Defending Messaging Apps Against Spyware Using Data Diodes

Peter Story
Clark University
Worcester, Massachusetts, USA
PeStory@clarku.edu

Abstract

Messaging apps like Signal, WhatsApp, and iMessage use end-to-end encryption to protect users' messages against passive government surveillance. However, these apps offer limited protection against targeted attacks: if a user's device is compromised by spyware, the attacker will gain access to all the user's messages. Furthermore, messaging apps are often used as an attack vector for spyware. First, we describe the threat model of encrypted messaging apps, and current defenses against spyware. Then, we propose a novel approach for hardening messaging apps against spyware using one-way network device, known as a data diode. We anticipate this technology will be valuable to journalists, politicians, and other targets of spyware. Finally, we enumerate obstacles to deploying our proposed defenses.

Keywords

encrypted messaging, spyware, data diodes, threat modeling

1 Introduction

Governments around the world have extensive intelligence gathering capabilities. Sometimes data access follows a well-defined process, such as legal requests for data as part of court proceedings. Other times data access is opaque, as when intelligence agencies covertly monitor network traffic. Encryption offers a defense of users' privacy, even against powerful entities like governments. For example, messaging apps like Signal, WhatsApp, and iMessage use end-to-end encryption (E2EE) to protect the confidentiality of users' messages. E2EE ensures that only the sender and recipient have the technical means to read the contents of messages, and that service providers only have access to metadata about messages. E2EE is a strong defense against passive government surveillance, and against legal requests for data. However, E2EE offers limited protection against targeted attacks: if a user's device is compromised by spyware, the attacker will gain access to all the user's messages. Spyware like the NSO Group's Pegasus have been used to target journalists, human rights workers, and high profile public figures [2–5, 16, 18, 28, 29]. Furthermore, messaging apps themselves are often used to deliver spyware payloads [3, 16, 19, 31].

In our Related Work (§ 2), we share more details about E2EE and spyware. Next, we describe the threat model of encrypted messaging apps, and current defenses against spyware (§ 3). Then, we propose a novel approach for hardening messaging apps against spyware (§ 4). In particular, we show how a one-way network device, known as a data diode, can protect the confidentiality of messages. We anticipate this technology will be valuable to journalists, politicians, and other targets of spyware. Finally, we enumerate obstacles to deploying our proposed defenses.

Free and Open Communications on the Internet 2026(1), 1–5

2 Related Work

First, we give a high-level overview of the protocols used by encrypted messaging apps (§ 2.1). Next, we describe how spyware can undermine the security of messaging apps (§ 2.2). Finally, we explain how one-way networking devices can defend against spyware (§ 2.3).

2.1 Encrypted Messaging Apps

Secure messaging apps use encryption to protect the confidentiality of users' messages. End-to-end encryption (E2EE) protocols ensure that messages cannot be read without decryption keys, and that decryption keys are only held by conversation participants. The Signal messaging app uses the Signal Protocol, and this protocol has been adopted by other messaging apps, including WhatsApp and Google Messages. When two parties communicate, the protocol starts with a key agreement protocol for mutual authentication, then uses a ratchet algorithm to update shared secret keys as the conversation progresses [32]. Through these mechanisms, the protocol provides forward secrecy and cryptographic deniability. The protocol was recently updated for quantum-resistance, replacing the X3DH key agreement protocol with the PQXDH protocol, and adding a third ratchet to the original Double Ratchet algorithm [22]. Apple's iMessage uses its own protocol [7], which was recently updated for quantum-resistance and forward secrecy [9, 20]. Although iMessage is not open source, iMessage's quantum-resistant protocol (PQ3) was formally verified [11, 33] and its source code was audited by a third party [9].

2.2 Spyware

Governments are capable of intercepting internet traffic at scale, but the E2EE protocols used by encrypted messaging apps render intercepted messages unreadable. However, intelligence agencies can use *spyware* to access a target's messages after they have been decrypted on the target's device. Spyware uses software exploits to take over a target's device [14]. These exploits are often sent through encrypted messaging apps like iMessage and WhatsApp, and have taken the form of malicious attachments, links, and calls [3, 16, 19, 31]. Exploits can even take over a device without user-interaction (e.g., not opening an attachment). After compromising a device, spyware can steal messages, track the user's location, and enable the device's microphone and camera. The NSO Group's Pegasus spyware has been involved in many highly publicized hacks [12], but other vendors offer similar products (e.g., Cytrox's

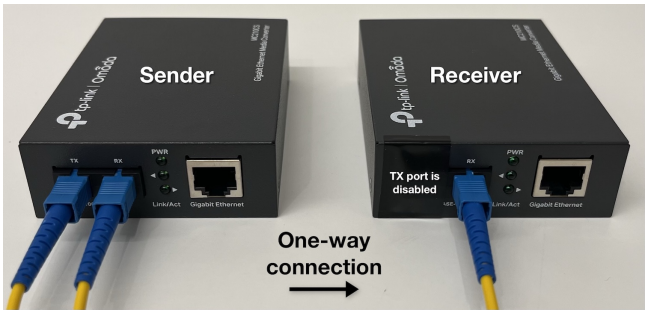


Figure 1: A data diode built with off-the-shelf network hardware. The Ethernet fiber media convertor on the left sends data to the convertor on the right. The convertor on the right physically cannot transfer data in the reverse direction, since its transmit port is taped over.

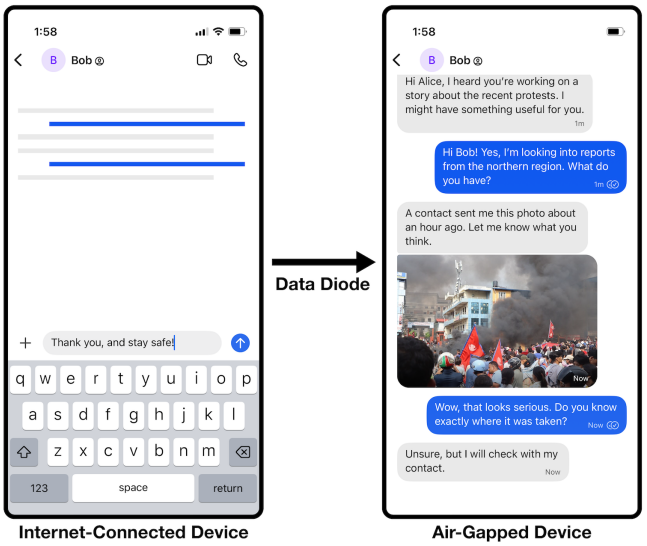


Figure 2: The internet-connected device on the left sends data through a data diode to the air-gapped device on the right. The air-gapped device has no wireless functionality, and the data diode physically prevents it from sending data to the internet-connected device. Incoming messages can only be decrypted on the air-gapped device, since only the air-gapped device holds the private key. Although the air-gapped device can be compromised by malicious attachments, hackers have no way to exfiltrate data. Protest photo credit: [36].

Predator [18]), and some nation states develop in-house capabilities. Spyware vendors claim their products facilitate legitimate crime-fighting operations, but spyware is routinely used against journalists, opposition politicians, and other members of civil society [2–5, 16, 18, 28, 29]. Spyware is not scalable as a form of mass-surveillance, since the more an exploit is used, the more likely it is to be discovered and patched. Nevertheless, spyware enables close observation of particular targets, and may result in a chilling effect for the wider population.

2.3 Data Diodes

Software-based protections, like sandboxing and malware scanning, can increase the difficulty of compromising a device, but are insufficient to stop spyware. In software, there is an inherent asymmetry between offense and defense: exploits are costly to discover, but the complexity of modern software ensures there is a steady supply for those who invest the resources [1, 6, 14]. Intelligence agencies are aware of this challenge, so they employ physical countermeasures to protect their own data. For example, cross-domain solutions (CDSs) are used to limit exchange of data between different security classification levels [10]. If data is physically prevented from leaving a device, there is no way for a remote attacker to exfiltrate data. This can be accomplished using one-way networking devices, known as *data diodes* [13, 34]. A data diode is a network device which is only physically capable of transferring data in one direction. A network firewall is not a data diode, since a software vulnerability could compromise the firewall and modify its rules. In contrast, a data diode should include hardware that only allows data to be transferred in one direction. For example, using a modified fiber optic link [10, 35, 37], as shown in Figure 1. These physical properties allow for guarantees about the direction of data flow [13, 34].

3 Encrypted Messaging Threat Model

In our threat model, we focus on nation state attackers who are operating remotely, such as foreign governments. Our goal is to protect the confidentiality of messages in encrypted messaging apps. The E2EE protocols used by encrypted messaging apps prevent attackers from reading any messages they intercept. However, spyware poses a significant threat: if a user's messaging client or smartphone is compromised, the attacker can exfiltrate data after it has been decrypted. Spyware can compromise a device by sending malicious attachments, links, and calls to a target, and can even take over a device without user-interaction. Since spyware must be targeted to individuals, messaging apps are a common attack vector [3, 16, 19], though email and browser-based attacks are also possible [18, 31].

Currently, protecting a user's smartphone from compromise is a prerequisite to preserving the confidentiality of messages. Smartphone operating systems offer settings to reduce their attack surface. For example, iOS and Android offer Lockdown Mode [8, 17] and Advanced Protection [21, 23], respectively. Both features limit JavaScript acceleration and disable 2G cellular networks, among other protections. Lockdown Mode trades reduced functionality for a smaller attack surface, disabling many attachment types in iMessage (e.g., PDFs), and blocking messages from new numbers [8, 27]. This loss of functionality might be unacceptable for some users. Furthermore, although these settings increase the challenge for spyware developers, they cannot completely stop spyware. We propose defenses that reduce the attack surface of messaging apps without disabling functionality, and which offer protection even if the smartphone is compromised.

4 Defending Against Spyware With Data Diodes

Encrypted messaging apps can be defended against spyware by splitting functionality across two devices. Figure 2 shows a user interface mockup. The device on the left is an internet-connected smartphone, which is used to send and receive messages — though the received messages cannot be decrypted on this device. The device on the right is an air-gapped device, which decrypts and displays messages. The devices are connected by a data diode, which allows data to enter but not leave the air-gapped device.

This architecture has two major benefits. First, potentially malicious messages can be opened safely. Incoming messages and attachments are only decrypted on the air-gapped device, which is physically prevented from compromising the internet-connected device. This reduces the attack surface of the messaging client on the internet-connected device, making it less likely to be compromised by spyware. Second, our Spyware Defense Protocol (§ 4.1) protects the confidentiality of messages, even if a user's devices are compromised by spyware. Incoming messages are encrypted using a public key, and the corresponding private key only exists on the air-gapped device. If the internet-connected device is compromised, incoming messages cannot be decrypted by the attacker. If the air-gapped device is compromised (e.g., by a malicious attachment), the data diode prevents exfiltrating data.

4.1 Spyware Defense Protocol

Next, we explain how this feature could be integrated into encrypted messaging apps, using Signal as an example. In our example, Alice is using the Spyware Defense feature, and Bob may or may not have the feature enabled for himself.

Initial Setup: Alice generates a public private key pair on her air-gapped device. The public key is displayed with a QR code. On her internet-connected device, Alice enables the Spyware Defense feature in Signal, and scans the public key's QR code to register it server-side as a Spyware Defense public key.

Incoming Messages: Bob composes a message for Alice. Prior to the standard Signal Protocol, Bob's client encrypts the message with Alice's Spyware Defense public key. Alice's client receives the incoming message using the standard Signal Protocol. At this point, the message is still encrypted with her Spyware Defense public key, and the corresponding private key only resides on her air-gapped device. Alice's client sends the message to her air-gapped device, which automatically decrypts and displays it.

Outgoing Messages: Alice composes a message for Bob. If Bob has enabled the Spyware Defense feature for himself, Alice's client will encrypt the message with Bob's Spyware Defense public key before sending it using the standard Signal Protocol. To protect the confidentiality of the outgoing message, Alice's client automatically deletes the message after sending it to her air-gapped device for display in the conversation thread.

4.2 Spyware Defense Threat Model

Our proposed feature offers a strong defense against spyware. The internet-connected messaging app's attack surface is smaller, since incoming messages are only decrypted on the air-gapped device. However, Alice's internet-connected device could be compromised through other means (e.g., visiting a malicious website). Encryption offers a second layer of defense: even if Alice's internet-connected device is infected, her private key only resides on the air-gapped device, preventing hackers from decrypting incoming messages. Despite these defenses, there are ways to steal conversation plaintext. First, if Alice's internet-connected device is compromised, an adversary could see who she is communicating with and read her outgoing messages. However, the Spyware Defense feature gives spyware a limited window to capture outgoing messages, since outgoing messages are deleted from a user's internet-connected device after they have been transferred to their air-gapped device. If Bob's internet-connected device is compromised, the attacker would gain access to Bob's outgoing messages, or the entire conversation if Bob is not using the Spyware Defense feature. The Spyware Defense feature multiplies the number of devices an attacker must compromise, offering the possibility of detecting attacks using honeypots. Second, if Alice's internet-connected device was infected prior to initial setup, an adversary could register their own public key. Third, if Alice's air-gapped device is compromised, perhaps by a malicious attachment, Alice could be tricked into exfiltrating data by scanning a QR code [15]. Finally, side-channel attacks could be used to exfiltrate data from her air-gapped device [24–26].

To summarize, although the Spyware Defense feature reduces the attack surface of users' messaging apps, users should still defend against other attack vectors: compromise of the internet-connected device could reveal outgoing messages and contacts' identities. Ideally, high-risk users would have a dedicated device for secure messaging, and would avoid web browsing and other risky activities on that device. Spyware often lacks persistence, so periodic reboots would further limit the window to capture outgoing messages [3, 19, 30]. For the strongest protection, both Alice and Bob would use the Spyware Defense feature.

4.3 Obstacles to Adoption

We foresee two main obstacles to adoption. First, users who want protection against spyware must use additional hardware: a data diode and an air-gapped device. Figure 3 shows our prototype, which uses a data diode built from off-the-shelf network hardware connected to an air-gapped laptop. Users may be hesitant to adopt bulky hardware, and configuring the hardware requires technical expertise. Ideally, this technology would be miniaturized into a smartphone-sized portable device. Second, the messaging app must be modified to support Spyware Defense public keys. Although only users who want protection from spyware require additional hardware, anyone they communicate with should use a client which supports encrypting outgoing messages using an optional Spyware Defense public key. Deploying this feature to all clients would require buy-in from the messaging app developer.

Without widespread support for our Spyware Defense feature, interested users can still run a modified client which offers a subset of our proposed protections. Since Signal is open source, implementing these protections wouldn't require resources from the Signal Foundation. Upon receiving an incoming message through the standard Signal Protocol, Alice's client could immediately encrypt the message using her Spyware Defense public key, ensuring messages are only displayed on her air-gapped device. This approach defends against spyware by reducing the messaging app's attack surface.

However, if Alice's device were compromised through another vector, incoming messages would be readable as they arrive. Thus, the full Spyware Defense protocol (§ 4.1) offers stronger protections.

5 Conclusion

Encrypted messaging apps defend against passive government surveillance, but offer little protection against spyware. In fact, messaging apps often serve as an attack vector for spyware [3, 16, 19, 31]. Spyware is a serious threat to journalists, politicians, and other members of civil society. In this paper, we describe a novel approach for hardening messaging apps against spyware: splitting functionality across two devices connected by a data diode (§ 4). Our approach shrinks the attack surface of messaging apps, reducing their potency as an attack vector. Also, even if a user's device is compromised through other means, our approach limits spyware's ability to exfiltrate messages. As shown in Figure 3, our lab is developing a prototype of this technology using off-the-shelf components and open source software. From a usability perspective, it would be preferable to miniaturize the hardware, perhaps into a smartphone-sized form factor. Long-term, messaging platforms could implement these spyware defenses in their apps, and could sell companion hardware to security-conscious users: a commercialized, miniaturized product would bring these defenses to the most users. If companies pursue this option, we recommend that they maintain a physically verifiable air gap and publish open-source hardware designs.

References

[1] Lillian Ablon and Andy Bogart. 2017. *Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits.* RAND Corporation. https://doi.org/10.7249/RR1751 http://www.rand.org/pubs/research_reports/RR1751.html.

[2] Azam Ahmed and Nicole Perlroth. 2017. Using Texts as Lures, Government Spyware Targets Mexican Journalists and Their Families. *The New York Times* (June 2017). https://www.nytimes.com/2017/06/19/world/americas/mexico-spyware-anticrime.html.

[3] Amnesty International. 2021. Forensic Methodology Report: How to Catch NSO Group's Pegasus. https://www.amnesty.org/en/latest/research/2021/07/forensic-methodology-report-how-to-catch-nso-groups-pegasus/.

[4] Amnesty International. 2022. Amnesty International Verifies Use of Pegasus Spyware against Journalists in El Salvador. https://www.amnesty.org/en/latest/news/2022/01/el-salvador-pegasus-spyware-surveillance-journalists/.

[5] Amnesty International. 2023. Pegasus Discovered on Journalist's Phone in Dominican Republic. https://www.amnesty.org/en/latest/news/2023/05/dominican-republic-pegasus-spyware-journalists-phone/.

[6] Ross Anderson. 2001. Why Information Security Is Hard - an Economic Perspective. In *Seventeenth Annual Computer Security Applications Conference*. IEEE Comput. Soc, New Orleans, LA, USA, 358–365. https://doi.org/10.1109/ACSAC.2001.991552 http://ieeexplore.ieee.org/document/991552/.

[7] Apple. 2024. *Apple Platform Security.* Technical Report. https://help.apple.com/pdf/security/en_US/apple-platform-security-guide.pdf.

[8] Apple. 2025. About Lockdown Mode. https://support.apple.com/en-us/105120

[9] Apple Security Engineering and Architecture. 2024. iMessage with PQ3: The New State of the Art in Quantum-Secure Messaging at Scale. https://security.apple.com/blog/imessage-pq3/.

[10] Ross D. Arnold. 2016. *Strategies for Transporting Data Between Classified and Unclassified Networks.* Technical Report. Defense Technical Information Center, Fort Belvoir, VA. https://doi.org/10.21236/AD1005160 https://apps.dtic.mil/sti/citations/AD1005160.

[11] David Basin, Felix Linker, and Ralf Sasse. 2024. *A Formal Analysis of the iMessage PQ3 Messaging Protocol.* Technical Report.

[12] Ronen Bergman and Mark Mazzetti. 2022. The Battle for the World's Most Powerful Cyberweapon. *The New York Times* (Jan. 2022). https://www.nytimes.com/2022/01/28/magazine/nso-group-israel-spyware.html.

[13] Fred Cohen. 1988. Designing Provably Correct Information Networks with Digital Diodes. *Computers & Security* 7, 3 (June 1988), 279–286. https://doi.org/10.1016/0167-4048(88)90034-X https://linkinghub.elsevier.com/retrieve/pii/016740488890034X.

[14] Mailyn Fidler. 2024. Zero Progress on Zero Days: How the Last Ten Years Created the Modern Spyware Market. *Nebraska Law Review* 103 (May 2024).

[15] Freedom of the Press Foundation. 2017. Security Advisory: Do not scan QR codes submitted through SecureDrop with connected devices. https://securedrop.org/news/security-advisory-do-not-scan-qr-codes-submitted-through-securedrop-connected-devices/

[16] Dan Goodin. 2020. Report: Bezos' Phone Uploaded GBs of Personal Data after Getting Saudi Prince's WhatsApp Message. https://arstechnica.com/information-technology/2020/01/report-bezos-phone-uploaded-gbs-of-personal-data-after-getting-saudi-princes-whatsapp-message/.

[17] Dan Goodin. 2022. Why Lockdown Mode from Apple Is One of the Coolest Security Ideas Ever. https://arstechnica.com/information-technology/2022/07/introducing-lockdown-from-apple-the-coolest-defense-youll-probably-never-use/.

[18] Dan Goodin. 2023. 3 iOS 0-Days, a Cellular Network Compromise, and HTTP Used to Infect an iPhone. https://arstechnica.com/security/2023/09/how-the-iphone-of-a-presidential-candidate-in-egypt-got-hacked-for-the-2nd-time/.

[19] Dan Goodin. 2023. 4-Year Campaign Backdoored iPhones Using Possibly the Most Advanced Exploit Ever. https://arstechnica.com/security/2023/12/exploit-used-in-mass-iphone-infection-campaign-targeted-secret-hardware-feature/.

[20] Dan Goodin. 2024. iMessage Gets a Major Makeover That Puts It on Equal Footing with Signal. https://arstechnica.com/security/2024/02/imessage-gets-a-major-makeover-that-puts-it-on-equal-footing-with-signal/.

[21] Dan Goodin. 2025. Google Introduces Advanced Protection Mode for Its Most At-Risk Android Users - Ars Technica. https://arstechnica.com/security/2025/05/google-introduces-advanced-protection-mode-for-its-most-at-risk-android-users/.

[22] Dan Goodin. 2025. Why Signal's Post-Quantum Makeover Is an Amazing Engineering Achievement. https://arstechnica.com/security/2025/10/why-signals-post-quantum-makeover-is-an-amazing-engineering-achievement/.

[23] Google. 2025. Improve device security with Advanced Protection. https://support.google.com/android/answer/16339980?hl=en

[24] Mordechai Guri and Yuval Elovici. 2018. Bridgeware: The Air-Gap Malware. *Commun. ACM* 61, 4 (March 2018), 74–82. https://doi.org/10.1145/3177230 https://dl.acm.org/doi/10.1145/3177230.

[25] Mordechai Guri, Yosef Solewicz, and Yuval Elovici. 2018. MOSQUITO: Covert Ultrasonic Transmissions Between Two Air-Gapped Computers Using Speaker-to-Speaker Communication. In *2018 IEEE Conference on Dependable and Secure Computing (DSC)*. IEEE, Kaohsiung, Taiwan, 1–8. https://doi.org/10.1109/DESEC.2018.8625124 https://ieeexplore.ieee.org/document/8625124/.

[26] Michael Hanspach and Michael Goetz. 2014. On Covert Acoustical Mesh Networks in Air. *Journal of Communications* (2014). https://doi.org/10.48550/ARXIV.1406.1213 https://arxiv.org/abs/1406.1213.

[27] hobbes444. 2024. 2 months of lockdown mode. https://www.reddit.com/r/ios/comments/1gkugen/2_months_of_lockdown_mode/

[28] Stephanie Kirchgaessner. 2023. Experts Warn of New Spyware Threat Targeting Journalists and Political Figures. *The Guardian* (April 2023). https://www.theguardian.com/technology/2023/apr/11/canadian-security-experts-warn-over-spyware-threat-to-rival-pegasus-citizen-lab.

[29] William R Marczak, John Scott-Railton, Vern Paxson, and Morgan Marquis-Boire. 2014. When Governments Hack Opponents: A Look at Actors and Technology. *23rd USENIX Security Symposium (USENIX Security 14)* (Aug. 2014).

[30] National Security Agency. 2020. Mobile Device Best Practices. https://www.documentcloud.org/documents/21018353-nsa-mobile-device-best-practices.

[31] David Pegg and Sam Cutler. 2021. What Is Pegasus Spyware and How Does It Hack Phones? *The Guardian* (July 2021). https://www.theguardian.com/news/2021/jul/18/what-is-pegasus-spyware-and-how-does-it-hack-phones.

[32] Signal Technology Foundation. 2025. Technical Information. https://signal.org/docs/.

[33] Douglas Stebila. 2024. *Security Analysis of the iMessage PQ3 Protocol.* Technical Report.

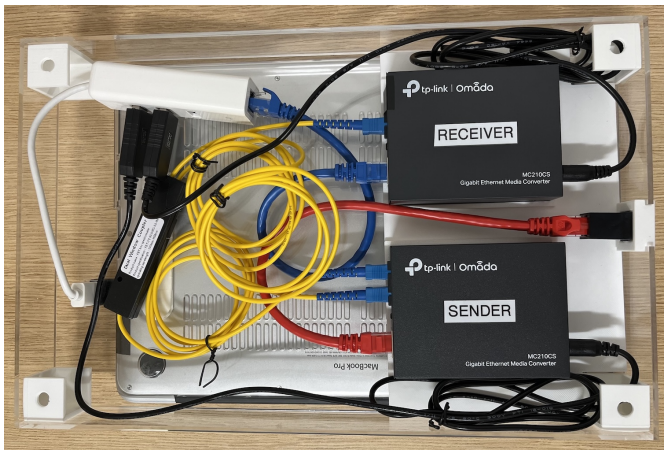
[34] Malcolm W Stevens. 1999. *An Implication of an Optical Data Diode.* Technical Report DSTO-TR-0785. Information Technology Division Electronics and Surveillance Research Laboratory.

[35] Peter Story. 2023. Building an Affordable Data Diode to Protect Journalists. In *Workshop on Privacy Engineering in Practice (PEP '23)*. https://peterstory.me/publications/story_pep_2023.pdf.

[36] Himal Subedi. 2025. 2025 Nepalese Gen Z protesters in Front of Bharatpur Mahanagarpalika Office. https://commons.wikimedia.org/wiki/File:2025_Nepalese_Gen_Z_protesters_infront_of_Bharatpur_mahanagarpalika_office.jpg

[37] Vrolijk. 2025. Get started with Data Diodes. https://github.com/Vrolijk/OSDD

(a) The internet-connected smartphone is connected to the air-gapped laptop through a data diode. Messages received on the smartphone can only be decrypted on the laptop. Data physically cannot leave the laptop, since its wireless radios, speakers, and USB ports are disabled: the laptop can only receive data through the data diode. Each device shows Signal user interface mockups.



(b) The enclosure underneath the air-gapped laptop contains a data diode. The data diode consists of two Ethernet fiber optic media converters, one of which has its transmit port taped over. This configuration physically enforces the direction of network traffic, preventing the laptop from sending data to the smartphone.

Figure 3: A portable data diode assembled from off-the-shelf hardware.