “Birdwatch” program aims to add crowdsourced “notes” to potentially misleading tweets (Coleman 2021). This program is the latest in a succession of related efforts (e.g., Roth and Pickles 2020; Ortutay 2021). Anita Butler, a Twitter designer who studied responses to potential “disputed” labels, summarized the mixed effectiveness of these tools across users, saying, “People were like, well, who’s disputing it?” (Ortutay 2021).

This observation aligns with a recent meta-analysis of fact-checking initiatives, which found that the degree to which such initiatives reduced a belief in or the likelihood of sharing false or misleading information was substantially attenuated by multiple factors (Walter et al. 2020). In general, information consumption and sharing is driven not just by a desire to be accurately informed and inform others, but

also to entertain, enhance social relationships, and/or provide an escape from the concerns of everyday life (Ruggero 2000). Past research identifies a wide range of social and psychological factors that predict misinformation sharing and spread, including individuals' emotional states (Wischniewski, Krämer, and McNamara 2021), analytic thinking habits (Pennycook and Rand 2019), social network composition (Young et al. 2021), perceptions of information novelty (Vosoughi, Roy, and Aral 2018), and level of trust in the information source (Sterrett et al. 2019). When presented with identical information, lay assessments of digital news content's credibility and truthfulness differ based on individual differences, and even expert evaluations vary according to discipline (Bhuiyan et al. 2020). In addition to more stable individual differences that influence the likelihood of spreading misinformation, an individual's emotional state also affects how they evaluate the factuality of information at a given moment (Tiedens and Linton 2001).

Ultimately, truth is an abstract concept that is difficult to ensure technically. Moreover, perceptions of message accuracy, factuality, or honesty are only one of many factors impacting the dissemination of misinformation. Due to the complexities underlying individuals' information evaluations, sharing decisions, and abilities to influence further spread of misinformation, interventions that focus solely on fact-checking offer limited efficacy for combating misinformation spread. Additionally, it is unlikely that any reactive approach to mitigating misinformation (i.e., via labeling) can keep pace with the exponential spread of online falsehoods (Vosoughi, Roy, and Aral 2018). It is necessary to develop proactive, scalable interventions that target multiple drivers of misinformation creation and spread.

Cryptographic Provenance Systems Are a Needed Misinformation Intervention

Cryptographic methods of authenticating specific attributes of digital messages represent a *proactive* approach to mitigating misinformation. Rather than attempting to arbitrate "truthiness," these technical approaches broadly seek to build *trust* in information sources and facilitate accountability through structural mechanisms. This section examines the contributions that cryptographic tools can make in the space of misinformation interventions.

Cryptographic Protocols Can Make Digital News More Transparent and Secure

In order to understand how cryptography can be applied to mitigate misinformation spread, it is imperative to recognize how provenance tools work and how they can be applied to digital news publishing. Cryptographic provenance tools produce a globally consistent log that permanently records *events* emitted by some *authority* (e.g., a news article is published), enabling third parties to monitor the log to verify the behavior of the authority (Chase and Meiklejohn 2016).

Several cryptographic protocols for provenance and transparency have been developed over the last decade. The most prominent, successfully deployed system is Certificate Transparency (CT), launched in 2013 (Laurie, Langley, and

Käser 2013; Laurie 2014). CT aims to improve public trust in certificate authorities (CAs), who issue x.509 certificates for use by TLS servers on the web. All issued certificates are permanently recorded in a public CT log which anybody can monitor, ensuring anyone can detect improperly issued certificates. Unlike many previous proposals for increasing trust in CAs, CT does not directly prevent misbehavior by CAs (which lacks a precise technical definition); CT only ensures a CA's behavior is observable and non-repudiable.

Following the successful deployment of CT, similar provenance and transparency systems have been proposed for monitoring other types of authorities, including user-key mappings in secure communication systems (Ryan 2014; Melara et al. 2015), cryptographic key usage (Yu, Ryan, and Cremers 2015), and distribution of software (Fahl et al. 2014; Nikitin et al. 2017; Al-Bassam and Meiklejohn 2018). Recent research proposes general-purpose transparency protocols (Meiklejohn et al. 2020; Chen et al. 2020; Tyagi et al. 2021), which can be adapted to any authority regularly publishing information and support updating that information after initial publication. Developers can directly adapt these tools to digital news publishers to create a transparent log of all news content published.

We argue that cryptographic provenance systems applied to digital publishing can and should provide four key assurances:

1. *Authenticating Provenance.* Cryptographic proofs offer persuasive evidence of publisher/broadcaster identity via digital signatures. While digital signatures can only guarantee that a particular public key signed content, we note that mapping public keys to real-world organizations or individuals (e.g., *The New York Times* or Walter Cronkite) requires a secondary layer of *public key infrastructure* (PKI). In practice, however, this challenge has been reasonably addressed via the PKIX system mapping public keys to URLs (e.g., *newyorktimes.com*). Users already rely upon URLs for authenticating the source of online information; therefore, this is a reasonable way to layer on an additional cryptographic assurance.
2. *Verifying Content is Unaltered.* Cryptography can provide strong evidence that information has not been modified or tampered with since the time of publication, as a timestamp is part of the generated signature. This verification is typically accomplished via collision-resistant hash functions. Moreover, authenticated data structures enable efficient *cryptographic commitments* to large amounts of information that parties can later verify. These techniques can be applied to verify that no changes have been made to a news article and authenticate embedded materials like photos and videos.
3. *Ensuring Users are Viewing the Same Content.* Cryptographic signatures can support consensus protocols that further establish a consistent global view of committed information, ensuring that all parties see the same information (Xiao et al. 2020). When applied to digital news, these protocols can make it possible to detect if a news publisher is showing two different versions of the same article to different users.

4. *Creating Unalterable Records of Changes.* Modern authenticated data structures support efficient versioning of information, enabling it to be updated and amended with short cryptographic proofs of the latest version of the information and all previous versions (Tyagi et al. 2021). This technique can be applied to create scalable tools that document every change to an article since it was published, creating an unalterable record.

Provenance systems that offer all the above-mentioned cryptographic assurances make digitally published news content more transparent and secure. Furthermore, when these cryptographic assurances are made salient and comprehensible to end users, cryptographic provenance systems can be powerful tools to combat misinformation.

Cryptographic Provenance Systems Proactively Counter Misinformation

Unlike defensive interventions like fact-checking, cryptographic provenance systems are proactive interventions for misinformation. By authenticating digital news artifacts, cryptographic systems refocus attention on provenance and transparency and foster improved trust between news consumers and digital news producers. Ultimately, these systems promote a misinformation-resistant online ecosystem by enabling structures that discourage dubious publishing practices, favor transparency, and build trust between news consumers and credible publishers.

By default, digital artifacts can be manipulated, updated, and removed without generating readily accessible evidence of those changes. Cryptographic tools facilitate a degree of accountability not currently endemic to digital messages. Provenance tools identify and record news artifacts and any subsequent changes on a publicly auditable database or cryptographically secure ledger. This documentation prevents retroactive modifications to the version history and can add a layer of confidence that the authentication is trustworthy. Leveraging cryptographic protocols to create a more secure news environment discourages malicious, questionable, and opaque publishing practices. Provenance systems can provide unalterable evidence if digital publishers engage in disreputable practices (e.g., ghost editing, tampering with embedded content). This discourages the creation and spread of manipulated content.

While cryptographic provenance systems make it easier for news consumers to identify questionable publishing practices and manipulated content, one of the major advantages of these systems is that they can help users recognize *credible* news publishers. For example, for an unaltered news article, effective cryptographic provenance systems will give users confidence that the article and any associated artifacts (e.g., embedded tweets) originated from the stated publishers, that there truly have been no changes to the information since its publication, and that all readers are viewing a consistent version. By providing infrastructure to generate positive indicators of news authenticity and transparency, cryptographic systems can encourage the identification and spread of credible information.

In turn, cryptographic provenance systems can establish

and reinforce public trust in news producers by enhancing transparency around news content's provenance and production history. For example, the unalterable record of changes afforded by such systems can corroborate journalists' transparency practices; readers can confirm that corrections acknowledged by a publisher match with the changes recorded in the ledger. Additionally, when cryptographic provenance systems are designed by a third party, it can reassure news consumers that the authentication indicators generated by the user interface are not biased by news organizations' potential conflicts of interest.

To reiterate, cryptographic provenance systems indiscriminately hold all publishers accountable for the content they alter or remove and make it easier for consumers to identify whether publishers' practices are transparent or opaque. These interventions target transparency, accountability, and trust as mechanisms for promoting the circulation of quality information. Moreover, the content-agnostic nature of cryptographic approaches allows them to scale seamlessly with the volume of content being produced. Unlike the many hours of human effort needed to fact-check even a single dubious claim, updates to the cryptographic record are immediately and automatically generated for every digital message produced or changed. Cryptographic provenance systems are promising as proactive misinformation interventions.

Cryptographic Provenance Systems for Digital News Are Technically Feasible

Existing technical solutions are already sufficient for building cryptographic provenance systems that offer meaningful transparency and accountability guarantees for digital information. For example, CT—which aims to ensure that SSL certificates are correctly and non-maliciously issued (Certificate Transparency Group)—has been successfully deployed at scale, with billions of certificates logged (Li et al. 2020) and built-in support in Chrome, Firefox and Safari browsers. With regard to news, although peer-reviewed research in this space is limited, we note that variations on the cryptographic approaches described are already being developed and tested in many research and industry settings. Notable examples of these efforts include the News Provenance Project (Koren 2019), Project Origin (Aythora et al. 2020), Project Starling (Takahashi 2021), the Credibility Coalition (Coalition 2017), Arweave (Project 2021; Library 2019), and the Content Authenticity Initiative (Initiative 2021). Details on these projects and their specific goals and approaches are shown in Table 1.

The prototypes developed by these projects prove it is possible to overcome the technical barriers to building cryptographic provenance solutions. However, from a human-centered computing perspective, it is necessary to consider the end users from the beginning of the design process, or else usability barriers will impede the system's effectiveness. The next section discusses the requirements, challenges, and opportunities of designing news provenance tools for journalists and news consumers.

| Project | Founders | Goal | Cryptographic Protocols |
|---------------------------------|--|---|--|
| News Provenance Project (NPP) | The New York Times' R&D lab and IBM Garage | To explore the relationship between cryptographic provenance and user trust with respect to digital images (Koren 2019). | Provenance authentication applied to digital images |
| Project Origin | New York Times, Canadian Broadcasting Corporation/ Radio-Canada, British Broadcasting Corporation (BBC), and Microsoft | To help publishers verify information that they sent to users. (This project combines the NPP, the BBC/CBC Provenance Project, and the Microsoft AMP system; Aythora et al. 2020.) | Provenance authentication applied to news articles, augmented with trust measures |
| Project Starling | USC Shoah Foundation and Stanford University's Department of Electrical Engineering | To develop an open-source, end-to-end framework for application developers to add image and document verification to the media generation process, using blockchain and distributed ledger technology (Labs 2022) | Provenance authentication and verification of tampering |
| Credibility Coalition | N/A | To develop common standards for information credibility by collaborating with a large team of partners from journalists, academics, policy-makers, and technologists to incubate projects and conduct research | N/A |
| Arweave | Minimum Spanning Technologies Limited | To solve the problem of 'link rot' and create what they call the 'permaweb: A global, permanent web of pages and applications that live forever' (Project 2021) | A protocol that allows for decentralized saving of websites to create a permanent record (Library 2019). |
| Content Authenticity Initiative | Adobe, Twitter, the New York Times | To counter the rise of misinformation by securely preserving provenance and attribution data for digital content, starting with photo and video content (Initiative 2021). | An end-to-end system for content provenance of digital content through open-source development, cross-industry collaboration, and interoperability of tools that can be integrated into a blockchain system (Content Authenticity Initiative). |

Table 1: Cryptographic Provenance Initiatives

Usability Challenges Pose the Greatest Barriers to Implementing Provenance Tools

Despite the technical feasibility of integrating cryptographic provenance into existing digital publishing pipelines, key challenges to ensuring their adoption and acceptance by both journalists and news consumers remain. In particular, there may be a mismatch between the degree of transparency journalists and news organizations deem worthwhile and what news consumers may prefer. Despite years of research on security indicators, improving the salience, usability, and acceptance of cryptographic evidence for end users remains an open challenge. In the following section, we detail core ar-

reas for consideration and begin to address areas for future research.

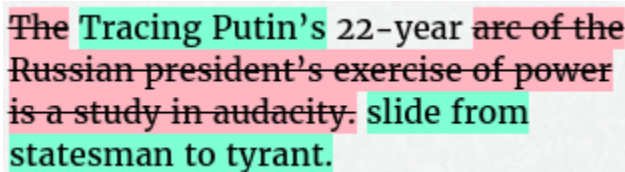
Attending to Journalists as Stakeholders Can Inform Effective Tool Design

Journalists currently seek to improve trust relationships with their readers through transparency practices (e.g., Moran 2021). Still, the impact of these efforts is questionable as news consumers may have diverging conceptualizations of transparency and can be wary of disclosures made by news organizations themselves. While news consumers are the targeted end users of cryptographic provenance systems, journalists are key stakeholders in these systems. Content au-

thentication facilitated by a secure, third-party system can allow journalists to corroborate their transparency initiatives and build trust with news consumers.

However, from an implementation perspective, some challenges need to be considered. One challenge is that, though journalists' content management systems (CMS) provide tools to annotate edits in the schema (e.g., adding a correction tag), journalists do not always take advantage of these opportunities. For example, the widely-used WordPress CMS supports version control, but it is unclear whether journalistic organizations widely use this feature. Cryptographic authentication of the back-end transparency practices of news organizations has the potential to provide positive indicators of trustworthiness. Still, more work is needed to understand the schema practices currently used by news organizations and encourage the utilization of schema that can be cryptographically verified to communicate transparency (Credibility Coalition and Nieman Foundation for Journalism).

Additionally, although formal academic programs exist at both the undergraduate and graduate levels, journalism in the United States retains a largely informal apprenticeship practice model. Many journalistic institutions operate without formalized (i.e., written) approaches to key practices, including issues like handling post-publication changes to news content. While some news organizations index their corrections in the article metadata (e.g., ProPublica n.d.; Press n.d.), news organizations likely limit their view of "corrections" to factual corrections. Not *every* change made to an article post-publication (e.g., typos or changes in framing) is considered a "correction". In contrast, the corrections suggested by readers include inaccurately listed dates, locations, or descriptions, as well as "a surprising variety of other flavors of mistakes" (Jacquette 2018). The @nyt.diff Twitter account (Editing TheGrayLady), which documents real-time changes to news published on the main page of *The New York Times*, also illustrates there is a mismatch between journalists' and readers' perceptions of the types of changes worth publicly acknowledging. The post shown in Figure 1 is an example of wording changes that are often not documented by publishers, but that readers want to be able to access.



The Tracing Putin's 22-year are of the Russian president's exercise of power is a study in audacity. slide from statesman to tyrant.

Figure 1: Sample post from the "Editing TheGrayLady" Twitter feed, describing a change in an article abstract.

The dissonance between readers' and publishers' conceptualizations of "corrections" creates a challenge for implementing a provenance system that does not inadvertently produce counterproductive effects on trust through the surfacing of corrections not flagged as corrections by the producer. Additionally, transparency information published

about content from reputable news sources may provide fodder for propagators of misinformation. Visibility of edits and corrections, for example, can be pointed to in an attempt to undermine faith in a news publisher's credibility, though this has long been accepted journalistic practice. Further exploration of the needs and practices of journalists as stakeholders is needed in order to implement usable provenance tools that encourage rather than penalize necessary updates to articles.

Designing for News Consumers' Needs Is Complicated but Vital

The considerable volume of research conducted on digital security indicators shows that the user experience of these indicators influences how users interact with them (Akhawe and Felt 2013; Egelman, Cranor, and Hong 2008; Felt et al. 2015). Like security indicators (e.g., SSL certificate warnings), news provenance tools aim to communicate the authenticity of cryptographically-verified digital artifacts. We argue that the difficulty of designing salient, interpretable, and, overall, *usable* indicators for end users is the most pressing obstacle to deploying effectual cryptographic provenance systems for digital news.

In particular, users' understanding of the meaning of security warnings impacts if and how they act upon those warnings (Felt et al. 2015). A study examining users' mental models of encryption found that users vary in their beliefs related to the level of security offered by encryption technologies, and users often misunderstand security indicators like browser warnings (Wu and Zappala 2018). Similarly, target users of news provenance tools may not understand what cryptographic authentication entails. While resources have been developed to teach principles of cryptography to a range of lay audiences (Bell et al. 2003), providing more detailed and accurate understandings of how these systems work does not necessarily increase user confidence (Wu and Zappala 2018). Instead, designers should "make efforts to align designs and communication efforts with the functional models users already possess" (Wu and Zappala 2018, p. 405). System developers should research users' mental models of the cryptographic provenance systems to navigate barriers to understanding authentication indicators. They can then use this research to develop user-informed explanations that effectively communicate the security of these systems and their significance for authenticating news content (Ngo and Krämer 2021; Wu and Zappala 2018).

As touched on in the section on journalists as stakeholders, a vital challenge for designing a usable cryptographic provenance system for news is determining the scope of transparency indicators that should be enabled through the system. While cryptographically documenting provenance and changes to articles is relatively straightforward, it is challenging to determine how much of this information should be presented to the user in the standard interface. Presenting too much information to users generally leaves them overwhelmed and unsure of which direction to take, which is especially detrimental when it comes to making trust decisions. Additionally, users have varying information needs and criteria for assessing news content (Bhuiyan

et al. 2020). This barrier can be navigated through usability studies that identify the optimal amount and presentation of transparency information for the average user and design choices that allow users to personalize further the information they want to see. Additionally, as observed by Pennycook et al. 2020, it is crucial to clarify how users should interpret the presence and absence of indicators, being mindful of any counterproductive effects of existing design elements.

Discussion

We outlined how cryptographic protocols can be applied to make digital news more transparent and secure, proactively mitigating misinformation harms by facilitating trust and accountability. We also addressed usability as the core challenge for implementing effective cryptographic provenance systems and provided considerations for designing for journalists and news consumers. In the following sections, we discuss the boundaries of cryptographic provenance systems as misinformation interventions, note limitations in current prototypes, and suggest paths forward for designing effective cryptographic provenance systems.

Noting Boundaries and Navigating Barriers

Just as the limits of fact-checking interventions should be acknowledged, it is important to note the limits and challenges related to cryptographic provenance measures as misinformation interventions. These include the bounded nature of the solution, challenges related to potentially differing notions of transparency, and ethical considerations. We suggest future research and design directions that can help navigate these challenges.

First, while the transparency afforded by cryptographic provenance systems lays a foundation for trust-building and accountability in digital publishing, it is important to recognize that transparency itself is not a panacea; also, transparency does not automatically increase trust (Ananny and Crawford 2018). Cryptographic provenance systems fill a critical gap in current misinformation intervention efforts through their potential to proactively discourage the creation and propagation of false or manipulated content. Even so, they are not meant to be a one-stop solution. Future research should assess how cryptographic provenance systems can best mobilize transparency to facilitate trust-building across diverse stakeholders. Additionally, cryptographic solutions should be integrated with and used to augment a variety of misinformation interventions. For example, existing tools that track changes to published information or that flag media that has likely been altered could be guaranteed to be reliable with cryptographic provenance.

Second, we note a specific barrier to communicating the advantages of these systems to end users. Namely, the transparency assurances provided by cryptographic systems may not overlap with colloquial understandings of what constitutes news transparency. For example, cryptographic transparency has a narrow technical definition (Chase and Meiklejohn 2016); cryptographic systems can guarantee that certain actions (e.g., issuing a certificate or publishing a news article) are permanently and publicly visible to all observers

but they make no *semantic* claims about the content. Cryptographic provenance systems do not provide assurances that news consumers might associate with the term “transparency”, such as disclosures of ideological biases, financial backing, or conflicts of interest relevant to an article. We argue that the narrow technical definition of cryptographic transparency (Chase and Meiklejohn 2016) is partly a strength in the digital publishing context. The provenance systems we have described do not attempt to assert that a news item is “true” versus “false” (e.g., Khodabakhsh, Busch, and Ramachandra 2018), nor do they directly identify dis- or misinformation (e.g., Kumar, West, and Leskovec 2016), thus sidestepping some of the semantic debates around fact-checking. Nevertheless, system designers will need to intentionally design the user-facing indicators in a manner that clearly communicates the purpose of cryptographic provenance systems, including what they can and cannot guarantee concerning digital news artifacts. We suggest that research on security indicators (e.g., Chase and Meiklejohn 2016; Akhawe and Felt 2013) and credibility indicators (e.g., Stomber et al. 2021; Sumpter and Neal 2021) provide a valuable foundation for informing the initial design of cryptographic provenance indicators and guiding usability research for these systems.

Third, as with any intervention, the design and implementation of cryptographic systems have ethical implications. Throughout the development and launch of cryptographic provenance systems, it is important that researchers and system developers attend to the ethicality of potential and actual outcomes relevant to misinformation and beyond, asking questions like: Who benefits from these systems and are these benefits distributed equitably across diverse groups of publishers and users? Is the user interface unbiased and transparent in how it surfaces information from the cryptographic ledger? How can large-scale provenance systems utilize environmentally sustainable blockchain technology? Proactively interrogating the ethical implications of these systems from multiple perspectives throughout their development will help these systems address existing problems without potentially creating new ones.

Paths Forward for Developing Cryptographic Provenance Systems

While we have noted promising initiatives that are currently prototyping cryptographic provenance tools for digitally published content, many existing projects apply only to select content types (see Table 1). For example, while the available documentation for efforts like Project Origin (Aythora et al. 2020) refer to the feasibility of applying their provenance indicators to diverse media formats, the primary focus of Project Origin—like the NPP (Koren 2019) upon which it draws—appears to be photographs; Project Starling (Labs 2022) and the Content Authenticity Initiative (Initiative 2021) extend this to video content. These tools can support provenance and verification against alteration for digital images and videos, and it is possible that some may provide information about multiple versions. For example, NPP provided readers with the multiple contexts and captions with which a given photo had been published. Still, it is not clear

that any of these projects are poised to address the types of text-based changes like those illustrated in Figure 1 that news consumers can deem meaningful. Generating a secure record of alterations to published text is important because there are instances where even small changes to framing and word choice can have monumental implications (e.g., Ecker et al. 2014; McBride 2020; Owen 2021). Though solutions like Arweave (Project 2021) could be used to *store* text-based content, it does not meet the usability requirements needed to scale to the speed, volume, and variety of platforms through which digital news is published.

While existing cryptographic provenance approaches for digital publishing are beginning to take shape across a range of initiatives, we argue it is vital to expand the types of assurances offered by these systems and to ensure the usability of these systems for a broad spectrum of news producers and consumers. In light of the limitations of existing solutions, we argue that cryptographic provenance systems for news need to be developed for *all* media formats. They should also support four important *technical* assurances. Cryptographic provenance systems for news should 1) cryptographically confirm the provenance of digital news artifacts, 2) verify articles and their embedded content have not been altered, 3) detect if a publisher is showing two different versions of the same article to other users, and 4) generate transparent records of all changes to news content from the initial point of publication. Just as importantly, these technical features must be paired with usable interfaces for news consumers that “meet them where they are.”

So, what might a system like this entail? At its most basic, cryptographic provenance for digital publishing would involve cryptographically signing items of digital content and committing the results to a public log. We believe that a more usable approach is likely to involve not just logging the hash and signature, but also publishing an inclusion proof alongside the content. A browser extension could then verify the inclusion proof and confirm to news consumers that the content has been signed and logged, without any additional user effort. Given the importance of enhancing public trust in news, we suggest that cryptographic provenance systems for news should also enumerate and surface the log history—and the expanded content at each step—in a usable format that would allow readers to easily view and evaluate *any* transformations within the published content from its initial publication until the present.

Conclusion

Interventions focused solely on fact-checking or verifying the “truth” of digital messages are insufficient to counter contemporary misinformation formats and difficult to scale. In contrast, cryptographic provenance systems can proactively mitigate misinformation harms through promoting *accountability* via technically-assured publisher transparency, thereby facilitating greater *trust* in participating publishers. Such systems can also readily scale to the volume of digitally published content.

As technical barriers to developing cryptographic provenance systems are low, usability barriers are the most pressing challenges that must be overcome to create productive

cryptographic provenance tools for digital news. To navigate these barriers, we find it imperative to understand key stakeholders’ needs and practices, including news producers and consumers. Designing for the varying needs of news consumers is complicated but vital. Efficacious provenance systems for digital news will apply multiple cryptographic solutions, design for usability, and complement a variety of misinformation interventions that target specific mechanisms of misinformation spread.

Acknowledgments

This research was supported by the National Science Foundation under award numbers 1940679, 1940670, and 1940713.

References

- Akhawe, D.; and Felt, A. P. 2013. Alice in Wonderland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *22nd USENIX Security Symposium (USENIX Security 13)*, 257–272. Washington, D.C.: USENIX Association. ISBN 978-1-931971-03-4. URL <https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/akhawe>.
- Al-Bassam, M.; and Meiklejohn, S. 2018. Contour: A Practical System for Binary Transparency. *DPM/CBT@ESORICS 11025*: 94–110.
- Ananny, M.; and Crawford, K. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20(3): 973–989.
- Aythora, J.; Burke-Agüero, R.; Chamayou, A.; S Clebsch, M. C.; Earnshaw, N.; Ellis, L.; England, P.; Fournet, C.; Gaylor, M.; Halford, C.; Horvitz, E.; Jenks, A.; Kane, K.; Lavallee, M.; Lowenstein, S.; MacCormack, B.; Malvar, H.; O’Brien, S.; Parnall, J.; Shamis, A.; Sharma, I.; Stokes, J.; Wenker, S.; and Zaman, A. 2020. MULTI-STAKEHOLDER MEDIA PROVENANCE MANAGEMENT TO COUNTER SYNTHETIC MEDIA RISKS IN NEWS PUBLISHING. *IBC* URL https://drive.google.com/file/d/11c41iTwq7z-nMMMPQLE5GZ6gJmc_IJ1/view.
- Bell, T.; Thimbleby, H.; Fellows, M.; Witten, I.; Koblitz, N.; and Powell, M. 2003. Explaining cryptographic systems. *Computers & Education* 40(3): 199–215.
- Bhuiyan, M. M.; Zhang, A. X.; Sehat, C. M.; and Mitra, T. 2020. Investigating differences in crowdsourced news credibility assessment: Raters, tasks, and expert criteria. *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW2): 1–26.
- Certificate Transparency Group. n.d. Certificate Transparency. <https://certificate.transparency.dev/>.
- Chase, M.; and Meiklejohn, S. 2016. Transparency Overlays and Applications. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 168–179.

- Chen, W.; Chiesa, A.; Dauterman, E.; and Ward, N. P. 2020. Reducing Participation Costs via Incremental Verification for Ledger Systems. *IACR Cryptol. ePrint Arch.* 2020: 1522.
- Coalition, C. 2017. Our goal: To understand the veracity, quality and credibility of online information. URL <https://credibilitycoalition.org/>.
- Coleman, K. 2021. URL https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation.
- Content Authenticity Initiative. 2022. Frequently asked questions. <https://contentauthenticity.org/faq>.
- Credibility Coalition and Nieman Foundation for Journalism. 2021. An Introduction to Schemas for Journalists. URL https://assets.ctfassets.net/tlowcqj4pb76/1edyyqNHwzMoJ9dzNsoLTh/29733395d8f3be00f020f89a614acfc8/An_Introduction_to_Schemas_for_Journalists.pdf.
- Ecker, U. K.; Lewandowsky, S.; Chang, E. P.; and Pillai, R. 2014. The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied* 20(4): 323.
- Editing TheGrayLady. 2016. Editing TheGrayLady. https://twitter.com/nyt_diff.
- Egelman, S.; Cranor, L. F.; and Hong, J. 2008. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 1065–1074. New York, NY, USA: Association for Computing Machinery. ISBN 9781605580111. doi: 10.1145/1357054.1357219. URL <https://doi-org.libproxy.clemson.edu/10.1145/1357054.1357219>.
- Fahl, S.; Dechand, S.; Perl, H.; Fischer, F.; Smrcek, J.; and Smith, M. 2014. Hey, NSA: Stay Away from my Market! Future Proofing App Markets against Powerful Attackers. In Ahn, G.-J.; Yung, M.; and Li, N., eds., *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, 1143–1155. Scottsdale, AZ, USA: ACM. doi: 10.1145/2660267.2660311. URL <https://doi.org/10.1145/2660267.2660311>.
- Felt, A. P.; Ainslie, A.; Reeder, R. W.; Consolvo, S.; Thyagaraja, S.; Bettes, A.; Harris, H.; and Grimes, J. 2015. Improving SSL Warnings: Comprehension and Adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 2893–2902. New York, NY, USA: Association for Computing Machinery. ISBN 9781450331456. doi:10.1145/2702123.2702442. URL <https://doi-org.libproxy.clemson.edu/10.1145/2702123.2702442>.
- Initiative, C. A. 2021. Addressing misinformation through digital content provenance. URL <https://contentauthenticity.org/>.
- Jacquette, R. 2018. We Stand Corrected: How The Times Handles Errors. *The New York Times* URL <https://www.nytimes.com/2018/06/07/reader-center/corrections-how-the-times-handles-errors.html>.
- Khodabakhsh, A.; Busch, C.; and Ramachandra, R. 2018. A taxonomy of audiovisual fake multimedia content creation technology. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 372–377. IEEE.
- Koren, S. 2019. Introducing the News Provenance Project. URL <https://open.nytimes.com/introducing-the-news-provenance-project-723dbaf07c44>.
- Kumar, S.; West, R.; and Leskovec, J. 2016. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th international conference on World Wide Web*, 591–602.
- Labs, S. 2022. Starling Lab for Data Integrity. URL <https://www.starlinglab.org/journalism/>.
- Laurie, B. 2014. Certificate transparency. *Commun. ACM* 57(10): 40–46.
- Laurie, B.; Langley, A.; and Käsper, E. 2013. Certificate Transparency. *RFC* 6962: 1–27.
- Lewandowsky, S.; Cook, J.; Ecker, U.; Albarracin, D.; Amazeen, M.; Kendou, P.; Lombardi, D.; Newman, E.; Pennycook, G.; Porter, E.; et al. 2020. *The debunking handbook 2020*. null. doi:10.17910/b7.1182.
- Li, B.; Li, F.; Ma, Z.; and Wu, Q. 2020. Exploring the security of certificate transparency in the wild. In *International Conference on Applied Cryptography and Network Security*, 453–470. Springer.
- Library, T. D. P. 2019. Link Rot: The Web is Decaying. URL <https://arweave.medium.com/link-rot-the-web-is-decaying-cc7d1c5ad48b>.
- McBride, K. 2020. 'unarmed black man' doesn't mean what you think it means. URL <https://www.npr.org/sections/publiceditor/2020/05/21/859498255/unarmed-black-man-doesnt-mean-what-you-think-it-means>.
- Meiklejohn, S.; Kalinnikov, P.; Lin, C. S.; Hutchinson, M.; Belvin, G.; Raykova, M.; and Cutter, A. 2020. Think Global, Act Local: Gossip and Client Audits in Verifiable Data Structures. *CoRR* abs/2011.04551.
- Melara, M. S.; Blankstein, A.; Bonneau, J.; Felten, E. W.; and Freedman, M. J. 2015. CONIKS: Bringing Key Transparency to End Users. In *USENIX Security Symposium*, 383–398. Washington, D.C.: USENIX Association.
- Moran, R. E. 2021. Subscribing to transparency: Trust-building within virtual newsrooms on slack. *Journalism Practice* 15(10): 1580–1596.
- Ngo, T.; and Krämer, N. 2021. It's Just a Recipe?—Comparing Expert and Lay User Understanding of Algorithmic Systems. *Technology, Mind, and Behavior* 2(4). doi:10.1037/tmb0000045. URL <https://tmb.apaopen.org/pub/9oppvmbi>. <https://tmb.apaopen.org/pub/9oppvmbi>.
- Nikitin, K.; Kokoris-Kogias, E.; Jovanovic, P.; Gailly, N.; Gasser, L.; Khoffi, I.; Cappos, J.; and Ford, B. 2017. CHAINIAC: Proactive Software-Update Transparency via Collectively Signed Skipchains and Verified Builds. In *USENIX Security Symposium*, 1271–1287. Vancouver, BC: USENIX Association.

- Ortutay, B. 2021. Twitter revamps warning labels for false and misleading tweets. *The Chicago Tribune* URL <https://www.chicagotribune.com/business/ct-biz-twitter-warning-misleading-labels-20210701-isdhxr4svfcbkh7henzyreusa-story.html>.
- Owen, Q. 2021. ICE to stop using the term 'illegal alien' referring to immigrants. URL <https://abcnews.go.com/Politics/ice-stop-term-illegal-alien-referring-immigrants/story?id=77165043>.
- Pennycook, G.; Bear, A.; Collins, E. T.; and Rand, D. G. 2020. The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science* 66(11): 4944–4957.
- Pennycook, G.; and Rand, D. G. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188: 39–50.
- Press, A. n.d. Corrections/Correctives. *Associated Press* URL <https://www.ap.org/about/news-values-and-principles/telling-the-story/corrections-correctives>.
- Project, T. A. 2021. What is Arweave? URL <https://arwiki.wiki/#/en/main>.
- ProPublica. n.d. Corrections. *Associated Press* URL <https://www.propublica.org/corrections/>.
- Roth, Y.; and Pickles, N. 2020. Updating our approach to misleading information. URL https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.
- Ruggiero, T. E. 2000. Uses and Gratifications Theory in the 21st Century. *Mass Communication and Society* 3(1): 3–37. ISSN 1520-5436, 1532-7825. doi:10.1207/S15327825MCS0301_02.
- Ryan, M. D. 2014. Enhanced Certificate Transparency and End-to-End Encrypted Mail. In *NDSS*. San Diego, CA: The Internet Society.
- SciLine. 2021. URL <https://sciline.org/social-sciences/misinformation/>.
- Sterrett, D.; Malato, D.; Benz, J.; Kantor, L.; Tompson, T.; Rosenstiel, T.; Sonderman, J.; and Loker, K. 2019. Who shared it?: Deciding what news to trust on social media. *Digital journalism* 7(6): 783–801.
- Stomber, J.; Gamage, D.; Skeet, B.; and Zhang, A. X. 2021. Towards a Unified Framework for the UX Design of News Credibility Tools. In *Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media*.
- Sumpter, M.; and Neal, T. 2021. User Perceptions of Article Credibility Warnings: Towards Understanding the Influence of Journalists and AI Agents. In *Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media*.
- Swift, J.; and Roscoe, T. 1841. *The Works of Jonathan Swift ...: Containing Interesting and Valuable Papers, Not Hitherto Published ... With Memoir of the Author*. Number v. 1 in *The Works of Jonathan Swift ...: Containing Interesting and Valuable Papers*. H. G. Bohn. URL <https://books.google.com/books?id=88REAAAAYAAJ>.
- Takahashi, D. 2021. Project Starling uses technology to preserve authenticity of Capitol riot images. URL <https://venturebeat.com/2021/02/10/project-starling-uses-technology-to-preserve-authenticity-of-capitol-riot-images/>.
- Tiedens, L. Z.; and Linton, S. 2001. Judgment under emotional certainty and uncertainty: the effects of specific emotions on information processing. *Journal of personality and social psychology* 81(6): 973.
- Tyagi, N.; Fisch, B.; Bonneau, J.; and Tessaro, S. 2021. Client-Auditable Verifiable Registries. *Cryptology ePrint Archive, Report 2021/627*. <https://ia.cr/2021/627>.
- Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science* 359(6380): 1146–1151. doi:10.1126/science.aap9559. URL <https://www.science.org/doi/abs/10.1126/science.aap9559>.
- Walter, N.; Cohen, J.; Holbert, R. L.; and Morag, Y. 2020. Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication* 37(3): 350–375. doi:10.1080/10584609.2019.1668894. URL <https://doi.org/10.1080/10584609.2019.1668894>.
- Wischnewski, M.; Krämer, N.; and McNamara, D. S. 2021. The Role of Emotions and Identity-Protection Cognition When Processing (Mis)Information. *Technology, Mind, and Behavior* 2(1). doi:10.1037/tmb0000029. URL <https://tmb.apaopen.org/pub/osng2517>. <https://tmb.apaopen.org/pub/osng2517>.
- Wu, J.; and Zappala, D. 2018. When is a tree really a truck? exploring mental models of encryption. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, 395–409.
- Xiao, Y.; Zhang, N.; Lou, W.; and Hou, Y. T. 2020. A survey of distributed consensus protocols for blockchain networks. *IEEE Communications Surveys & Tutorials* 22(2): 1432–1465.
- Young, L. E.; Sidnam-Mauch, E.; Twyman, M.; Wang, L.; Xu, J. J.; Sargent, M.; Valente, T. W.; Ferrara, E.; Fulk, J.; and Monge, P. 2021. Disrupting the COVID-19 misinfodemic with network interventions: network solutions for network problems. *American journal of public health* 111(3): 514–519.
- Yu, J.; Ryan, M.; and Cremers, C. 2015. How to detect unauthorised usage of a key. *IACR Cryptol. ePrint Arch.* 2015: 486.